Practical Considerations in Direct Detection Under Tukey Signalling

Amir Tasbihi, Member, IEEE, and Frank R. Kschischang, Fellow, IEEE

Abstract—The deliberate introduction of controlled intersymbol interference (ISI) in Tukey signalling enables the recovery of signal amplitude and (in part) signal phase under direct detection, giving rise to significant data rate improvements compared to intensity modulation with direct detection (IMDD). The use of an integrate-and-dump detector makes precise waveform shaping unnecessary, thereby equipping the scheme with a high degree of robustness to nonlinear signal distortions introduced by practical modulators. Signal sequences drawn from star quadrature amplitude modulation (SQAM) formats admit an efficient trellis description that facilitates codebook design and low-complexity near maximum-likelihood sequence detection in the presence of both shot noise and thermal noise. Under the practical (though suboptimal) allocation of a 50% duty cycle between ISI-free and ISI-present signalling segments, at a symbol rate of 50 Gbaud and a launch power of −10 dBm the Tukey scheme has a maximum theoretically achievable throughput of 200 Gb/s with an (8,4)-SQAM constellation, while an IMDD scheme achieves about 145 Gb/s using PAM-8. Note that the two mentioned constellations have the same number of magnitude levels and the difference in throughput is resulting from exploiting phase information under using a complex-valued signal constellation.

Index Terms—Direct detection, modulator nonlinearity, short-haul fiber-optic communication, intersymbol interference, Tukey window.

I. INTRODUCTION

Tukey waveforms are time-limited signals that allow amplitude and (to some extent) phase recovery of complex-valued symbols under direction detection by the deliberate introduction of inter-symbol interference (ISI) [1], in a manner that is reminiscent of partial-response signalling schemes dating back to the early 1960s [2], [3]. Under Tukey signalling, ISI is controlled so that only adjacent symbols interfere, resulting in an alternation between ISI-free and ISI-present signalling segments at the receiver, where detection is accomplished by a low-complexity integrate-and-dump architecture.

Tukey signalling with direct detection is an alternative to intensity-modulation with direct detection (IMDD), which encodes only the intensity, i.e., squared magnitude, of the transmitted waveform by transmitting only positive real symbols. In contrast to IMDD, Tukey signalling with direct detection allows information to be encoded in the phase, not only the intensity, of the transmitted symbols; however, this increased capability necessitates implementation of an in-phase/quadrature (IQ) modulator at the transmitter and the use of two (rather than one) analog-to-digital (A/D) converters (each operating at the symbol rate) at the receiver. The requirement of having at least two real-valued samples per complex-valued symbol is inevitable in all schemes that extract phase (in addition to magnitude).

Another class of communication schemes that can exploit phase under direct detection is based on the so-called Kramers-Kronig (KK) (or Hilbert transform) relationship between the phase and the logarithm of the magnitude of a minimum-phase complex-valued signal [4], [5], [6]. The use of this relationship in communication systems dates back to the early 1960s [7], [8], [9], [10], [11], [12], where it was used to extract phase information from an envelope detector. To satisfy the minimum-phase constraint, KK-based schemes must add a non-information-bearing tone to the complex-valued waveform (either at the transmitter or at the receiver), resulting in a substantial inefficiency in overall transceiver power consumption. Furthermore, because they must perform bandwidth-broadening nonlinear operations, KK receivers often demand an over-sampling of about three times the minimum required sampling rate, i.e., six times the symbol rate. Although there are KK schemes that enable phase recovery without such oversampling, these require a very high carrier-to-signal power ratio [6].

In this paper, we address a number of practical concerns that arise with Tukey signalling under direct detection. In particular,

- we consider the influence of the nonlinearity of the IQ modulator (which was idealized in [1]) and show that the scheme is quite robust to modulator imperfections;
- we fix the duty-cycle between ISI-free and ISI-present signalling segments to 50% (unlike in [1], where it was allowed to range to values as low as 10%, which leads to unrealistically short integration intervals in high baud-rate systems);
- we introduce a low-complexity near maximum-likelihood (ML) trellis-based decoding algorithm in the presence of both shot noise and thermal noise encountered at the output of a p-i-n photodiode under practical parameter settings (replacing the brute-force search decoding and avalanche photodiode analysis of [1]);
- we consider a wider range of constellation sizes compared with [1].

Manuscript received 3 March 2022; revised 5 October 2022; accepted 1 December 2022. Date of publication 5 December 2022; date of current version 16 March 2023. This work was supported by a grant from Huawei Technologies Canada. (Corresponding author: Amir Tasbihi.)

The authors are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: tasbihi@ece.utoronto.ca; frank@ece.utoronto.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2022.3272006.

Digital Object Identifier 10.1109/JLT.2022.3272006
The rest of the paper is organized as follows. Section II describes the system model for use in the C band where chromatic dispersion is significant; in particular, Section II-B addresses the nonlinearity of the IQ modulator. Section III shows how trellis diagrams can be used for codebook design and decoding. While typically trellises are associated with a given codebook, in Section III-D trellises are used to find codebooks composed of symbols drawn from the class of signal constellations described in Section III-B. Section III-E provides a branch-metric function that enables near-ML trellis-based decoding using the Viterbi algorithm. Section IV provides extensive numerical simulation results, while Section V compares Tukey signaling under direct detection with IMDD. The system model of Section II requires link-distance knowledge to precompensate for chromatic dispersion. In Section VI, we remove this requirement by operating in the O band, instead of the C band, leaving residual chromatic dispersion uncompensated and making the transmitter agnostic to the transmission-link length. Finally, Section VII provides various concluding remarks.

Throughout this paper, the integers, real numbers, and complex numbers are denoted, respectively, by \( \mathbb{Z} \), \( \mathbb{R} \), and \( \mathbb{C} \). The set of real-valued and complex-valued functions over \( \mathbb{R} \) are denoted as \( \mathbb{R}^x \) and \( \mathbb{C}^x \), respectively. For any \( x \in \mathbb{R} \), \( \mathbb{R}^{>x} \) denotes the set of real numbers strictly greater than \( x \). The sets \( \mathbb{R}^{>x} \), \( \mathbb{Z}^{>x} \), and \( \mathbb{Z}^{>2x} \) are defined similarly. For any \( a, b \in \mathbb{Z} \) with \( a \leq b \), let \( [a, b] = \{ a, a + 1, \ldots, b \} \). The cardinality of a set \( S \) is denoted by \( |S| \); thus, e.g., \( |[0, m - 1]| = m \). For any \( r \in \mathbb{R} \), \( [r] \) and \( \{ r \} \) denote, respectively, the smallest integer greater than or equal to \( r \) and the largest integer less than or equal to \( r \). For a complex number \( w \in \mathbb{C} \), \( |w| \), \( \arg(w) \), and \( w^* \) denote the magnitude, phase, and complex conjugate of \( w \), respectively; furthermore, we assume that \( \arg(w) \in [0, 2\pi) \) and \( i = \sqrt{-1} \). The Fourier transform of a complex-valued function \( x(t) \) is denoted as \( \mathcal{F}[x(t)] \); the inverse Fourier transform of \( X(f) \) is denoted as \( \mathcal{F}^{-1}[X(f)] \). For a random variable \( X \), \( X \sim \mathcal{N}(\mu, \sigma^2) \) denotes that \( X \) has a Gaussian distribution with mean \( \mu \) and variance \( \sigma^2 \). Finally, for any \( k \in \mathbb{Z}^{>0} \), \( I_k \) denotes the \( k \times k \) identity matrix, while for any \( k \) and \( \ell \in \mathbb{Z}^{>0} \), \( I_{k \times \ell} \) denotes the all-one matrix of size \( k \times \ell \).

II. System Model

In this section, we describe the system model, as shown in Fig. 1. For simplicity of computations and ease of reading, we assume transmission over a single polarization; however, with an obvious modification, the scheme can be adapted to work in a dual-polarized system.

A. The Transmission Medium

We assume data transmission over a standard single-mode fiber (SSMF) at a wavelength in the C band, i.e., close to 1550 nm, where the fiber has the lowest loss, i.e., about 0.2 dB km\(^{-1}\). The lower loss comes at the expense of a higher chromatic dispersion compared to operating at the zero-dispersion wavelength of SSMF, i.e., about 1310 nm in the O band.

A key application of direct detection is in short-haul data transmission, e.g., intra-data-center communication. Therefore, power loss and chromatic dispersion are the only fiber imperfections that we take into account. Other defects of the channel, e.g., Kerr effect or polarization mode dispersion (for a dual-polarized transmission), are not considered. Accordingly, if the waveform \( x(t) \in \mathbb{C}^R \) is launched into the fiber, the waveform received at the fiber output is modelled as [13, Section 4.3.3]

\[
r(t) = \mathcal{F}^{-1}\left[ X(f)e^{-\frac{a}{2}e^{-i\beta_0L}e^{-2\pi f\beta_1L}e^{-i2\beta_2L\pi^2f^2}} \right],
\]

where \( X(f) = \mathcal{F}[x(t)] \), \( L \in \mathbb{R}^{>0} \) is the fiber length, \( \rho \in \mathbb{R}^{>0} \) is the loss factor and \( \beta_2 \in \mathbb{R} \) is the group-velocity dispersion parameter. At the wavelength of 1550 nm, \( \rho = 0.046 \text{ km}^{-1} \) (equivalent to a loss of 0.2 dB km\(^{-1}\)) and \( \beta_2 = -2.167 \times 10^{-23} \text{ s}^2 \cdot \text{km}^{-1} \). At the operating wavelength \( \lambda \), we have \( \beta_0 = 2\pi/\lambda \).

Furthermore, \( \beta_1 \) is the group delay per unit of length [13, Section 4.3].

B. IQ Modulator

The IQ modulator comprises two Mach-Zehnder modulators (MZMs) operating in the push-pull regime. For a modulating waveform \( v(t) \in \mathbb{R}^R \), the output of an MZM in this regime is \( E_{\text{out}}(t) = E_{\text{in}} \sin(\kappa \nu t) \), where \( E_{\text{in}} \) is an unmodulated continuous-wave narrow-band optical input and \( \kappa \in \mathbb{R} \) is a constant depending on the operating wavelength and physical properties of the MZM [13, Section 7.5.3]. In this paper, we consider a baseband model; thus, we assume that \( E_{\text{in}} \in \mathbb{R}^{>0} \).

The IQ modulator uses separate MZMs for the in-phase and the quadrature components, producing the optical waveform [13, Section 7.5.3]

\[
x(t) = E_{\text{in}} \left( \sin (\kappa \Re(u(t))) + i \sin (\kappa \Im(u(t))) \right)
\]

in response to the complex-valued electrical waveform \( u(t) \in \mathbb{C}^R \).

C. Signalling

For any \( \beta \in [0, 1] \) the Tukey (or time-domain raised cosine [14]) waveform \( w_{\beta}(t) \) is defined as [1]

\[
w_{\beta}(t) \triangleq \begin{cases} 
\alpha_{\beta}, & \text{if } |t| \leq \frac{1-\beta}{2} \\
\alpha_0 \left( 1 - \sin \left( \frac{\pi(2|t|-1)}{2\beta} \right) \right), & \text{if } \left| |t| - \frac{1}{2} \right| \leq \frac{\beta}{2} \\
0, & \text{otherwise,}
\end{cases}
\]

where \( \alpha_0 \triangleq \frac{1}{\sqrt{4\beta}} \). For a \( T \in \mathbb{R}^{>0} \) and a block length \( n \in \mathbb{Z}^{>1} \), the signalling block in Fig. 1 accepts \( n \) complex numbers
\[ x_0, \ldots, x_{n-1} \in \mathbb{C} \] as its input and produces
\[ g(t) = \sum_{k \in [0, n-1]} x_k w(t - kT), \] (2)
where the signalling waveform \( w(t) \) is a scaled, dilated Tukey waveform; in particular, \( w(t) = \frac{1}{\sqrt{T}} w_\beta \left( \frac{t}{T} \right) \) for some \( \beta \in [0, 1] \).

\[ x_0, \ldots, x_{n-1} \in \mathbb{C} \] as its input and produces
\[ g(t) = \sum_{k \in [0, n-1]} x_k w(t - kT), \] (2)

\[ \text{D. Dispersion Precompensation} \]

Chromatic dispersion is precompensated in the electrical domain by a filter with frequency response \( H(f) = e^{i2\beta_L f^2} \).

Note that this linear precompensation does not take into account the nonlinearity of the IQ modulator and therefore it does not perform ideal precompensation. Nevertheless, when the power of the modulating waveform, \( u(t) \), is relatively small, the IQ modulator operates almost linearly; thus, the dispersion precompensating filter with frequency response \( H(f) \) compensates the chromatic dispersion of the fiber to a large degree.

\[ \text{E. Photodiode and Noise} \]

The output \( s(t) \in \mathbb{R}^R \) of the photodiode, in response to the input waveform \( r(t) \in \mathbb{C}^R \), is given as
\[ s(t) = \Re [r(t)]^2 + |r(t)| \mathcal{N}(t) + \mathcal{N}(t), \]
where \( \Re \) is the photodiode responsivity, and where \( \mathcal{N}(\cdot) \) and \( \mathcal{N}(\cdot) \) are zero-mean white Gaussian random processes with constant two-sided power spectral densities \( \sigma^2 \) and \( \sigma^2 \), respectively.

The two terms, \( |r(t)| \mathcal{N}(t) \) and \( \mathcal{N}(t) \), are referred to as shot noise and thermal noise, respectively.

\[ \text{F. Integrate & Dump} \]

Note that \( g(t) \) depends only on \( x_k, k \in [0, n-1] \), whenever \( |t - kT| \leq \frac{1}{2} \) and depends only on \( x_{k+1} \) for \( k \in [0, n-2] \), whenever \( |t - (k + 1/2)| \leq \frac{1}{2} \). Thus, as in [1], we define the \( \ell \)-th ISI-free interval as
\[ \gamma_k \triangleq \{ t \in \mathbb{R} : |t - kT| \leq (1 - \beta)T/2 \} \]
and the \( \ell \)-th ISI-present interval as
\[ Z_k \triangleq \{ t \in \mathbb{R} : |t - \ell T/2| < \beta T/2 \}. \]
The integrate-and-dump unit accepts \( s(t) \) as its input and produces
\[ y_k = \int_{\gamma_k} s(t) \, dt \quad \text{and} \quad z_k = \int_{\gamma_k} s(t) \, dt \] (3)
for \( k \in [0, n-1] \) and \( \ell \in [0, n-2] \).

As we do not compensate for the nonlinearity of the IQ modulator, determining the exact distribution of \( y_k \) and \( z_k \) given the transmitted symbols, \( x_0, \ldots, x_{n-1} \), is not straightforward. Thus, the implementation of a true ML receiver seems intractable. Instead, we use
\[ x(t) \approx E_{\text{AWG}}u(t), \] (4)
as an approximation for (1) at the decoder, which becomes increasingly accurate for low-power modulating waveforms. Note that this approximation is applied at the receiver, and not at the transmitter, to simplify computations. In all simulation results reported in Section IV, the nonlinearity of the IQ modulator, as given in (1), has been properly taken into account when simulating the transmitter.

Using (4), one may approximate \( y_k \) in (3) as
\[ y_k \approx \Re \mathcal{N} + \Re \mathcal{N} \mathcal{N} + m_k, \] (5)
where
\[ \mathcal{N} \mathcal{N} = \left( E_{\text{AWG}}K_\alpha e^{-i2\beta_L f^2} \right)^2 (1 - \beta) \in \mathbb{R}^{n, n}, \] (6)
and where \( n_k \sim \mathcal{N}(0, \sigma^2), m_k \sim \mathcal{N}(0, (1 - \beta)T \sigma^2), \) and \( k \in [0, n-1] \).

Define \( \psi : \mathbb{C}^2 \to \mathbb{R}^{n, n} \) such that, for all \( (a, b) \in \mathbb{C}^2 \),
\[ \psi(a, b) = \frac{1}{4}|a + b|^2 + \frac{1}{8}|a - b|^2. \] (7)
Note that \( \psi(a, b) \) is a function of the magnitudes of \( a \) and \( b \) and the cosine of their phase difference; thus it follows that
\[ \psi(ae^{i\theta}, be^{i\theta}) = \psi(a, b), \] (8)
for any \( \theta \in \mathbb{R} \). By using (3) and (4), one may approximate \( z_k \) as
\[ z_k \approx \Re \mathcal{N} + \Re \mathcal{N} \mathcal{N} + \psi(m_k, \mathcal{N}), \] (9)
where
\[ \mathcal{N} = \left( \alpha_\beta E_{\text{AWG}}e^{-2i\beta_L f^2} \right)^2 \psi(x_k, x_{k+1}) + \beta \in \mathbb{R}^{n, n}, \] (10)
and where \( p_k \sim \mathcal{N}(0, \sigma^2), q_k \sim \mathcal{N}(0, (1 - \beta)T \sigma^2), \) and \( k \in [0, n-2] \). Furthermore, given the transmitted symbols \( x_0, \ldots, x_{n-1} \), for any \( k, k' \in [0, n-1] \) and \( \ell, \ell' \in [0, n-2] \), \( m_k, m_{k'}, \ell, \ell' \), and \( q_k, q_{k'}, \ell, \ell' \) are mutually independent random variables.

Under the approximations (5) and (9), the conditional distribution of \( y_k \) given \( x_k \) is
\[ y_k | x_k \sim \mathcal{N} (\Re \mathcal{N}, \mathcal{N} \mathcal{N} \mathcal{N} \mathcal{N} + (1 - \beta)T \sigma^2), \] (11)
and the conditional distribution of \( z_k \) given \( x_k \) and \( x_{k+1} \) is
\[ z_k | x_k, x_{k+1} \sim \mathcal{N} (\Re \mathcal{N} \mathcal{N}, \mathcal{N} \mathcal{N} \mathcal{N} \mathcal{N} + (1 - \beta)T \sigma^2), \] (12)
where “\( \sim \)” should be read as “approximately distributed as.” From (11) and (12), it follows that the detection performance—i.e., bit error rate (BER) or achievable information rate—depends on the symbol rate. This is in contrast with communication over a classical additive white Gaussian noise (AWGN) channel, where the detection performance does not depend on the symbol rate. In the communication scenarios considered in this paper, shot noise dominates thermal noise at sufficiently high symbol rates.

\[ \text{G. Codebook} \]

Similar to [1], we define the function \( \Upsilon : \mathbb{C}^n \to \mathbb{R}^{2n-1} \) as
\[ \Upsilon(x_0, \ldots, x_{n-1}) = (|x_0|^2, \psi(x_0, x_1), |x_1|^2, |x_1|^2, \psi(x_1, x_2), \ldots, |x_{n-2}|^2, |x_{n-2}|^2, \psi(x_{n-2}, x_{n-1}), |x_{n-1}|^2, \ldots), \]
and, for any \( x \in \mathbb{C}^n \), we refer to \( \Upsilon(x) \) as the signature of \( x \).

From (5) and (9) one may see that any two transmitted complex-valued \( n \)-vectors having the same signature are indistinguishable by our receiver, even in the absence of noise.
We may define an equivalence relation on \( \mathbb{C}^n \), deeming two vectors, \( \mathbf{x} \) and \( \tilde{x} \in \mathbb{C}^n \) as square-law equivalent, denoted \( \mathbf{x} \equiv \tilde{x} \), if \( \Upsilon (\mathbf{x}) = \Upsilon (\tilde{x}) \). When \( \mathbf{x} \) and \( \tilde{x} \) are not square-law equivalent, they are called square-law distinct (SLD). We refer to any set \( S \subset \mathbb{C}^n \) of pairwise SLD \( n \)-tuples as a codebook, and we will always assume that the transmitted \( n \)-tuples, \((x_0, \ldots, x_{n-1})\), in (2) are chosen from some fixed codebook.

### H. Decoder

The decoder block in Fig. 1 accepts \( y_0, \ldots, y_{n-1} \) and \( z_0, \ldots, z_{n-2} \) and produces the detected \( n \)-tuple \( \hat{\mathbf{x}} \), an element of the codebook \( S \). Note that the ISI-present intervals at the beginning and at the end of the received waveform, which overlap, respectively, with the previous and the next blocks, are ignored in detection. Therefore, no time-guard is needed between consecutive blocks.

From (5) and (9) one may conclude that, independent of the decoder block, any pair of codebooks \( S_1 \) and \( S_2 \) where \( \Upsilon(S_1) = \Upsilon(S_2) \), i.e., where the codewords produce the same set of signatures, are equivalent from the perspective of the detector and thus have the same error rate.

In [1], the decoder performs ML detection by a brute-force search over all \( |S| \) elements of the codebook. We describe a more practical lower-complexity trellis-based decoding algorithm in the next section.

### III. THE TRELILS DIAGRAM AND STAR QAM CONSTELLATIONS

In this section we show how trellis diagrams that result from the use of star quadrature amplitude modulation (SQAM) can be used to determine SLD sequences for codebook design and for decoding.

#### A. Trellises

Recall that a trellis \( \mathcal{T} = (\mathcal{V}, \mathcal{A}, \mathcal{E}) \) is an edge-labelled directed graph having a vertex set \( \mathcal{V} \), an edge-label set \( \mathcal{A} \), and a set \( \mathcal{E} \subset \mathcal{V} \times \mathcal{A} \times \mathcal{V} \) of labelled edges, satisfying (15):

1) \( \mathcal{V} \) contains distinct vertices \( v_r \) (the root, having zero in-degree) and \( v_g \) (the goal, having zero out-degree);
2) every vertex in \( \mathcal{V} \) is reachable by a directed path from \( v_r \);
3) \( v_g \) is reachable by a directed path from every vertex in \( \mathcal{V} \);
4) every directed path from \( v_r \) to \( v_g \) has the same length, called the length of \( \mathcal{T} \) and denoted by \( |\mathcal{T}| \).

These properties imply that every directed path from \( v_r \) to any fixed vertex \( v \in \mathcal{V} \) has the same length, \( d(v) \), called the depth of \( v \). Thus \( d(v_r) = 0 \) and \( d(v_g) = |\mathcal{T}| \). The vertex set \( \mathcal{V} \) can then be partitioned as \( \mathcal{V}_0 \cup \cdots \cup \mathcal{V}_{|\mathcal{T}|} \), where \( \mathcal{V}_j \) denotes the set of vertices having depth \( j \). The subgraph of \( \mathcal{T} \) induced by edges incident from \( \mathcal{V}_j \) is called the \( j \)th trellis section of \( \mathcal{T} \).

An edge \( e = (v_1, a, v_2) \in \mathcal{E} \) (directed from \( v_1 \) to \( v_2 \) with label \( a \)) is said to have label \( \lambda(e) = a \). If

\[
p = ((v_1, a_1, v_2), (v_2, a_2, v_3), \ldots, (v_{m-1}, a_{m-1}, v_m))
\]

is a directed path from a vertex \( v_1 \) to a vertex \( v_m \) in \( \mathcal{T} \), we denote by \( \lambda(p) \) the label sequence \((a_1, \ldots, a_{m-1})\) associated with that path. The set of all directed paths from \( v_r \) to \( v_g \) in \( \mathcal{T} \) is denoted as \( \mathcal{P}_T \), and the set of associated label sequences is denoted as

\[
\mathcal{L}_T \triangleq \{ \lambda(p) \mid p \in \mathcal{P}_T \}.
\]

In effect, a trellis \( \mathcal{T} \) is simply a convenient graphical representation of its set of label sequences, \( \mathcal{L}_T \).

#### B. Star QAM Constellation

In this paper we consider a class of signal constellations called star-quadrature-amplitude modulation (SQAM) [14] having particularly tractable trellis representations which enable their investigation. Star QAM constellations have points in the complex plane that lie at the intersections of a number of uniformly spaced rays and a number of concentric rings, as illustrated in Fig. 2.

More formally, the number of rings is denoted as \( n_r \). The ring radii are determined by a radius set \( \mathcal{D} = \{d_1, d_2, \ldots, d_{n_r}\} \) of \( n_r \) distinct positive real numbers. The number of phase angles (or rays) is denoted as \( n_p \). The phase angles are drawn from the set

\[
\Phi_{n_p} \triangleq \left\{ \frac{2\pi}{n_p}, \frac{2\pi}{n_p-1}, \ldots, \frac{2\pi(n_p-1)}{n_p} \right\},
\]

and the \((n_r, n_p)\)-SQAM constellation with radius set \( \mathcal{D} \) and \( n_p \) phases is given as

\[
\mathcal{K}_{\mathcal{D}, n_p} = \{ d \cdot e^{i\theta} : d \in \mathcal{D}, \theta \in \Phi_{n_p} \}.
\]

For later convenience, let

\[
\mathcal{D}_j \triangleq \{ d_j e^{i\theta} : \theta \in \Phi_{n_p} \}
\]

be the set of constellation points at radius \( d_j \). In [1], \( \mathcal{K}_{\mathcal{D}, n_p} \) is referred to as an \( n_r \)-ring/\( n_p \)-ary phase constellation. For brevity, when \( \mathcal{D} \) is clear from context or irrelevant, we may refer to an \((n_r, n_p)\)-SQAM constellation without mentioning the radius set \( \mathcal{D} \) explicitly.

#### C. A Trellis for SQAM Constellations

In this subsection, we provide a trellis diagram, \( \mathcal{T} \), that describes the signatures of SLD sequences of length \( n \) drawn from \( \mathcal{K}_{\mathcal{D}, n_p} \), an \((n_r, n_p)\)-SQAM constellation with radius set \( \mathcal{D} = \{d_1, \ldots, d_{n_r}\} \). The trellis \( \mathcal{T} \) will be used both to design a codebook \( S \) and for decoding.
The length of $T$ is $2n - 1$. The vertex set of $T$ is given as $V = \{v_r\} \cup V_1 \cup V_2 \cup \ldots \cup V_{2n-2} \cup \{v_0\}$, where, for $j \in [1, 2n - 2]$, $V_j$ is the $n_r$-set $V_j = \{v_{j,1}^1, v_{j,2}^1, \ldots, v_{j,n_r}^1\}$. Thus, except at depth 0 and at depth $|T|$, the trellis $T$ has exactly $n_r$ vertices at each depth. To define the edge-label alphabet of $T$, for any (not necessarily distinct) $j$ and $k \in [1, n_r]$ let

$$\Psi_{j,k} = \{\psi(a, b) : a \in D_j, b \in D_k\}. \quad (13)$$

By symmetry, $\Psi_{j,k} = \Psi_{k,j}$. Let $A_{\text{SI-present}} = \bigcup_{j=1}^{n_r} \bigcup_{k=1}^{n_r} \Psi_{j,k}$ and let $A_{\text{SI-free}} = \{d_1^2, \ldots, d_{n_r}^2\}$. The edge-label alphabet of $T$ is then given as $A = A_{\text{SI-present}} \cup A_{\text{SI-free}}$.

Finally, to define the edge set of $T$, let

$$E_r = \{(v_{j,1}^1, v_{j,2}^1) : j \in [1, n_r]\},$$

$$E_g = \{(v_{2n-2,j}^1, v_{2n-2,j}^2) : j \in [1, n_r]\},$$

$$E_{\text{SI-present}} = \{(v_{2b+1,j}^1, \zeta, v_{2b+2,j}^2) : b \in [0, n-2], j, k \in [1, n_r], \zeta \in \Psi_{j,k}\},$$

$$E_{\text{SI-free}} = \{(v_{2b+1,j}^1, v_{2b+2,j}^2) : b \in [1, n-2], j \in [1, n_r]\};$$

then $E = E_r \cup E_g \cup E_{\text{SI-present}} \cup E_{\text{SI-free}}$. Note that $E_r$ is the set of edges in $T$ incident from the root vertex $v_r$, $E_g$ is the set of edges incident to the goal vertex $v_g$, $E_{\text{SI-present}}$ are edges incident from vertices at odd depth, while $E_{\text{SI-free}}$ are edges incident from vertices at even depth. There are $|\Psi_{j,k}|$ parallel edges between $v_{2b+1,j}^1$ and $v_{2b+2,j}^2$ for $b \in [0, n-2]$ and $j, k \in [1, n_r]$.

**Example 1:** In this example, we sketch the trellis diagram for the (3,3)-SQAM constellation with radius set $D = \{2\sqrt{2}, 4\sqrt{2}, 6\sqrt{2}\}$, as shown in Fig. 2(a), when the length of SLD sequences is $n = 3$. We have $\Psi_{1,1} \equiv \{8, 5\}$, $\Psi_{1,2} \equiv \{19, 13\}$, $\Psi_{1,3} \equiv \{27, 36\}$, $\Psi_{2,2} \equiv \{20, 32\}$, $\Psi_{2,3} \equiv \{33, 51\}$, and $\Psi_{3,3} \equiv \{45, 72\}$. The resulting trellis is shown in Fig. 3. One may see that for $b \in [1, 3]$, the edge labels between, for example, $v_{b+1}^1$ and $v_{b+1}^2$ belong to $\Psi_{1,3}$. Similarly, the edge label between, for example, $v_{2}^2$ and $v_{3}^2$ is equal to $(4\sqrt{2})^2$. For ease of reading, the radius set $D$ for this example was judiciously chosen to result in integer edge labels.

**Theorem 1:** For any $k$ and $\ell \in [1, n_r]$, $|\Psi_{k,\ell}| = \left[\frac{n_r+1}{2}\right]$.

**Proof:** The phase difference between any $a \in D_k$ and $b \in D_\ell$ is in $\Phi_{n_r}$ and, since $\cos(\theta) = \cos(2\pi - \theta)$, for any $\theta \in \mathbb{R}$, we leave it to the reader to check that there are exactly $\left[\frac{n_r+1}{2}\right]$ distinct values for $\cos(\arg(ab))$.

For example, Fig. 4 illustrates the situation that arises when $n_p = 8$ and $n_p = 9$, for which $|\Psi_{k,\ell}| = 5$.

The following theorem counts $|P_T|$, the number of distinct directed paths from root to goal in $T$.

**Theorem 2:** For the trellis $T$ of length $2n - 1$ described in Section III-C,

$$|P_T| = \left[\frac{n_p+1}{2}\right]^{n-1}n_r^n.$$

**Proof:** Let $A_j$ denote the adjacency matrix of the $j$th trellis section of $T$. We have

$$A_j = \begin{cases} I_{1 \times n_r}, & j = 0, \\ I_{n_r \times n_r}, & j = \{2, 4, \ldots, 2n-4\}, \\ \left[\frac{n_p+1}{2}\right]I_{n_r \times n_r}, & j = \{1, 3, \ldots, 2n-3\}, \\ I_{n_r \times 1}, & j = 2n-2, \end{cases}$$

where the case for odd $j$ follows from Theorem 1. The number of paths from root to goal in $T$ is then given as

$$|P_T| = \prod_{j=0}^{2(n-1)-1} A_j = \left[\frac{n_p+1}{2}\right]^{n-1}n_r^n.$$

The next theorem shows that there is a one-to-one correspondence between $P_T$ and $\mathcal{L}_T$.

**Theorem 3:** $|P_T| = |\mathcal{L}_T|$. 

**Proof:** The map $\lambda$ taking paths in $P_T$ to label sequences in $\mathcal{L}_T$ is surjective by definition. For $\lambda$ to be injective, every label sequence $\{(\ell_0, \ell_1, \ldots, \ell_{2n-2})\} \in \mathcal{L}_T$ must have a unique pre-image $p \in P_T$. The set of vertices through which $p$ traverses is uniquely determined by $\{(\ell_0, \ell_2, \ldots, \ell_{2n-2})\}$. Furthermore, the selection from among parallel edges in the odd trellis sections is uniquely determined by $\{(\ell_1, \ell_3, \ldots, \ell_{2n-3})\}$, since such parallel edges have distinct labels.

Fig. 3. The trellis for Example 1.

Fig. 4. Illustrating Theorem 1. The filled circles show the possible values for $\cos(\arg(ab))$. 
D. Finding Square-Law Distinct Sequences Using the Trellis

Finding SLD sequences for a relatively small constellation size and a small block length \( n \) is feasible by computer search. For example, for a constellation \( K \), one may produce two matrices, \( Q \in K_{|K| \times n} \) and \( M \in \mathbb{R}^{(|K| \times (2n-1))} \) where the rows of \( Q \) are all possible \( n \)-vectors defined over \( K \) and the rows of \( M \) are the corresponding signatures. Then, one may find all SLD sequences (which are some rows of \( Q \)) by deleting the duplicated rows of \( M \).

For large values of constellation size or block length \( n \), finding those sequences by exhaustive search is computationally intractable. Fortunately, SLD sequences drawn from an \((n_r, n_p)\)-SQAM constellation can be found using the trellis described in Section III-C.

Recall that two \( n \)-tuples \( s \) and \( \tilde{s} \) in \( K_{n_r, n_p}^n \) are square-law equivalent, written \( s \equiv \tilde{s} \), if \( \Upsilon(s) = \Upsilon(\tilde{s}) \), i.e., if the two \( n \)-tuples have the same signature. The set of such \( n \)-tuples is partitioned into disjoint equivalence classes. A transversal of this collection of classes is then any subset \( T \subset K_{n_r, n_p}^n \) containing exactly one element from each equivalence class. We call such a transversal an SLD-transversal since, by definition, the elements of any such transversal are SLD. A codebook \( S \) is, by definition, a subset of an SLD-transversal; therefore, SLD-transversals are the largest possible codebooks that can be used for the system described in Section II.

**Theorem 4:** For every SLD-transversal \( T \),

\[
\{ \Upsilon(s) : s \in T \} = \mathcal{L}_T.
\]

In other words, the signature of each \( n \)-vector in an SLD-transversal is a label sequence in \( \mathcal{L}_T \), and each label sequence in \( \mathcal{L}_T \) is the signature of an \( n \)-vector in the SLD-transversal.

**Proof:** This follows from the definitions of \( \mathcal{A}_{\text{List-free}} \) and \( \mathcal{A}_{\text{ISL-present}} \) and Theorem 3.

**Corollary 4.1:** The maximum rate that can be achieved by an \((n_r, n_p)\)-SQAM constellation with the system described in Section II, as \( n \to \infty \), is \( \log_2(n_r \left\lceil \frac{n_p+1}{2} \right\rceil) \) b/sym.

**Proof:** This follows Theorems 2, 3, and 4.

Note that the maximum achievable rate for an \((n_r, n_p)\)-SQAM constellation under coherent detection is \( \log_2(n_r) \) and

\[
0 < \log_2(n_r) - \log_2 \left( n_r \left\lceil \frac{n_p+1}{2} \right\rceil \right) < 1.
\]

According to Theorem 4, SLD-transversals are all mapped to the set \( \mathcal{L}_T \) by the function \( \Upsilon \). Thus every SLD-transversal is equivalent from the perspective of the detector, i.e., all of SLD-transversals will have the same error rate. In the following, we describe an algorithm that uses the trellis \( T \) to determine a particular SLD-transversal.

An \( n \)-vector \((s_0, \ldots, s_{n-1}) \in K_{n_r, n_p}^n \) is called standard if

1) \( s_0 \in \mathbb{R}^{>0} \), and
2) \( \arg(s_{\ell+1} s_{\ell}^*) \in [0, \pi], \) for every \( \ell \in [0, n-2] \).

In other words, the first component of a standard vector is a positive real number, and the argument of each component is obtained from the argument of the previous component by adding (modulo \( 2\pi \)) an angle in the range \([0, \pi] \).

Define a family of functions

\[
\varphi(\cdot; a, b) : \left[ \frac{3}{8}(a^2 + b^2) - \frac{ab}{4}, \frac{3}{8}(a^2 + b^2) + \frac{ab}{4} \right] \rightarrow [0, \pi],
\]

indexed by positive real numbers \( a \) and \( b \), and given as

\[
\varphi(\cdot; a, b) = \arccos \left( \frac{8c - 3(a^2 + b^2)}{2ab} \right).
\]

Given two complex numbers \( z_a \) and \( z_b \) such that \( |z_a| = a \), \( |z_b| = b \), and \( \psi(z_a, z_b) = \zeta \), there are at most two possible values of \( \arg(z_a z_b^*) \), namely \( \varphi(z_a, a, b) \in [0, \pi] \) and \( 2\pi - \varphi(z_a, a, b) \in [\pi, 2\pi] \). The function \( \varphi(\cdot; a, b) \) is defined to select the first of these possibilities.

**Theorem 5:** Each equivalence class (under square-law equivalence) \( C \subset K_{n_r, n_p}^n \) contains exactly one standard vector.

**Proof:** The signature of each vector in \( C \) is the same, given as, say, \( \zeta = (\zeta_0, \ldots, \zeta_{2n-2}) \in \mathbb{R}^{2n-1} \). Let

\[
s_0 = \sqrt{\zeta_0},
\]

and, for \( \ell \in [0, n-2] \), let

\[
s_{\ell+1} = \frac{\sqrt{2c} z_\ell}{s_\ell} \exp \left( i \varphi \left( \zeta_{2\ell+1}; \sqrt{\zeta_{2\ell}}, \sqrt{\zeta_{2\ell+2}} \right) \right).
\]

Let \( s = (s_0, \ldots, s_{n-1}) \). From (14) and (15) we have \( \Upsilon(s) = \zeta \); thus, \( s \in C \). Furthermore, by construction, \( s \) is standard. Uniqueness follows from the fact that the function \( \varphi \) always chooses (from two possibilities) the unique phase difference between successive components that results in a standard vector.

Define \( \Upsilon^{-1} : \mathcal{L}_T \to \mathcal{T} \), that, given a signature \((\zeta_0, \ldots, \zeta_{2n-2}) \in \mathcal{L}_T \) returns the standard vector with that signature, as determined by (14) and (15).

The (unique) SLD-transversal containing only standard vectors, denoted by \( \mathcal{T} \), is called the standard SLD-transversal. The elements of \( \mathcal{T} \) can be computed by applying \( \Upsilon^{-1} \) to the label sequences obtained by following every possible path from root to goal in the trellis \( T \). When \( |\mathcal{T}| \) is sufficiently small, the elements of \( \mathcal{T} \) are stored in a lookup table, thereby allowing for an efficient mapping between messages and codewords. This is the approach followed for the codebooks \( S \) studied in this paper, which are obtained as power-of-two subsets of \( \mathcal{T} \).

For larger codebooks, the mapping between messages and codewords becomes more complicated, possibly requiring enumerative encoding techniques to select a suitable path through \( T \). However, once such a path is selected, computation of a standard vector is readily accomplished by applying \( \Upsilon^{-1} \) to the corresponding label sequence.

E. Trellis Decoding

In addition to determining \( \mathcal{T} \), the trellis \( T \) of Section III-C can be used with the Viterbi algorithm to detect the sequence \( \hat{x} \in S \) from the noisy received vectors \( y = (y_0, \ldots, y_{n-1}) \) and \( z = (z_0, \ldots, z_{2n-2}) \). For the reasons of simplicity noted in Section II-F, we use (11) and (12) in branch-metric computation to approximate true ML detection.
Consider the transmission of the standard vector corresponding to the label sequence associated with some trellis path \( p = (e_0, \ldots, e_{2n-2}) \in \mathcal{T}. \) From (5), (6), (9), (10), and Theorem 4 one may see that, in the absence of noise, the received values are (to a close approximation) a scaled version of the edge labels. In particular, in the noise-free case, \( y_k \simeq \Re e^{-e^{L}(E_{in} \kappa \alpha \beta)^2 \lambda(e_k)} \) and \( z_k \simeq \Re (1 - \beta)e^{-e^{L}(E_{in} \kappa \alpha \beta)^2 \lambda(e_{2k+1})}, \) for any \( k \in [0, n - 1] \) and \( \ell \in [0, n - 2]. \)

Now in the presence of noise, from (11) and (12) it follows that the mean and the variance of the received \( y \) and \( z \) entries are, respectively, linear and affine functions of the entries of the label sequence \( \lambda(p) \) in our approximation. In particular, for any edge \( e \in \mathcal{E}, \) let \( \text{Mean} : \mathcal{E} \to \mathbb{R}, \) and \( \text{Var} : \mathcal{E} \to \mathbb{R} \) be defined as
\[
\text{Mean}(e) = \Re e^{-e^{L}(E_{in} \kappa \alpha \beta)^2 \lambda(e)}, \\
\text{Var}(e) = e^{-e^{L}(E_{in} \kappa \alpha \beta)^2 \sigma_{\text{sh}}^2 \lambda(e)} + T \sigma_{\text{sh}}^2.
\]

Thus for an edge \( e_{2j} \) in the \( (2j)\text{th} \) trellis section, \( j \in [0, n - 1], \) incident from a vertex of even depth, the mean and the variance of \( y_j \) are given as \( (1 - \beta)\text{Mean}(e_{2j}) \) and \( (1 - \beta)\text{Var}(e_{2j}), \) respectively. Similarly for an edge \( e_{2j+1} \) in the \( (2j + 1)\text{th} \) trellis section, \( j \in [0, n - 2], \) incident from a vertex of odd depth, the mean and the variance of \( z_j \) are given as \( \beta \text{Mean}(e_{2j+1}) \) and \( \beta \text{Var}(e_{2j+1}), \) respectively.

We define two branch-metric functions, \( w_{\text{ ISI-free }} : \mathbb{R} \times \mathcal{E} \to \mathbb{R}, \) and \( w_{\text{ ISI-present }} : \mathbb{R} \times \mathcal{E} \to \mathbb{R}, \) as follows. Let
\[
w_{\text{ ISI-free }}(y, e) = \frac{(y - (1 - \beta)\text{Mean}(e))}{(1 - \beta)\text{Var}(e)} + \ln \left( \text{Var}(e) \right),
\]
and let
\[
w_{\text{ ISI-present }}(z, e) = \frac{(z - \beta\text{Mean}(e))}{\beta\text{Var}(e)} + \ln \left( \text{Var}(e) \right),
\]
for any \( y \) and \( z \in \mathbb{R} \) and any \( e \in \mathcal{E}. \)

To detect \( \hat{x} \) from \( y \) and \( z, \) one may run the Viterbi algorithm on \( \mathbb{T} \) with branch metrics defined above to determine the path \( p_{\text{mw}} \in \mathcal{T} \) of minimum total weight, computing
\[
p_{\text{mw}} = \arg \min_{(e_0, \ldots, e_{2n-2}) \in \mathcal{T}} \sum_{j \in [0, n - 1]} w_{\text{ ISI-free }}(y_j, e_{2j}) + \sum_{j \in [0, n - 2]} w_{\text{ ISI-present }}(z_j, e_{2j+1}).
\]

The detected codeword is then \( \hat{x} = T^{-1}(\lambda(p_{\text{mw}})), \) provided that \( \hat{x} \in \mathcal{S}. \) In case \( \hat{x} \not\in \mathcal{S}, \) a decoding failure is declared. In practice, \( \mathcal{S} \) is selected to have as many codewords from \( \mathcal{T} \) as possible, while still having a size that is a power of two. Thus the majority of paths in \( \mathcal{T} \) are codewords, and decoding failures are rare. Indeed, the numerical simulations of Section IV do indeed confirm that, at the launch powers required to achieve reasonably small decoding error rates, the probability of decoding failure is negligible.

IV. NUMERICAL SIMULATION

In this section, the schemes described in Sections II and III are verified by numerical simulations.

A. Simulation Setup

Because we consider square-law detection with both shot noise and thermal noise, a notion of signal-to-noise ratio (SNR), as would be appropriate for linear additive white Gaussian channels, is not appropriate here. Accordingly, the various figures of merit considered here are shown as functions of launch power. We use practical values for \( \sigma_{\text{sh}}^2 \) and \( \sigma_{\text{th}}^2 \) in all cases.

In all simulations we have assumed a transmission length of \( L = 10 \text{ km}. \) Furthermore, laser power refers to the mean-squared power of the unmodulated waveform, i.e., \( \frac{1}{2} E_{in}^2. \)

In [1] it is shown that \( \beta = 0.9 \) provides near-optimal error rates and achievable rate performance versus launch power. However, \( \beta = 0.9 \) requires integration intervals as short as 10% of a symbol period, which may be difficult to implement at high symbol rates. To tackle this issue, we have used \( \beta = 0.5 \) in all simulations of this paper, corresponding to an integration period that is 50% of the symbol period for both the ISI-free and ISI-present signalling segments.

The practical ease of \( \beta = 0.5 \) comes at the expense of a worse error rate and a larger bandwidth. Note that Tukey waveforms are time-limited, thus, band-unlimited. Therefore, the typical notion of bandwidth for time-limited waveforms, e.g., sinc and raised-cosine waveforms, does not apply to Tukey waveforms. For this reason, we use 95% in-band-energy as the bandwidth criterion for comparing Tukey waveforms with each other. While \( w_{0.5} \) has a 15% spectral overhead compared to the minimum required bandwidth for Nyquist signalling, the spectral overhead of \( w_{0.5} \) is 33.6%. However, spectral efficiency may not be an important performance criterion in some applications, e.g., in data-center interconnects, and so the spectral overhead of Tukey waveforms may not be an important concern in such applications.

B. The Photodiode

The simulations in this chapter are done by assuming the use of an InGaAs p-i-n photodiode at the receiver. This choice gave better performance than avalanche photodiodes in our simulations. The reason for its better performance is that, as explained in Section II-F, the performance of direct detection schemes depends on the symbol rate and at sufficiently large symbol rates shot noise dominates thermal noise. As noted by Agrawal, “the SNR of APD receivers is worse than that of p-i-n receivers when shot noise contribution is concerned.” [16].

The two-sided power spectral density of the thermal noise is [16]
\[
\sigma_{\text{th}}^2 = \frac{2kT_k}{R_L},
\]
where \( k \) is the Boltzmann constant, \( T_k \) is the absolute temperature, and \( R_L \) is the load resistor. In the simulations we
have used $T_k = 300$ K and $R_L = 300$ Ω. Furthermore, the two-sided power spectral density of $n_{th}(t)$ of a p-i-n photodiode is $\sigma_{\text{th}}^2 = eR$, where $e$ is the elementary charge [16]. The typical range for the responsivity of an InGaAs p-i-n photodiode is $0.6–0.9$ mA/mW [16]; in the simulations, we have used $R = 0.75$ mA/mW.

### C. Codebook-Size Constraint

In all of our numerical simulations, we have assumed that $|S| = M$ is a power of two. For example, for the (8,4)-SQAM constellation and with a block length of 5, we have chosen $2^{21} = 2097152$ of the $|T_s| = 2654208$ available SLD sequences to form $S$. Under this power-of-two restriction, the maximum achievable rate of the proposed scheme using $(n_r, n_p)$-SQAM with a block length $n$ is

$$\frac{1}{n} \log_2 \left( \frac{(n_p + 1)^{n-1} - n_r^n}{2} \right) \text{ b/sym.}$$

### D. Optimizing Ring Spacing

A pair of $n_r$ and $n_p \in \mathbb{Z}^{\geq 2}$ does not uniquely specify an $(n_r, n_p)$-SQAM constellation; one must also specify the radius set $D$. In this section, for a fixed $n_r$ and $n_p$ we try to find the best value of $\delta \in \mathbb{R}^{\geq 0}$ such that

$$D = \left\{ 1, 1 + \delta, 1 + 2\delta, \ldots, 1 + (n_r - 1)\delta \right\}. \quad (16)$$

Fig. 5 shows the BER of (8,4)-SQAM constellations for different values of $\delta$. For these constellations $M = |S| = 2^{12}$ and $n = 3$; thus, their maximum achievable rate is 4 b/sym. One can see that among the tested $\delta$ values, $\delta = 0.2$ has the best performance.

Table I shows the best $\delta$ for different constellations. The tested values of $\delta$ are similar to those of Fig. 5. The criteria for choosing the best $\delta$ is the one which results in the least launch power at BER of $10^{-3}$. The rows which are not highlighted are those for which we can achieve at least as great a rate, in b/sym, with another $(n_r, n_p)$-SQAM constellation at the same symbol rate, yet with smaller launch power. For example, the row of (2.5)-SQAM constellation is not highlighted as at 50 Gbaud it achieves a rate of about 2.2 b/sym at a launch power of $-9$ dBm while (3.2)-SQAM constellation achieves 2.43 b/sym at a launch power of $-12.2$ dBm. Thus, we are interested only in highlighted rows.

Theorem 2 implies that the maximum codebook size, $|T_s|$, depends on $\frac{n_r+1}{2}$. As a result, for a fixed block length $n$ and a radius set $D$, an $(n_r, n_p)$-SQAM constellation with an even $n_p$ results in the same transversal cardinality as an $(n_r, n_p+1)$-SQAM constellation. Therefore, as the ring points of the former constellation are further apart than those of the latter one, we expect the $(n_r, n_p)$-SQAM constellation to have a better performance than the $(n_r, n_p+1)$ constellation. Table I does indeed support this claim, as all highlighted rows have an even $n_p$.

### E. Bit Error Rate

Fig. 6 shows the trellis decoding bit error rate of the proposed scheme for the constellations highlighted in Table I at 50 Gbaud and with 1 dBm laser power. As it is apparent, the BER curves have two parts: 1) a roll-down part where the BER decreases with a higher launch power, and 2) a roll-up part which behaves in the opposite manner. The reason for the roll-up part is that the proposed decoder approximates the transmitted waveform as (4), while the waveform actually suffers from nonlinear distortion according to (1). As mentioned in II-F, this is a good approximation when the power of the modulating waveform, $u(t)$, is relatively small. For a fixed laser power, the launch power is controlled by the modulating-waveform power; i.e., a high launch-power demands a higher modulating-waveform.
power which degrades the accuracy of approximating (1)
with (4).

Fortunately, a practical raw BER, e.g., about \(10^{-3}\), can
be achieved in the roll-down area for all considered constellations,
except (12,4)-, (13,4)-, and (16,4)-SQAM constellations. This
is the reason that some cells corresponding to those particular
constellations in Table I are left blank. For example, the smallest
BER obtained by a (16,4)-SQAM constellation at 50 Gbaud is
about 0.011, is achieved at a launch power about \(-7.9\) dBm and
with \(\delta = 0.2\).

As explained in Section III-E, the power-of-two constraint
may cause the trellis decoder to declare a decoding failure.
Fig. 7 shows the decoding failure probability for the same
constellations as Fig. 6. The constellations which are absent in
Fig. 7 but are present in Fig. 6 had no decoding failure at the
tested launch power values. By comparing Fig. 7 with 6 one may
see that at launch power values with a practically small BER,
e.g., \(10^{-3}\), the decoding failure rate is negligible.

As noted in Section II-F, unlike a typical AWGN communica-
tion scheme, the performance under direct detection depends
on the symbol rate. This fact is illustrated in Fig. 8 which provides
the BER curves in the same setup as Fig. 6, except that it is at
25 Gbaud. By comparing these two figures one may see that
for a fixed constellation, a BER of \(10^{-3}\) can be achieved with
approximately 1.5 dB less launch power at 25 Gbaud than at
50 Gbaud.

Note that in computing BER we treat decoding failures as
equivalent to a momentary BER of 0.5. We made no attempt
to optimize (e.g., by Gray labelling) the mapping of sequences
\(x \in S\) to binary vectors of length \(\log_2(|S|)\).

F. Achievable Data Rate

Fig. 9 shows the theoretically maximum achievable data
rate for a few SQAM constellations at 50 Gbaud and with a
1 dBm laser power, obtained by Monte Carlo simulations. Except
the (16,4)-SQAM constellation, other constellations are highlighted in Table I, but
with a possibly different block length. Curves are obtained at 50 Gbaud and with
1 dBm laser power.
maximum achievable rate of each constellation can be increased up to $\frac{1}{4}$ bits. For example, for the (5,2)-SQAM constellation with $n = 4$, the maximum achievable rate by using all SLD sequences is $\log_2 |T_n| \simeq 3.07$ b/sym. However, as only $2^{12} = 4096$ symbols have been chosen out of 5000 SLD sequences, the rate has decreased to 3 b/sym.

Fig. 9 shows the necessity of using error correcting codes from two perspectives. First, a desired rate can be achieved using a higher-order constellation along with an error correcting code at a lower launch power. For example, while the uncoded (8,4)-SQAM constellation with $n = 3$ can achieve 4 b/sym at a launch power of $-9.5$ dBm, the same rate can be achieved with a (10,4)-SQAM constellation with $n = 3$ and an error correcting code of rate $\frac{12}{15}$ at a launch power of $-12$ dBm; resulting in a coding gain of about 2.5 dBm. Secondly, at a fixed launch power, a higher data rate can be achieved by using a higher-order constellation along with an error correcting code. For example, at a launch power of $-12$ dBm, the uncoded (5,2)-SQAM constellation with $n = 4$ can achieve a rate of 3 b/sym, while by using the (8,4) constellation with $n = 3$ along with an error correcting code of rate $\frac{15}{16}$ one may achieve 3.75 b/sym.

Another fact which is shown in Fig. 9 is the role of the block length $n$ on the performance. While the maximum achievable rate for the (3,2)-SQAM constellation is 2 b/sym with a block length of $n = 3$, it is 2.2 b/sym with $n = 5$. Note that, in contrast to [1], where brute-force search was used, the increase in data rate resulting from a larger $n$ does not come with a huge increase in decoding complexity as the complexity of trellis decoding is linear in $n$.

G. Effect of Laser Power

Fig. 10(a) shows the BER of the (8,4)-SQAM constellation with $\delta = 0.2$, as recommended by Table I, $n = 5$, and at 50 Gbaud. Note that the BER curves corresponding to a laser power of $\geq -1$ dBm superimpose in the region of interest, i.e., the region with BER about $10^{-3}$. However, the BER performance degrades as laser power drops further. The reason for this behavior is again in approximating (1) with (4). For a fixed launch power, a lower laser power demands a higher modulating-waveform power and it violates the condition for an accurate approximation of (1) by (4).

From Fig. 10(a) one may conclude that the laser power determines the BER at which the roll-down part of the BER curve meets the roll-up part. In other words, the larger is the laser power, the smaller is the BER at the transition point. For example, while the roll-down to roll-up transition occurs at a BER of about $2 \times 10^{-2}$ for a laser power of $-3$ dBm, it happens at a BER of $10^{-3}$ and $<10^{-4}$ for a laser power of $-1$ dBm and 1 dBm, respectively. Therefore, depending on the minimum required uncoded BER, one may determine the minimum required laser power.

Fig. 10(b) shows the decoding failure rate with the same simulation setup as in Fig. 10(a). Interestingly, in contrast to the BER curves, the smaller is the laser power, the smaller is the decoding failure rate. However, the smaller decoding failure rate comes with a higher bit error rate. From the spacing between the decoding failure rate curves it is apparent that the curves approach a limiting curve as the laser power increases. However, note that launch power values which result a target BER, e.g., about $10^{-3}$, the decoding failure rate is almost negligible.

Fig. 11 shows the BER curves for the (2,2)-SQAM constellation with $\delta = 0.2$ and block length $n = 7$. The maximum achievable rate for this constellation and $n$ is $\frac{7}{2}$ b/sym. By comparing Figs. 10(a) with Fig. 11 one may see that increasing constellation size, thus rate, increases the minimum required laser power to get a target BER. For example, while we can achieve a BER of $10^{-3}$ with a $-5$ dBm or even $-7$ dBm laser power with the described (2,2)-SQAM constellation, the minimum required laser power to achieve the same BER with the described (8,4)-SQAM constellation is $-1$ dBm.

In short-haul applications, e.g., intra data-center communication, power consumed by lasers play an important component of the overall power budget. One may decide to allocate less power to lasers in the price of a lower data rate, in b/sym, or higher BER for a fixed constellation.
Fig. 11. The BER for the (2,2)-SQAM constellation with $n = 7$ and $\delta = 0.2$ (see (16)), and at 50 Gbaud, at different laser powers. Legends show the laser power in dBm.

Fig. 12. Throughput for IMDD and the proposed scheme at 50 Gbaud.

**V. COMPARING WITH IMDD**

In contrast to the majority of the literature, instead of experimental results all of our figures of merit are obtained by computer-based simulations. There are many practical issues, e.g., splicing, scattering, etc., which have been ignored in our numerical simulations. Therefore, to have a fair comparison, we have simulated IMDD using the same photodiode parameters that were used in our system.

Fig. 12 shows the throughput of the proposed and the IMDD schemes at 50 Gbaud for a single wavelength and a single polarization. While the (8,4)-SQAM Tukey signalling achieves a throughput of 200 Gb/s at a launch power of about $-10$ dBm, IMDD with PAM-16 achieves this rate at about 0 dBm; thus, by using the proposed scheme a gain of about $8$ dB can be achieved compared to the IMDD scheme. We can achieve the same throughput by using a (16,4)-SQAM constellation with $n = 3$ and an error correcting code of rate 0.8 at a launch power about $-12$ dBm, as well, i.e., a 2.9 dB coding gain. Furthermore, Fig. 12 shows that for a fixed launch power and a fixed symbol rate one may achieve higher throughputs by using the proposed scheme rather than IMDD. For example, while a throughput of 200 Gb/s is achievable with the proposed scheme at a launch power of $-10$ dBm and at 50 Gbaud with an (8,4)-SQAM constellation of block length $n = 3$, one may achieve only about 145 Gb/s with an IMDD scheme using PAM-8 constellation at the same symbol rate and the same launch power. Note that these two constellations have the same number of magnitude levels; the proposed scheme has a higher throughput since it is able to extract phase information from a complex-valued constellation.

**VI. O-BAND OPERATION**

So far we have assumed operation in the C band according to the system model shown in Fig. 1. Because of the non-negligible chromatic dispersion in this band and the difficulty of post-detection dispersion compensation after signal detection using a single photodiode, dispersion was (partially) precompensated at the transmitter, as discussed in Section II-D. However, pre-compensation of chromatic dispersion requires knowledge of the link length, which is generally not possible without a feedback channel.

To remedy this issue, one may operate in the O band, near the zero-dispersion wavelength of SSMF, omitting transmitter-side dispersion precompensation, and making the transmitter agnostic to the transmission distance. This changes the system model to that of Fig. 13. The advantage of having a link-length agnostic transmitter comes at the expense of having channel with greater loss compared to a C-band channel. In particular, the fiber loss in O band is about $0.5$ dB · km$^{-1}$.

Fig. 14 shows the O-band BER for the constellations highlighted in Table I, at 50 Gbaud, and at a laser power of 2 dBm. The chromatic dispersion is $1$ ps · nm$^{-1}$ · km$^{-1}$. Simulations performed with a chromatic dispersion of $-1$ ps · nm$^{-1}$ · km$^{-1}$ gave identical results.
exactly zero dispersion; instead we simulate operation at two wavelengths assumed to have chromatic dispersions of \(-1\) and \(1 \text{ ps} \cdot \text{nm}^{-1} \cdot \text{km}^{-1}\), respectively, leaving the channel with some uncompensated residual dispersion.

For a 10 km length of SSMF, the total power loss of the transmission link in the O band is 3 dB more than the total power loss in the C band. We see this by comparing the roll-down parts of the curves in Fig. 14 with their corresponding curves in Fig. 6. For example, one may see from Fig. 14 that the (8,4)-SQAM constellation with block length \(n = 3\) achieves a BER of \(10^{-3}\) at a launch power of \(-5.8 \text{ dBm}\) in the O band; while, from Fig. 6, this constellation achieves the same BER at a launch power of \(-8.8 \text{ dB}\) in the C band.

Note that this “3 dB difference” is violated for the (10,4)- and (13,4)-SQAM constellations. These large constellations achieve a target BER at a higher launch power compared to the other smaller constellations highlighted in Table I. Furthermore, due to the greater loss in the O band, a higher launch power is required to achieve a target BER. However, as explained in Section II-F and for a fixed laser power, the higher launch power results in a higher modulator nonlinearity. Thus, it is the nonlinearity of the modulator that hinders the performance of these large constellations in the O band. This issue may be remedied by operating at a higher laser power.

VII. DISCUSSION

In contrast to a long-haul or wireless communication system where the power budget of the transmitter and the receiver are decoupled, in short-haul applications the power budget is shared between the transmitter and receiver. As KK schemes use an unsynchronized laser at the receiver— send a tone along with the signal at transmitter—a fair comparison between KK and direct detection under Tukey signalling or with IMDD should be made under a fixed total power, i.e., including the receiver laser power (or the transmitter tone), as well. As a result, due to the high laser power at the receiver (or high tone power at the transmitter), the KK schemes are not power efficient for communication over short distances, i.e., <10 km. However, for longer distances, e.g., about 100 km, other imperfections of optical fiber become substantial, necessitating post-detection compensation. In those cases, by recovery of the complex-valued at the symbol rate, at the receiver.

IQ modulator at the transmitter and two ADCs, each operating at the symbol rate, at the receiver.

REFERENCES

[1] A. Tasbihi and F. R. Kschischang, “Direct detection under tukey signalling,” J. Lightw. Technol., vol. 39, no. 21, pp. 6845–6857, Nov. 2021.
[2] A. Lender, “The duobinary technique for high-speed data transmission,” IEEE Trans. Commun. Electron., vol. 82, no. 2, pp. 214–218, May 1963.
[3] P. Kabal and S. Pasupathy, “Partial-response signalling,” IEEE Trans. Commun., vol. 23, no. 9, pp. 921–934, Sep. 1975.
[4] A. Mecozzi, C. Antonelli, and M. Shtaif, “Kramers-Kronig coherent receiver,” Optica, vol. 3, no. 11, pp. 1220–1227, 2016.
[5] A. Mecozzi, C. Antonelli, and M. Shtaif, “Kramers-Kronig receivers,” Adv. Opt. Photon., vol. 11, no. 3, pp. 480–517, 2019.
[6] T. Bo and H. Kim, “Kramers-Kronig receiver operable without digital up-sampling,” Opt. Exp., vol. 26, no. 11, pp. 13810–13818, 2018.
[7] K. H. Powers, “The compatibility problem in single-sideband transmission,” Proc. IEEE, vol. 48, no. 8, pp. 1431–1435, Aug. 1960.
[8] B. Logan and M. Schroeder, “A solution to problem of compatible single-sideband transmission,” IRE Trans. Inf. Theory, vol. 8, no. 5, pp. 252–259, Sep. 1962.
[9] H. B. Voelcker, “Toward a uniform theory of modulation Part I: Phase-envelope relationship,” Proc. IEEE, vol. 54, no. 3, pp. 340–353, Mar. 1966.
[10] H. B. Voelcker, “Toward a uniform theory of modulation Part II: Zero manipulation,” Proc. IEEE, vol. 54, no. 5, pp. 735–755, May 1966.
[11] H. B. Voelcker, “Demodulation of single-sideband signals via envelope detection,” IEEE Trans. Commun., vol. 14, no. 1, pp. 22–30, Feb. 1966.
[12] G. B. Lockhart, “A spectral theory for hybrid modulation,” IEEE Trans. Commun., vol. 21, no. 7, pp. 790–800, Jul. 1973.
[13] G. C. Papen and R. E. Blahut, Lightweight Communications, Cambridge, U.K.: Cambridge Univ. Press, 2019.
[14] D. Plabst et al., “Achievable rates for short-reach fiber-optic channels with direct detection,” J. Lightw. Technol., vol. 40, no. 12, pp. 3602–3613, Jun. 2022.
[15] F. R. Kschischang and V. Sorokine, “On the trellis structure of block codes,” IEEE Trans. Inf. Theory, vol. 41, no. 6, pp. 1924–1937, Nov. 1995.
[16] G. P. Agrawal, Fiber–Optic Communication Systems, 4th ed. NJ, USA: Hoboken, NJ, USA: Wiley, 2010.

Amir Tasbihi (Member, IEEE) received the bachelor’s degree in electrical engineering from Tehran Polytechnic, Tehran, Iran, in 2014, and the M.A.Sc. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 2016, and the Ph.D. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 2022. He is currently a Postdoctoral Fellow. His research interests include deep-space optical communication, data-center interconnects, information theory, coding theory, and all aspects of physical layer communications.

Frank R. Kschischang (Fellow, IEEE) received the B.A.Sc. degree (with Hons.) in electrical engineering from the University of British Columbia, Vancouver, BC, Canada, in 1985, and the M.A.Sc. and Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1988 and 1991, respectively. Since 1991, he has been a Faculty Member of electrical and computer engineering with the University of Toronto, where he currently holds the title of Distinguished Professor of Digital Communication. His research interests include the area of channel coding techniques, applied to wireline, wireless and optical communication systems, and networks.

During 1997–2000, he was an Associate Editor for Coding Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY, and from 2014–16, he was the Editor-in-Chief of IEEE TRANSACTIONS ON INFORMATION THEORY. He was General Co-chair for the 2008 IEEE International Symposium on Information Theory, and he was the 2010 President of the IEEE Information Theory Society. He was recipient several awards for teaching and research, including the 2010 Communications Society and Information Theory Society Joint Paper Award and the 2018 IEEE Information Theory Society Paper Award, IEEE Richard W. Hamming Medal in 2023 and the Society’s Aaron D. Wyner Distinguished Service Award in 2016. He is a Fellow of the Engineering Institute of Canada, Canadian Academy of Engineering, and Royal Society of Canada.