

# Rate-Adaptive Protograph MacKay-Neal Codes

Ayman Zahr, Balazs Matuz, Gianluigi Liva

**Abstract**—A class of rate-adaptive protograph MacKay-Neal (MN) codes is introduced and analyzed. The code construction employs an outer distribution matcher (DM) to adapt the rate of the scheme. The DM is coupled with an inner protograph-based low-density parity-check (LDPC) code, whose base matrix is optimized via density evolution analysis to approach the Shannon limit of the binary-input additive white Gaussian noise (biAWGN) channel over a given range of code rates. The density evolution analysis is complemented by finite-length simulations, and by a study of the error floor performance.

## I. INTRODUCTION

MacKay-Neal (MN) codes were introduced in [1, Sec. VI] as a class of error correcting codes based on sparse matrices for nonuniform sources. They can be seen as multi-edge-type low-density parity-check (LDPC) codes [2], [3] defined by an extended bipartite graph with a set of variable nodes (VNs) associated with the source bits, and the remaining VNs associated with the codeword bits. LDPC code constructions closely related to MN codes were proposed in [4], [5] for joint source and channel coding. While originally introduced to deal with nonuniform sources, it was pointed out in [1] that MN can be employed also with uniform sources, by introducing a nonlinear block code that turns the uniform source output sequence into a sequence with a prescribed distribution. In [1] it was also recognized that this option may be appealing from a rate flexibility viewpoint, since the rate of the scheme may be modified (e.g., adapted to varying channel conditions) by changing the statistics of the sequences produced by the nonlinear block encoder, hence without modifying the underlying LDPC code. This approach is interesting from two viewpoints. First, the possibility of adapting the rate without changing the underlying LDPC encoder/decoder is important to limit the implementation complexity (modern communication standards that seek rate flexibility often specify a large number of LDPC code parity-check matrices to serve different rates, see e.g. [6]). Second, a rate-adaptive scheme based on MN codes allows to keep the block length constant, allowing the introduction of periodic synchronization markers, with benefits in terms of frame synchronization performance. Despite these appealing properties, MN codes as a means to achieve rate flexibility received little attention. A notable exception is the probabilistic amplitude shaping (PAS) scheme introduced in [7], where a construction reminiscent of MN codes was proposed. In PAS, the sequence output by a uniform (binary) source is also processed by the nonlinear block encoder,

referred to as distribution matcher (DM) [7], generating a sequence of amplitude symbols with a given empirical distribution. The binary labels associated with the amplitude symbols are encoded through a nonsystematic LDPC code encoder, producing a parity bit vector. The amplitude symbols together with the parity bits are then mapped onto pulse amplitude modulation (PAM) symbols. While the main result achieved by PAS is to provide sizable shaping gains, it was quickly recognized that, as a byproduct, PAS is naturally rate-adapting thanks to the possibility of tuning the DM rate [7], as originally hypothesized in [1]. This aspect makes PAS attractive for optical long haul transmission systems [8].

In this paper, we analyze and design a class of rate-adaptive MN codes for the binary-input additive white Gaussian noise (biAWGN) channel. The class is based on protograph LDPC codes [9] coupled with a constant composition distribution matcher (CCDM) [7], [10]. Noting that the concatenation of the CCDM with the inner linear block code results in a nonlinear code, we introduce an equivalent decoding problem. The equivalent problem resorts to the study of the performance of the protograph LDPC code over the communication channel, where side information is provided to the decoder by observing the LDPC encoder input through a binary-input, binary-output channel. By means of protograph extrinsic information transfer (PEXIT) analysis [11], [12], we design rate-adaptive MN codes that are capable of operating close to the biAWGN channel capacity over a wide range of code rates. A quantized density evolution (DE) analysis [13] is used to confirm the accuracy of the thresholds computed via PEXIT analysis. Finite-length performance results are finally included, and complemented by an error floor analysis that (for the designed codes) yields precise estimates of the code performance at large signal-to-noise ratios (SNRs).

## II. PRELIMINARIES

In the following, we denote random variables (r.v.s) by uppercase letters and their realisations by lowercase letters. We denote the probability mass function (p.m.f.) of a discrete r.v.  $X$  as  $P_X(x) = \mathbb{P}[X = x]$ , and the probability density function (p.d.f.) of discrete r.v.  $X$  as  $p_X(x)$ . In either case, the subscript will be dropped whenever the context allows it, i.e.,  $P(x) = P_X(x)$  and  $p(x) = p_X(x)$ . We use  $H_b(\omega)$  to denote the binary entropy function, i.e.,  $H_b(\omega) = -\omega \log_2 \omega - (1-\omega) \log_2 (1-\omega)$  for  $0 < \omega < 1$  and  $H_b(0) = H_b(1) = 0$ . We refer to vectors as row vectors denoted by bold letters, e.g.,  $\mathbf{x}$ , whereas matrices are denoted by uppercase bold letters, e.g.,  $\mathbf{X}$ . We denote by  $\mathbb{F}_2$  the order-2 finite field. We use  $w_H(\mathbf{x})$  and  $d_H(\mathbf{x}, \mathbf{y})$  to be respectively the Hamming weight of a vector  $\mathbf{x}$  and the Hamming distance between two vectors  $\mathbf{x}$  and  $\mathbf{y}$ .

We consider transmission over the biAWGN channel defined by  $Y = X + N$  where  $Y$  is the channel output,  $X \in \{-1, +1\}$

A. Zahr, B. Matuz, and G. Liva are with the Institute of Communications and Navigation, German Aerospace Center (DLR), Wessling, Germany (email: {ayman.zahr,balazs.matuz,gianluigi.liva}@dlr.de).

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany in the programme of "Souverän. Digital. Vernetzt." Joint project 6G-RIC, project identification number: 16KISK022.

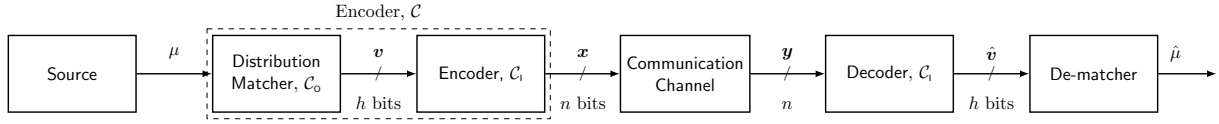


Fig. 1. System model, where a MN code is used to communicate over the biAWGN channel (communication channel).

is the channel input, and where  $N \sim \mathcal{N}(0, \sigma^2)$  is the additive white Gaussian noise term. The channel SNR is defined as  $E_s/N_0 = 1/(2\sigma^2)$  where  $E_s$  is the energy per symbol and  $N_0$  is the single-sided noise power spectral density.

### A. Protograph LDPC codes

A protograph  $\mathcal{P}$  is a small bipartite graph consisting of a set of  $n_0$  VNs, a set of  $m_0$  check nodes (CNs), and  $e$  edges. VNs in the protograph are numbered from 1 to  $n_0$ . Similarly, protograph CNs are numbered from 1 to  $m_0$ . Each VN/CN/edge in a protograph defines a VN/CN/edge type. The bipartite graph  $\mathcal{G}$  of an LDPC code can be derived by lifting the protograph. In particular, the protograph is copied  $\ell$  times (where  $\ell$  is referred to as the lifting factor), and the edges of the protograph copies are permuted under the following constraint: if an edge connects a type- $j$  VN to a type- $i$  CN in  $\mathcal{P}$ , after permutation the edge should connect one of the  $\ell$  type- $j$  VN copies with one of the  $\ell$  type- $i$  CN copies in  $\mathcal{G}$ . We denote by  $\{v_1, v_2, \dots, v_n\}$  the set of  $n = \ell n_0$  VNs in  $\mathcal{G}$ , and by  $\{c_1, c_2, \dots, c_m\}$  the set of  $m = \ell m_0$  CNs. The lifted graph  $\mathcal{G}$  defines the  $m \times n$  parity-check matrix  $\mathbf{H}$ . The base matrix of a protograph is an  $m_0 \times n_0$  matrix  $\mathbf{B} = [b_{i,j}]$  where  $b_{i,j}$  is the number of edges that connect VN  $j$  to CN  $i$  in  $\mathcal{P}$ . We will make use of LDPC codes with *punctured* (or *state*) VNs. A punctured VN is associated with a codeword bit that is not transmitted through the communication channel. We will assume that all the VNs of a given type are either punctured, or they are not.

## III. PROTOGRAPH MACKAY-NEAL CODES

A MN code can be constructed by concatenating an outer nonlinear code  $\mathcal{C}_o$  with an inner code  $\mathcal{C}_i$  [1]. More specifically, we consider the setting depicted in Figure 1. Here, an independent and identically-distributed (i.i.d.) uniform source generates a message  $\mu \in \{1, 2, \dots, M\}$ . The message is input to the encoder of an length- $h$  outer code  $\mathcal{C}_o$  whose task is to generate an output sequence with a prescribed empirical distribution. Following [7], we refer to such a device as the DM. We restrict our attention to DMs based on constant composition (CC) codes which admit low-complexity implementation via arithmetic coding [1], [10]. Let  $\omega \in (0, 1)$  denote the fractional Hamming weight of the  $h$ -bits vector  $\mathbf{v}$ , i.e.,  $\omega = w_h(\mathbf{v})/h$ . We have that  $M = |\mathcal{C}_o| = \binom{h}{\omega h}$ . Hence, the rate of the outer code is  $R_o = h^{-1} \log_2 M = h^{-1} \log_2 \binom{h}{\omega h}$ , which converges to  $H_b(\omega)$  for large  $h$ . The output of the DM is then input to the encoder of an inner  $(n, h)$  binary linear block code  $\mathcal{C}_i$ . The inner code  $\mathcal{C}_i$  is defined by

$$\mathcal{C}_i = \{c | c\mathbf{H}_2^T = \mathbf{v}\mathbf{H}_1^T, \mathbf{v} \in \mathbb{F}_2^h\}$$

where  $\mathbf{H}_1$  is an  $n \times h$  sparse binary matrix, and  $\mathbf{H}_2$  is an  $n \times n$  sparse binary matrix. Note that, strictly speaking,  $\mathcal{C}_i$

may not be an LDPC code, i.e., the code may not possess a sparse parity-check matrix. Nevertheless,  $\mathcal{C}_i$  can be seen as the code obtained by puncturing an  $(n+h, h)$  LDPC code with  $n \times (h+n)$  parity-check matrix

$$\mathbf{H} = [\mathbf{H}_1 | \mathbf{H}_2] \quad (1)$$

where puncturing is applied to the first  $h$  coordinates. We refer to the  $(n+h, h)$  LDPC code with parity-check matrix in the form (1) as the (inner) *mother code*  $\mathcal{C}_{\text{im}}$ . The inner code rate is  $R_i = h/n$ , whereas the mother code rate is  $R_{\text{im}} = h/(n+h) = R_i/(1+R_i)$ . We assume  $\mathbf{H}_2$  to possess rank  $n$ , implying that  $\mathbf{H}_2$  is invertible. We denote by  $\mathbf{G} = \mathbf{H}_1\mathbf{H}_2^{-T}$  the generator matrix of the inner code  $\mathcal{C}_i$ . Note that if the inverse of  $\mathbf{H}_2$  is dense, then the generator matrix  $\mathbf{G}$  is dense. It follows that the concatenation of the outer CC code with the inner linear block code yields a marginal distribution of the codeword bits that is close to uniform, as required by the capacity-achieving input distribution of the biAWGN channel [7]. The overall code obtained by the concatenation of the outer CC code and the inner binary linear block code is denoted by  $\mathcal{C}$ , and its rate is  $R = R_o R_i \approx H_b(\omega) R_i$ . We remark that all rates  $0 < R < R_i$  can be achieved simply by fixing the DM parameter, hence its rate, without performing any modification (e.g., puncturing/shortening) to the inner code.

We consider MN codes based on a protograph mother LDPC code  $\mathcal{C}_{\text{im}}$ . In particular, the mother code parity-check matrix (1) is obtained by lifting a protograph whose  $n_0 \times (h_0 + n_0)$  base matrix takes the form  $\mathbf{B} = [\mathbf{B}_1 | \mathbf{B}_2]$  where  $\mathbf{B}_1$  has dimensions  $n_0 \times h_0$  and  $\mathbf{B}_2$  has dimensions  $n_0 \times n_0$ , where  $n_0 = n/\ell$  and  $h_0 = h/\ell$  are positive integers and where  $\ell$  is the integer lifting factor. Note that all VNs of type  $i$  with  $i = 1, \dots, h_0$  are punctured. An MN code is fully defined by the parity-check matrix of the inner mother code  $\mathcal{C}_{\text{im}}$  and by the DM parameter  $\omega$ . We refer to a MN code *family* as the set of MN codes with fixed mother code, obtained for all  $\omega \in (0, 1)$  that yield an integer  $\omega h$ . Similarly, we refer to a (protograph) MN code ensemble  $\mathcal{E}_\omega(\mathcal{P})$  as the set of MN codes whose inner mother code bipartite graph  $\mathcal{G}$  is obtained by lifting  $\mathcal{P}$ , and where the DM parameter is  $\omega$ . The MN code ensemble family  $\mathcal{E}(\mathcal{P})$  is the set of ensembles  $\{\mathcal{E}_\omega(\mathcal{P})\}_{\omega \in (0,1)}$ .

### A. Belief Propagation Decoding

At the channel input, each codeword  $\mathbf{c}$  is mapped onto  $\{-1, +1\}^n$  via binary antipodal modulation, i.e.,  $x_i = 1 - 2c_i$  for  $i = 1, \dots, n$ . With a slight abuse of notation, we will refer to the modulated codeword  $\mathbf{x}$  as the codeword. We will also use  $\mathcal{C}_i$  and  $\mathcal{C}$  to denote the modulated codebook of the inner code and of the overall code, respectively. We assume next that decoding is performed via belief propagation (BP) applied to the bipartite graph of the mother LDPC code. We initialize the BP decoder as follows. Let us denote by  $L_i$  the  $L$ -value at

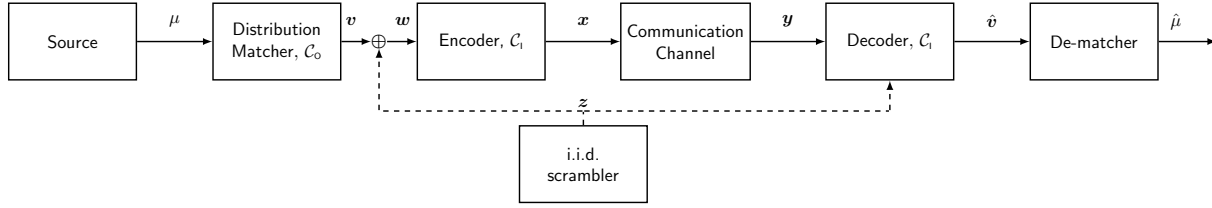


Fig. 2. Modification of the system model of Figure 1, where an i.i.d. scrambler is introduced.

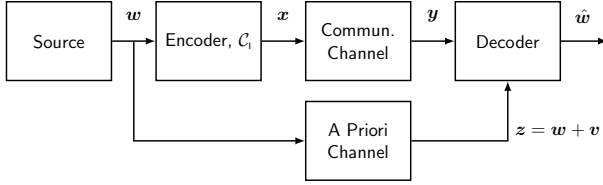


Fig. 3. Equivalent parallel channel model.

the input of the  $i$ th VN. Moreover, define  $\Delta = \ln[(1 - \omega)/\omega]$ . We set  $L_i = \Delta$  if  $i \leq h$ , whereas we set  $L_i = 2y_{i-h}/\sigma^2$  if  $i > h$ . In other words, for the punctured VNs we provide prior information obtained from the marginal distribution of the CC code codeword  $v$ , while for VNs associated with the codeword bits that are transmitted through the biAWGN channel, we input the corresponding channel log-likelihood ratios (LLRs). The BP decoder output  $\hat{v}$  is then processed by the de-matcher [10], producing the estimate  $\hat{\mu}$  of the transmitted message (see Figure 1). It is interesting to note that this decoder (proposed already in [1]) employs the same layered decoding architecture [14] adopted by PAS [7]: the BP decoder does not have any information on the outer CC constraints, and it exploits only the knowledge of the marginal distribution of the bits in  $v$ .

### B. More on Decoding Metrics

We introduce next two decoding approaches which, albeit impractical, will be useful in the analysis of MN codes. We first consider the maximum likelihood (ML) decoder

$$\hat{x}_{\text{ML}} = \arg \max_{x \in \mathcal{C}} p(\mathbf{y}|\mathbf{x}) \quad (2)$$

where  $p(\mathbf{y}|\mathbf{x})$  is the probability density of the biAWGN channel output  $\mathbf{y}$  conditioned on the input  $\mathbf{x}$ . Note that the ML criterion (2) can be rephrased as

$$\hat{x}_{\text{ML}} = \arg \max_{x \in \mathcal{C}_1} p(\mathbf{y}|\mathbf{x}(v))P(v). \quad (3)$$

In (3), the dependency of  $\mathbf{x}$  (inner encoder output) on  $v$  (inner encoder input) is emphasized. Note in particular that the search is over the inner code  $\mathcal{C}_1$ , hence, over an enlarged set compared to (2): the overall code structure is conveyed by the prior  $P(v)$  taking value  $1/|\mathcal{C}_o|$  if  $v \in \mathcal{C}_o$ , and taking value zero otherwise. Following the spirit of the BP decoder, which employs the marginal distribution of the bits composing  $v$  to bias the decoder operating over the bipartite graph of the mother code, we consider also a decoder that outputs

$$\hat{x}_{\text{MM}} = \arg \max_{x \in \mathcal{C}_1} p(\mathbf{y}|\mathbf{x}(v))Q(v). \quad (4)$$

where  $Q(v) = \prod_{i=1}^h P(v_i)$  denotes the product of the marginal distributions of the bits  $v_1, v_2, \dots, v_h$ . The decoding metric adopted in (4) is clearly suboptimal compared with the one of (3). In fact, the term  $Q(v)$  acts as a mismatched prior, yielding a nonzero probability also for  $v \notin \mathcal{C}_o$ . We refer to the decoding metric  $p(\mathbf{y}|\mathbf{x}(v))Q(v)$  as the *mismatched metric*.

### C. Equivalent Parallel Channel Model

The analysis of the scheme described in the previous sections presents some challenges. For example, being the code  $\mathcal{C}$  nonlinear, the block error probability under (2)–(3) will depend upon the transmitted codeword. This hinders the use of a reference codeword to compute bounds on the block error probability. A similar issue arises when attempting a DE evolution analysis under BP decoding, where typically the all-zero codeword is used as reference. The issue can be circumvented by resorting to alternative communication models that can be proved to be equivalent to the one depicted in Figure 1 [5], [15]. Consider first the scheme depicted in Figure 2, where a scrambling block is introduced. The block generates a sequence  $\mathbf{z} = (z_1, z_2, \dots, z_h)$  where each element is picked independently and uniformly at random in  $\{0, 1\}$ . The sequence  $\mathbf{z}$  is then added (in  $\mathbb{F}_2$ ) to  $v$  prior to encoding with  $\mathcal{C}_1$ . The same sequence is made available to the decoder. Considering either (2) or (3), and owing to the symmetry of the biAWGN channel, we observe that the presence of the scrambler is irrelevant to the analysis of the error probability, since the addition of  $\mathbf{z}$  at the transmitter side can be compensated at the decoder by computing first  $\mathbf{b} = \mathbf{z}\mathbf{G}$ , and then flipping the sign of the observations  $y_i$  for all  $i \in \text{supp}(\mathbf{b})$ . The model of Figure 2 admits an equivalent model, provided in Figure 3: here, an i.i.d. uniform binary source generates an  $h$ -bits vector  $\mathbf{w}$ , which is encoded via  $\mathcal{C}_1$  yielding a codeword  $\mathbf{x}$  that is transmitted through the biAWGN channel. The decoder obtains also an observation of  $\mathbf{w}$  via a so-called *a priori channel*. The a priori channel adds (in  $\mathbb{F}_2$ ) a weight- $\omega h$  binary vector  $\mathbf{v}$  to  $\mathbf{w}$ , where  $v$  is picked uniformly at random in  $\mathcal{C}_o$ , resulting in the observation  $\mathbf{z}$ . Upon observing  $\mathbf{y}$  and  $\mathbf{z}$ , the decoder produces a decision on  $\mathbf{w}$ , or, equivalently, a decision on  $v$  since  $\mathbf{w} = \mathbf{v} + \mathbf{z}$ . ML decoding will produce

$$\hat{x}_{\text{ML}} = \arg \max_{x \in \mathcal{C}_1} p(\mathbf{y}|\mathbf{x}(\mathbf{w}))P(\mathbf{z}|\mathbf{w}) \quad (5)$$

where  $P(\mathbf{z}|\mathbf{w}) = 1/|\mathcal{C}_o|$  if  $(\mathbf{z} - \mathbf{w}) \in \mathcal{C}_o$ , and  $P(\mathbf{z}|\mathbf{w}) = 0$  otherwise. The decoding problem can be immediately recognized to be equivalent to the one in (3). The decoder may also resort to a mismatched model for the a priori channel,

treating it as a binary symmetric channel (BSC) with crossover probability  $\omega$  and resulting in

$$\hat{\mathbf{x}}_{\text{MM}} = \arg \max_{\mathbf{x} \in \mathcal{C}_1} p(\mathbf{y}|\mathbf{x}(\mathbf{w}))Q(\mathbf{z}|\mathbf{w}) \quad (6)$$

where  $Q(\mathbf{z}|\mathbf{w}) = \prod_{i=1}^h P(z_i|w_i)$ , i.e., the solution of (6) is equivalent to the solution of (4). We refer to the model of Figure 3 as the *equivalent parallel channel (EPC) model*. The convenience of the EPC model stems from the fact that, owing to the symmetry of the communication and of the a priori channels, and to the linearity of the code  $\mathcal{C}_1$ , the error probability is independent on  $\mathbf{w}$ : we can analyze the error probability of the scheme of Figure 3 by fixing as reference the allzero codeword. The resulting analysis will characterize exactly the performance of the original scheme of Figure 1.

#### IV. DENSITY EVOLUTION ANALYSIS

Protograph-based MN code ensembles can be analyzed via DE by resorting to the EPC model introduced in Section III-C. The analysis shares several commonalities with the DE of LDPC code ensembles designed for joint source and channel coding [5]. We performed the analysis both via quantized DE [13], as well as by means of the Gaussian approximation in the form of PEXIT analysis [11], [12]. In either case, with reference to the EPC model, we replaced the a priori channel (that introduces a constant number of errors  $\omega h$  in  $\mathbf{z}$ ) with a BSC with crossover probability  $\omega$ . The choice is justified by observing that, as  $h$  (and  $n$ ) grows large, the fraction of errors introduced by the BSC concentrates around  $\omega$ . We describe next how to perform the PEXIT analysis, following the steps in [12]. Denote by  $C_{\text{BSC}}(\omega)$  the capacity of a BSC with crossover probability  $\omega$ , and by  $C_{\text{AWGN}}(E_s/N_0)$  the capacity of a biAWGN channel with SNR  $E_s/N_0$ . The PEXIT analysis is initialized by setting the mutual information (MI) at the input of type- $i$  VNs,  $i = 1, \dots, h_0$ , to  $C_{\text{BSC}}(\omega)$ , whereas for  $i = h_0+1, \dots, h_0+n_0$  the MI at the input of type- $i$  VNs is initialized to  $C_{\text{AWGN}}(E_s/N_0)$ . The analysis is then carried out via the recursions provided in [12], and it allows to determine (for fixed  $\omega$ ) the iterative decoding threshold over the biAWGN channel, that is the minimum  $E_s/N_0$  for which the MI values tracked by the PEXIT analysis converge to 1. We denote the threshold value as  $\gamma^*(R)$ , where we emphasize the dependency on  $R$  (and, hence, on  $\omega$ ). Due to the faster computations entailed by the PEXIT analysis with respect to quantized DE, PEXIT analysis will be used for protograph optimization, while quantized DE will be used to verify the accuracy of the thresholds computed for the designed protographs.

The threshold  $\gamma^*$  can be used as target measure to be minimized in the design of protographs. In particular, suppose we are interested in finding an inner mother code protograph that allows to operate close to capacity over a range  $[\underline{R}, \overline{R}]$  of rates  $R$ , i.e., over a range  $[\underline{\omega}, \overline{\omega}]$  of values of  $\omega$ . Fix the protograph parameters  $h_0, n_0$ . By doing so, the inner code rate  $R_1$  is fixed. A set of target rates  $\mathcal{R} \subset [\underline{R}, \overline{R}]$  is selected. For each target rate  $R \in \mathcal{R}$  we can derive the rate of the DM as  $R_0 = R/R_1$ , out of which the DM parameter  $\omega$  is obtained. Given a protograph  $\mathcal{P}$  we define the worst-case loss (WCL)

$$\Delta(\mathcal{P}) := \max_{R \in \mathcal{R}} [\gamma^*(R) - C_{\text{AWGN}}^{-1}(R)] \quad (7)$$

TABLE I  
THRESHOLDS (dB) COMPUTED BY QUANTIZED DE AND BY PEXIT FOR DIFFERENT OVERALL CODE RATE  $R$

$R$	0.5	0.4	0.3	0.2	0.1
$\gamma^*$ dB (quant. DE)	-2.01	-3.38	-5.02	-7.00	-10.32
$\gamma^*$ dB (PEXIT)	-2.06	-3.42	-5.05	-7.14	-10.49

that is the maximum gap between the protograph iterative decoding threshold and the biAWGN Shannon limit for the rates in  $\mathcal{R}$ . A search for the protograph with parameters  $h_0, n_0$  that minimizes the WCL in (7) can be carried out, for example, via differential evolution [16]. We provide next some examples of application to the design of protograph-based MN code ensemble families addressing different rate regimes. In all examples, the search space was limited by setting the maximum number of parallel edges between protograph VN/CN pairs (i.e., value of the base matrix elements) to 3.

**Example 1.** Consider a code rate range  $[0.1, 0.5]$ . An MN code family addressing this range can be derived from an inner  $R_1 = 1/2$  code, i.e., the mother code has rate  $R_{\text{IM}} = 1/3$ . We search for protographs with 6 VNs and 4 CNs, minimizing the WCL over  $\mathcal{R} = \{0.1, 0.3, 0.5\}$ . We obtain the base matrix

$$\mathbf{B}_{1/2} = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 3 & 0 & 1 \\ 2 & 0 & 1 & 1 & 1 & 0 \\ 1 & 2 & 1 & 2 & 0 & 0 \end{pmatrix}.$$

where the first two columns are associated with punctured VNs. The decoding thresholds for the code ensemble defined by  $\mathbf{B}_{1/2}$  are provided in Table I for various rates. The values are computed with both quantized DE and PEXIT analysis. The results confirm the accuracy of the latter, with thresholds that are within 0.05 dB from the quantized DE ones for rates in the ranging from 0.3 to 0.5. The accuracy reduces for the lowest rates, however still yielding acceptable estimates.

**Example 2.** Consider a code rate range  $[0.1, 0.666]$ . An MN code family addressing this range can be derived from an inner  $R_1 = 2/3$  code. Accordingly, fix the base matrix to be a  $3 \times 5$  matrix. We search for protographs minimizing the WCL over  $\mathcal{R} = \{0.1, 0.3, 0.666\}$ . We obtain the base matrix

$$\mathbf{B}_{2/3} = \begin{pmatrix} 1 & 0 & 0 & 3 & 1 \\ 1 & 1 & 0 & 3 & 0 \\ 1 & 2 & 2 & 1 & 0 \end{pmatrix}.$$

Remarkably, the iterative decoding thresholds for both ensembles of Examples 1 and 2 (displayed in Figure 4) are within 1 dB from the Shannon limit over a wide range of rates.

#### V. FINITE LENGTH PERFORMANCE ANALYSIS

In this section, we provide numerical results on finite-length constructions of protograph MN codes. We complement Monte Carlo simulation results with an error floor analysis based on the union bound (UB) on the block error probability. In particular, we analyze mismatched decoding as defined in (6). This choice follows the observation that the BP decoder does

not exploit the joint p.m.f.  $P(\mathbf{v})$ , but rather the marginal distribution  $P(v_i)$  for all  $i = 1, \dots, h$ . By resorting to the EPC setting, the derivation of bounds on the error probability under (5) reduces to the analysis of the error probability under (6). We first derive the pairwise error probability

$$\text{PEP}(\mathbf{x}') = \mathbb{P} \left[ p(\mathbf{Y}|\mathbf{x})Q(\mathbf{Z}|\mathbf{w}) \leq p(\mathbf{Y}|\mathbf{x}')Q(\mathbf{Z}|\mathbf{w}') \right]. \quad (8)$$

In (8), the codeword transmitted over the communication channel is  $\mathbf{x}$ , and it is the result of the encoding of  $\mathbf{w}$ , where the vector  $\mathbf{w}$  is transmitted over the a priori channel. The competing codeword is  $\mathbf{x}'$ , and it is the result of the encoding of  $\mathbf{w}'$ . Note that in (8) ties are broken in favor of the competing codeword. Owing to the symmetry of the communication and a priori channels, and to the linearity of  $\mathcal{C}_i$ , we assume without loss of generality that  $\mathbf{w} = (0, 0, \dots, 0)$  and hence  $\mathbf{x} = (+1, +1, \dots, +1)$ . Conditioned on  $\mathbf{X} = \mathbf{x}$  and  $\mathbf{W} = \mathbf{w}$ ,  $\mathbf{Z}$  is uniformly distributed over the set of  $h$ -bit sequences with Hamming weight  $\omega h$ , whereas  $Y_1, Y_2, \dots, Y_n$  are i.i.d.  $\sim \mathcal{N}(+1, \sigma^2)$ . We can rewrite (8) as

$$\begin{aligned} \text{PEP}(\mathbf{x}') &= \mathbb{P} \left[ \sum_{i=1}^n \ln \frac{p(Y_i|x_i)}{p(Y_i|x'_i)} \leq \sum_{i=1}^h \ln \frac{Q(Z_i|w'_i)}{Q(Z_i|w_i)} \right] \\ &= \mathbb{P} \left[ \sum_{i \in \mathcal{D}(\mathbf{x}')} L_i \leq - \sum_{i \in \mathcal{D}(\mathbf{w}')} T_i \right]. \end{aligned} \quad (9)$$

In (9),  $\mathcal{D}(\mathbf{x}') = \{i | x'_i \neq x_i\}$ ,  $\mathcal{D}(\mathbf{w}') = \{i | w'_i \neq w_i\}$ , whereas  $L_i := \ln[p(Y_i|+1)/p(Y_i|-1)]$  and  $T_i := \ln[p(Z_i|0)/p(Z_i|1)]$ . Denote by  $L = \sum_{i \in \mathcal{D}(\mathbf{x}')} L_i$  and by  $T = \sum_{i \in \mathcal{D}(\mathbf{w}')} T_i$ . Moreover, let  $\delta_1 = d_H(\mathbf{w}, \mathbf{w}')$  and  $\delta_2 = d_H(\mathbf{x}, \mathbf{x}')$ . Conditioned on  $\mathbf{X} = \mathbf{x}$  and  $\mathbf{W} = \mathbf{w}$ , we have that  $L \sim \mathcal{N}(2\delta_2/\sigma^2, 4\delta_2/\sigma^2)$ . Recalling  $\Delta = \ln[(1-\omega)/\omega]$ , we have that  $T_i = \Delta$  if  $Z_i = 0$ , whereas  $T_i = -\Delta$  if  $Z_i = 1$ , i.e.,

$$T = (\delta_1 - E)\Delta - E\Delta = \delta_1 - 2E\Delta$$

where  $E$  follows an hypergeometric distribution with parameters  $(h, \omega h, \delta_1)$ . After a few simple manipulations, we obtain

$$\text{PEP}(\mathbf{x}') = \mathbb{E} \left[ Q \left( \frac{2\delta_2/\sigma^2 + \delta_1 - 2E\Delta}{2\sqrt{\delta_2}/\sigma} \right) \right] \quad (10)$$

where  $Q(x)$  is the well-known Gaussian  $Q$ -function. By observing that  $\text{PEP}(\mathbf{x}')$  depends on  $\mathbf{x}'$  only through its Hamming distance from  $\mathbf{x}$ , and on the Hamming distance between the corresponding information sequence  $\mathbf{w}'$  and  $\mathbf{w}$ , we can upper bound the block error probability under (6) as

$$P_B \leq \sum_{\delta_1=1}^h \sum_{\delta_2=1}^n A_{\delta_1, \delta_2}^{\text{IO}} \mathbb{E} \left[ Q \left( \frac{2\delta_2/\sigma^2 + \delta_1 - 2E\Delta}{2\sqrt{\delta_2}/\sigma} \right) \right] \quad (11)$$

where  $A_{\delta_1, \delta_2}^{\text{IO}}$  is the input-output weight enumerator of  $\mathcal{C}_i$ . At large SNRs, (11) can be approximated by truncating the summation to the dominant term  $A_{\delta_1^*, \delta_2^*}^{\text{IO}}$ , yielding

$$P_B \approx A_{\delta_1^*, \delta_2^*}^{\text{IO}} \mathbb{E} \left[ Q \left( \frac{2\delta_2^*/\sigma^2 + \delta_1^* - 2E\Delta}{2\sqrt{\delta_2^*}/\sigma} \right) \right].$$

A possible strategy to identify the dominant term  $A_{\delta_1^*, \delta_2^*}^{\text{IO}}$  can be to enumerate low-weight codewords of  $\mathcal{C}_i$  (e.g., via the algorithm proposed in [17]). Once a sufficient number of low-weight codewords is collected, their contribution to the error probability can be measured via (10).

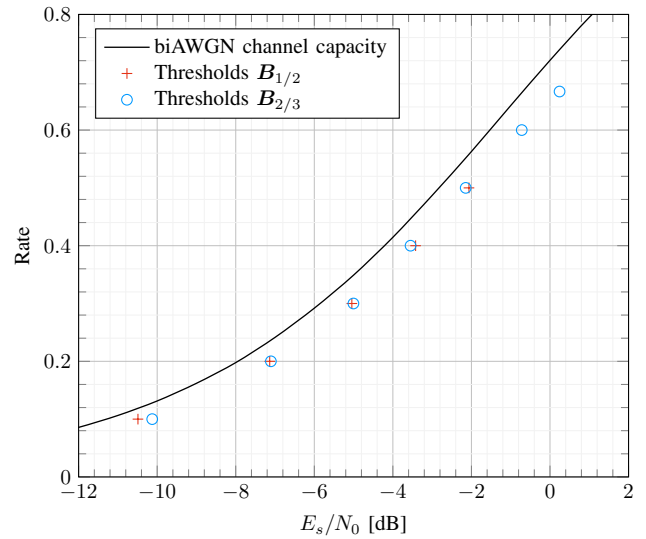


Fig. 4. Iterative decoding thresholds computed for the MN code ensembles defined by the base matrices described in Example 1 and in Example 2.

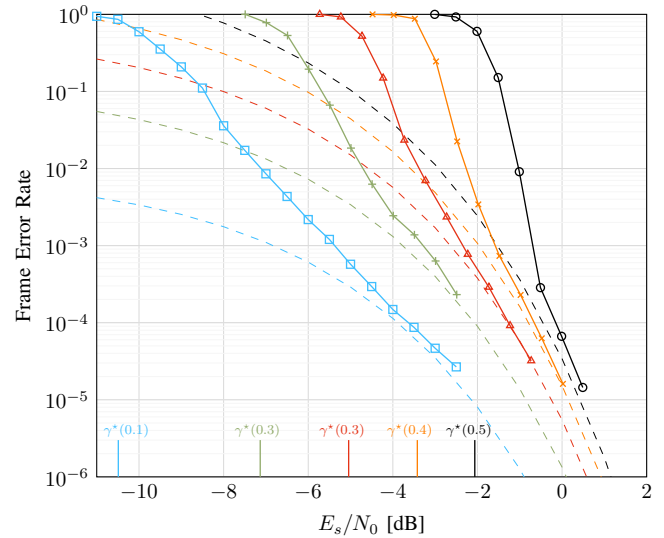


Fig. 5. FER vs.  $E_s/N_0$  dB for length-1200 protograph MN code with inner mother code base matrix  $\mathbf{B}_{1/2}$  and for rates  $R = 0.5$  ( $\bullet$ ),  $R = 0.4$  ( $\blacktriangle$ ),  $R = 0.3$  ( $\blacktriangleright$ ),  $R = 0.2$  ( $\blacktriangleleft$ ), and  $R = 0.1$  ( $\blacksquare$ ). The TUB of the block error probability for each rate is provided (dashed lines) as well as the corresponding iterative decoding thresholds.

Figure 5 reports simulation results for a length-1200 MN code family, for different rates  $R \in [0.1, 0.5]$ . The base matrix of  $\mathcal{C}_{\text{IM}}$  is  $\mathbf{B}_{1/2}$ . The protograph was lifted through a circulant version of the progressive edge growth (PEG) algorithm [18]. The truncated union bounds (TUBs) provide an excellent prediction of the frame error rate (FER) at large SNRs. Interestingly, the TUBs indicate at large SNR a diminishing return in coding gain when the rate of the outer CC is reduced, whereas the coding gains at moderate FER are more sizeable.

#### ACKNOWLEDGMENT

The authors would like to thank Prof. Gerhard Kramer (Technical University of Munich) for his constructive comments and suggestions.

## REFERENCES

- [1] D. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.
- [2] R. Gallager, "Low-density parity-check codes," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 1963.
- [3] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge University Press, 2008.
- [4] M. J. Wainwright and E. Martinian, "Low-density graph codes that are optimal for binning and coding with side information," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1061–1079, Mar. 2009.
- [5] M. Fresia, F. Perez-Cruz, H. V. Poor, and S. Verdú, "Joint source and channel coding," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 104–113, Nov. 2010.
- [6] A. Morello and V. Mignone, "DVB-S2: The second generation standard for satellite broad-band services," *Proc. IEEE*, vol. 94, no. 1, pp. 210–227, Jan. 2006.
- [7] 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, Dec. 2015.
- [8] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *IEEE/OSA J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.
- [9] J. Thorpe, "Low-density parity-check (LDPC) codes constructed from protographs," NASA JPL, IPN Progr. Rep. 42-154, Aug. 2003.
- [10] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.
- [11] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.
- [12] G. Liva and M. Chiani, "Protograph LDPC codes design based on EXIT analysis," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2007.
- [13] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.
- [14] G. Böcherer, "Achievable rates for probabilistic shaping," *arXiv preprint arXiv:1707.01134*, 2017.
- [15] A. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 1, pp. 2–10, Jan. 1974.
- [16] A. Shokrollahi and R. Storn, "Design of efficient erasure codes with differential evolution," in *Differential Evolution*. Springer, 2005.
- [17] X.-Y. Hu, M. P. C. Fossorier, and E. Eleftheriou, "On the computation of the minimum distance of low-density parity-check codes," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2004.
- [18] X.-Y. Hu, E. Eleftheriou, and D. Arnold, "Regular and irregular progressive edge-growth Tanner graphs," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 386–398, Jan. 2005.