Feedforward Carrier Recovery for Coherent Optical Communications

Ezra Ip and Joseph M. Kahn, *Fellow, IEEE*

***Abstract*—We study a carrier-synchronization scheme for coherent optical communications that uses a feedforward architecture that can be implemented in digital hardware without a phase-locked loop. We derive the equations for maximum *a posteriori* joint detection of the transmitted symbols and the carrier phase. The result is a multidimensional optimization problem that we approximate with a two-stage iterative algorithm: The first stage is a symbol-by-symbol soft detector of the carrier phase, and the second stage is a hard-decision phase estimator that uses prior and subsequent soft-phase decisions to obtain a minimum mean-square-error phase estimate by exploiting the temporal correlation in the phase-noise process. The received symbols are then derotated by the hard-decision phase estimates, and maximum-likelihood sequence detection of the symbols follows. As each component in the carrier-recovery unit can be separately optimized, the resulting system is highly flexible. We show that the optimum hard-decision phase estimator is a linear filter whose impulse response consists of a causal and an anticausal exponential sequence, which we can truncate and implement as an finite-impulse-response filter. We derive equations for the phase-error variance and the system bit-error ratio (BER). Our results show that 4, 8, and 16 quadrature-amplitude-modulation (QAM) transmissions at 1 dB above sensitivity for BER $= 10^{-3}$ is possible with laser beat linewidths of $\Delta\nu T_{\mathrm{b}} = 1.3 \times 10^{-4}$, 1.3×10^{-4}, and 1.5×10^{-5} when a decision-directed soft-decision phase estimator is employed.**

***Index Terms*—Carrier synchronization, coherent detection, digital signal processing, feedforward, optical communications, phase locked loop.**

I. Introduction

PHASE NOISE is a major impairment in coherent optical communications. The traditional method of demodulating coherent optical signals is to use an optical or electrical phase-locked loop (PLL) that synchronizes the frequency and phase of the local oscillator (LO) with the transmitter laser. Owing to high data throughput, PLLs have generally been implemented with analog components. The performance of PLL-based receivers for the detection of optical phase-shift keying (PSK) and quadrature-amplitude modulation (QAM) has been studied in [1] and [2]. PLLs are sensitive to propagation delay inside the loop. For 10-Gb/s transmission, it has been shown that delays greater than a few tens of nanoseconds will lead to loop instability. Even if careful circuit design reduces propagation delay to no more than a few symbol periods, coherent detection with a PLL still has stringent linewidth requirements. In [1] and [2], it was found that the maximum tolerable linewidth-to-bit-rate ratio $(\Delta\nu T_{\mathrm{b}})$ for 4-, 8-, and 16-QAMs are 3.5×10^{-5}, 1.8×10^{-5}, and 1.3×10^{-6}, respectively, if a bit-error ratio (BER) of 10^{-9} is to be achieved with less than 0.5-dB power penalty. Since current-generation external-cavity tunable lasers (ECLs) have linewidths on the order of tens to hundreds of kilohertz [3], 8-QAM transmission at 10 Gb/s is only barely feasible, while 16-QAM transmission at 10 Gb/s is beyond the scope of current lasers.

Advances in high-speed very large-scale integration (VLSI) technology promise to change the paradigm of coherent optical receivers. By converting the in-phase (I) and quadrature (Q) outputs of a homodyne receiver to the digital domain using high-speed analog-to-digital converters (ADCs), digital-signal-processing (DSP) algorithms can be used to compensate channel impairments. Provided that the ADC sampling rate satisfies Nyquist's criterion, the digitized signals contain all the information in the received electric field. DSP compensation of linear impairments, such as chromatic dispersion (CD) [4]–[6] and polarization-mode dispersion (PMD) [7], [8], is possible. In addition, a digital coherent receiver also allows new techniques for carrier synchronization. PLLs are no longer required, as frequency mismatch between the transmitter and the LO laser only results in a constant rotation of the received constellation, which can be tracked by DSP and compensated in a feedforward architecture. Recent experiments have shown that feedforward carrier-recovery schemes are more tolerant to laser phase noise than PLL-based receivers. Coherent detection of optical binary PSK that yielded a Q factor of 8.5 dB at a laser linewidth of $\Delta\nu T_{\mathrm{b}} = 0.032$ was reported in [9], while coherent detection of 4-PSK that yielded a 1.5-dB power penalty at a BER of 10^{-9} with $\Delta\nu T_{\mathrm{b}} = 10^{-4}$ was reported in [10]. Feedforward carrier recovery for optical 8-PSK was studied in [11].

It is noted that the above experimental results were obtained for constant-amplitude M-PSK constellations. While various carrier-recovery schemes have been proposed for QAM in digital subscriber lines [12]–[14], the proposed algorithms take thousands of symbols to estimate the carrier phase, which is too slow for the linewidth–bit-rate ratios typically encountered in optical transmission. To our knowledge, no comprehensive study has been undertaken to determine the optimum receiver structures for optical QAM. Even for optical PSK, no theoretical analysis of system performance has been carried out.

Manuscript received November 27, 2006; revised May 10, 2007. This work was supported by the Defense Advanced Research Projects Agency under TACOTA Program Prime Contract W911-QX-O6-C-0101 via a Subcontract from CeLight, Inc.

The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9515 USA (e-mail: wavelet@stanford.edu; jmk@ee.stanford.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2007.902118

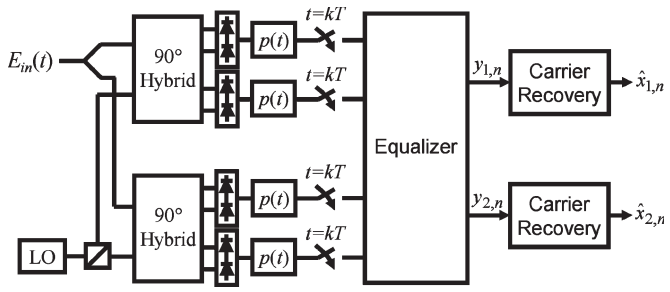

Fig. 1. Polarization-diversity homodyne receiver.

The objective of this paper is to provide a theoretical basis for feedforward carrier recovery in coherent optical communications. We will approach the problem in a systematic manner, starting with a derivation of the maximum *a posteriori* probability (MAP) detector. We will propose a two-stage phase estimator that is computationally simple enough to be implemented at gigabits-per-second data rates.

Our paper is organized as follows. In Section II, we derive formulas for MAP detection of the transmitted symbols and the carrier phases. We will show that MAP detection is a joint optimization problem that is numerically too complicated to solve. In Section III, we propose an approximate procedure consisting of two iterative stages: A soft-decision phase-estimator stage, followed by a hard-decision phase-estimator and symbol detector. We will show that the hard-decision phase estimator can be implemented as a finite-impulse-response (FIR) filter. We will analyze system performance in terms of the phase-error standard deviation and the system BER. We will also investigate the impact of nonzero frequency offset between the signal and LO lasers. Simulation results will be presented in Section IV, and implementation issues will be discussed in Section V.

II. MAP Symbol/Phase Detector

A. Receiver Model

In this paper, we shall assume the use of a polarization diversity homodyne receiver shown in Fig. 1. The received optical signal $E_{\rm in}(t)$ is combined with a LO laser using two 90° hybrids and a polarization beam splitter [4], [15]. The outputs of the balanced photodetectors are the baseband electrical signals corresponding to the I and Q signals in the two reference polarizations of the LO laser. These signals are low-pass filtered and sampled using high-speed ADC. The I and Q of the transmitted polarizations may then be recovered by a complex-valued matrix multiplication [16]. Provided that sampling is performed above the Nyquist rate, linear channel impairments, such as CD and PMD, may be compensated using a linear equalizer with low power penalty [8]. In this paper, we shall assume that the equalizer outputs have no ISI and no nonlinear distortion. We shall further assume that symbol synchronization has been achieved and that the equalizer correctly rotates the reference LO polarizations into the polarizations of the transmitted signal so that $y_{1,k}$ and $y_{2,k}$ have low-polarization crosstalk and are at one sample per symbol.

We assume that carrier recovery is performed separately for each channel.[1] All the quantities expressed, including the transmitted power, noise variance, and bit rate, are therefore referenced for the one polarization considered. The input signal to the carrier-recovery unit, then, has the form

$$y_k = x_k e^{j\theta_k} + n_k \tag{1}$$

where x_k is a complex-valued symbol transmitted at the kth symbol period, θ_k is the carrier phase, and n_k is the additive white Gaussian noise[2] (AWGN) with a circular Gaussian distribution with zero mean and variance $N_0/2$ per dimension. The goal of this paper is to find an algorithm that can estimate θ_k, which will allow derotation of y_k by multiplying it with $e^{-j\theta_k}$. Maximum-likelihood (ML) detection can then follow to estimate the transmitted sequence $\hat{x}_k$.

We will initially assume that the frequencies of signal and LO lasers are synchronized so that the time variation in θ_k is due to phase noise only, which can be modeled as a Wiener process. In Section III-D, we will investigate the impact of a small frequency difference δf between the carrier and LO lasers, where we will propose algorithms for estimating and compensating δf.

B. MAP Symbol-by-Symbol Detection

The conditional probability density function (pdf) of the received signal y_k is given by

$$p(y_k|x_k,\theta_k) = \frac{1}{\pi N_0}\exp\left(-\frac{\left|y_k - x_k e^{j\theta_k}\right|^2}{N_0}\right). \tag{2}$$

Since x_k and θ_k are independent, the MAP estimates of the transmitted symbol and the carrier phase are

$$(\hat{x}_k,\hat{\theta}_k) = \max_{x_k,\theta_k} p(y_k|x_k,\theta_k)p(x_k)p(\theta_k). \tag{3}$$

In general, this joint maximization problem does not yield a closed-form solution. One method of evaluating (3) is to discretize phase to a set of values $[\theta]$ and then compute the cost function (3) for different pairs of $(x_m,[\theta])$; the pair that maximizes (3) will be the MAP estimate. This approach is computationally expensive as θ_k needs to be known accurately (particularly for dense modulation formats like 16-QAM) to avoid large power penalties. It is, thus, unsuitable for high-bit-rate systems. In data-assisted (DA) phase estimation, training symbols are transmitted. Provided θ_k is nearly static and its *a priori* pdf is uniform, the MAP estimate has a simple analytical formula

$$\hat{\theta}_k = \max_{\theta_k} p(y_k|x_k,\theta_k) = \arg\{y_k\} - \arg\{x_k\}. \tag{4}$$

[1]It is possible to generalize our results for two polarizations by employing separate soft-decision phase estimators for each polarization channel and, then, using a hard-decision phase estimator that combines the soft-phase estimates from the two channels.

[2]This includes LO spontaneous beat noise in a long-haul transmission system with inline optical amplifiers, as well as shot and thermal noises produced by the receiver.

During data transmission, we can approximate the DA algorithm by replacing the known symbols x_k with the output of a decision device, which we denote as $\hat{x}_k$. This approach approximates (3) by first maximizing the cost function over $\hat{x}_k$ and, then, over $\hat{\theta}_k$

$$\max_{x_k,\theta_k} p(y_k|x_k,\theta_k) \approx \max_{\theta_k} \max_{x_k} p(y_k|x_k,\theta_k). \tag{5}$$

Provided that the system is operating at low BER, x_k equals $\hat{x}_k$ with high probability, and it has been shown that this algorithm, known as decision-directed (DD) carrier recovery, is asymptotically optimal for high signal-to-noise ratio (SNR) [17].

Another approach is to reverse the order of the approximations used in evaluating (3)

$$\max_{x_k,\theta_k} p(y_k|x_k,\theta_k) \approx \max_{x_k} \max_{\theta_k} p(y_k|x_k,\theta_k). \tag{6}$$

This approach leads to the nondecision-aided (NDA) algorithm, which first computes $\hat{\theta}_k$ that maximizes the average of the cost function over all possible transmitted symbols $\{x_m : 0 \le m \le M-1\}$

$$p(y_k|\theta_k) = \sum_{m=0}^{M-1} p(y_k|x_k = x_m,\theta_k)p(x_k = x_m). \tag{7}$$

The NDA algorithm is especially well suited to M-ary PSK transmission since raising the received signal to the Mth power removes the data modulation, allowing θ_k to be estimated without any symbol decisions. The received signal y_k can then be derotated by $\hat{\theta}_k$, and $\hat{x}_k$ may be found by a decision device. It has been shown that the NDA algorithm is asymptotically optimal for low SNR [17].

C. MAP Sequence Detection

In sequence detection, the receiver takes a vector $\mathbf{y} = (y_k, y_{k-1}, \ldots, y_{k-L+1})$ and computes the best estimates of the transmitted symbols $\mathbf{x} = (x_k, x_{k-1}, \ldots, x_{k-L+1})$ and the carrier phases $\boldsymbol{\theta} = (\theta_k, \theta_{k-1}, \ldots, \theta_{k-L+1})$. In most semiconductor lasers, phase noise can be modeled as a Wiener process

$$\theta_k = \sum_{m=-\infty}^{k} \nu_m \tag{8}$$

where the ν_m's are independently identically distributed (i.i.d.) Gaussian random variables with zero mean and variance $\sigma_p^2 = 2\pi\Delta\nu T$, $\Delta\nu$ is the sum of the 3-dB linewidths of the signal and LO lasers (also known as the beat linewidth), and T is the symbol period.

The MAP estimation procedure is given by

$$(\hat{\mathbf{x}},\hat{\boldsymbol{\theta}}) = \max_{\mathbf{x},\boldsymbol{\theta}} p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})p(\mathbf{x})p(\boldsymbol{\theta}) \tag{9}$$

where

$$p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}) = \prod_{l=0}^{L-1} \frac{1}{\pi N_0} \exp\left(-\frac{\left|y_{k-l} - x_{k-l}e^{j\theta_{k-l}}\right|^2}{N_0}\right) \tag{10}$$

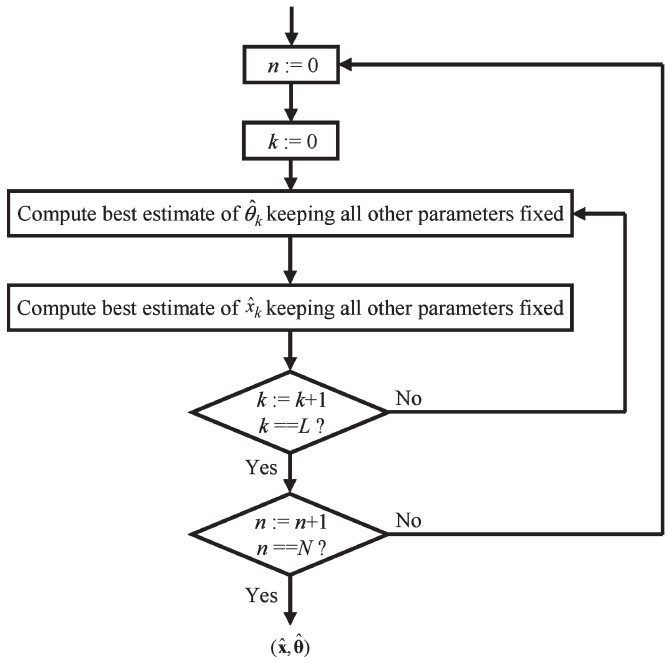

Fig. 2. Iterative procedure for estimating $\mathbf{x}$ and $\boldsymbol{\theta}$.

and $p(\mathbf{x})$ and $p(\boldsymbol{\theta})$ are the *a priori* probabilities of the transmitted sequences $\mathbf{x}$ and carrier phases $\boldsymbol{\theta}$. Since phase noise is a Wiener process, we have

$$p(\boldsymbol{\theta}) = \frac{p(\theta_{k-L+1})}{\left(2\pi\sigma_p^2\right)^{\frac{L}{2}}} \prod_{l=0}^{L-2} \exp\left(-\frac{(\theta_{k-l} - \theta_{k-l-1})^2}{2\sigma_p^2}\right). \tag{11}$$

We note that the evolution of the carrier phase as a Wiener process performs a similar function to coding in communication systems. In coding, the transmitter imposes temporal correlation in the transmitted signal that permits the receiver to correct symbol-detection errors. Here, temporal correlation in the carrier phase is introduced by the Wiener process, as the phase at any symbol period is likely to have a value similar to the phases at adjacent symbols. In (9), the temporal correlation in the transmitted symbols and carrier phases is represented by the Lth order joint pdfs $p(\mathbf{x})$ and $p(\boldsymbol{\theta})$. MAP sequence detection yields performance superior to the MAP symbol-by-symbol detection discussed in Section II-B.

It is possible to approximate (9) using a DD approach

$$(\hat{\mathbf{x}},\hat{\boldsymbol{\theta}}) \approx \max_{\boldsymbol{\theta}} \max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})p(\mathbf{x})p(\boldsymbol{\theta}) \tag{12}$$

or an NDA approach

$$(\hat{\mathbf{x}},\hat{\boldsymbol{\theta}}) \approx \max_{\mathbf{x}} \max_{\boldsymbol{\theta}} p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})p(\mathbf{x})p(\boldsymbol{\theta}). \tag{13}$$

In both of these approximations, however, each maximization is over an L-dimensional space so a simple closed-form solution like (5) does not exist. It is possible to approximate (9) with the approach shown in Fig. 2, where the parameters are optimized one at a time. Such a solution is clearly suboptimal. However, if we iterate the algorithm enough times, subsequent passes can correct decision errors made earlier, and the

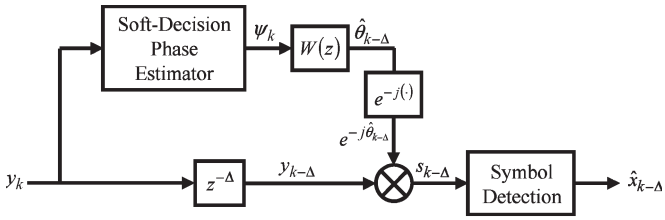

Fig. 3. Two-stage iterative carrier-phase estimator.

algorithm may converge to an optimal solution (or possibly a local minimum).

III. Practical Implementation

A. Two-Stage Iterative Carrier-Phase Estimator

In this paper, we propose an algorithm, based on Fig. 3, which employs a two-stage iteration process for finding $\hat{\mathbf{x}}$ and $\hat{\boldsymbol{\theta}}$. The first-stage (soft-decision) phase estimator is shown in the upper branch. It computes soft estimates of the carrier phase (ψ_k) for each symbol period without taking temporal correlation in θ_k into account. The details of its operation are explained in Section III-B. In the second stage of phase estimation, ψ_k is passed through a linear filter $W(z)$ whose output is the minimum mean-square-error (mmse) estimate of θ_k. We denote this output by $\hat{\theta}_{k-\Delta}$, where Δ is the delay of the filter. The selection of its coefficients and the resulting performance is analyzed in Section III-C. Finally, the received signal is delayed by Δ and multiplied with $e^{-j\hat{\theta}_{k-\Delta}}$ to produce derotated symbols $s_{k-\Delta}$ that are then detected. The output sequence $\hat{x}_{k-\Delta}$ is the receiver's estimate of the transmitted symbols.

Our structure is highly flexible, as each functional block in Fig. 3 can be separately designed for tradeoff between implementation complexity and performance. For example, the soft-decision phase estimator may employ a DD or NDA design; $W(z)$ may be a FIR filter, an infinite-impulse-response (IIR) filter, or a combination thereof; and the output stage can be as simple as a symbol-by-symbol detector, or it may be an ML-sequence detector.

B. Soft-Decision Phase Estimator

A block diagram of the soft-decision phase estimator is shown in Fig. 4(a) and (b) for NDA and DD designs, respectively. Both take y_k as their input, and both produce a soft estimate ψ_k of the carrier phase θ_k.

NDA Soft-Decision Phase Estimator: The NDA soft-decision phase estimator exploits the M-fold rotational symmetry of an M-PSK constellation. We can write the set of transmitted signals as: $x_m = \sqrt{P_{tx}}e^{j2m\pi/M}$, $m = 1, \ldots, M$. Raising the received signal (1) to the M-th power, we get

$$y_k^M = \left(x_k e^{j\theta_k} + n_k\right)^M = P_{tx}^{M/2} e^{jM\theta_k} + m_k \tag{14}$$

where $x_k^M = P_{tx}^{M/2} e^{jM\theta_k}$ is the desired term depending on θ_k, and $m_k = \sum_{p=1}^{M} \binom{M}{p} (x_k e^{j\theta_k})^{M-p} n_k^p$ is a sum of the unwanted cross terms between the signal and AWGN. Fig. 5(a) shows the impact of m_k, which adds a noise vector to $P_{tx}^{M/2} e^{jM\theta_k}$.

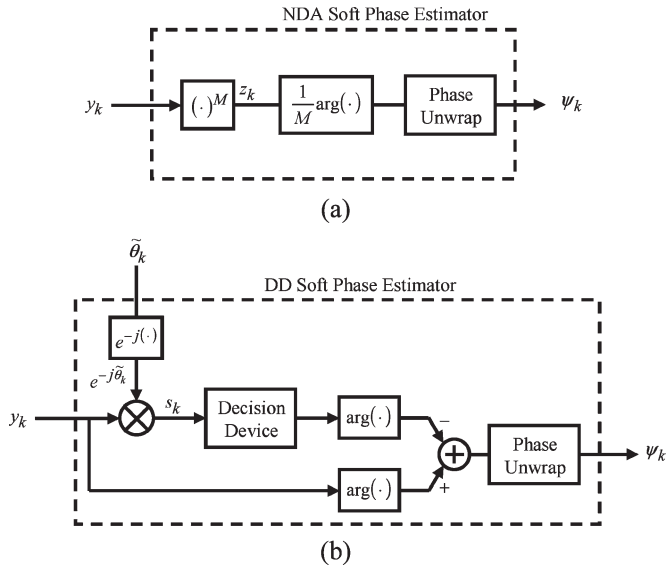

Fig. 4. Soft-decision phase-estimator structures for (a) NDA algorithm and (b) DD algorithm.

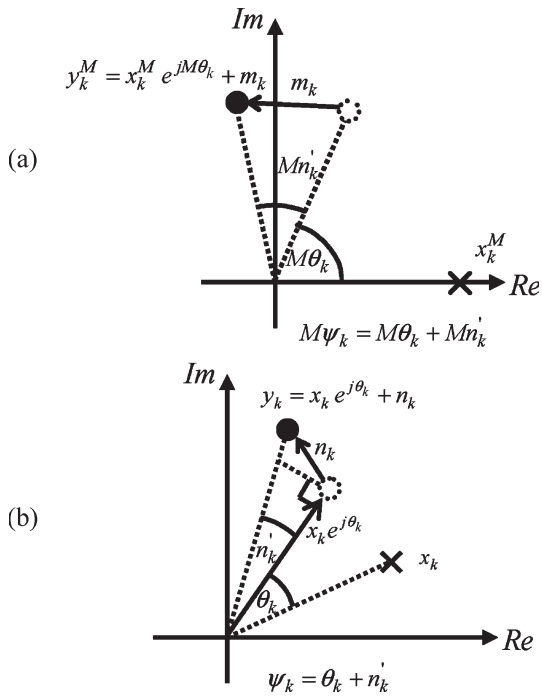

Fig. 5. Signal models of the soft-decision phase estimator. (a) NDA algorithm. (b) DD algorithm.

Consequently, when we take the argument of y_k^M and scale the result by $1/M$, we get

$$\psi_k = \frac{1}{M} \arg\left\{y_k^M\right\} \approx \theta_k + n_k' \tag{15}$$

where Mn_k' is the angular projection of m_k, which is corrupting our estimate of θ_k. At high SNR, it can be shown that n_k' is approximately i.i.d. Gaussian with zero mean and variance

$$\sigma_{n'}^2 = \eta(M, \gamma)\frac{1}{\gamma} \tag{16}$$

where γ is the SNR per symbol,[3] and $\eta(M,\gamma)$ is a function derived in Appendix A.

DD Soft-Decision Phase Estimator: When a non-PSK constellation is transmitted, the carrier phase can be estimated with a DD soft-decision phase estimator, as shown in Fig. 4(b). A key difference between the NDA and the DD phase estimator is that an initial estimate of the phase (denoted as $\tilde{\theta}_k$) is required to find ψ_k.[4] The principle of the DD phase estimator is as follows: The received signal y_k is first derotated by $\tilde{\theta}_k$, and the result is passed through a decision device

$$\hat{x}_k = \left[y_k e^{-j\tilde{\theta}_k} \right]_{\mathrm{D}} \tag{17}$$

where $[\cdot]_{\mathrm{D}}$ denotes the output of the decision device. At high SNR, $\hat{x}_k = x_k$ with high probability, so our soft estimate of the received phase is

$$\psi_k = \arg\{y_k\} - \arg\{\hat{x}_k\}. \tag{18}$$

Fig. 5(b) shows (17) and (18) graphically. The received signal y_k is the sum of x_k rotated by θ_k plus a noise vector n_k (1). The angular projection of n_k in the direction orthogonal to $x_k e^{j\theta_k}$ is the phase-estimation noise n'_k. We can, therefore, write

$$\psi_k = \theta_k + n'_k. \tag{19}$$

At high SNR, n'_k is approximately i.i.d. Gaussian with zero mean and variance

$$\sigma^2_{n'} = \eta \frac{1}{\gamma} \tag{20}$$

where $\eta = (1/2)E[|x|^2]E[1/|x|^2]$ is equal to half the "constellation penalty," which is a function of the modulation format [2], and γ is the SNR per symbol.

Phase Unwrapping: Both the NDA and DD phase estimators require phase unwrapping. This is because the $\arg\{\cdot\}$ function only returns values between $-\pi$ and π, as angles that differ by integer multiples of 2π are indistinguishable. In the absence of the phase unwrapper, the NDA phase-estimator output is constrained between $-\pi/M$ and π/M, whereas the DD output is constrained between $-\pi$ and π.[5] Our carrier-phase model in (8) assumes θ_k to be a Wiener process whose value is unconstrained. Phase unwrapping has been studied in [9]. Its purpose is to add an integer multiple of $2\pi/M$ to ensure that the magnitude of the phase difference between adjacent symbols is always less than π/M. Let $\widetilde{\psi}_k$ be the carrier phase prior to the unwrapper. We compute

$$\psi_k = \tilde{\psi}_{k+1} + p\, 2\pi/M. \tag{21}$$

[3]Since carrier recovery is performed symbol by symbol, it is more convenient to use the SNR per symbol γ in our theoretical treatment. If the transmission format encodes b bits per symbol, the SNR per bit is given by $\gamma_b = \gamma/b$.

[4]In Section III-B3, we will show how $\tilde{\theta}_k$ can be obtained.

[5]For convenience, we may treat the DD phase estimator as an NDA with $M = 1$.

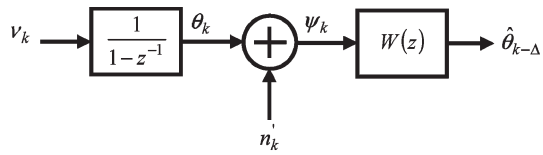

Fig. 6. Carrier-phase-estimator signal model.

It can be shown that the required p is[6]

$$p = \left\lfloor \frac{1}{2} + \frac{\psi_{k-1} - \widetilde{\psi}_k}{2\pi/M} \right\rfloor. \tag{22}$$

Phase unwrapping can cause cycle slips, which is a highly nonlinear phenomenon [17]. We shall discuss how this impacts the system's performance in Section III-D.

C. Hard-Decision Phase Estimator

Wiener Filter: Using (15) and (19), we have the linear-phase-estimator model shown in Fig. 6. The input frequency noise ν_k is a white Gaussian process whose running sum is the phase noise θ_k. At the output of the soft-decision phase estimator, θ_k is corrupted by noise n'_k to produce ψ_k. We can pass ψ_k through a Wiener filter $W(z)$ whose output is the mmse estimate $\hat{\theta}_k$ of the actual carrier phase θ_k. We will refer to this filter as the hard-decision phase estimator. Writing all the quantities in Fig. 6 in terms of their two-sided z-transforms, we have

$$X(z) = \lim_{K\to\infty} \sum_{n=-K}^{+K} x_n z^{-n}. \tag{23}$$

It can be shown that the Wiener solution $W(z)$ is given by

$$W(z) = \frac{E\left[\theta(z)\psi(z^{-1})\right]}{E\left[\psi(z)\psi(z^{-1})\right]} = \frac{rz^{-1}}{-1 + (2+r)z^{-1} - z^{-2}} \tag{24}$$

where $r = \sigma^2_p/\sigma^2_{n'} > 0$. This filter has two poles at

$$z_1, z_2 = \left(1 + \frac{r}{2}\right) \pm \sqrt{\left(1 + \frac{r}{2}\right)^2 - 1} \tag{25}$$

that are inverses of each other, with z_1 inside the unit circle mapping to a causal sequence and z_2 outside the unit circle mapping to an anticausal sequence. The inverse z-transform of (24) is

$$w_n = \begin{cases} \frac{\alpha r}{1-\alpha^2}\alpha^n, & n \geq 0 \\ \frac{\alpha r}{1-\alpha^2}\alpha^{-n}, & n < 0 \end{cases} \tag{26}$$

[6]At low SNR or high phase noise, it is possible that an inaccurate value of ψ_{k-1} may lead to the wrong value of p being computed. Consequently, a more robust algorithm can use several prior phases to compute p. In our simulations in Section V, we found that $p = \lfloor (1/2) + ((1/3)(\psi_{k-1} + \psi_{k-2} + \psi_{k-3}) - \widetilde{\psi}_k/2\pi/M) \rfloor$ yields reliable performance.

where $\alpha = (1 + r/2) - \sqrt{(1 + r/2)^2 - 1}$. The mmse hard-decision phase estimator, thus, computes

$$\hat{\theta}_k = \sum_{l=-\infty}^{+\infty} w_l \psi_{k-l}. \tag{27}$$

We note that the Wiener-filter coefficients in (26) consist of two exponentially decaying sequences that are symmetric about $n = 0$. This result is intuitively satisfying: First, the hard-decision phase estimator gives decreasing emphasis to soft decisions that are far away from symbol k whose phase is estimated in (27), as phase noise makes far away symbols inaccurate estimators of θ_k. Second, the symmetry of the filter means that the causal and anticausal soft-phase estimates are weighted equally, which is expected, since the accuracy of ψ_{k-l}, as an estimator of θ_k, depends only on $|l|$. Finally, the decay rate of the Wiener coefficients depends only on the ratio between the intensity of phase noise and AWGN. In the limit of low phase noise ($\sigma_p^2 \ll \sigma_{n'}^2$), $\alpha \to 1$ and the decay rate is slow because of the long coherence time of the phase-noise process. Conversely, in the limit of high phase noise ($\sigma_p^2 \gg \sigma_{n'}^2$), soft-decision phase estimates at symbols far from period k are poor estimators of θ_k, so $\alpha \to 0$ and the Wiener coefficients rapidly decay. It can be shown that $\sum_{n=-\infty}^{\infty} w_n = 1$, so (27) is an unbiased estimator of θ_k.

FIR Approximation[7]*:* Since the coefficients of the Wiener filter are nonzero over $-\infty$ to $+\infty$, it is not possible to implement (27) unless the system has infinite delay. However, the exponentially decaying coefficients become negligible for sufficiently large $|n|$, so we can truncate (27) without significant performance degradation. For example, if we neglect the coefficients that are less than a fraction f of the value of the largest coefficient w_0. We can approximate (26) with a FIR filter of length

$$L = \left\lceil 2\frac{\log(f)}{\log(1/\alpha)} \right\rceil. \tag{28}$$

Let $W_{\mathrm{FIR}}(z)$ be the filter that takes as its input $\boldsymbol{\psi} = (\psi_k, \psi_{k-1}, \ldots, \psi_{k-L+1})$ and produces $\hat{\theta}_{k-\Delta}$ that is the best estimate of the carrier phase at symbol $k - \Delta$ (Fig. 5)

$$\hat{\theta}_{k-\Delta} = \sum_{l=0}^{L-1} w_l \psi_{k-l}. \tag{29}$$

It can be shown that the conditional pdf of $\boldsymbol{\psi}$, given $\theta_{k-\Delta}$, is

$$p(\boldsymbol{\psi}|\theta_{k-\Delta}) = \frac{1}{(2\pi)^{L/2}|K|^{1/2}} \cdot \exp\left(-\frac{1}{2}(\boldsymbol{\psi} - \theta_{k-\Delta}\mathbf{1})^{\mathrm{T}}\mathbf{K}^{-1}(\boldsymbol{\psi} - \theta_{k-\Delta}\mathbf{1})\right) \tag{30}$$

[7]It is possible to approximate (26) with other filter structures. For instance, if a zero-delay filter is desired (e.g., for $W_{\mathrm{sd}}(z)$ in Section III-D), we can discard the noncausal Wiener filter coefficients. The remaining causal coefficients can then be implemented as an IIR filter with a single feedback tap of α. We can also use an FIR–IIR hybrid for (26), where the causal coefficients are implemented as an IIR filter, and $L/2$ noncausal coefficients (L given by (28) with $f = 0.05$) are implemented as an FIR filter optimized for delay $\Delta = L/2$. The IIR output can then be delayed by Δ, and the signals in the two branches are optimally combined with splitting ratios equal to the inverse of their respective phase-error variances.

where $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}$ is a vector of L ones, and $\mathbf{K}$ is an autocorrelation matrix, which is a sum of two components

$$\mathbf{K} = \mathbf{K_p} + \mathbf{K_n} \tag{31}$$

$$\mathbf{K_p} = \sigma_p^2 \cdot \begin{bmatrix} \Delta & \cdots & 3 & 2 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots & \vdots \\ 3 & \cdots & 3 & 2 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 2 & \cdots & 2 & 2 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & 0 & 1 & 1 & 1 & \cdots & 1 \\ 0 & \cdots & 0 & 0 & 0 & 0 & 1 & 2 & 2 & \cdots & 2 \\ 0 & \cdots & 0 & 0 & 0 & 0 & 1 & 2 & 3 & \cdots & 3 \\ \vdots & \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & 1 & 2 & 3 & \cdots & L-\Delta-1 \end{bmatrix} \tag{32}$$

and

$$\mathbf{K_n} = \sigma_{n'}^2 \cdot \mathbf{I}_{L\times L}. \tag{33}$$

$\mathbf{I}_{L\times L}$ is the $L \times L$ identity matrix. The terms $\mathbf{K_p}$ and $\mathbf{K_n}$ arise from the autocorrelations of θ_k and n'_k, respectively. The upper right and lower left corners of $\mathbf{K_p}$ are zeros because of the conditional independence between the carrier phases before and after symbol $k - \Delta$, given $\theta_{k-\Delta}$.

We can find the ML estimator (which also happens to be the mmse estimator) for $\theta_{k-\Delta}$ by minimizing the argument in the exponential in (30). Since $\mathbf{K}^{-1}$ is symmetric, it can be shown that

$$\mathbf{W}_{\mathrm{FIR}} = \frac{\mathbf{K}^{-1}\mathbf{1}}{\mathbf{1}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{1}}. \tag{34}$$

The coefficients in (34), again, sum to one, so (29) is an unbiased estimator of $\theta_{k-\Delta}$.

There are a number of alternative settings for $W(z)$ that are commonly used in practice. For instance, $\mathbf{W} = (1/L) \cdot [1, \ldots, 1]$ is a uniform filter that gives equal weighting to all soft decisions in the computation of $\hat{\theta}_{k-\Delta}$ and was used in [10]. This strategy is asymptotically optimal for low phase noise ($\sigma_p^2 \ll \sigma_{n'}^2$). Alternatively, a one-shot estimator $\mathbf{W} = [0, \ldots, 0, 1, 0, \ldots, 0]$ with a single tap of one at w_Δ is asymptotically optimal for high phase noise ($\sigma_p^2 \gg \sigma_{n'}^2$). As in the case of the infinite-length Wiener filter in (26), the coefficients in (34) converge to these two limiting cases for low and high phase noise, respectively. In Fig. 7, the coefficients of $\mathbf{W}_{\mathrm{FIR}}$ are shown for different values of $r = \sigma_p^2/\sigma_{n'}^2$.

MSE Performance: In a system where AWGN and phase noise are the only channel impairments, the mean and variance of the phase error $\varepsilon_k = \hat{\theta}_{k-\Delta} - \theta_{k-\Delta}$ determine the system power penalty at a given BER [18]. Consider the error signal

$$\varepsilon_k = \sum_{l=0}^{L-1} w_l \psi_{k-l} - \theta_{k-\Delta}. \tag{35}$$

Regardless of whether a DD or NDA soft-decision phase estimator is used, from (15) or (19), ψ_{k-l} can be written as

$$\psi_{k-l} = \theta_{k-\Delta} + (\theta_{k-l} - \theta_{k-\Delta}) + n'_{k-l}. \tag{36}$$

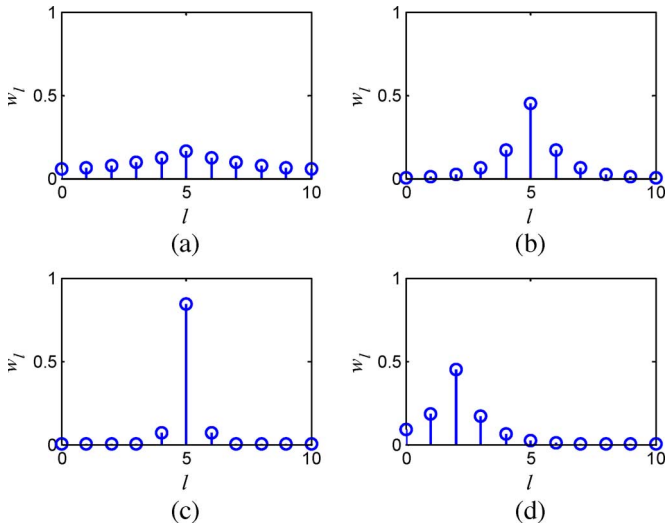

Fig. 7. Wiener filter coefficients for (a) $L = 11$, $\Delta = 5$, $\sigma_p^2/\sigma_{n'}^2 = 0.1$; (b) $L = 11$, $\Delta = 5$, $\sigma_p^2/\sigma_{n'}^2 = 1$; (c) $L = 11$, $\Delta = 5$, $\sigma_p^2/\sigma_{n'}^2 = 10$; and (d) $L = 11$, $\Delta = 2$, $\sigma_p^2/\sigma_{n'}^2 = 1$.

Since the carrier phase is a Wiener process given by (8)

$$\theta_{k-l} - \theta_{k-\Delta} = \begin{cases} \sum_{m=k-\Delta+1}^{k-l} \nu_m, & l < \Delta \\ \sum_{m=k-l+1}^{k-\Delta} (-\nu_m), & l > \Delta. \end{cases} \tag{37}$$

Substituting (36) and (37) into (29), the hard-decision phase estimate $\hat{\theta}_{k-\Delta}$ is given by

$$\begin{aligned} \hat{\theta}_{k-\Delta} = & \left[\theta_{k-\Delta} \cdot \sum_{l=0}^{L-1} w_l\right] \\ & + \left[\sum_{m'=0}^{\Delta-1} \nu_{k-m'} \left(\sum_{l'=0}^{m'} w_{l'}\right) \right. \\ & \left. + \sum_{m'=\Delta+1}^{L-1} (-\nu_{k-m'+1}) \left(\sum_{l'=m'}^{L-1} w_{l'}\right)\right] \\ & + \left[\sum_{l=0}^{L-1} w_l n'_{k-l}\right]. \end{aligned} \tag{38}$$

The three terms enclosed in square brackets are as follows: 1) the carrier phase being estimated, weighted by the sum of the filter coefficients, 2) a weighted sum of the frequency noises before and after symbol period $k - \Delta$, and 3) a weighted sum of AWGN. In an unbiased estimator, $\sum_{l=0}^{L-1} w_l = 1$ [this is the case with the coefficients in (34)], so ε_k has zero mean, and its variance is given by

$$\begin{aligned} \sigma_\varepsilon^2(\mathbf{W}, \Delta) = & \sigma_p^2 \cdot \left[\sum_{m'=0}^{\Delta-1} \left(\sum_{l'=0}^{m'} w_{l'}\right)^2 + \sum_{m'=\Delta+1}^{L-1} \left(\sum_{l'=m'}^{L-1} w_{l'}\right)^2\right] \\ & + \sigma_{n'}^2 \cdot \left[\sum_{l=0}^{L-1} w_l^2\right]. \end{aligned} \tag{39}$$

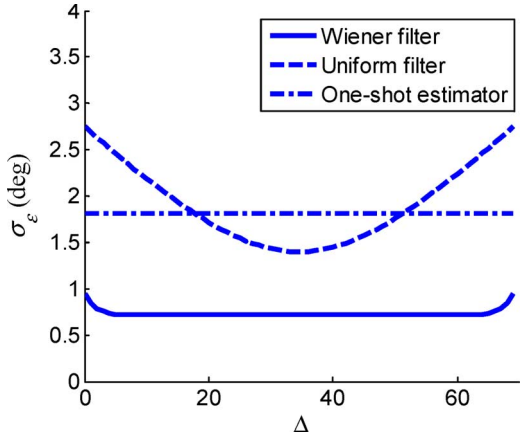

Fig. 8. Phase-error standard deviation of the hard-decision phase estimator versus filter delay for $L = 70$, $\sigma_{n'}^2 = 10^{-3}$, and $\sigma_p^2 = 10^{-4}$.

For any given L, Δ, σ_p^2, and $\sigma_{n'}^2$, the phase-error variance is minimized when the coefficients in (34) are used.

We intuitively expect that the lowest mse is obtained when the delay is equal to half the filter length, i.e., $\Delta = \lfloor L - 1/2 \rfloor$. If $\Delta = 0$, for example, the soft decisions ψ_{k-l} for large l will be poor estimators of $\theta_{k-\Delta}$ due to phase noise. A similar argument holds for $\Delta = L - 1$. Setting $\Delta = \lfloor L - 1/2 \rfloor$ results in the same number of soft phases used from either side of symbol period $k - \Delta$ for estimating $\theta_{k-\Delta}$. A plot of phase-error standard deviation versus delay is shown in Fig. 8 for a fixed filter length of 70. As expected, the FIR Wiener filter has the best performance for all delay values. The one-shot estimator's performance is independent of Δ, since it always considers only one soft-phase sample at the desired symbol period. The uniform filter has poor performance when the delay Δ is either too small or too large.

In Fig. 9, we set the filter delay to its optimal value of $\Delta = \lfloor L - 1/2 \rfloor$ and plot the phase-error standard deviation versus L for 4-, 8-, and 16-QAMs when the receiver optical SNR (OSNR) is 1 dB above sensitivity. In Fig. 9(a)–(c), a target BER of 10^{-3} is used, while in Fig. 9(d)–(f), the target BER is 10^{-9}. We assume the use of a DD soft-decision phase estimator in all plots [$\sigma_{n'}^2$ given by (20)].Column three of Table I shows the OSNRs needed to achieve these target BERs when there is no phase error.

To see how the results in Fig. 9 compare with a PLL, the maximum tolerable phase-error standard deviations to ensure less than 0.5-dB power penalty at the target BER are listed in columns four and five of Table I.[8] These results assume that phase error has a Tikhonov distribution in the case of a PLL [18] and a Gaussian distribution in the case of feedforward carrier-phase estimation. The maximum tolerable laser linewidth for a PLL is shown in column six.[9]

[8]See [2, Fig. 11] for the BER curves for 8- and 16-QAM when phase error has a Tikhonov distribution. The maximum tolerable σ_ε at BER $= 10^{-9}$ for Tikhonov-distributed phase error were shown in [2, Table 1]. To obtain σ_ε for BER $= 10^{-3}$, we extended our numerical results to low SNR. To obtain σ_ε for Gaussian-distributed phase error (column five of Table I), we used (69) as the pdf of the received signal instead [2, eq. (7)].

[9]Using the values of σ_ε shown in column four in Table I, we invoke (46) to compute the maximum tolerable linewidth.

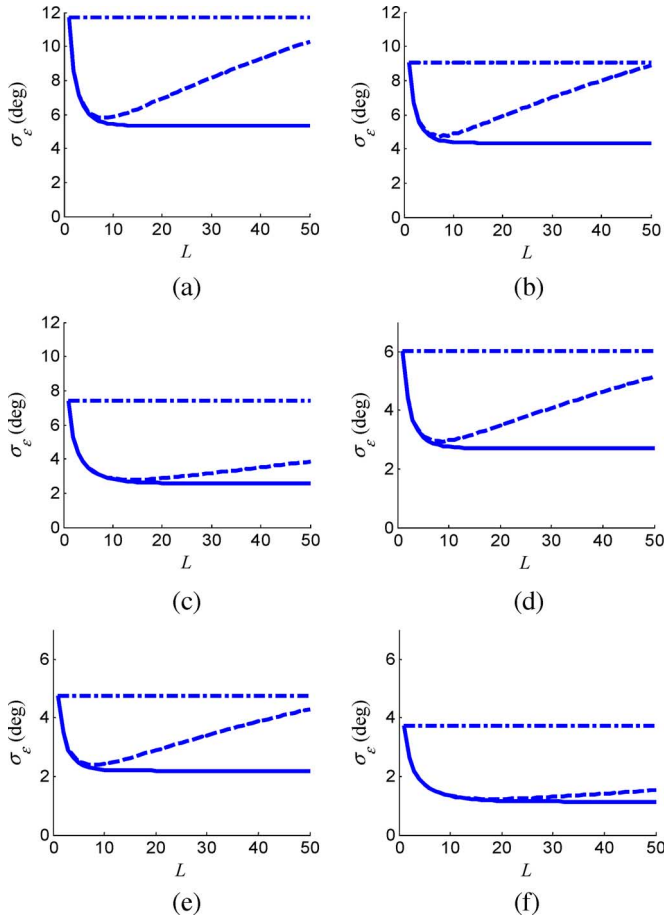

Fig. 9. Phase-error standard deviation σ_ε versus filter length L for the hard-decision phase estimator whose delay is assumed to be $\Delta = \lfloor L - 1/2 \rfloor$. The cases shown are as follows: (a) 4-QAM at $\gamma_\mathrm{b} = 7.79$ dB, $\Delta\nu T_\mathrm{b} = 6.0 \times 10^{-4}$, (b) 8-QAM at $\gamma_\mathrm{b} = 10.03$ dB, $\Delta\nu T_\mathrm{b} = 3.0 \times 10^{-4}$, (c) 16-QAM at $\gamma_\mathrm{b} = 11.52$ dB, $\Delta\nu T_\mathrm{b} = 4.0 \times 10^{-5}$, (d) 4-QAM at $\gamma_\mathrm{b} = 13.55$ dB, $\Delta\nu T_\mathrm{b} = 1.5 \times 10^{-4}$, (e) 8-QAM at $\gamma_\mathrm{b} = 15.60$ dB, $\Delta\nu T_\mathrm{b} = 7.0 \times 10^{-5}$, and (f) 16-QAM at $\gamma_\mathrm{b} = 17.46$ dB, $\Delta\nu T_\mathrm{b} = 6.0 \times 10^{-6}$. The solid line "-," the dashed line "- -," and the dotted–dashed line "-.-" in each plot are the results for $W(z)$ being a Wiener filter, a uniform filter, and a one-shot estimator.

TABLE I
SENSITIVITY AND PHASE-ERROR REQUIREMENTS FOR 4-, 8-, AND 16-QAMS

Format	Target BER	Sensitivity SNR per bit (γ_b) (dB)	Max. Tolerable σ_ε for 0.5 dB power penalty		Max. Linewidth ($\Delta\nu T_b$)
			Tikhonov	Gaussian	
4-QAM	10^{-3}	6.79	5.45°	5.46°	1.05×10^{-4}
	10^{-9}	12.55	2.89°	2.89°	3.11×10^{-5}
8-QAM	10^{-3}	9.03	4.46°	4.46°	5.23×10^{-5}
	10^{-9}	14.60	2.30°	2.31°	1.33×10^{-5}
16-QAM	10^{-3}	10.52	2.65°	2.70°	7.86×10^{-6}
	10^{-9}	16.46	1.20°	1.20°	1.20×10^{-6}

We define 8-QAM in this paper to be the constellation shown in Fig. 10 of [2].

In each of the plots in Fig. 9(a)–(f), if we substitute the value of L, then we would obtain for the Wiener filter length using (28) with $f = 0.05$; the achievable σ_ε is less than the values shown in column five of Table I. Thus, feedforward carrier-phase estimation allows the linewidths shown in the caption of Fig. 9. In comparison with the PLL's linewidths shown in column six, feedforward carrier-phase estimation appears to be five to six times more phase-noise tolerant than a PLL. In particular, the results in Fig. 9(a)–(c) suggests that 4-, 8-, and 16-QAMs transmissions at 10 Gbps at a BER of 10^{-3} is possible with laser beat linewidths as large as 6 MHz, 3 MHz, and 400 kHz, respectively. These numbers are well within the capabilities of current ECLs [3]. However, our results have not taken soft-decision error propagation into account. We shall outline the procedure in estimating the system BER in Section III-E after discussion of soft-decision errors.

D. Soft-Decision Phase-Estimator Input Phase

It was noted in Section III-B that the DD soft-decision phase estimator[10] requires an input $\widetilde{\theta}_k$. If we assume that phase noise is sufficiently low so that the carrier phase drifts slowly, we may take $\widetilde{\theta}_k$ to be the hard-decision phase-estimator output from a recent symbol. One solution is shown in Fig. 10(a), where $\widetilde{\theta}_{k+1} = \hat{\theta}_{k-\Delta}$. There is an inherent tradeoff in selecting Δ in this setup. Although it was shown in Fig. 8 that the mse for $\hat{\theta}_{k-\Delta}$ is minimized by setting $\Delta = \lfloor L - 1/2 \rfloor$, the benefit gained with having an accurate $\hat{\theta}_{k-\Delta}$ is compromised by the fact that, if Δ is large, $\hat{\theta}_{k-\Delta}$ will not be an accurate initial phase for the soft-decision phase estimator.

We can improve system performance by using a two-filter structure, as shown in Fig. 10(b). A dedicated hard-decision filter $\mathbf{W_{hd}}$ is optimized for delay $\Delta = \lfloor L - 1/2 \rfloor$, while for the feedback path, we use a second filter $\mathbf{W_{sd}}$ that is optimized for $\Delta = 0$ for the prediction of θ_{k+1}.

When an inaccurate $\widetilde{\theta}_k$ causes the decision device in Fig. 5(b) to make an error, the soft-decision estimate ψ_k will acquire an offset equal to the angular difference between the detected symbol and the actual transmitted symbol (Fig. 11). From (29), we see that a soft-decision phase error of $\Delta\varepsilon$ at symbol $k - l$ will cause $\hat{\theta}_{k-\Delta}$ to acquire a mean offset of $w_l\Delta\varepsilon$ if we assume that there are no further errors in the block $\{\psi_m\}_{m=k-L+1}^{k}$. Thus, $\varepsilon_{k-\Delta}$ will be distributed as $\mathrm{N}(w_l\Delta\varepsilon, \sigma_\varepsilon^2)$, where σ_ε^2 is given by (39).

Soft-decision errors can potentially cause catastrophic failure, as the mean phase shift in $\hat{\theta}_{k-\Delta}$ degrades future soft decisions through the feedback $\widetilde{\theta}_{k+1}$ (Fig. 10): The derotated constellation s_k in Fig. 4(b) will be misaligned by $w_l\Delta\varepsilon$. Consider an example where $\Delta\varepsilon = \pi/2$. If the coefficient w_l, which is multiplying $\Delta\varepsilon$ in (29), is 0.1, $w_l\Delta\varepsilon = 9°$. A phase offset as large as this will likely cause severe degradation to the performance of the decision device. Subsequent soft-decision errors will be more likely, leading to catastrophic failure. It is, thus, necessary to take error propagation into account when computing the system BER.[11]

[10]The NDA soft-decision phase estimator does not require an initial phase estimate [Fig. 5(a)].

[11]It will turn out that catastrophic failure actually leads to cycle slip, where differential bit encoding can be employed to avert catastrophic bit errors. This is discussed in Section III-F.

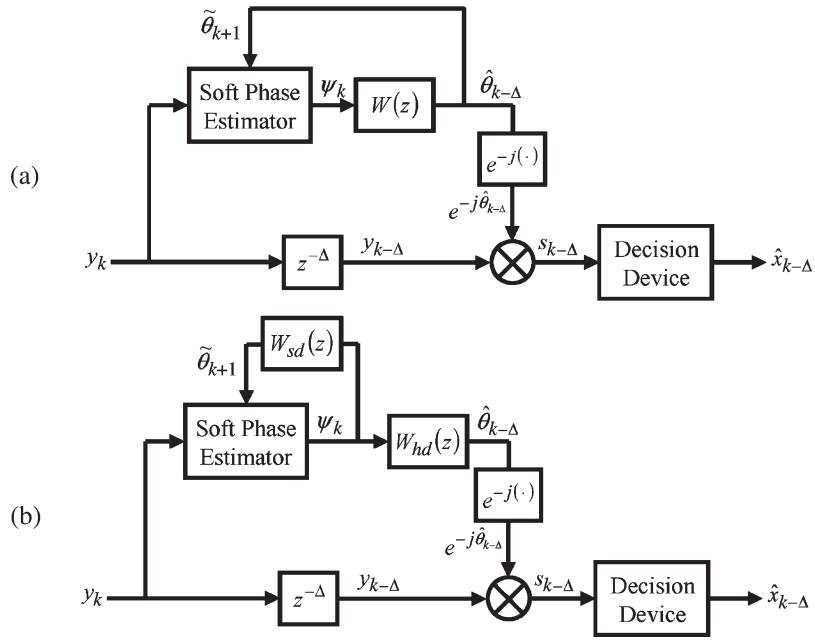

Fig. 10. Carrier-phase-estimator structure employing a DD soft-decision phase estimator. (a) One-filter approach. (b) Two-filter approach.

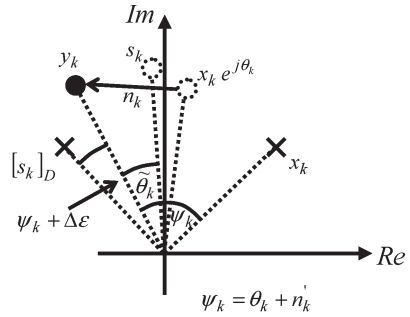

Fig. 11. Effect of a soft-decision phase error. The received signal y_k is derotated by $\widetilde{\theta}_k$ to produce s_k. Due to large phase noise and AWGN, s_k is in the wrong decision region, causing the output of the decision device to produce $[s_k]_{\rm D} \neq x_k$. The phase estimator returns $\psi_k + \Delta\varepsilon$, where $\Delta\varepsilon$ is the angular difference between $[s_k]_d$ and x_k.

We note that catastrophic error propagation can only occur when a DD soft-decision phase estimator is used, since the NDA phase estimator has no feedback. Instead, the NDA phase estimator may suffer from cycle slips due to phase unwrapping (Section III-B).[12] To appreciate its impact, suppose a cycle-slip event occurred at symbol k. If another cycle slip does not occur within the next $L-1$ symbols, the soft phases from ψ_{k+1} onwards will have a constant error of $\Delta\varepsilon = \pm 2\pi/M$, depending on the sign of the cycle-slip event. Substituting this result into (29), we observe that the hard-decision phase-estimator outputs $\{\hat{\theta}_{k-\Delta+m}\}_{m=0}^{L-1}$ will be shifted by $\Delta\varepsilon \sum_{l=0}^{m} w_l$; while outputs $\{\hat{\theta}_{k-\Delta+m}\}_{m=L}^{\infty}$ will have an error of $\Delta\varepsilon$ since the filter coefficients sum to one. A cycle slip thus causes the derotated constellation $s_{k-\Delta}$ (Fig. 3) to have large misalignment from symbol periods $k-\Delta$ to $k-\Delta+L-1$, so symbol-detection errors in this period are likely. For subsequent symbols, however, the phase error will be $\pm 2\pi/M$. Since an M-PSK constellation is rotationally invariant to phase shifts of $\pm 2\pi/M$, low probability of symbol error can still be achieved, provided that differential-bit coding is used [17].[13] A cycle-slip event will result in a finite number of bit errors between symbols $k-\Delta$ and $k-\Delta+L-1$. Catastrophic failure will not occur because the NDA phase estimator has no feedback.

E. System Probability of Error

To compute the overall system BER, we first need to evaluate the probabilities of symbol error for the two decision devices 1) inside the soft-decision phase estimator [Fig. 4(b)] and 2) at the output-symbol detector (Fig. 3).The parameters that determine these error probabilities are the AWGN variance and the mean and variance of phase error. In the setup shown in Fig. 10(b), the phase-error variances are

$$\sigma^2_{\varepsilon,\rm sd} = E\left[(\widetilde{\theta}_{k+1} - \theta_{k+1})^2\right] = \sigma^2_{\varepsilon}(\mathbf{W_{sd}}, 0) + \sigma^2_p \tag{40}$$

$$\sigma^2_{\varepsilon,\rm hd} = E\left[(\hat{\theta}_{k-\Delta} - \theta_{k+\Delta})^2\right] = \sigma^2_{\varepsilon}(\mathbf{W_{hd}}, \Delta). \tag{41}$$

For the special case where the setup of Fig. 10(a) is used, we have $\mathbf{W_{sd}} = \mathbf{W_{hd}} = \mathbf{W}$.

To obtain the probability of symbol error, we define two quantities whose closed-form expressions will be given in Appendix B: 1) $P_e(N_0, \mu_\varepsilon, \sigma^2_\varepsilon)$: The probability of symbol

[12]The DD phase estimator can also suffer from cycle slip.

[13]Note that this increases the BER, as discussed in Section III-F.

error when AWGN and carrier-phase error are distributed as $\mathrm{N}(0, N_0)$ and $\mathrm{N}(\mu_\varepsilon, \sigma_\varepsilon^2)$, respectively. 2) $P(\Delta\varepsilon|N_0, \mu_\varepsilon, \sigma_\varepsilon^2)$: The probability that the soft-decision phase estimator makes an error of $\Delta\varepsilon$ when AWGN and carrier-phase error are distributed as $\mathrm{N}(0, N_0)$ and $\mathrm{N}(\mu_\varepsilon, \sigma_\varepsilon^2)$.

The probability of symbol error for the system may be written as

$$\begin{aligned} P_e = {} & P_e\left(N_0, 0, \sigma_{\varepsilon,\mathrm{hd}}^2\right) \times \prod_{j=0}^{L-1} P\left(0|N_0, 0, \sigma_{\varepsilon,\mathrm{sd}}^2\right) \\ & + \sum_{l=0}^{L-1} \sum_{\Delta\varepsilon} P_e\left(N_0, w_{\mathrm{hd},l}\Delta\varepsilon, \sigma_{\varepsilon,\mathrm{hd}}^2\right) \\ & \times \left[\left(\prod_{j=l+1}^{L-1} P\left(0|N_0, 0, \sigma_{\varepsilon,\mathrm{sd}}^2\right) \right) \times P\left(\Delta\varepsilon|N_0, 0, \sigma_{\varepsilon,\mathrm{sd}}^2\right) \right. \\ & \left. \times \left(\prod_{j=0}^{l-1} P\left(0|N_0, w_{\mathrm{sd},l}\Delta\varepsilon, \sigma_{\varepsilon,\mathrm{sd}}^2\right) \right) \right] \\ & + \text{higher order terms.} \end{aligned} \tag{42}$$

The first term in (42) is the probability that the output-symbol detector makes an error when there are no soft-decision errors in $\{\psi_m\}_{m=k-L+1}^{k}$.

The second term is the probability of the output-symbol detector making an error when there is one soft-decision error in $\{\psi_m\}_{m=k-L+1}^{k}$. Suppose the soft-decision error occurred at symbol $k-l$ and its magnitude is $\Delta\varepsilon$. The hard-decision phase-estimator output will have a mean offset of $\mu_{\hat{\varepsilon}_{k-\Delta}} = w_{\mathrm{hd},l}\Delta\varepsilon$, and its probability of symbol error is $P_e(N_0, w_{\mathrm{hd},l}\Delta\varepsilon, \sigma_{\varepsilon,\mathrm{hd}}^2)$, where $\sigma_{\varepsilon,\mathrm{hd}}^2$ is given by (41). This error probability is then multiplied by the likelihood of a soft-decision phase error of $\Delta\varepsilon$ occurring at symbol $k-l$, which requires:

1) correct estimation of ψ_m from symbols $k-L+1$ to $k-l-1$;
2) a decision error at symbol $k-l$ causing a phase-estimation error of $\Delta\varepsilon$; and
3) correct estimation of ψ_m from symbols $k-l+1$ to k.

The probabilities 1)–3) are the quantities inside the square brackets in (42).

Finally, the "higher order terms" represent the probability that the output-symbol detector makes an error when there are two or more soft-decision errors in $\{\psi_m\}_{m=k-L+1}^{k}$. It is possible to enumerate these terms exactly by using an exhaustive approach as in the previous paragraph. However, the resulting expression becomes unwieldy. If we assume that the higher order terms are negligible at sufficiently high BER, we can write the upper and lower bounds. First, the higher order terms will always sum to a value greater than zero. Second, the occurrence of two or more soft-decision errors will, at most, cause all subsequent symbols to be in error, so we have an upper bound of $1 - P(n_{\mathrm{sd}} = 0) - P(n_{\mathrm{sd}} = 1)$, where $P(n_{\mathrm{sd}} = i)$ is the probability of i soft-decision errors in $\{\psi_m\}_{m=k-L+1}^{k}$. It can be shown that

$$P\left(n_{\mathrm{sd}} = 0\right) = \prod_{j=0}^{L-1} P\left(0|N_0, 0, \sigma_{\varepsilon,\mathrm{sd}}^2\right) \tag{43}$$

$$\begin{aligned} P(n_{\mathrm{sd}} = 1) = {} & \sum_{l=0}^{L-1} \sum_{\Delta\varepsilon} \left(\prod_{j=l+1}^{L-1} P\left(0|N_0, 0, \sigma_{\varepsilon,\mathrm{sd}}^2\right) \right) \\ & \times P\left(\Delta\varepsilon|N_0, 0, \sigma_{\varepsilon,\mathrm{sd}}^2\right) \\ & \times \left(\prod_{j=0}^{l-1} P\left(0|N_0, w_{\mathrm{sd},j}\Delta\varepsilon, \sigma_{\varepsilon,\mathrm{sd}}^2\right) \right). \end{aligned} \tag{44}$$

TABLE II
AVERAGE NUMBER OF BIT ERRORS TO SYMBOL ERRORS (n_b) FOR GRAY-CODED AND DIFFERENTIAL-BIT-ENCODED QAM

	4-QAM	8-QAM	16-QAM
Gray-Code	$\frac{1}{2}$	$\frac{11}{24}$	$\frac{1}{4}$
Differential Bit Encoding	1	$\frac{2}{3}$	$\frac{13}{32}$

The BER of the system can then be found using (45) in conjunction with the average nearest neighbor counts listed in Table II.

In Fig. 12(a), we plot BER versus Δ for 16-QAM for the single-filter carrier-recovery structure shown in Fig. 10(a), assuming an OSNR of 17.46 dB/b (1 dB above sensitivity at a BER of 10^{-9}) and a laser beat linewidth of $\Delta\nu T_{\mathrm{b}} = 7.5 \times 10^{-6}$. A filter length of $L = 50$ was selected. The two curves shown are the upper and lower bounds for BER. As expected, there is a tradeoff in Δ with this carrier-recovery setup. For Δ greater than eight, $\hat{\theta}_{k-\Delta}$ becomes an increasingly worse estimator of θ_{k+1}, so the higher order terms in (42) become significant, and the lower and upper bounds for P_b diverges. Using the BER upper bound as a conservative assumption, a delay of $\Delta = 8$ is optimum, and the resulting BER is 2.0×10^{-10}. The phase-error standard deviations at the soft- and hard-decision phase-estimator outputs [(40) and (41)] are shown in Fig. 12(b). As expected, $\Delta = \lfloor L - 1/2 \rfloor$ optimizes the performance of the hard-decision phase estimator while $\Delta = 0$ optimizes the performance of the soft-decision phase estimator.

In Fig. 13, we plot the performances of 4-, 8-, and 16-QAMs versus linewidth assuming the use of the two-filter carrier-receiver structure, as shown in Fig. 10(b). The length of the filter was determined as per (28) with $f = 0.05$. Fig. 13(a) assumes a receiver OSNR of 1 dB above sensitivity at a target BER of 10^{-9}, while Fig. 13(b) assumes an OSNR of 1 dB above sensitivity at a target BER of 10^{-3} (see column three of Table I). In each plot, the dotted and solid curves are the lower and upper bounds for BER. We observe that while the BER upper bound in Fig. 13(a) can be made smaller than the target value of 10^{-9} for small enough linewidth, the same is not true for a high target BER, such as in Fig. 13(b). This is because, at low OSNR, soft-decision errors occur so frequently that error propagation cannot be neglected. Catastrophic failure of the soft-decision phase estimator causes the theoretical BER upper bound to lie

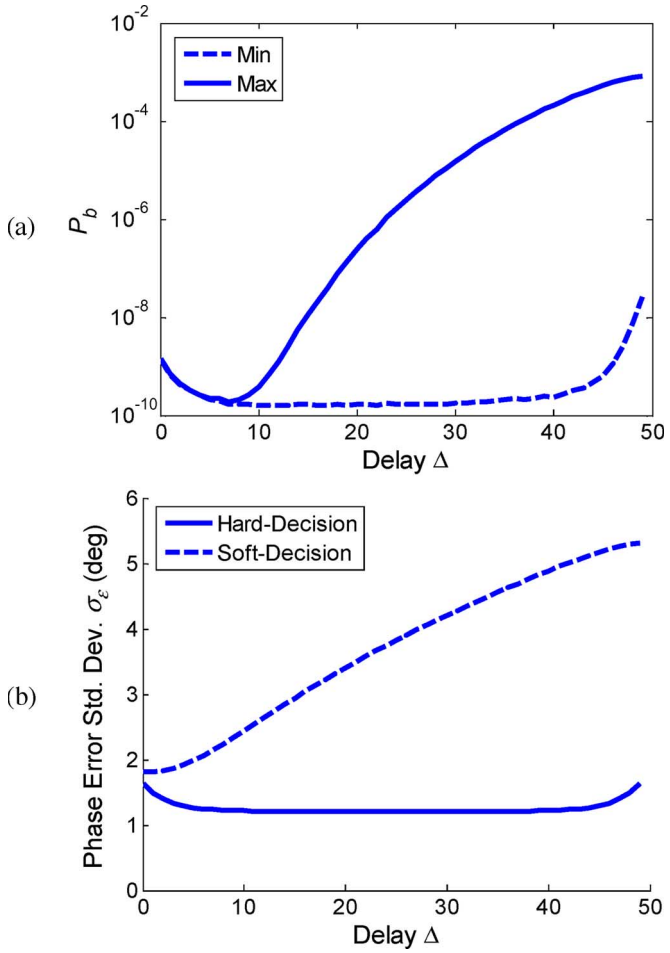

Fig. 12. (a) Probability of bit-error and (b) phase-error standard deviation at the soft- and hard-decision phase-estimator outputs for 16-QAM at OSNR $=$ 17.46 dB/b and $\Delta\nu T_{\rm b} = 7.5 \times 10^{-6}$. A single-filter structure is assumed, and $W(z)$ is a FIR filter of length $L = 50$.

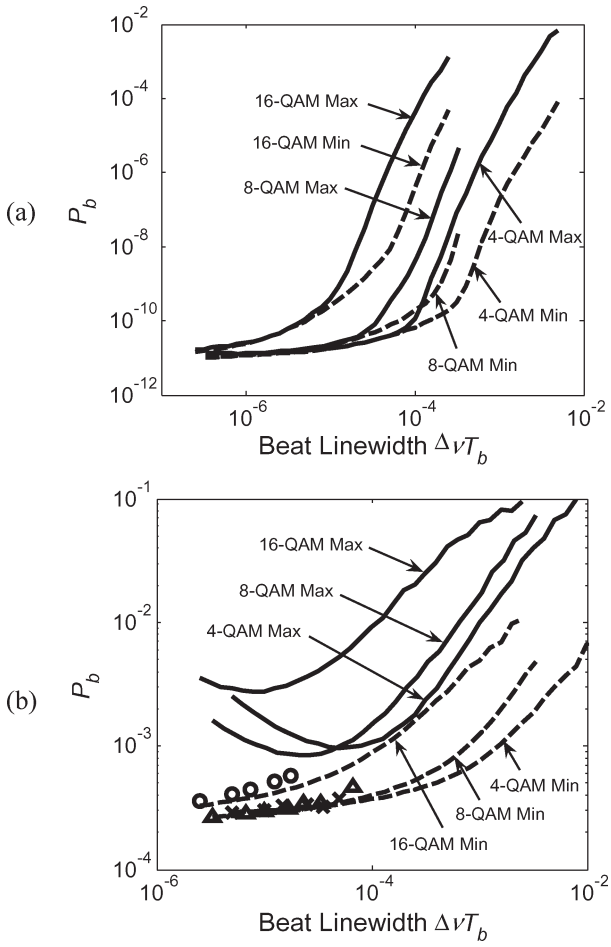

Fig. 13. BER versus linewidth for 4-, 8-, and 16-QAM. The dotted and solid curves are the theoretical lower and upper bounds of the BER. An OSNR of 1 dB above sensitivity is assumed for a target BER of (a) 10^{-9} and (b) 10^{-3}. In (b), the "x," "Δ," and "o" are the simulated BER for 4-, 8-, and 16-QAM.

above the target value for all linewidths. It is also observed in Fig. 13(b) that the BER upper bound actually increases at low linewidths. This is because, according to (28), a longer filter length L is used. The consequence is that multiple soft-decision errors within the observation window becomes more likely, and the "higher order terms" in (42) actually increases.

To confirm the occurrence of catastrophic failure of the soft-decision phase estimator, we performed Monte Carlo simulations for our carrier-recovery scheme (see Section IV). The results are shown as the "x," "Δ," and "o" in Fig. 13(b) for 4-, 8-, and 16-QAMs. One million symbols were used in each of these simulations. When the BER gets above $\sim 3 \times 10^{-4}$, catastrophic failure occurred, causing all subsequent bits in the simulation to be in error. We were, therefore, unable to obtain data points for larger linewidths. Our theoretical BER model outlined in (42), thus, fails at this point.

F. Catastrophic Failure of the Soft-Decision Phase Estimator

In our simulations, we found that, whenever catastrophic failure occurred, $\hat{\theta}$ always locked onto a new phase that is an integer multiple of $\pi/2$ from the correct phase. Although this result is difficult to prove, it is intuitively sensible since 4-, 8-, and 16-QAMs are invariant to rotation by angles that are integer multiples of $\pi/2$. Thus, phase offsets of these values are stable operating points for our system. Catastrophic failure of the soft-decision phase estimator thus manifests as cycle slips. We can avert the problem of catastrophic bit errors if we employ differential-bit encoding. Consider the bits-to-symbols mappings shown in Fig. 14. We divide the constellation into four quadrants. All symbols in the same quadrant are assigned the same most significant bits (MSB) $\rm b_0$ and $\rm b_1$. To uniquely identify each symbol within a quadrant, we employ a Gray-code map for the least significant bits (LSB)—e.g., for 16-QAM, the four points are labeled in a counterclockwise fashion starting from the innermost point. Provided that the LSB labels are properly rotated for the four quadrants, cycle slips will not affect the correct detection of the LSB. To prevent error propagation for the MSB, we can differentially encode $\rm b_0$ and $\rm b_1$.

Cycle slipping not only affects feedforward carrier recovery but also for PLLs at low OSNR. Differential-bit encoding can also be used in PLL-based systems to mitigate catastrophic failure. An unintended consequence of the bits-to-symbols mappings used in Fig. 14 is that they increase the ratio between bit errors and symbol errors (when no cycle slip occurs). In the

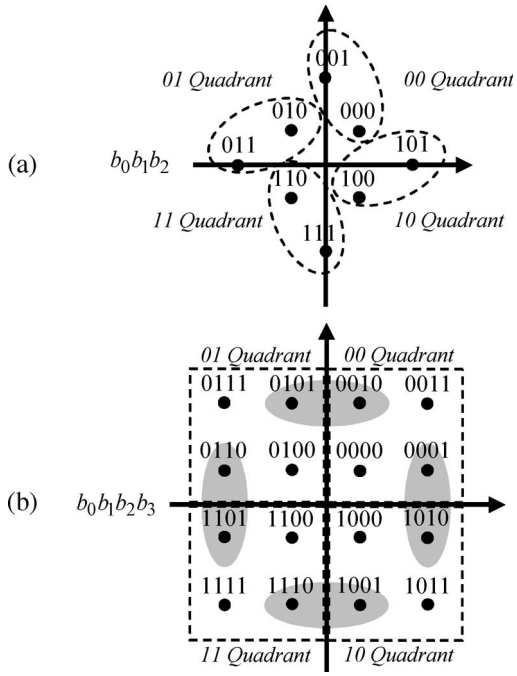

Fig. 14. Differential-bit encoding for (a) 8- and (b) 16-QAM.

16-QAM constellation, as shown in Fig. 14(b) for instance, there are four pairs of neighbors (shown shaded) whose bit labels differ by three bits. This occurs because our bit-labeling scheme is no longer a Gray code when we force the LSB to exhibit rotational invariance. In addition, differential encoding of the MSB also doubles the effective BER for b_0 and b_1 whenever a symbol-detection error occurs that does not produce a cycle slip. Provided that cycle slips occur infrequently as compared to single-detection errors, we can count only the latter in our estimation of BER. Thus, for both Gray-coded and differential-bit-encoded QAM, we have

$$P_{\rm b} = n_{\rm b} P_e \tag{45}$$

where $P_{\rm b}$ and P_e are the probabilities of bit error and symbol error, respectively. The values of $n_{\rm b}$ for different-modulation formats are shown in Table II. The price of employing differential-bit encoding to mitigate catastrophic failure is higher BER compared to Gray coding with no cycle slips. At a BER of 10^{-3}, the power penalties for differential bit-encoded 4-, 8-, and 16-QAM are 0.57, 0.27, and 0.39 dB, respectively.

In Fig. 15, we simulated the achievable BER versus linewidth for 4-, 8-, and 16-QAM using the differential-bit-encoding scheme discussed. Catastrophic bit-error propagation was successfully averted. We were able to detect the transmitted bits at BERs well above 10^{-2} over a simulation period of one million symbols. The OSNR for this simulation was 1 dB above the receiver sensitivity for BER $= 10^{-3}$ assuming no differential-bit encoding (column two of Table III). The target BER is indeed achieved for small enough linewidth. If the 1-dB power penalty was fully allocated to phase-noise mitigation, linewidths of $\Delta\nu T_{\rm b} = 1.3 \times 10^{-4}$, 1.3×10^{-4}, and 1.5×10^{-5} can be tolerated for 4-, 8-, and 16-QAMs using a DD soft-decision phase estimator, and $\Delta\nu T_{\rm b} = 8 \times 10^{-5}$

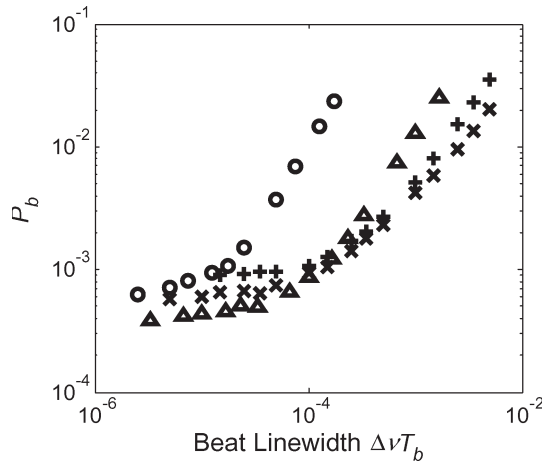

Fig. 15. BER versus linewidth for differential-bit-encoded QAM. "x," "Δ," and "o" are the simulated BER for 4-, 8-, and 16-QAM when a DD soft-decision phase estimator is used. "+" is the simulated BER for 4-QAM when an NDA soft-decision phase estimator is used.

TABLE III
COMPARISON OF THE LINEWIDTH REQUIREMENT FOR FEEDFORWARD CARRIER RECOVERY AND FOR A PLL

Format	OSNR per bit (dB)	Max. Tolerable σ_ε for BER = 10^{-3}	Max. Linewidth using a PLL ($\Delta\nu T_b$)	Max. Linewidth using Feedforward ($\Delta\nu T_b$)
4-QAM (NDA)	7.79	4.91°	4.9×10^{-5}	8.0×10^{-5}
4-QAM (DD)			6.9×10^{-5}	1.3×10^{-4}
8-QAM (DD)	10.03	5.01°	8.3×10^{-5}	1.3×10^{-4}
16-QAM (DD)	11.52	2.70°	7.9×10^{-6}	1.5×10^{-5}

is acceptable for 4-QAM using an NDA soft-decision phase estimator.

We compare the phase-noise tolerance of feedforward carrier recovery with a PLL in Table III. Column two lists the OSNRs used in the simulations of Fig. 15. We make the optimistic assumption for the PLL that symbol errors occur independently. By ignoring burst errors caused by cycle slips, we can compute the value of $\sigma_{\varepsilon,\max}$ that yields the target BER of 10^{-3}, and the results are shown in column three.[14] The maximum tolerable linewidth for the PLL is then given by [1]

$$(\Delta\nu T_{\rm b})_{\max} = \frac{\sigma^4_{\varepsilon,\max} 4\zeta^2 \gamma_{\rm b}}{(1+4\zeta^2)2\pi\eta} \tag{46}$$

where $\gamma_{\rm b}$ is the OSNR per bit, ζ is the damping factor of the second-order PLL, η is the quantity defined in (16) or (20), and $(\Delta\nu T_{\rm b})_{\max}$ is the maximum tolerable linewidth-to-bit-rate ratio listed in column four. The corresponding linewidths found for feedforward carrier recovery in Fig. 15 are shown in column five. Even with our optimistic assumptions for the PLL, the realistic linewidth requirement of feedforward carrier recovery is still 50%–100% better for a target BER of 10^{-3}. If, on the other hand, we use phase-error variance as our criteria for evaluation (Fig. 9 and Table I), then feedforward is up to five times better.

[14]This requires solving $P_{\rm b,target} = n_{\rm b} P_e(N_0, 0, \sigma^2_{\varepsilon,\max})$, where $P_e(N_0, 0, \sigma^2_{\varepsilon,\max})$ is given by (70).

G. Frequency Mismatch

In the previous sections, frequency lock was assumed between the signal and LO lasers. This allowed θ_k to be modeled as a Wiener process in (8). If we allow a frequency offset $\delta f = f_{\mathrm{TX}} - f_{\mathrm{LO}}$ between the two lasers, θ_k becomes a biased random walk with a mean phase change of $2\pi\delta fT$ per symbol. We can write the soft-decision phase estimates as

$$\psi_k = (2\pi\delta fkT + \xi) + \theta'_k + n'_k \tag{47}$$

where θ'_k is the Wiener process previously considered. The carrier phase to be tracked is then $\theta_k = (2\pi\delta fkT + \xi) + \theta'_k$, and n'_k is the impact of AWGN corrupting our phase measurement, whose variance is given by (16) or (20).

We can use the same analysis as Section III-C to find the Wiener filter $W(z)$ that minimizes phase mse, assuming that δf is known. For the FIR approximation, it can be shown that the conditional pdf of $\mathbf{\Psi}$, given $\theta_{k-\Delta}$, is the Gaussian distribution given by (30), and its autocorrelation matrix has the form

$$\mathbf{K} = \mathbf{K_{\delta f}} + \mathbf{K_p} + \mathbf{K_n}. \tag{48}$$

The matrices $\mathbf{K_p}$ and $\mathbf{K_n}$ are the same as (32) and (33), while the frequency-offset matrix has elements:

$$(\mathbf{K_{\delta f}})_{mn} = (2\pi\delta fT)^2(\Delta - m)(\Delta - n). \tag{49}$$

Using this adjusted form of $\mathbf{K}$, the FIR Wiener filter coefficients can again be found using (34). The phase-error variance at the hard-decision phase-estimator output can similarly be derived by taking δf into account

$$\sigma_\varepsilon^2(\mathbf{W},\Delta) = \left[(2\pi\delta fT)^2\left(\sum_{l=0}^{L-1} w_l(\Delta - l)\right)^2\right] + \sigma_p^2\cdot\left[\sum_{m'=0}^{\Delta-1}\left(\sum_{l'=0}^{m'} w_{l'}\right)^2 + \sum_{m'=\Delta+1}^{L-1}\left(\sum_{l'=m'}^{L-1} w_{l'}\right)^2\right] + \sigma_{n'}^2\cdot\left[\sum_{l=0}^{L-1} w_l^2\right]. \tag{50}$$

In comparison with (39), the additional first term is caused by frequency offset. The extra phase-error variance is proportional to δf^2; thus, a nonzero value δf degrades performance in general unless $\Delta = (L-1)/2$. This special case arises because the frequency-offset term is proportional to the square of the first moment of the filter coefficients about Δ. In a symmetric filter, the extra phase of $2\pi\delta fT(\Delta - l)$ at symbol $k - l$ is exactly cancelled by a phase of $2\pi\delta fT(l - \Delta)$ at symbol $k - 2\Delta + l$. Hence, when a two-filter carrier-recovery structure [Fig. 10(b)] is used, frequency offset will not impact the performance of the hard-decision phase estimator. For the soft-decision phase estimator, however, where the optimal Δ is zero, frequency offset will degrade performance.

Consider Fig. 16 where we plot phase-error standard deviation versus δf for 16-QAM at an OSNR of 17.46 dB/b (1 dB above sensitivity at a BER of 10^{-9}) and a laser beat linewidth of

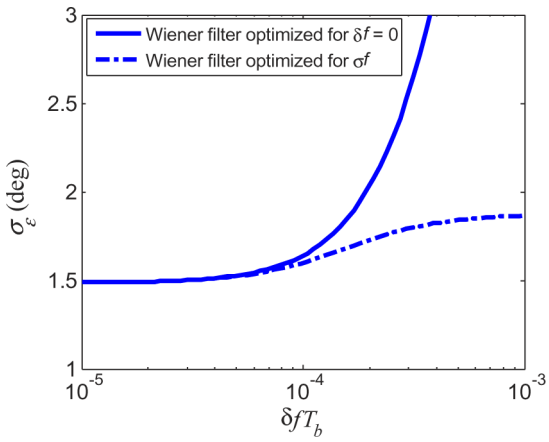

Fig. 16. Phase-error standard deviation versus frequency offset for 16-QAM at $\gamma_{\mathrm{b}} = 17.46$ dB and $\Delta\nu T_{\mathrm{b}} = 5\times10^{-6}$. The filter length is 15, and the coefficients are optimized for a delay (Δ) of zero.

$\Delta\nu T_{\mathrm{b}} = 5\times10^{-6}$. The length of the soft-decision filter is $L = 15$ [half the value of (28)] and the delay Δ is zero. The solid curve is the phase-error standard deviation that results when the filter coefficients are optimized without taking frequency offset into account [assumes $\delta f = 0$ when invoking (48) and (34)]. The dotted curve results when the filter coefficients are optimized for the correct δf. We observe that the increase in σ_ε is much slower for the dotted curve.

Since uncompensated frequency offset leads to poorer performance, we can measure δf and take its value into account when computing the filter coefficients (dotted curve of Fig. 16). Alternatively, we can use the modified receiver shown in Fig. 17. The latter approach is superior since, according to (50) and the dotted curve shown in Fig. 16, a nonzero δf always leads to a higher σ_ε. In addition, removing frequency offset prior to carrier recovery allows us to use fixed coefficients for $W(z)$. The numerical oscillator in Fig. 17 multiplies the incoming symbols by $e^{-j2\pi\delta\hat{f}kT}$, where $\delta\hat{f}$ is the receiver's estimate of the instantaneous frequency offset. Although the feedback structure looks like a PLL, its operating principles are different as the numerical oscillator only needs to track slow frequency drifts associated with temperature or mechanical variations. The rapid fluctuations in frequency caused by phase noise are tracked by the feedforward carrier-recovery structure. The design of the frequency-offset compensator is, therefore, much less stringent than in a PLL.

Estimating Frequency Offset: Consider the difference between soft estimates of the carrier phase separated by L' symbols. We have

$$\psi_{k+L'} - \psi_k = 2\pi\delta fL'T + \sum_{m=1}^{L'}\nu_m + \left(n'_{k+L'} - n'_k\right). \tag{51}$$

The first term in (51) is a deterministic phase rotation proportional to δf. The remaining terms are related to phase noise and AWGN, both having zero mean. We may therefore construct the estimator

$$\delta\hat{f} = \frac{\psi_{k+L'} - \psi_k}{2\pi L'T}. \tag{52}$$

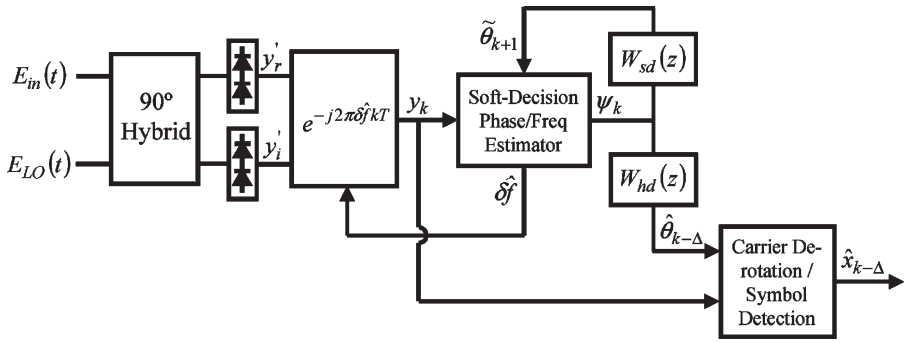

Fig. 17. Coherent receiver with frequency offset compensator (only one polarization is shown).

The frequency estimator in (52) has a mean of δf, and its variance is

$$\left\langle \left(\delta\hat{f} - \langle\delta\hat{f}\rangle\right)^2 \right\rangle = \frac{L'\sigma_p^2 + 2\sigma_{n'}^2}{(2\pi L'T)^2}. \tag{53}$$

For large L', the variance of $\delta\hat{f}$ decreases as $1/L'$: This occurs because phase noise is a sluggish process. By lengthening the duration between measurements, the AWGN and phase-noise terms in (51) become negligible in comparison with the linear growth in ψ_k due to frequency offset. Although it appears beneficial to make L' arbitrarily large, in practice, the benefit of doing so is offset by the time-varying nature of δf due to temperature or mechanical fluctuations. There exists a tradeoff between tracking ability and accurate estimation of δf. Optimization of L' will depend on the rate at which δf evolves.

However, it is possible to establish benchmarks for L'. Suppose the magnitude of the frequency-offset term in (50) is to be no more than 1/10th the value of σ_ε^2 when $\delta f = 0$, there is a maximum frequency estimation error $\delta f_{e,\max}$ that the receiver can tolerate. In the example shown in Fig. 16, σ_ε at $\delta f = 0$ is 1.49°. If we allow frequency offset to increase the phase-error variance by 10%, we can tolerate $\sigma_\varepsilon = 1.56°$. From the solid curve in Fig. 16, this occurs at $\delta f_{e,\max} T_\mathrm{b} = 7.0 \times 10^{-5}$. Suppose we establish the criterion that $\delta\hat{f}$ must be within $\delta f_{e,\max}$ of the true frequency offset 99% of the time. This requires the variance of $\delta\hat{f}$ to satisfy $\sqrt{\langle(\delta\hat{f} - \langle\delta\hat{f}\rangle)^2\rangle} < \delta f_{e,\max}/2.45$. For our example, this leads to $\langle(\delta\hat{f} - \langle\delta\hat{f}\rangle)^2\rangle < 8.2 \times 10^{-10}/T_\mathrm{b}^2 = 1.3 \times 10^{-8}/T^2$. Substituting this into (53) and using values of σ_p^2 and $\sigma_{n'}^2$ corresponding to $\Delta\nu T_\mathrm{b} = 5 \times 10^{-6}$ and SNR = 17.46 dB/b, we get $L' = 299$ symbol periods. This corresponds to an update period of 119 ns at 10 Gb/s, which is significantly shorter than the kilohertz fluctuations expected for mechanical vibrations.

IV. Simulation Results

In Fig. 18, we simulated 16-QAM transmission at an OSNR of 11.52 dB/b (1 dB above sensitivity at a BER of 10^{-3}) at a laser beat linewidth of $\Delta\nu T_\mathrm{b} = 1.5 \times 10^{-5}$. We employed the two-filter phase-estimator structure in Fig. 10(b) with $W_\mathrm{sd}(z)$ and $W_\mathrm{hd}(z)$, both are FIR filters. The lengths of these filters were chosen as per (28) with $f = 0.05$, resulting in $L_\mathrm{hd} = 40$ and $\Delta_\mathrm{hd} = 19$, and $L_\mathrm{sd} = 20$ and $\Delta_\mathrm{sd} = 0$, respectively. Fig. 18(a) shows the simulated carrier phase, and Fig. 18(b) shows the received symbols y_k. It is clear that the received symbols cannot be detected at low BER without carrier derotation. After employing our carrier-recovery algorithm, the hard-decision phase estimate $\hat{\theta}_k$ and the derotated symbols s_k are shown in Fig. 18(c) and (d). Using the differential-bit-encoding scheme discussed in Section III-G, a BER of 1.0×10^{-3} was recorded over a simulation period of 50 000 symbols (200 000 b). The phase-error standard deviation measured at the hard-decision phase-estimator output was 2.10°, which is in agreement with the theoretical value of 2.03° predicted by (41). The phase-error standard deviations measured for $\widetilde{\theta}_k$ and ψ_k (3.07° and 7.51°) were also in agreement with theoretical values (2.98° and 7.39°).

In Fig. 19, we simulated 4-QAM transmission at an OSNR of 7.79 dB/b (1 dB above sensitivity at a BER of 10^{-3}) at a laser beat linewidth of $\Delta\nu T_\mathrm{b} = 8 \times 10^{-5}$. An NDA soft-decision phase estimator was used so only one filter $W(z)$ was required [see Fig. 10(a)]. We chose a FIR filter with $L = 37$ [(28) with $f = 0.05$] and $\Delta = 18$. Fig. 19(a)–(d) shows the simulated carrier phase, the received symbols, the hard-decision phases, and the derotated symbols, respectively. The difference between Fig. 19(a) and (c) is due to cycle slips. Our differential-bit-encoding scheme successfully prevented catastrophic bit-error propagation, and a BER of 1.0×10^{-3} was obtained over the simulation period of 50 000 symbols (100 000 b). The tight clustering of points observed in the derotated constellation shows that the feedforward carrier-phase estimator always reacquired lock after a cycle-slip event. To measure the phase-error standard deviation at the hard-decision phase-estimator output, we take the difference between $\hat{\theta}_k$ and θ_k modulo $\pi/2$. The measured value of 3.71° agrees with the theoretical value of 3.56°.

V. Discussion

Two major implementation issues for feedforward carrier recovery are the number of bits required for the ADC and the system complexity. In practical systems, the use of forward error correction allows the uncoded BER to be around 10^{-3}. From Table I, the OSNR requirement for 16-QAM is 11.52 dB/b (16.54 dB/symbol). Suppose that after allowances are made for CD/PMD compensation, nonlinear effects, and carrier recovery, the OSNR becomes 20 dB. In the example provided in [8], 8-b resolution for the ADC is sufficient to ensure

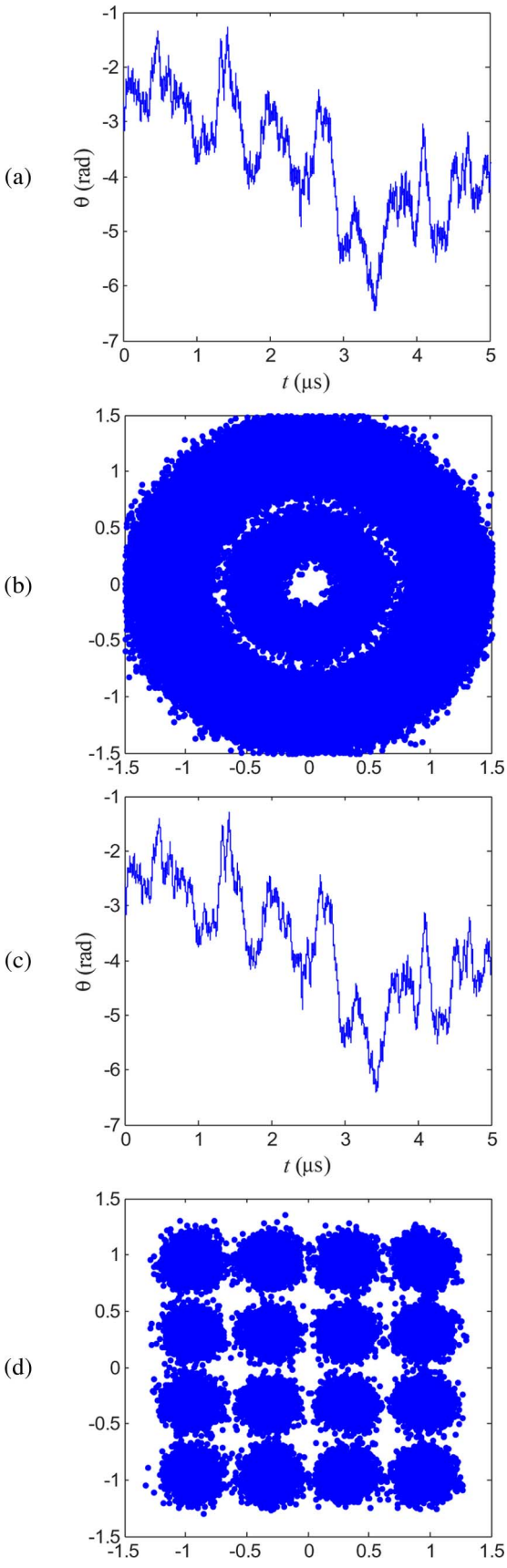

Fig. 18. Simulation of 16-QAM transmission at OSNR $= 11.52$ dB/b and a linewidth of $\Delta\nu T_{\mathrm{b}} = 1.5 \times 10^{-5}$. (a) Simulated carrier phase θ_k. (b) Received symbols y_k. (c) Recovered carrier phase $\hat{\theta}_k$. (d) Symbols after carrier derotation s_k.

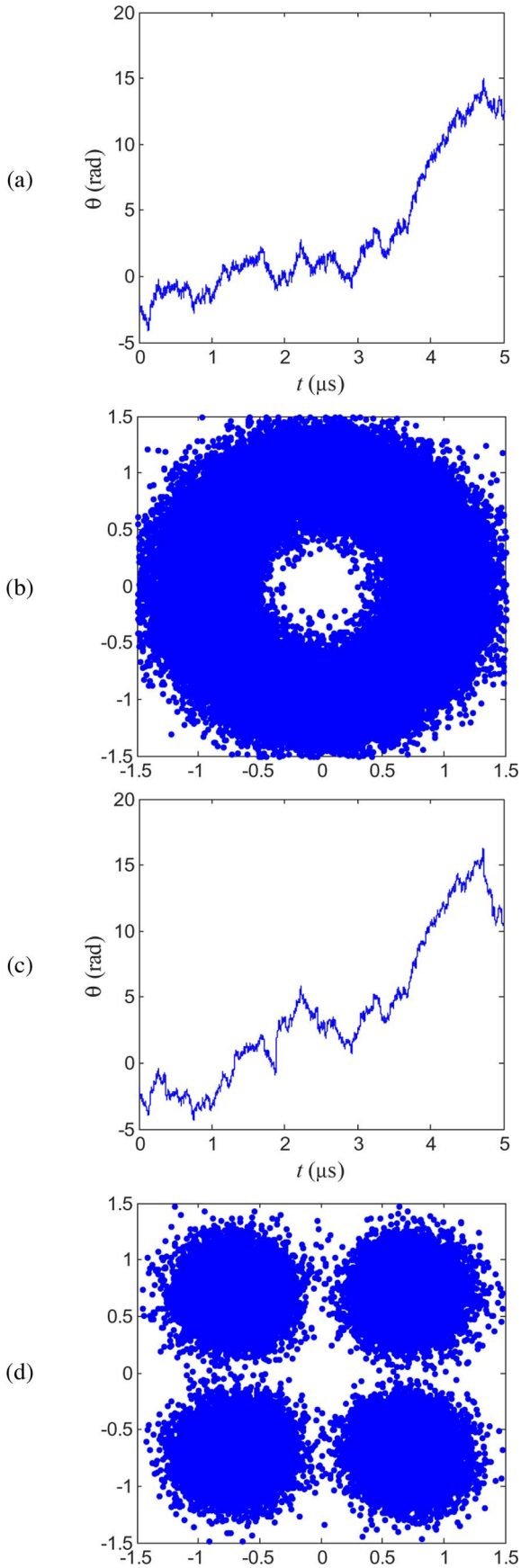

Fig. 19. Simulation of 4-QAM transmission at OSNR $= 7.79$ dB/b and a linewidth of $\Delta\nu T_{\mathrm{b}} = 8 \times 10^{-5}$. (a) Simulated carrier phase θ_k. (b) Received symbols y_k. (c) Recovered carrier phase $\hat{\theta}_k$. (d) Symbols after carrier derotation s_k.

the quantization noise variance is less than 10% of the AWGN variance per sample, even when we consider the modulation scheme with the largest dynamic range: 33%-return-to-zero 16-QAM.

System complexity can be determined by counting the number of operations required to implement Figs. 4 and 10. If an NDA soft-decision phase estimator is used in conjunction with 4-QAM, the input quadrupler in Fig. 4(a) requires three complex multiplications (CMs). If a DD soft-decision phase estimator is used, regardless of the modulation format, the symbol derotation takes one CM. Both the decision device and the $\arg(\cdot)$ operations can be implemented using lookup tables. The phase-unwrapping formula given by (22) requires one real multiplication (RM). Counting each CM as four RMs, the complexity of the soft-decision phase estimator is 13 RMs for the NDA design and five RMs for the DD design. The most numerically intensive operation in Fig. 10 are the two filters $W_{\rm sd}(z)$ and $W_{\rm hd}(z)$. In the example given in Fig. 18 for 16-QAM, $L_{\rm hd}$ and $L_{\rm sd}$ needed to be 40 and 20, respectively, yielding a complexity of 60 RMs per symbol. The carrier derotation operation at the output of Fig. 10 requires one further CM. Hence, the total complexity of the 16-QAM example is $5 + 60 + 4 = 69$ RMs per-symbol per polarization. This is approximately half the number of operations needed for CD/PMD compensation, where in [8], we showed that 60 CMs (240 RMs) are sufficient to compensate 2000 km of single-mode fiber at 10-GHz symbol rate for two polarizations, yielding 120 RMs per symbol. Achieving such performance would require a VLSI circuit with extensive pipelining and parallelization.

VI. Conclusion

We have demonstrated the feasibility of feedforward carrier recovery for coherent optical communications. Our proposed receiver consists of a soft-decision phase-estimation stage followed by hard-decision estimation of the carrier phase and the transmitted symbols. Two implementations of the soft-decision phase estimator were proposed—these were the NDA phase estimator and the DD phase estimator. The former is suitable for M-PSK transmission formats, while the latter is suitable for any general constellation. Although the DD algorithm is more flexible, errors in the soft-decision stage can lead to error propagation by the presence of feedback. In our simulations, we showed that the propagation of soft-decision errors lead to cycle slip. Catastrophic bit-error propagation can be avoided with an appropriately designed bits-to-symbols encoding scheme in conjunction with differential-bit encoding.

We modeled phase noise as a Wiener process and showed that the optimal hard-decision phase estimator is a linear filter, which can be approximated by a FIR filter with a sufficient number of taps. We investigated the effect of a nonzero frequency offset between the transmitter and LO lasers and found that uncompensated frequency offset leads to higher phase-error variance. To mitigate this, we proposed a numerical oscillator be used to track the slow components of laser frequency drift caused by temperature fluctuation or mechanical vibrations, while letting the feedforward structure to compensate for rapidly varying phase noise.

Our simulation results show that feedforward carrier recovery can tolerate at least 1.5–2 times larger linewidth than a PLL even when we used an optimistic assumption for computing the maximum tolerable linewidth for a PLL. If the criterion for comparison is phase-error variance, feedforward carrier recovery is up to five times better. The increased linewidth tolerance enables the use of presently available ECLs for 16-QAM transmission at 10 Gb/s.

Appendix A
Phase-Error Variance of Soft-Decision Phase Estimator

NDA Phase Estimator

Consider the noise signal m_k corrupting the soft-decision phase estimator's estimate of θ_k in (14)

$$m_k = \sum_{p=1}^{M} \binom{M}{p} (x_k e^{j\theta_k})^{M-p} n_k^p. \tag{54}$$

Although the terms in this summation are not Gaussian in general except for $p = 1$, in the limit of high SNR, the first term dominates. If we assume m_k to be circularly Gaussian, we overestimate the tail probability in its pdf so our prediction for system BER will be slightly pessimistic. It can be shown that the m_k's are i.i.d. with zero mean, with variance given by

$$E\left[|m_k|^2\right] = \sum_{p=1}^{M}\sum_{q=1}^{M} \binom{M}{p}\binom{M}{q} \times E\left[(xe^{j\theta_k})^{M-p}(x^* e^{-j\theta_k})^{M-q}\right] E\left[n_k^p\,(n_k^*)^q\right]. \tag{55}$$

Consider the expected value of the signal correlation term in (55)

$$E\left[(xe^{j\theta_k})^{M-p}(x^* e^{-j\theta_k})^{M-q}\right] = P_t^{(2M-p-q)/2} \cdot E\left[e^{j(q-p)\varphi_k}\right] \cdot E\left[e^{j(q-p)\theta_k}\right] \tag{56}$$

where $\varphi_k = 2\pi m/M$, $m \in \{0, \ldots, M-1\}$ is the phase of the transmitted signal. Since raising an M-PSK signal to an integer power produces another zero-mean M-PSK signal unless the exponent is an integer multiple of M, in our summation, we have $E[e^{j(q-p)\varphi_k}] = \delta_{pq}$, and (55) simplifies to

$$E\left[m_k^2\right] = \sum_{p=1}^{M} \binom{M}{p}^2 P_t^{M-p} \cdot E\left[|n_k|^{2p}\right] = P_t^M \sum_{p=1}^{M} \binom{M}{p}^2 E\left[|\tilde{n}_k|^{2p}\right] \tag{57}$$

where $\tilde{n}_k = n_k/\sqrt{P_t}$ is a circular Gaussian variable with zero mean and variance $1/\gamma$, where γ is the SNR per symbol. The

variance of m_k is, therefore, a sum of the even moments of a Gaussian distribution, which is given by

$$E\left[|n_k|^{2p}\right] = (p!)\frac{1}{\gamma^p}. \tag{58}$$

Finally, we recall that the soft-phase estimator computes

$$\psi_k = \frac{1}{M}\arg\{e^{jM\theta_k} + m_k\} = \theta_k + n_k'. \tag{59}$$

Provided that the receiver SNR is sufficiently high, such that the amplitude of m_k is small compared to $e^{jM\theta_k}$, the phase uncertainty, according to Fig. 5(a), is

$$n_k' \approx \frac{1}{M}\mathrm{Im}\left\{\frac{m_k}{x^M e^{jM\theta_k}}\right\}. \tag{60}$$

The n_k' are, thus, approximately i.i.d. Gaussian variables with zero mean and variance

$$\sigma_{n'}^2 = \frac{1}{2M^2}\sum_{p=1}^{M}\binom{M}{p}^2 p!\frac{1}{\gamma^p} = \eta(M,\gamma)\frac{1}{\gamma} \tag{61}$$

where the multiplicative factor η is a function of M and γ. As an example, 4-QAM transmission at 1 dB above sensitivity for BER $= 10^{-9}$ ($\gamma = 45.3$ [1]) has $\eta(M,\gamma) = 0.552$.

DD Phase Estimator

Provided that the decision device in Fig. 4(b) recovers the correct symbol $\hat{x}_k$, according to Fig. 5(b), the phase uncertainty in ψ_k is given by

$$n_k' \approx \mathrm{Im}\left\{\frac{n_k}{xe^{j\theta_k}}\right\}. \tag{62}$$

The n_k' are, therefore, approximately i.i.d. Gaussian variables with zero mean and variance

$$\sigma_{n'}^2 = \frac{1}{2}E\left[\frac{|n_k|^2}{|x|^2}\right] = \eta\frac{1}{\gamma} \tag{63}$$

where $\gamma = E[|x|^2]/E[|n_k|^2]$ is the SNR per symbol, and $\eta = (1/2)E[|x|^2]E[1/|x|^2] = (1/2)\eta_c$ is one half of the "constellation penalty" defined in [2]. For constant amplitude-modulation formats (M-PSK), $\eta_c = 1$. It can be shown that, for 8- and 16-QAM, $\eta_c = 1.5$ and 1.889, respectively.

Appendix B
Conditional Probability of y_k in the Presence of AWGN and Phase Error

Consider the received signal after carrier derotation

$$ye^{-j\hat{\theta}} = xe^{j(\theta-\hat{\theta})} + n'e^{-j\hat{\theta}}. \tag{64}$$

Since n' has a circular Gaussian distribution, phase rotation does not change its statistics. In addition, we showed in Section III-C that phase error $\varepsilon = \theta - \hat{\theta}$ is Gaussian distributed with mean μ_ε and variance σ_ε^2. The conditional pdf of y is, therefore

$$\begin{aligned} p\left(y|x,N_0,\mu_\varepsilon,\sigma_\varepsilon^2\right) &= \int_{-\infty}^{+\infty}\frac{1}{\pi N_0}\exp\left(-\frac{|ye^{-j\varepsilon}-x|^2}{N_0}\right) \\ &\quad\times\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left(-\frac{(\varepsilon-\mu_\varepsilon)^2}{2\sigma_\varepsilon^2}\right)d\varepsilon \\ &= \frac{1}{\pi N_0}\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left(-\frac{|x|^2+|y|^2}{N_0}\right) \\ &\quad\times\int_{-\infty}^{+\infty}\exp(p\cos(\varepsilon')) \\ &\quad\times\exp\left(-\frac{(\varepsilon'+\theta_y-\theta_x-\mu_\varepsilon)^2}{2\sigma_\varepsilon^2}\right)d\varepsilon' \end{aligned} \tag{65}$$

where $p = 2|x||y|/N_0$, and $\varepsilon' = \varepsilon - \theta_y + \theta_x$. Since $\exp(p\cos(\varepsilon'))$ is periodic with period 2π, it can be expressed as

$$\exp\left(p\cos(\varepsilon')\right) = \frac{a_0}{2} + \sum_{k=1}^{\infty}a_k\cos(k\varepsilon') \tag{66}$$

where the Fourier coefficients are

$$a_k = \frac{1}{\pi}\int_{-\pi}^{+\pi}\exp\left(p\cos(\varepsilon')\right)\cos(k\varepsilon')d\varepsilon' = 2I_k(p) \tag{67}$$

where $I_k(p)$ is the kth-order modified Bessel function of the first kind. Substituting (66) into (65) and using [19]

$$\begin{aligned} &\int_{-\infty}^{+\infty}e^{-(ax^2+bx+c)}\cos(kx)dx \\ &\qquad= \sqrt{\frac{\pi}{a}}\exp\left(\frac{b^2-ac-k^2/4}{a}\right)\cos\left(-\frac{bk}{a}\right) \end{aligned} \tag{68}$$

we can write the conditional pdf for y_k in the presence of AWGN and phase error as an infinite Bessel series

$$\begin{aligned} p\left(y|x,N_0,\mu_\varepsilon,\sigma_\varepsilon^2\right) &= \frac{1}{\pi N_0}\exp\left(-\frac{|x|^2+|y|^2}{N_0}\right) \\ &\quad\times\left[I_0(p) + 2\sum_{k=1}^{\infty}I_k(p)\exp\left(-\frac{k^2\sigma_\varepsilon^2}{2}\right)\right. \\ &\quad\left.\times\cos\left(-k(\theta_y-\theta_x-\mu_\varepsilon)\right)\right]. \end{aligned} \tag{69}$$

To compute the probability of symbol error $P_e(N_0,\mu_\varepsilon,\sigma_\varepsilon^2)$ and the probability that the decision device at the soft-decision phase estimator makes an error of $\Delta\varepsilon$, i.e., $P(\Delta\varepsilon|N_0,\mu_\varepsilon,\sigma_\varepsilon^2)$,

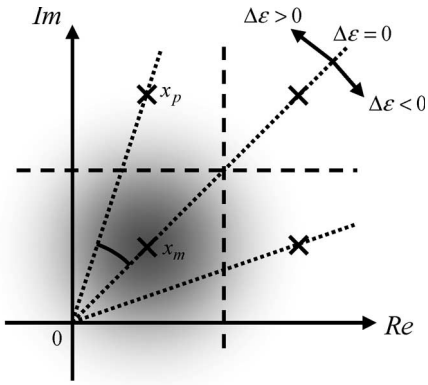

Fig. 20. Decision regions for the computation of the probability of symbol error and the probability that the angle of the detected symbol differs from that of the transmitted symbol by $\Delta\varepsilon$.

we integrate (69) over appropriate decision regions. Let $\mathbf{D}_x$ be the decision region associated with symbol x (Fig. 20). We have

$$P_e\left(N_0, \mu_\varepsilon, \sigma_\varepsilon^2\right) = \frac{1}{M} \sum_{x_m} \int_{y \notin \mathbf{D}_{x_m}} p\left(y|x_m, \mu_\varepsilon, \sigma_\varepsilon^2\right) dy \qquad (70)$$

where the summation is over all possible transmitted symbols, where M is the alphabet size (we have assumed that all symbols are equally likely to be transmitted). We can similarly express $P(\Delta\varepsilon|N_0, \mu_\varepsilon, \sigma_\varepsilon^2)$ by modifying the region of integration to be the union of the decision regions of all symbols x_p whose angle differs with that of the transmitted symbol x_m by $\Delta\varepsilon$. If we define $\mathbf{D}_{x_m}^{\Delta\varepsilon} \in \bigcup_{x_p:\arg\{x_p\}-\arg\{x_m\}=\Delta\varepsilon} \mathbf{D}_{x_p}$, we have

$$P\left(\Delta\varepsilon|N_0, \mu_\varepsilon, \sigma_\varepsilon^2\right) = \frac{1}{M} \sum_{x_m} \int_{y \in \mathbf{D}_{x_m}^{\Delta\varepsilon}} p\left(y|x_m, \mu_\varepsilon, \sigma_\varepsilon^2\right) dy. \qquad (71)$$

We note that $P_e(N_0, \mu_\varepsilon, \sigma_\varepsilon^2) \neq P(\Delta\varepsilon = 0|N_0, \mu_\varepsilon, \sigma_\varepsilon^2)$ in general, as there may exist more than one signal point sharing the same angle, e.g., the innermost and outermost points of the same quadrant in 16-QAM have the same angle.

ACKNOWLEDGMENT

The authors would like to thank D.-S. Ly-Gagnon for the discussions.

REFERENCES

[1] J. R. Barry and J. M. Kahn, "Carrier synchronization for homodyne and heterodyne detection of optical quadriphase-shift keying," *J. Lightw. Technol.*, vol. 10, no. 12, pp. 1939–1951, Dec. 1992.

[2] E. Ip and J. M. Kahn, "Carrier synchronization for 3- and 4-bit-per-symbol optical transmission," *J. Lightw. Technol.*, vol. 23, no. 12, pp. 4110–4124, Dec. 2005.

[3] E. Ip, J. M. Kahn, D. Anthon, and J. Hutchins, "Linewidth measurements of MEMS-based tunable lasers for phase-locking applications," *IEEE Photon. Technol. Lett.*, vol. 17, no. 10, pp. 2029–2031, Oct. 2005.

[4] S. Tsukamoto, K. Katoh, and K. Kikuchi, "Unrepeated transmission of 20-Gb/s optical quadrature phase-shift-keying signal over 200-km standard single-mode fiber based on digital processing of homodyne-detected signal for group-velocity dispersion compensation," *IEEE Photon. Technol. Lett.*, vol. 18, no. 9, pp. 1016–1018, May 1, 2006.

[5] S. Tsukamoto, K. Katoh, and K. Kikuchi, "Coherent demodulation of optical multilevel phase-shift-keying signals using homodyne detection and digital signal processing," *IEEE Photon. Technol. Lett.*, vol. 18, no. 10, pp. 1131–1133, May 15, 2006.

[6] M. G. Taylor, "Coherent detection method using DSP for demodulation of signal and subsequent equalization of propagation impairments," *IEEE Photon. Technol. Lett.*, vol. 16, no. 2, pp. 674–676, Feb. 2004.

[7] M. Secondini, E. Forestieri, and G. Prati, "PLC optical equalizer for chromatic and polarization-mode dispersion compensation based on MSE control," *IEEE Photon. Technol. Lett.*, vol. 16, no. 4, pp. 1173–1175, Apr. 2004.

[8] E. Ip and J. M. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *J. Lightw. Technol.*, vol. 25, no. 8, pp. 2033–2043, Aug. 2007.

[9] M. G. Taylor, "Accurate digital phase estimation process for coherent detection using a parallel digital processor," in *Proc. ECOC*, Sep. 2005, vol. 2, pp. 263–264.

[10] D.-S. Ly-Gagnon, S. Tsukamoto, K. Katoh, and K. Kikuchi, "Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation," *J. Lightw. Technol.*, vol. 24, no. 1, pp. 12–21, Jan. 2006.

[11] K. Kikuchi, "Coherent detection of phase-shift keying signals using digital carrier-phase estimation," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2006.

[12] H. Xu, Q. Huang, and H. Zheng, "On the blind carrier phase estimation for large QAM constellation," in *Proc. 17th Int. Conf. AINA*, 2003, pp. 551–554.

[13] L. Chen and H. Kusaka, "Blind phase recovery in QAM communication systems using higher order statistics," *IEEE Signal Process. Lett.*, vol. 3, no. 5, pp. 147–149, May 1996.

[14] C. Georghiades, "Blind carrier phase acquisition for QAM constellations," *IEEE Trans. Commun.*, vol. 45, no. 11, pp. 1477–1486, Nov. 1997.

[15] M. G. Taylor, "Measurement of phase diagrams of optical communication signals using sampled coherent detection," in *Proc. Symp. Opt. Fiber Meas. Tech. Dig.*, 2004, pp. 163–166.

[16] R. Noe, "Phase noise-tolerant synchronous QPSK/BPSK baseband-type intradyne receiver concept with feedforward carrier recovery," *J. Lightw. Technol.*, vol. 23, no. 2, pp. 802–808, Feb. 2005.

[17] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers*. New York: Wiley, 1997.

[18] V. K. Prabhu, "PSK performance with imperfect carrier phase recovery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-12, no. 2, pp. 275–285, Mar. 1976.

[19] I. S. Gradshteyn, I. M. Ryzhik, and A. Jeffrey, *Table of Integrals, Series and Products*, 5th ed. New York: Academic, Jan. 1994.

[20] D.-S. Ly-Gagnon, "Information recovery using coherent detection and digital signal processing for phase-shift-keying modulation formats in optical communication systems," M.S. thesis, Univ. Tokyo, Tokyo, Japan, 2004.

Ezra Ip received the B.E. degree (with honors) in electrical and electronics engineering from the University of Canterbury, Christchurch, New Zealand, in 2002 and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 2004, where he is currently working toward the Ph.D. degree in electrical engineering.

In 2002, he was a Research Engineer with Industrial Research Ltd., New Zealand. His research interests include single-mode optical-fiber communications, free-space optical communications, and nonlinear optics.

Joseph M. Kahn (M'90–SM'98–F'00) received the A.B., M.A., and Ph.D. degrees in physics from the University of California, Berkeley, in 1981, 1983, and 1986, respectively.

From 1987 to 1990, he was with AT&T Bell Laboratories, Crawford Hill Laboratory, Holmdel, NJ. He demonstrated multigigabits-per-second coherent optical-fiber transmission systems, setting world records for receiver sensitivity. From 1990 to 2003, he was a faculty member with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, performing research on optical and wireless communications. In 2000, he helped found StrataLight Communications, where he served as Chief Scientist from 2000 to 2003. Since 2003, he has been a Professor of electrical engineering with Stanford University, Stanford, CA. His current research interests include single- and multimode optical-fiber communications, free-space optical communications, and microelectromechanical systems for optical communications.

Prof. Kahn was the recipient of the National Science Foundation Presidential Young Investigator Award in 1991. From 1993 to 2000, he served as a Technical Editor of *IEEE Personal Communications Magazine.*