Optimal Shaping for the Stokes Vector Receiver

Hongxiang Jia, Ethan Liang, *Graduate Student Member, IEEE*, and Joseph M. Kahn, *Fellow, IEEE*

***Abstract*—We discuss shaping with continuous and discrete input distributions for the Stokes vector receiver (SVR). We use the Forney method to derive tight analytical approximations to the optimal continuous input distribution at high signal-to-noise ratio (SNR). The Forney method analysis finds that an exponential distribution in the intensity (corresponding to the Stokes parameter S_0) is the optimal continuous input distribution for thermal noise-limited and amplifier noise-limited SVRs, providing ultimate shape gains of 1.056 dB and $\pi e/6 \approx 1.533$ dB, respectively. We also perform numerical studies of discrete input distributions obtained using the Blahut-Arimoto method or by sampling analytically derived continuous distributions. We find that a sampled exponential distribution in the intensity provides 0.078 dB and 0.165 dB higher shape gains than a sampled Gaussian distribution of the SV in the thermal noise-limited and amplifier noise-limited regimes, respectively. We also compare the performance of the SVR to that of a dual-polarization coherent receiver. We show that, owing to noise enhancement, the SVR incurs an SNR penalty of about 5.5 dB with respect to a coherent system that modulates three of the four available degrees of freedom.**

***Index Terms*—Fiber optic communication, probabilistic shaping, stokes vector receiver.**

I. Introduction

A STOKES vector receiver (SVR) decodes information by detecting the Stokes parameters of a received electric field. While a local oscillator (LO) can be used to detect the Stokes parameters of optical signals [1], this paper studies a receiver that does not employ an LO [2]. The SVR provides up to three independent degrees of freedom per symbol [3], [4] and is a hybrid of noncoherent and differentially coherent detection [5]. The SVR provides an intermediate solution between one-dimensional (1-D) intensity modulation with direct detection (IM/DD) and four-dimensional (4-D) standard coherent detection (SCD), as the SVR enables more degrees of freedom than IM/DD while avoiding the complexity associated with an LO at the receiver. These performance and complexity tradeoffs have led the SVR to be considered for short-reach, high-capacity links, such as inter- or intra-data center interconnects [6].

Probabilistic shaping (PS) involves the optimization of an input distribution to maximize the achievable information rate on a given channel. Shannon analyzed the average power-constrained additive white Gaussian noise (AWGN) channel and found that the channel capacity is achieved using a Gaussian input distribution [7]. Forney introduced an alternate method for computing the optimal input distribution based on constructing a high-dimensional lattice code and choosing the set of minimal energy signal points [8]. Forney's method yields an analytical approximation to the optimal continuous input distribution in the high signal-to-noise (SNR) regime. For the average power-constrained AWGN channel, Forney's method finds the optimized input distribution to be Gaussian.

Shaping for optical communications has been studied since the late 1990s. Some early works used the Forney method to find the optimal input distributions for IM/DD channels subject to constraints on the average transmitted power. For thermal noise-limited IM/DD channels, the optimal distribution at high SNR is an exponential distribution in the intensity [9], while for amplifier noise-limited IM/DD channels, it is a half-Gaussian distribution in the electric field magnitude [10]. Another early work studied lattice codes and average power-minimizing bounding regions for thermal noise-limited IM/DD channels [11]. In coherent optical systems using digital signal processing, efficient joint shaping and error correction encoding enabled by probabilistic amplitude shaping (PAS) [12] has facilitated the widespread adoption of shaping [13].

Shaping for SVR systems that encode information in the three Stokes parameters $\{S_1, S_2, S_3\}$, which is the most common practice, was studied in [2], [14]. The thermal noise-limited regime was considered in [2] and the amplifier noise-limited regime was considered in [14]. These two papers found the optimal input distribution considering a constraint on the trace of the covariance matrix of the transmitted Stokes vector (SV), which is equivalent to a constraint on the average detected electrical power. A multivariate Gaussian distribution of the transmitted SV was found to be the optimal input distribution in both noise regimes.

In this paper, we study shaping for systems that encode information in the three Stokes parameters $\{S_1, S_2, S_3\}$, under constraints on the average transmitted optical power, for both thermal noise-limited and amplifier noise-limited SVRs. Using the Forney method, we derive tight analytical approximations to the optimal continuous input distribution at high SNR. We find that an exponential distribution in the intensity (corresponding to the Stokes parameter S_0) is the optimal continuous input distribution for both thermal noise-limited and amplifier noise-limited SVRs, providing ultimate shape gains of 1.056 dB and $\pi e/6 \approx 1.533$ dB, respectively. We also perform numerical studies of finite discrete input distributions obtained by

Manuscript received 23 February 2023; revised 13 June 2023 and 18 July 2023; accepted 20 July 2023. Date of publication 27 July 2023; date of current version 16 November 2023. This work was supported by the National Science Foundation under Grant DGE-1656518. *(Hongxiang Jia and Ethan Liang contributed equally to this work.)* (Corresponding author: Ethan Liang.)

The authors are with the E. L. Ginzton Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: jiahx20@stanford.edu; emliang@stanford.edu; jmk@ee.stanford.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2023.3299306.

Digital Object Identifier 10.1109/JLT.2023.3299306

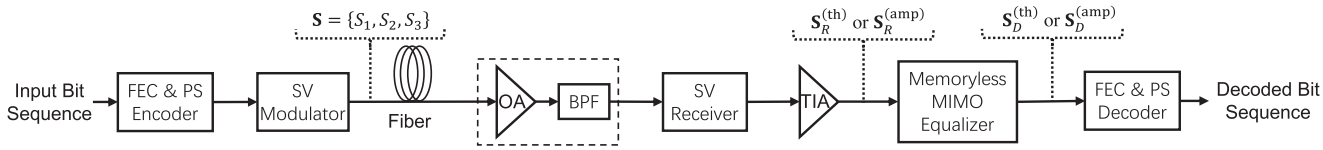

Fig. 1. Block diagram of an optical communication system that modulates and detects symbols using a Stokes vector modulator and Stokes vector receiver, respectively. The components in the dashed box are used only in the amplifier noise-limited case and are omitted in the thermal noise-limited case. The symbols with the "(th)" and "(amp)" superscripts denote the symbols for the thermal and amplifier noise-limited cases, respectively. FEC: forward error correction, PS: probabilistic shaping, SV: Stokes vector, OA: optical amplifier, BPF: bandpass filter, MIMO: multiple-input multiple-output.

the Blahut-Arimoto (BA) method or by sampling analytically derived continuous distributions. We find that a sampled exponential distribution in the intensity provides 0.078 dB and 0.165 dB higher shape gain than a sampled multivariate Gaussian distribution in $\{S_1, S_2, S_3\}$ in the thermal noise-limited and amplifier noise-limited regimes, respectively. We compare the performance of the SVR to a reference dual-polarization coherent receiver, showing that because of noise enhancement, the SVR incurs an SNR penalty of about 5.5 dB with respect to a coherent system that modulates only three of the four available degrees of freedom. Noise enhancement in the SVR occurs because each detected output contains beat-noise terms from both quadratures in both polarizations, whereas each output of the reference coherent receiver is corrupted only by noise in the corresponding quadrature and polarization.

While we study systems that modulate information onto all three Stokes parameters $\{S_1, S_2, S_3\}$, alternative SVR encoding and detection strategies have been studied. Some system architectures transmit an unmodulated carrier on one of the polarizations and modulate information onto the other polarization. The carrier is used to recover the transmitted electric field on the other polarization, achieving 2-D modulation and enabling techniques such as electronic compensation of group velocity dispersion (GVD) [3], [15], [16]. An additional degree of freedom can be achieved by differentially encoding and decoding successive three-dimensional (3-D) Stokes symbols [17]. These alternative schemes are beyond the scope of this work.

The remainder of this paper is organized as follows. Section II presents the SVR system architecture with shaping and derives discrete-time channel models for the thermal noise-limited and amplifier noise-limited regimes. Section III uses the Forney method to analytically approximate the optimal continuous input distribution in both noise regimes. Section IV presents numerical results on finite input constellations for the optimal input distribution and various sampled analytical distributions. Section V analyzes the properties of the optimized discrete input distribution and compares the SVR against a reference coherent receiver. Section VI concludes the paper.

II. SVR System Architecture With Shaping

This section presents our model for optical communication systems employing SVRs. Section II-A provides an overview of the model while Section II-B derives the discrete-time channel model and introduces the associated notation.

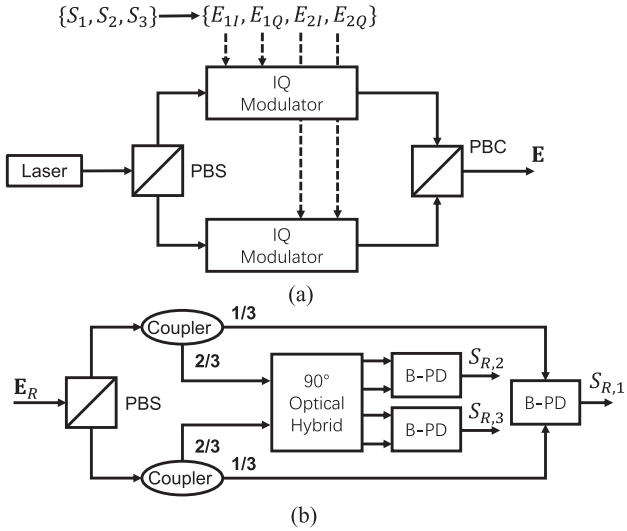

Fig. 2. Architecture of the (a) transmitter and (b) receiver for the SVR. IQ: in-phase and quadrature, PBS/PBC: polarization beam splitter/combiner, B-PD: balanced photodetector.

A. Overview of the Stokes Vector Receiver

Fig. 1 shows a block diagram of a generic optical communication system that transmits and receives information using the Stokes parameters. The system begins by performing forward error correction (FEC) and PS encoding. Constant composition distribution matching and sphere shaping are two commonly used shaping methods for mapping information bits into a sequence of symbols with a specified target probability distribution [18]. This paper does not presume the use of a particular shaping architecture but assumes that the PS encoder can achieve any desired target probability distribution.

The sequence of shaped symbols is encoded onto the Stokes parameters of the electric field using the SV modulator. Fig. 2(a) shows the structure of the SV modulator. The SV $\mathbf{S} = [S_1, S_2, S_3]^T$ is modulated by setting the corresponding Jones vector $\mathbf{E} = [E_1, E_2]^T = [E_{1I}, E_{1Q}, E_{2I}, E_{2Q}]^T$. The SV can be expressed in terms of its corresponding Jones vector by

$$S_0 = |E_1|^2 + |E_2|^2, \tag{1}$$

$$S_1 = |E_1|^2 - |E_2|^2, \tag{2}$$

$$S_2 = 2\text{Re}\left(E_1 E_2^*\right), \tag{3}$$

$$S_3 = 2\text{Im}\left(E_1 E_2^*\right). \tag{4}$$

Only three components of the SV are independent, providing up to three degrees of freedom for modulation. $\{S_1, S_2, S_3\}$ are real, bipolar parameters and are used for modulation [2]. S_0 is the optical power of the SV, and is related to the other three Stokes parameters by $S_0 = \sqrt{S_1^2 + S_2^2 + S_3^2}$. S_0 is assumed to be non-negative throughout the paper, i.e., $S_0 \geq 0$. The three modulated SV parameters are encoded onto minimum-bandwidth pulses in the optical electric field and therefore occupy a bandwidth B about the carrier frequency. The modulated signal is coupled into the single-mode fiber and is transmitted to the receiver. The received signal is optionally amplified by an optical amplifier (OA), followed by an optical bandpass filter (BPF) with bandwidth B about the carrier frequency.

The SV receiver transforms the Stokes parameters of the received electric field into three currents. Fig. 2(b) shows the structure of the SV receiver. In the absence of noise, this specific SV receiver configuration generates three currents proportional to the three received Stokes parameters. While alternative SVR architectures exist [2], [15], [17], we analyze the structure in Fig. 2(b) for the following reasons. This receiver configuration has the fewest electrical outputs and consequently uses the fewest analog-to-digital converters. This configuration also directly outputs the received SV without additional processing, resulting in identical noise distributions on each of the detected Stokes parameters. The other SV receivers analyzed in [2], [15], [17] do not directly output signals proportional to the received Stokes parameters or may have different noise distributions.

The currents generated by the SV receiver are subsequently converted to voltages by a transimpedance amplifier (TIA), which is typically the dominant source of thermal noise. Each electrical current is filtered by an electrical low-pass filter with cutoff frequency B and sampled at the symbol rate. A memoryless multiple-input multiple-output (MIMO) equalizer compensates for the effects of the fiber. The equalizer output is transformed to the decoded bit sequence by the FEC & PS decoder.

The model does not include distortion arising from GVD, polarization mode dispersion, Kerr nonlinearity, polarization-dependent loss, or polarization-dependent gain. The SVR system model does not transmit a carrier and is therefore especially suited to short-reach links that do not require equalization of GVD. We study the regimes in which the dominant noise is either (a) thermal noise from the TIA or (b) amplified spontaneous emission (ASE) from the optical amplifier.

B. Channel Model and SNR Definition

In the continuous-time thermal noise-limited model, the TIA output voltage includes a continuous-time Gaussian noise with one-sided power spectral density N_0. The electrical signal is filtered by an ideal low-pass filter with a cutoff frequency equal to the symbol rate B and sampled at the symbol rate. The variance of the discrete-time noise sample is $\sigma_{\text{th}}^2 = N_0 B$. The discrete-time thermal noise-limited model for the received SV $\mathbf{S}_R^{(\text{th})}$ is given by

$$\mathbf{S}_R^{(\text{th})} = \mathbf{H}\mathbf{S} + \mathbf{N}^{(\text{th})}, \tag{5}$$

where $\mathbf{H}$ is a 3×3 orthogonal rotation matrix, $\mathbf{S} = [S_1, S_2, S_3]^T$ is the transmitted SV, and $\mathbf{N}^{(\text{th})} = [N_1^{(\text{th})}, N_2^{(\text{th})}, N_3^{(\text{th})}]^T$ is an AWGN vector with covariance matrix $\text{cov}\{\mathbf{N}^{(\text{th})}\} = \sigma_{\text{th}}^2 \mathbf{I}_3$. $\mathbf{H}$ is assumed to have channel gain factor ς, which is defined by $\mathbf{H}^H\mathbf{H} = \varsigma \mathbf{I}_3$.

The effects of the channel gain and rotation matrix can be inverted by left multiplying the detected signal by $\mathbf{H}^H/\varsigma$. It is assumed that the receiver has perfect knowledge of $\mathbf{H}$, which can be obtained in practice by blind channel estimation [2]. After scaling and de-rotation, the demodulated vector $\mathbf{S}_D^{(\text{th})} = [S_{D,1}^{(\text{th})}, S_{D,2}^{(\text{th})}, S_{D,3}^{(\text{th})}]^T$ is given by

$$\mathbf{S}_D^{(\text{th})} = \frac{1}{\varsigma}\mathbf{H}^H\mathbf{S}_{\mathbf{R}}^{(\text{th})} = \mathbf{S} + \frac{1}{\varsigma}\mathbf{H}^H\mathbf{N}^{(\text{th})}. \tag{6}$$

The SNR in the thermal noise-limited regime, $\text{SNR}^{(\text{th})}$, is defined as the ratio of the average detected electrical signal power to average noise power per real dimension. The average received electrical signal power is $\langle \|\mathbf{S}\|_2^2 \rangle = \langle S_0^2 \rangle$, where $\langle \cdot \rangle$ denotes the expected value operator. The average noise power per real dimension is $\frac{1}{3}\,\text{tr}\{\text{cov}\{\frac{1}{\varsigma}\mathbf{H}^H\mathbf{N}^{(\text{th})}\}\} = \frac{1}{\varsigma}\sigma_{\text{th}}^2$, resulting in

$$\text{SNR}^{(\text{th})} = \frac{\varsigma\langle S_0^2 \rangle}{\sigma_{\text{th}}^2}. \tag{7}$$

We now define the discrete-time channel model in the amplifier noise-limited regime. A similar derivation of the discrete-time channel model in the amplifier noise-limited regime is given in [14]. The dominant noise source is ASE noise at the output of the optical amplifier, which is modeled by a complex white Gaussian noise process with one-sided power spectral density S_{ASE} in each of the two polarizations. Suppose an optical signal with optical power P_{in} and ASE noise are input to the SVR depicted in Fig. 2(b). Using the monochromatic signal approximation, one can show that the signal-spontaneous beat noise at the output of each balanced photodetector has a one-sided power spectral density $4R_d^2 P_{\text{in}} \frac{S_{\text{ASE}}}{2}$ and is uniform over the frequency interval $[0, B/2]$ [19]. R_d is the responsivity of the photodetector, P_{in} is the power of the signal input to the photodetector, and B is the bandwidth of the optical bandpass filter following the amplifier. We assume throughout this paper that $R_d = 1$ A/W. The current generated by the photodetector is filtered using an ideal electrical low-pass filter with cutoff frequency B. The total power of the signal-spontaneous beat-noise after the electrical low-pass filter is given by $4R_d^2 P_{\text{in}} \frac{S_{\text{ASE}}}{2} \frac{B}{2}$.

Now suppose a discrete-time optical signal with power P_{in} and a circularly symmetric Gaussian random vector with variance σ_{amp}^2 in each real dimension are input to the SVR depicted in Fig. 2(b). The power of the discrete-time signal-spontaneous beat noise at the output of each photodetector would be $4R_d^2 P_{\text{in}} \sigma_{\text{amp}}^2$. Equating the powers of the signal-spontaneous beat noise under the monochromatic approximation in the continuous-time and discrete-time models, we set $\sigma_{\text{amp}}^2 = \frac{1}{4} S_{\text{ASE}} B$.

In the amplifier noise-limited discrete-time model, the ASE noise is modeled as a circularly symmetric complex AWGN vector $\mathbf{N}^{(\text{amp})} = [N_1, N_2]^T = [N_{1I} + jN_{1Q}, N_{2I} + jN_{2Q}]$ with $\text{cov}\{\mathbf{N}^{(\text{amp})}\} = 2\sigma_{\text{amp}}^2 \mathbf{I}_2$. The received electric field vector

$\mathbf{E}_R^{(\text{amp})}$ is given by

$$\mathbf{E}_R^{(\text{amp})} = \mathbf{U}\mathbf{E} + \mathbf{N}^{(\text{amp})} \tag{8}$$

$$= \mathbf{E}' + \mathbf{N}^{(\text{amp})}, \tag{9}$$

where $\mathbf{E} = [E_1, E_2]^T$ is the transmitted electric field, $\mathbf{E}' = [E_1', E_2']^T$ is the noiseless received electric field, and $\mathbf{U}$ is a scaled unitary matrix satisfying $\mathbf{U}\mathbf{U}^H = \sqrt{\varsigma}\mathbf{I}_2$.

The SNR in the amplifier noise-limited regime $\text{SNR}^{(\text{amp})}$ is defined as the average received optical power divided by optical noise power in one polarization and is given by

$$\text{SNR}^{(\text{amp})} = \frac{\sqrt{\varsigma}\left\langle\sqrt{S_1^2+S_2^2+S_3^2}\right\rangle}{2\sigma_{\text{amp}}^2} = \frac{\sqrt{\varsigma}\langle S_0\rangle}{2\sigma_{\text{amp}}^2}. \tag{10}$$

After detection, the received Stokes parameters are

$$\mathbf{S}_R^{(\text{amp})} = \mathbf{H}\mathbf{S} + \check{\mathbf{N}}^{(\text{amp})} \tag{11}$$

$$= \mathbf{S}' + \check{\mathbf{N}}^{(\text{amp})}, \tag{12}$$

where $\mathbf{S}' = [S_1', S_2', S_3']^T$ is the SV of the noiseless received electric field after transformation by $\mathbf{U}$ and $\check{\mathbf{N}}^{(\text{amp})} = [\check{N}_1^{(\text{amp})}, \check{N}_2^{(\text{amp})}, \check{N}_3^{(\text{amp})}]^T$ is the noise vector after passing both the signal and the noise through the 90° optical hybrid. Unitary matrices scaled by gain factor $\sqrt{\varsigma}$ transform to Muller matrices $\mathbf{H}$ with channel gain factor ς [20].

To derive an expression for $\check{\mathbf{N}}^{(\text{amp})}$, we examine how the additive noise beats with the received signal electric field. Let $[E_1', E_2']^T = [E_{1I}' + jE_{1Q}', E_{2I}' + jE_{2Q}']^T$. Expanding $\mathbf{S}_R^{(\text{amp})}$ as $\mathbf{S}_R^{(\text{amp})} = [S_{R,1}^{(\text{amp})}, S_{R,2}^{(\text{amp})}, S_{R,3}^{(\text{amp})}]^T$, we have

$$S_{R,1}^{(\text{amp})} = S_1' + \check{N}_1^{(\text{amp})}, \tag{13}$$

$$S_{R,2}^{(\text{amp})} = S_2' + \check{N}_2^{(\text{amp})}, \tag{14}$$

$$S_{R,3}^{(\text{amp})} = S_3' + \check{N}_3^{(\text{amp})}, \tag{15}$$

where

$$\check{N}_1^{(\text{amp})} = 2N_{1I}E_{1I}' + 2N_{1Q}E_{1Q}' - 2N_{2I}E_{2I}' - 2N_{2Q}E_{2I}' + N_{1I}^2 + N_{1Q}^2 - N_{2I}^2 - N_{2Q}^2, \tag{16}$$

$$\check{N}_2^{(\text{amp})} = 2N_{1I}E_{2I}' + 2N_{1Q}E_{2Q}' + 2N_{2I}E_{1I}' + 2N_{2Q}E_{1Q}' + 2N_{1I}N_{2I} + 2N_{1Q}N_{2Q}, \tag{17}$$

$$\check{N}_3^{(\text{amp})} = -2N_{1I}E_{2Q}' + 2N_{1Q}E_{2I}' + 2N_{2I}E_{1Q}' - 2N_{2Q}E_{1I}' + 2N_{1Q}N_{2I} - 2N_{1I}N_{2Q}, \tag{18}$$

$$S_1' = E_{1I}'^2 + E_{1Q}'^2 - E_{2I}'^2 - E_{2Q}'^2, \tag{19}$$

$$S_2' = 2E_{1I}'E_{2I}' + 2E_{1Q}'E_{2Q}', \tag{20}$$

$$S_3' = 2E_{1I}'E_{2Q}' + 2E_{1Q}'E_{2I}'. \tag{21}$$

The random variables $\check{N}_1^{(\text{amp})}$, $\check{N}_2^{(\text{amp})}$, and $\check{N}_3^{(\text{amp})}$ each contain terms consisting of the product of two Gaussian random variables, which results in a non-Gaussian noise distribution. In practical optically amplified IM/DD systems, the signal-spontaneous beat noise dominates and the spontaneous-spontaneous beat noise can often be ignored. Under this assumption, the noise vector $\check{\mathbf{N}}^{(\text{amp})}$ can be approximated by the noise vector $\tilde{\mathbf{N}}^{(\text{amp})} = [\tilde{N}_1^{(\text{amp})}, \tilde{N}_2^{(\text{amp})}, \tilde{N}_3^{(\text{amp})}]^T$, where

$$\tilde{N}_1^{(\text{amp})} = 2N_{1I}E_{1I}' + 2N_{1Q}E_{1Q}' - 2N_{2I}E_{2I}' - 2N_{2Q}E_{2I}', \tag{22}$$

$$\tilde{N}_2^{(\text{amp})} = 2N_{1I}E_{2I}' + 2N_{1Q}E_{2Q}' + 2N_{2I}E_{1I}' + 2N_{2Q}E_{1Q}', \tag{23}$$

$$\tilde{N}_3^{(\text{amp})} = -2N_{1I}E_{2Q}' + 2N_{1Q}E_{2I}' + 2N_{2I}E_{1Q}' - 2N_{2Q}E_{1I}'. \tag{24}$$

Computing the covariance matrix of $\tilde{\mathbf{N}}^{(\text{amp})}$, we get

$$\text{cov}\left\{\tilde{\mathbf{N}}^{(\text{amp})}\right\} = 4\sigma_{\text{amp}}^2\langle\|\mathbf{U}\mathbf{E}\|_2^2\rangle\mathbf{I}_3 = 4\sigma_{\text{amp}}^2\sqrt{\varsigma}\langle S_0\rangle\mathbf{I}_3, \tag{25}$$

where $S_0 = E_{1I}^2 + E_{1Q}^2 + E_{2I}^2 + E_{2Q}^2$ and $\mathbf{I}_3$ is the 3×3 identity matrix. Including the effects of spontaneous-spontaneous beat noise and after compensating for the rotation matrix, the demodulated vector $\mathbf{S}_D^{(\text{amp})} = [S_{D,1}^{(\text{amp})}, S_{D,2}^{(\text{amp})}, S_{D,3}^{(\text{amp})}]^T$ is given by

$$\mathbf{S}_D^{(\text{amp})} = \frac{1}{\varsigma}\mathbf{H}^H\mathbf{S}_R^{(\text{amp})} = \mathbf{S} + \frac{1}{\varsigma}\mathbf{H}^H\check{\mathbf{N}}^{(\text{amp})}. \tag{26}$$

In the high-SNR regime, $\check{\mathbf{N}}^{(\text{amp})} \approx \tilde{\mathbf{N}}^{(\text{amp})}$. The covariance matrix of the approximate noise vector $\frac{1}{\varsigma}\mathbf{H}^H\tilde{\mathbf{N}}^{(\text{amp})}$ is

$$\text{cov}\left\{\frac{1}{\varsigma}\mathbf{H}^H\tilde{\mathbf{N}}^{(\text{amp})}\right\} = 4\frac{\sigma_{\text{amp}}^2\langle S_0\rangle}{\sqrt{\varsigma}}\mathbf{I}_3. \tag{27}$$

III. Shaping With Continuous Unconstrained Constellations

In this section, we use the Forney method [8] to derive analytical approximations to the optimal continuous input distribution in the thermal noise-limited and amplifier noise-limited regimes. We begin with a brief overview of the Forney method. More thorough presentations can be found in [8] and [10]. We then apply the technique to the discrete-time SVR channel models derived in Section II-B in both noise regimes.

A. Forney Method Overview

The Forney method yields tight analytical approximations to the optimal continuous input distribution at high SNR [8]. The first step of the method is to choose the coordinate system in which to construct a high-dimensional lattice code. The signal points that constitute the lattice code are the N-fold Cartesian product of N symbols from consecutive, disjoint time intervals. The set of possible symbols in any one symbol interval is described in a set of natural coordinates [10]. This is a coordinate system in which the binary error probability between adjacent signal points is a decreasing function of the Euclidean distance between them, at least asymptotically at high SNR. Signal points are uniformly spaced in the natural coordinates to equalize the

binary error probabilities between pairs of nearest neighbors. Since the points are equally spaced, when we describe a sequence of N symbols by an N-fold Cartesian product of the natural coordinates, each point occupies an equal volume.

The natural coordinates are typically determined by the physical signal attribute being detected and by the noise statistics and their dependence on the signal. In the coherent AWGN channel, the natural coordinate for a symbol interval is the (complex-valued) field amplitude [8], since the noise is Gaussian and signal-independent. In the thermal noise-limited IM/DD channel, the natural coordinate for a symbol interval is the (non-negative) intensity [9], since the noise is Gaussian and signal-independent in the domain of photocurrent, which is proportional to the intensity. In the amplifier noise-limited IM/DD channel, the dominant noise at high SNR is signal-spontaneous beat noise, whose variance scales linearly with signal intensity. The binary error probabilities between neighboring signals are equalized if the intensities form a quadratic progression, i.e., if the field magnitudes form a linear progression. Hence, the natural coordinate for a symbol interval is the (non-negative) field magnitude [10].

After constructing an infinite lattice, the code is constrained to include only points lying within a bounding region chosen to minimize the average physical energy of the enclosed signal points. The physical energy associated with a signal point is a function of its coordinates. For example, in the thermal noise-limited IM/DD channel, the physical energy corresponds to the L1 norm of the intensity [9], while in the amplifier noise-limited IM/DD channel, it corresponds to the squared L2 norm of the field magnitude [10]. An optimal bounding region encloses a set of signal points with the property that the energy of each point within or on the bounding region is no greater than any point outside the bounding region.

The continuous approximation replaces a finite-dimensional lattice code enclosed by a bounding region with a uniform continuous distribution over the region enclosed by the same bounding region [8]. The number of points enclosed in the lattice code is approximated by the volume enclosed by the bounding region and determines the maximum achievable information rate. The average power of the lattice code can also be approximated by a function of the enclosed continuous region.

Under the continuous approximation, the total coding gain of the lattice code can be written as the product of a coding gain and a shape gain [8]. The shape gain is a measure of the improvement in energy efficiency obtained by constellation shaping and is defined as the ratio of the average power of a shaped constellation to a reference constellation under the constraint that the volumes occupied by the two constellations are equal. We compute the shape gain using the continuous approximation in the limit $N \to \infty$, where the lattice code describes an infinitely long symbol sequence.

B. Thermal Noise-Limited Regime

In the thermal noise-limited regime, the natural coordinates for a symbol interval are $[S_1, S_2, S_3]^T$, since the noise in this domain is signal-independent and additive, and the binary error probability for a maximum-likelihood (ML) receiver can be expressed as a function of the Euclidean distance:

$$P_e(\mathbf{S}_1, \mathbf{S}_2) = Q\left(\frac{\|\mathbf{S}_1 - \mathbf{S}_2\|_2^2}{2\sigma_{\text{th}}^2}\right), \tag{28}$$

where $\mathbf{S}_1$ and $\mathbf{S}_2$ are two arbitrary 3-D Stokes symbols.

Let $\mathbf{S}^{(N)}$ be a length-$3N$ vector of Stokes parameters formed by concatenating N time-disjoint 3-D Stokes symbols, i.e., $\mathbf{S}^{(N)} = [S_{1,1}, S_{1,2}, S_{1,3}, S_{2,1}, \ldots, S_{N,2}, S_{N,3}]$. In addition, let $C^{(\text{th})}$ be a $3N$-D lattice of signal points embedded in a $3N$-D vector space with Stokes parameters coordinates. The total optical power of $\mathbf{S}^{(N)}$ is $\sum_{i=1}^{N} S_{i,0}$. The optimal bounding region encloses all signal points with power at most $P_{\text{peak}}^{(\text{th})}$. The set of minimum-power signal points is the subset of signal points $\{\mathbf{S}^{(N)} \subseteq C^{(\text{th})}\}$ enclosed by the bounding region

$$\mathcal{R}_{\text{shaped}}^{(\text{th})}\left(N, P_{\text{peak}}^{(\text{th})}\right) = \left\{\mathbf{S}^{(N)} \,\middle|\, \sum_{i=1}^{N} S_{i,0} = P_{\text{peak}}^{(\text{th})}\right\}. \tag{29}$$

For a fixed N and $P_{\text{peak}}^{(\text{th})}$, the set of signal points enclosed by $\mathcal{R}_{\text{shaped}}^{(\text{th})}(N, P_{\text{peak}}^{(\text{th})})$ can be partitioned by conditioning on $S_{N,0} = s_0$. The volume of the $3N$-D space bounded by $\mathcal{R}_{\text{shaped}}^{(\text{th})}(N, P_{\text{peak}}^{(\text{th})})$ can therefore be recursively computed by

$$\begin{aligned} &V_{\text{shaped}}^{(\text{th})}\left(N, P_{\text{peak}}^{(\text{th})}\right) \\ &\quad = \int_0^{P_{\text{peak}}^{(\text{th})}} V_{\text{shaped}}^{(\text{th})}\left(N-1, P_{\text{peak}}^{(\text{th})} - s_0\right) 4\pi s_0^2 ds_0. \end{aligned} \tag{30}$$

The closed-form solution for (30) is given by

$$\begin{aligned} V_{\text{shaped}}^{(\text{th})}\left(N, P_{\text{peak}}^{(\text{th})}\right) &= \left(P_{\text{peak}}^{(\text{th})}\right)^{3N} (4\pi)^N \prod_{i=1}^{N} B(3, 3i-2) \\ &= \left(P_{\text{peak}}^{(\text{th})}\right)^{3N} \frac{(8\pi)^N}{\Gamma(3N+1)}, \end{aligned} \tag{31}$$

where B is the beta function (also sometimes called the Euler integral of the first kind).

The induced probability density function (PDF) in $f_{\mathbf{S}_{1,0}^{(N)}}(s_0)$ is given by $f_{\mathbf{S}_{1,0}^{(N)}}(s_0) = \frac{V_{\text{shaped}}^{(\text{th})}(N-1, P-s_0) 4\pi s_0^2}{V_{\text{shaped}}^{(\text{th})}(N,P)}$. By symmetry, $f_{\mathbf{S}_{1,0}^{(N)}}(s_0) = f_{\mathbf{S}_{2,0}^{(N)}}(s_0) = \cdots = f_{\mathbf{S}_{N,0}^{(N)}}(s_0)$. The average power per basic 3-D SV symbol is calculated to be

$$\begin{aligned} P_{\text{shaped}}^{(\text{th})} &= \frac{\int_0^{P_{\text{peak}}^{(\text{th})}} V_{\text{shaped}}^{(\text{th})}\left(N-1, P_{\text{peak}}^{(\text{th})} - s_0\right) 4\pi s_0^3 \, ds_0}{V_{\text{shaped}}^{(\text{th})}\left(N, P_{\text{peak}}^{(\text{th})}\right)} \\ &= \frac{3}{3N+1} P_{\text{peak}}^{(\text{th})}. \end{aligned} \tag{32}$$

We now proceed to compute the induced marginal distribution of each basic 3-D SV symbol. Let $f_{\mathbf{S}_i^{(N)}}(s_1, s_2, s_3)$, $i = 1, \ldots, N$ be the induced marginal distribution of $\mathbf{S}_i^{(N)}$, where $\mathbf{S}_i^{(N)} = [S_{i,1}, S_{i,2}, S_{i,3}]$. Due to the symmetry of each 3-D Stokes symbol in $\mathbf{S}^{(N)}$, $f_{\mathbf{S}_1^{(N)}}(s_1, s_2, s_3) = f_{\mathbf{S}_2^{(N)}}(s_1, s_2, s_3) =$

$\cdots = f_{\mathbf{S}_N^{(N)}}(s_1, s_2, s_3)$. To simply the notation, the following analysis focuses on $f_{\mathbf{S}_1^{(N)}}(s_1, s_2, s_3)$, but the results apply equally to all 3-D symbols, $\mathbf{S}_i^{(N)} \in \mathbf{S}^{(N)}, i = 1, \ldots, N$.

Conditioned on a fixed optical power for a basic 3-D SV symbol, the probability density assigned to 3-D symbols with the same optical power must be equal. This can be seen from the recursive construction in (30). We therefore have $f_{\mathbf{S}_1^{(N)}|\mathbf{S}_{1,0}^{(N)}}(s_1, s_2, s_3|s_0) = f_{\mathbf{S}_1^{(N)}|\mathbf{S}_{1,0}^{(N)}}(s_1', s_2', s_3'|s_0) = 1/4\pi s_0^2$. Expanding $f_{\mathbf{S}_1^{(N)}}(s_1, s_2, s_3)$, we get

$$f_{\mathbf{S}_1^{(N)}}(s_1, s_2, s_3) = f_{\mathbf{S}_1^{(N)}|\mathbf{S}_{1,0}^{(N)}}(s_1, s_2, s_3|s_0) f_{\mathbf{S}_{1,0}^{(N)}}(s_0) = \frac{V_{\text{shaped}}^{(\text{th})}\left(N-1, P_{\text{peak}}^{(\text{th})} - s_0\right)}{V_{\text{shaped}}^{(\text{th})}\left(N, P_{\text{peak}}^{(\text{th})}\right)}. \quad (33)$$

Multiplying (33) by the normalization constant $V_{\text{shaped}}^{(\text{th})}(N, P_{\text{peak}}^{(\text{th})})/V_{\text{shaped}}^{(\text{th})}(N-1, P_{\text{peak}}^{(\text{th})})$ and substituting in (32), we get

$$f_{\mathbf{S}_1^{(N)}}(s_1, s_2, s_3) \propto \frac{V_{\text{shaped}}^{(\text{th})}\left(N-1, P_{\text{peak}}^{(\text{th})} - s_0\right)}{V_{\text{shaped}}^{(\text{th})}\left(N-1, P_{\text{peak}}^{(\text{th})}\right)} = \left(1 - \frac{3s_0}{P_{\text{shaped}}^{(\text{th})} \times (3N+1)}\right)^{3N-3}. \quad (34)$$

As $N \to \infty$, the induced probability $f_{\mathbf{S}_1^{(N)}}(s_1, s_2, s_3) \propto \exp(-3s_0/P_{\text{shaped}}^{(\text{th})})$. The induced marginal distribution is therefore an exponential distribution in S_0, and is given by

$$f_{\mathbf{S}_1^{(N)}}(s_1, s_2, s_3) \propto \exp\left(-\frac{s_0}{\beta}\right). \quad (35)$$

The shape parameter β is a constant that modifies the PDF by dividing the argument of the exponent, enabling the generation of a family of PDFs that converge to the uniform distribution as β increases.

1) Shape Gain: In the absence of shaping, the distribution of the constituent 3-D constellation is a discrete uniform distribution. Under the continuous approximation, this corresponds to a continuous uniform distribution in Stokes spaces on each dimension of the transmitted vector, i.e., $S_{i,j} \sim U(-A, A), i = 1, \ldots, N, j = 1, 2, 3$. The reference bounding region in the thermal noise-limited case is given by

$$\mathcal{R}_{\text{ref}}^{(\text{th})}(N, A) = \left\{ \mathbf{S}^{(N)} \Big| \max_{\substack{i \in \{i, \ldots, N\} \\ j \in \{1,2,3\}}} |S_{i,j}| = A \right\}, \quad (36)$$

where A is the largest absolute value taken by any of the Stokes parameters. The volume enclosed by $\mathcal{R}_{\text{ref}}^{(\text{th})}(N, A)$ is

$$V_{\text{ref}}^{(\text{th})}(N, A) = (2A)^{3N}. \quad (37)$$

Equating (31) and (37) and using Stirling's approximation $(\Gamma(3N+1))^{1/3N} \sim (6\pi N)^{1/6N}(\frac{3N}{e})$ yields

$$P_{\text{peak}}^{(\text{th})} = 2A \frac{(6\pi N)^{1/6N}\left(\frac{3N}{e}\right)}{\sqrt[3]{8\pi}}. \quad (38)$$

The optical power per 3-D SV symbol of a length-$3N$ vector of Stokes parameters is given by

$$P_{\text{ref}}^{(\text{th})} = \mathbb{E}\left[\sqrt{S_{1,1}^2 + S_{1,2}^2 + S_{1,3}^2}\right]. \quad (39)$$

We were unable to find a simple, closed-form expression for (39) [21], [22]. Using Jensen's inequality to compute an upper bound yields

$$P_{\text{ref}}^{(\text{th})} \le \sqrt{\mathbb{E}\left[S_{1,1}^2 + S_{1,2}^2 + S_{1,3}^2\right]} = A. \quad (40)$$

In addition to this upper bound, we estimated $P_{\text{ref}}^{(\text{th})}$ by numerically approximating the distribution of $S_{1,1}^2 + S_{1,2}^2 + S_{1,3}^2$ and computing the expectation and by performing a Monte Carlo simulation. The results of both methods coincided with $P_{\text{ref}}^{(\text{th})} \approx 0.961A$. Using (32) and (38), the ratio of $P_{\text{ref}}^{(\text{th})}$ to $P_{\text{shaped}}^{(\text{th})}$ as $N \to \infty$ yields

$$\lim_{N\to\infty} \frac{P_{\text{ref}}^{(\text{th})}}{P_{\text{shaped}}^{(\text{th})}} \approx 0.961 \frac{\sqrt[3]{\pi}e}{3} = 1.056\,\text{dB} \quad (41)$$

$$\le \frac{\sqrt[3]{\pi}e}{3} = 1.23\,\text{dB}. \quad (42)$$

(41) results from the numerical approximation of $P_{\text{ref}}^{(\text{th})}$ while (42) uses (40).

C. Amplifier Noise-Limited Regime

We begin by defining a nonlinear transformation from 3-D SV space to a 3-D scaled SV space. Our analysis assumes operation at high SNR and that the Gaussian noise approximation holds, i.e., $\tilde{\mathbf{N}}^{(\text{amp})} \approx \tilde{\mathbf{N}}^{(\text{amp})}$. We show that the 3-D scaled SV space is the natural coordinate system in the amplifier noise-limited regime. We then find the induced distribution in 3-D scaled SV space, derive the induced distribution in the original SV space, and compute the ultimate shape gain.

1) Natural Coordinates: We define $\mathbf{X} = [X_1, X_2, X_3]$ to be a 3-D scaled Stokes symbol, where X_1, X_2, X_3 are the corresponding scaled Stokes parameters. $\mathbf{X}$ is defined in terms of $\mathbf{S}$ by the following transformation:

$$[X_1, X_2, X_3]^T = \frac{[S_1, S_2, S_3]^T}{\sqrt{S_0}} = \frac{[S_1, S_2, S_3]^T}{\sqrt[4]{S_1^2 + S_2^2 + S_3^2}}. \quad (43)$$

Consider two arbitrary 3-D scaled Stokes symbols $\mathbf{X}_1$ and $\mathbf{X}_2$ that correspond to the 3-D Stokes symbols $\mathbf{S}_1$ and $\mathbf{S}_2$. Using a correlation receiver with the equal-crossover error-probability decision threshold

$$\frac{\|\mathbf{S}_2\|_2^{1/2}\left(\mathbf{S}_2^T\mathbf{S}_1 - \|\mathbf{S}_1\|_2^2\right) + \|\mathbf{S}_1\|_2^{1/2}\left(\|\mathbf{S}_2\|_2^2 - \mathbf{S}_1^T\mathbf{S}_2\right)}{\|\mathbf{S}_1\|_2^{1/2} + \|\mathbf{S}_2\|_2^{1/2}}, \quad (44)$$

the binary error probability is given by

$$P_e\left(\mathbf{S}_1,\mathbf{S}_2\right) = Q\left(\frac{\|\mathbf{S}_1-\mathbf{S}_2\|_2}{2\sigma_{\text{amp}}(\sqrt{\mathbf{S}_{1,0}}+\sqrt{\mathbf{S}_{2,0}})}\right). \tag{45}$$

We further assume that the signal points are sufficiently far from the origin that the square roots of the transmitted optical powers of $\mathbf{S}_1$ and $\mathbf{S}_2$ are approximately equal, i.e., $\sqrt{\mathbf{S}_{1,0}} \approx \sqrt{\mathbf{S}_{2,0}}$. Using this approximation, we get

$$P_e\left(\mathbf{S}_1,\mathbf{S}_2\right) \approx Q\left(\frac{1}{4\sigma_{\text{amp}}}\|\mathbf{X}_1-\mathbf{X}_2\|_2\right). \tag{46}$$

The binary error probability (46) depends on $\mathbf{S}_1$ and $\mathbf{S}_2$ in terms of the Euclidean distance between the corresponding scaled symbols $\mathbf{X}_1$ and $\mathbf{X}_2$. We conclude that $[X_1, X_2, X_3]^T$ are the natural coordinates to employ in the Forney method.

2) Induced 3-D Distribution: Let $\mathbf{X}^{(N)}$ be a length-$3N$ vector of scaled Stokes parameters formed by concatenating N time-disjoint 3-D scaled SV symbols, i.e., $\mathbf{X}^{(N)} = [\mathbf{X}_1^{(N)}, \mathbf{X}_2^{(N)}, \ldots, \mathbf{X}_N^{(N)}] = [X_{1,1}, X_{1,2}, X_{1,3}, X_{2,1}, \ldots, X_{N,2}, X_{N,3}]$. The total optical power of $\mathbf{X}^{(N)}$ is $\sum_{i=1}^{N} X_{i,0}^2$, where $X_{i,0} = \sqrt{X_{i,1}^2 + X_{i,2}^2 + X_{i,3}^2}$.

Let $C^{(\text{amp})}$ be a $3N$-D lattice of signal points embedded in a $3N$-D vector space with scaled Stokes parameters coordinates. The set of length-$3N$ vectors of scaled Stokes parameters with peak optical power at most $P_{\text{peak}}^{(\text{amp})}$ is the subset of signal points $\{\mathbf{X} \subseteq C^{(\text{amp})}\}$ enclosed by the bounding region

$$\mathcal{R}_{\text{shaped}}^{(\text{amp})}\left(N, P_{\text{peak}}^{(\text{amp})}\right) = \left\{\mathbf{X}^{(N)} \middle| \left\|\mathbf{X}^{(N)}\right\|_2^2 = P_{\text{peak}}^{(\text{amp})}\right\}. \tag{47}$$

The spherical bounding region described in (47) is analogous to the bounding region described in [10], except for the nonnegative orthant constraint. Slightly adapting the procedure described in [10], one can show that the induced marginal distribution on $\mathbf{X}_1^{(N)}$ is a Gaussian distribution as $N \to \infty$.

The transformation from Stokes space to the natural coordinates defined in (43) has an inverse transformation

$$[S_1, S_2, S_3]^T = \sqrt{X_1^2 + X_2^2 + X_3^2}\,[X_1, X_2, X_3]^T. \tag{48}$$

Using (48), the induced marginal PDF in Stokes space can be computed as

$$f_{\mathbf{S}_1^{(N)}}(s_1, s_2, s_3) = f_{X_1,X_2,X_3}\left(\frac{1}{s_0}[s_1, s_2, s_3]^T\right)|\mathbf{J}|, \tag{49}$$

where $|\mathbf{J}|$ is the Jacobian determinant of (43). Using the matrix determinant lemma, it can be shown that

$$|\mathbf{J}| = \frac{1}{s_0^{3/2}}\left|\mathbf{I}_3 - \frac{1}{2s_0^2}\mathbf{S}^{(N)}\left(\mathbf{S}^{(N)}\right)^T\right| = \frac{1}{2s_0^{3/2}}. \tag{50}$$

The induced distribution on the basic 3-D SV symbol is given by

$$f_{\mathbf{S}_1^{(N)}}(s_1, s_2, s_3) \propto \frac{\exp(-s_0/\beta)}{{s_0}^{3/2}}. \tag{51}$$

We note that the previous analysis uses the approximations given in (25) and (46), which are valid in the high-SNR regime. In this regime, the exponential factor in (51) dominates. We therefore conclude that the induced distribution is closely approximated by an exponential distribution of S_0.

3) Shape Gain: Without shaping, each Stokes symbol is chosen uniformly and independently along each dimension in the natural coordinates over the interval $[-A, A]$. The reference bounding region in the amplifier noise-limited case is given by

$$\mathcal{R}_{\text{ref}}^{(\text{th})}(N, A) = \left\{\mathbf{X}^{(N)} \middle| \max_{\substack{i\in\{i,\ldots,N\}\\ j\in\{1,2,3\}}} |X_{i,j}| = A\right\}. \tag{52}$$

The volume enclosed by the reference constellation bounded by (52) is given by

$$V_{\text{ref}}^{(\text{amp})}(N, A) = (2A)^{3N}. \tag{53}$$

The volume enclosed by the shaped constellation bounded by (47) is given by [10]

$$V_{\text{shaped}}^{(\text{amp})}\left(N, P_{\text{peak}}^{(\text{amp})}\right) = \frac{\pi^{3N/2}}{\Gamma(3N/2+1)}\left(P_{\text{peak}}^{(\text{amp})}\right)^{3N/2}. \tag{54}$$

Equating (53) and (54) and using Stirling's approximation $(\Gamma(3N/2+1))^{2/3N} \sim (3\pi N)^{1/3N}\frac{3N}{2e}$, we get

$$P_{\text{peak}}^{(\text{amp})} = \frac{6N}{\pi e}A^2(3\pi N)^{1/3N}. \tag{55}$$

The average power per basic 3-D symbol of a $3N$-D vector drawn uniformly from the signal points enclosed by $\mathcal{R}_{\text{shaped}}^{(\text{amp})}(N, P_{\text{peak}}^{(\text{amp})})$ is given by

$$P_{\text{shaped}}^{(\text{amp})} = \frac{3}{3N+2}P_{\text{peak}}^{(\text{amp})}. \tag{56}$$

The average power per basic 3-D symbol of a $3N$-D vector drawn uniformly from the signal points enclosed by $\mathcal{R}_{\text{ref}}^{(\text{amp})}(N, A)$ is $P_{\text{ref}}^{(\text{amp})} = A^2$. Taking the ratio of $P_{\text{ref}}^{(\text{amp})}$ to $P_{\text{shaped}}^{(\text{amp})}$, substituting in (55), and taking the limit, we get

$$\lim_{N\to\infty}\frac{P_{\text{ref}}^{(\text{amp})}}{P_{\text{shaped}}^{(\text{amp})}} = \frac{\pi e}{6} \approx 1.533\,\text{dB}, \tag{57}$$

which coincides with the ultimate shape gain in [10].

IV. Shaping With Discrete Finite Constellations

The Forney method analysis in Section III presumes the use of high-density discrete constellations that can be approximated well by continuous distributions. In practice, however, many optical links are subject to strong complexity constraints and use constellations with a limited number of points. In this section, we conduct numerical studies of PS on finite input distributions. We begin by explaining the fundamentals of the BA method. We then distinguish 3-D PS from independent 1-D PS and present several different shaping schemes. We conclude the section with a presentation of the shape gain achieved under various input distribution and encoding schemes.

A. Numerical Computation of the Optimal Input Distribution

The BA method is a numerical algorithm to compute an input distribution that maximizes the mutual information (MI) for arbitrary discrete memoryless channels [23], [24]. The BA method can be extended to include an average power constraint and subsequent references to the BA method presume an average power constraint.

The BA method requires a discrete-time channel model. Equations (6) and (26) are used as the models for the thermal noise-limited and amplifier noise-limited cases, respectively. Under the use of a finite input constellation, each of the two channels is described by a conditional PDF. An $(8 \times 8 \times 8)$ input constellation is adopted for all numerical simulations using 3-D modulation, resulting in a maximum MI of 9 bits per 3-D SV symbol.

We quantize the continuous-valued output of the channel using a $(24 \times 24 \times 24)$ grid. The number of intervals in the grid is obtained using the optimization procedure described in [25] to reduce estimation error. With a fixed input and output constellation, the corresponding conditional probability mass function (PMF) is estimated by performing a Monte Carlo simulation with 2^{18} 3-D SV symbols.

The coded modulation capacity under a uniformly spaced constellation is defined as the maximum MI over (a) the set of all input PMFs and (b) the distance between input constellation points subject to a fixed average power constraint [12]. We assume the signal points of the input constellation are uniformly spaced when described in terms of the natural coordinates in thermal noise-limited and amplifier noise-limited cases. The set of possible input constellations is parameterized by a scale factor λ, which scales a reference constellation to set the spacing between the constellation points. Let $\mathcal{S}^{(m)} = \{-3.5, -2.5, -1.5, -0.5, 0.5, 1.5, 2.5, 3.5\}^{\bigotimes m}$ be the reference input constellation, where $\bigotimes m$ denotes the m-fold Cartesian product. The set of possible input constellations is given by $\{\lambda \mathcal{S}^{(m)}, \lambda \in \mathbb{R}^+\}$.

The power-limited BA method computes the input distribution that maximizes the MI for a fixed λ, and is subject to an average power constraint. The coded modulation capacity can be expressed as the joint optimization over the input PMF $P_S(s_1, s_2, s_3)$ and scale factor λ. The optimized input PMF $P_S^*(s_1, s_2, s_3)$ and optimized scale factor λ^* that achieve the coded modulation capacity are given by

$$P_{\mathbf{S}}^*(s_1, s_2, s_3), \lambda^* = \arg\max_{P_{\mathbf{S}}, \lambda} I(\mathbf{S}, \mathbf{S}_D)$$
$$\text{subject to} \quad \lambda \langle [\|\mathbf{S}\|_2] \rangle \leq P. \tag{58}$$

Here, P is the average power per 3-D SV symbol, I denotes the MI, and $\mathbf{S}_D = \mathbf{S}_D^{(\text{th})}$ and $\mathbf{S}_D = \mathbf{S}_D^{(\text{amp})}$ in the thermal noise-limited and amplifier noise-limited cases, respectively. All MI values using the BA method presented in this paper were obtained by a joint optimization of the input distribution and scale factor.

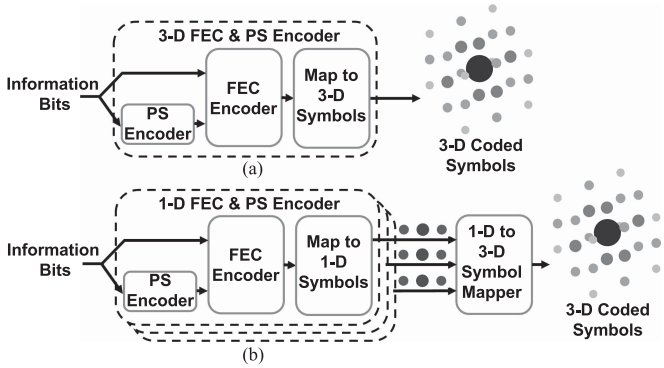

Fig. 3. Joint PS/FEC encoders architectures for (a) 3-D encoding and (b) independent 1-D encoding, respectively. PS: probabilistic shaping, FEC: forward error correction.

B. Sampling Continuous Distributions to Obtain Optimized Discrete Distribution

While the BA method numerically computes the optimal input distribution, for several reasons it is often desirable to generate the input distributions using an analytic expression. The relative simplicity of an analytic expression facilitates a performance comparison among several analytic distributions and against the distribution found using the BA method. Some of these analytic distributions can be derived from the continuous distributions found in Section III, providing a complementary analysis of the distributions.

The Gaussian and exponential distributions can be expressed in terms of the scale parameter λ and shape parameter β. A fixed constellation of points is required to transform a 3-D PDF in SV space $f_{\mathbf{S}}(s_1, s_2, s_3; \beta)$ to a discrete PMF $P_{\mathbf{S}}[s_1, s_2, s_3]$. The set of possible input constellations in 3-D SV space can be parameterized by λ and reference constellation $\mathcal{S}^{(3)}$. For a fixed λ, β controls the probability density assigned to each point in the constellation. For fixed λ and β, the induced PMF is given by

$$P_{\mathbf{S}}[s_1, s_2, s_3] = \frac{f_{\mathbf{S}}(s_1, s_2, s_3; \beta)}{\sum_{(\dot{s}_1, \dot{s}_2, \dot{s}_3) \in \lambda \mathcal{S}^{(3)}} f_{\mathbf{S}}(\dot{s}_1, \dot{s}_2, \dot{s}_3; \beta)}. \tag{59}$$

The scale and shape parameters that optimize the MI are $\{\lambda^*, \beta^*\}$, which are given by

$$\{\lambda^*, \beta^*\} = \arg\max_{\lambda, \beta} I(\mathbf{S}, \mathbf{S}_D). \tag{60}$$

$\mathbf{S}_D = \mathbf{S}_D^{(\text{th})}$ and $\mathbf{S}_D = \mathbf{S}_D^{(\text{amp})}$ in the thermal noise-limited and amplifier noise-limited cases, respectively.

All MI values using an input distribution derived from an analytical PDF are obtained by joint optimization of λ and β at each SNR. The MI of the continuous-valued output of the channel is estimated using a $(24 \times 24 \times 24)$ grid of uniformly spaced intervals whenever 3-D modulation is used.

C. 3-D vs. 1-D Probabilistic Shaping Encoders

Two different PAS-based architectures for FEC/PS encoders are shown in Fig. 3. Each FEC/PS module consists of a PS

encoder, an FEC encoder, and a symbol mapper. The PS encoder maps a subset of the information bits to a sequence of bits that index output symbols with a desired target probability distribution. The output of the PS encoder and the subset of information bits not used by the PS encoder are input to an FEC encoder. The information bits input to the FEC encoder and the parity bits generated by the FEC encoder encode the sign of the output symbols.

In Fig. 3(a), the output of the FEC/PS module is a sequence of 3-D symbols. By contrast, the output of each of the three FEC encoders in Fig. 3(b) is mapped to an independent sequence of 1-D symbols. An additional symbol mapper is required to transform the three independent sequences of 1-D symbols into a single 3-D symbol sequence. The 1-D to 3-D symbol mapper performs this transformation by mapping each of the three input sequences to a different SV parameter.

The encoder shown in Fig. 3(a) is assumed capable of generating a sequence of SV symbols with an arbitrary 3-D PMF. The encoder in Fig. 3(b) generates a sequence of SV symbols whose PMF can be described as the product of 3 independent 1-D distributions. The structure of Fig. 3(b) may be preferred in some cases due to its lower complexity.

The constraint on the output distribution in Fig. 3(b) implies that the independent architecture can only generate a subset of the distributions that can be generated using the architecture in Fig. 3(a). This implies the achievable MI using Fig. 3(a) is lower bounded by the MI achievable using Fig. 3(b). The relative performance degradation resulting from independent shaping of the SV parameters depends on the design of the input distribution.

We now list the 1-D and 3-D distributions used in our numerical analyses. We use italicized font to provide unique labels for each of the probability distributions that appear in the Figs. of Sections IV and V.

1) Sampling the Gaussian Distribution: The joint PDF of three uncorrelated Gaussian random variables can be written as a product of the marginal PDFs of the three independent Gaussian random variables. This property is preserved under the sampling process described in (59) and the reference constellation $\mathcal{S}^{(3)}$ is symmetric with respect to the three SV parameters. The 1-D and 3-D Gaussian input PMFs are equivalent and are labeled using the term *Gaussian*.

2) Sampling the Exponential Distribution: The 3-D continuous distribution described in (35) or (51) cannot be written as a product of three independent 1-D continuous exponential distributions. This property extends to the 3-D PMF induced via the sampling process in (59), as the non-separability of the continuous function is not affected by the discrete nature of the sampled points or the normalization factor.

The induced PDF given in (51) is proportional to an exponential distribution in S_0 multiplied by a factor of $1/S_0^{3/2}$. Because the Forney method assumes a high SNR, the exponential factor in (51) is dominant, and we drop the factor $1/S_0^{3/2}$ in the amplifier noise-limited case. Dropping the factor $1/S_0^{3/2}$ in (51) results in the same 3-D PDF for both the thermal noise-limited and amplifier noise-limited cases. We use the term *3-D exponential* to identify an input PMF that is obtained by sampling (35).

We can alternatively design an input distribution by sampling the product exponential PDF. The term *1-D exponential* identifies a joint PMF that is designed by sampling the 3-D product exponential PDF.

3) Blahut-Arimoto Method: As previously explained in Section IV-A, the BA method can numerically compute the capacity-achieving distribution for an average power-limited discrete memoryless channel. The term *3-D optimal* is used to refer to a 3-D PMF of the transmitted SV designed using the BA method where a nonzero probability can be assigned to each of the values in the input constellation.

For any 3-D optimal distribution, we can compute the marginal PMF in S_1. We can marginalize over S_1 without loss of generality because the marginal distribution for any of the three SV parameters is the same. Using this marginal distribution, we can generate a 3-D product PMF by assuming S_2 and S_3 have the same distribution as S_1 and computing the product of the three 1-D PMFs. We use the term *marginal of 3-D optimal* to refer to any 3-D product PMF where the distributions of the SV parameters are equal and computed by marginalizing a 3-D optimal PMF.

We can directly compute a 1-D PMF using the BA method. During the design of the 1-D PMF, we set $S_2 = S_3 = 0$ and optimize the 1-D PMF of S_1. The average power constraint for the 1-D PMF is set to $1/3$ of the target average power constraint of the 3-D PMF. The optimized 1-D PMF of S_1 is then transformed to a 3-D product PMF by choosing the 1-D PMFs of S_2 and S_3 to be identical to the optimized 1-D PMF of S_1 and then computing the product of the three PMFs. The term *1-D Optimal* refers to a 3-D product PMF of the transmitted SV designed by using the BA method on a single SV component.

D. Reference Systems

We analyze two additional systems to compare the performance of the aforementioned shaped distributions. In a system not using PS, all signal points in the 1-D constituent constellation are transmitted with equal probability. The term *uniform* denotes an SVR-based system that uses a uniform input distribution.

We also compare the performance of the SVR under various shaping distributions against a reference communication system employing shaping and standard coherent detection (SCD) [5]. The term *4-D Coherent* denotes the MI achieved by a system employing SCD and PS using the optimal input distribution. We assume the standard coherent system is LO-spontaneous beat noise-limited and uses an optical phase-locked loop, homodyne detection, active polarization control, and two 90° optical hybrids for simultaneous detection of the in-phase and quadrature components of the two polarizations. The received electric field vector $\mathbf{E}_R^{(\mathrm{co})}$ is given by

$$\mathbf{E}_R^{(\mathrm{co})} = \sqrt{\varsigma}\mathbf{E} + \mathbf{U}^H\mathbf{N}^{(\mathrm{co})}, \tag{61}$$

where $\mathbf{N}^{(\mathrm{co})} = [N_{1I}^{(\mathrm{co})}, N_{1Q}^{(\mathrm{co})}, N_{2I}^{(\mathrm{co})}, N_{2Q}^{(\mathrm{co})}]^T$ is a Gaussian random vector with covariance matrix

$$\mathrm{cov}\left\{\mathbf{N}^{(\mathrm{co})}\right\} = \sigma_{\mathrm{co}}^2\mathbf{I}_4. \tag{62}$$

Similar to Section II-B, we analyze the dominant noise source in the continuous-time model to derive the noise covariance matrix in the discrete-time model. The analysis of the LO-spontaneous beat noise is identical to the analysis of the signal-spontaneous beat noise in Section II-B under the monochromatic signal and noise approximation [19]. The power of the LO-spontaneous beat noise after an ideal electrical low-pass filter with bandwidth $B/2$ is therefore given by $4R_d^2 P_{\mathrm{LO}} S_{\mathrm{ASE}} \frac{B}{2}$, where P_{LO} is the LO power. To equate the noise powers of the discrete-time and continuous-time models at the output of the photodetector, we have $\sigma_{\mathrm{co}}^2 = S_{\mathrm{ASE}} \frac{B}{4}$. Because expressions for σ_{co}^2 and σ_{amp}^2 are identical, the standard coherent reference system uses the same definition of SNR as the amplifier noise-limited SVR, which is given by $\frac{\sqrt{\zeta}\langle S_0 \rangle}{2\sigma_{\mathrm{co}}^2}$.

The BA method is used to find the optimal input distribution for the standard coherent reference system. An input constellation of $(8 \times 8 \times 8 \times 8)$ is used, providing a maximum MI of 12 bits per symbol.

E. Hard-Decision Decoding

Hard-decision (HD) FEC is widely used in optical links owing to its low complexity as compared to soft-decision (SD) FEC. It is therefore of interest to analyze the MI achievable using HD decoding at the receiver.

The design procedures described in Sections IV-A or IV-B can be straightforwardly adapted to accomodate HD decoders by mapping the continuous-valued output of the channel to a discrete set of ML decision thresholds. The conditional noise distribution is recomputed at each SNR analyzed. For a fixed value of σ_{th} or σ_{amp}, the variance of the noise conditioned on the transmission of a specific signal point is numerically estimated using a Monte Carlo simulation. A separate estimate is performed for each signal point in the transmitted constellation. The noise distribution is then approximated as a Gaussian random vector using the conditional variances from the Monte Carlo simulations. In our simulations, we employ quasi-ML decoding by performing ML decoding only using the transmitted constellation point and its nearest neighbors (at most eight other points).

Under quasi-ML decoding, the discrete-input continuous-output channel is transformed into a discrete-input discrete-output channel. The MI can then be computed without using histogram-based techniques. Thus, the optimization procedures described in Sections IV-A and IV-B can be adopted to design input distributions for HD-based systems by replacing the continuous-valued channel output with the output of the quasi-ML decoder. The phrase *HD* is added to each of the terms introduced in Section IV-C to refer to input distributions that were designed for systems using an HD decoder.

F. 3-D Shaping Under Soft-Decision Decoding

In this section, we study the performance of SVR systems that perform joint shaping of the three SV parameters in each SV symbol, comparing them to SVR systems using uniform distributions and to SCD systems using optimized shaping. Fig. 4 shows MI vs. SNR in (a) the thermal noise-limited regime and (b) the amplifier noise-limited regime. Numerical results for the 3-D optimal distribution, the 3-D exponential distribution, the Gaussian distribution, the uniform distribution, and the standard coherent system are provided.

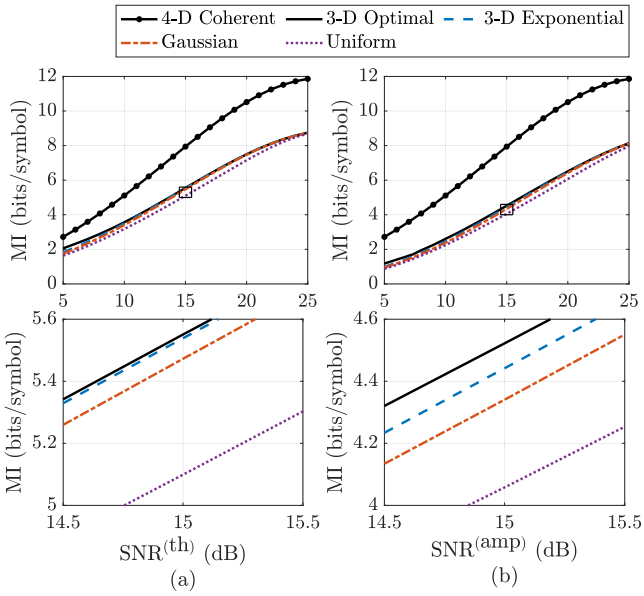

Fig. 4. Mutual information (MI) vs. SNR for several 3-D input distributions for SVR using soft-decision decoding in the (a) thermal noise-limited regime and (b) amplifier noise-limited regime. Plots in the bottom row are restricted to the rectangular regions indicated in the top row. The MI vs. SNR of the reference coherent receiver using standard coherent detection and soft-decision decoding is also included.

Fig. 4(a) and (b) clearly show that the Gaussian distribution is not the optimal sampled input distribution. The exponential distribution provides higher MI than the Gaussian distribution over the entire SNR range studied. This is consistent with the analysis in Section III.

The exponential distribution is within 0.01 bits/symbol of the optimal distribution in the thermal noise-limited regime. By contrast, the gap between the exponential distribution and the optimal distribution is larger in the amplifier noise-limited regime. We attribute this gap, in part, to the approximation in (46) and to the dropped multiplicative factor in (51).

The shape gains of the 3-D optimal distribution at MI values of 5.5 bits/symbol and 4.4 bits/symbol are 1.105 dB and 1.178 dB in the thermal noise-limited and amplifier noise-limited regimes, respectively. The relative magnitudes of the shape gain are consistent with the analytical values (41) and (57). There is a substantial MI gap between the SVR using the optimal input distribution and the SCD system, due to the additional degrees of freedom of SCD and different noise statistics. We further compare SVR and SCD systems in Section V-B.

G. 1-D Shaping Under Soft-Decision Decoding

In this section, we analyze the performance of PS in SVR-based systems that perform independent shaping of the three SV parameters. These 1-D distributions can be generated by the independent encoder structure shown in Fig. 3(b). Fig. 5(a) and (b) show the MI vs. SNR for four different 1-D distributions

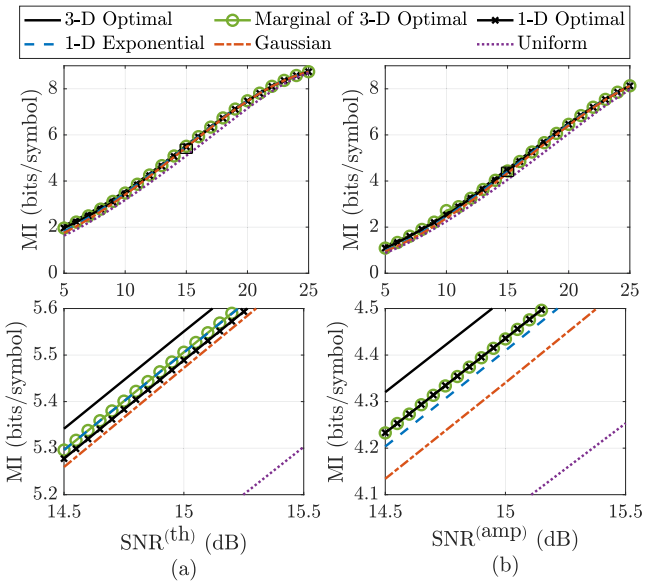

Fig. 5. Mutual information (MI) vs. SNR for the 3-D optimal input distribution and several 1-D input distributions for SVR using soft-decision decoding in the (a) thermal noise-limited regime and (b) amplifier noise-limited regime. Plots in the bottom row are restricted to the rectangular regions indicated in the top row.

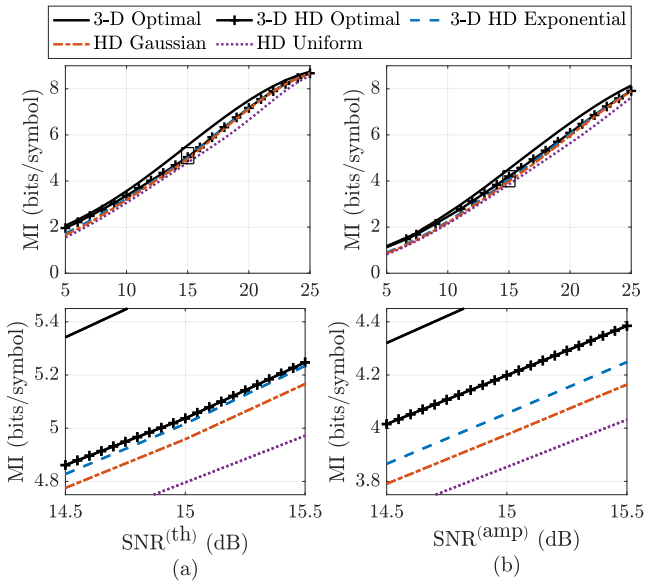

Fig. 6. Mutual information (MI) vs. SNR for several 3-D input distributions for SVR using hard-decision decoding in the (a) thermal noise-limited regime and (b) amplifier noise-limited regime. Plots in the bottom row are restricted to the rectangular regions indicated in the top row.

in the thermal noise-limited and amplifier noise-limited cases, respectively. The 3-D optimal input distribution is also included for comparison.

The 1-D exponential distribution outperforms the Gaussian distribution under independent encoding in both the thermal noise-limited and amplifier noise-limited cases over the range of SNRs analyzed. At MI values of 5.5 and 4.4 bits/symbol, the SNR gaps between the 1-D exponential distribution and Gaussian distribution are 0.078 dB and 0.165 dB in the thermal noise-limited and amplifier noise-limited cases, respectively. This mirrors the result under 3-D encoding and shows that the Gaussian distribution is not the optimal sampled analytic distribution under a product distribution constraint.

A higher SNR is required to achieve a target MI under independent encoding as compared to the joint encoding shown in Fig. 4. The SNR gaps at 5.5 bits/symbol and 4.4 bits/symbol between the optimal 3-D distribution and optimal 1-D distribution are 0.148 dB and 0.214 dB in the thermal noise-limited regime and amplifier noise-limited regime, respectively. We note that the 1-D optimal distribution does not necessarily provide the highest MI among all 1-D distributions in the thermal noise-limited regime. While the 1-D optimal distribution provides the highest MI at low SNR, the 1-D distribution is outperformed by the 1-D exponential and marginal of 3-D optimal distributions at $\mathrm{SNR} = 15$ dB.

The marginal of 3-D optimal distribution is more closely matched by a different 1-D distribution depending on the noise regime. The MI of the 1-D exponential distribution is within 0.01 bits/symbol of the marginal of 3-D optimal distribution for $\mathrm{SNR} \geq 15$ dB in the thermal noise-limited regime. There is also a visible gap in MI between the marginal of 3-D optimal distribution and 1-D optimal distribution for the thermal noise-limited regime. By contrast, the MI of the 1-D optimal distribution is within 0.01 bits/symbol of the marginal of 3-D optimal distribution for $\mathrm{SNR} \geq 15$ dB for the amplifier noise-limited regime. The larger gap in the amplifier noise-limited regime is between the 1-D exponential distribution and marginal of 3-D optimal distribution.

H. Shaping Under Hard-Decision Decoding

We now study the shape gain for an SVR using HD decoding at the receiver. Fig. 6 compares the MI achieved for the 3-D HD optimal, 3-D HD exponential, and 3-D HD Gaussian distributions. We also provide the MI achieved by the 3-D optimal distribution using SD decoding and uniform signaling using HD decoding as references.

The MI achieved using HD decoding is less than that achieved using SD decoding. There are 1.22 dB and 0.98 dB SNR gaps at an MI of 5 bits/symbol between the 3-D optimal distribution using SD decoding and the 3-D optimal distribution using HD decoding in the thermal noise-limited and amplifier noise-limited cases, respectively. This is expected, as quantization at the receiver results in an information loss. Despite this information loss, the 3-D HD optimal distribution still provides shape gains of 0.686 dB and 1.068 dB at an MI of 5 bits/symbol over the HD uniform distribution in the thermal noise-limited and amplifier noise-limited cases, respectively.

The relative ordering of the distributions at fixed SNR does not change when using HD decoding. The 3-D HD optimal distribution provides the highest performance at all SNRs analyzed and is most closely approached by the 3-D HD exponential distribution in both the thermal noise-limited regime and amplifier noise-limited regime. The MI gap between the 3-D HD exponential distribution and 3-D HD optimal is larger in the

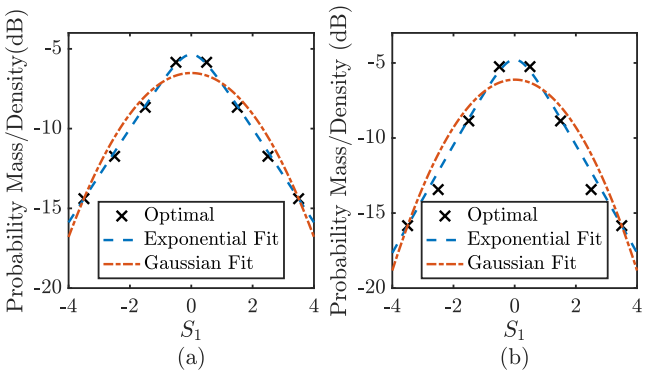

Fig. 7. Probability mass function of S_1 vs. the amplitude of S_1 when $S_2 = S_3 = 0.5$, SNR $= 20$ dB, and $\lambda = 1$ in (a) thermal noise-limited regime and (b) amplifier noise-limited regime. The exponential and Gaussian probability density functions are fitted to the 8 probability masses to minimize the L1 norm of the error.

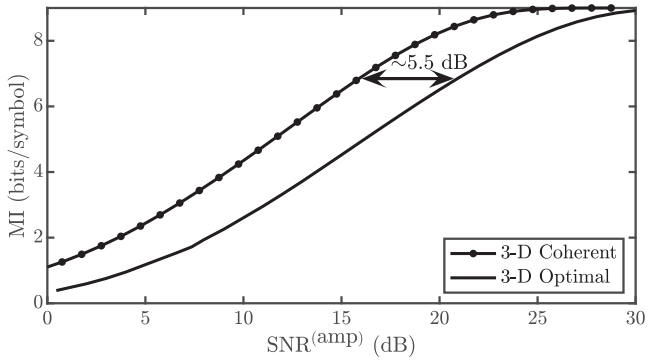

Fig. 8. Mutual information (MI) vs. SNR$^{(amp)}$ for the reference coherent receiver and for the SVR using 3-D input distributions optimized for SCD and SVR, respectively, in the amplifier noise-limited regime.

amplifier noise-limited case than in the thermal noise-limited case at SNR $= 15.5$ dB, which mirrors the result found for SD decoding. The HD exponential distribution outperforms the HD Gaussian distribution over the range of SNRs analyzed in both the thermal noise-limited and amplifier noise-limited cases.

V. Discussion of Optimized Discrete Distributions

In previous sections, we presented analytic results for shaping with continuous, unconstrained constellations and numerical results for shaping with discrete, constrained constellations. Our results for discrete constellations have focused on comparing the MI achieved using different input PMFs. We now study other important properties of the SVR and various optimized discrete input distributions.

A. Exponential vs. Gaussian Distribution

In this section, we compare the PMF designed using the BA method to fitted exponential or Gaussian distributions. Fig. 7 shows the probability mass of the eight-point PMF of S_1 with SNR $= 20$ dB and $S_2 = S_3 = 0.5$. Fig. 7(a) and (b) describe the thermal noise-limited and amplifier noise-limited cases, respectively. The scale parameter is fixed by the procedure described in Section IV-A. A single-parameter search over β yields the minimum L1-norm fit (in dB) of the exponential distribution and Gaussian distribution.

The exponential distribution produces an L1-norm fit of the probability masses with lower residual error than the Gaussian distribution. In addition, the general shape of the 8 probability masses more closely follows the straight lines given by the exponential fit than the quadratic shape of the Gaussian distribution. We observe that the exponential distribution better fits the optimal distribution in the thermal noise-limited regime than in the amplifier noise-limited regime. This arises, in part, from the factor of $1/S_0^{3/2}$ in (51) that is missing in the standard exponential distribution. The results in Fig. 7(a) and (b) are consistent with the Forney method analysis for continuous, unconstrained constellations in Sections III-B and III-C.

B. Noise Enhancement in the Stokes Vector Receiver

We previously compared the MI obtained using the 3-D optimal distribution to the MI achieved by the SCD reference receiver in Fig. 4. We now proceed to limit the degrees of freedom achievable by the SCD receiver to match the SVR in order to study the impact of noise enhancement in the SVR. This study is confined to the amplifier noise-limited regime.

Fig. 8 shows the MI vs. SNR for two different systems. The SVR system uses the 3-D optimal input distribution. The SCD reference system uses the BA method to compute the optimal input distribution subject to the constraint that $E_{2Q} = 0$. We refer to the resulting input distribution as the 3-D optimal coherent distribution. This optimization equates the number of degrees of freedom for the two systems. Any SNR penalty remaining for the SVR must therefore arise from noise enhancement. In Fig. 8, there is roughly a 5.5 dB SNR penalty for the SVR employing the 3-D optimal distribution in comparison to a coherent system employing the 3-D optimal coherent distribution. The SNR gap is 4.56 dB at an MI of 4.5 bits/symbol and reaches a maximum of 5.75 dB at an MI of 8.8 bits/symbol.

In order to explain this SNR gap, we consider the following discrete-time model for a 3-D reference coherent system, which is derived from the 4-D reference coherent system described in Section IV-D. The vector of currents output by the first three of the four balanced photodetectors $\mathbf{I} = [I_{1I}, I_{1Q}, I_{2I}]^T$, after compensating for the channel loss, is given by

$$\mathbf{I} = \sqrt{P_L}\,[E_{1I}, E_{1Q}, E_{2I}]^T + \tilde{\mathbf{N}}^{(\mathrm{co})}. \tag{63}$$

$\tilde{\mathbf{N}}^{(\mathrm{co})} = [\tilde{N}_{1I}^{(\mathrm{co})}, \tilde{N}_{1Q}^{(\mathrm{co})}, \tilde{N}_{2I}^{(\mathrm{co})}]^T$ is the noise vector added to the detected currents, which has covariance matrix

$$\mathrm{cov}\left\{\tilde{\mathbf{N}}^{(\mathrm{co})}\right\} = \frac{1}{\varsigma} P_L\,\mathrm{cov}\left\{U^H \mathbf{N}^{(\mathrm{co})}\right\} = \frac{1}{\sqrt{\varsigma}} P_L \sigma_{\mathrm{co}}^2 \mathbf{I}_3. \tag{64}$$

We have assumed $E_{2Q} = 0$ to limit the number of degrees of freedom to 3.

We model the SVR system using the discrete-time model (26). We assume $\tilde{\mathbf{N}}^{(\mathrm{amp})} \approx \tilde{\mathbf{N}}^{(\mathrm{amp})}$.

In comparing the SVR system and the 3-D coherent system, we focus on a single dimension in each system, as in each system, the noise statistics are identical for all three available

degrees of freedom. For the SVR and 3-D reference systems, the dimensions we analyze are S_1 and the in-phase quadrature of the first polarization, respectively.

The average electrical signal powers in one dimension of the SVR and 3-D coherent systems are $\langle S_0^2\rangle/3$ and $P_L\langle S_0\rangle/3$, respectively. The electrical noise powers in the SVR and 3-D coherent systems are $4\sigma_{\text{amp}}^2\langle S_0\rangle/\sqrt{\varsigma}$ and $P_L\sigma_{\text{co}}^2/\sqrt{\varsigma}$, respectively. The resulting ratios of the electrical signal and noise powers in the SVR and 3-D coherent systems are $\sqrt{\varsigma}\langle S_0^2\rangle/(12\langle S_0\rangle\sigma_{\text{amp}}^2)$ and $\sqrt{\varsigma}\langle S_0\rangle/(3\sigma_{\text{co}}^2)$, respectively. Assuming the two systems have the same optical SNR, i.e., $\sigma_{\text{amp}}^2=\sigma_{\text{co}}^2$, and using Jensen's inequality, the ratio of the electrical signal-to-noise power ratios in the two systems is upper-bounded as follows:

$$\frac{\langle I_{1I}^2\rangle \Big/ \left\langle \left(\tilde{N}_{1I}^{(\text{co})}\right)^2\right\rangle}{\langle S_1^2\rangle \Big/ \left\langle \left(\tilde{N}_1^{(\text{amp})}\right)^2\right\rangle} = 4\frac{\langle S_0\rangle^2}{\langle S_0^2\rangle} \le 4 \approx 6.02\,\text{dB}. \tag{65}$$

The 6.02 dB upper bound in (65) is consistent with the SNR gap of about 5.5 dB shown in Fig. 8.

The physical basis for the SNR gap is as follows. In the SVR, the signal and noise in both polarizations are input into a balanced photodetector to generate S_1 and input into the 90° optical hybrid to generate S_2 and S_3. Each of the resulting SV noise components (22), (23), and (24) includes contributions from both quadratures in both polarizations. The SCD reference system, by contrast, separates the two received polarizations and inputs them into separate 90° hybrids. Each component of the detected current $[I_{1I}, I_{1Q}, I_{2I}]^T$ is only corrupted by the optical noise present in the corresponding polarization and quadrature $[N_{1I}^{(\text{co})}, N_{1Q}^{(\text{co})}, N_{2I}^{(\text{co})}]^T$. The SNR gap for the SVR is analogous to the SNR gaps for non-coherent and differentially coherent detection schemes relative to coherent detection [5], [26].

C. Optimal Scale Factor and Peak Power Constraint

Explicit peak power constraints were not considered in the PMFs analyzed in Section IV. However, the optimization of the scale factor in the BA method from Section IV-A and the sampling procedure from Section IV-B produces an implicit peak power constraint on the finite constellation. We therefore study how the optimized scale factor λ^* changes with SNR.

Fig. 9 shows the optimal scale factor vs. SNR for SNRs between 10 and 30 dB. An initial scale factor of 1 is used and the noise power is chosen to achieve the target SNR assuming a uniform input distribution. The 3-D optimal input distribution is then computed using the BA method. The optimized scale factor is highest at SNRs toward the lower end of the range examined. At the higher SNRs, the scale factor converges to 1 and the input distribution converges to the uniform distribution.

The maximum scale factors are 1.88 and 1.26 in the thermal noise-limited and amplifier noise-limited regimes, respectively. The larger value in the thermal noise-limited regime is due, in part, to the signal dependence of the noise in (25). Increasing the scale factor favors constellation points closer to the origin to reduce the impact of signal-dependent noise.

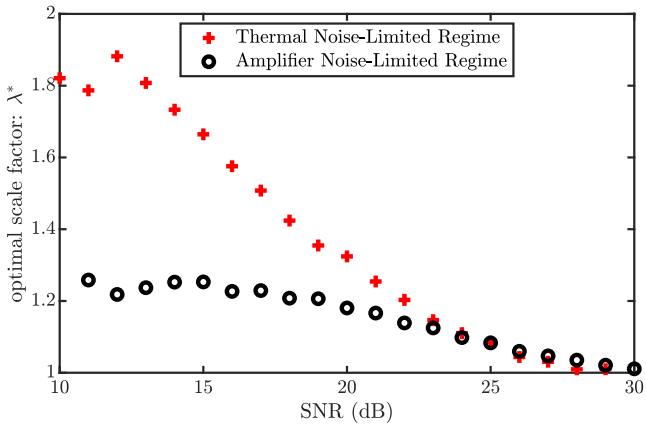

Fig. 9. Optimal scale factor vs. SNR for SVR for the 3-D optimal input distribution. The SNR corresponds to $\text{SNR}^{(\text{th})}$ and $\text{SNR}^{(\text{amp})}$ in the thermal noise-limited and amplifier noise-limited regimes, respectively.

VI. Conclusion

We discussed shaping for the SVR for continuous and discrete input distributions. The Forney method analysis found that an exponential distribution in the intensity is the optimal continuous input distribution in both the thermal noise-limited and amplifier noise-limited regimes at high SNR. Numerical analysis using discrete input constellations found that a sampled exponential distribution in the intensity outperformed a sampled Gaussian input distribution of the transmitted SV in both noise regimes. The optimized discrete input distribution designed using the BA method outperformed both a sampled exponential distribution in the intensity and a sampled Gaussian distribution of the transmitted SV. Comparison of the SVR to a reference coherent system modulating three of the four available degrees of freedom shows that the SVR is subject to an optical power penalty of approximately 5.5 dB. This power penalty arises from noise enhancement due to the beating of signal and noise components from different polarizations in the SVR.

Acknowledgment

Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] S. Benedetto and P. Poggiolini, "Multilevel polarization shift keying: Optimum receiver structure and performance evaluation," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 1174–1186, Feb.–Apr. 1994.
[2] W. Shieh, H. Khodakarami, and D. Che, "Invited article: Polarization diversity and modulation for high-speed optical communications: Architectures and capacity," *APL Photon.*, vol. 1, no. 4, 2016, Art. no. 040801.
[3] D. Che, A. Li, X. Chen, Q. Hu, Y. Wang, and W. Shieh, "Stokes vector direct detection for short-reach optical communication," *Opt. Lett.*, vol. 39, no. 11, pp. 3110–3113, 2014.
[4] K. Kikuchi and S. Kawakami, "Multi-level signaling in the stokes space and its application to large-capacity optical communications," *Opt. Exp.*, vol. 22, no. 7, pp. 7374–7387, 2014.
[5] E. Ip, A. P. T. Lau, D. J. Barros, and J. M. Kahn, "Coherent detection in optical fiber systems," *Opt. Exp.*, vol. 16, no. 2, pp. 753–791, 2008.
[6] J. K. Perin, A. Shastri, and J. M. Kahn, "Data center links beyond 100 Gbit/s per wavelength," *Opt. Fiber Technol.*, vol. 44, pp. 69–85, 2018.

[7] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
[8] G. Forney and L.-F. Wei, "Multidimensional constellations. I. Introduction, figures of merit, and generalized cross constellations," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 877–892, Aug. 1989.
[9] D.-S. Shiu and J. M. Kahn, "Shaping and nonequiprobable signaling for intensity-modulated signals," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2661–2668, Nov. 1999.
[10] W. Mao and J. M. Kahn, "Lattice codes for amplified direct-detection optical systems," *IEEE Trans. Commun.*, vol. 56, no. 7, pp. 1137–1145, Jul. 2008.
[11] S. Hranilovic and F. R. Kschischang, "Optical intensity-modulated direct detection channels: Signal space and lattice codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1385–1399, Jun. 2003.
[12] G. Böcherer, "Probabilistic signal shaping for bit-metric decoding," in *Proc. IEEE Int. Symp. Inf. Theory*, 2014, pp. 431–435.
[13] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 230–244, Jan. 2019.
[14] H. Khodakarami, D. Che, and W. Shieh, "Information capacity of polarization-modulated and directly detected optical systems dominated by amplified spontaneous emission noise," *J. Lightw. Technol.*, vol. 35, no. 14, pp. 2797–2802, Jul. 2017.
[15] P. Dong, X. Chen, K. Kim, S. Chandrasekhar, Y.-K. Chen, and J. H. Sinsky, "128-gb/s 100-km transmission with direct detection using silicon photonic stokes vector receiver and I/Q modulator," *Opt. Exp.*, vol. 24, no. 13, pp. 14208–14214, 2016.
[16] T. Tanemura and Y. Nakano, "Compact InP stokes-vector modulator and receiver circuits for short-reach direct-detection optical links," *IEICE Trans. Electron.*, vol. 101, no. 7, pp. 594–601, 2018.
[17] M. Morsy-Osman, M. Chagnon, and D. V. Plant, "Four-dimensional modulation and stokes direct detection of polarization division multiplexed intensities, inter polarization phase and inter polarization differential phase," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1585–1592, Apr. 2016.
[18] Y. C. Gültekin, T. Fehenberger, A. Alvarado, and F. M. J. Willems, "Probabilistic shaping for finite blocklengths: Distribution matching and sphere shaping," *Entropy*, vol. 22, no. 5, 2020, Art. no. 581. [Online]. Available: https://www.mdpi.com/1099-4300/22/5/581
[19] P. Becker, N. Olsson, and J. Simpson, "Chapter 7 - Optical amplifiers in fiber optic communication systems – Theory," in *Erbium-Doped Fiber Amplifiers*, P. Becker, N. Olsson, and J. Simpson, Eds. San Diego, CA, USA: Academic Press, 1999, pp. 201–250.
[20] S. N. Savenkov, *Jones and Mueller Matrices: Structure, Symmetry Relations and Information Content*. Berlin, Germany: Springer, 2009, pp. 71–119.
[21] I. Weissman, "Sum of squares of uniform random variables," *Statist. Probability Lett.*, vol. 129, pp. 147–154, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167715217302018
[22] P. J. Forrester, "Comment on "sum of squares of uniform random variables" by I. Weissman," *Statist. Probability Lett.*, vol. 142, pp. 118–122, 2018.
[23] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.
[24] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.
[25] T. Fehenberger, F. Kristl, C. Behrens, A. Ehrhardt, A. Gladisch, and N. Hanik, "Estimates of constrained coded modulation capacity for optical networks," in *Proc. Photonic Netw.; 15. ITG Symp.*, 2014, pp. 1–6.
[26] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

Hongxiang Jia received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2021. He is currently working toward the Ph.D. degree with the Department of Electrical Engineering, Stanford University, Stanford, CA, USA. His research interests include optical communications, fiber sensors, fiber optic gyroscopes, lasers, and LiDAR.

Ethan Liang (Graduate Student Member, IEEE) received the A.S. degree in computer science from El Camino College, Torrance, CA, USA, in 2016, and the B.S. degree in electrical engineering from the University of California, Los Angeles, Los Angeles, CA, USA, in 2019. He is currently working toward the Ph.D. degree with the Department of Electrical Engineering, Stanford University, Stanford, CA, USA. His research interests include optical communications and coding theory. He is a National Science Foundation Graduate Research Fellow.

Joseph M. Kahn (Fellow, IEEE) received the A.B., M.A., and Ph.D. degrees in physics from the University of California, Berkeley, Berkeley, CA, USA, in 1981, 1983, and 1986, respectively. From 1987 to 1990, he was with AT&T Bell Laboratories, Holmdel, NJ, USA. In 1989, he demonstrated the first successful synchronous (i.e., coherent) detection using semiconductor lasers, achieving record receiver sensitivity. From 1990 to 2003, he was on the Electrical Engineering and Computer Sciences Faculty with Berkeley. He demonstrated coherent detection of QPSK in 1992. In 1999, D. S. Shiu and Kahn authored or coauthored the first work on probabilistic shaping for optical communications. In the 1990s and early 2000s, Kahn and collaborators performed seminal work on indoor and outdoor free-space optical communications and multi-input multi-output wireless communications. In 2000, Kahn and K. P. Ho founded StrataLight Communications, whose 40 Gb/s-per-wavelength long-haul fiber transmission systems were deployed widely by AT&T, Deutsche Telekom, and other carriers. In 2002, Ho and Kahn applied to patent the first electronic compensation of fiber Kerr nonlinearity. StrataLight was acquired by Opnext in 2009. In 2003, he became a Professor of electrical engineering with the E. L. Ginzton Laboratory, Stanford University, Stanford, CA, USA. Kahn and collaborators have extensively studied rate-adaptive coding and modulation, and digital signal processing for mitigating linear and nonlinear impairments in coherent systems. In 2008, E. Ip and Kahn (and G. Li independently) invented simplified digital backpropagation for compensating fiber Kerr nonlinearity and dispersion. Since 2004, Kahn and collaborators have been studied propagation, modal statistics, spatial multiplexing and imaging in multimode fibers, elucidating principal modes and demonstrating transmission beyond the traditional bandwidth-distance limit in 2005, deriving the statistics of coupled modal group delays and gains in 2011, and deriving resolution limits for imaging in 2013. His research interests include addresses optical frequency comb generators, coherent data center links, rate-adaptive access networks, fiber Kerr nonlinearity mitigation, ultra-long-haul submarine links, and optimal free-space transmission through atmospheric turbulence. He was the recipient of the National Science Foundation Presidential Young Investigator Award in 1991.