

# Huffman-Coded Sphere Shaping for Extended-Reach Single-Span Links

Pavel Skvortcov, *Student Member, IEEE*, Ian Phillips, *Member, IEEE*, Wladek Forysiak, *Member, IEEE*, Toshiaki Koike-Akino, *Senior Member, IEEE*, Keisuke Kojima, *Senior Member, IEEE*, Kieran Parsons, *Senior Member, IEEE*, and David S. Millar, *Member, IEEE*

**Abstract**—Huffman-coded sphere shaping (HCSS) is an algorithm for finite-length probabilistic constellation shaping, which provides nearly optimal energy efficiency at low implementation complexity. In this paper, we experimentally study the nonlinear performance of HCSS employing dual-polarization 64-ary quadrature amplitude modulation (DP-64QAM) in an extended-reach single-span link comprising 200 km of standard single-mode fiber (SSMF). We investigate the effects of shaping sequence length, dimensionality of symbol mapping, and shaping rate. We determine that the naïve approach of Maxwell–Boltzmann distribution matching — which is optimal in the additive white Gaussian noise channel — provides a maximum achievable information rate (AIR) gain of 0.18 bits/4D-symbol with respect to uniform signaling at optimum launch power in the infinite length regime. Conversely, HCSS can achieve a gain of 0.37 bits/4D-symbol over uniform signaling using amplitude sequence length of 32, which may be implemented without multiplications, using integer comparison and addition operations only. Coded system performance, with a net data rate of approximately 425 Gb/s for both shaped and uniform inputs, is also analyzed.

**Index Terms**—Optical fiber communication, probabilistic shaping, sphere shaping, nonlinear fiber channel, single-span links.

## I. INTRODUCTION

IN the last decade, coherent detection has been a key enabling technology for high-throughput optical fiber communication systems. Initially, the adoption of coherent detection with high-speed digital signal processing (DSP) provided immediate four-fold improvement in spectral efficiency compared to direct detection systems, as all four dimensions of the optical field could be detected. The first coherent 40 Gb/s transmission systems using dual polarization (DP) quadrature phase-shift keying (QPSK) [1] were therefore able to operate without significant increase in bandwidth compared with 10 Gb/s on-off keying systems. DP-QPSK 100 Gb/s systems were subsequently introduced [2], and are widely deployed in commercial long-haul networks. In the presence of increasingly stringent bandwidth constraints, Nyquist pulse shaping and  $M$ -ary quadrature amplitude modulation (MQAM) have been employed in the next generations of coherent systems to increase spectral efficiency, and therefore achieve the required growth in per-wavelength bit rates [3].

P. Skvortcov, I. Phillips, W. Forysiak are with Aston University, Birmingham, B4 7ET, UK. E-mails: {skvortcp, i.phillips, w.forysiak}@aston.ac.uk.

D. S. Millar, T. Koike-Akino, K. Kojima and K. Parsons are with Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA. E-mails: {millar, koike, kojima, parsons}@merl.com.

Due to the reduction in noise tolerance of high spectral efficiency transmission systems, coded modulation (CM) schemes have been proposed to maximize the transmission system performance given the constraints of the physical channel.

By considering the optical field as a 4D signal space, power efficiency may be improved in a number of ways (e.g., by considering constellation points on an optimal lattice bound by a maximum power) [4], [5]. High-dimensional modulation (HDM) based on utilization of multiple time-slots (consequent transmitted symbols) or multiple carriers to utilize more efficient lattices in larger numbers of dimensions can provide further improvement in power efficiency [6], [7]. However, coded performance using binary forward error correction (FEC) codes and bit-interleaved coded modulation (BICM) implementation for such systems may be challenging due to their lack of Gray-coded bit labeling. Additionally, while much research in this area has focused on power efficiency, several proposed HDM techniques specifically target improved nonlinear tolerance in the optical fiber channel [8]–[10].

Constellation shaping may be considered as the optimization of transmitted symbols distribution in terms of either location or probability in the signal space, such that the transmitted signal has improved power efficiency or nonlinear tolerance for an optical channel. In the additive white Gaussian noise (AWGN) channel, constellation shaping provides a gain of up to 1.53 dB in power efficiency over uniform signaling [11].

Finite-length probabilistic shaping may be considered as HDM — i.e., the mapping of a block of  $k$  input bits to a point on an  $L$ -dimensional constellation (e.g., on a square lattice), where an  $L$ -dimensional constellation point is then mapped to a sequence of 4D symbols in time (e.g., DP-MQAM) [12].

The key enabler for probabilistic shaping in practical systems was the introduction of a probabilistic amplitude shaping (PAS) framework for QAM [13]. The primary advantage of PAS over other proposed probabilistic shaping architectures is the use of reverse concatenation of the shaping and FEC coding stages (coding is performed on the bit labels of probabilistically shaped amplitudes), which allows for low-complexity integration with BICM systems and enables transmission rate adaptivity with fixed FEC code rate.

Since the proposal of PAS, many implementations utilizing this structure have been investigated. Initially, an analysis based on an asymptotic infinite-length shaping approach with an ideal Maxwell–Boltzmann (MB) distribution of amplitudes was commonly performed for transmission over a nonlinear optical channel. MB shaping achieves the maximum power ef-

efficiency for a given entropy, which is therefore the optimal distribution in the AWGN channel [14]. However, for an optical channel, MB shaping can enhance fiber Kerr nonlinearities, leading to a decrease in shaping gain, which was shown for long-haul transmission [15] and unrepeatable transmission [16]. Importantly, while considerable performance improvement can be achieved using infinite-length MB shaping, this approach is fundamentally impossible to implement.

More recently, research has focused on finite-length shaping architectures which can be realizable in the hardware. One class, referred to as distribution matching (DM), is based on obtaining a fixed target distribution. First, constant-composition DM (CCDM) was introduced [17], whereby all transmitted amplitude sequences are permutations of a single composition, defined by the target distribution. While CCDM can provide asymptotically low rate-loss, and therefore high power efficiency, it requires long sequences (typically on the order of several hundred amplitudes) to achieve it. Multiset-partition DM (MPDM) was proposed in [18], [19] and provides lower rate loss at a fixed shaping sequence length. MPDM is based on multiple complimentary compositions, which on average result in the desired distribution. Other DM implementations were also proposed, such as product DM [20], hierarchical DM [21], prefix-free code DM [17], and parallel-amplitude DM [22].

Another class of shaping architectures is based on sphere shaping. Sphere shaping introduces an optimal sphere bound in multi-dimensional signal space — for a target transmission rate and finite sequence length, the optimal set of constellation points from a fixed lattice is chosen such that the geometry is bounded by a hyper-sphere [12], [14]. This scheme has intuitively optimal power efficiency — for any rate, lattice and sequence length, we can define a smallest possible sphere which contains the correct number of points: any alternative set of constellation points will contain points outside the sphere, leading to degraded power efficiency. Sphere shaping-based algorithms include shell mapping (SM) [23], [24], enumerative sphere shaping (ESS) [25]–[27], and Huffman-coded sphere shaping (HCSS) [28], [29].

The dependence of nonlinear tolerance on shaping length was investigated using constant composition distribution matching (CCDM) [30], [31] and ESS [27], [32] for multi-span long-haul links. Also, the advantage of 2D symbol mapping for nonlinearity tolerance has been mentioned previously in [31], while the advantage of the short-length shaping regime for a nonlinear optical channel was theoretically investigated in [33], [34]. Significant shaping gain exceeding the theoretical gain for the AWGN channel was demonstrated in numerical simulations for single-span links by optimally combining linear and nonlinear shaping gains using SM with very short shaping lengths [34]. In [35], [36] the authors investigated shaping of a single 4D quadrant using SM and demonstrated increased nonlinear tolerance in single-span links. In our previous work, we demonstrated a significant nonlinear shaping gain with short-length HCSS and 4D symbol mapping for extended-reach single-span links [37].

Nonlinearity tolerant shaping can be of particular interest for short-distance transmission with low accumulated chro-

matic dispersion (CD). Low CD leads to highly correlated short-memory nonlinear interactions. Hence, for short-distance systems improved nonlinear tolerance can be achieved with short-length shaping [34], which is attractive in terms of implementation complexity [37]. In contrast, for long-haul systems nonlinearities become significantly decorrelated (turning into Gaussian-like noise) and longer length shaping is required, as for AWGN channels [27].

At present, demand for high capacity single-span transmission systems is driven by cloud and inter-data-center traffic. After successful introduction of the 400G ZR standard [38], which specifies transmission of up to 120 km, there is great interest in increasing system reach and bit rate. Increasing the transmission distance of single-span systems leads to the signal being significantly impaired by fiber nonlinearities due to increased optimal launch power. We previously demonstrated that for these systems short-length shaping can offer significant performance gain at reasonably low implementation complexity [37].

We extend our previous study on nonlinear performance of HCSS in extended-reach single-span links presented in [37], where we studied the short-length shaping regime of HCSS with fixed shaping rate and symbol mapping dimensionality. In this work we present a comprehensive analysis of HCSS performance in extended-reach single-span links comprising investigations of the impact of shaping length, dimensionality of symbol mapping, and shaping rate on nonlinear tolerance in comparison with uniform signaling and infinite-length shaping with an ideal MB distribution. Also, coded performance is analyzed using low-density parity-check (LDPC) codes, which provide matching data rates for shaped and uniform signals.

The remainder of this paper is structured as follows. In Section II we give an overview of the PAS framework, HCSS architecture, multi-dimensional symbol mapping strategies, associated energy efficiency and rate loss of the shaping/mapping scheme. Section III provides the description of the transmission experiment, DSP and performance metrics used for evaluating system performance and comparing shaping methods. In Section IV we present the experimental results for back-to-back and single-span transmission, analyze coded performance and discuss the nonlinear system performance, while in Section V we draw our conclusions. Appendices A and B give detailed explanations of probability mass function (PMF) calculation for high-dimensional symbol mapping strategies, and experimental data fitting used for the analysis of the results, respectively.

## II. PROBABILISTIC SHAPING: HUFFMAN-CODED SPHERE SHAPING (HCSS)

### A. Probabilistic amplitude shaping (PAS)

The diagram of PAS architecture at the transmitter and receiver is shown in Fig. 1. The PAS architecture calls for blocks of uncoded information bits to be mapped onto probabilistically shaped amplitude sequences [13]. The bit labels corresponding to the shaped amplitudes are then encoded with a systematic FEC code, and the parity bits assigned to the sign bits of the pulse amplitude modulation (PAM) constellation.

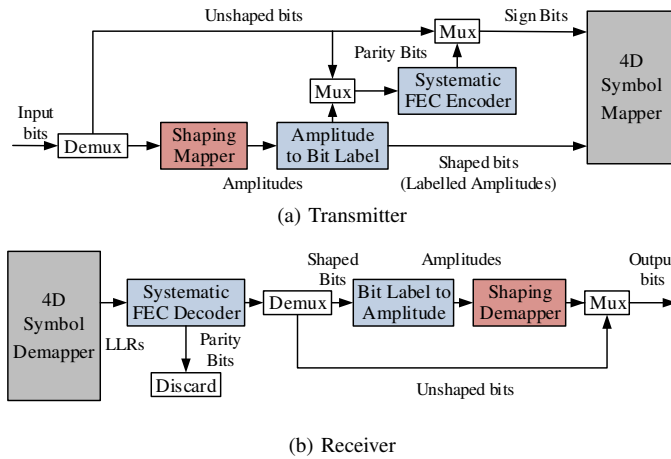


Fig. 1. PAS architecture: (a) transmitter, (b) receiver.

Some uncoded information bits (referred to as unshaped bits) may also be carried on the sign of the constellation if required. We note that parity and information bits are considered to be uniformly distributed, and therefore the resulting encoded distribution is symmetrical about zero. Finally, signed amplitudes are mapped onto the 4D optical carrier for transmission.

The primary advantage of the PAS architecture is that at the receiver the FEC decoder operates on the bit labels of the shaped amplitude sequences, enabling FEC decoding to be performed before shaping demapping. The demapping of amplitude sequences onto blocks of information bits is then performed without noise or bit errors, greatly reducing the complexity of both the demapping procedure and the shaping system design. Also, we note that the overall transmission rate of the system may be tuned by adjusting the rate of the shaping algorithms, while the FEC code rate remains the same.

We define the shaping rate in bits per amplitude (b/Amp) as

$$R_S = \frac{k}{L}, \quad (1)$$

where  $k$  is the number of uniform input bits, and  $L$  is the length of the shaped amplitude sequence.

The set of unique amplitude sequences, which are available for mapping/demapping blocks of information bits and forms a high-dimensional constellation, is defined by the chosen shaping architecture. We note that the shaping architecture determines the energy efficiency and rate loss of the system, while the mapping and demapping algorithms determine the implementation complexity.

### B. Huffman-coded sphere shaping (HCSS)

Sphere shaping architecture introduces a sphere bound on power efficiency for finite-length shaping — i.e., it defines all constellation points for signaling on a specified multi-dimensional lattice having a certain energy constraint. By definition, this architecture achieves the best possible energy efficiency for a given rate (i.e., number of constellation points), lattice (e.g., the square lattice modulation based on uniform QAM), and dimension (i.e., amplitude sequence length) [14].

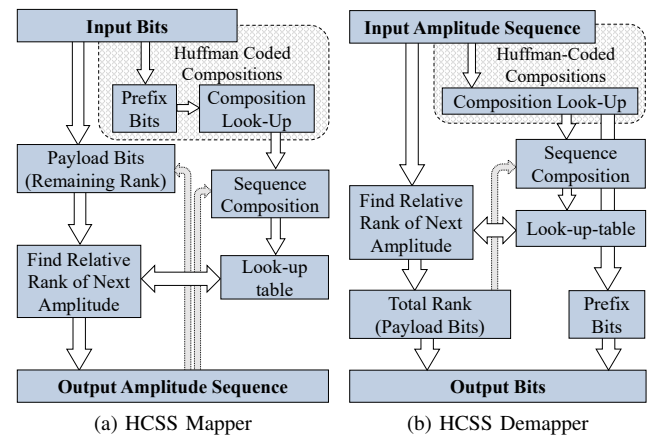


Fig. 2. Block diagram of HCSS algorithm using multiset ranking and LUT [37].

Examples of algorithms which closely approximate the sphere bound are ESS and SM.

HCSS introduces some additional structure into the spherical shaping architecture in exchange for a small reduction in power efficiency and increase in rate loss. In particular, HCSS restricts the number of constellation points utilized for each unique composition within spherical architecture to be a power of two, and then introduces a minimal number of additional compositions with higher power to ensure a dyadic distribution of compositions [28]. This enables the use of a variable length binary prefix (Huffman code) to uniquely address compositions in the shaping architecture. The remaining payload bits in a binary input word are then used to address a unique permutation (i.e., amplitude sequence) of the specified composition. Therefore, the mapping/demapping task is divided into two parts — addressing composition using a Huffman code and a CCDM mapping/demapping problem.

If the multiset ranking (MR) algorithm is used for CCDM mapping/demapping as described in [28], [29], the lexicographical rank of the selected sequence corresponds to the payload bits. Additionally, we note that sequence ranks may be computed without multiplication operations by pre-computing multinomial coefficients and storing them in a look-up table (LUT). For short sequence lengths, the coefficients required for MR mapping and demapping can be stored in moderately sized LUTs — for example, sequence length of 32 requires only 100 kbits of memory [29]. Therefore, both mapping in Fig. 2(a) and demapping in Fig. 2(b) are performed iteratively on a per-symbol basis, using LUTs, requiring integer comparison and addition operations only. We note that while LUT size and algebraic complexity estimation may provide a basic idea of the feasibility of hardware implementation, quantitative comparison of the hardware complexity of different algorithms must be done in dimensions such as power consumption and circuit footprint, and requires a process-specific netlist-level hardware design [39].

The number of available sequences for the  $i^{\text{th}}$  composition in the Huffman-coded structure is

$$N_{\text{seq}}^i = 2^{\lfloor \log_2 N_{\text{perm}}^i \rfloor}, \quad (2)$$

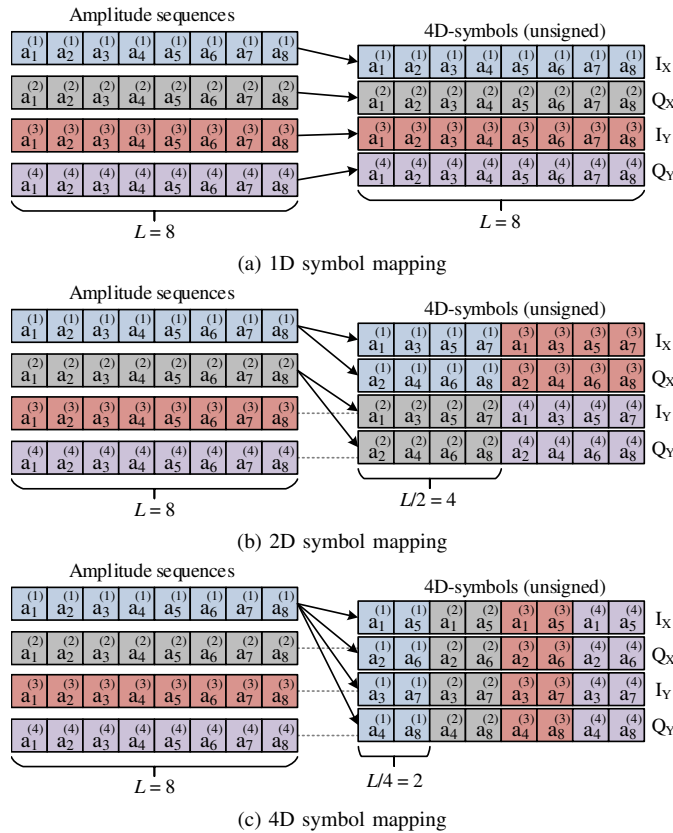


Fig. 3. Strategies for mapping amplitude sequences into modulated 4D symbols (corresponding signs are not reflected): (a) 1D symbol mapping, (b) 2D symbol mapping, (c) 4D symbol mapping.

where  $N_{\text{perm}}^i$  is the number of possible permutations of the  $i^{\text{th}}$  composition, which is given by multinomial coefficient:

$$N_{\text{perm}}^i = \frac{L!}{\prod_a c^i(a)!}, \quad (3)$$

where  $c^i(a)$  is the number of occurrences in the  $i^{\text{th}}$  composition of the amplitude  $a$  for  $L = \sum_a c^i(a)$ . The probability of the  $i^{\text{th}}$  composition is therefore given by:

$$p_i = \frac{N_{\text{seq}}^i}{2^k}. \quad (4)$$

### C. Symbol mapping strategies

We studied three strategies for mapping shaped sequences of amplitudes into the modulated 4D symbols of the DP-64QAM format. These mapping strategies are illustrated in Fig. 3. We note that a sign is assigned to each amplitude according to the sign bit during the mapping process (see Fig. 1 (a)), however, added signs are not reflected in Fig. 3 for simplicity of consideration.

We refer to 1D symbol mapping when four independent shaped amplitude sequences of length  $L$  are sequentially (on an amplitude-by-amplitude basis) mapped into four simultaneous quadratures (in-phase and quadrature signal components in both polarizations) constructing a single 4D-symbol sequence of length  $L$ , as shown in Fig. 3 (a).

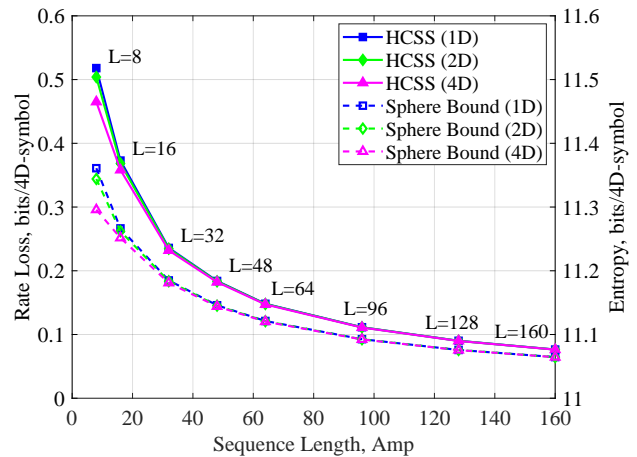


Fig. 4. Rate loss and entropy vs. shaping sequence length ( $R_S = 1.75$ ) for 1D, 2D and 4D symbol mapping strategies.

In the case of 2D symbol mapping, as shown in Fig. 3 (b), two independent shaped amplitude sequences of length  $L$  are mapped into a single 4D-symbol sequence of length  $L/2$  — two consecutive amplitudes from each amplitude sequence are mapped into four simultaneous quadratures.

In the case of 4D symbol mapping, as shown in Fig. 3 (c), a single shaped amplitude sequence of length  $L$  is mapped in a single 4D-symbol sequence of length  $L/4$  — four consecutive amplitudes are mapped into four simultaneous quadratures.

By increasing symbol mapping dimensionality we can effectively reduce the time-domain length of the output 4D-symbol sequence and the number of simultaneously interacting sequences, while maintaining the same power efficiency. The derivation of the resulting multi-dimensional PMF for different symbol mapping strategies is described in detail in Appendix A. We note that 1D, 2D and 4D mapping strategies result in 1D, 2D and 4D distributions, respectively.

### D. Rate loss and power efficiency

Rate loss due to the use of a finite-length shaping scheme is calculated in bits per 4D-symbol (b/4D) as

$$R_{\text{loss}} = H(X) - D \cdot (R_S + 1), \quad (5)$$

where  $H(X)$  denotes the entropy of the 4D output signal  $X$  (calculated according to PMF in Appendix A),  $D$  accounts for symbol mapping dimensionality (in the case of 4D symbol mapping  $D = 4$ ) and “1” accounts for the sign bit per dimension.

Fig. 4 demonstrates the rate loss and corresponding entropy for HCSS in comparison with the sphere bound on fixed length shaping, in the case of 1D, 2D and 4D mapping based on DP-64QAM at  $R_S = 1.75$ . HCSS has slightly higher rate loss compared to the sphere bound due to additional constraints introduced in its structure. This difference becomes negligible at long shaping lengths. Also, we note that higher dimensional mapping demonstrates slightly reduced rate loss at short shaping lengths.

Fig. 5 shows the power penalty for HCSS, the sphere bound and infinite-length MB shaping (also, based on DP-64QAM)

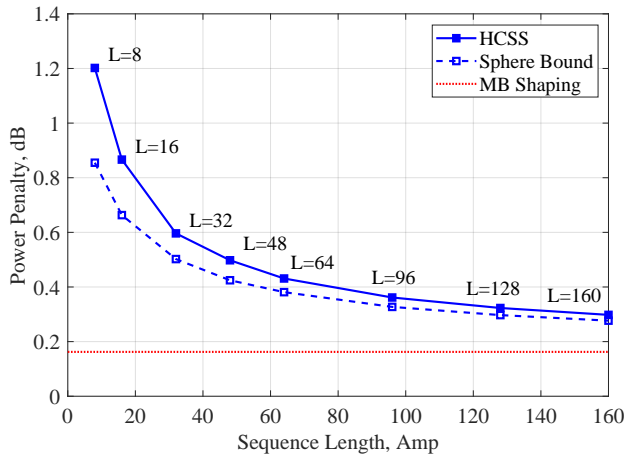


Fig. 5. Power penalty vs. shaping sequence length ( $R_S = 1.75$ ). Reference level is the MB shaping with unconstrained cardinality.

at  $R_S = 1.75$ . The power penalty is calculated with respect to the MB distribution with unconstrained cardinality, which provides the maximum power efficiency for a given entropy [14]. HCSS and sphere shaping demonstrate the same trend as for the rate loss — the difference in power penalty becomes negligible at long shaping lengths. We note that power efficiency does not depend on symbol mapping dimensionality.

### E. AWGN performance

The performance of HCSS in terms of AIR in comparison with the sphere bound and uniform signaling in simulated AWGN channel is shown in Fig. 6 as a function of signal-to-noise ratio (SNR). Similarly to the rate loss and power efficiency analysis, both for HCSS and sphere shaping the shaping was based on DP-64QAM at  $R_S = 1.75$  (the dimensionality of symbol mapping does not have any impact as for power efficiency) with shaping lengths of  $L = 32, 160$ , and the uniform case is also represented by DP-64QAM. The chosen shaping rate can be considered near-optimal for the operational SNR range of approximately 11–13 dB, since the shaping gain is maximized. In this SNR range (see inset in Fig. 6), for  $L = 32$  HCSS provides the sensitivity improvement in terms of SNR of 0.4 dB, while sphere shaping achieves an additional 0.1 dB — equivalently, AIR gains are 0.25 b/4D for HCSS and an additional 0.05 b/4D for sphere shaping. For  $L = 160$ , the difference between HCSS and sphere shaping becomes negligible and the sensitivity is improved by 1 dB, or AIR gain is 0.45 b/4D. We note that other previously mentioned sphere shaping-based algorithms, i.e., ESS and SM, closely approach the sphere bound on performance, since they differ from the sphere bound in only a small number of constellation points.

## III. TRANSMISSION EXPERIMENT

We investigated the system performance using HCSS in comparison with uniform signaling and infinite-length MB shaping. For HCSS we varied the shaping sequence length ( $L$  in the range of 8–160), dimensionality of amplitude-to-symbol

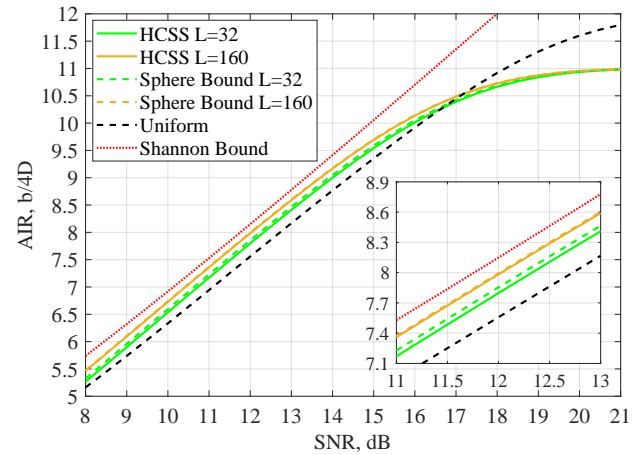


Fig. 6. Performance in AWGN channel: AIR vs. SNR. For shaping schemes  $R_S = 1.75$ .

mapping (1D, 2D and 4D) and shaping rate ( $R_S$  in the range of 1.625–1.875 b/Amp).

For the MB shaping case, signals were drawn from an MB distribution with the entropy matching the shaping rate of HCSS [15]. We note that symbols in the transmitted signal were drawn independently and identically on the underlying PMF. This method may be considered to give a finite-length sample of an infinite-length shaped sequence, which incurs no rate loss.

### A. Experimental setup

The experimental setup is shown in Fig. 7. Wavelength-division multiplexed (WDM) transmission of 9 DP-64QAM channels operating at 56 GBd (with root-raised cosine pulse shaping with 10% roll-off factor) on a 62.5 GHz grid was carried out over a 200 km single-span link of SSMF. All channels used the same shaping scheme under investigation.

The central channel-under-test (CUT) was generated using a 92 GSa/s digital-to-analog converter (DAC) with nominal effective number of bits (ENOB) of 6 and bandwidth of more than 25 GHz [40], followed by a pluggable CFP2 analog coherent optics (ACO) transceiver [41], [42] (integrated laser with nominal 100 kHz linewidth and bandwidth exceeding 25 GHz). The 8 interfering channels were divided into two groups of 4 channels, and generated by two pairs of DP in-phase/quadrature-phase (IQ) Mach-Zehnder modulators and DACs (for one group of channels a 120 GSa/s DAC with nominal ENOB of 5.5 and bandwidth exceeding 30 GHz, and modulator with bandwidth exceeding 35 GHz were used; for the other group of channels a 64 GSa/s DAC with nominal ENOB of 5.5 and bandwidth exceeding 20 GHz, and modulator with bandwidth exceeding 25 GHz were used), where 4 external cavity lasers with 100 kHz linewidth combined with a polarization maintaining coupler were used for each modulator/DAC pair. The two groups of interfering channels were then pre-amplified by erbium-doped fiber amplifiers (EDFAs), and spectrally interleaved and combined with the CUT via a programmable wavelength selective switch (WSS).

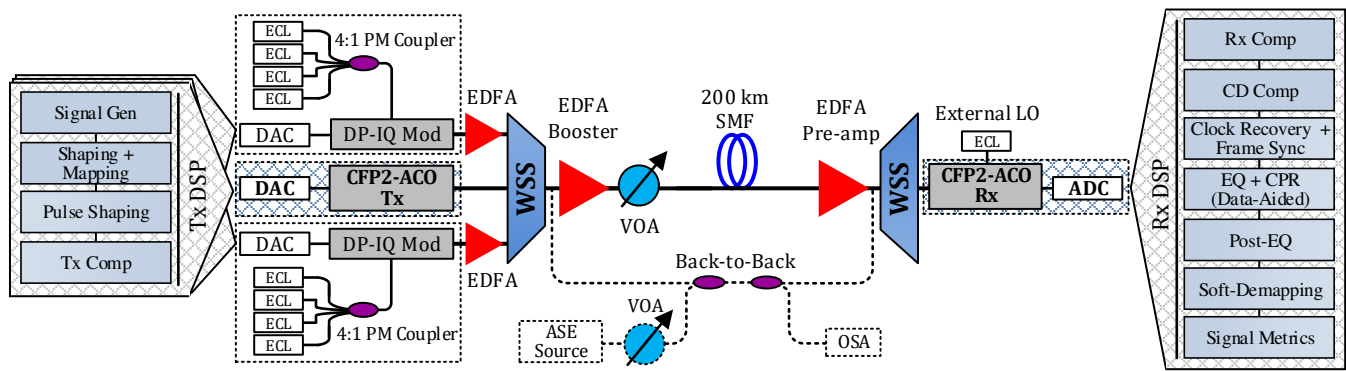


Fig. 7. Block diagram of transmission experimental setup. Block diagram of offline DSP for transmitter and receiver.

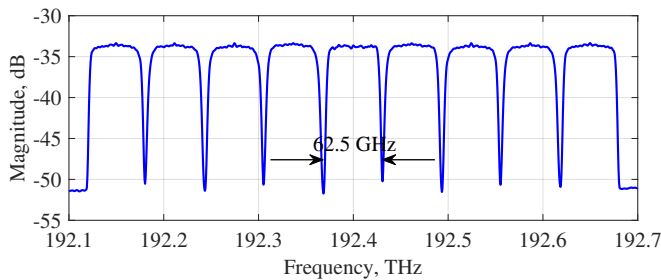


Fig. 8. Signal spectrum (measured with 2 GHz resolution). 9 channels (root-raised cosine pulse shaping with 10% roll-off factor) at 56 Gbd on a 62.5 GHz grid.

The resulting spectrum of the generated WDM signal is shown in Fig. 8.

At the link input the WDM signal was amplified by a booster EDFA, then a variable optical attenuator (VOA) was used to control the total launched power. After transmission over 200 km of SSMF, the WDM signal was pre-amplified by an EDFA with a noise figure of 5.5 dB and the CUT was filtered by a WSS. Finally, the CUT was received by a CFP2-ACO transceiver (an external laser with 100 kHz linewidth was used as a local oscillator) and its analog output was sampled and digitized by a 92 GSa/s analog-to-digital converter (ADC) with nominal ENOB of 6 and bandwidth of more than 25 GHz.

In back-to-back configuration, the output of the transmitter-side WSS was connected to the receiver-side WSS via a pair of couplers used for noise loading. An amplified spontaneous emission (ASE) noise source and VOA were used for varying optical signal-to-noise ratio (OSNR), while an optical spectrum analyzer (OSA) was used for OSNR measurement.

### B. Digital signal processing (DSP)

DSP was performed offline according to the generic flow in Fig. 7. At the transmitter-side, firstly, signals were randomly generated according to the shaping and mapping scheme under consideration. Then, root-raised cosine pulse shaping was applied followed by transmitter pre-emphasis (to compensate for frequency response and skews). Finally, the signals were uploaded to the DACs.

At the receiver-side the signal recovery was implemented as follows. Firstly, the received data was extracted from the

ADC and receiver pre-compensation was applied (compensation of frequency response, skews and I/Q imbalance). Then, CD was compensated, clock recovery was performed by a frequency-domain Gardner algorithm, conventional complex-valued decision-directed least-mean squares (DD-LMS)  $2 \times 2$  multiple-input multiple-output (MIMO) equalization (35 taps) was performed in conjunction with carrier phase recovery (CPR) in fully data-aided mode, and post-equalization was done using real-valued DD-LMS  $2 \times 2$  MIMO equalizers (5 taps) to enable compensation of residual transmitter IQ impairments [43]. Soft-demapping assumed a circularly symmetric Gaussian channel, and transmission performance metrics were averaged over approximately  $5 \times 10^6$  symbols.

### C. Performance metrics

We denote the transmitted signal by  $X$  and received signal after DSP algorithms by  $Y$ . The transmitted signal  $X$  takes values from the 4D-constellation  $\mathcal{X} = \{x_1, \dots, x_M\}$ . We define the adjusted transmitted signal  $X'$  to take values from  $\mathcal{X}' = \{x'_1, \dots, x'_M\}$ , such that the new constellation points  $x'_i$  represent the centroids of the received symbols  $y$  corresponding to the original constellation points  $x_i$ , which can be expressed as

$$x'_i = E[y|x_i], \quad (6)$$

where  $E[\cdot]$  denotes the expectation.

The adjusted constellation  $\mathcal{X}'$  is subsequently used for estimation of effective SNR and soft-demapping. This reduces the impact of impairments, which represent geometrical distortions of the constellation (e.g., transceiver nonlinearity and uncompensated modulation impairments).

The effective SNR [15] of the received signal is calculated as

$$\text{SNR}_{\text{eff}} = \frac{\text{Var}[X']}{\text{Var}[Y - X']}, \quad (7)$$

where  $\text{Var}[\cdot]$  denotes the variance. Effective SNR accounts for both linear and nonlinear noise contributions accumulated during signal propagation over an optical fiber, as well as transceiver noise floor (i.e., implementation penalty and DSP imperfections).

The soft-demapper calculates log-likelihood ratios (LLRs) for binary labels  $B_i$  ( $i = 1, \dots, m$ ) as

$$\text{LLR}_i = \log \frac{\sum_{x \in \mathcal{X}'_{1,i}} f_{Y|X'}(y|x') P_{X'}(x')}{\sum_{x \in \mathcal{X}'_{0,i}} f_{Y|X'}(y|x') P_{X'}(x')}, \quad (8)$$

where  $\mathcal{X}'_{1,i}$  and  $\mathcal{X}'_{0,i}$  are the subsets of constellation  $\mathcal{X}'$ , which represent  $B_i$  being equal to 1 or 0, respectively;  $f_{Y|X'}(y|x')$  is the transition probability density function of the auxiliary channel used for mismatched decoding;  $P_{X'}(x')$  is the 4D PMF (calculated according to Appendix A).

We consider a memoryless 4D circularly symmetric Gaussian auxiliary channel and assume that the noise in each dimension is independent and identically distributed [44]. In this case, the channel can be described as

$$f_{Y|X'}(y|x') = \frac{1}{(\pi\sigma^2)^2} \exp\left(-\frac{\|y-x'\|^2}{\sigma^2}\right), \quad (9)$$

where  $\sigma^2$  is the noise variance. We note that lower-dimensional soft-demapping can be done analogously for the case of 1D and 2D mapping.

The achievable information rate (AIR) for bit-metric decoding (BMD) impacted by the rate loss associated with finite-length shaping [18] is calculated in b/4D as

$$\text{AIR} = \underbrace{\left[ H(X) - \sum_{i=1}^m H(B_i|Y) \right]}_{R_{\text{BMD}}} - R_{\text{loss}}, \quad (10)$$

where  $R_{\text{BMD}}$  is the BMD rate, which is given by the generalized mutual information (GMI). We note that for both uniform signaling and infinite-length MB shaping, we have  $R_{\text{loss}} = 0$ .

Coded performance was analyzed based on normalized GMI (nGMI), which can be calculated for a uniform signal as

$$\text{nGMI} = \frac{\text{GMI}}{m} = \frac{\text{AIR}}{m}, \quad (11)$$

while for shaped signals it is calculated as

$$\text{nGMI} = 1 - \frac{H(X) - \text{GMI}}{m} = 1 - \frac{4 \cdot (R_S + 1) - \text{AIR}}{m}. \quad (12)$$

#### IV. EXPERIMENTAL RESULTS

##### A. Back-to-back

Fig. 9 shows a back-to-back characterization of the system under consideration. We note that these results are partly reproduced from [37]. We compared uniform signaling, MB shaping at  $R_S = 1.75$  b/Amp, and HCSS with  $L = 16, 32, 48$  using 4D symbol mapping at the same shaping rate. Fig. 9(a) illustrates the variation of AIR with OSNR. We observe that infinite-length MB shaping achieves superior performance with a 0.48 b/4D AIR gain compared with uniform signaling over an operating OSNR range of 18–21 dB. HCSS achieves gains of 0.08, 0.23 and 0.32 b/4D for  $L = 16, 32, 48$ , respectively. This shows the trend that longer length shaping achieves better performance (eventually, approaching that of MB shaping), as is expected for AWGN channels. For higher OSNR values, shaping gain reduces both for HCSS and MB shaping. We note that HCSS performance in back-to-back configuration is

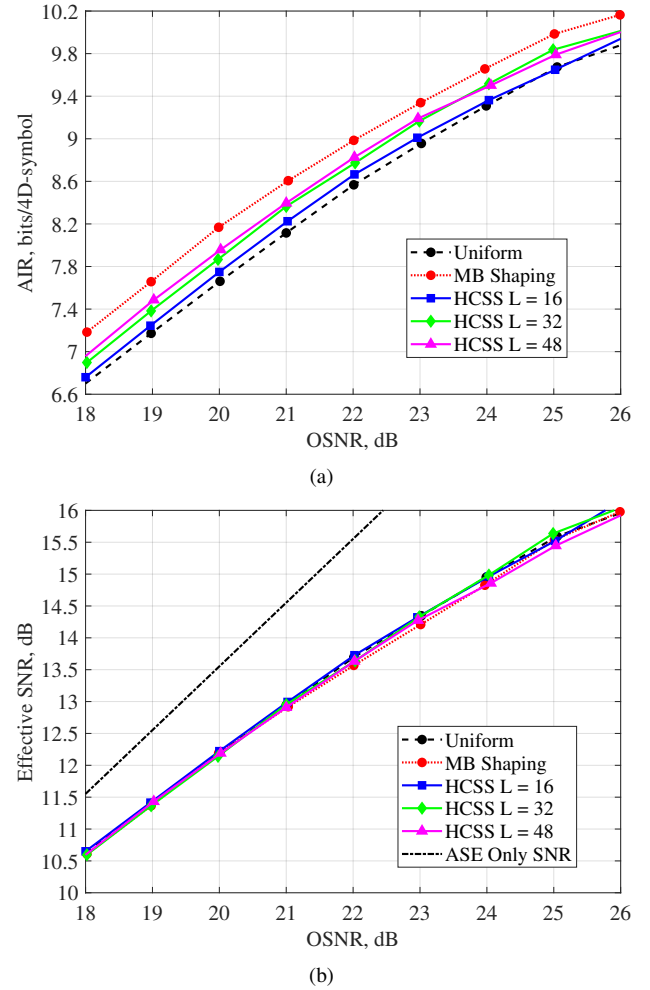


Fig. 9. Performance vs. OSNR in back-to-back configuration ( $R_S = 1.75$ ;  $L = 16, 32, 48$ ; 4D symbol mapping): (a) AIR, (b) Effective SNR.

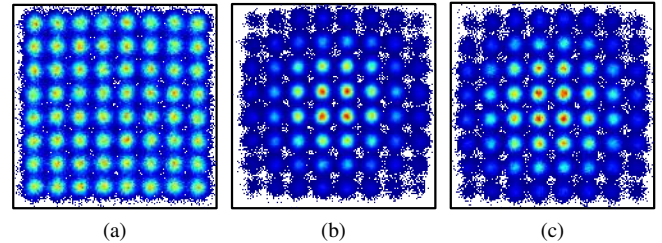
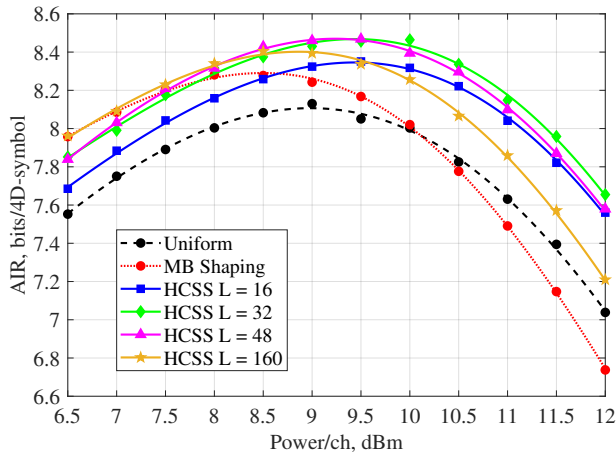


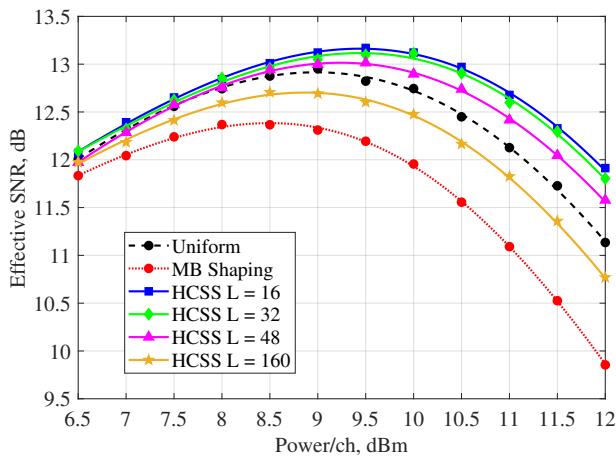
Fig. 10. Experimental constellations in back-to-back under high OSNR: (c) Uniform, (b) Maxwell-Boltzmann shaping, (c) Huffman-coded sphere shaping ( $L = 32$ ).

well matched with AWGN performance shown in Fig. 6 (e.g., for  $L = 32$ , where OSNR range of 18–21 dB corresponds to SNR range of 10.5–13 dB).

Fig. 9 (b) illustrates the variation of effective SNR with OSNR. We observe that there is no noticeable difference in terms of effective SNR over an operating OSNR range of 18–21 dB (SNR range of 10.5–13 dB) for the shaping schemes under consideration. Therefore, we conclude that shaping algorithms do not introduce additional implementation penalty over uniform signaling over the range of SNRs under consideration, hence, the relative performance of different



(a)



(b)

Fig. 11. Performance vs. optical launch power for single-span transmission ( $R_S = 1.75$ ;  $L = 16, 32, 48, 160$ ; 4D symbol mapping): (a) AIR, (b) Effective SNR.

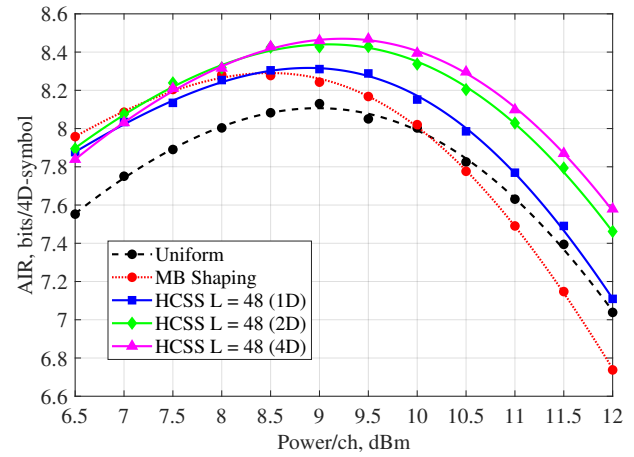
shaping schemes in the nonlinear regime is not impacted by their respective implementation penalties. We also note that shaping sequence length and symbol mapping strategy do not impact the SNR in back-to-back.

Fig. 10 (a)–(c) shows constellation diagrams for uniform signaling, MB shaping and HCSS ( $L = 32$ , 4D amplitude-to-symbol mapping) at  $R_S = 1.75$  b/Amp under high OSNR (about 35 dB). We observe no notable visual difference between HCSS and MB shaped constellations.

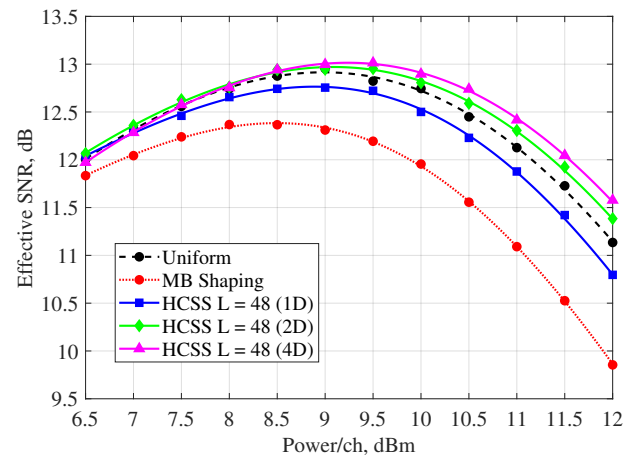
### B. Single-span transmission

For analysis of the experimental data for single-span transmission, data fitting based on a Gaussian noise (GN)-model [45] was performed. A detailed explanation of the data fitting approach can be found in Appendix B.

1) *Optical launch power sweep*: Figs. 11 and 12 demonstrate system performance characterization in terms of AIR and effective SNR as a function of optical launch power for uniform signaling, MB shaping and HCSS at  $R_S = 1.75$  b/Amp. Individual data points represent measured experimental data, while corresponding smooth lines represent



(a)



(b)

Fig. 12. Performance vs. optical launch power for single-span transmission ( $R_S = 1.75$ ;  $L = 48$ ; 1D, 2D and 4D symbol mapping): (a) AIR, (b) Effective SNR.

the data fit. We note that the model described in Appendix B provides a very good fit to the measured experimental data.

Fig. 11 (a) and (b) show AIR and effective SNR for uniform signaling, MB shaping and HCSS using  $L = 16, 32, 48, 160$  with 4D symbol mapping. We note that these results are in good agreement with [37], while also demonstrating asymptomatic performance in the long length shaping regime. In the linear regime, performance is consistent with that of the back-to-back measurements. HCSS with longer sequence length achieves higher AIR (approaching the performance of MB shaping with  $L = 160$ ), while variation in effective SNR is relatively small. At the optimal launch power, we observe that MB shaping achieves a gain over uniform signaling of 0.18 b/4D, while HCSS exhibits gains of 0.24, 0.37, 0.38, 0.30 b/4D for  $L = 16, 32, 48, 160$ , respectively. MB shaping suffers from severe nonlinear impairment, which can be seen as effective SNR degradation of 0.52 dB compared to uniform signaling at optimum power, while HCSS demonstrated improved nonlinearity tolerance. HCSS with  $L = 16, 32, 48$  elicits SNR gains of 0.3, 0.17, 0.06 dB compared to uniform signaling at optimal power, while for  $L = 160$  the effective SNR is reduced by 0.2 dB. In the



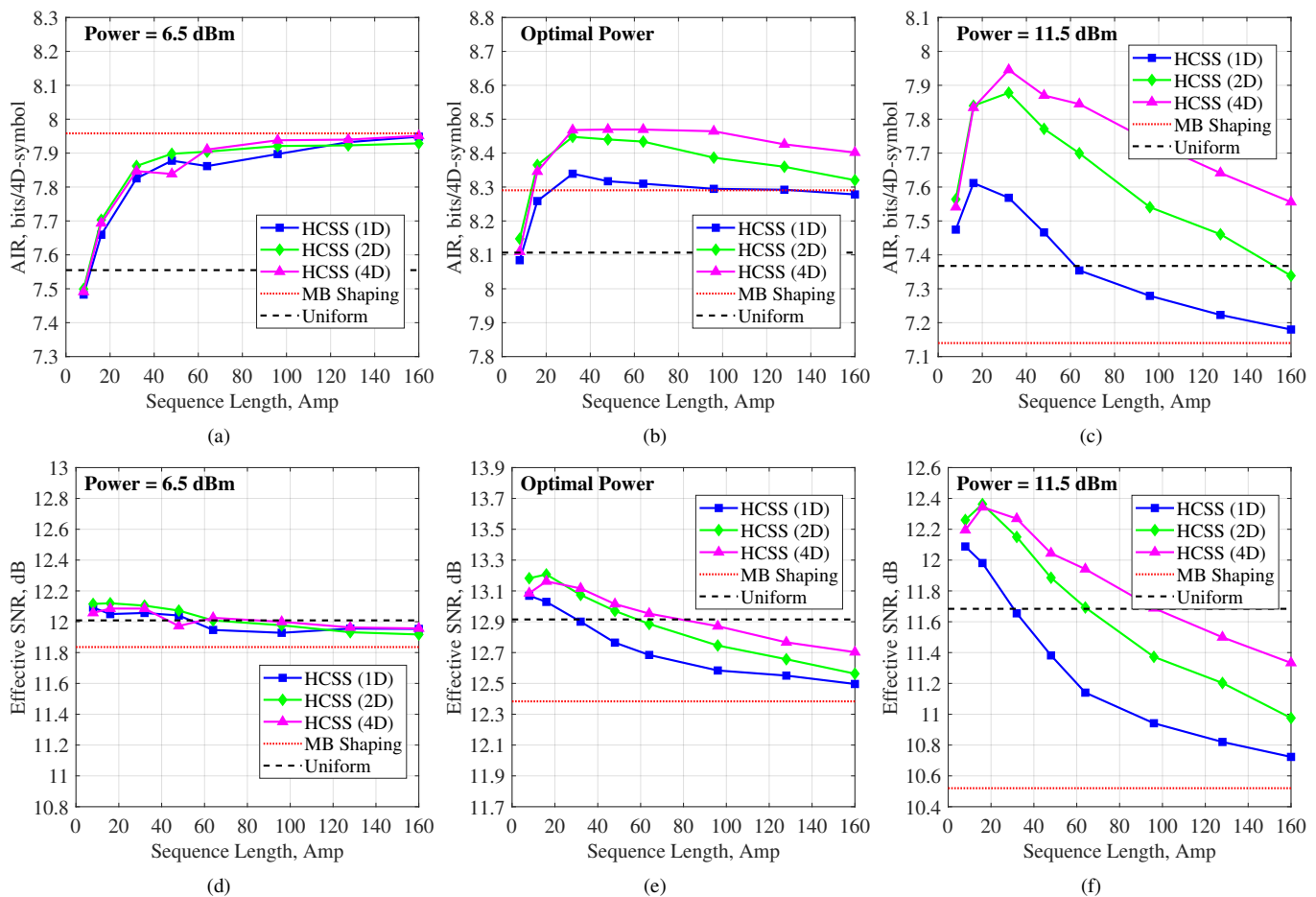


Fig. 13. Performance vs. shaping sequence length ( $R_S = 1.75$ ; 1D, 2D and 4D symbol mapping): (a) and (d) AIR and effective SNR for launch power of 6.5 dBm, (b) and (e) AIR and effective SNR for optimal launch power, (c) and (f) AIR and effective SNR for launch power of 11.5 dBm.

highly nonlinear regime (launch power above  $\sim 11$  dBm), the performance degradation for MB shaping and HCSS with  $L = 160$  is more significant, whereas the gain for HCSS with shorter shaping length is increased.

Fig. 12 (a) and (b) show AIR and effective SNR for uniform signaling, MB shaping and HCSS using  $L = 48$  with 1D, 2D and 4D symbol mapping strategies. We observe that HCSS with higher-dimensional symbol mapping achieves better performance in the nonlinear regime. At the optimal launch power, HCSS with 4D symbol mapping achieves AIR gains of 0.04, 0.15 b/4D with corresponding SNR gain of 0.05, and 0.25 dB compared to 2D and 1D symbol mapping, respectively. In the linear regime, the performance difference is negligible among all symbol mapping strategies.

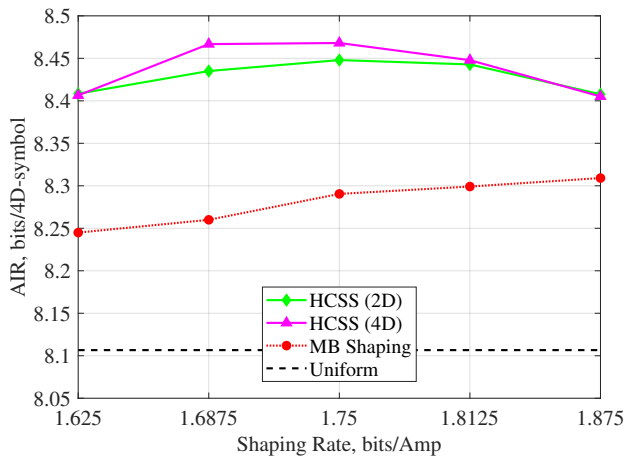
2) *Optimal shaping length*: In Fig. 13 we examine the performance in terms of AIR and effective SNR when varying the shaping sequence length of HCSS ( $L = 64, 96, 128$  are added into consideration) with 1D, 2D and 4D symbol mapping strategies in linear, optimal launch power and highly nonlinear regimes. We note that the results shown in Fig. 13 are based on fitting of launch power sweep measurements.

As discussed previously, in the linear regime, which is shown in Fig. 13 (a) and (d), HCSS with longer shaping sequence length provides higher AIR and closely approaches MB shaping performance with  $L = 160$ . No significant

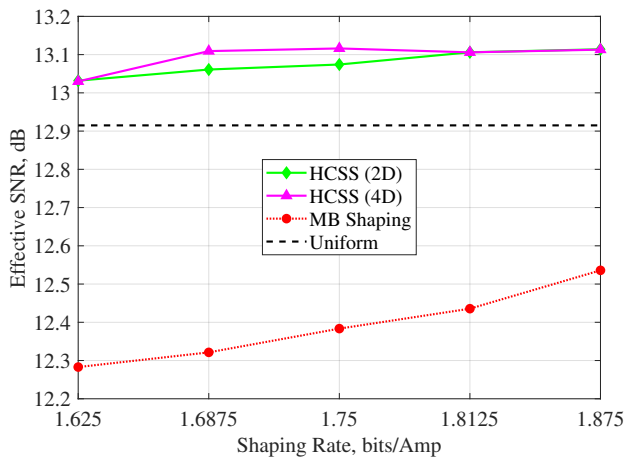
difference is observed among all symbol mapping strategies. Effective SNR also varies insignificantly — there is minor SNR gain at shorter sequence lengths due to weak presence of nonlinearities.

In the optimal launch power regime, which is shown in Fig. 13 (b) and (e), HCSS using 2D and 4D symbol mapping demonstrates significant performance gain in terms of AIR with the shaping sequence length  $L$  in the range of 32–96, while HCSS using 1D symbol mapping provides performance close to MB shaping (with  $L \geq 16$ ). The sequence length  $L = 32$  can be considered optimal (achieving the highest AIR) for all symbol strategies — AIR gain over uniform signaling is 0.37, 0.34, 0.22 b/4D (and 0.19, 0.16, 0.06 b/4D over MB shaping) for 4D, 2D and 1D symbol mapping, respectively. AIR gain is supported by the improvement in effective SNR — 4D mapping provides the highest SNR gain compared to 1D and 2D mapping with  $L \geq 32$ , while 2D mapping slightly outperforms with  $L \leq 16$ . For optimal sequence length of  $L = 32$ , SNR gain over uniform signaling is 0.2, 0.16, 0 dB (and 0.72, 0.68, 0.52 dB over MB shaping) for 4D, 2D and 1D symbol mapping, respectively.

In the highly nonlinear regime, which is shown in Fig. 13 (c) and (f), a similar performance trend can be observed as for optimal launch power regime, however, the gain (both



(a)



(b)

Fig. 14. Performance vs. shaping rate ( $L = 32$ ; 2D and 4D symbol mapping): (a) AIR, (b) Effective SNR.

in AIR and SNR) with short shaping length is exaggerated. For instance, 0.58 b/4D AIR gain and 0.6 dB SNR gain over uniform signaling (0.83 b/4D and 1.75 dB over MB shaping) can be observed with  $L = 32$  and 4D symbol mapping.

In general, we note that HCSS with longer shaping sequence length can achieve better power efficiency for a fixed rate, and therefore provide higher shaping gain in a linear channel. This is clearly seen in the back-to-back case and linear regime of single-span transmission. In the presence of significant fiber nonlinearities, shorter shaping sequences provide higher nonlinear tolerance by improving the effective SNR in the received signal. By choosing appropriate shaping length for the operating regime, an optimal trade-off of shaping gain and effective SNR gain can be achieved. Also, increasing the dimensionality of the symbol mapping can improve nonlinear performance by shortening the shaped symbol sequence length in the time-domain and reducing probabilities of high peak power values, while maintaining the same power efficiency.

3) *Optimal shaping rate:* Next, we studied the impact of the shaping rate on the transmission performance. Fig. 14 shows a characterization of the AIR and effective SNR as a function of the shaping rate. For HCSS we considered shaping sequence length  $L = 32$  with 2D and 4D symbol mapping.

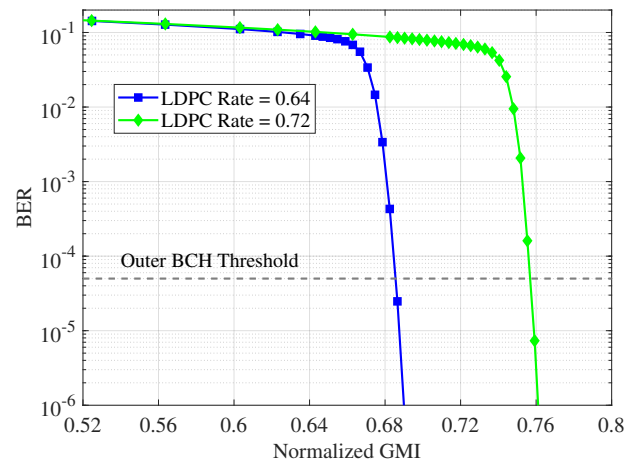


Fig. 15. Characterization of LDPC codes: BER vs. normalized GMI.

Each data point is based on the data fit of launch power sweep measurements and represents the performance at the optimal launch power.

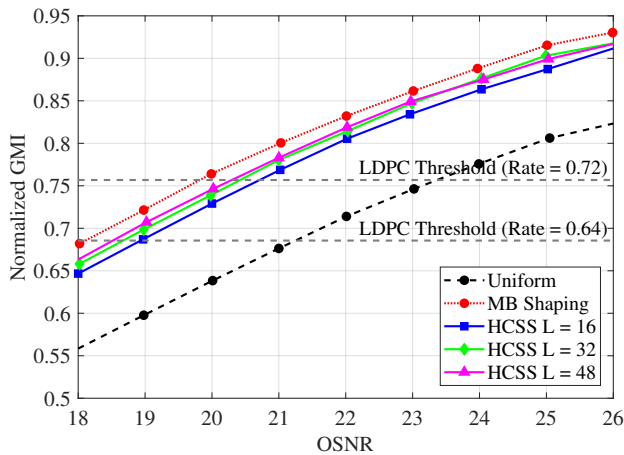
From Fig. 14 (a) we observe that  $R_S = 1.75$  b/Amp is the optimal shaping rate for HCSS — AIR varies over 0.07 b/4D in  $R_S$  range of 1.625–1.875 b/Amp achieving the highest value at  $R_S = 1.75$  b/Amp, the corresponding variation in effective SNR is 0.1 dB. For MB shaping, the AIR increases with the shaping rate by 0.06 b/4D (within  $R_S$  of consideration), which is supported by the corresponding increase in effective SNR of 0.25 dB.

In the case of HCSS with fixed shaping sequence length, the optimal effective SNR does not depend significantly on the shaping rate. Therefore, we may infer that the optimal shaping rate is mostly affected by the linear shaping gain contribution. Conversely, MB shaping demonstrates stronger dependence of both effective SNR and AIR on the shaping rate, indicating that the increase in AIR with shaping rate is associated with nonlinear transmission gain. We note that with increasing shaping rate, MB shaping will converge to uniform signaling, while HCSS will exhibit some rate loss due to the dyadic distribution of compositions constraint.

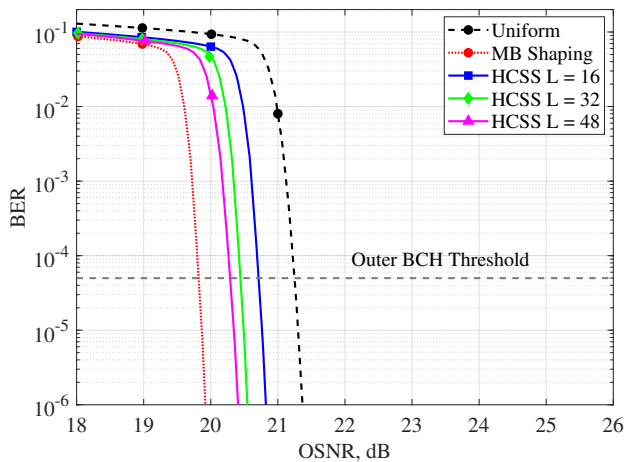
### C. Coded performance

Coded performance analysis is based on mapping nGMI into bit error ratio (BER) after pre-characterized inner LDPC codes. For the shaped signals, an inner LDPC code with rate 0.72 and length 52,800 was used, and decoding was performed over 32 iterations of the sum-product algorithm. For uniform signals an LDPC code with rate 0.64 and length 52,800 was used, and decoding was performed in the same manner. In both cases, we assume the use of an outer Bose–Chaudhuri–Hocquenghem (BCH) code with rate 0.9922, which achieves an output BER below  $10^{-15}$  given an input BER of  $5 \times 10^{-5}$  [46]. We considered MB shaping and HCSS with 4D symbol mapping at  $R_S = 1.75$  b/Amp. In that case the net bit rate after shaping and coding is 424.1 Gb/s for both MB shaping and HCSS, while for uniform signaling the net rate is 426.7 Gb/s.

The characterization of LDPC codes in terms of BER as a function of nGMI is shown in Fig. 15. The LDPC nGMI



(a)



(b)

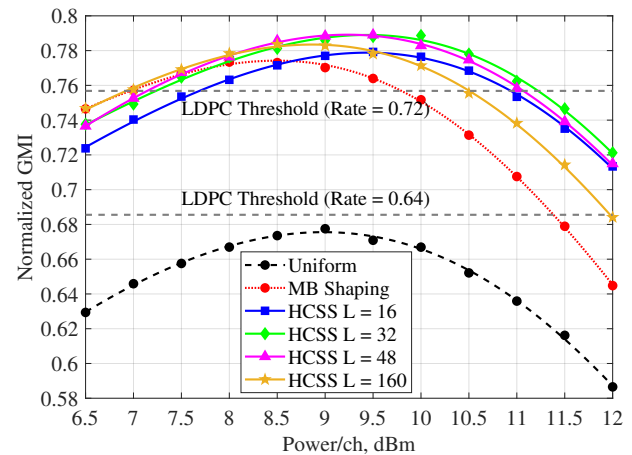
Fig. 16. Coded performance analysis in back-to-back configuration ( $R_S = 1.75$ ;  $L = 16, 32, 48$ ; 4D symbol mapping): (a) normalized GMI vs. OSNR, (b) predicted BER after LDPC vs. OSNR.

thresholds for the outer BCH code (such that the BER at the output of LDPC code is below the BCH code threshold of  $5 \times 10^{-5}$ ) are 0.757 and 0.686 for code rates of 0.72 and 0.64, respectively.

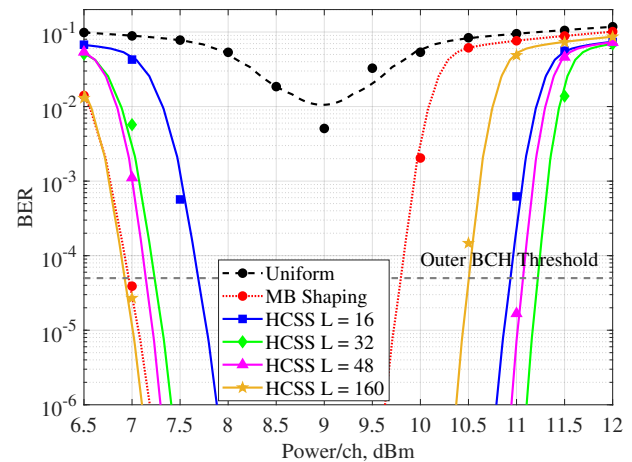
BER after LDPC decoding shown in Fig. 16 (b) and Fig. 17 (b) is predicted based on nGMI shown in Fig. 16 (a) and Fig. 17 (a). The individual points represent predicted BER based on measured nGMI values, while corresponding smooth lines represent predicted BER based on fitted nGMI values, which are calculated from AIR fit according to (11) and (12).

Fig. 16 demonstrates coded performance for the back-to-back configuration. MB shaping achieves the best sensitivity, while HCSS with longer shaping length achieves better sensitivity than with shorter length — for MB shaping OSNR margin is improved by 1.3 dB, for HCSS the OSNR margin is improved by 0.5, 0.7, 0.7 dB for  $L = 16, 32, 48$ , respectively.

Coded performance for single-span transmission is shown in Fig. 17. In the linear regime the sensitivity shows the similar trend as for back-to-back configuration, while in nonlinear regime  $L = 32$  demonstrated the best sensitivity. With  $L = 32$  the launch power margin is improved by 1.3 dB compared to the MB shaping. We note that in case of uniform signaling,



(a)



(b)

Fig. 17. Coded performance analysis for single span transmission ( $R_S = 1.75$ ;  $L = 16, 32, 48$ ; 4D symbol mapping): (a) normalized GMI vs. optical launch power, (b) predicted BER after LDPC vs. optical launch power.

desired performance is not achievable.

## V. DISCUSSION AND CONCLUSIONS

We have investigated HCSS as a method for probabilistic constellation shaping for application in optical fibre communication systems comprising extended-reach single-span links, subject to strong nonlinearities. We have demonstrated that the naïve approach of optimizing the signal PMF and attempting to achieve this with a distribution matcher, while being optimal in the AWGN channel, is highly suboptimal in this case, significantly reducing the ultimate shaping gain. Such a system can achieve only a maximum 0.18 b/4D gain in 200 km single-span SSMF links. Conversely, HCSS achieves a gain of 0.37 b/4D with a shaping sequence length of only  $L = 32$  and 4D symbol mapping. Such an HCSS system can be implemented without multiplications, and LUT size of no more than 100 kbit.

In extended-reach single-span transmissions, signals may suffer from strong highly correlated short-memory nonlinearities. Hence, the temporal structure of the signal (e.g., ordering of transmitted symbols) may have a significant impact on nonlinear performance. By shortening the shaping sequence

length, it is possible to introduce some advantageous changes in the temporal structure — e.g., reduced concentration and probability of high-power symbols [31]. Also, by increasing the dimensionality of symbol mapping, the effective shaped symbol sequence length can be reduced as well as the number of simultaneously interacting independent shaped sequences, which leads to reduced probability of high peak power values.

## APPENDIX A

### PROBABILITY MASS FUNCTION (PMF) CALCULATION

To construct a 4D DP-64QAM format, we consider the alphabet of amplitudes  $\mathcal{A} = \{a_1, a_2, a_3, a_4\} = \{1, 3, 5, 7\}$  and the alphabet of signs  $\mathcal{S} = \{s_1, s_2\} = \{-1, 1\}$ . The alphabet of a 4D-signal (4D-constellation) can be expressed as the Cartesian product  $\mathcal{X} = \mathcal{S}^4 \times \mathcal{A}^4$ , where  $\mathcal{S}^4$  and  $\mathcal{A}^4$  are 4-fold Cartesian products with themselves.

A composition is defined as  $C_i = \{c_1^i, c_2^i, c_3^i, c_4^i\}$ , where  $i$  is the index of the composition in the Huffman-coded structure, and  $c_k^i$  is the number of instances of amplitude  $a_k$  in the shaped amplitude sequence of length  $L = \sum_{k=1}^4 c_k^i$ .

We postulate that the probability distribution of signs  $P_S(s)$  is uniform. Hence,  $P_S(s) = 1/2$  for  $\forall s \in \mathcal{S}$  and  $P_{S^4}(s^4) = 1/16$  for  $\forall s^4 \in \mathcal{S}^4$ . The PMF of a 4D-signal can be therefore expressed as

$$P_X(x) = P_{S^4}(s^4) \cdot P_{A^4}(a^4) = 1/16 \cdot P_{A^4}(a^4), \quad (13)$$

where  $P_{A^4}(a^4)$  can be considered as the PMF of a 4D-quadrant and  $a^4 = [a_{k_1}, a_{k_2}, a_{k_3}, a_{k_4}]$  is a 4D-amplitude vector.

#### A. 1D symbol mapping

Since all components of a 4D-amplitude vector are mapped from independent shaped amplitude sequences, the PMF for a 4D-quadrant is a product of 1D-PMFs:

$$P_{A^4}(a^4) = P_A(a_{k_1}) \cdot P_A(a_{k_2}) \cdot P_A(a_{k_3}) \cdot P_A(a_{k_4}). \quad (14)$$

The 1D-PMF which results from the  $i^{\text{th}}$  composition can be calculated as  $P_A^i(a_k) = c_k^i/L$ , while the total 1D-PMF is

$$P_A(a_k) = \sum_{i=1}^N p_i P_A^i(a_k) = \sum_{i=1}^N p_i \frac{c_k^i}{L}, \quad (15)$$

where  $p_i$  is the probability of occurrence of the  $i^{\text{th}}$  composition, and  $N$  is the total number of compositions.

#### B. 2D symbol mapping

In the case of 2D symbol mapping, two consecutive amplitudes from two independent shaped amplitude sequences are used to construct a 4D-amplitude vector. Therefore, the 4D-quadrant PMF can be expressed as the product of 2D-PMFs:

$$P_{A^4}(a^4) = P_{A^2}(a^2) \cdot P_{A^2}(a^{2*}), \quad (16)$$

where  $a^2 = [a_{k_1}, a_{k_2}]$  and  $a^{2*} = [a_{k_3}, a_{k_4}]$  are 2D-amplitude vectors,  $a^4 = [a^2, a^{2*}]$ .

Since components of 2D-amplitude vectors are mapped from a single sequence and not independent, the 2D-PMF resulting from the  $i^{\text{th}}$  composition can be calculated as

$$P_{A^2}^i(a^2) = \left[ \frac{c_{k_1}^i}{L} \cdot \frac{c_{k_2}^i - \delta}{L-1} \right]^+, \quad (17)$$

where  $[\cdot]^+$  denotes  $\max\{\cdot, 0\}$  operator,  $\delta$  is

$$\delta = \begin{cases} 0, & k_1 \neq k_2, \\ 1, & k_1 = k_2. \end{cases} \quad (18)$$

The total 2D-PMF is

$$P_{A^2}(a^2) = \sum_{i=1}^N p_i P_{A^2}^i(a^2), \quad (19)$$

#### C. 4D symbol mapping

Since four consecutive amplitudes from a single shaped amplitude sequence are used to produce a 4D-amplitude vector (all components of 4D-amplitude vector are not independent), the 4D-quadrant PMF resulting from the  $i^{\text{th}}$  composition is given by

$$P_{A^4}^i(a^4) = \left[ \frac{c_{k_1}^i}{L} \cdot \frac{c_{k_2}^i - \delta_1}{L-1} \cdot \frac{c_{k_3}^i - \delta_2}{L-2} \cdot \frac{c_{k_4}^i - \delta_3}{L-3} \right]^+, \quad (20)$$

where  $\delta_l$  is

$$\delta_l = \sum_{m=1}^l \mathbf{I}_{k_{l+1}=k_m}, \quad (21)$$

where  $\mathbf{I}_{k_{l+1}=k_m}$  is the indicator function which equals 1 when the condition  $k_{l+1} = k_m$  is true and equals 0 otherwise.

The total 4D-PMF is

$$P_{A^4}(a^4) = \sum_{i=1}^N p_i P_{A^4}^i(a^4). \quad (22)$$

We note that the 4D-PMF can not be decomposed into the product of lower-dimensional PMFs.

#### D. Example

Consider a shaper with a single composition  $C = \{6, 5, 3, 2\}$  of length  $L = 16$ . The probability of 4D-symbol  $x_1 = [+7, +7, +7, +7]$  resulting from composition  $C$  using 1D, 2D and 4D symbol mapping will be

$$P_X^{1D}(x_1) = \frac{1}{16} \cdot \frac{2}{16} \cdot \frac{2}{16} \cdot \frac{2}{16} = 0.000015,$$

$$P_X^{2D}(x_1) = \frac{1}{16} \cdot \frac{2}{16} \cdot \frac{1}{15} \cdot \frac{2}{16} = 0.000004,$$

$$P_X^{4D}(x_1) = \frac{1}{16} \cdot \frac{2}{16} \cdot \frac{1}{15} \cdot \frac{0}{14} = 0.$$

For  $x_2 = [+1, +3, +5, +7]$  the probabilities will be

$$P_X^{1D}(x_2) = \frac{1}{16} \cdot \frac{6}{16} \cdot \frac{5}{16} \cdot \frac{3}{16} = 0.000172,$$

$$P_X^{2D}(x_2) = \frac{1}{16} \cdot \frac{6}{16} \cdot \frac{5}{15} \cdot \frac{3}{16} = 0.000195,$$

$$P_X^{4D}(x_2) = \frac{1}{16} \cdot \frac{6}{16} \cdot \frac{5}{15} \cdot \frac{3}{14} = 0.000258.$$

We further note that higher-dimensional mapping reduces the probability of having equal amplitudes and increases the probability of having unequal amplitudes in simultaneous quadratures of a 4D-symbol. This leads to less probable high peak power values in the case of higher-dimensional mapping while maintaining the same power efficiency.

## APPENDIX B EXPERIMENTAL DATA FITTING

Data fitting based on the GN-model [45] was performed for each set of experimental points, which measure effective SNR and AIR at a specified optical launch power.

### A. Effective SNR Fitting

Effective SNR is the combination of ASE, nonlinear and transceiver SNR terms. It can be expressed as

$$\begin{aligned} \frac{1}{\text{SNR}_{\text{eff}}} &= \frac{1}{\text{SNR}_{\text{ASE}}} + \frac{1}{\text{SNR}_{\text{NLin}}} + \frac{1}{\text{SNR}_{\text{Tr}}} \\ &= \frac{a}{P} + \frac{b \cdot P^3}{P} + c, \end{aligned} \quad (23)$$

where  $\text{SNR}_{\text{ASE}}$  is a linear SNR term due to ASE in EDFAs,  $\text{SNR}_{\text{NLin}}$  is a nonlinear SNR term due to nonlinear fiber Kerr nonlinearity (nonlinear noise power is assumed to have cubic dependence on power [45]) and  $\text{SNR}_{\text{Tr}}$ , which is the constant SNR term due to transceiver electrical noise and component imperfections;  $P$  is the launch power;  $a, b, c$  are the fitting parameters. Effective SNR can be rewritten as follows:

$$\text{SNR}_{\text{eff}} = \frac{P}{a + c \cdot P + b \cdot P^3}. \quad (24)$$

Initial estimation of fitting parameters is performed as follows:  $c$  is estimated according to (25) using BtB measurements,  $a$  is estimated using launch power sweep measurements neglecting the nonlinear term at low launch power values (26), and finally  $b$  is estimated using launch power sweep measurements at high launch power values (27).

$$\begin{aligned} c &= \frac{1}{\text{SNR}_{\text{eff}}^{\text{BtB}}} - \frac{1}{\text{SNR}_{\text{ASE}}^{\text{BtB}}} \\ &= \frac{1}{\text{SNR}_{\text{eff}}^{\text{BtB}}} - \frac{BW}{\text{OSNR} \cdot 12.5 \text{ GHz}}, \end{aligned} \quad (25)$$

$$a = \frac{P}{\text{SNR}_{\text{eff}}^{\text{Lin}}} - c \cdot P, \quad (26)$$

$$b = \frac{1}{\text{SNR}_{\text{eff}}^{\text{NLin}} \cdot P^2} - \frac{a}{P^3} - \frac{c}{P^2}, \quad (27)$$

where  $BW$  is the signal bandwidth.

Refinement of the estimates of fitting parameters may be carried out, for example, using the least squares method [47].

### B. AIR Fitting

For AIR fitting we use a linear mapping with regard to SNR in a logarithmic scale

$$\text{AIR} = k \cdot \log_{10}(\text{SNR}_{\text{eff}}), \quad (28)$$

where  $k$  is the fitting parameter. For the range of SNR values considered in this work, linear mapping between AIR and SNR (in logarithmic scale) shows very good agreement. However, we note that for very low or very high SNR values (when AIR saturates) this approach is unsuitable.

## ACKNOWLEDGMENT

The experimental work carried out at Aston University was supported by UK EPSRC grant EP/M009092/1. We thank Lumentum UK for loan of the CFP2-ACO and studentship support of Pavel Skvortcov, and Socionext for loan of the DAC/ADC DKs used in this work.

## REFERENCES

- [1] H. Sun, K.-T. Wu, and K. Roberts, "Real-time measurements of a 40 Gb/s coherent system," *Opt. Express*, vol. 16, no. 2, pp. 873–879, Jan 2008.
- [2] E. Yamazaki, S. Yamanaka, Y. Kisaka, T. Nakagawa, K. Murata, E. Yoshida, T. Sakano, M. Tomizawa, Y. Miyamoto, S. Matsuoka, J. Matsui, A. Shibayama, J. ichi Abe, Y. Nakamura, H. Noguchi, K. Fukuchi, H. Onaka, K. Fukumitsu, K. Komaki, O. Takeuchi, Y. Sakamoto, H. Nakashima, T. Mizuochoi, K. Kubo, Y. Miyata, H. Nishimoto, S. Hirano, and K. Onohara, "Fast optical channel recovery in field demonstration of 100-Gbit/s ethernet over OTN using real-time DSP," *Opt. Express*, vol. 19, no. 14, pp. 13 179–13 184, Jul 2011.
- [3] Y. R. Zhou, K. Smith, R. Payne, A. Lord, G. Whalley, T. Bennett, E. Maniloff, S. Alexander, and D. Boymel, "Real-time gridless 800G super-channel transport field trial over 410km using coherent DP-16 QAM," in *Proc Opt. Fiber Commun. Conf. (OFC)*. Optical Society of America, 2014, p. Tu2B.3.
- [4] E. Agrell and M. Karlsson, "Power-efficient modulation formats in coherent transmission systems," *J. Lightw. Technol.*, vol. 27, no. 22, pp. 5115–5126, 2009.
- [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, 2015.
- [6] D. S. Millar, T. Koike-Akino, S. O. Arik, K. Kojima, K. Parsons, T. Yoshida, and T. Sugihara, "High-dimensional modulation for coherent optical communications systems," *Opt. Express*, vol. 22, no. 7, pp. 8798–8812, Apr 2014.
- [7] D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Multidimensional modulation for next-generation transmission systems," in *Next-Generation Optical Communication: Components, Sub-Systems, and Systems VI*, G. Li and X. Zhou, Eds., vol. 10130, International Society for Optics and Photonics. SPIE, 2017, pp. 85 – 90.
- [8] 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, 2017.
- [9] A. D. Shiner, M. Reimer, A. Borowiec, S. O. Gharan, J. Gaudette, P. Mehta, D. Charlton, K. Roberts, and M. O'Sullivan, "Demonstration of an 8-dimensional modulation format with reduced inter-channel nonlinearities in a polarization multiplexed coherent system," *Opt. Express*, vol. 22, no. 17, pp. 20 366–20 374, Aug 2014.
- [10] K. Kojima, K. Parsons, T. Koike-Akino, and D. S. Millar, "Mapping options of 4D constant modulus format for multi-subcarrier modulation," in *Proc Conf. on Las. and Elect.-Opt. (CLEO)*, 2018, pp. 1–2.
- [11] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 913–929, 1993.
- [12] 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, 1994.
- [13] G. Böcherer, 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, 2015.
- [14] R. F. H. Fischer, *Precoding and signal shaping for digital transmission*. New York, New York, USA: John Wiley & Sons, 2005.
- [15] 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, 2016.

- [16] J. Renner, T. Fehenberger, M. P. Yankov, F. Da Ros, S. Forchhammer, G. Böcherer, and N. Hanik, "Experimental comparison of probabilistic shaping methods for unrepeatable fiber transmission," *J. Lightw. Technol.*, vol. 35, no. 22, pp. 4871–4879, 2017.
- [17] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, 2016.
- [18] 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, 2019.
- [19] D. S. Millar, T. Fehenberger, T. Koike-Akino, K. Kojima, and K. Parsons, "Distribution matching for high spectral efficiency optical communication with multiset partitions," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 517–523, 2019.
- [20] F. Steiner, P. Schulte, and G. Bocherer, "Approaching waterfilling capacity of parallel channels by higher order modulation and probabilistic amplitude shaping," in *Ann. Conf. on Inf. Scien. and Sys. (CISS)*, 2018, pp. 1–6.
- [21] T. Yoshida, M. Karlsson, and E. Agrell, "Hierarchical distribution matching for probabilistically shaped coded modulation," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1579–1589, 2019.
- [22] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Parallel-amplitude architecture and subset ranking for fast distribution matching," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 1981–1990, 2020.
- [23] A. K. Khandani and P. Kabal, "Shaping multidimensional signal spaces. I. optimum shaping, shell mapping," *IEEE Trans. Inf. Theory*, vol. 39, no. 6, pp. 1799–1808, 1993.
- [24] P. Schulte and F. Steiner, "Divergence-optimal fixed-to-fixed length distribution matching with shell mapping," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 620–623, 2019.
- [25] Y. C. Gültekin, W. J. van Houtum, S. Şerbetli, and F. M. J. Willems, "Constellation shaping for IEEE 802.11," in *Proc. IEEE PIMRC*, Montreal, QC, Canada, 2017.
- [26] Y. C. Gültekin, W. J. van Houtum, A. G. C. Koppelaar, and F. M. J. Willems, "Enumerative sphere shaping for wireless communications with short packets," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1098–1112, 2020.
- [27] A. Amari, S. Goossens, Y. C. Gültekin, O. Vassilieva, I. Kim, T. Ikeuchi, C. M. Okonkwo, F. M. J. Willems, and A. Alvarado, "Introducing enumerative sphere shaping for optical communication systems with short blocklengths," *J. Lightw. Technol.*, vol. 37, no. 23, pp. 5926–5936, 2019.
- [28] D. S. Millar, T. Fehenberger, T. Yoshida, T. Koike-Akino, K. Kojima, N. Suzuki, and K. Parsons, "Huffman coded sphere shaping with short length and reduced complexity," in *Proc. Europ. Conf. Opt. Commun. (ECOC)*, Dublin, Ireland, 2019.
- [29] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, K. Parsons, and H. Griesser, "Huffman-coded sphere shaping and distribution matching algorithms via lookup tables," *J. Lightw. Technol.*, vol. 38, no. 10, pp. 2826–2834, 2020.
- [30] T. Fehenberger, H. Griesser, and J.-P. Elbers, "Mitigating fiber nonlinearities by short-length probabilistic shaping," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Diego, CA, USA, 2020.
- [31] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, K. Parsons, and H. Griesser, "Analysis of nonlinear fiber interactions for finite-length constant-composition sequences," *J. Lightw. Technol.*, vol. 38, no. 2, pp. 457–465, 2020.
- [32] S. Goossens, S. Van der Heide, M. Van den Hout, A. Amari, Y. C. Gültekin, O. Vassilieva, I. Kim, T. Ikeuchi, F. M. J. Willems, A. Alvarado, and C. Okonkwo, "First experimental demonstration of probabilistic enumerative sphere shaping in optical fiber communications," in *Proc. Opto-Electron. Commun. Conf. (OECC) and Int. Conf. on Phot. in Switch. and Comp. (PSC)*, Fukuoka, Japan, 2019.
- [33] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel," in *IEEE Int. Symp. on Inf. Theor.*, 2014, pp. 2794–2798.
- [34] 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, 2016.
- [35] P. Schulte, F. Steiner, and G. Bocherer, "Four dimensional probabilistic shaping for fiber-optic communication," in *Advanced Photonics 2017 (IPR, NOMA, Sensors, Networks, SPPCom, PS)*. Optical Society of America, 2017, p. SpM2F.5.
- [36] F. Steiner, F. Da Ros, M. P. Yankov, G. Böcherer, P. Schulte, S. Forchhammer, and G. Kramer, "Experimental verification of rate flexibility and probabilistic shaping by 4D signaling," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2018, pp. 1–3.
- [37] P. Skvortcov, I. Phillips, W. Forsysiak, T. Koike-Akino, K. Kojima, K. Parsons, and D. S. Millar, "Nonlinearity tolerant LUT-based probabilistic shaping for extended-reach single-span links," *IEEE Photon. Technol. Lett.*, pp. 1–1, 2020.
- [38] "OIF 400G ZR implementation agreement." [Online]. Available: [http://oiforum.com/wp-content/uploads/OIF-400ZR-01.0\\_reduced2.pdf](http://oiforum.com/wp-content/uploads/OIF-400ZR-01.0_reduced2.pdf)
- [39] C. Fougstedt, O. Gustafsson, C. Bae, E. Börjeson, and P. Larsson-Edefors, "ASIC design exploration for DSP and FEC of 400-Gbit/s coherent data-center interconnect receivers," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2020, pp. 1–3.
- [40] T. Drenski and J. C. Rasmussen, "ADC & DAC — technology trends and steps to overcome current limitations," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2018, pp. 1–3.
- [41] W. Forsysiak and D. S. Govan, "Progress toward 100-G digital coherent pluggables using inp-based photonics," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2925–2934, 2014.
- [42] H. Chaouch, M. Filer, and A. Bechtolsheim, "Lessons learned from CFP2-ACO system integrations, interoperability testing and deployments," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2017, pp. 1–3.
- [43] P. Skvortcov, C. Sanchez-Costa, I. Phillips, and W. Forsysiak, "Receiver DSP highly tolerant to transmitter IQ impairments," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Diego, CA, USA, 2019.
- [44] T. A. Eriksson, T. Fehenberger, P. A. Andrekson, M. Karlsson, N. Hanik, and E. Agrell, "Impact of 4D channel distribution on the achievable rates in coherent optical communication experiments," *J. Lightw. Technol.*, vol. 34, no. 9, pp. 2256–2266, 2016.
- [45] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 694–721, 2014.
- [46] D. S. Millar, R. Maher, D. Lavery, T. Koike-Akino, M. Pajovic, A. Alvarado, M. Paskov, K. Kojima, K. Parsons, B. C. Thomsen, S. J. Savory, and P. Bayvel, "Design of a 1 Tb/s superchannel coherent receiver," *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1453–1463, 2016.
- [47] R. W. Hamming, *Numerical Methods for Scientists and Engineers*. Dover Press, 1986.

**Pavel Skvortcov** (S'18) received the B.Sc. and M.Sc degrees in applied physics and mathematics from Moscow Institute of Physics and Technology (MIPT), Moscow, Russia, in 2012 and 2014. During 2014–2016 he was working as an Optical System Engineer in "T8", Moscow, Russia. Currently, he is a Ph.D candidate in fiber-optical communications at Aston Institute of Photonic Technologies (AIPT), Aston University, Birmingham, UK. His research interests include high-speed fiber-optic communications, digital signal processing, coded modulation, transceiver optimization.

**Ian Phillips** (M'98) received M.Eng. degree in Electronic Engineering in 1994, and his Ph.D. degree in optical communications from Aston University, Birmingham, UK in 1997 for his study of high-speed all-optical regeneration. He continued his research career in optical communications at British Telecom Research Laboratories and later Corning Research Centre. In 2000 he joined Marconi-Solstis working on ultra-long haul (3000km) DWDM transmission where his worked focused on transceiver design. He continued this work at Marconi, and later at Ericsson before moving into a broader role of Photonic System Test based in Genoa, Italy. He returned to Aston University in 2012 as a Research Fellow where his work focused on compensating nonlinear transmission effects with an overall goal to increase system capacity. He became a Teaching Fellow in 2018, where he continues his research in optical communications.

**Wlodek Forsysiak** holds a Ph.D. in physics and has a research background in nonlinear photonics and high-speed optical fibre communication systems. He was a co-founder of Marconi-Solstis in 2000 and spent 15 years in WDM system related product development with Marconi, Ericsson, and Oclaro. His current research interests are in wideband optical fibre communication systems, optical devices and subsystems, and the impact and mitigation of device and fibre nonlinearities.

**Toshiaki Koike-Akino** (M'05–SM'11) received the B.S. degree in electrical and electronics engineering, M.S. and Ph.D. degrees in communications and computer engineering from Kyoto University, Kyoto, Japan, in 2002, 2003, and 2005, respectively. During 2006–2010 he was a Postdoctoral Researcher at Harvard University, and joined Mitsubishi Electric Research Laboratories, Cambridge, MA, USA, in 2010. His research interests include digital signal processing for data communications and sensing. He received the YRP Encouragement Award 2005, the 21st TELECOM System Technology Award, the 2008 Ericsson Young Scientist Award, the IEEE GLOBECOM'08 Best Paper Award in Wireless Communications Symposium, the 24th TELECOM System Technology Encouragement Award, and the IEEE GLOBECOM'09 Best Paper Award in Wireless Communications Symposium.

**Keisuke Kojima** (S'82–M'84–SM'13) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Tokyo, Tokyo, Japan, in 1981, 1983, and 1990, respectively. He also received the M.S. degree from the University of California, Berkeley, CA, USA, in 1982. He worked for eight years at the Central Research Laboratory, Mitsubishi Electric Corp., from 1983 on the research of narrow linewidth DFB and DBR lasers, and optical neural networks. He spent nine years at AT&T/Lucent Bell Laboratories on the R&D of uncooled Fabry–Perot and DFB lasers, vertical-cavity surface-emitting lasers, passive optical network systems, and metro optical systems. He also worked at Agere Systems, Denselight Semiconductors, and TriQuint Semiconductors on optical devices and modules, and optical systems testbed. He has been with Mitsubishi Electric Research Laboratories, Cambridge, MA, USA, since 2005, where he is currently working on the design of photonic-integrated circuits and coherent optical systems using deep learning, as a Distinguished Research Scientist. He has more than 200 publications in journals and conference proceedings. He is a Fellow of the Optical Society of America, and Associate Editor of IEEE/OSA Journal of Lightwave Technology.

**Kieran Parsons** (M'07–SM'09) received the B.Eng. and Ph.D. degrees in electronic and communications engineering from the University of Bristol, Bristol, U.K., in 1992 and 1996, respectively. During 1997–2002, he was with Nortel Networks, Ottawa, Canada, where he worked on wireless and long-haul optical communications system architecture and design. From 2004–2006, he worked on carrier-grade mesh WiFi RF system design at BelAir Networks (now part of Ericsson) and from 2006–2009 on 10-G PHY device development with Applied Micro, both in Kanata, Canada. In 2009, he joined Mitsubishi Electric Research Laboratories, Cambridge, MA, USA, where he is currently Senior Team Leader and Senior Principal Research Scientist. His research interests include digital signal processing and coherent optical transmission systems.

**David S. Millar** (M'08) received the Ph.D degree from University College London (UCL), London, UK in 2011. From 2011-2012 he was a post-doctoral researcher at UCL. Since 2012 he has been with Mitsubishi Electric Research Laboratories (MERL) in Cambridge, Massachusetts, USA. His research interests include digital communications, optics, sensing and imaging systems.