Introducing enumerative sphere shaping for optical communication systems with short blocklengths

Document status and date:
Published: 01/12/2019

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Introducing Enumerative Sphere Shaping for Optical Communication Systems with Short Blocklengths

Abdelkerim Amari, *Member, IEEE*, Sebastiaan Goossens, *Student Member, IEEE*, Yunus Can Gültekin, *Student Member, IEEE*, Olga Vassilieva, *Senior Member, IEEE*, Inwoong Kim, *Member, IEEE*, Tadashi Ikeuchi, *Member, IEEE*, Chigo Okonkwo, *Senior Member, IEEE*, Frans M. J. Willems, *Fellow, IEEE*, and Alex Alvarado, *Senior Member, IEEE*

Abstract—Probabilistic shaping based on constant composition distribution matching (CCDM) has received considerable attention as a way to increase the capacity of fiber optical communication systems. CCDM suffers from significant rate loss at short blocklengths and requires long blocklengths to achieve high shaping gain, which makes its implementation very challenging. In this paper, we propose to use enumerative sphere shaping (ESS) and investigate its performance for the nonlinear fiber optical channel. ESS has lower rate loss than CCDM at the same shaping rate, which makes it a suitable candidate to be implemented in real-time high-speed optical systems. In this paper, we first show that finite blocklength ESS and CCDM exhibit higher effective signal-to-noise ratio than their infinite blocklength counterparts. These results show that for the nonlinear fiber optical channel, large blocklengths should be avoided. We then show that for a 400 Gb/s dual-polarization 64-QAM WDM transmission system, ESS with short blocklengths outperforms both uniform signaling and CCDM. Gains in terms of both mutual information and bit-error rate are presented. ESS with a blocklength of 200 is shown to provide an extension reach of about 200 km in comparison with CCDM with the same blocklength. The obtained reach increase of ESS with a blocklength of 200 over uniform signaling is approximately 450 km (approximately 19%).

Index Terms—Constant composition distribution matching, enumerative sphere shaping, fiber nonlinearity, optical communication systems, probabilistic shaping.

I. INTRODUCTION

OPTICAL communication systems have evolved to ensure the growth of Internet traffic, which has exponentially increased in recent years due to multiple modern bandwidth-hungry applications. Fiber optical systems are approaching their capacity limits and an optimal exploitation of the installed network resources is highly desirable to keep up with the traffic demands [1]. In this context, several approaches have been investigated in the literature to increase the spectral efficiency. One of the most popular techniques is fiber nonlinearity mitigation, which can be implemented using different approaches [2], [3]. The main drawback of fiber nonlinearity compensation is its high complexity, which makes its real-time implementation very challenging. Another way to increase the spectral efficiency is by employing sophisticated forward error correcting (FEC) combined with high-order modulation formats, a combination known as coded modulation [4], [5]. Coded modulation can be taken one step further via constellation shaping, a technique that dates back to the early 1990s [6]–[8].

Constellation shaping has recently received a lot of attention in the fiber optical communications community. Two types of constellation shaping have been proposed: geometric shaping [9]–[13], and probabilistic shaping [14]–[23] (but also hybrid approaches). While probabilistic shaping consists of using uniformly spaced constellation points with different probabilities, geometric shaping uses nonuniformly spaced constellation points with same probabilities. In the additive white Gaussian noise (AWGN) channel, both techniques can provide gains up to 1.53 dB. This so-called "ultimate shaping gain" is achieved at infinite blocklengths (for probabilistic shaping) or infinitely dense constellations (for geometrical shaping). The ultimate shaping gain can in principle be exceeded for the nonlinear fiber optical channel, as shown in [24]. Probabilistic shaping has been shown to give larger gains in comparison with geometric shaping, especially when bit-metric decoding is used [4, Ch. 4], [25].

Probabilistic shaping has been widely investigated in the context of optical communication systems not only because of its shaping gain but also because it provides rate adaptivity for systems with fixed FEC [19]. Distribution matching (DM) is considered as a key technique for the implementation of the shaping and deshaping of probabilistically shaped sequences. So far, constant composition DM (CCDM) with long blocklengths is the preferred alternative to achieve the Maxwell-Boltzmann (MB) distribution, which maximizes the shaping gain in AWGN channel. However, it has been observed in [16] that CCDM-based probabilistic shaping has lower tolerance to the fiber nonlinearity in comparison with uniform distribution. Nevertheless, the combination of shaping gain and nonlinear penalty results in an overall increase of performance with respect to uniform signaling [16]. The total gain, however, is lower than for the AWGN channel due to the nonlinear penalty. CCDM with long blocklengths is also very difficult to implement in high-speed communications because it is based on *sequential* arithmetic coding. A way to decrease the blocklength by reducing the rate loss, known as multiset-partition DM, has been recently presented in [23].

A. Amari, Y. C. Gültekin, and F. M. J. Willems, and A. Alvarado are with the Information and Communication Theory Lab, Signal Processing Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven 5600 MB, The Netherlands. Email: a.amari@tue.nl.

S. Goossens and C. Okonkwo are with the Institute for Photonic Integration, Eindhoven University of Technology, Eindhoven 5600 MB, The Netherlands.

O. Vassilieva, I. Kim, and T. Ikeuchi are with Fujitsu Laboratories of America, Inc., Richardson, TX 75082 USA.

In this paper, we propose to use enumerative sphere shaping (ESS) [26]–[32] as an alternative for CCDM. ESS was introduced in the context of wireless communication systems to improve the system capacity in [27]. ESS achieves the Maxwell-Boltzmann (MB) distribution at infinite shaping blocklength and performs better than CCDM at short blocklengths [29], [32].

In this paper, we compare the performance of ESS and CCDM through numerical simulations. To the best of our knowledge, this is the first time such comparison is made in the context of the nonlinear fiber optical channel. Both shaping approaches improve the performance in comparison with uniform signaling at long blocklengths. At short blocklengths, ESS is shown to outperform CCDM due to its lower rate loss. In addition, we show that short blocklengths ESS and also CCDM provide *higher* effective signal-to-noise ratio (SNR) than uniform 64-QAM. This is in contrast to the long blocklengths case, where probabilistic shaping causes an effective SNR reduction.

The key observation in this paper is that short blocklengths ESS should be the preferred alternative for shaping in the nonlinear fiber optical channel. This is due to a balance between linear and nonlinear gains, low rate loss, and high effective SNR. Mutual information (MI) results confirm that the best performance is obtained for short blocklengths ESS. The MI of short blocklength ESS is shown to be higher than long blocklengths ESS, short and long blocklengths CCDM, and uniform signaling. For a 400 Gb/s dual-polarization 64-QAM WDM transmission system, the effecive SNR and MI gains are translated to considerable transmission reach increases. For ESS with a blocklength of 200, gains of about 100 km, 200 km, and 450 km are reported in comparison with CCDM with blocklength of 3600, CCDM with blocklength of 200 and uniform, resp. End-to-end results including shaping and FEC are also presented in this paper.

The remainder of the paper is organized as follow. The system model and signaling schemes are discussed in Sec. II. The working principles of CCDM and ESS techniques are reviewed in Sec. III. The simulation setup, performance metrics, and numerical results are presented in Secs. V and VI. Conclusions are drawn in Sec. VII.

II. SYSTEM MODEL AND SIGNALING SCHEMES

In this work, we consider the probabilistic amplitude shaping (PAS) scheme as an alternative to uniform signaling schemes. We assume transmission using polarization multiplexed (PM) square QAM constellations. The four-dimensional signal space (PM-QAM signals) is generated by the Cartesian product of four identical 2^m-ary pulse amplitude modulation (PAM) constellations using the same binary labeling. Without loss of generality, we consider the underlying 2^m-PAM constellation. The alphabet (constellation) has the form $\mathcal{X} = \{\pm 1, \pm 3, \cdots, \pm 2^m - 1\}$ where $m \geq 2$. These alphabets can be factorized as $\mathcal{X} = \mathcal{S} \times \mathcal{A}$, where $\mathcal{S} = \{-1, +1\}$ and $\mathcal{A} = \{1, 3, \cdots, 2^m - 1\}$. We refer to $\mathcal{S}$ and $\mathcal{A}$ the sign and amplitude alphabets, resp., where $|\mathcal{S}| = 2$ and $|\mathcal{A}| = 2^{m-1}$.

One requirement of the PAS scheme is that the binary label of the PAM constellation can be separated into two parts: a sign bit and $m - 1$ amplitude bits. In this paper we consider the binary reflected Gray coded (BRGC) [35], which due to its recursive construction [36, Sec. IV] satisfies this condition. The BRGC is also known to be asymptotically the best binary labeling in terms of maximizing the generalized mutual information (GMI) [37, Sec. IV].

A. Uniform Signaling

In uniform signaling, a k-bit uniform information sequence $\boldsymbol{u} = (u_1, u_2, \ldots, u_k)$ is encoded by a rate $R_c = k/n_c$ FEC code.[1] The n_c-bit coded sequence $\boldsymbol{c} = (c_1, c_2, \ldots, c_n)$ is then divided into length-m binary vectors $\boldsymbol{b} = (b_1, b_2, \ldots, b_m)$, each of which is mapped to a channel input x via the BRGC. The sequence of symbols transmitted through the channel is $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$, where N is the number of transmitted symbols. The information rate of this scheme is $R = k/N = mR_c$ [bits/1D-sym], or equivalently, $2mR_c$ [bits/2D-sym]. This scheme is shown in Fig. 1 (a).

B. Nonuniform Signaling

The PAS architecture shown in Fig. 1 (b) is a reverse concatenation structure, where the shaping operation precedes FEC encoding. First, an amplitude shaping block transforms the uniform information sequence $\boldsymbol{u}$ into a shaped amplitude sequence $\boldsymbol{a} = (a_1, a_2, \ldots, a_N)$, where $a_n \in \mathcal{A}$ with $n = 1, 2, \ldots, N$. The sequences $\boldsymbol{a}$ are chosen to satisfy a predefined condition. The rate of this shaper, referred to as the shaping rate, is $R_s = k/N$ [bits/amp]. This conversion is an invertible mapping which can be realized by various shaping approaches (e.g., CCDM or ESS). Next, the amplitudes $\boldsymbol{a}$ are mapped into bits using the last $m - 1$ bits of the BRGC. This is shown in Fig. 1 (b) as amplitude-to-bit conversion, which generates $N(m - 1)$ nonuniform bits.

In the simplest form of the PAS architecture, the $N(m-1)$ nonuniform bits are fed to a rate $R_c = (m-1)/m$ systematic FEC encoder, which generates Nm code bits and N parity bits. As shown in Fig. 1 (b), this can accomplished using a parity-check matrix $\mathbb{P}$ of dimensions $(m - 1)N \times N$. The N parity bits are mapped to the first bit of the 2^m-PAM mapper, and thus, they select the signs (they are mapped to $\mathcal{S}$). These are called sign bits. The $N(m - 1)$ nonuniform bits are used as bits 2 to m the 2^m-PAM mapper, and thus, they are referred to as the shaping bits. Finally, the sequence $\boldsymbol{x} \subset S^N \times A^N$ is transmitted over the channel. The information rate of this scheme is $R = R_s = k/N$ [bits/1D-sym].

The simplest form of the PAS scheme in Fig. 1 (b) assumes a FEC rate of $R_c = (m-1)/m$. Lower FEC rates can be used if the bit-level shaping schemes in [33], [38] are considered. Higher FEC code rates ($R_c > (m-1)/m$) can be used via a straightforward modification of the PAS architecture. This is shown in Fig. 1 (c). The FEC rate in this case is assumed to be $R_c = (m - 1 + \gamma)/m$, where $0 < \gamma < 1$ represents additional information bits. As shown in Fig. 1 (c), the information sequence is of length $k + \gamma N$. In the lower branch, k bits

---

[1]Throughout this paper, boldface letters $\boldsymbol{x}$ denote vectors and blackboard letters $\mathbb{X}$ denote matrices.
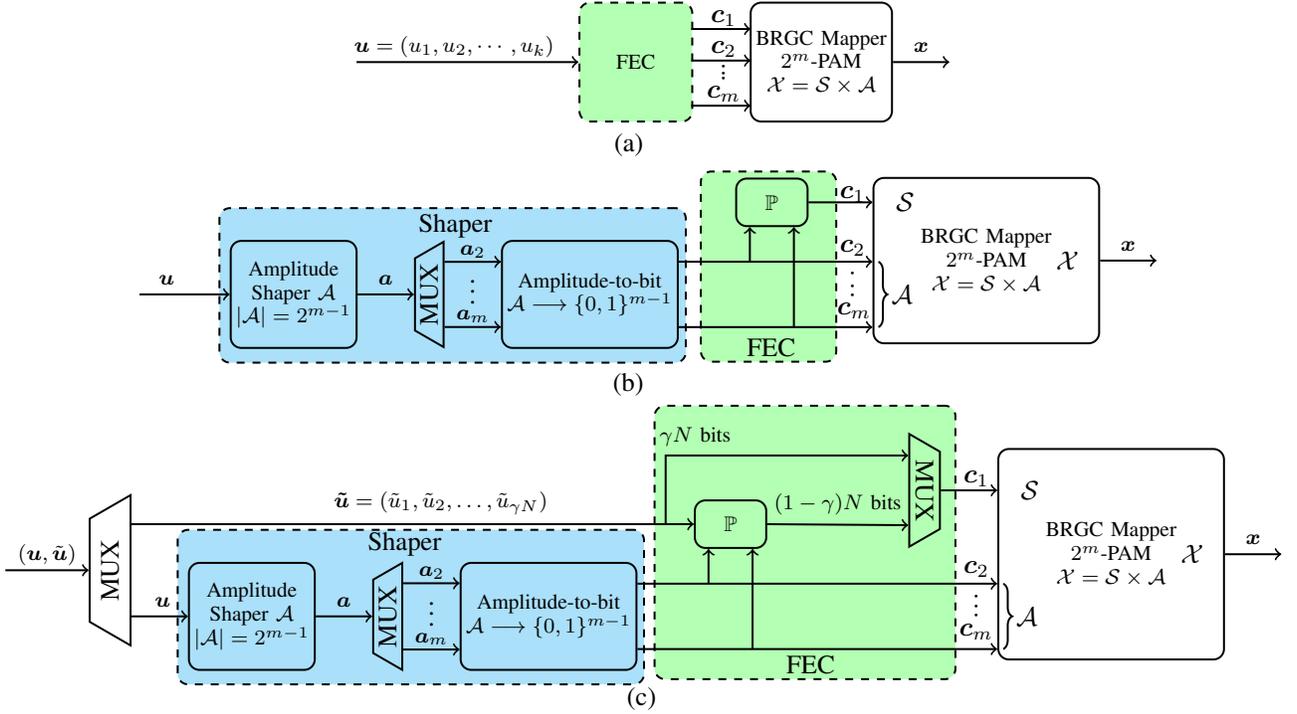
Fig. 1: Signaling alternatives: (a) uniform signaling, (b) PAS architecture with FEC rate $R_c = (m-1)/m$, and (c) PAS architecture with FEC rate $R_c > (m-1)/m$.

are shaped in the same way as in Fig. 1 (b). In the upper branch, γN information bits are used together with $(m-1)N$ shaped bits to generate $(1-\gamma)N$ parity bits. These $(1-\gamma)N$ parity bits are multiplexed with the uniform γN bits to create the vector c_1 of N sign bits.

In the modified architecture in Fig. 1 (c), the parity-check matrix $\mathbb{P}$ is of dimensions $(m-1+\gamma)N \times N$. The information rate of this scheme is $R = (k + \gamma N)/N = R_s + \gamma$ [bits/1D-sym]. Note that the case $\gamma = 0$ corresponds to the scheme in Fig. 1 (b), while $\gamma = 1$ corresponds to uncoded transmission.

Example 1: Consider the uniform signaling scheme which combines a rate $R_c = 3/4$ FEC code with 8-PAM ($m = 3$). The transmission rate this scheme is $R = R_c m = 2.25$ [bits/1D-sym]. To achieve the same rate with a shaped 8-PAM constellation, we need to somehow compensate for the decrease in rate caused by shaping. Thus, we use a higher rate FEC code, more precisely $R_c = 4/5$, which leads to $\gamma = 0.4$. Combining this code with a shaping rate $R_s = 1.85$ bits/amp shaper, we obtain the same rate, i.e., $R = R_s + \gamma = 2.25$ [bits/1D-sym].

III. AMPLITUDE SHAPING: CONSTANT COMPOSITION AND SPHERE CODING

A nonuniform channel input distribution can be obtained by changing the bounding geometry of the multidimensional signal space. Consider for example the transmission of N amplitudes $\boldsymbol{a} = (a_1, a_2, \ldots, a_N)$, where each amplitude is taken from a set $\mathcal{A}$. If no coding is applied, the transmitted codewords are all the $|\mathcal{A}|^N$ N-dimensional vectors from the set $\mathcal{A}^N$. If FEC is applied, some sequences are never transmitted, and thus, the codewords belong to a subset of $\mathcal{A}^N$. Probabilistic shaping can be interpreted as the process of selecting the subset of *allowed* sequences from $\mathcal{A}^N$. In what follows, we describe two approaches to do this: constant composition and sphere coding.

A. Constant Composition Coding

CCDM consists in generating probabilistically shaped sequences, with a given probability distribution, from uniformly distributed bits. It is inspired by arithmetic coding for data compression, in which the sequences are represented by intervals [14]. The performance of CCDM in terms of shaping gain is optimum at infinite blocklengths $N \to \infty$. In this regime, any desired distribution can be achieved, including the MB distribution, which is optimal for the AWGN channel. However, CCDM suffers from high rate loss at short blocklength [23], [29].

CCDM only allows sequences $\mathcal{A}^\circ \subset \mathcal{A}^N$ that satisfy a particular composition. A sequence $\boldsymbol{a}$ is valid only if out of the N amplitudes, the sequence contains exactly n_1 amplitudes one, n_2 amplitudes three, n_3 amplitudes five, etc. The composition is therefore $(n_1, n_2, \ldots, n_{2^m-1})$. By definition, all the sequences generated by CCDM have the same per-codeword energy E°, which depends on the targeted composition.

Example 2 (Geometry of Constant Composition Coding): The geometrical interpretation of CCDM is therefore sequences on an N-dimensional sphere of radius $\sqrt{E^\circ}$. This is schematically shown in Fig. 2 (a), where a 2D projection is presented. The fact that CCDM does not fully cover the sphere comes from the fact that not all sequences on the shell of the N-dimensional sphere of radius $\sqrt{E^\circ}$ satisfy the composition constraint.
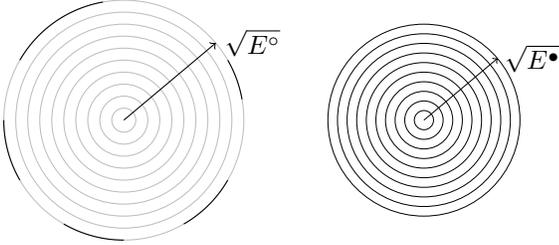
Fig. 2: Geometric representation in 2D of shaping schemes: (a) Constant composition, and (b) Sphere shaping.

B. Sphere Coding

Sphere coding is the process of constraining the codewords to be selected from within an N-sphere. In this case, the induced distribution in one real dimension converges to a MB distribution as $N \to \infty$ (or to a Gaussian distribution if the set $\mathcal{A}$ has a continuous support).

The idea of sphere shaping is to utilize the set of bounded-energy (i.e., spherically-constrained) amplitude sequences

$$\mathcal{A}^\bullet = \left\{ a_1, a_2, \ldots, a_N \middle| \sum_{n=1}^{N} a_n^2 \leqslant E^\bullet \right\} \subset \mathcal{A}^N, \quad (1)$$

where $E^\bullet$ is a real constant. The value of $E^\bullet$ determines the "amount of shaping": the smaller the value of $E^\bullet$, the more shaped the signal will be. Note that if the alphabet $\mathcal{A}$ is discrete, the energy levels will be quantized.

Example 3 (Geometry of Sphere Coding): The set $\mathcal{A}^\bullet$ in (1) consists of all amplitude sequences having an energy no greater than $\sqrt{E^\bullet}$. Geometrically, this corresponds to all points inside or on the surface of an n-sphere of radius $\sqrt{E^\bullet}$ are employed. This is shown schematically in Fig. 2 (b). This figure shows schematically that sphere coding can transmit more sequences (i.e., $|\mathcal{A}^\bullet| > |\mathcal{A}^\circ|$). Furthermore, the energy of the inner shells in $\mathcal{A}^\bullet$ is smaller, and thus, the spheres in Fig. 2 (b) are smaller than those in Fig. 2 (a) (energy can be saved via sphere shaping).

Due to the sphere hardening, all sequences will concentrate near the surface of the N-sphere as $N \to \infty$ which means that CCDM and sphere shaping are asymptotically equivalent for large N. However, for finite blocklengths, there is a rate loss of CCDM with respect to sphere shaping. The rate loss is defined as [29]

$$R_L \triangleq H(P_A) - \frac{k}{N} = H(P_A) - R_s \text{ [bits/amp]}, \quad (2)$$

where P_A is the targeted probability distribution and $H(P_A)$ is the entropy in bits/amp.

Example 4 (Rate Losses): The rate losses of CCDM and ESS are shown in Fig. 3. At a rate loss of 0.02 bits/amp., ESS reduces the shaping blocklength by about 70% when used instead of CCDM. Similar results for an entropy of 1.75 bits/amp. were presented in [32, Fig. 2].

C. Enumerative Sphere Shaping

To efficiently index amplitude sequences bounded by a maximum energy constraint, i.e., sphere constraint, the enu-
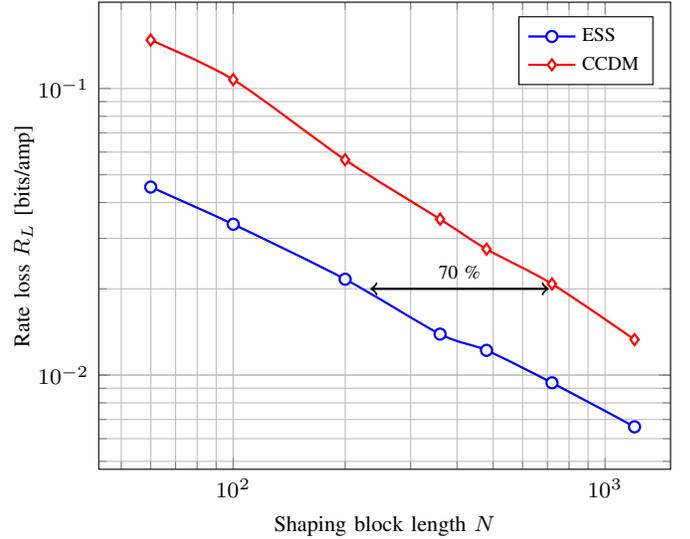

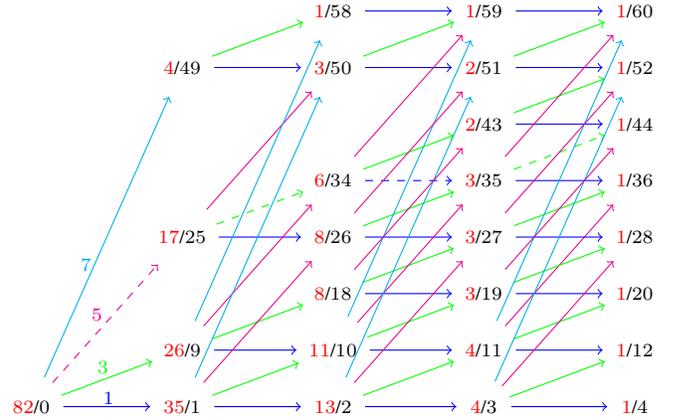Fig. 3: Rate loss vs. shaping blocklength for ESS and CCDM.

Fig. 4: Enumerative trellis for $N = 4$, $\mathcal{A} = \{1, 3, 5, 7\}$ and $E^\bullet = 60$. Each state shows the number of unique paths advancing from that state to one of the states in the final stage (red numbers), and the accumulated energy of the paths arriving to that state (black numbers).

merative approach was introduced in [27]. Enumerative sphere shaping (ESS) starts from the assumption that these sequences can be ordered lexicographically. To create this ordering, a bounded-energy amplitude trellis is constructed, as shown in the following example.

Example 5 (ESS Trellis): Fig. 4 shows a bounded-energy amplitude trellis with $N = 4$, $\mathcal{A} = \{1, 3, 5, 7\}$ and $E^\bullet = 60$. In this trellis, each state represents an energy level, which is indicated using black numbers. Each branch designates an amplitude from $\mathcal{A}$. Each path, starting from the zero-energy state (i.e., bottom left) and ending in one of the states in the final stage, represents an amplitude sequence. These sequences are composed of the corresponding amplitude values (different colors) of the branches of their path. The energy level at stage n that a path passes through, is the accumulated energy of the sequence over the first n components.

A key number for the enumerative shaping and deshaping algorithms is T_n^e which is the number of total paths advancing from a state of energy e in stage n to one of the final states. T_0^0 is therefore the total number of sequences represented in the trellis, i.e., $T_0^0 = |\mathcal{A}^\bullet|$, which is 82 in Example 5, see Fig. 4. The rate of the shaper that indexes sequences from $\mathcal{A}^\bullet$ can then be expressed as

$$R_s = \frac{\log_2 T_0^0}{n} \text{ [bits/amp]}. \tag{3}$$

The values of T_n^e (shown with red in Fig. 4), can be calculated in a recursive manner as

$$T_n^e \triangleq \sum_{a \in \mathcal{A}} T_{n+1}^{e+a^2}, \tag{4}$$

where the states in the last stage are initialized with ones, hence

$$T_N^e = \begin{cases} 1, & \text{if } e = N, N + 8, N + 16, \ldots, E^\bullet\} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Finding the index i of a sequence $\boldsymbol{a} = (a_1, a_2, \ldots, a_N)$ is equivalent to count the number of sequences which are lexicographically smaller than $\boldsymbol{a}$. This can be implemented by considering the path in the trellis corresponding to $\boldsymbol{a}$. Following this path, at stage n for $n = 0, 1, \cdots, N - 1$, we accumulate the number of sequences which have their first n elements identical to $\boldsymbol{a}$ and are lexicographically smaller than $\boldsymbol{a}$. The procedure starts in the zero-energy state. All sequences which start with an amplitude $a < a_1$ have a smaller index than $\boldsymbol{a}$. Thus we sum the corresponding $T_1^{a^2}$ values and arrive at the state of energy a_1^2 in the first stage. At this point, all sequences which start with a_1 and continue with an amplitude $a < a_2$ have a smaller index than $\boldsymbol{a}$. Thus we add the corresponding $T_2^{a_1^2 + a^2}$ values to our accumulating sum and arrive at the state of energy $a_1^2 + a_2^2$ in the second stage. Repeating this procedure recursively, we arrive at the state of energy $\sum_i a_i^2$ at final stage, and the accumulated sum is the index $i(\boldsymbol{a})$ of our sequence $\boldsymbol{a}$. This leads to Cover's indexing formula, see [34],

$$i(\boldsymbol{a}) = \sum_{n=1}^{N} \sum_{b \in \mathcal{A}: b < a_n} T_n^{b^2 + \sum_{m=1}^{n-1} a_m^2}. \tag{6}$$

The following example demonstrates (6) based on the trellis in Fig. 4.

Example 6 (ESS Indexing): Consider the sequence $\boldsymbol{a} = (5, 3, 1, 3)$, which is represented by a path passing through energy levels $\boldsymbol{e} = (0, 25, 34, 35, 44)$ in stages $n = 0, 1, 2, 3, 4$, respectively. This path is shown with dashed lines in Fig. 4. At first, we count the number of sequences which start either with a 1 or a 3 since all these will have a smaller index than $\boldsymbol{a}$. Thus we add the numbers $T_1^1 = 35$ and $T_1^9 = 26$ to get 61, and arrive at the node of energy $a_1^2 = 25$ in the first stage. Then we need to add the number of sequences which start with $(5, 1)$ since all these will have a smaller index than $\boldsymbol{a}$. Therefore we add $T_2^{26} = 8$ to our accumulating sum to get 69. Our sequence have a 1 in the third position which makes it lexicographically the first, given that the first two elements are fixed. Thus we add nothing in this stage. Finally, we need

TABLE I: WDM simulation parameters.

Number of WDM channels	11
Number of polarizations	2
Symbol rate	45 GBd
Channel spacing	50 GHz
Modulation	64 QAM
RRC roll off ρ	0.1
Span length L	80 km
EDFA noise figure	5 dB
Number of LDPC blocks	100

to account for the sequence $(5, 3, 1, 1)$ since it is the only sequence which has its first three components identical to $\boldsymbol{a}$ and has a smaller index. Thus we add $T_4^{36} = 1$ to find the total index which is 70.

The indexing algorithm (shaping) and its inverse (deshaping) can be implemented in a recursive way as N-step operations, see Algorithms 1 and 2 in [32].

IV. SIMULATION SETUP AND PERFORMANCE METRICS

A. Simulation Setup

The performance of the ESS algorithm, which is implemented following [30], is compared to state-of-the-art CCDM-based probabilistic shaping and uniform signaling through numerical simulations. Single-polarization single-channel and dual-polarization wavelength-division multiplexing (WDM) long-haul transmission configurations are considered.

The simulation parameters for the WDM transmission are given by Table I. The net bit rate per WDM channel is 405 Gb/s. 64-QAM (8-PAM per real dimension) is used as modulation formats for both shaped and uniform signaling. We consider low-density parity check (LDPC) codes with coding rates $R_c = 3/4$ and $R_c = 4/5$, as described in Example 1. This ensures two transmission scenarios (uniform and shaped) with same net bit rate.

The transmission link consists of multi-span standard single mode fiber (SSMF) with an attenuation coefficient $\alpha = 0.2 \text{ dB} \cdot \text{km}^{-1}$, a dispersion parameter $D = 17 \text{ ps} \cdot \text{nm}^{-1} \cdot \text{km}^{-1}$, and a nonlinear coefficient $\gamma = 1.3 \text{ W}^{-1} \cdot \text{km}^{-1}$. An erbium-doped fiber amplifier (EDFA) with a 5 dB noise figure and 16 dB gain is used to periodically amplify the signal after each span of $L = 80$ km.

Each real-dimension of the nonuniform signaling simulation setup is shown in Fig. 5. Firstly the information bits $\boldsymbol{u}$ are shaped using ESS or CCDM. The shaping and LDPC blocks operate as explained in Sec. II (see Fig. 1). After PAM mapping, the signal is oversampled with 16 samples/symbol and passed through a root-raised cosine (RRC) filter, with roll-off factor $\rho = 0.1$, for spectrum shaping.

At the receiver side, after channel selection and downsampling the signal to 2 samples/symbol, chromatic dispersion compensation is performed. After that, an RRC matched filter is applied before downsampling to 1 sample/symbol. Phase rotation is ideally compensated. At this stage, the effective SNR and MI are estimated (see next section). Then, the log-likelihood ratios (LLRs) $\lambda_1, \lambda_2, \ldots, \lambda_m$ are calculated for the
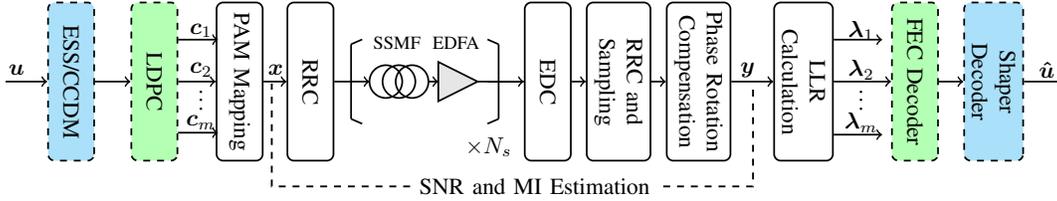
Fig. 5: Transmission diagram: ESS: Enumerative sphere shaping, CCDM: Constant composition distribution matching, LDPC: Low-density parity check, PAM: Pulse amplitude modulation, RRC: Root-raised cosine, SSMF: standard single mode fiber, EDFA: Erbium-doped fiber amplifier, EDC: Electronic dispersion compensation, SNR: Signal-to-noise ratio, MI: Mutual information, N_s: number of spans, $\lambda_1, \cdots, \lambda_m$: Log-likelihood ratios.

bits $c_1, c_2, \ldots, c_m$, which are then passed to the soft-decision LDPC decoder. Finally, ESS or CCDM deshaping is performed to recover the transmitted data bits. In the case of uniform signaling, the shaping and deshaping blocks are removed from the simulation setup (see Fig. 1 (a)).

B. Performance metrics

In this work, we mainly use two metrics to evaluate the performance of ESS in comparison with CCDM and uniform signaling: effective SNR, and mutual information (MI).

The effective SNR is defined as [17]

$$\text{SNR}_{\text{eff}} \approx \frac{\mathbb{E}[|X|^2]}{\mathbb{E}[|Y - X|^2]} \quad (7)$$

where $\mathbb{E}[\cdot]$ represents expectation and X and Y are the transmitted and received symbols respectively. The effective SNR in (7) includes the amplified spontaneous emission (ASE) noise and the nonlinear noise, takes into accounts the probability of each constellation points. The effective SNR was calculated per QAM symbol.

The second metric we use is a "finite blocklength MI", which is the symbol-wise analogous of [23, eq. (15)]. Our MI is calculated per QAM symbol (2D-sym), and and defined as

$$I_N(X;Y) = H(X) - H(X|Y) - 2 \cdot R_L \text{ [bits/2D-sym]}, \quad (8)$$

where H is the entropy. The MI in (8) is the difference between the classic MI (i.e., $I(X;Y) = H(X) - H(X|Y)$) and the rate loss per QAM symbol. This definition allows us to have fair comparisons for short blocklengths, where the rate loss could be significant. The MI in (8) converges to the classic MI for infinite blocklengths.

V. SIMULATION RESULTS: SINGLE-POLARIZATION SINGLE-CHANNEL TRANSMISSION

Here evaluate the performance of ESS and CCDM and uniform signaling in the context of single-polarization single-channel transmission. To understand the effect of the shaping blocklength in the nonlinear fiber channel, we start by studying the linear channel, in which the fiber nonlinearity is turned off. In Fig. 6, we plot the MI versus the SNR for different shaping blocklengths N. ESS performs better than CCDM at short blocklengths due to its lower rate loss. The gain for a blocklength $N = 200$ is about 0.1 bits/2D-sym at an SNR of
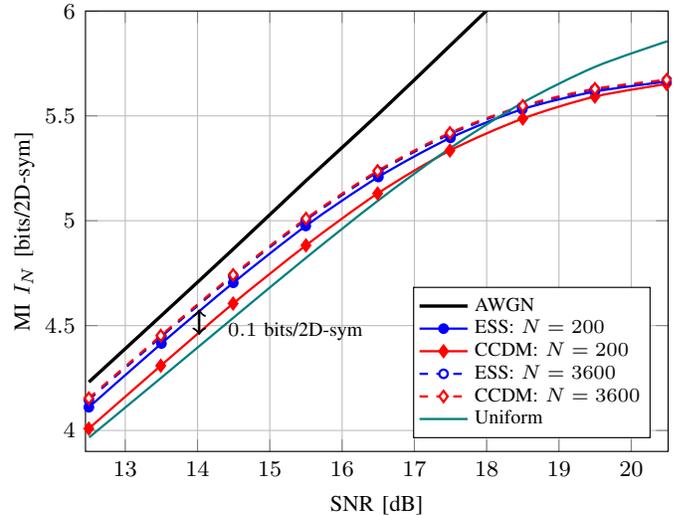

Fig. 6: MI vs. SNR for linear fiber channel and different shaping blocklengths.

14 dB. ESS and CCDM at $N = 3600$ exhibit similar performances and show a gain of about 0.03 bits/2D-sym and 0.19 bits/2D-sym in comparison with ESS at $N = 200$ and uniform, respectively at 14 dB SNR. Increasing the blocklength in the linear channel improves the performance and reduces the gap with the AWGN capacity. For high SNR regime (SNR > 18 dB), uniform signaling gives better performance than shaped constellations.

In Fig. 7, we evaluate the MI as a function of the blocklength N for both the linear and nonlinear channels. We first show the performance for the nonlinear fiber channel at optimal input power. Then, we fix the SNR for the linear channel so that the MI performance for uniform signaling in linear and nonlinear channels are the same. In this case, in addition to the comparison of ESS, CCDM and uniform signaling in the nonlinear channel, we can also compare the gain/penalty of shaped schemes due to the fiber nonlinearity. As shown in Fig. 7, for the linear channel case (dashed lines), ESS outperforms CCDM at short blocklengths due to its low rate loss. Then, at long blocklength ($N = 3600$) in which the rate losses became negligible, ESS and CCDM converges to their maximum MI and exhibits similar performances.

In the nonlinear channel case (solid lines in Fig. 7), we observe that in comparison with the linear channel, ESS and
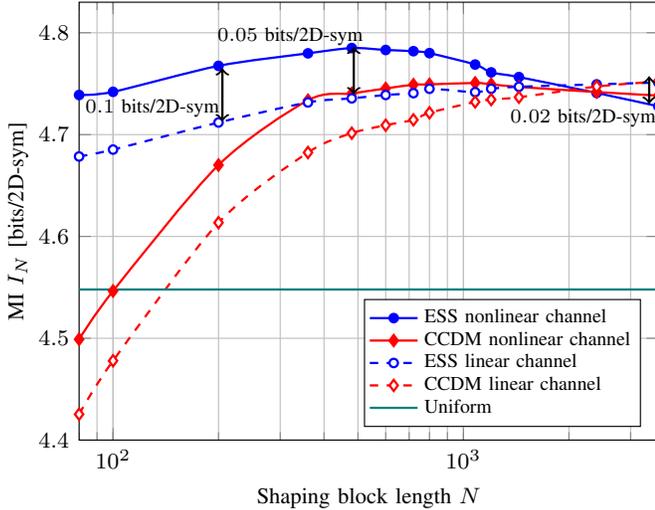
Fig. 7: MI vs. shaping blocklength N for linear and nonlinear fiber channel at optimal input power.
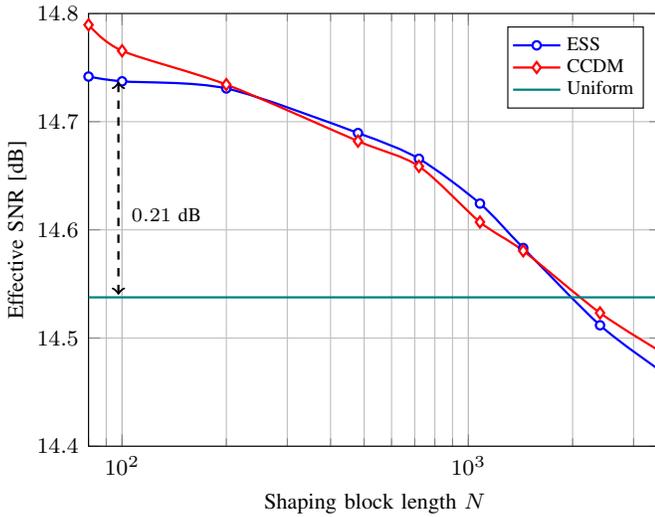

Fig. 8: Effective SNR vs. shaping blocklength N for nonlinear fiber channel at optimal input power. The effecive SNR gap between ESS and unfirom at $N = 100$ is shown.

CCDM reach their maximum performance in terms of MI at $N = 480$ and $N = 1200$, respectively. In addition, the ESS performance at $N = 480$ increases by 0.05 bits/2D-sym in comparison with linear channel performance. ESS still outperforms CCDM at short blocklength and the gain is 0.1 bits/2D-sym at $N = 200$. At long blocklengths, the performances of ESS and CCDM are slightly decreased unlike the linear channel transmission case. These results can be explained by the effective SNR performance, which measures the nonlinear tolerance of the different constellations.

In Fig. 8, we plot the effective SNR versus the shaping blocklength N for nonlinear fiber channel at optimal input power. We observe that short blocklengths ESS and CCDM have better performance in terms of the effective SNR, and consequently higher nonlinear tolerance, than long blocklengths case. The gain for ESS at $N = 100$ is about
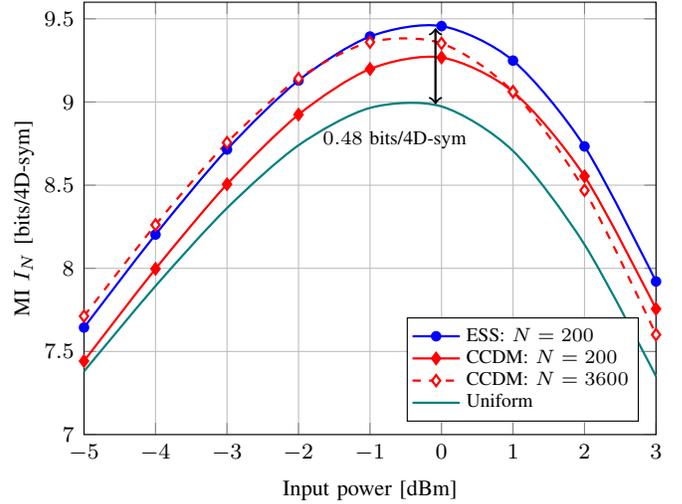

Fig. 9: MI vs. input power for different blocklengths. At optimal input power, ESS at $N = 200$ shows the highest MI performance in comparison with CCDM at $N = 200$ and $N = 3600$, and uniform.

0.27 dB in comparison with ESS at $N = 3600$. ESS and CCDM at the same blocklength exhibit similar performance in terms of effective SNR, and their performances are inversely proportional to the blocklength.

An important observation from Fig. 8 is the fact that ESS and CCDM at short blocklengths give better SNR performance with respect to uniform signaling. Then, this performance gain is decreased for longer blocklengths. For blocklengths $N > 2000$, uniform signaling gives better performance than CCDM, which coincides with the state of the art results [16]. The same behavior is observed with ESS. Our results therefore show that short-blocklength ESS (which exhibits the best performance) and CCDM, provide a combination of linear shaping gain and nonlinear tolerance gain in comparison with uniform. Long-blocklength ESS and CCDM provide higher linear shaping gain but at the same time, they are affected by a nonlinear penalty. The overall gain is lower than the short blocklength shaping schemes, and this, finite short-blocklength shaping is the best alternative.

VI. SIMULATION RESULTS: DUAL-POLARIZATION WDM TRANSMISSION

In the following, we focus on the dual-polarization WDM configuration for which the simulation parameters are shown in Table I. We compare ESS and CCDM at a blocklength of $N = 200$, as well as CCDM with long blocklength ($N = 3600$), and uniform signaling.

In Fig. 9, the MI is plotted as a function of input power. At optimal input power 0 dBm, ESS at $N = 200$ shows a gain of about 0.1 bits/4D-sym, 0.19 bits/4D-sym and 0.48 bits/4D-sym in comparison with CCDM at $N = 3600$, CCDM at $N = 200$, and uniform signaling, respectively. In the linear regime at low input powers, CCDM at $N = 3600$ exhibits the highest linear shaping gain because it has the lowest rate loss. ESS at $N = 200$ shows similar MI performance to CCDM at
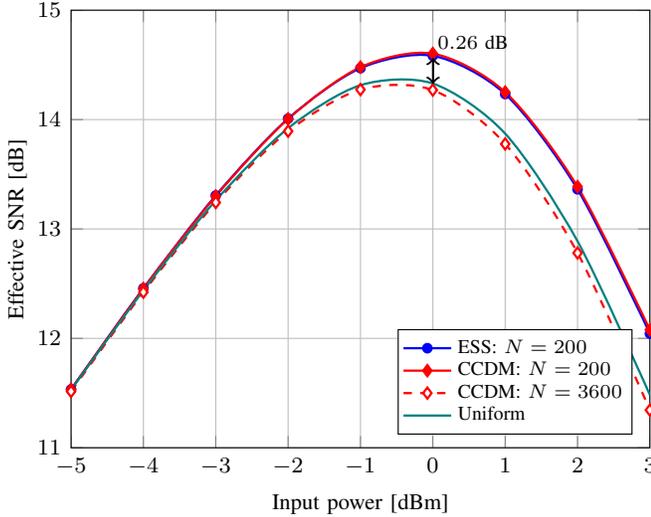
Fig. 10: SNR vs. input power for different blocklengths. Short blocklengths ESS and CCDM exhibit higher nonlinear tolerance than uniform signaling and long blocklength CCDM.
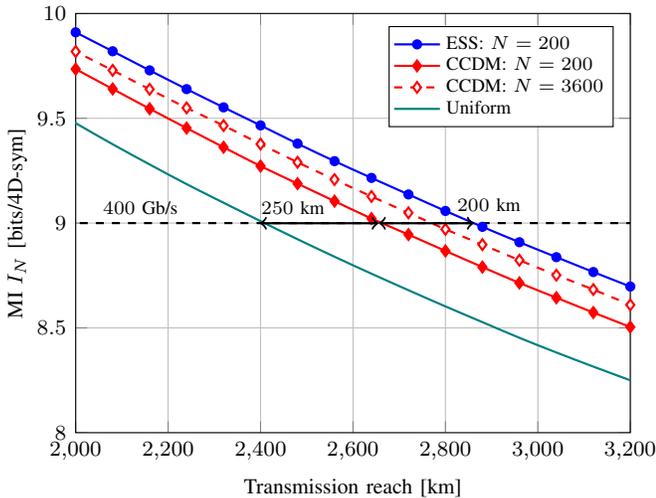

Fig. 12: BER post-LDPC vs. transmission reach for different blocklengths at optimal input power. Square markers indicate noise-loaded results.

Fig. 11: MI vs. transmission reach for different blocklengths at optimal input power.

$N = 3600$ and better performance than CCDM at $N = 200$. In the nonlinear regime, ESS and CCDM at $N = 200$ show higher MI than CCDM at $N = 3600$. The performance gap between uniform signaling and CCDM at $N = 3600$ is also reduced. This can be explained by the nonlinear penalty that long blocklengths CCDM suffers from, in comparison with uniform signaling and the short blocklengths case.

The results in Fig. 10, in which the effective SNR is plotted as a function of the input power, confirms that ESS and CCDM at $N = 200$ are more tolerant to the fiber nonlinearity. The gain at optimal input power is about 0.32 dB and 0.26 dB in comparison with long blocklength CCDM and uniform, respectively. In the linear regime at low input powers, ESS, CCDM, and uniform signaling exhibits similar performance as expected for the case of linear channel.

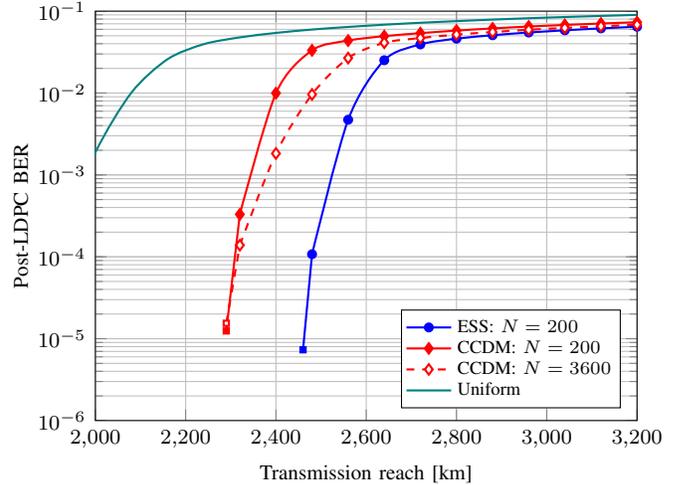We also evaluate the transmission reach increase obtained by using ESS and CCDM wth respect to the uniform signaling baseline. In Fig. 11, we plot MI versus transmission reach at optimal input power. At the net bit rate of 400 Gb/s, it is observed that ESS at $N = 200$ outperforms CCDM at $N = 3600$, CCDM at $N = 200$, and uniform signaling in terms of the transmission reach. The gains are about 100 km, 200 km and 450 km, respectively. This gain, in comparison with CCDM at $N = 200$, is due to the linear shaping gain at short blocklength that ESS offers (see Figs. 8–10). For the uniform signaling case, ESS at $N = 200$ provides a combination of linear shaping gain and nonlinear tolerance. Long blocklength CCDM with $N = 3600$ exhibits lower performance in terms of transmission reach than ESS at $N = 200$ due to the nonlinear penalty. CCDM at $N = 200$ shows lower performance than CCDM at $N = 3600$. This means that the large rate loss that CCDM at $N = 200$ presents in comparison with $N = 3600$, compensates the nonlinear gain provided by the use of short blocklength.

The transmission reach is also plotted as a function of the BER post-LDPC, as given in Fig. 12. Noise loading is used to obtain data in the lower BER regime (indicated by square markers). The performance improvement of ESS at $N = 200$ in terms of MI are confirmed by the BER post-LDPC results.

VII. CONCLUSIONS

We have proposed to use enumerative sphere shaping to increase to capacity of fiber optical communication systems. In the context of 200 Gb/s 64-QAM single-polarization single channel and 400 Gb/s 64-QAM dual-polarization WDM systems, we have shown that ESS improves the performance for short blocklengths in comparison with CCDM, due to its low rate loss. Short blocklengths ESS was shown to provide the best performance in terms of mutual information and transmission reach. Short blocklengths ESS also exhibit better effective SNR performance than uniform signaling and long blocklengths ESS and CCDM. It combines both linear shaping

gain and nonlinear tolerance, and has lower complexity than long blocklengths ESS and CCDM, which make it a promising candidate to be implemented in real-time systems. An experimental validation of ESS shaping and an investigation of the effective SNR improvement allowed by using short blocklength shaping is left for further investigation.

VIII. Acknowledgements

References

[1] P. Bayvel, R. Maher, T. Xu, G. Liga, N. A. Shevchenko, D. Lavery, A. Alvarado, and R. I. Killey, "Maximizing the optical network capacity," *Philosoph. Trans. Roy. Soc.*, vol. 374, no. 2062, Jan. 2016, Art. no. 20140440.

[2] R. Dar and P. J. Winzer, "Nonlinear interference mitigation: Methods and potential gain," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 903–930, Feb. 2017.

[3] A. Amari, O. A. Dobre, R. Venkatesan, O. S. S. Kumar, P. Ciblat, Y. Jaouën "A survey on fiber nonlinearity compensation for 400 Gb/s and beyond optical communication systems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 3097–3113, 4th Quart., 2017.

[4] L. Szczecinski and A. Alvarado "*Bit-Interleaved Coded Modulation: Fundamentals, Analysis and Design*," John Wiley & Sons, Chichester, UK, 2015.

[5] A. Alvarado and E. Agrell, "Four-Dimensional Coded Modulation with Bit-Wise Decoders for Future Optical Communications," *J. Lightw. Technol.*, vol. 33, no. 10, pp. 1993–2003, May 2015.

[6] A.R. Calderbank and L.H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 726–740, Jul. 1990.

[7] G. David Forney Jr, "Trellis Shaping," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 281–300, Mar. 1992.

[8] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 913–929, May 1993.

[9] K. Kojima, T. Yoshida, T. Koike-Akino, D. S. Millar, K. Parsons, M. Pajovic, and V. Arlunno, "Nonlinearity-tolerant four-dimensional 2A8PSK family for 5-7 bits/symbol spectral efficiency," *J. Lightw. Technol.*, vol. 35, no. 8, pp. 1383–1391, Apr. 2017.

[10] F. Jardel, T. A. Eriksson, C. Masson, A. Ghazisaeidi, F. Buchali, W. Idler, and J. J. Boutros, "Exploring and experimenting with shaping designs for next-generation optical communications," *J. Lightw. Technol.*, vol. 36, no. 22, pp. 5298–5308, Nov. 2018.

[11] B. Chen, C. Okonkwo, H. Hafermann, and A. Alvarado," Increasing achievable information rates via geometric shaping", in Proc. ECOC, Rome, Italy, Nov. 2018.

[12] T. H. Lotz, X. Liu, S. Chandrasekhar, P. J. Winzer, H. Haunstein, S. Randel, S. Corteselli, B. Zhu, and D. W. Peckham, "Coded PDM-OFDM transmission with shaped 256-iterative-polar-modulation achieving 11.15-b/s/Hz intrachannel spectral efficiency and 800-km reach," *J. Lightw. Technol.*, vol. 31, no. 4, pp. 538–545, Feb. 2013.

[13] T. Liu and I. B. Djordjevic, "Multidimensional optimal signal constellation sets and symbol mappings for block-interleaved coded-modulation enabling ultrahigh-speed optical transport," *IEEE Photon. J.*, vol. 6, no. 4, Aug. 2014.

[14] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.

[15] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: an experimental demonstration," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.

[16] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel," *J. Lightw. Technol.*, vol. 34, no. 21, pp. 5063–5073, Nov. 2016.

[17] J. Renner, T. Fehenberger, M. P. Yankov, F. Da Ros, S. Forchhammer, G. Böcherer, and N. Haniku, "Experimental comparison of probabilistic shaping methods for unrepeated fiber transmission," *J. Lightw. Technol.*, vol. 35, no. 22, pp. 4871–4879, Nov. 2017.

[18] C. A. S. Diniz, C. J. Helio, A. L. N. Souza, T. C. Lima, R. R. Lopes, S. M. Rossi, A. M. Garrich, J. D. Reis, D. S. Arantes, J. R. F. Oliveira, and D. A. A. Mello, "Network cost savings enabled by probabilistic shaping in DP-16QAM 200-Gb/s systems," In proc OFC, Anaheim, CA, USA, Mar. 2016, Paper Tu3F.7.

[19] G. Bocherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651-4665, Dec. 2015.

[20] M. P. Yankov, F. Da Ros, E. P. da Silva, S. Forchhammer, K. J. Larsen, L. K. Oxenlwe, M. Galili, and D. Zibar, "Constellation shaping for WDM systems using 256QAM/1024QAM with probabilistic optimization," *J. Lightw. Technol.*, vol. 34, no. 22, pp. 5146–5156, Nov. 2016.

[21] R. F. H. Fischer,"*Precoding and signal shaping for digital transmission*," J. Wiley-Interscience, 2002.

[22] O. Geller, R. Dar, M. Feder, and M. Shtaif, "A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems," *J. Lightw. Technol.*, vol. 34, no. 16, pp. 3884–3889, Aug. 2016.

[23] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Multiset-partition distribution matching," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1885–1893, Mar. 2019.

[24] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel," in Proc. IEEE Int. Symp. Inform. Theory (ISIT), Honolulu, HI, USA, Aug. 2014, pp. 2794-2798.

[25] F. Steiner and G. Böcherer, "Comparison of geometric and probabilistic shaping with application to ATSC 3.0", arXiv:1608.00474 [cs.IT] (2017).

[26] T. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory*, vol. 19, no. 1, pp. 73–77, Jan. 1973.

[27] F. M. J. Willems, and J. Wuijts, "A pragmatic approach to shaped coded modulation,"in IEEE 1st Symp. on Commun. and Veh. Technol. in the Benelux, 1993.

[28] R. Laroia, N. Farvardin, and S. A. Tretter,"On optimal shaping of multidimensional constellations," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1044–1056, Jul. 1994.

[29] Y. Can Gültekin, W. van Houtum, and F. M. J. Willems, "On constellation shaping for for short blocklengths," in Proc. 2018 Symposium on Information Theory and Signal Processing in the Benelux, Enschede, the Netherlands, Jun. 2018.

[30] Y. Can Gültekin, W. van Houtum, S. Serbetli, and F. M. J. Willems, "Constellation shaping for IEEE 802.11," in 2017 IEEE 28th Int. Symp. on Personal, Indoor, and Mobile Radio Commun. (PIMRC), Montreal, QC, Canada, Oct. 2017.

[31] Y. Can Gültekin, F. M. J. Willems, W. van Houtum, and S. Serbetli, "Approximate enumerative sphere shaping," in Proc. IEEE Int. Symp. Inform. Theory (ISIT), Vail, CO, USA, Jun. 2018, pp. 676-680, Jun. 2018.

[32] Y. C. Gültekin, W. J. van Houtum, A. Koppelaar, F. M. J. Willems, "Enumerative Sphere Shaping for Wireless Communications with Short Packets", arXiv:1903.10244 [eess.SP] (2019).

[33] Y. C. Gültekin, W. J. van Houtum, A. Koppelaar, F. M. J. Willems, "Partial Enumerative Sphere Shaping", arXiv:1904.04528 [eess.SP] (2019).

[34] T. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory*, vol. 19, no. 1, pp. 73–77, Jan. 1973.

[35] F. Gray, "Pulse code communications," U. S. Patent 2 632 058, Mar. 1953.

[36] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.

[37] A. Alvarado, F. Brännström, E. Agrell, and T. Koch, "High-SNR asymptotics of mutual information for discrete constellations with applications to BICM," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1061–1076, Feb. 2014.

[38] F. Steiner, P. Schulte and G. Bocherer, "Approaching waterfilling capacity of parallel channels by higher order modulation and probabilistic amplitude shaping," 2018 52nd Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, 2018, pp. 1-6.