MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Dual Coding Concatenation for Burst-Error Correction in Probabilistic Amplitude Shaping

Skvortcov, Pavel; Koike-Akino, Toshiaki; Millar, David S.; Kojima, Keisuke; Parsons, Kieran

TR2022-145 November 17, 2022

Abstract

We propose the use of dual coding concatenation for mitigation of post-shaping burst errors in probabilistic amplitude shaping (PAS) architectures. The proposed dual coding concatenation for PAS is a hybrid integration of conventional re- verse concatenation and forward concatenation, i.e., post-shaping forward error correction (FEC) layer and pre-shaping FEC layer, respectively. A low-complexity architecture based on parallel Bose–Chaudhuri–Hocquenghem (BCH) codes is introduced for the pre-shaping FEC layer. Proposed dual coding concatenation can relax bit error rate (BER) requirement after post-shaping soft-decision (SD) FEC codes by an order of magnitude, resulting in a gain of up to 0.25 dB depending on the complexity of post- shaping FEC. Also, combined shaping and coding performance was analyzed based on sphere shaping and the impact of shaping length on coding performance was demonstrated.

IEEE Journal of Lightwave Technology 2022

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Dual Coding Concatenation for Burst-Error Correction in Probabilistic Amplitude Shaping

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

***Abstract*—We propose the use of dual coding concatenation for mitigation of post-shaping burst errors in probabilistic amplitude shaping (PAS) architectures. The proposed dual coding concatenation for PAS is a hybrid integration of conventional reverse concatenation and forward concatenation, i.e., post-shaping forward error correction (FEC) layer and pre-shaping FEC layer, respectively. A low-complexity architecture based on parallel Bose–Chaudhuri–Hocquenghem (BCH) codes is introduced for the pre-shaping FEC layer. Proposed dual coding concatenation can relax bit error rate (BER) requirement after post-shaping soft-decision (SD) FEC codes by an order of magnitude, resulting in a gain of up to 0.25 dB depending on the complexity of post-shaping FEC. Also, combined shaping and coding performance was analyzed based on sphere shaping and the impact of shaping length on coding performance was demonstrated.**

***Index Terms*—Probabilistic amplitude shaping, forward error correction, parallel codes, optical fiber communications.**

I. Introduction

In recent years, probabilistic constellation shaping for coherent optical communication systems has been widely studied and became a mature technology for deployment [1]. The key enabler for probabilistic shaping in practical systems was the introduction of probabilistic amplitude shaping (PAS) [2]. PAS allows for low-complexity implementation of probabilistic shaping into modern bit-interleaved coded modulation (BICM) systems with a capability of decoupled design of channel coding and shaping stages.

Many practical shaping approaches have been proposed since the initial proposal of PAS. One class is based on targeting fixed distribution of amplitudes, referred to as distribution matching (DM) — examples are constant-composition DM (CCDM) [3], multiset-partion DM (MPDM) [4], [5], product DM [6] and hierarchical DM [7]. Another class is based on defining the most energy-efficient signal space for finite-length shaping, referred to as spherical shaping (SS) [8], [9] — examples are shell mapping (SM) [10], [11], enumerative sphere shaping (ESS) [12]–[14], and Huffman-coded sphere shaping (HCSS) [15], [16]. Typically, DM-based approaches require longer shaping lengths compared to SS-based approaches to achieve the same performance. While the optimal performance for linear channels is achieved with asymptomatic infinite-length shaping, for nonlinear fiber-optical channels the best performance is achieved with finite-length shaping due to complimentary nonlinearity mitigation gains. Specifically, for SS-based approaches the optimal shaping length is in the range of 100–200 amplitudes for long-haul links [14], [17] and 20–50 amplitudes for highly nonlinear short-distance links [18]–[20]. We note that short-length shaping is also advantageous in terms of hardware complexity and latency.

Within conventional PAS architecture, which typically employs *reverse* concatenation of shaping and forward error correction (FEC) stages, FEC encoding is performed on shaped bits. We note that a common approach for FEC in PAS is a concatenation of a powerful soft-decision (SD) code and a low-complexity hard-decision (HD) code with a low overhead, e.g., low-density parity-check (LDPC) code and Bose–Chaudhuri–Hocquenghem, (BCH) code, respectively.

At the receiver-side, shaped bits are first decoded and then demapped (i.e., de-shaped) to information bits under the assumption that all bits are correctly decoded (no errors within shaping sequences). While state-of-the-art FEC schemes achieve low output bit error rates (BERs), they always have a non-zero probability of errors in practice. The most commonly considered post-FEC BER threshold is 10^{-15}, which is used to determine acceptable performance after FEC decoding for conventional uniform signalling. However, for systems utilizing PAS architecture, BER *enhancement* may occur after shaping demapping due to uncorrected errors within shaped sequences after FEC decoding — a single bit error within a shaped sequence will induce multiple errors of bits (i.e., burst error) after shaping demapper. While some specific shaping mappers/demappers are proposed to reduce the effect of BER enhancement [7], [21], they are not universally applicable to all mapping/demapping schemes.

Fig. 1 shows BER enhancement as a function of shaping length for 64-ary quadrature-amplitude modulation (64-QAM), where all information bits are shaped and 50 % bits will be in error on average when burst errors occur after shaping demapping (note that BER enhancement varies with shaping algorithm and error distributions). The BER enhancement after de-shaping is approximately linear with the shaping length. Accordingly, it is in favour of shorter-length shaping in addition to aforementioned nonlinear gains in optical channels.

For conventional PAS with *reverse* concatenation, a typical solution against the BER enhancement is simply to use stronger SD-FEC codes to decrease post-FEC BER such that BER enhanced after shaping demapping is kept below

P. Skvortcov is presentnly with Infinera Corporation, San Jose, CA 95119, USA. E-mail: pskvortcov@infinera.com.

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

D. S. Millar is presently with Infinera Corporation, Kanata, ON K2K 2X3, Canada E-mail: dmillar@infinera.com.

This work is done when P. Skvortcov and D. S. Millar worked for MERL.

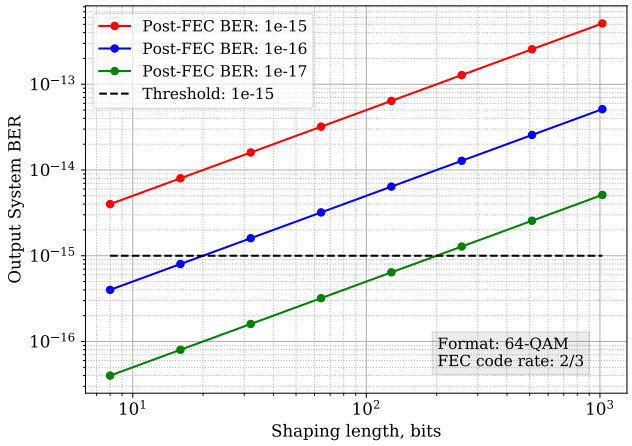


Fig. 1. BER enhancement: system output BER vs. shaping length.

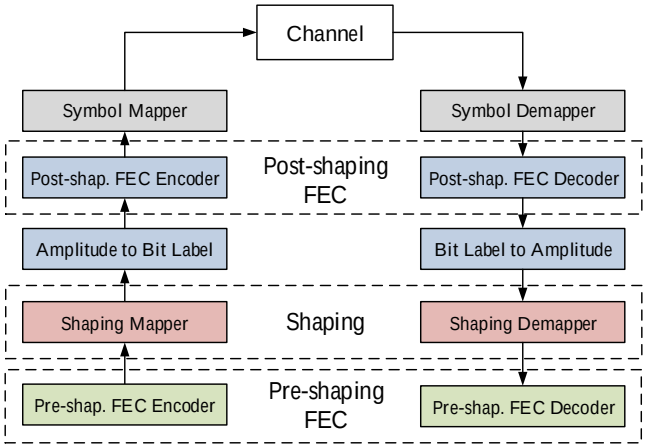


Fig. 2. PAS architecture with dual concatenation having post-shaping inner and pre-shaping outer FEC layers.

10^{-15}. We propose an alternative solution based on *dual concatenation* with pre-shaping (forward concatenation) and post-shaping (reverse concatenation) FEC layers, which can effectively mitigate the BER enhancement.

In this work, we extend our preliminary analysis of *dual concatenation* for PAS, presented in [22]. First, we introduce the concept of pre-shaping FEC for PAS and consider various burst error correction coding bounds. We then propose a low-complexity scalable architecture for correction of burst errors in the pre-shaping FEC layer, which is based on parallel BCH codes. Subsequently, a comprehensive performance analysis for PAS is provided. The coding performance is characterized in terms of the the impact of various shaping and coding parameters. The combined shaping and coding performance is then analyzed using SS-based shaping in both additive white Gaussian noise (AWGN) and fiber-optical channels.

II. Preliminaries

A. PAS architecture

Fig. 2 shows a diagram of PAS architecture with pre-shaping and post-shaping FEC layers. This architecture employs a dual concatenation framework which uses a joint forward and reverse concatenation of the shaping and FEC coding. Specifically, shaping precedes FEC coding in reverse concatenation, while opposite in forward concatenation. We note that typical PAS systems use only reverse concatenation without a dedicated pre-shaping FEC layer.

The key components of the PAS architecture are the shaping mapper and demapper which perform mapping and demapping of the stream of uniform input bits into unsigned probabilistically shaped amplitudes. Post-shaping systematic FEC coding is performed on the bit-labels of the shaped amplitudes and parity bits are added as signs of amplitudes. Signed amplitudes which represent amplitude-shift keying (ASK) are then concatenated into symbols of a QAM format. The transmission rate can be adapted by changing the amplitude shaping rate, while keeping the post-shaping FEC code rate fixed if desired. The pre-shaping FEC layer is placed outside the shaping layer and designed to correct burst errors after shaping demapper due to residual errors after post-shaping FEC decoding.

Fig. 3 shows framing and mapping of shaping sequences and FEC codewords within PAS. The shaping rate (in bits per unsigned amplitude, b/Amp) is defined as

$$R_{\mathrm{Sh}} = \frac{L_{\mathrm{DeSh}}}{L_{\mathrm{Sh}}^{\mathrm{Amp}}} = \frac{L_{\mathrm{DeSh}}}{L_{\mathrm{Sh}}}(m-1), \tag{1}$$

where $L_{\mathrm{Sh}}^{\mathrm{Amp}}$ is the shaped sequence length in amplitudes, L_{Sh} is the shaped sequence length in bits, L_{DeSh} is the de-shaped sequence length in bits, and m is the number of bit labels per amplitude (i.e., $m = \log_2 M$ of M-ASK format). Typically, L_{Sh} is fixed and L_{DeSh} is varied with the shaping rate for transmission rate. The maximum shaping rate is $R_{\mathrm{Sh}} = m - 1$, which corresponds to uniform shaping, and the de-shaped sequence is not longer than shaped sequence ($L_{\mathrm{DeSh}} \leq L_{\mathrm{Sh}}$).

For post-shaping FEC, we consider a typical concatenation of powerful SD LDPC code and low-complexity HD BCH codes (we note that post-LDPC interleaver can be also employed to distribute LDPC burst errors). The code rate is lower-bounded by $(m-1)/m$ at which all parity bits are assigned to amplitude signs. To have higher code rates, some information bits (referred to as unshaped bits) may be carried out on signs of amplitudes besides parity bits. The overall code rate of post-shaping FEC is given as

$$R_{\mathrm{FEC}}^{\mathrm{PostSh}} = R_{\mathrm{LDPC}}^{\mathrm{PostSh}} \cdot R_{\mathrm{BCH}}^{\mathrm{PostSh}} = \frac{m-1+\gamma}{m}, \tag{2}$$

where $R_{\mathrm{LDPC}}^{\mathrm{PostSh}}$ is the code rate of post-shaping LDPC code, $R_{\mathrm{BCH}}^{\mathrm{PostSh}}$ is the code rate of post-shaping BCH code, and $0 \leq \gamma \leq 1$ specifies the portion of unshaped bits. The length of unshaped bits is $L_{\mathrm{UnSh}} = \gamma L_{\mathrm{Sh}}/(m-1)$ and the length of parity bits is $L_{\mathrm{Par}} = (1-\gamma)L_{\mathrm{Sh}}/(m-1)$.

B. Pre-shaping FEC

1) Burst error correction bounds: First, we consider bounds on burst error correction — theoretical bounds: Rieger bound and Hamming bound; as well as bounds on performance of practical codes: Fire code, Reed–Solomon code, and BCH code. These bound are summarized in Table I, where n is the

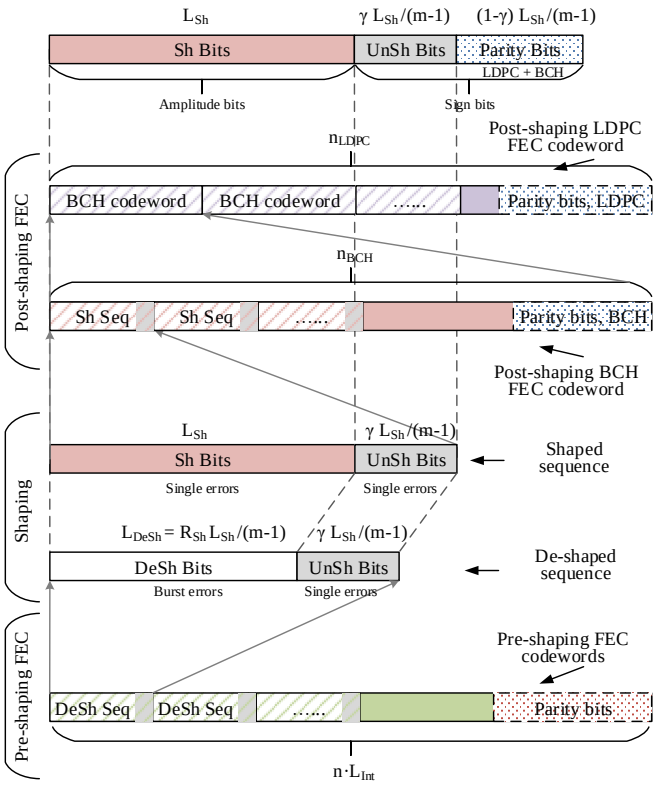


Fig. 3. Framing and mapping within PAS architecture.

TABLE I
BURST ERROR CORRECTION BOUNDS

Type	Bound
Rieger Bound	$k = n - 2l$
Fire Code	$k = n - 3l$
Reed–Solomon Code	$k^{\text{sym}} = n^{\text{sym}} - 2l^{\text{sym}}$, $l^{\text{sym}} = 1 + \lfloor \frac{l+s-2}{s} \rfloor$ with s being symbol size
Hamming Bound	$k = n - \lceil \log_2(\sum_{l=0}^{t} \binom{n}{l}) \rceil$
BCH Code	$k = n - l \log_2(n+1)$

length of the code, k is the information length within codeword (i.e., payload) and l is the length of the burst.

For Hamming bound and BCH code, we consider parallel structures based on block interleaving to enhance burst error correction ability of random error correction approaches. Interleaving is performed in the following manner: block of input bits of length n' are written sequentially in rows of length L_{Int} (referred to as interleaving order) and output bits are read sequentially from columns of length $n = n'/L_{\text{Int}}$. Block interleaving is used to spread the burst errors among multiple codewords, such that short code with reduced error correcting ability can be used. E.g., for the burst of length l we can use the code of length n' with l-error correcting ability, while with L_{Int}-way interleaving we can use the code of length $n = n'/L_{\text{Int}}$ with error correcting ability of l/L_{Int}. When the interleaving order equals the burst length ($L_{\text{Int}} = l$) we refer to it as a fully-parallel structure.

Fig. 4 shows comparison of burst error correction bounds in terms of required code overhead for single burst correction.

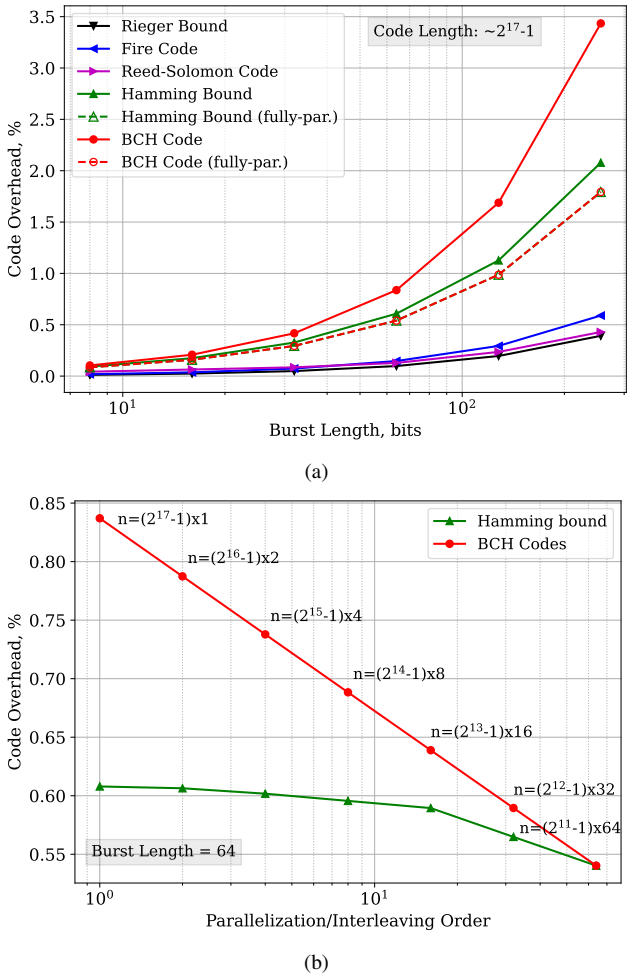


Fig. 4. Burst error correction approaches for single burst correction.

Fig. 4 (a) demonstrates the required code overhead as a function of burst length. Reed–Solomon code provides excellent burst error correction capability and can closely approach Rieger bound. Fire code is also an efficient approach, which requires about 30 % higher overhead compared to Rieger bound. For random error correction approaches (Hamming bound and BCH code), required overheads are significantly higher. However, fully-parallel/interleaved structures can substantially improve burst error correction ability.

Fig. 4 (b) shows the required code overhead as a function of interleaving/parallelization order for BCH code and Hamming bound. We note that required overhead for fully-parallel BCH code is matched with Hamming bound, since only single error needs to be corrected per each codeword. While random error correction approaches are in general less efficient for burst error correction, parallel BCH code can offer low complexity, scalability and ability to correct multiple bursts.

2) Parallel BCH code: For pre-shaping FEC layer we consider parallel BCH structure based on block interleaving of shaping sequences according to Fig. 5. Each column in the L_{Int}-way interleaved structure is a codeword of a BCH code of length $n = n'/L_{\text{Int}}$. We consider fully parallel structure, where we utilize $L_{\text{Int}} = L_{\text{DeSh}} + L_{\text{UnSh}}$-way interleaving. We note that in the case of full interleaving we have no more than

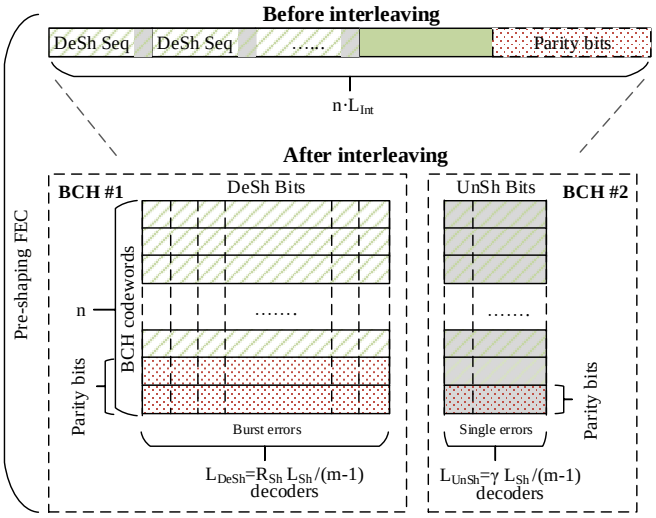


Fig. 5. Parallel structure for pre-shaping BCH code.

$t = T$ errors per parallel codeword after interleaving when T bursts occurs within a non-interleaved structure.

This structure enables parallel encoding/decoding, suited for high-throughput systems. It can be easily scaled according to the shaping rate. This parallel structure is initially designed for the maximum shaping rate (i.e., $L_{\text{DeSh}} = L_{\text{Sh}}$). Then, adaptivity is achieved by subsequently disabling parallel encoders. Furthermore, separate BCH codes can be used for shaped and unshaped bits, since the requirement for error correcting ability can be different — for unshaped bits there is no BER enhancement and, hence, potentially weaker BCH code may be used to minimize the overall overhead. In Fig. 5 greyed blocks represent unshaped bits. Respective parity bits of BCH codes are kept within unshaped and shaped blocks of bits.

We also note that the latency of the pre-shaping parallel BCH structure is expected to be relatively low compared to the post-shaping FEC. The pre-shaping latency is mainly due to the interleaver length, while the decoding latency for short parallel BCH codes can be maintained low. The post-shaping latency is mainly dominated by iterative LDPC decoding and post-LDPC interleaver, and expected to be significantly higher.

C. Transmission rate

The transmission rate in bits per one-dimensional symbol (b/1D) for conventional PAS systems with reverse concatenation can be expressed as

$$\begin{aligned} R_{\text{Tr}} &= \frac{L_{\text{DeSh}} + L_{\text{UnSh}}}{L_{\text{Sh}}^{\text{Amp}}} \\ &= \frac{R_{\text{Sh}} L_{\text{Sh}}/(m-1) + \gamma L_{\text{Sh}}/(m-1)}{L_{\text{Sh}}/(m-1)} = R_{\text{Sh}} + \gamma, \end{aligned} \tag{3}$$

where $\gamma = 1 - m(1 - R_{\text{FEC}}^{\text{PostSh}}) = 1 - m(1 - R_{\text{LDPC}}^{\text{PostSh}} \cdot R_{\text{BCH}}^{\text{PostSh}})$. For PAS with dual concatenation based on BCH codes in pre-shaping FEC layer transmission rate is expressed as

$$\begin{aligned} R_{\text{Tr}} &= \frac{L_{\text{DeSh}} \cdot R_{\text{BCH}}^{\text{PreSh}} + L_{\text{UnSh}} \cdot R_{\text{BCH}}^{\text{PreSh}'}}{L_{\text{Sh}}^{\text{Amp}}} \\ &= R_{\text{Sh}} \cdot R_{\text{BCH}}^{\text{PreSh}} + \gamma \cdot R_{\text{BCH}}^{\text{PreSh}'}, \end{aligned} \tag{4}$$

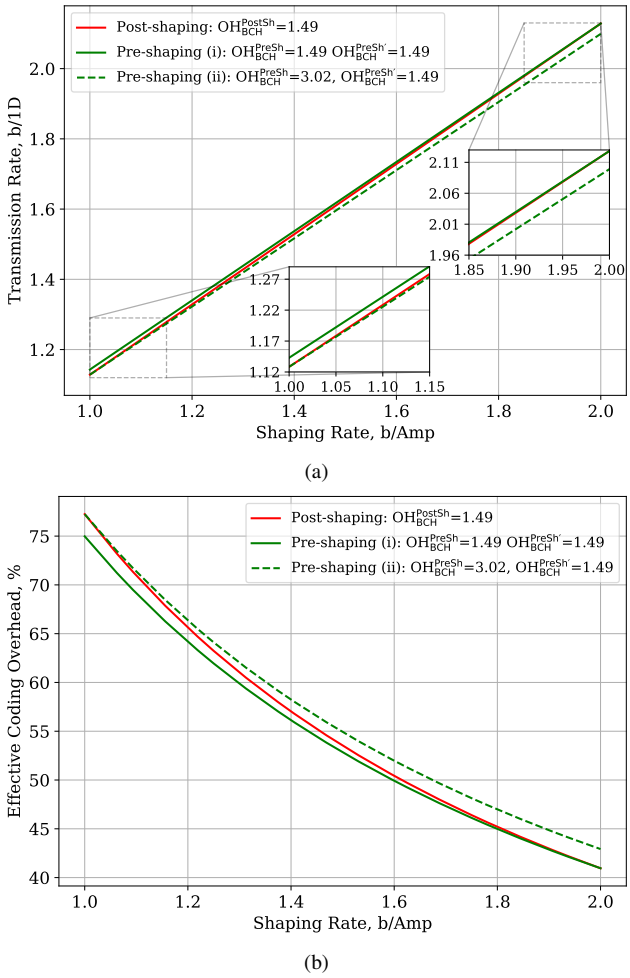


Fig. 6. Transmission rate and effective coding overhead vs. shaping rate for different coding options: (a) transmission rate; (b) effective coding overhead

where $R_{\text{BCH}}^{\text{PreSh}}$ is the rate of the pre-shaping BCH code for shaped bits, and $R_{\text{BCH}}^{\text{PreSh}'}$ is the rate of the pre-shaping BCH code for unshaped bits.

We also define effective coding overhead for PAS as

$$\text{OH}_{\text{Eff}} = \frac{1 - R_{\text{Eff}}}{R_{\text{Eff}}}, \tag{5}$$

where R_{Eff} is the effective coding rate calculated as

$$R_{\text{Eff}} = \frac{R_{\text{Tr}}}{R_{\text{Sh}} + 1}, \tag{6}$$

where denominator represents the uncoded transmission rate (i.e., maximum transmission rate for a given shaping rate).

Fig. 6 shows the comparison of effective coding overhead and corresponding transmission rate as a function of shaping rate for the cases without pre-shaping or post-shaping BCH code. The base modulation format for PAS is 64-QAM (i.e., $m = 3$), and post-shaping LDPC code is used for both cases at a rate of $R_{\text{LDPC}}^{\text{PostSh}} = 0.72$. As shown in Fig. 6 (b), post-shaping coding is less efficient compared to pre-shaping coding, since not the full alphabet of bit sequences of length L_{Sh} is utilized for signalling, and coding protects all possible sequences, causing an extra overhead. Specifically, pre-shaping BCH codes having an overhead of 1.49 % can achieve lower

effective coding overhead than post-shaping BCH code having the same code overhead of 1.49 %, for a shaping rate range of $1 \leq R_{\mathrm{Sh}} < 2$. We note that post-shaping coding is performed on bit sequences with extended length due to shaping, whereas pre-shaping coding is done on shorter bit sequences.

The above-mentioned gain motivates us to consider the use of stronger pre-shaping BCH codes having higher overhead, without sacrificing the total transmission rate. For example, the pre-shaping BCH code can have an overhead up to 3.02 % to achieve higher transmission rates than the case of the conventional reverse concatenation for a shaping rate of $R_{\mathrm{Sh}} = 1$ b/Amp. The optimization of the overhead adjustment will be discussed further in the following sections.

III. Error rates evaluation

We evaluate the performance of post-/pre-shaping BCH codes. As simulating the whole systems at a target BER of 10^{-15} is not practical and the performance may highly depend on particular LDPC codes and shaping algorithms, we mainly focus on analytical derivation to give more generic insights. Note that the post-shaping FEC layer uses a concatenation of LDPC and BCH codes. We consider numerically pre-characterized LDPC codes and assume that post-LDPC errors are distributed randomly (with the use of post-LDPC interleaver). We consider BCH codes (n, k, t), where n is the code length ($n = 2^s - 1$), k is the information length within codeword ($k = n - t \cdot s$), and t is the error correction ability of the code. The code rate of BCH codes is then defined as

$$R_{\mathrm{BCH}} = \frac{k}{n} = \frac{n - s \cdot t}{n}, \tag{7}$$

while corresponding code overhead is expressed as

$$\mathrm{OH}_{\mathrm{BCH}} = \frac{1 - R_{\mathrm{BCH}}}{R_{\mathrm{BCH}}} = \frac{s \cdot t}{n - s \cdot t}. \tag{8}$$

A. Shaped sequence error probability and BER enhancement

Probability of having at least one bit error in the shaped sequence (before shaping demapping) is given as

$$P_{\mathrm{Err}}^{\mathrm{ShSeq}} = 1 - (1 - \mathrm{BER}_{\mathrm{Sh}})^{L_{\mathrm{Sh}}} \approx \mathrm{BER}_{\mathrm{Sh}} \cdot L_{\mathrm{Sh}}, \tag{9}$$

where $\mathrm{BER}_{\mathrm{Sh}}$ is the expected BER at the output of post-shaping FEC. After de-shaping the length of the burst error is L_{DeSh}. The total BER after de-shaping is a combination of enhanced BER of shaped bits and non-enhanced BER of un-shaped bits, as follows:

$$\mathrm{BER}_{\mathrm{DeSh}} = \frac{L_{\mathrm{DeSh}} \cdot \mathrm{BER}_{\mathrm{DeSh}}^{\mathrm{Sh}} + \gamma L_{\mathrm{Sh}}/(m-1) \cdot \mathrm{BER}_{\mathrm{DeSh}}^{\mathrm{UnSh}}}{L_{\mathrm{DeSh}} + \gamma L_{\mathrm{Sh}}/(m-1)}, \tag{10}$$

where $\mathrm{BER}_{\mathrm{DeSh}}^{\mathrm{UnSh}} = \mathrm{BER}_{\mathrm{Sh}}$ and enhanced BER after de-shaping can be expressed as

$$\begin{aligned} \mathrm{BER}_{\mathrm{DeSh}}^{\mathrm{Sh}} &= \frac{1}{N}\Bigg[\underbrace{p \cdot L_{\mathrm{DeSh}}}_{\text{Errors per seq.}} \cdot \underbrace{P_{\mathrm{Err}}^{\mathrm{ShSeq}} \frac{N}{L_{\mathrm{DeSh}}}}_{\text{Number of seq.}}\Bigg] \\ &= p \cdot P_{\mathrm{Err}}^{\mathrm{ShSeq}} \approx p \cdot \mathrm{BER}_{\mathrm{Sh}} \cdot L_{\mathrm{Sh}}, \end{aligned} \tag{11}$$

where p is the probability of flipped bit in a burst after de-shaping and N is the length of observation window for BER measurements. Note that while with small shaping rate the length of the bursts after de-shaping reduces, the number of de-shaped sequences within the same observation window increases. Therefore, according to (11), $\mathrm{BER}_{\mathrm{DeSh}}$ is independent of L_{DeSh} and the shaping rate.

B. Output BER for post-shaping BCH code

We here estimate BER at the output of post-shaping BCH code. We assume that BCH code corrects all l errors for $l \leq t$ within codeword and no errors are corrected if $l > t$. Probability of having l bit errors within codeword is expressed with binomial distribution probability mass function (PMF) as

$$P_{\mathrm{Errors}}(l) = \binom{n}{n-l}(1 - \mathrm{BER}_{\mathrm{in}})^{n-l}(\mathrm{BER}_{\mathrm{in}})^{l}, \tag{12}$$

where $\mathrm{BER}_{\mathrm{in}}$ is the input BER for BCH decoding. Output BER after the post-shaping BCH code is then given as

$$\begin{aligned} \mathrm{BER}_{\mathrm{out}} &= \frac{1}{n}\underbrace{\sum_{l>t} l \cdot P_{\mathrm{Errors}}(l)}_{\text{Uncorrected errors}} \\ &= \mathrm{BER}_{\mathrm{in}} - \underbrace{\frac{1}{n}\sum_{l\leq t} l \cdot P_{\mathrm{Errors}}(l)}_{\text{Corrected BER}}. \end{aligned} \tag{13}$$

Note that for our case we have $\mathrm{BER}_{\mathrm{Sh}} = \mathrm{BER}_{\mathrm{out}}$ and $\mathrm{BER}_{\mathrm{in}}$ is post-LDPC BER. The above binomial distribution assumes sufficiently long interleaving to realize random bit errors after LDPC decoding. We have verified that our analysis agrees well with simulation results for such a case. However, when the post-LDPC interleaver order is limited, it is known that the error patterns are correlated to cause burst errors in practice. Even for that case, correlated bit errors can be dealt with a slight modification of the binomial distribution. In addition, (13) still holds given a true distribution $P_{\mathrm{Errors}}(l)$.

C. Output BER for pre-shaping BCH code

For the estimation of BER performance of pre-shaping FEC layer, we consider fully parallel/interleaved structure with (n, k, t) BCH codes. First, we consider the worst case when all bits are flipped in a burst. Probability of having T errors per each parallel codeword (equivalently, T bursts of length L_{DeSh} within non-interleaved structure of length $n \cdot L_{\mathrm{DeSh}}$) can be calculated using binomial distribution PMF as

$$P_{\mathrm{Bursts}}(T) = \binom{N_{\mathrm{seq}}}{N_{\mathrm{seq}} - T}\left(1 - P_{\mathrm{Err}}^{\mathrm{ShSeq}}\right)^{N_{\mathrm{seq}}-T}\left(P_{\mathrm{Err}}^{\mathrm{ShSeq}}\right)^{T}, \tag{14}$$

where N_{seq} is the number of de-shaped sequences within non-interleaved structure (which equals to the length of the code n in the case of full interleaving). The output BER is then calculated as

$$\mathrm{BER}_{\mathrm{out}} = P_{\mathrm{Error}}^{\mathrm{Sh}} - \underbrace{\frac{1}{n}\sum_{T\leq t} T \cdot P_{\mathrm{Bursts}}(T)}_{\text{Corrected BER}}. \tag{15}$$

Next, considering the realistic case that probability of flipped bit in a burst is p, the probability of having l flipped bits

per parallel codeword when T bursts occurs can be expressed using binomial distribution PMF as

$$P_{\text{Errors}}(l|T) = \begin{cases} \binom{T}{T-l}(1-p)^{T-l}(p)^{l}, & l \leq T, \\ 0, & l > T. \end{cases} \tag{16}$$

The total probability of having l errors per codeword in the parallel structure can be expressed as

$$P_{\text{Errors}}(l) = \sum_{T=0}^{N_{\text{seq}}} P_{\text{Errors}}(l|T) \cdot P_{\text{Bursts}}(T). \tag{17}$$

The output BER for pre-shaping BCH code is given as

$$\begin{aligned} \text{BER}_{\text{out}} &= p \cdot P_{\text{Err}}^{\text{ShSeq}} - \underbrace{\frac{1}{n}\sum_{l \leq t} l \cdot P_{\text{Errors}}(l)}_{\text{Corrected BER}} \\ &= p \cdot P_{\text{Err}}^{\text{ShSeq}} - \underbrace{\frac{1}{n}\left[\sum_{l \leq t} l \cdot \sum_{T=0}^{N_{\text{seq}}} P_{\text{Errors}}(l|T) \cdot P_{\text{Bursts}}(T)\right]}_{\text{Corrected BER}}, \end{aligned} \tag{18}$$

where the first term represents enhanced BER after shaping demapper ($\text{BER}_{\text{DeSh}}^{\text{Sh}}$). We note that this analysis is specifically for shaped bits, whereas BER evaluation for unshaped bits is equivalent to post-shaping BCH analysis.

D. Output background block error rate (BBER)

We consider optical transport network (OTN) framing for systems operating beyond 100 Gbit/s — optical transport unit Ck (OTUCk, where Ck index refers to the approximate bit-rate of $k \times 100$ Gbit/s) framing is defined in ITU-T G.709 standard [23]. The length of OTUCk frames is $L_{\text{OTU}} = 130560 \times k$ bits ($4080 \times 4 \times k$ bytes). Note that FEC for the OTUCk signals is interface/vendor specific and not included in the OTUCk definition in ITU-T G.709 standard.

The performance of systems with OTN framing is typically evaluated using BBER or severely errored second rate (SESR) [24]. SESR is based on a one-second period which contains $\geq 15\%$ errored frames. BBER is the ratio of errored OTU frames to the total number of transmitted frames, excluding those occurred during severely errored seconds. Throughout the paper we focus on BBER metric besides BER.

OTUCk frames may consist of multiple de-inteleaved blocks (which consist of multiple pre-shaping BCH codewords). In such case output BBER after de-shaping and pre-shaping decoding is calculated as

$$\text{BBER} = 1 - \left[P_{\text{Errors}}^{\text{PreSh}}(l \leq t) \cdot P_{\text{Errors}}^{\text{PreSh}\prime}(l \leq t)\right]^{N_{\text{block}}^{\text{OTU}}}, \tag{19}$$

where $N_{\text{block}}^{\text{OTU}} = L_{\text{OTU}}/(nL_{\text{DeSh}} + nL_{\text{UnSh}})$ is the number of de-interleaved blocks, $P_{\text{Errors}}^{\text{PreSh}}(l \leq t)$ is the probability of decoding success of pre-shaping BCH code for shaped bits, and $P_{\text{Errors}}^{\text{PreSh}\prime}(l \leq t) = P_{\text{Errors}}(l \leq t)^{L_{\text{UnSh}}}$ is the probability of decoding success of pre-shaping BCH code for unshaped bits. When outer coding is not used, BBER is calculated with (19) by setting error-correcting abilities of the codes to zero.

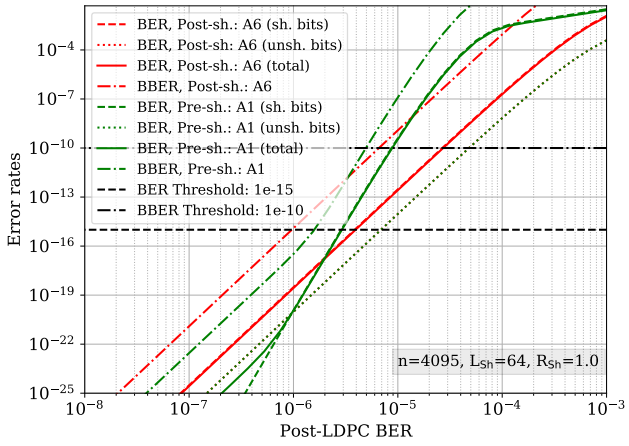


Fig. 7. Post-shaping and pre-shaping BCH codes performance. Total, shaped and unshaped BER and BBER.

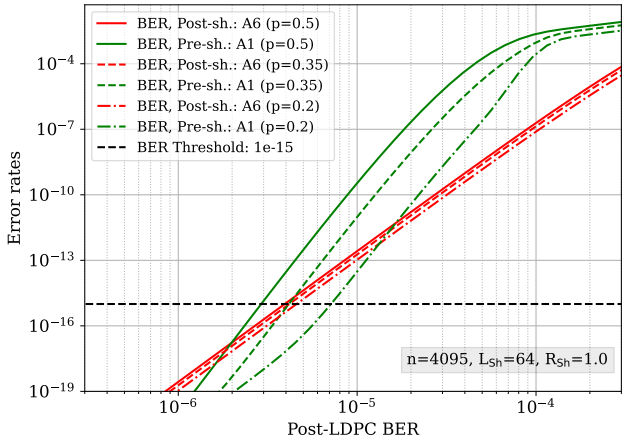


Fig. 8. Post-shaping and pre-shaping BCH codes performance. Total BER with various probabilities of flipped bit p.

E. Error rates analysis

Fig. 7 shows BER analysis of post-shaping and pre-shaping BCH codes, respectively, for the configurations "A1" and "A6" in Table II. The modulation format is 64-QAM, shaping length is $L_{\text{Sh}} = 64$, and shaping rate is $R_{\text{Sh}} = 1$. The considered code length is $n = 4095$. Probability of flipped bit in a burst after shaping demapping is $p = 0.5$.

We present the dependence of output de-shaped BERs for shaped bits, unshaped bits, and also total resulted BER as a function of input BER (post-LDPC BER). Here, we choose overheads for outer BCH code for unshaped bits such that BER for unshaped bits is kept below BER for shaped bits. It is seen in Fig. 7 that post-shaping BCH code is not effective compared to pre-shaping BCH code when LDPC codes show a moderate BER below about 2×10^{-6}.

Fig. 8 shows performance of the same BCH configurations with probabilities of flipped bit $p = 0.5, 0.35, 0.2$. Lower flipping probability results in reduced BER enhancement, which relaxes post-LDPC BER requirement to achieve output BER below 10^{-15} for both post-shaping and pre-shaping BCH codes. However, lower flipping probability is more beneficial

TABLE II
CODE PARAMETERS: $n = 4095$

Configuration	$\mathrm{OH}_{\mathrm{BCH}}^{\mathrm{PostSh}}$	$\mathrm{OH}_{\mathrm{BCH}}^{\mathrm{PreSh}}$	$\mathrm{OH}_{\mathrm{BCH}}^{\mathrm{PreSh}'}$	R_{Tr}, b/1D
A1 (pre-sh.)	0.00	3.02	1.49	1.1284
A2	0.29	2.40	1.19	1.1284
A3 (optimal)	0.59	1.79	0.89	1.1285
A4	0.89	0.89	0.89	1.1310
A5	1.19	0.29	0.29	1.1314
A6 (post-sh.)	1.49	0.00	0.00	1.1284

TABLE III
CODE PARAMETERS: $n = 8191$

Configuration	$\mathrm{OH}_{\mathrm{BCH}}^{\mathrm{PostSh}}$	$\mathrm{OH}_{\mathrm{BCH}}^{\mathrm{PreSh}}$	$\mathrm{OH}_{\mathrm{BCH}}^{\mathrm{PreSh}'}$	R_{Tr}, b/1D
B1 (pre-sh.)	0.00	2.77	2.44	1.1292
B2	0.16	2.44	2.27	1.1293
B3	0.32	2.27	1.12	1.1292
B4	0.48	1.94	0.96	1.1292
B5	0.64	1.61	0.80	1.1293
B6	0.80	1.29	0.64	1.1293
B7 (optimal)	0.96	0.96	0.48	1.1292
B8	1.12	0.48	0.48	1.1306
B9	1.29	0.16	0.16	1.1308
B10 (post-sh.)	1.45	0.00	0.00	1.1291

for pre-shaping BCH code. With $p = 0.35$ the required post-LDPC BER is approximately the same for both post-shaping and pre-shaping BCH codes, while with $p = 0.2$ higher post-LDPC BER can be tolerated with pre-shaping BCH.

We note that probability of flipped bit and related BER enhancement depend on the specific design of shaping mapping/demapping (i.e., specific algorithm) [7], [21]. For further analysis, we assume $p = 0.5$, which represents the worst case of BER enhancement — i.e., a single error at the input of a shaping demapper results in the whole sequence demapped incorrectly, and therefore on average half of the bits are in error within de-shaped sequence. We also emphasize that lower probability of flipped bit can be more advantageous for pre-shaping BCH configurations.

IV. OPTIMAL DUAL CODING CONCATENATION

The proposed dual coding concatenation for PAS enables flexible configurations and heterogeneous assignments for pre-shaping and post-shaping error patterns. The joint use of pre-shaping and post-shaping BCH codes shows a remarkable performance improvement by optimizing the trade-off between overhead and error correction ability for respective BCH codes to deal with post-shaping burst errors. In this section we analyze the coding performance of dual concatenation systems to optimize the configuration in terms of BER and required SNR. We specifically investigate the impact of various shaping and coding parameters: shaping length, shaping rate, BCH code length and overhead.

Performance analysis is mainly based on two metrics: post-LDPC BER threshold, which is a required BER after post-shaping LDPC code; and required SNR (per bit) at the input of LDPC code for a target system performance. For target system performance we use a BER of 10^{-15} and BBER of 10^{-10}. The base modulation format for PAS is 64-QAM. For

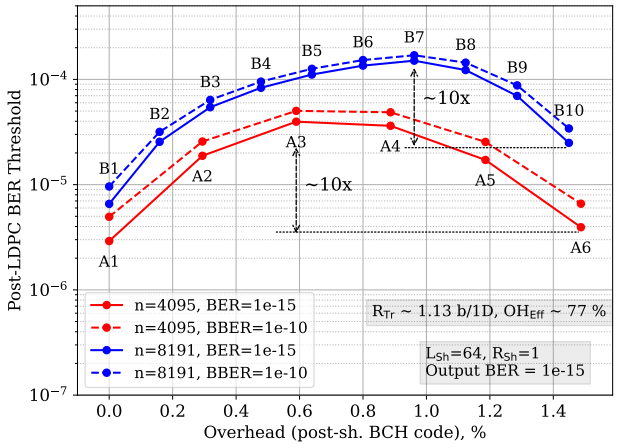


Fig. 9. Optimal dual-concatenation of pre-/post-shaping BCH codes.

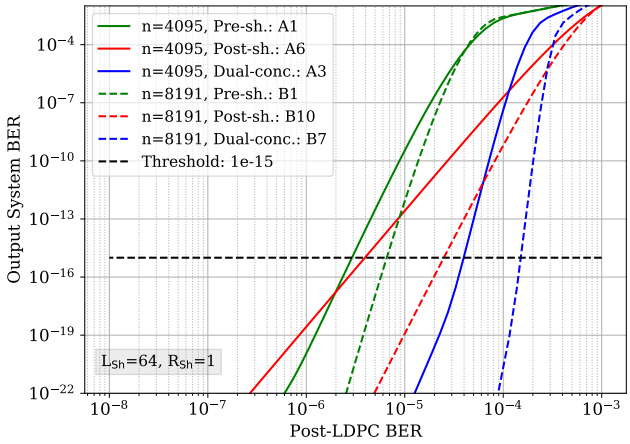


Fig. 10. Output system BER for post-shaping BCH, pre-shaping BCH, and optimal dual-concatenated BCH configurations.

the results in Sections IV-A – IV-D and IV-F, the considered code rate of post-shaping LDPC codes is 0.72.

A. Optimal dual-concatenation of post-shaping and pre-shaping BCH codes

We consider various configurations for concatenation of pre-/post-shaping BCH codes. The overhead adjustment is listed in Tables II and III for BCH codes of $n = 4095$ and $n = 8191$ bits, respectively. For the conventional reverse concatenation (i.e., "A6" and "B10" without pre-shaping BCH code), we consider an overhead for post-shaping BCH code of about 1.5 % (minor deviation due to finite granularity in adjustment of BCH parameters). We then varied the overhead for post-shaping BCH code (up to 1.5 %) and defined complimentary overheads for pre-shaping BCH codes such that the transmission rate (and effective coding overhead) is closely matched with the case of the conventional reverse concatenation. Note that higher overheads can be used for pre-shaping BCH codes compared to post-shaping BCH code, while the transmission rate is always not lower than the reverse concatenation case.

Fig. 9 shows the required post-LDPC BER threshold over different configurations of BCH concatenation. Here, we con-

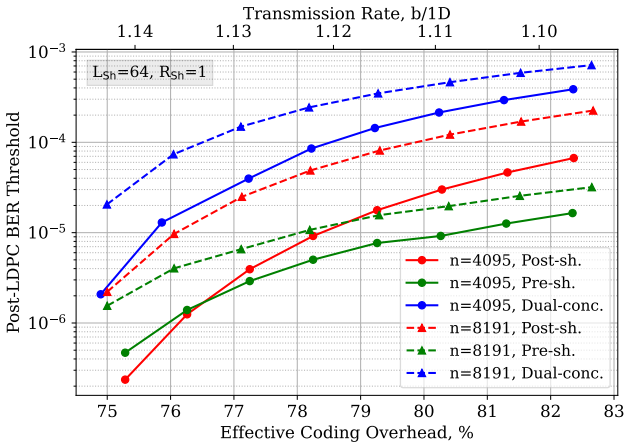


Fig. 11. Threshold for post-LDPC BER. Effective coding overhead sweep.

sider a shaping rate of $R_{\mathrm{Sh}} = 1$ b/Amp and a shaping length of $L_{\mathrm{Sh}} = 64$. The left-end points (A1 and B1) and right-end points (A6 and B10) represent post-shaping BCH only and pre-shaping BCH only configurations, respectively, while points in the middle represent dual coding concatenation. When comparing end points, we can observe that post-shaping BCH slightly outperforms pre-shaping BCH configuration (for shorter code length the difference is negligible). With the optimal dual concatenation, the post-LDPC BER threshold can be improved approximately by an order of magnitude compared to only post-shaping BCH configuration — from 3×10^{-6} to 3×10^{-5} for $n = 4095$ and from 2×10^{-5} to 2×10^{-4} for $n = 8191$ for a target output BER of 10^{-15}. Similar trends are observed for target output BBER of 10^{-10}.

Fig. 10 shows overall output system BER as a function of post-LDPC BER for post-shaping BCH only, pre-shaping BCH only, and optimal dual-BCH concatenation. We can see that the gain of the proposed dual concatenation is more significant for a system that requires more stringent BER target.

B. Impact of code overhead

Fig. 11 shows performance of post-shaping, pre-shaping and optimally dual-concatenated BCH coding configurations across effective coding overheads and corresponding transmission rates. In general, increased effective overhead result in more relaxed post-LDPC BER threshold — e.g., for optimally dual-concatenated configuration the post-LDPC BER threshold is increased from 2×10^{-6} to 3×10^{-4} when increasing effective overhead from 75 % to 82.5 % for $n = 4095$. It is important to note that for lower overheads pre-shaping BCH configuration becomes more advantageous — e.g., pre-shaping BCH outperforms post-shaping BCH at low overheads for $n = 4095$. Similar trends are observed for $n = 8191$.

C. Impact of shaping length and shaping rate

Fig. 12 shows performance of BCH coding configurations in terms of shaping lengths and shaping rates. In Fig. 12 (a) we considered shaping rates of $R_{\mathrm{Sh}} = 0.5$, 1 and 1.5 b/Amp for a shaping length from $L_{\mathrm{Sh}} = 8$ to 1024 (equivalently,

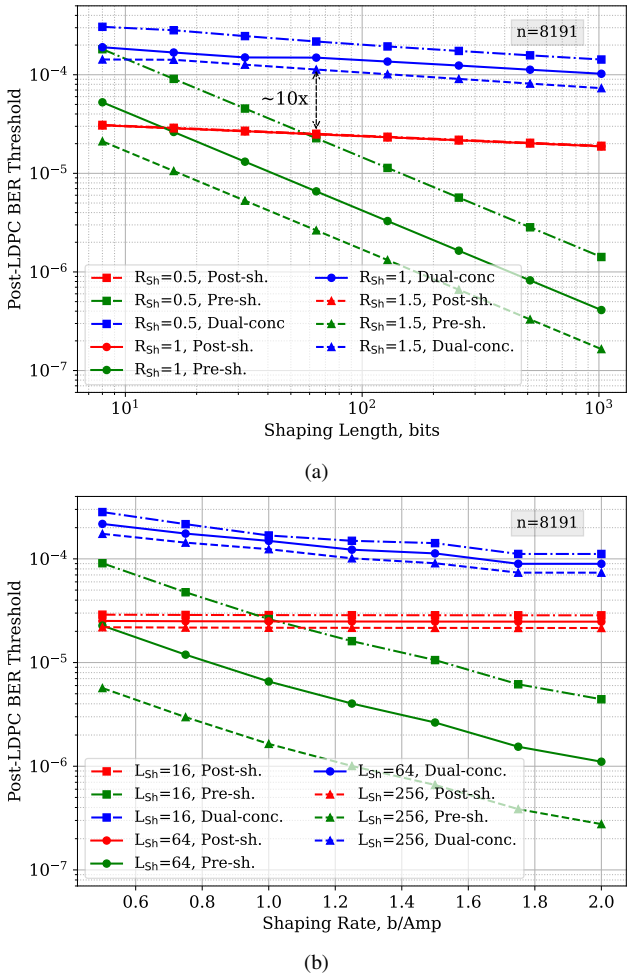


Fig. 12. Threshold for post-LDPC BER. Shaping length and rate sweep.

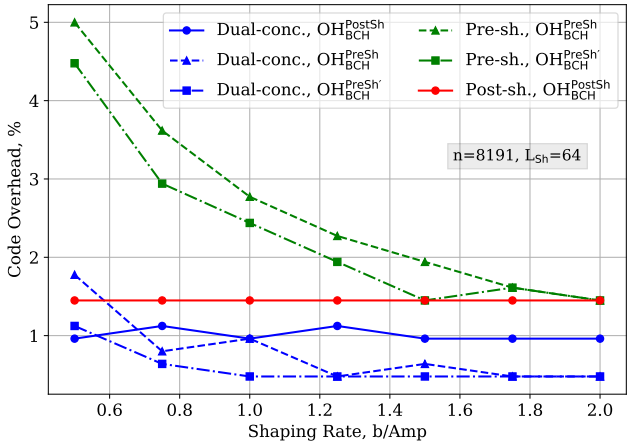


Fig. 13. Threshold for post-LDPC BER. De-shaped BER threshold sweep.

$L_{\mathrm{Sh}}^{\mathrm{Amp}} = 4$ through 512). In Fig. 12 (b) we considered shaping lengths of $L_{\mathrm{Sh}} = 16$, 64 and 256 for a shaping rate from $R_{\mathrm{Sh}} = 0.5$ to 2 b/Amp (where $R_{\mathrm{Sh}} = 2$ b/Amp corresponds to uniform shaping). We note that shaping rates below $R_{\mathrm{Sh}} = 0.5$ b/Amp are not justified with 64-QAM base modulation format as lower-order QAM may be more relevant for such cases.

For short-length shaping, performance of pre-shaping BCH

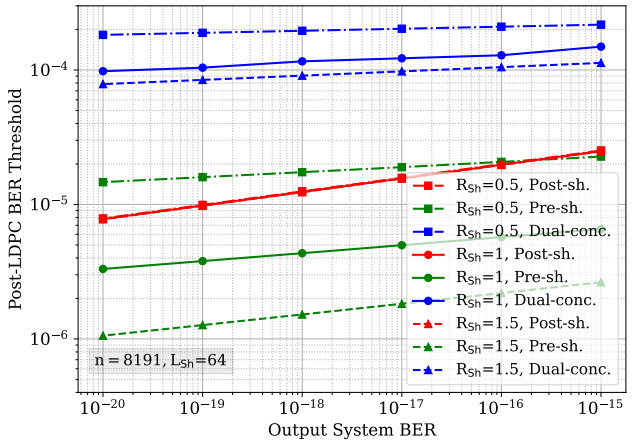


Fig. 14. Threshold for post-LDPC BER. De-shaped BER threshold sweep.

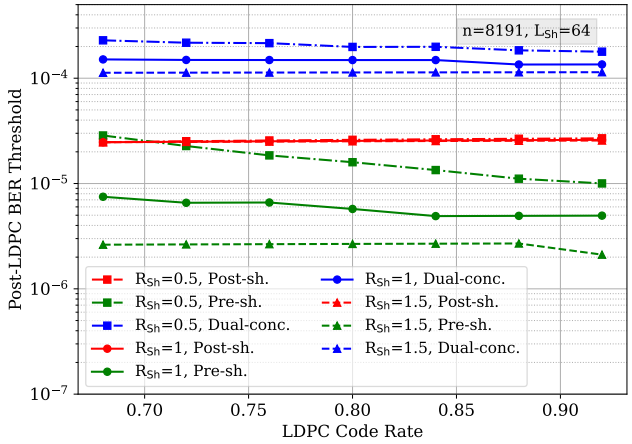


Fig. 15. Threshold for post-LDPC BER with varying the LDPC code rate.

is close to post-shaping BCH and even superior for R_{Sh} below 1 b/Amp, while for long-length shaping the performance of pre-shaping BCH is considerably worse. With optimally dual-concatenated BCH systems, the performance can be substantially improved and the gain over post-shaping BCH case is nearly constant within the considered shaping length range.

With respect to the shaping rate, for lower shaping rate we observe improved performance of pre-shaping and dual-concatenated BCH configurations. This is because of the increased code overheads for pre-shaping BCH codes due to the transmission rate gain over the conventional reverse concatenation. We note that this effect is more impactful for pre-shaping BCH configuration.

Fig. 13 demonstrates the respective code overheads of pre-shaping BCH codes (for shaped and unshaped bits) and post-shaping BCH codes as a function of shaping rate when $L_{\mathrm{Sh}} = 64$. For dual-concatenated and pre-shaping configurations, optimal code overheads for pre-shaping BCH codes are increased with lower shaping rate (transmission rate is kept the same). We note that impact of shaping length on code overheads is insignificant.

TABLE IV
LDPC CODE PARAMETERS

Parameter	Value
Code rate	0.72
Code length, bits	1650, 3300, 6600, 13200, 26400, 52800
Net coding gain	9.1, 10.1, 11.1, 12.3, 12.7, 13.0dB (32-ite)
Decoding iterations	4–32
Degree design	Pareto-optimal [25]
Variable-node degree	$\lambda(x) = 0.45x^2 + 0.44x^4 + 0.11x^{16}$
Check-node degree	$\rho(x) = 0.11x^{15} + 0.89x^{16}$
Girth	Cycle-8 designed by PEG [26]

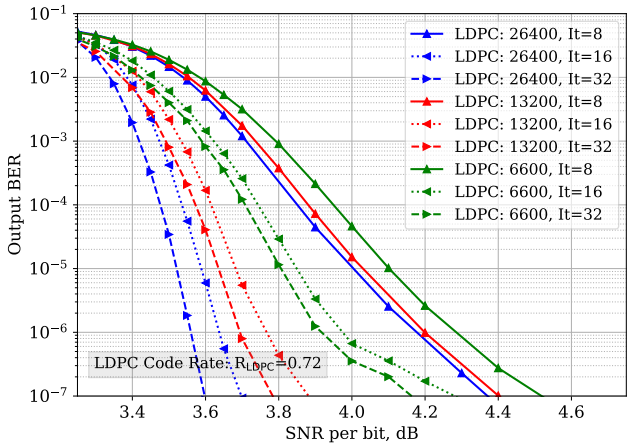


Fig. 16. Characterization of LDPC codes: BER vs. SNR (per bit).

D. Impact of post-shaping BER threshold

Fig. 14 shows performance in terms of system target BER threshold. For conventional post-shaping BCH configuration we observe a stronger dependence on the system target BER, requiring lower post-LDPC BER for more stringent system BER target. Dual-concatenated BCH configuration provides more significant gain for a system that requires stringent system BER target and is robust across wide target BER range.

E. Impact of post-shaping LDPC code rate

Fig. 15 shows performance vs. post-shaping LDPC code rate. The LDPC code rate affects the proportion of unshaped bits in relation to shaped bits — i.e., higher code rate results in larger portion of unshaped bits, which results in slightly reduced BER enhancement. This has negligible impact on post-LDPC BER threshold for post-shaping and dual-concatenated BCH configurations. For pre-shaping BCH configuration, a non-significant decrease in post-LDPC BER is observed for higher LCPC code rates and lower shaping rates.

F. Required SNR for post-shaping LDPC codes

For required SNR analysis, we consider numerically pre-characterized state-of-the-art post-shaping LDPC codes based on the iteration-aware Pareto-optimal irregular design method proposed in [25]. The Pareto-optimal code was demonstrated to be effective for realizing Tb/s experimental systems [27]. The girth is optimized by progressive edge-growth (PEG) [26].

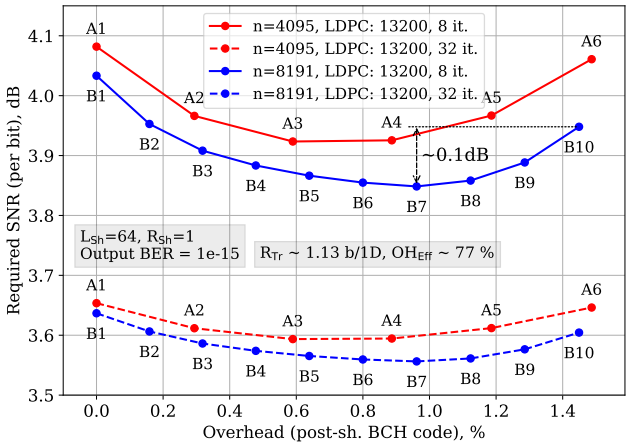


Fig. 17. Required SNR for post-shaping LDPC code.

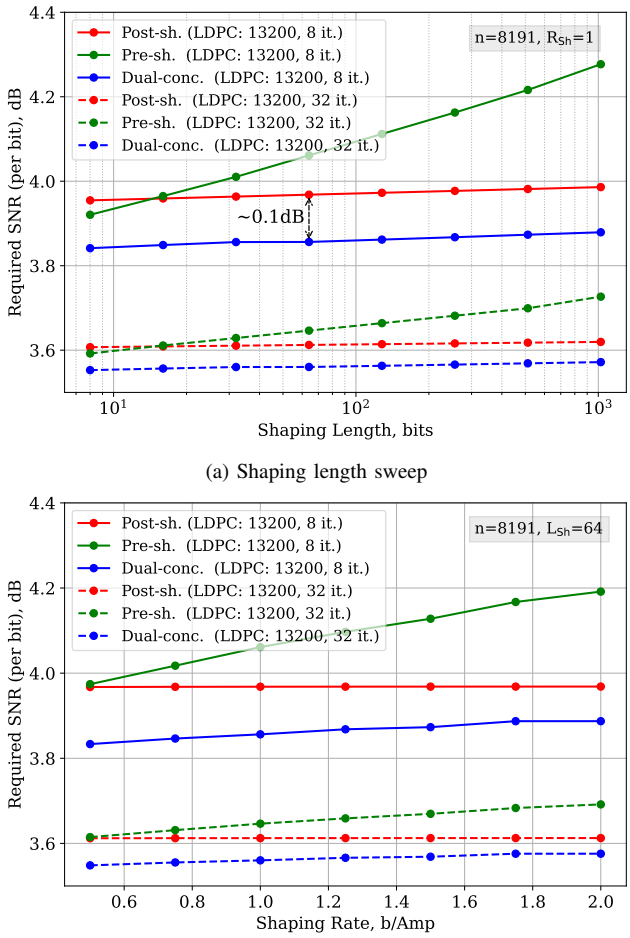
Code parameters are summarized in Table IV. The considered code rate is 0.72, code lengths are 1650, 3300, 6600, 13200, 26400, 52800, and LDPC decoding is performed using the layered sum-product algorithm over 4–32 iterations. The characterization of some LDPC code configurations in terms of BER as a function of SNR per bit is shown in Fig. 16. Note that those codes achieve an excellent performance close to the Gallager's random coding bound (RCB) [28], [29] with a gap of about 0.3 dB for 32-iteration decoding.

Figs. 17 and 18 show required SNR per bit for LDPC decoding with BCH concatenation configurations corresponding to Figs. 9 and 12, respectively. We considered the code length of 13200 and decoding over 8 and 32 iterations. In Fig. 17, for 8-iteration decoding, the SNR gain is greater than 0.1 dB, while for 32 decoding iterations the gain is around 0.05 dB. Fig. 18 (a) shows constant SNR gain with dual-concatenation over the whole shaping length range — around 0.1 dB with decoding over 8 iterations and 0.05 dB with 32 iterations. Fig. 18 (b) shows SNR variations of 0.05 dB and 0.02 dB for dual-concatenation with 8 and 32 decoding iterations.

Fig. 19 shows performance over the number of decoding iterations and code length for post-shaping LDPC code. In general, higher SNR gains with dual-concatenation can be achieved when using lower-complexity LDPC codes — i.e. shorter code length and less decoding iterations. For the code length of 13200, the SNR gain is increased to 0.25 dB with 4 decoding iterations compared to 0.1 dB with 8 decoding iterations. With relation to the code length, the SNR gain is increased to 0.2 dB from 0.05 dB when reducing the code length from 13200 to 1650 with 32 decoding iterations.

V. Sphere Shaping with Dual Coding Concatenation in Practical Channels

We previously focused on coding gain for burst-error correction, but shaping gain was omitted for analysis. In this section, we further evaluate the combined shaping and coding performance in a system utilizing the state-of-the-art energy-efficient SS-based shaping in AWGN and fiber-optical channels. The base modulation format is dual-polarization (DP) 64-QAM.



(a) Shaping length sweep

(b) Shaping rate sweep

Fig. 18. Required SNR for post-shaping LDPC code.

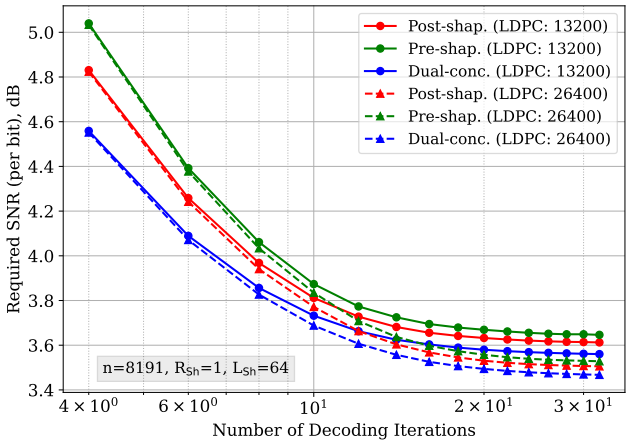
A. Linear AWGN channels

First, we consider numerically simulated AWGN channels. Fig. 20 shows required channel SNR as a function of shaping length. The shaping rate was set to $R_{\mathrm{Sh}} = 1.7$ b/Amp. The dual-concatenated coding configuration provides SNR gain of 0.1 dB for 8 iterations of LDPC decoding and 0.05 dB for 32 iterations. Also, it is important to note that maximum performance for pre-shaping coding configuration is observed at a finite shaping length. Classic assumption is that longer length shaping offers better performance in AWGN channel. However, as shown, coding aspect can affect this statement in reality. It is expected that for post-shaping and dual-concatenated configurations the similar trends will be seen beyond simulation range of Fig. 20.

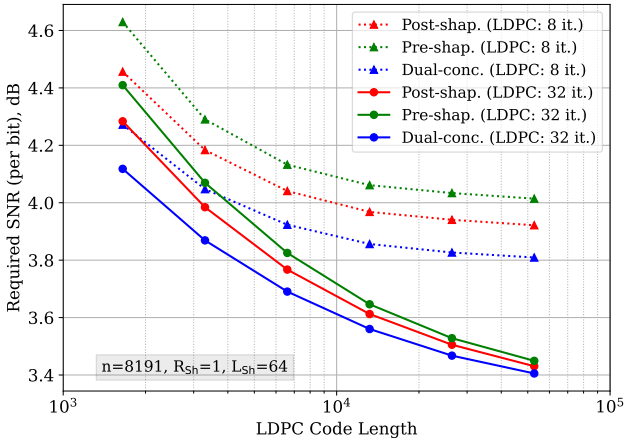
B. Nonlinear fiber-optic channels

For fiber-optical channels, we consider long-haul transmission link consisting of 30 spans of 100 km standard single-mode fiber, simulated with a full-field split-step Fourier method (SSFM). The parameters of the simulated system are listed in Table V. Laser phase noise and polarization mode dispersion effects are excluded from the simulated model.

Fig. 21 shows SNR margin as a function of shaping length. The shaping rate is $R_{\mathrm{Sh}} = 1.7$ b/Amp, which is the optimal



(a)



(b)

Fig. 19. Required SNR for post-shaping LDPC code: (a) number of decoding iterations; and (b) code length sweep.

TABLE V
SSFM SIMULATION PARAMETERS

Parameter	Value
Modulation Format	DP 64-QAM
Symbol Rate	28 GBd
Pulse Shaping	Nyquist
Number of Channels	9
Channel Spacing	30 GHz
Nonlinear Coefficient	1.3 /W/km
Dispersion Coefficient	16 ps/nm/km
Fiber Attenuation	0.2 dB/km
Amplification	Lumped (EDFA)
Noise Figure	4 dB
SSFM Step-size	0.1 km
Oversampling Factor	32

rate for considered transmission link. The SNR margin is defined as the difference between received SNR (calculated on received symbols) and required SNR (from Fig. 20). Negative SNR margin means that desired target BER of 10^{-15} cannot be achieved. Maximum performance is observed at finite shaping length, which was also demonstrated in [14], [20]. More importantly, we can observe that the optimal shaping length depends on number of decoding iterations for LDPC

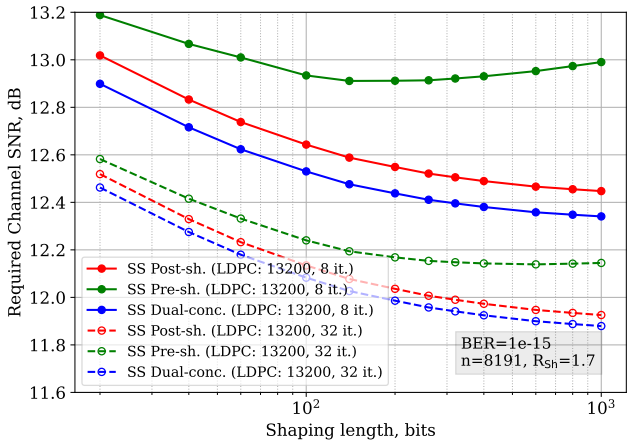


Fig. 20. AWGN Performance of sphere shaping with LDPC codes for a shaping rate of 1.7 b/Amp and a target BER of 10^{-15}. Shaping length sweep.

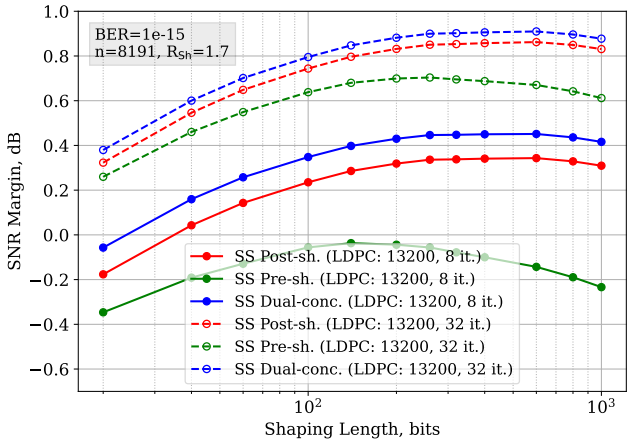


Fig. 21. Fiber-optical channel performance: 9 channels (28 GBaud on 30 GHz grid) 30 spans of 100 km SSMF. Sphere shaping with LDPC codes for shaping rate of 1.7 b/Amp. Shaping length sweep.

code. With lower-complexity LDPC decoding the optimal range is shifted towards shorter shaping length, which is more pronounced for pre-shaping coding case and less pronounced for post-shaping and dual-concatenated coding systems.

VI. DISCUSSION AND CONCLUSIONS

In this work we demonstrated the concept of dual coding concatenation for PAS, which can effectively mitigate the impact of post-shaping burst errors to improve overall system performance. Specifically, dual-concatenation architecture based on parallel BCH codes for pre-shaping and post-shaping FEC layers was proposed and analyzed. This architecture offers flexibility for shaping rate adaptation and potentially reduced implementation complexity.

We studied coding performance in terms of the impact of various shaping and coding parameters. In the case of strong LDPC codes with steep waterfall curve, pre-shaping BCH configuration (without post-shaping BCH) can be beneficial due to lower bit throughput through pre-shaping FEC layer, while the SNR variation in performance is negligible. In the

case of low-complexity LDPC codes, SNR gain of up to 0.25 dB is demonstrated with dual-concatenated configuration. Also, for short-length shaping pre-shaping-only configuration can provide similar performance to that of the standard reverse concatenation. Furthermore, we demonstrated that for mapping/demapping algorithms with reduced BER enhancement pre-shaping BCH configuration may offer even higher improvement. In addition, we note that pre-shaping BCH configuration may further benefit from simplified design of post-shaping FEC. While dual-concatenated and post-shaping BCH configuration require post-LDPC interleaver, pre-shaping configuration does not necessarily require post-LDPC interleaver, leading to a significant reduction in overall latency. These scenarios are the subject for the further analysis.

Finally, we analyzed combined shaping and coding system performance based on sphere shaping in AWGN and fiber-optical channels. For AWGN channel, we demonstrated that maximum performance can be achieved using finite shaping length, while typical assumption is that maximum performance is related to infinite length shaping. For fiber-optical channel, we demonstrated that optimal shaping length range depends on the performance of post-shaping SD FEC code (e.g. LDPC code). Less powerful post-shaping SD code results in optimal shaping length to be shifted towards shorter lengths.

REFERENCES

[1] H. Sun, M. Torbatian, M. Karimi, R. Maher, S. Thomson, M. Tehrani, Y. Gao, A. Kumpera, G. Soliman, A. Kakkar, M. Osman, Z. A. El-Sahn, C. Doggart, W. Hou, S. Sutarwala, Y. Wu, M. R. Chitgarha, V. Lal, H. Tsai, S. Corzine, J. Zhang, J. Osenbach, S. Buggaveeti, Z. Morbi, M. I. Olmedo, I. Leung, X. Xu, P. Samra, V. Dominic, S. Sanders, M. Ziari, A. Napoli, B. Spinnler, K. Wu, and P. Kandappan, "800G DSP ASIC design using probabilistic shaping and digital sub-carrier multiplexing," *IEEE/OSA Journal of Lightwave Technology*, vol. 38, no. 17, pp. 4744–4756, 2020.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] R. F. H. Fischer, *Precoding and signal shaping for digital transmission*. New York, New York, USA: John Wiley & Sons, 2005.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] P. Skvortcov, I. Phillips, W. Forysiak, 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.*, vol. 32, no. 16, pp. 967–970, 2020.

[20] P. Skvortcov, I. Phillips, W. Forysiak, T. Koike-Akino, K. Kojima, K. Parsons, and D. S. Millar, "Huffman-coded sphere shaping for extended-reach single-span links," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 27, no. 3, pp. 1–15, 2021.

[21] T. Yoshida, M. Karlsson, and E. Agrell, "Technologies toward implementation of probabilistic constellation shaping," in *2018 European Conference on Optical Communication (ECOC)*, 2018, pp. 1–3.

[22] P. Skvortcov, T. Koike-Akino, D. S. Millar, K. Kojima, and K. Parsons, "Dual coding concatenation for burst-error correction in probabilistic amplitude shaping," in *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1–4.

[23] OTU-T, "Interfaces for the optical transport network." [Online]. Available: www.itu.int/rec/T-REC-G.709

[24] ——, "Error performance parameters and objectives for multioperator international paths within optical transport networks." [Online]. Available: www.itu.int/rec/T-REC-G.8201

[25] T. Koike-Akino, D. S. Millar, K. Kojima, K. Parsons, Y. Miyata, K. Sugihara, and W. Matsumoto, "Iteration-aware LDPC code design for low-power optical communications," *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 573–581, 2016.

[26] X.-Y. Hu, E. Eleftheriou, and D.-M. Arnold, "Regular and irregular progressive edge-growth tanner graphs," *IEEE transactions on information theory*, vol. 51, no. 1, pp. 386–398, 2005.

[27] D. S. Millar, R. Maher, D. Lavery, T. Koike-Akino, M. Pajovic, A. Alvarado, M. Paskov, K. Kojima, K. Parsons, B. C. Thomsen *et al.*, "Design of a 1 Tb/s superchannel coherent receiver," *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1453–1463, 2016.

[28] R. Gallager, "The random coding bound is tight for the average code," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 244–246, 1973.

[29] Y. Domb, R. Zamir, and M. Feder, "The random coding bound is tight for the average linear code or lattice," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 121–130, 2015.