Spectral Efficiency of Coded Phase-Shift Keying for Fiber-Optic Communication

Gerhard Kramer, *Member, IEEE*, Alexei Ashikhmin, *Member, IEEE*, Adriaan J. van Wijngaarden, *Member, IEEE*, and Xing Wei, *Member, OSA*

***Abstract*—Several optical modulation and detection schemes are compared by computing their spectral efficiencies over additive white Gaussian noise channels. The bandwidth savings of differential quadrature phase-shift keying (D-QPSK) over both direct-detection on–off keying and differential binary phase-shift keying suggest that D-QPSK can improve the reach and efficiency of wavelength-division multiplexing systems. To test the theory, Reed–Solomon and low-density parity-check forward error correction codes are designed and evaluated. The codes generally behave as expected, except that for D-QPSK the gains are hampered by the differential detector. It is further shown that neither multiple-symbol differential detection nor decision-feedback detection is attractive when using strong codes.**

***Index Terms*—Error correcting coding, optical signal detection, optical transmitters, phase-shift keying.**

I. Introduction

DIFFERENTIAL binary phase-shift keying (D-BPSK) has recently attracted renewed attention for use in ultra-long-haul optical wavelength-division multiplexed (WDM) systems [1]. The main reason is that D-BPSK gives a substantial (≈2.5 dB) detection energy gain over direct-detection on–off keying (DD-OOK) when using a delay interferometer and a balanced receiver for differential detection. Furthermore, D-BPSK tends to suppress fiber nonlinear penalties such as those due to cross-phase modulation better than OOK [1]–[4]. Although D-BPSK is an old technique (see, e.g., [5, p. 527]), the importance of its combined linear and nonlinear gains has been recognized only recently.

This paper investigates the performance of several coded modulation and detection schemes that are suited for high-speed fiber-optic communication. For instance, Table I compares the spectral efficiencies of DD-OOK, D-BPSK, and differential quadrature PSK (D-QPSK) on an additive white Gaussian noise (AWGN) channel. All schemes have a data rate of 40 Gb/s. Also, all schemes use *noncoherent* receivers having only one decision threshold per detector arm, also known as *hard* detection.

Consider the first, second, and seventh rows of numbers in Table I. The DD-OOK and D-BPSK schemes use a 6.7% overhead code at a symbol rate of 42.7 GHz. The D-QPSK scheme uses a 14.3% overhead code at a symbol rate of 22.9 GHz. The "energy gain" refers to how much energy is saved over either DD-OOK or D-BPSK when capacity-achieving codes are used, and a 3-dB gain means that half the energy is required. We see that D-QPSK requires 2.41 dB less energy than DD-OOK, and a mere 0.05 dB more energy than D-BPSK, but with 46% less bandwidth. This suggests that D-QPSK can improve the efficiency of WDM systems.

TABLE I
COMPARISON OF MODULATIONS FOR 40-Gb/s TRANSMISSION

Modulation	Code Rate	Symbol Rate [GHz]	Energy Gain [dB] vs. DD-OOK	Energy Gain [dB] vs. D-BPSK
DD-OOK	15/16	42.7	–	−2.46
D-BPSK	15/16	42.7	2.46	–
D-QPSK	15/32	42.7	4.43	1.97
	1/2	40.0	4.37	1.91
	2/3	30.0	3.83	1.37
	3/4	26.7	3.40	0.94
	7/8	22.9	2.41	−0.05
	15/16	21.3	1.52	−0.94

There are many practical issues to resolve before one can realize the gains of Table I. For example, the gains are for extremely long capacity-achieving forward error correction (FEC) codes. We address this issue in Section V and show that the gains remain intact for practical codes. The main exception is that the D-QPSK differential detector causes additional losses if the code is not close to optimal. A second important issue is that our analysis does not include nonlinear fiber effects. Thus, our conclusions will apply primarily to transmission in a linear regime, e.g., low power or short distances. We concentrate on PSK to keep the uniform envelope advantages of D-BPSK.

This paper is organized as follows. In Section II we derive and compute the spectral efficiencies of several modulation and detection schemes. This section organizes known results and should be considered a review of communications theory for optical high-speed transmission where one uses noncoherent and hard detection (see, e.g., [6, pp. 282–284, 380–387]). This theoretical work was done in conjunction with the experiments of [1]. Section III discusses the behavior of spectral efficiency at low signal-to-noise ratios. Section IV presents results on multiple-symbol differential detectors and decision-feedback detectors, and suggests that these are not so useful when combined with strong codes. Section V considers the performance of two code families used in practice: Reed–Solomon (RS) codes and low-density parity-check (LDPC) codes. The latter are decoded with a suboptimal but high-speed and low-power decoder. Section VI summarizes our results.

Manuscript received November 22, 2002; revised June 18, 2003.

The authors are with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: gkr@bell-labs.com; aea@bell-labs.com; alw@bell-labs.com; xingwei@bell-labs.com).

Digital Object Identifier 10.1109/JLT.2003.817704

0733-8724/03$17.00 © 2003 IEEE

II. Modulation/Detection Spectral Efficiency

The spectral efficiency of a modulation set is the number of bits per second per hertz that the set can support. A first-order approach is to compute spectral efficiency without FEC coding [7]. However, in practice, codes are a necessity so one should take them into account. To define spectral efficiency, we follow an approach similar to [8, Section VIII] and [5, Section 7.2] and make the following idealizations.

i) The channel passes frequencies f (measured in hertz) in the range
$$f_0 - W/2 < |f| < f_0 + W/2$$
where the center frequency f_0 is much larger than W. The equivalent baseband representations of the channel input and output signals are therefore complex and bandlimited to $|f| < W/2$. We say that the channel bandwidth is W.

ii) The channel is used for a period of T seconds. We describe the channel input and output signals by TW complex samples spaced $1/W$ seconds apart, i.e., we can reconstruct the baseband input signal $X(t)$ as
$$X(t) = \sum_{\ell} X_\ell \cdot \frac{\sin(\pi W t - \pi \ell)}{\pi W t - \pi \ell}$$
where X_ℓ is a complex number whose squared amplitude and phase are the respective energy and phase of the ℓth input sample. The average energy per sample is
$$E_s = \sum_{\ell} \mathrm{E}[|X_\ell|^2]/(TW).$$
We will consider primarily PSK for which $|X_\ell|^2 = E_s$ for all ℓ.

iii) The channel is modeled as adding complex Gaussian noise of energy N to each input sample, i.e., the ℓth channel output sample is $Y_\ell = X_\ell + Z_\ell$, where $Z_\ell = Z_{R\ell} + jZ_{I\ell}$ and $Z_{R\ell}$ and $Z_{I\ell}$ are independent Gaussian random variables each having variance $N/2$. Usually N increases proportionally with W, and we write $N = WN_0$, where $N_0/2$ is the noise power per hertz.

We remark that the above setup considers transmission using only one polarization of light. A model that includes both polarizations can be found, e.g., in [9].

We define the modulation rate in bits as $R_m = \log_2(M)$, where M is the number of values that X can take on. For instance, three-phase PSK has $R_m = \log_2(3)$ bits and QPSK has $R_m = 2$ bits. We assume that FEC is done with a binary code of rate R_c so the overall rate is $R = R_c R_m$ bits. The energy sent per information bit is $E_b = E_s/R$ and we define the *information bit* signal-to-noise ratio as E_b/N_0. We write the bit-error probability out of the decoder as P_b. It is now natural to define the *spectral efficiency* of the modulation as
$$\eta(E_b/N_0, P_b) = R_c^* R_m \text{ bits/s/Hz} \tag{1}$$
where R_c^* is the *maximum code rate* for which one can achieve a bit-error probability of P_b when the information bit signal-to-noise ratio is E_b/N_0. For optical transmission, P_b can be required to be 10^{-16} or less. We will in fact consider only the case $P_b \to 0$ because the spectral efficiency hardly changes from that when $P_b = 10^{-16}$. We denote the resulting spectral efficiency by $\eta(E_b/N_0)$.

Note that (1) defines spectral efficiency without taking into account the spectral guard bands that are necessary in practice. The reason for doing this is that the amount of guard band needed, in proportion to the bandwidth, will depend primarily on the *pulse shape* and less on the number of points in the signal set. We will not consider pulse shaping here.

A. Computing Spectral Efficiency

Suppose we have chosen a modulation set and a detector, and we use the channel $L = TW$ times. For large T, the system *capacity* is
$$C = \max_{P(\underline{x})} \frac{1}{T} I(\underline{X}; \underline{Y}) \text{ bits/s} \tag{2}$$
where $\underline{X}$ is the length L random vector of channel inputs, $\underline{Y}$ is the random vector of channel outputs, $P(\underline{x})$ is the probability distribution of $\underline{X}$, and $I(\underline{X};\underline{Y}) = \sum P(\underline{x},\underline{y}) \log_2[P(\underline{x},\underline{y})/(P(\underline{x})P(\underline{y}))]$ is the *mutual information* between $\underline{X}$ and $\underline{Y}$ ([10, Section 2.3]). In many cases of interest, the channel is memoryless and the best $P(\underline{x})$ is the one that chooses every symbol of $\underline{x}$ independently. For such cases, we can rewrite (2) as
$$C/W = \max_{P(x)} I(X;Y) \text{ bits/s/Hz} \tag{3}$$
where X and Y are the input and output symbols for one use of the channel. We will often drop the maximization and simply assume that $P(x)$ is a uniform distribution.

The mutual information in (3) is a function of E_s/N, i.e., we have
$$C/W = f(E_s/N) = f((RE_b)/(WN_0)) \tag{4}$$
where $f(\cdot)$ is some nondecreasing function and where we have used $E_s = RE_b$ and $N = WN_0$. Note further that $R \le C$, so that the best C/W is the largest C^*/W satisfying
$$C^*/W = f(C^*/W \cdot E_b/N_0). \tag{5}$$
This equation gives the ultimate E_b/N_0 in terms of C^*/W, or the ultimate C^*/W in terms of E_b/N_0. We thus have
$$\eta(E_b/N_0) = C^*/W \text{ bits/s/Hz}. \tag{6}$$
Some properties of $\eta(E_b/N_0)$ are developed in Section III.

We illustrate how to compute $\eta(E_b/N_0)$ by considering the case where arbitrary complex X are permitted. Since R_m is in principle infinite, one now maximizes the overall rate R without separating it into R_m and R_c. Shannon showed that for this case the capacity is ([11, Section 25])
$$C = W \log_2(1 + E_s/N) \text{ bits/s}. \tag{7}$$
We set $E_s = RE_b, N = WN_0$, and use (7) to obtain [8, Section VIII]
$$\frac{2^\eta - 1}{\eta} \le \frac{E_b}{N_0}. \tag{8}$$
We plot the η that satisfy (8) with equality as the curve labeled "Gauss, soft" in Fig. 1. Observe that as $\eta \to 0$, we have $E_b/N_0 \to \ln(2) \approx 0.6931$, which is -1.59 dB. This ultimate

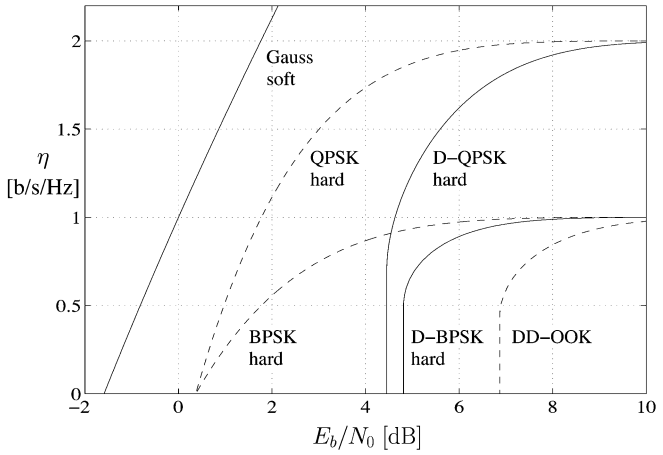


Fig. 1. Spectral efficiencies of various modulations.

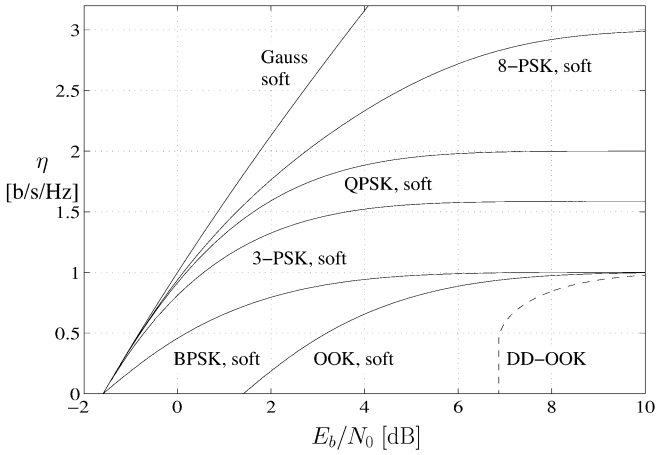


Fig. 2. Spectral efficiencies for soft coherent detection.

energy limit will decrease slowly as P_b increases, but it remains at essentially -1.59 dB for $P_b \leq 10^{-6}$.

The other curves in Fig. 1 are those for various modulation and detection schemes. We review how to compute these curves in the following sections. We are interested mainly in PSK modulation sets, i.e., X takes on the values $x_k = \sqrt{E_s}e^{j2\pi k/M}$, where $k = 0, 1, \ldots, M-1$. We will consider three types of detection schemes: 1) *soft coherent detection*, meaning the detector passes Y to the decoder, 2) *hard coherent detection*, meaning the detector decides which of the M phases was sent and passes this decision to the decoder, and 3) *hard differential detection*, meaning differential detection is used *and* the detector passes one of M decisions to the decoder. We begin, however, with the commonly used direct-detection OOK method for purposes of comparison.

B. Direct-Detection OOK

On–off keying involves transmitting either $X = 0$ or $X = \sqrt{2E_s}$ so that the average energy is E_s. Direct detection means that the decoder sees either $\tilde{Y} = 0$ or $\tilde{Y} = 1$ depending on whether $|X + Z|$ is smaller or larger, respectively, than some threshold τ. The capacity will be a function of the probabilities that the detector puts out $\tilde{Y} = 1$ when $X = 0$, and $\tilde{Y} = 0$ when $X = \sqrt{2E_s}$. These probabilities are the respective (see [6, p. 47])

$$P_{d,0} = e^{-\tau^2/N} \tag{9}$$

$$P_{d,1} = \int_0^{\tau} \frac{2y}{N} e^{\frac{-y^2-2E_s}{N}} I_o\left(\frac{2y\sqrt{2E_s}}{N}\right) dy \tag{10}$$

where $I_o(\cdot)$ is the zeroth-order modified Bessel function of the first kind.

The threshold τ is usually chosen so that $P_{d,0}$ and $P_{d,1}$ are the same value P_d. This type of direct detection creates a binary symmetric channel (BSC) from the encoder output to the decoder input. The capacity of the BSC with crossover probability P_d is known to be (see, e.g., [10, p. 187])

$$f(E_s/N) = 1 - h(P_d) \text{ bits} \tag{11}$$

where $f(\cdot)$ is the same function as in (4) and where $h(p) = -p\log_2(p) - (1-p)\log_2(1-p)$ is the binary entropy function. We now fix E_b/N_0 and numerically find the $\eta = C^*/W$ that satisfies (5). Alternatively, fix C^*/W and find the E_b/N_0 that satisfies (5). The resulting curve is depicted in Fig. 1 as the curve labeled DD-OOK. Observe the sharp rise of this curve from $\eta = 0$ to approximately $\eta = 0.5$. This means that it makes little sense to use a code with rate less than 1/2 with DD-OOK. We discuss this behavior in more detail in Section III.

Consider next OOK with soft coherent detection. The spectral efficiency is easily seen to be 3 dB worse than BPSK with soft coherent detection, so we proceed to consider PSK.

C. PSK With Soft Coherent Detection

The modulation capacity of PSK with soft coherent detection is $I(X;Y)$, where X takes on one of M equally likely phases. We have $I(X;Y) = H(Y) - H(Y \mid X)$, where $H(Y)$ is the entropy of Y and $H(Y \mid X)$ is the conditional entropy of Y given X. We further have $H(Y \mid X) = W\log_2(\pi eN)$ bits/s and (see [6, p. 92] and [11, Section 20, 25])

$$H(Y) = W\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} -p(y)\log_2(p(y))\, dy_R\, dy_I \tag{12}$$

where $y = y_R + jy_I$ and

$$p_{Y\mid X}(y \mid x_k) = \frac{1}{\pi N}\exp\left(-\frac{|y - x_k|^2}{N}\right) \tag{13}$$

$$p_Y(y) = \sum_{k=0}^{M-1} \frac{1}{M} p_{Y\mid X}(y \mid x_k). \tag{14}$$

Recall that $x_k = \sqrt{E_b R}e^{j2\pi k/M}$. We again fix E_b/N_0 and numerically find the $\eta = C^*/W$ that satisfies (5). The results for $M = 2, 3, 4, 8$ are shown in Fig. 2. It is clear that all modulations have a spectral efficiency at most $\log_2(M)$ bits/s/Hz. Furthermore, note that at low rates, BPSK performs worse than all other PSK sets. This behavior is discussed in detail in [12]. The BPSK curve is, in fact, exactly one-half the QPSK curve. Finally, note that OOK with soft detection loses exactly 3 dB as compared to BPSK.

D. PSK With Hard Coherent Detection

Soft detection is preferable to hard detection, but at very high speeds, no A/D (analog-to-digital) converters exist that are fast enough to give multiple-bit resolution. Thus, from here on we

will concentrate on hard detection. This section treats hard *coherent* detection, and the next section treats hard *differential* detection.

We compute the capacity of PSK with hard coherent detection by using $I(X;\hat{Y})$ where the detector output $\hat{Y}$ takes on one of M decisions. We have $I(X;\tilde{Y}) = H(\tilde{Y}) - H(\tilde{Y}\,|\,X)$ and use the symmetry of the modulation to write $H(\tilde{Y}) = W\log_2(M)$ bits/s and

$$H(\tilde{Y}\,|\,X) = -W\sum_{k=0}^{M-1} P_{\tilde{Y}\,|\,X}(\tilde{y}_k\,|\,x_0)\log_2(P_{\tilde{Y}\,|\,X}(\tilde{y}_k\,|\,x_0)) \tag{15}$$

where

$$P_{\tilde{Y}\,|\,X}(\tilde{y}_k\,|\,x_0) = \int_{2\pi(k-0.5)/M}^{2\pi(k+0.5)/M} \left\{\int_0^\infty \frac{1}{\pi N}\cdot\exp\left(-\frac{|re^{j\phi}-\sqrt{E_s}|^2}{N}\right) r\,dr\right\} d\phi. \tag{16}$$

The integral inside the curly brackets is the probability distribution of the phase ϕ and evaluates to (see, e.g., [13, (253)], [14, p. 659], or [15, (68)])

$$P(\phi) = \frac{1}{2\pi}e^{-E_s/N} + \sqrt{\frac{E_s}{\pi N}}\cos(\phi)e^{-(E_s/N)\sin^2(\phi)} \cdot\left[1-Q\left(\sqrt{\frac{2E_s}{N}}\cos(\phi)\right)\right] \tag{17}$$

where $Q(x) = (1/\sqrt{2\pi})\int_x^\infty e^{-t^2/2}\,dt$. Alternatively, we can use the somewhat simpler expression of [14, (1)], [15, (70)], or [16, (5)] to compute

$$P_{\tilde{Y}\,|\,X}(\tilde{y}_k\,|\,x_0) = F\left(\frac{2\pi(k+0.5)}{M}\right) - F\left(\frac{2\pi(k-0.5)}{M}\right) \tag{18}$$

for $k = 1,\ldots,M-1$, where

$$F(\psi) = \frac{-1}{2\pi}\int_0^{\pi-\psi}\exp\left[\frac{-(E_s/N)\sin^2(\psi)}{\sin^2(\theta)}\right]d\theta. \tag{19}$$

Of course, $P_{\tilde{Y}\,|\,X}(\tilde{y}_0\,|\,x_0)$ is simply $1-\sum_{k=1}^{M-1}P_{\tilde{Y}\,|\,X}(\tilde{y}_k\,|\,x_0)$.

For example, for $M=2$ the channel becomes a BSC with crossover probability $P_d = Q(\sqrt{2E_bR/N})$. The normalized capacity of BPSK with hard coherent detection is thus

$$C/W = 1 - h(Q(\sqrt{2E_bR/N})) \text{ bits/s/Hz}. \tag{20}$$

As another example, if $M=4$ and a Gray mapping from the coded bits to the phases is used, i.e., x_0,x_1,x_2,x_3 represent the coded bits 00, 01, 11, 10, respectively, then the channel is equivalent to two uses of a BSC with crossover probability $P_d = Q(\sqrt{E_bR/N})$. The normalized capacity is thus

$$\begin{aligned} C/W &= 2[1-h(Q(\sqrt{E_bR/N}))] \\ &= 2[1-h(Q(\sqrt{2E_bR_c/N}))] \text{ bits/s/Hz} \end{aligned} \tag{21}$$

where we have used $R = 2R_c$. We see that QPSK with Gray mapping is exactly twice as spectrally efficient as BPSK by inserting $R = R_c$ into (20).

As in the previous section, the spectral efficiency of general M-phase PSK with hard coherent detection is computed by numerically finding the relation between $\eta = C^*/W$ and E_b/N_0 in (5). The results for $M = 2,3,4,8$ are shown in Fig. 3. We remark that with hard coherent detection one cannot transmit below an E_b/N_0 of 0.37 dB with either BPSK or QPSK, i.e., hard coherent detection is about 2 dB worse than soft coherent detection for these modulations. However, as a notable difference to soft detection, BPSK outperforms 3-PSK at low rates. This happens because 3-PSK exhibits the same abrupt behavior as DD-OOK. We discuss this behavior in Section III.

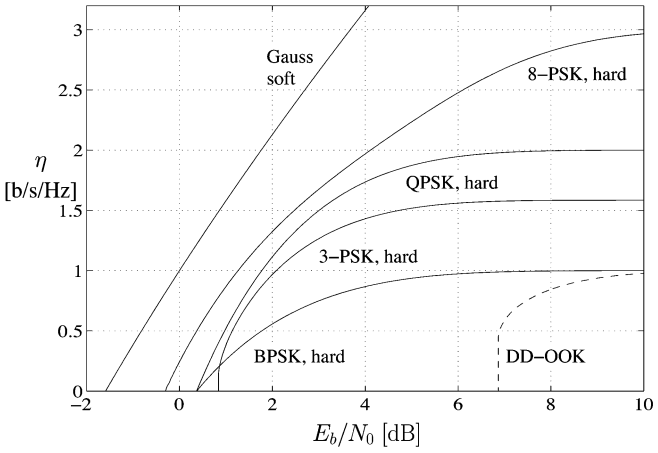


Fig. 3. Spectral efficiencies for hard coherent detection.

E. PSK With Hard Differential Detection

Suppose we employ differential encoding so that $X_\ell = U_\ell X_{\ell-1}$ where $X_0 = \sqrt{E_s}$ and U_ℓ carries the information on one of M phases (with $|U_\ell| = 1$) for $\ell = 1,2,3,\ldots$. A differential detector puts out

$$S_\ell = Y_\ell Y_{\ell-1}^* = (U_\ell X_{\ell-1} + Z_\ell)(X_{\ell-1}+Z_{\ell-1})^* \tag{22}$$

where Y_ℓ is the ℓth channel output and $Y_{\ell-1}^*$ is the complex conjugate of $Y_{\ell-1}$. Suppose the receiver uses a *soft* detector with $Y_0 = 1$. The receiver can now recover Y_ℓ recursively via $Y_\ell = S_\ell Y_{\ell-1}/|Y_{\ell-1}|^2$ for $\ell = 1,2,3,\ldots$. Thus, a soft differential detector can in principle perform as well as a soft coherent detector (see, e.g., [17] and the references therein).

Consider next a *hard* differential detector, which puts out one of M decisions $\tilde{Y}_\ell$ based on S_ℓ. The channel from U_ℓ to $\tilde{Y}_\ell$ is *not* memoryless because S_ℓ is a function of successive channel outputs Y_ℓ and $Y_{\ell-1}$. Thus, one should use (2) rather than (3) to compute η. However, for practical reasons such as encoding and decoding complexity, one usually treats the channel as memoryless. This approach will be accurate if one places a long (random-like) interleaver between the encoder and modulator and the appropriate deinterleaver between the detector and decoder. The deinterleaver breaks the statistical dependence between successive symbols. We will adopt this approach and will therefore compute $I(U;\tilde{Y})$ rather than $I(\underline{U};\underline{\tilde{Y}})$.

We compute $I(U;\tilde{Y})$ by using symmetry to write $H(\tilde{Y}) = W\log_2(M)$ bits/s, and write $H(\tilde{Y}\,|\,U)$ as in (15). However, instead of (16) or (18), we now use (see [15, (11)] or [16, (5)])

$$\begin{aligned} P_{\tilde{Y}\,|\,U}(\tilde{y}_k\,|\,x_0) =& F_d\left(\frac{2\pi(k+0.5)}{M}\right) \\ &- F_d\left(\frac{2\pi(k-0.5)}{M}\right) \end{aligned} \tag{23}$$

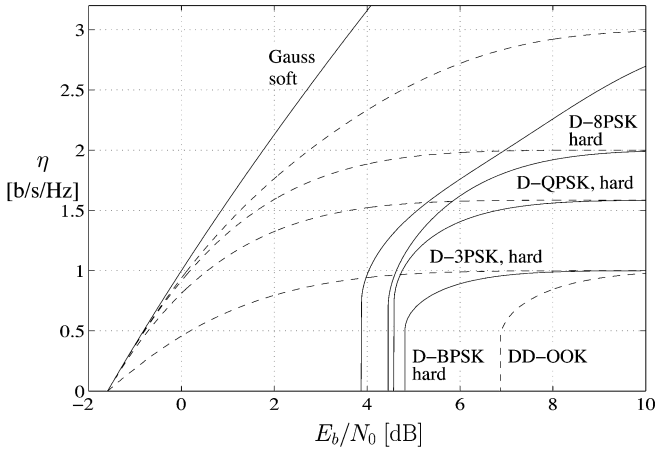


Fig. 4. Spectral efficiencies for hard (solid lines) and soft (dashed lines except for DD-OOK) differential detection.

for $k = 1, \ldots, M-1$, where

$$F_d(\psi) = \frac{-1}{2\pi} \int_0^{\pi-\psi} \exp\left[\frac{-(E_s/N)\sin^2(\psi)}{1+\cos(\psi)\cos(\theta)}\right] d\theta. \quad (24)$$

Again, $P_{\tilde{Y}\,|\,U}(\tilde{y}_0\,|\,x_0)$ is simply $1 - \sum_{k=1}^{M-1} P_{\tilde{Y}\,|\,U}(\tilde{y}_k\,|\,x_0)$.

For example, for $M = 2$ the detector error probability is $P_d = \exp(-E_s/N)/2$ (see [5, p. 527] and [6, p. 276]). The normalized capacity of D-BPSK with hard detection (and ignoring the channel memory) is thus

$$C/W = 1 - h\left(\frac{1}{2}e^{-E_s/N}\right) \text{ bits/s/Hz}. \quad (25)$$

The spectral efficiency for general M is again computed by numerically finding the relation between $\eta = C^*/W$ and E_b/N_0 in (5). The results for $M = 2, 3, 4, 8$ are plotted in Fig. 4 together with the soft detection curves from Fig. 2 as dashed lines. Observe that with hard differential detection, one cannot transmit below an E_b/N_0 of about 4.8 dB for $M = 2$, and about 4.5 dB for $M = 3$ and $M = 4$. Observe also the sharp rise in all the DPSK curves. We further remark that D-QPSK can outperform D-BPSK both in terms of bandwidth and energy efficiency by choosing appropriate code rates R_c. For example, if $E_b/N_0 = 6$ dB, then one could improve the bandwidth efficiency by choosing R_c from one-half the D-BPSK spectral efficiency, which is $R_c \approx 0.9/2 = 0.45$, to one-half the D-QPSK spectral efficiency, which is $R_c \approx 1.6/2 = 0.8$.

Finally, note that hard differential detection is much less efficient than soft coherent detection and, as discussed above, the soft detection curves can be approached by using soft differential detection. This suggests that large energy gains are possible once high-speed analog-to-digital converters with better than 1-bit resolution become available.

III. Spectral Efficiency at Low E_b/N_0

An important parameter is the smallest E_b/N_0 for which $\eta(E_b/N_0) > 0$. Let $\dot{f}(x)$ denote the derivative of $f(\cdot)$ at x. Reference [12, (30)] shows that we can transmit at

$$E_b/N_0 \geq [\dot{f}(0)]^{-1} \quad (26)$$

as long as $\dot{f}(0) > 0$ and $f(\cdot)$ is concave. For example, the soft coherent Gaussian curve of (7) has

$$E_b/N_0 \geq \left[\frac{d}{dx}\log_2(1+x)\Big|_{x=0}\right]^{-1} = \ln(2) \quad (27)$$

and $10 \cdot \log_{10}(\ln(2)) = -1.59$ dB. One gets the same result of -1.59 dB for soft coherent PSK. We can also apply (26) to hard coherent BPSK and QPSK to obtain

$$E_b/N_0 \geq \ln(2)\frac{\pi}{2} \quad (28)$$

i.e., one loses $10\log_{10}(\pi/2) = 1.96$ dB using hard rather than soft detection (see, e.g., [18, p. 155]).

Consider next hard coherent 3-PSK, DD-OOK, and DPSK. For these schemes $f(\cdot)$ is *not* concave so that (26) does not give the right answer [12, Section V.E]. Instead, the minimum E_b/N_0 is found by using the following graphical procedure: 1) plot $f(x)$ against x and 2) draw a line through the origin and tangent to $f(\cdot)$ so that all points of $f(\cdot)$ lie on or below the line. The inverse of the slope of this line gives the minimum E_b/N_0 we are seeking. Furthermore, the size of the η jump at this E_b/N_0 is simply $f(x^*)$, where x^* is the point where the line is tangent to $f(\cdot)$. Observe that if $f(\cdot)$ is concave, this graphical procedure is equivalent to using (26).

The above discussion implies that for nonconcave $f(\cdot)$ the minimum E_b/N_0 is not characterized by $\eta \to 0$ or $E_s/N \to 0$. This is the reason for the sharp rises in the hard 3-PSK, DD-OOK, and DPSK spectral efficiency curves. The consequence of the abrupt behavior is that if one lowers R_c beyond some threshold, then one pays more in the code's $10\log_{10}(R_c)$ energy loss than one can gain back with error control capabilities. Thus, it makes little sense to choose R_c less than this threshold. For instance, it makes little sense to use strong codes with hard D-BPSK and $R_c < 1/2$.

IV. Other Differential Detectors

We discuss two techniques for improving a differential detector. The first is *multiple-symbol* differential detection [21, Fig. 3], [22], i.e., the hard decisions $\hat{Y}_\ell$ are functions of more than just $S_\ell = Y_\ell Y_{\ell-1}^*$. For example, [22, (20)] uses the detector

$$S_\ell = Y_\ell Y_{\ell-1}^* \hat{U}_\ell^* + Y_{\ell-1} Y_{\ell-2}^* \hat{U}_{\ell-1}^* + Y_\ell Y_{\ell-2}^* \hat{U}_{\ell-1}^* \hat{U}_\ell^* \quad (29)$$

where $\hat{U}_\ell$ is a decision about the ℓth phase change U_ℓ. Observe that (29) requires two types of differential outputs: $Y_\ell Y_{\ell-1}^*$ and $Y_\ell Y_{\ell-2}^*$. These outputs could be realized with two delay interferometers: one having a delay of one symbol and the other having a delay of two symbols.

According to [22], (29) achieves a detector gain of about 1 dB for D-QPSK and for large E_b/N_0 (see also [21, Table I]). However, [22, Fig. 5] shows that this gain disappears at low E_b/N_0 where $P_d \geq 10^{-2}$. This means that using (29) will not increase the spectral efficiency by much at rates where $R_c \leq 1 - h(0.01) \approx 0.92$. We do emphasize, however, that this statement is valid only if one uses *strong* codes. When using weak codes, or even no coding at all, one will in fact get an energy gain of 1 dB over uncoded D-QPSK because one must operate at large E_b/N_0.

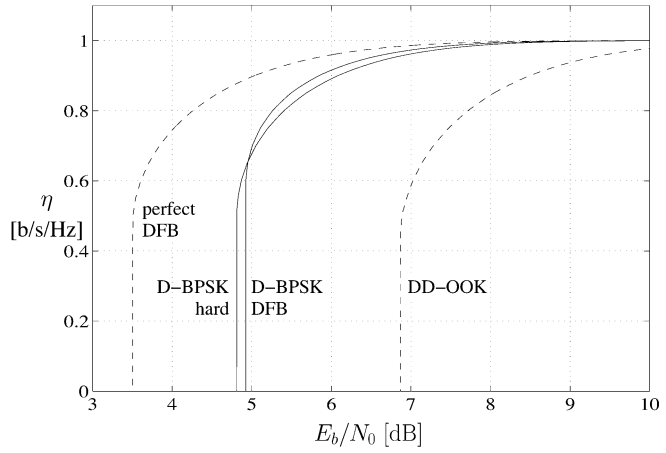


Fig. 5. Spectral efficiencies for decision-feedback hard differential detection.

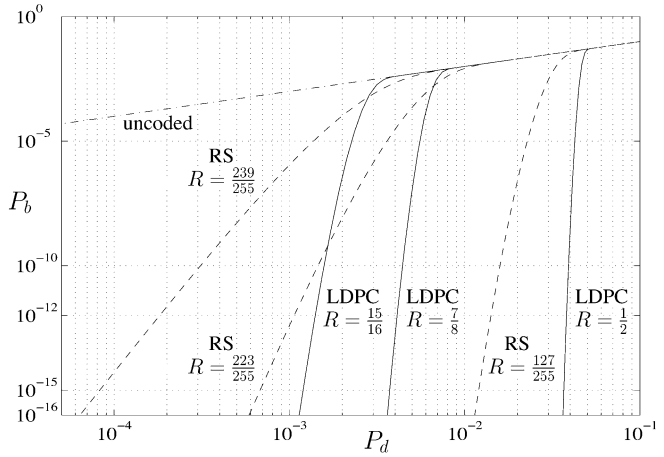


Fig. 6. Code bit-error probability transfer curves.

A second approach related to multiple-symbol differential detection is *decision-feedback* detection (see [21] and [23]). However, as we proceed to show, the advantages of this technique are also small if one uses strong codes. The simplest decision-feedback structure of [21] puts out the decision statistic

$$S_\ell = Y_\ell (Y_{\ell-1} + Y_{\ell-2}\hat{U}_{\ell-1})^* \tag{30}$$

where $\hat{U}_{\ell-1}$ is computed from $S_{\ell-1}$ only. In other words, $\hat{U}_{\ell-1}$ is the $(\ell-1)$st hard decision that is fed back to provide a better phase estimate than $Y_{\ell-1}$ alone. Observe that if $\hat{U}_{\ell-1} = U_{\ell-1}$, then the term in brackets in (30) is

$$\begin{aligned}(X_{\ell-1} + Z_{\ell-1}) + (X_{\ell-2} + Z_{\ell-2})U_{\ell-1} \\ = 2X_{\ell-1} + (Z_{\ell-1} + Z_{\ell-2}U_{\ell-1}).\end{aligned} \tag{31}$$

Note that (31) has twice the signal-to-noise ratio of $Y_{\ell-1}$ alone. Thus, as long as $\hat{U}_{\ell-1}$ is correct, (30) will give a less noisy estimate of U_ℓ. However, if $\hat{U}_{\ell-1}$ is incorrect, one will have a poor phase reference. For instance, if $M = 2$, the term in brackets in (30) becomes

$$\begin{aligned}(X_{\ell-1} + Z_{\ell-1}) + (X_{\ell-2} + Z_{\ell-2})(-U_{\ell-1}) \\ = Z_{\ell-1} - Z_{\ell-2}U_{\ell-1}\end{aligned} \tag{32}$$

i.e., the phase reference is completely noisy. In this case $\hat{U}_\ell$ will be in error exactly one-half the time. For $M > 2$ the error analysis is similar, although somewhat more complicated.

We continue to consider the case $M = 2$. Let $P_d = \Pr(\hat{U}_\ell \neq U_\ell)$ be the raw bit-error rate out of the detector, and let $q = \Pr(\hat{U}_\ell \neq U_\ell \,|\, \hat{U}_{\ell-1} = U_{\ell-1})$. We compute

$$\begin{aligned}P_d &= \Pr(\hat{U}_\ell \neq U_\ell) \\ &= \Pr(\hat{U}_{\ell-1} = U_{\ell-1})q + \Pr(\hat{U}_{\ell-1} \neq U_{\ell-1})\frac{1}{2} \\ &= (1 - P_d)q + P_d/2.\end{aligned} \tag{33}$$

We thus have $P_d = 2q/(1+2q)$ and, ignoring the channel memory as in (25), we have

$$C/W = 1 - h(2q/(1+2q)) \text{ bits/s/Hz}. \tag{34}$$

We can compute q using the techniques of [15] or [16]. More precisely, we use [16, (5)] with $\rho_1 = 2E_s/N, \rho_2 = E_s/N$ and $\psi = \pi/2$ to obtain

$$q = \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} \exp\left[\frac{-4(E_s/N)}{3 + \sin(\theta)}\right] d\theta. \tag{35}$$

The numerical results are plotted in Fig. 5. We see that the decision-feedback scheme performs virtually the same as D-BPSK. Fig. 5 also shows the hypothetical curve if all the feedback decisions are correct. The large loss of imperfect feedback is not surprising because at low rates the noise energy is large and the decisions are incorrect with high probability. In summary, the decision-feedback detector (30) holds little promise if one uses *strong* codes. At the same time, we emphasize that when using weak codes, or no coding, it might very well make sense to use (30) rather than the usual DPSK detector.

V. Comparison of Codes

The spectral efficiencies of Figs. 1–4 are ultimate limits on performance if one uses capacity-approaching codes. Although practical codes cannot exactly reach these limits, the *relative* comparisons between the modulation and detection schemes will hopefully be accurate when one uses a particular code or code family. We consider two families of codes: RS codes [24] and LDPC codes [25], [26].

The RS codes have parameters (n, k), where n is the number of 8-bit symbols in a code word and k is the number of 8-bit symbols containing information bits. We consider RS codes with parameters (255, 239), (255, 223), and (255, 127). The code words thus have 2040 bits and the respective rates are 0.937, 0.875, and 0.498. We assume that all error patterns with $t = \lfloor (n-k+1)/2 \rfloor$ or fewer symbol errors are corrected and that no miscorrections occur. The last assumption is optimistic, but pessimistic bounds show the approximation to be accurate.

The LDPC codes have parameters (n, k), where n is now the number of *bits* in a code word and k is the number of information bits. We consider three length 32 000 LDPC codes whose rates are 15/16, 7/8, and 1/2. The codes are iteratively decoded, and we determine their performance by a combination of simulation and analysis.

Fig. 6 plots the decoder output bit-error rate P_b versus the decoder input bit-error rate P_d. Table II lists the corresponding operating points when $P_b = 10^{-16}$. For example, uncoded D-BPSK and D-QPSK lose 0.29 and 2.33 dB, respectively,

TABLE II
CODE OPERATING POINTS: P_d AND E_b/N_0 FOR $P_b = 10^{-16}$.

Code	R_c	P_d	E_b/N_0 [dB] BPSK	D-BPSK	D-QPSK	DD-OOK
uncoded	1	10^{-16}	15.29	15.58	17.62	18.47
LDPC	1/2	0.0363	5.08	7.20	7.66	9.49
	7/8	0.00362	6.15	7.51	8.59	10.02
	15/16	0.00114	6.96	8.12	9.37	10.69
RS	0.498	0.0114	7.16	8.80	9.64	11.23
	0.875	$5.84\cdot10^{-4}$	7.80	8.88	10.20	11.47
	0.937	$6.35\cdot10^{-5}$	8.94	9.81	11.32	12.49

versus coherent BPSK. These numbers are close to the asymptotic degradations given in [27, p. 253]

$$10\log_{10}\left[\frac{\sin^2(\pi/M)}{2\sin^2(\pi/M/2)}\right]\ \text{dB}. \tag{36}$$

Inserting $M = 2$ and $M = 4$ into (36), we find that the respective asymptotic losses are 0 and 2.32 dB. However, these losses degrade to the respective 1.16 and 2.41 dB at the $P_d = 0.001\,14$ operating point of the rate 15/16 LDPC code. In fact, uncoded D-QPSK outperforms uncoded D-BPSK for very high P_d. This is why D-QPSK is superior to D-BPSK at low rates.

A. Code Spectral Efficiencies

We plot the operating points of our codes in Fig. 7. There are six sets of code curves corresponding to the two code families and the three modulations D-BPSK, D-QPSK, DD-OOK. We have also included some of the ultimate operating points from Table I. Observe that these points (even the LDPC points) are rather far from the ultimate operating curves. The main reason is that we cannot use optimal message passing iterative decoders for high decoder speeds such as 10 and 40 Gb/s.

In the following, we will compare a few of the operating points of Fig. 7. Many further comparisons can be made using Fig. 6 and Table II. Suppose we choose as our "standard" the $R_c = 15/16$ LDPC code with DD-OOK. We see from Fig. 7 and Table II that this combination achieves $P_b = 10^{-16}$ at a symbol rate of 42.7 GHz and a signal-to-noise ratio of $E_b/N_0 = 10.69$ dB. Alternatively, one could use D-BPSK to operate at $E_b/N_0 = 8.12$ dB, thereby saving 2.57 dB in energy (this does not include the observed fiber nonlinearity gains of D-BPSK).

Suppose next that one uses D-QPSK with the $R_c = 7/8$ LDPC code. One can now send at a symbol rate of 22.9 GHz and $E_b/N_0 = 8.59$ dB, i.e., one uses 46% less bandwidth than our standard and still gains 2.10 dB in energy. Moreover, the bandwidth savings over D-BPSK are also 46% but one sacrifices 0.47 dB in energy. This 0.47 dB loss is 0.42 dB greater than the 0.05-dB loss of Table I. The main factor contributing to the 0.42 dB is a 0.36-dB loss because a D-QPSK detector performs progressively worse than a D-BPSK detector as E_b/N_0 increases. This means that a better code will achieve a double

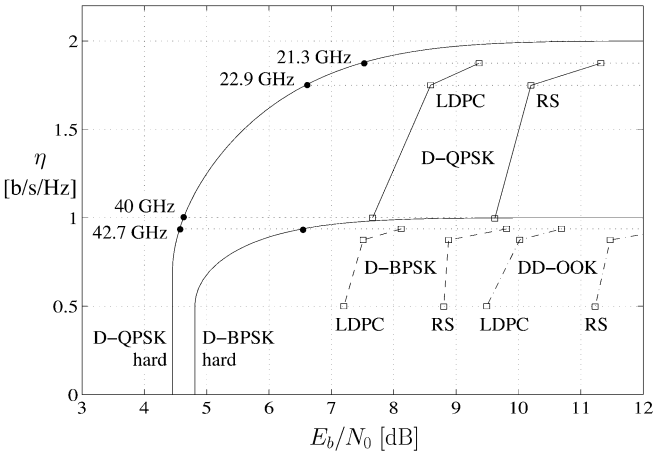


Fig. 7. Code operating points for $P_b = 10^{-16}$ and a data rate of 40 Gb/s. The point labels refer to the symbol rates.

gain as compared to the current code: the coding gain and a detector gain.

Finally, suppose one uses D-QPSK with the $R_c = 1/2$ LDPC code. The symbol rate is now 40 GHz and $E_b/N_0 = 7.66$ dB. Thus, one saves a small amount of bandwidth and still saves 3.03 dB in energy over our standard DD-OOK scheme. However, the energy savings over D-BPSK are only 0.46 dB, which is much less than the 1.97 dB expected from Table I. The main factor contributing to the large degradation is that, for hard detection BPSK, our rate 1/2 LDPC code gains only 1.88 dB rather than the expected 3.25 dB over our rate 15/16 LDPC code. This suggests that one can design a better rate 1/2 LDPC code.

VI. SUMMARY

This paper has investigated the performance of several combinations of codes, modulations, and detectors. We have concentrated on PSK because of its linear gains and its potential fiber nonlinearity gains. The main points we wish to make are as follows.

1) D-BPSK outperforms DD-OOK by about 2.5 dB in energy efficiency with strong codes.
2) D-QPSK with low-rate codes outperforms D-BPSK with high-rate codes, in terms of both bandwidth and energy. The bandwidth gains are particularly large.
3) Real codes can realize the gains predicted by the information-theoretic analysis, even for high-speed systems where optimal decoding is not possible.
4) Soft differential detection will give further large gains. However, this requires A/D converters with multibit resolution at high speeds such as 10 and 40 Gb/s.

ACKNOWLEDGMENT

The LDPC code performance analysis was based in part on unpublished work by T. J. Richardson and R. L. Urbanke. The authors also wish to thank S. Hunsche, J. E. Mazo, J. Salz, and R. E. Slusher for their support of this work.

REFERENCES

[1] A. Gnauck, G. Raybon, S. Chandrasekhar, J. Leuthold, C. R. Doerr, L. W. Stulz, A. Agarwal, S. Banerjee, D. F. Grosz, S. Hunsche, A. P. Kung, A. Marhelyuk, D. Maywar, M. Movassaghi, X. Liu, C. Xu, X. Wei, and D. M. Gill, "2.5 Tb/s (64×42.7 Gb/s) transmission over 40×100 km NZDSF using RZ-DPSK format and all-Raman-amplified link," in *Proc. OFC 2002*. paper FC2.

[2] M. Rohde, C. Caspar, N. Heimes, M. Konitzer, E.-J. Bachus, and N. Hanik, "Robustness of DPSK direct detection transmission format in standard fiber WDM systems," *Electron. Lett.*, vol. 36, no. 17, pp. 1483–1484, Aug. 17, 2000.

[3] T. Miyano, M. Fukutoku, K. Hattori, and H. Ono, "Suppression of degradation induced by SPM/XPM+GVD in WDM transmission using a bit-synchronous intensity modulated DPSK signal," presented at the OECC, 2000. 14D3-3.

[4] J. Leibrich, C. Wree, and W. Rosenkranz, "CF-RZ-DPSK for suppression of XPM on dispersion-managed long-haul optical WDM transmission on standard single-mode fiber," *IEEE Photon. Technol. Lett.*, vol. 14, pp. 155–157, Feb. 2002.

[5] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.

[6] J. G. Proakis, *Digital Communications*, 3rd ed: McGraw-Hill, 1995.

[7] S. Betti, F. Curti, G. De Marchis, and E. Iannone, "A novel multilevel coherent optical system: 4-quadrature signaling," *J. Lightwave Technol.*, vol. 9, pp. 514–523, Apr. 1991.

[8] C. E. Shannon, "Communication in the presence of noise," in *Proc. IRE*, vol. 37, Jan. 1949, pp. 10–21. Reprinted in *Claude Elwood Shannon: Collected Papers*, pp. 160–172, (N. J. A. Sloane and A. D. Wyner, Eds.) Piscataway, NJ: IEEE Press, 1993.

[9] P. Poggiolini and S. Benedetto, "Theory of polarization spreading techniques—Part I," *IEEE Trans. Commun.*, vol. 42, pp. 2105–2118, May 1994.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[11] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948. Reprinted in *Claude Elwood Shannon: Collected Papers*, pp. 5–83, (N. J. A. Sloane and A. D. Wyner, Eds.) Piscataway, NJ: IEEE Press, 1993.

[12] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319–1343, June 2002.

[13] W. R. Bennet, "Methods of solving noise problems," in *Proc. IRE*, vol. 44, May 1956, pp. 609–638.

[14] F. S. Weinstein, "Simplified relationships for the probability distribution of the phase of a sine wave in narrow-band normal noise," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 658–661, Sept. 1974.

[15] R. F. Pawula, S. O. Rice, and J. H. Roberts, "Distribution of the phase angle between two vectors perturbed by Gaussian noise," *IEEE Trans. Commun.*, vol. 30, pp. 1828–1841, Aug. 1982.

[16] R. F. Pawula, "A new formula for MDPSK symbol error probability," *IEEE Commun. Lett.*, vol. 2, pp. 271–272, Oct. 1998.

[17] P. Hoeher and J. Lodge, ""Turbo DPSK": Iterative differential PSK demodulation and channel decoding," *IEEE Trans. Commun.*, vol. 47, pp. 837–843, June 1999.

[18] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[19] S. Yamazaki and K. Emura, "Feasibility study on QPSK optical-heterodyne detection system," *J. Lightwave Technol.*, vol. 8, pp. 1646–1653, Nov. 1990.

[20] S. Norimatsu, K. Iwashita, and K. Noguchi, "An 8 Gb/s QPSK optical homodyne detection experiment using external-cavity laser diodes," *IEEE Photon. Technol. Lett.*, vol. 4, no. 7, pp. 765–767, 1992.

[21] H. Leib and S. Pasupathy, "The phase of a vector perturbed by Gaussian noise and differentially coherent receivers," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1491–1501, Nov. 1988.

[22] D. Divsalar and M. K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Trans. Inform. Theory*, vol. 38, pp. 300–308, Mar. 1990.

[23] H. Leib, "Data-aided noncoherent demodulation of DPSK," *IEEE Trans. Commun.*, vol. 43, pp. 722–725, Feb./Mar./Apr. 1995.

[24] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. Soc. Ind. Appl. Math.*, vol. 8, pp. 300–304, June 1960.

[25] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 21–28, Jan. 1962.

[26] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.

[27] R. W. Lucky, J. Salz, and E. J. Weldon Jr., *Principles of Data Communication*. New York: McGraw-Hill, 1968.



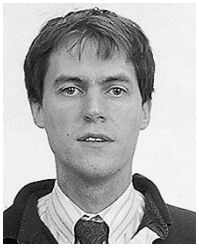
Gerhard Kramer (S'91–M'94) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Manitoba, Winnipeg, MB, Canada, in 1991 and 1992, respectively, and the Dr.Sc.Techn. degree from the Swiss Federal Institute of Technology (ETH), Zürich, Switzerland, in 1998.

From August 1998 to March 2000, he was with Endora Tech AG, Basel, Switzerland, as a Communications Engineering Consultant. Since May 2000, he has been a Member of Technical Staff in the Mathematics of Communications Research Department, Bell Laboratories, Lucent Technologies, Murray Hill, NJ.



Alexei Ashikhmin (M'00) received the Ph.D. degree in information theory from the Institute of Information Transmission Problems, Moscow, Russia, in 1994.

From 1995 to 1996, he was with the Delft University of Technology, Delft, The Netherlands. From January 1997 to July 1999, he was with Los Alamos National Laboratory, Los Alamos, NM. Since 1999, he has been with the Mathematics of Communications Research Department, Bell Laboratories, Lucent Technologies, Murray Hill, NJ. His research interests include the theory of error-correcting codes, as well as information and communication theory.



Adriaan J. van Wijngaarden (S'87–M'98) received the engineering degree in electrical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1992 and the doctorate degree in engineering from the University of Essen, Essen, Germany, in 1998.

From 1992 to 1998, he was a Research Engineer and Member of Staff of the Digital Communications Group, Institute for Experimental Mathematics, University of Essen, Germany. Since 1998, he is a Member of Technical Staff in the Mathematics of Communications Research Department, Bell Laboratories, Lucent Technologies, Murray Hill, NJ. His current research interests include communications, in particular error control coding, synchronization, and modulation coding with applications in recording, optical, and wireless communications.



Xing Wei received the B.S. and M.S. degrees in physics from Fudan University, Shanghai, China, in 1990 and 1992, respectively, and the Ph.D. degree in physics from the University of California at Berkeley in 2000.

While at Berkeley, he studied molecular structures at surfaces and interfaces of various materials including ice and polymers using nonlinear optical spectroscopy. He joined Bell Laboratories, Lucent Technologies, Murray Hill, NJ, as a Member of Technical Staff in the Physical Sciences Research Division in January 2001. Since then, he has engaged in basic research related to optical fiber communications with focus on novel modulation techniques.

Dr. Wei is a Member of the American Physical Society and the Optical Society of America (OSA).