On the Application of LDPC Codes to Arbitrary Discrete-Memoryless Channels *

Amir Bennatan and David Burshtein

Dept. of Electrical Engineering Systems

Tel Aviv University

Tel Aviv 69978, Israel

Email: abn@eng.tau.ac.il, burstyn@eng.tau.ac.il

Abstract

We discuss three structures of modified low-density parity-check (LDPC) code ensembles designed for transmission over arbitrary discrete memoryless channels. The first structure is based on the well known binary LDPC codes following constructions proposed by Gallager and McEliece, the second is based on LDPC codes of arbitrary (q-ary) alphabets employing modulo-q addition, as presented by Gallager, and the third is based on LDPC codes defined over the field GF(q). All structures are obtained by applying a quantization mapping on a coset LDPC ensemble. We present tools for the analysis of non-binary codes and show that all configurations, under maximum-likelihood decoding, are capable of reliable communication at rates arbitrarily close to channel capacity of any discrete memoryless channel. We discuss practical iterative decoding of our structures and present simulation results for the AWGN channel confirming the effectiveness of the codes.

Index Terms - Belief propagation, Coset codes, q-ary LDPC, Iterative decoding, Low density parity check (LDPC) codes, Turbo Codes.

*IEEE Transactions on Information Theory, volume 50, no. 3, pp. 417-438, March 2004. This research was supported by the Israel Science Foundation, grant no. 22/01–1 and by a fellowship from The Yitzhak and Chaya Weinstein Research Institute for Signal Processing at Tel Aviv University. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Yokohama, Japan, June 2003.

I Introduction

Low-density parity-check codes and iterative decoding algorithms were proposed by Gallager [12] four decades ago, but were relatively ignored until the introduction of Turbo Codes by Berrou *et al.* [2] in 1993. In fact there is a close resemblance between LDPC and Turbo Codes. Both are characterized by a low-density parity-check matrix, and both can be decoded iteratively using similar algorithms. Turbo-like codes have generated a revolution in coding theory, by providing codes that have capacity approaching transmission rates with practical iterative decoding algorithms. Yet Turbo-like codes are usually restricted to single-user discrete-memoryless binary-input symmetric-output channels.

Several methods for the adaptation of Turbo-like codes to more general channel models have been suggested. Wachsmann *et al.* [31] have presented a multilevel coding scheme using Turbo Codes as components. Robertson and Wörts [25] and Divsalar and Pollara [9] have presented Turbo decoding methods that replace binary constituent codes with TCM codes. Kavčić, Ma, Mitzenmacher and Varnica [16], [19] and [29] have suggested a method that involves a concatenation of a set of LDPC codes as outer codes with a matched-information rate (MIR) inner trellis code, which performs the function of shaping, an essential ingredient in approaching capacity [11].

In his book, Gallager [13], described a construction that can be used to achieve capacity under ML decoding, on an arbitrary discrete-memoryless channel, using a uniformly distributed parity check matrix (i.e. each element is set to $\{0, 1\}$ uniformly, and independently of the other elements). In his construction, Gallager used the coset-ensemble, i.e. the ensemble of all codes obtained by adding a fixed vector to each of the codewords of any given code in the original ensemble. Gallager also proposed mapping groups of bits from a binary code into channel symbols. This mapping is required in order to obtain the shaping gain. Coset LDPC codes were also used by Kavčić *et al.* [15] in the context of intersymbol-interference (ISI) channels. McEliece [20] discussed the application of Turbo-like codes to a wider range of channels, and proposed to use Gallager-style mappings. A uniform mapping from n-tuples to a constellation of 2^n signals was used in [28] and [5] in the contexts of multi-input, multi-output (MIMO) fading channels, and interference mitigation at the transmitter, respectively.

In [12] Gallager proposed a generalization of standard binary LDPC codes to arbitrary alphabet codes, along with a practical iterative soft decoding algorithm. Gallager presented an analysis technique on the performance of the code that applies to a symmetric channel. LDPC codes over an arbitrary alphabet were used by Davey and Mackay [8] in order to improve the performance

of standard binary LDPC codes.

In this paper we discuss three structures of modified LDPC code ensembles designed for transmission over arbitrary (not necessarily binary-input) discrete memoryless channels. The first structure, called binary quantized coset (BQC) LDPC, is based on [13] and [20]. The second structure is a modification of [12], that utilizes cosets and a quantization map from the code symbol space to the channel symbol space. This structure, called modulo quantized coset (MQC) LDPC, assumes that the parity-check matrix elements remain binary, although the codewords are defined over a larger alphabet. The third structure is based on LDPC codes defined over Galois fields GF(q) and allows parity-check matrices over the entire range of field elements. This last structure is called GF(q) quantized coset (GQC) LDPC.

We present tools for the analysis of non-binary codes and derive upper bounds on the maximum likelihood decoding error probability of the three code structures. We then show that these configurations are good in the sense that for regular ensembles with sufficiently large connectivity, a typical code in the ensemble enables reliable communication at rates arbitrarily close to channel capacity, on an arbitrary discrete-memoryless channel. Finally, we discuss the practical iterative decoding algorithms of the various code structures and demonstrate their effectiveness.

Independently of our work, Erez and Miller [10] have recently examined the performance, under ML decoding, of standard q-ary LDPC codes over modulo-additive channels, in the context of lattice codes for additive channels. In this case, due to the inherent symmetry of modulo-additive channels, neither cosets nor quantization mapping are required.

Our work is organized as follows: In Section II we discuss the BQC-LDPC codes, in Section III we discuss the MQC-LDPC codes and in Section IV we discuss GQC-LDPC codes. Section V discusses the iterative decoding of these structures. It also presents simulation results and a comparison with existing methods for transmission over the AWGN channel. Section VI concludes the paper.

II Binary Quantized Coset (BQC) Codes

We begin with a formal definition of coset-ensembles. Our definition is slightly different from the one used by Gallager [13] and is more suitable for our analysis (see [26]).

Definition 1 *Let $\mathcal{C}$ be an ensemble of equal block length binary codes. The ensemble* ***coset-$\mathcal{C}$*** *(denoted $\hat{\mathcal{C}}$) is defined as the output of the following two steps:*

1. *Generate the ensemble $\mathcal{C}'$ from $\mathcal{C}$ by including, for each code C in $\mathcal{C}$, codes generated by all*

possible permutations σ on the order of bits in the codewords.

2. *Generate ensemble $\hat{\mathcal{C}}$ from $\mathcal{C}'$ by including, for each code C' in $\mathcal{C}'$, codes $C'_{\mathbf{v}}$ of the form $\{\mathbf{c}' \oplus \mathbf{v} | \mathbf{c}' \in C'\}$ ($\oplus$ denotes bitwise modulo-2 addition) for all possible vectors $\mathbf{v} \in \{0,1\}^N$.*

Given a code $C'_{\mathbf{v}} \in \hat{\mathcal{C}}$, we refer to $\mathbf{v}$ as the *coset vector* and C as the *underlying* code.

Note: Step 1 can be dropped when generating the coset-ensemble of LDPC codes, because LDPC code ensembles already contain all codes obtained from a permutation of the order of bits.

We proceed to formally define the concept of *quantization*.

Definition 2 *Let $Q(\cdot)$ be a rational probability assignment of the form $Q(a) = N_a/2^T$, defined over the set $a \in \{0, ..., A-1\}$. A **quantization** $\delta(\mathbf{x}) = \delta_Q(\mathbf{x})$ associated with $Q(a)$, of length T, is a mapping from the set of vectors $\mathbf{u} \in \{0,1\}^T$ to $\{0, ..., A-1\}$ such that the number of vectors mapped to each $a \in \{0, ..., A-1\}$ is $2^T Q(a)$.*

Note that in the above definition, the mapping of vectors to channel symbols is in general not one-to-one. Hence the name "quantization". Applying the quantization to a vector $\mathbf{u}$ of length NT is performed by breaking the vector into N sub-vectors of T bits each, and applying $\delta(\mathbf{x})$ to each of the sub-vectors. That is:

$$\delta(\mathbf{u}) \triangleq < \delta(u_1, ..., u_T), \delta(u_{T+1}, ..., u_{2T}), ..., \delta(u_{(N-1)T+1}, ..., u_{NT}) > \tag{1}$$

The quantization of a code is the code obtained from applying the quantization to each of the codewords. Likewise, the quantization of an ensemble is an ensemble obtained by applying the quantization to each of the ensemble's codes. A BQC ensemble is obtained from a binary code ensemble by applying a quantization to the corresponding coset-ensemble.

It is useful to model BQC-LDPC encoding as a sequence of operations, as shown in Figure 1. An incoming message is encoded into a codeword of the underlying LDPC code C. The vector $\mathbf{v}$ is then added, and the quantization mapping applied. Finally, the resulting codeword is transmitted over the channel.

We now introduce the following notation, which is a generalization of similar notation taken from literature covering binary codes. For convenience, we formulate our results for the discrete output case. The conversion to continuous output is immediate.

$$D_{\mathbf{x}} \triangleq \frac{1}{2^T} \sum_{\boldsymbol{\xi} \in \{0,1\}^T} \sum_{y} \sqrt{\Pr[y|\delta(\boldsymbol{\xi})] \Pr[y|\delta(\boldsymbol{\xi} \oplus \mathbf{x})]} \tag{2}$$

where $P[y|u]$ denotes the transition probabilities of a given channel. Using the Cauchy-Schwartz inequality, it is easy to verify that $D_{\mathbf{0}} = 1$ and $D_{\mathbf{x}} \leq 1$ for all $\mathbf{x} \neq \mathbf{0}$. This last inequality becomes

strict for non-degenerate quantizations and channels, as explained in Appendix A.1. We assume these requirements of non-degeneracy throughout the paper.

We now define

$$D \triangleq \max_{\mathbf{x}\in\{0,1\}^T,\mathbf{x}\neq\mathbf{0}} D_{\mathbf{x}} \tag{3}$$

Analysis of the ML decoding properties of BQC-LDPC codes is based on the following theorem, which relates the decoding error to the underlying code's spectrum.

Theorem 1 *Let $p[y|u]$ be the transition probability assignment of a discrete memoryless channel with an input alphabet $\{0,...,A-1\}$. Let $R < C$ be a rate below channel capacity (the rate being measured in bits per channel use), $Q(a)$ an arbitrary distribution, and $\delta(\mathbf{x})$ a quantization associated with $Q(a)$ of length T. Let $\mathcal{C}$ be an ensemble of linear binary codes of length NT and rate at least R/T, and $\overline{S}_l$ the ensemble average number of words of weight l in $\mathcal{C}$. Let $\hat{\mathcal{D}}$ be the BQC ensemble corresponding to $\mathcal{C}$.*

Let $U \subset \{1,...,NT\}$ be an arbitrary set of integers. The ensemble average error under ML decoding of $\hat{\mathcal{D}}$, $\overline{P}_e$, satisfies:

$$\overline{P}_e \leq \sum_{l\in U} \overline{S}_l D^{l/T} + 2^{-NE_Q(R+\log\alpha/N)} \tag{4}$$

where D is given by (3), $E_Q(R)$ is the random coding exponent as defined by Gallager in [13] for the input distribution $Q(a)$

$$\begin{aligned} E_Q(R) &\triangleq \max_{0\leq\rho\leq 1}\{E_0(\rho,Q)-\rho R\} \\ E_0(\rho,Q) &\triangleq -\log\sum_y\left[\sum_{u=0}^{A-1} Q(u)P(y|u)^{1/(1+\rho)}\right]^{1+\rho} \end{aligned} \tag{5}$$

and α is given by:

$$\alpha = \max_{l\in U^c} \frac{\overline{S}_l}{(M-1)\binom{NT}{l}2^{-NT}} \tag{6}$$

$U^c \triangleq \{1,...,NT\}\backslash U$, and $M = 2^{NR}$.

The proof of this theorem is similar to the proof of Theorem 3 and to proofs provided in [26] and [21]. A sketch of the proof is provided in Appendix A.2. The proof of Theorem 3 is provided in detail in Appendix B.1.

We now proceed to examine BQC-LDPC codes, constructed based on the ensemble of (c,d)-regular binary LDPC codes. It is convenient to define a (c,d)-regular binary LDPC code of length

N by means of a bipartite (c, d)-regular Tanner graph [27]. The graph has N *variable* (left) nodes, corresponding to codeword bits, and L *check* (right) nodes. A word $\mathbf{x}$ is a codeword if at each check node, the modulo-2 sum of all the bits at its adjacent variable nodes is zero.

The following method, due to Luby *et al.* [18], is used to define the ensemble of (c, d)-regular binary LDPC codes. Each variable node is assigned c *sockets*, and each check node is assigned d sockets. The Nc variable sockets are matched (that is, connected using graph edges) to $Ld(= Nc)$ check sockets by means of randomly selected permutation of $\{1, 2, ..., Nc\}$. The ensemble of (N, c, d) LDPC codes consists of all the codes constructed from all possible graph constellations.

The mapping from a bipartite graph to the parity-check matrix is performed by setting each matrix element $A_{i,j}$, corresponding to the ith check node and the jth variable node, to the number of edges connecting the two nodes modulo-2 (this definition is designed to account for the rare occurrence of parallel edges between two nodes). The *design rate* of a (c, d) regular LDPC code is defined as $R = 1 - L/N = 1 - c/d$. This value is a lower bound on the true rate of the code, measured in bits per channel use.

In [21], it is shown there exists $0 < \gamma < 1$ such that the minimum distance $d_{\min}$ of a randomly selected code of a (c, d)-regular, NT-length, binary LDPC ensemble satisfies:

$$\lim_{N\to\infty} \Pr[d_{\min} \leq \gamma \cdot NT] = 0 \tag{7}$$

The meaning of the above observation is that the performance of the ensemble is cluttered by a small number of codes (at worst). Removing these codes from the ensemble, we obtain an *expurgated* ensemble with improved properties.

Formally, given an ensemble $\mathcal{C}$ of (c, d)-regular LDPC codes of length NT and an arbitrary number $\delta > 0$, we define the expurgated ensemble $\mathcal{C}^x$ as the ensemble obtained by removing all codes of minimum-distance less than or equal to $\delta \cdot NT$ from $\mathcal{C}$. If $\delta \leq \gamma$ and N is large enough, this does not reduce the size of the ensemble by a factor greater than 2. Thus the expurgated ensemble average spectrum satisfies:

$$\begin{aligned} \overline{S}_l^x &= 0, \quad 0 < l \leq \delta \cdot NT \\ \overline{S}_l^x &\leq 2\overline{S}_l, \quad l > \delta \cdot NT \end{aligned} \tag{8}$$

We refer to the BQC ensemble corresponding to an expurgated (c, d)-regular ensemble as an expurgated BQC-LDPC ensemble. We now use the above construction in the following theorem:

Theorem 2 *Let $p[y|u]$ be the transition probability assignment of a discrete memoryless channel with an input alphabet $\{0, ..., A-1\}$. Let R be some rational positive number, and let $Q(a)$, T, $\delta(\mathbf{x})$ and N be defined as in Theorem 1. Let $\epsilon_1, \epsilon_2 > 0$ be two arbitrary numbers. Then for d and N large enough there exists a (c,d)-regular expurgated BQC-LDPC ensemble $\hat{\mathcal{D}}^x$ (containing all but a diminishingly small proportion of the (c,d)-regular BQC-LDPC codes, as discussed above) of length N and design rate R bits per channel use (the rate of each code in the ensemble is at least R), satisfying $R/T = 1 - c/d$, such that the ML decoding error probability $\overline{P}_e^x$ over $\hat{\mathcal{D}}^x$ satisfies:*

$$\overline{P}_e^x \leq 2^{-NE_Q(R+\epsilon_1)} + D^{(1-\epsilon_2)N} \tag{9}$$

The proof of this theorem relies on results from [21] and is provided in Appendix A.3.

Gallager [13] defined the exponent $E_r(R)$ to be the maximum, for each R, of $E_Q(R)$, evaluated over all possible input distributions $Q(a)$. Let $Q_R(a)$ be the distribution that attains this maximum for a given R. By the nature of the quantization concept, we are restricted to input distributions $Q(a)$ that assign rational probabilities of the form $N_a/2^T$ to each $a \in \{0, ..., A-1\}$. Nevertheless, by selecting $Q(a)$ to approximate $Q_R(a)$, we obtain $E_Q(R)$ which approaches $E_r(R)$ equivalently close. $E_r(R) > 0$ for all $R < C$ and thus, by appropriately selecting $Q(a)$ (recalling that $D < 1$), we obtain BQC-LDPC codes that are capable of reliable transmission (under ML decoding) at any rate below capacity. Furthermore, we have that the ensemble decoding error decreases exponentially with N. At rates approaching capacity, $E_Q(R)$ approaches zero and hence (9) is dominated by the random-coding error exponent.

Figure 2 compares our bound on the threshold of several BQC-LDPC code ensembles, as a function of the SNR, over the AWGN channel. The quantizations used are

$$\begin{aligned}
\delta_1 &= [-2, -1.5, -1, -0.5, -0.5, -0.5, 0, 0, 0, 0, 0.5, 0.5, 0.5, 1, 1.5, 2], \quad T = 4 \\
\delta_2 &= [-1.92, -1.49, -1.07, -0.64, -0.64, -0.213, -0.213, -0.213, 0.213, 0.213, 0.213, 0.64, \\
&\quad 0.64, 1.07, 1.49, 1.92], \quad T = 4 \\
\delta_3 &= [-2.08, -1.84, -1.59, -1.35, -1.1, -0.858, -0.858, -0.613, -0.613, -0.368, -0.368, \\
&\quad -0.368, -0.123, -0.123, -0.123, -0.123, 0.123, 0.123, 0.123, 0.123, 0.368, 0.368, 0.368, \\
&\quad 0.613, 0.613, 0.858, 0.858, 1.1, 1.35, 1.59, 1.84, 2.08], \quad T = 5 \\
\delta_4 &= [\text{32 uniformly spaced values ascending from -1.68 to 1.68 }], \quad T = 5
\end{aligned} \tag{10}$$

The values of $\delta(\mathbf{x})$ are listed in ascending binary order, e.g. $[\delta(0000), \delta(0001), ..., \delta(1111)]$ (for the case of $T = 4$). The quantizations were chosen such that the mean power constraint is satisfied. Note that quantization δ_4 corresponds to equal-spaced 32-PAM transmission, effectively

representing transmission without quantization. Thus, the gap between codes $(12, 30, \delta_3)$ and $(12, 30, \delta_4)$ illustrates the importance of quantization.

To obtain the threshold we use Theorem 1, employing methods similar to the ones discussed in Section V of [21]. For a given (c, d)-regular code and distribution $Q(a)$ we seek the minimal SNR that satisfies the following (for simplicity we assume that d is even, and therefore $\overline{S}_l = \overline{S}_{NT-l}$). We first determine the maximal value of δ such that $\overline{S}_{\delta NT} D^{\delta N} \to 0$, we then define $U = \{1, \ldots, \delta NT\} \cup \{(1-\delta)NT, \ldots, NT\}$, and require that $E_Q(R + \log \alpha / N)$ be positive (the maximum point R yielding positive $E_Q(R)$ is evaluated using the mutual information as described in [13][Section 5.6]).

III Modulo-q Quantized Coset (MQC) LDPC Codes

The definition of (c, d)-regular, modulo-q LDPC codes is adapted from its binary equivalent (provided in Section II) and is a slightly modified version of Gallager's definition [12][Chapter 5]. Bipartite (c, d)-regular graphs are defined and constructed in the same way as in Section II, although variable nodes are associated with q-ary symbols rather than bits. A q-ary word $\mathbf{x}$ is a codeword if at each check node, the modulo-q sum of all symbols at its adjacent variable nodes is zero. The mapping from a bipartite graph to the parity-check matrix is performed by setting each matrix element $A_{i,j}$, corresponding to the ith check node and the jth variable node, to the number of edges connecting the two nodes modulo-q (we thus occasionally obtain matrix elements of a nonbinary alphabet).

As in the binary case, the design rate of a (c, d) regular modulo-q LDPC code is defined as $R = 1 - L/N = 1 - c/d$. Assuming that q is prime, this value is a lower bound on the true rate of the code, measured in q-ary symbols per channel use. The prime q assumption does not pose a problem, as the theorems presented in this section generally require q to be prime.

We proceed by giving the formal definitions of coset-ensembles and quantization for arbitrary-alphabet codes. Note that the definitions used here are slightly different to the ones used with BQC codes.

Definition 3 *Let $\mathcal{C}$ be an ensemble of equal block length codes over the alphabet $\{0, ..., q-1\}$. The ensemble **coset-$\mathcal{C}$** (denoted $\hat{\mathcal{C}}$) is generated from $\mathcal{C}$ by adding, for each code C in $\mathcal{C}$, codes $C_{\mathbf{v}}$ of the form $\{\mathbf{c} + \mathbf{v} | \mathbf{c} \in C\}$ for all possible vectors $\mathbf{v} \in \{0, ..., q-1\}^N$.*

Definition 4 *Let $Q(\cdot)$ be a rational probability assignment of the form $Q(a) = N_a/q$, defined over the set $a \in \{0, ..., A-1\}$. A **quantization** $\delta(x) = \delta_Q(x)$ associated with $Q(a)$ is a mapping from*

a set $\{0,...,q-1\}$ *to* $\{0,...,A-1\}$ *where the number of elements mapped to each* $a \in \{0,...,A-1\}$ *is* $q \cdot Q(a)$.

A quantization is applied to a vector by applying the above mapping to each of its elements. As in the binary case, the quantization of a code is the code obtained from applying the quantization to each of the codewords. The quantization of an ensemble is an ensemble obtained by applying the quantization to each of the ensemble's codes. As with BQC codes, a MQC ensemble is obtained from a modulo-q code ensemble by applying a quantization to the corresponding coset-ensemble.

Analysis of ML decoding of binary-based BQC codes is focused around the weight distribution of codewords. With q-ary based MQC codes, weight is replaced by the concept of *type* (note that in [7][Section 12.1] the type is defined as a normalized value).

Definition 5 *The* ***type*** $\mathbf{t} =< t_0,...,t_{q-1} >$ *of a vector* $\mathbf{v} \in \{0,...,q-1\}^N$ *is a* q*-dimensional vector of integers such that* t_i *is the number of occurrences of the symbol* i *in* $\mathbf{v}$. *We denote the set of all possible types by* $\mathcal{T}$.

The spectrum of a q-ary code is defined in a manner similar to that of binary codes.

Definition 6 *The* ***spectrum*** *of a code* C, *is defined as* $\{S_{\mathbf{t}}\}_{\mathbf{t}\in\mathcal{T}}$, *where* $S_{\mathbf{t}}$ *is the number of words of type* $\mathbf{t}$ *in* C.

We now introduce the notation $\mathbf{D} =< D_0,...,D_{q-1} >$ (not to be confused with the similar definition of $D_{\mathbf{x}}$ in (2)) defined by

$$D_i \triangleq \frac{1}{q}\sum_{k=0}^{q-1}\sum_{y}\sqrt{P[y|\delta(k)]P[y|\delta(k+i)]} \tag{11}$$

where $P[y|u]$ denotes the transition probabilities of a given channel, and $\delta(x)$ is a quantization. Using the Cauchy-Schwartz inequality, it is easy to verify that $D_0 = 1$ and $D_i \leq 1$ for all $i > 0$. As in the case of BQC codes, the last inequality becomes strict for non-degenerate quantizations and channels (defined as in Appendix A.1, replacing 2^T with q).

Given a type $\mathbf{t} \in \mathcal{T}$, we define

$$\mathbf{D}^{\mathbf{t}} = \prod_{i=0}^{q-1} D_i^{t_i}$$

The q-ary (uniform distribution) random coding ensemble is created by randomly selecting each codeword and each codeword symbol independently and with uniform probability. The ensemble average spectrum (average number of codewords of type $\mathbf{t}$) of the random coding ensemble is given by:

$$\overline{S}_{\mathbf{t}} = M \cdot \binom{N}{t_0,...,t_{q-1}} q^{-N}$$

where M is the number of codewords in each code and N is the codeword length.

The importance of the random-coding spectrum is given by the following theorem:

Theorem 3 *Let $p[y|u]$ be the transition probability assignment of a discrete memoryless channel with an input alphabet $\{0, ..., A-1\}$. Let $R < C$ be a rate below channel capacity (the rate being measured in q-ary symbols per channel use), $Q(a)$ an arbitrary distribution and $\delta(x)$ a quantization associated with $Q(a)$ over the alphabet $\{0, ..., q-1\}$. Let $\mathcal{C}$ be an ensemble of linear modulo-q codes of length N and rate at least R, and $\overline{S}_{\mathbf{t}}$ the ensemble average of the number of words of type $\mathbf{t}$ in $\mathcal{C}$. Let $\hat{\mathcal{D}}$ be the MQC ensemble corresponding to $\mathcal{C}$.*

Let $U \subset \mathcal{T}$ be a set of types. Then the ensemble average error under ML decoding of $\hat{\mathcal{D}}$, $\overline{P}_e$, satisfies:

$$\overline{P}_e \leq \sum_{\mathbf{t} \in U} \overline{S}_{\mathbf{t}} \mathbf{D}^{\mathbf{t}} + q^{-NE_Q(R+(\log \alpha)/N)} \tag{12}$$

where $\mathbf{D}$ is defined as in (11), $E_Q(R)$ is given by (5) and α is given by:

$$\alpha = \max_{\mathbf{t} \in U^c} \frac{\overline{S}_{\mathbf{t}}}{(M-1)\binom{N}{t_0, ..., t_{q-1}} q^{-N}} \tag{13}$$

$U^c \triangleq \mathcal{T} \backslash \{\mathbf{0}\} \backslash U$, where $\mathbf{0}$ is the type of the all-zeros word, and $M = q^{NR}$.

The proof of this theorem is provided in Appendix B.1.

The *normalized ensemble spectrum* of q-ary codes is defined in a manner similar to that of binary codes:

$$B(\boldsymbol{\theta}) \triangleq \lim_{N \to \infty} \frac{1}{N} \log \overline{S}_{\boldsymbol{\theta} \cdot N} \tag{14}$$

where $\boldsymbol{\theta}$ denotes a q-dimensional vector of rational numbers satisfying $\sum_{i=0}^{q-1} \theta_i = 1$. Throughout the rest of this paper, we adopt the convention that the base of the log function is always q.

The normalized spectrum of the random coding ensemble is given by:

$$\mathcal{R}(\boldsymbol{\theta}) = H(\boldsymbol{\theta}) - (1 - R)$$

where $H(\boldsymbol{\theta})$ denotes the entropy function, and R is the code rate (in q-ary symbols per channel use). The normalized spectrum of modulo-q LDPC codes is given by the following theorem.

Theorem 4 *The asymptotic normalized ensemble spectrum of a (c, d)-regular modulo-q LDPC code over the alphabet $\{0, ..., q-1\}$ is given by:*

$$B(\boldsymbol{\theta}) = (1-c)H(\boldsymbol{\theta}) + (1-R) \log \inf_{sgn(\mathbf{x}) = sgn(\boldsymbol{\theta})} \frac{A(\mathbf{x})}{\mathbf{x}^{d\boldsymbol{\theta}}} \tag{15}$$

where $\mathbf{x}^{d\boldsymbol{\theta}} \triangleq \prod_{i=0}^{q-1} x_i^{d\theta_i}$, *sgn*$(x)$ *is given by:*

$$sgn(x) \triangleq \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases}$$

$sgn(\mathbf{x}) \triangleq < sgn(x_0), ..., sgn(x_{q-1}) >$, *and* $A(\mathbf{x})$ *is given by:*

$$A(\mathbf{x}) = \frac{1}{q} \sum_{l=0}^{q-1} \left(\sum_{i=0}^{q-1} x_i e^{j\frac{2\pi li}{q}} \right)^d \tag{16}$$

The proof of this theorem is provided in Appendix B.3.

We now show that the modulo-q LDPC normalized spectrum is upper-bounded arbitrarily closely by the random-coding normalized spectrum.

Theorem 5 *Let* q *be a prime number, let* $0 < \delta < 1$ *be an arbitrarily chosen number and*

$$J_\delta = \{(\theta_0, ..., \theta_{q-1}) : \quad 0 \le \theta_i \le 1 - \delta \ \forall i, \quad \sum_{i=0}^{q-1} \theta_i = 1\} \tag{17}$$

Let $R < 1$ *be a given rational positive number and* $\mathcal{R}(\boldsymbol{\theta})$ *be the random-coding normalized spectrum corresponding to the rate* R. *Then for any* $\epsilon > 0$ *there exists a number* $d_0 > 0$ *such that:*

$$B(\boldsymbol{\theta}) < \mathcal{R}(\boldsymbol{\theta}) + \epsilon \tag{18}$$

for all $\boldsymbol{\theta} \in J_\delta$, *and all* c, d *satisfying* $d > d_0$ *and* $R = 1 - c/d$.

The proof of this theorem is provided in Appendix B.4.

Figure 3 presents the set J_δ for a ternary alphabet. The x and y axis represent variables (θ_1, θ_2) with θ_0 being implied by the relation $\theta_0 = 1 - \theta_1 - \theta_2$. A triangle outlines the region of valid values for (θ_1, θ_2) (that is $\theta_1, \theta_2 \ge 0$ and $\theta_1 + \theta_2 \le 1$).

Figure 4 presents the normalized spectrums of a ternary $(3, 6)$-regular LDPC code (the upper surface) and of the ternary rate $1/2$ random-coding ensemble (the lower surface). The normalized spectrums are plotted as functions of parameters (θ_1, θ_2), with $\theta_0 = 1 - \theta_1 - \theta_2$.

The above theorem provides uniform convergence only in a subset of the space of valid values of $\boldsymbol{\theta}$. This means that the maximum α (see (13)), evaluated over *all* values of $\boldsymbol{\theta}$, may be large regardless of the values of c and d. The solution to this problem follows in lines analogous to the binary case. We begin with the following lemma.

Lemma 1 *Let* $\mathbf{w}$ *be a word of weight* $l > 0$ *(*l *non-zero elements). We now bound the probability of* $\mathbf{w}$ *being a codeword of a randomly selected code,* C, *of the* (c, d)*-regular modulo-*q *ensemble,* $\mathcal{C}$.

1. *if* $l < \frac{2}{d}N$ *then*

$$\Pr[\mathbf{w} \in C] \leq \binom{L}{\lfloor \frac{lc}{2} \rfloor} \left(\frac{lc}{2L}\right)^{lc} \tag{19}$$

where $L = \frac{c}{d}N$ *is the number of check nodes in the parity-check matrix of a* (c, d)*-regular code and* $\lfloor x \rfloor$ *denotes the largest integer smaller than or equal to* x*.*

2. *For all* l*, assuming prime* q

$$\Pr[\mathbf{w} \in C] \leq (cN + 1)^q q^{N(1-R)\log A(\lambda)} \tag{20}$$

where $\lambda \triangleq l/N$*,* $A(\lambda)$ *is given by*

$$A(\lambda) \triangleq \frac{1}{q}\left\{1 + (q-1)[1 - 2(1-\rho)\lambda(1-\lambda)]^{\frac{d}{2}}\right\} \tag{21}$$

and ρ *is some constant dependent on* q *alone satisfying* $\rho < 1$*.*

The proof of this lemma is provided in Appendix B.5.

We now build on this lemma to examine the probability of there being *any* low-weight words in a randomly selected code for an LDPC ensemble.

Theorem 6 *Let* $R = 1 - c/d$ *be fixed,* $c \geq 3$ *and assume* q *is prime. Then there exists* $\gamma > 0$*, dependent on* R *and* q *alone, such that*

$$\Pr[d_{min} \leq \gamma N] = O(N^{1-c/2}) \tag{22}$$

where d_{min} *is the minimum distance of a randomly selected* (c, d)*-regular modulo-*q *code of length* N [1]*.*

The proof of this theorem is provided in Appendix B.6.

Given an ensemble $\mathcal{C}$ of (c, d)-regular LDPC codes of length N and an arbitrary number $\delta > 0$, we define the expurgated ensemble $\mathcal{C}^x$ as the ensemble obtained by removing all codes of minimum-distance less than or equal to δN from $\mathcal{C}$. If $\delta \leq \gamma$ (where γ is given by Theorem 6) and N is large enough, this does not reduce the size of the ensemble by a factor greater than 2. Thus the expurgated ensemble average spectrum satisfies:

$$\begin{aligned} \overline{S}^x_{\mathbf{t}} &= 0, \quad 0 < \mathrm{wt}(\mathbf{t}) \leq \delta N \\ \overline{S}^x_{\mathbf{t}} &\leq 2\overline{S}_{\mathbf{t}}, \quad \mathrm{wt}(\mathbf{t}) > \delta N \end{aligned} \tag{23}$$

[1] In fact, it can be shown that fixing $R = 1 - c/d$ and letting $c, d \to \infty$, the minimum distance of a randomly selected code is, with high probability, lower bounded by a value arbitrarily close to the Varshamov-Gilbert bound.

where $\mathrm{wt}(\mathbf{t})$ is the number of nonzero elements in a word of type $\mathbf{t}$. That is, $\mathrm{wt}(\mathbf{t}) = \sum_{i=1}^{q-1} t_i = N - t_0$. We refer to the MQC ensemble corresponding to an expurgated (c, d)-regular ensemble as an expurgated MQC-LDPC ensemble.

Theorem 7 *Let $p[y|u]$ be the transition probability assignment of a discrete memoryless channel with an input alphabet $\{0, ..., A-1\}$. Assume q is prime. Let R be some rational positive number, and let $Q(a)$, $\delta(x)$ and N be defined as in Theorem 3. Let $\epsilon_1, \epsilon_2 > 0$ be two arbitrary numbers. Then for d and N large enough there exists a (c, d)-regular expurgated MQC-LDPC ensemble $\hat{\mathcal{D}}^x$ (containing all but a diminishingly small proportion of the (c, d)-regular MQC-LDPC codes, as shown in Theorem 6) of length N and design rate R, satisfying $R = 1 - c/d$, such that the ML decoding error probability $\overline{P}_e^x$ over $\hat{\mathcal{D}}^x$ satisfies:*

$$\overline{P}_e^x \leq q^{-NE_Q(R+\epsilon_1)} + \sum_{k=1}^{q-1} D_k^{(1-\epsilon_2)N} \tag{24}$$

The proof of this theorem is provided in Appendix B.7.

Applying the same arguments as with BQC-LDPC codes, we obtain that MQC-LDPC codes can be designed for reliable transmission (under ML decoding) at any rate below capacity. Furthermore, at rates approaching capacity, equation (24) is dominated by the random-coding error exponent.

Producing bounds for individual MQC-LDPC code ensembles, similar to those provided in Figures 2 and 5, is difficult due to the numerical complexity of evaluating (13), for large q, over sets J_δ given by (17).

Note: In our constructions, we have restricted ourselves to prime values of q. In Appendix B.8 we show that this restriction is necessary, at least for values of q that are multiples of 2.

IV Quantized Coset LDPC Codes over GF(q) (GQC)

We begin by extending the definition of LDPC codes to a finite field GF(q) [2]. In our definition of modulo-q LDPC codes in Section III, the enlargement of the code alphabet size did not extend to the parity-check matrix. The parity-check matrix was designed to contain binary digits alone (with rare exceptions involving parallel edges). We have seen in Appendix B.8 that for nonprime q, this construction results in codes that are bounded away from the random-coding spectrum. The ideas presented in Appendix B.8 are easily extended to ensembles over GF(2^m) for $m > 1$.

[2]Galois fields GF(q) exist for values of q satisfying $q = p^m$ where p is prime and m is an arbitrary positive integer. See [3].

We therefore define the GF(q) parity-check matrix differently, employing elements from the entire GF(q) field.

Bipartite (c, d)-regular graphs for LDPC codes over GF(q) are constructed in the same way as described in Sections II and III, with the following addition: At each edge (v, c), a random, uniformly distributed *label* $g_{v,c} \in \mathrm{GF}(q)\backslash\{0\}$ is selected. A word $\mathbf{x}$ is a codeword if at each check node c the following equation holds:

$$\sum_{v \in \mathcal{N}(c)} g_{v,c} x_v = 0$$

where $\mathcal{N}(c)$ is the set of variable nodes adjacent to c.

The mapping from the bipartite graph to the parity-check matrix proceeds as follows: Element $A_{i,j}$ in the matrix, corresponding to the ith check node and the jth variable node, is set to the GF(q) sum of all labels g_e corresponding to edges connecting the two nodes. As before, the rate of each code is lower bounded by the design rate $R = 1 - L/N = 1 - c/d$ q-ary symbols per channel use.

GF(q) coset-ensembles and quantization mappings are defined in the same way as with modulo-q codes. Thus we obtain GF(q) quantized coset (GQC) code ensembles.

As with MQC codes, analysis of ML decoding properties of GQC LDPC codes involves the types rather that the weights of codewords. Nevertheless, the analysis is simplified using the following lemma:

Lemma 2 *Let $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ be two equal weight words. The probabilities of the words belonging to a randomly selected code C from a GF(q), (c, d)-regular ensemble satisfy:*

$$\Pr[\mathbf{w}^{(1)} \in C] = \Pr[\mathbf{w}^{(2)} \in C]$$

The proof of this lemma relies on the following two observations. First, by the symmetry of the construction of the LDPC ensemble, any reordering of a word's symbols has no effect on the probability of the word belonging to a randomly selected code. Second, if we replace a nonzero symbol a by another a', we can match any code $C \in \mathcal{C}$ with a unique code $C' \in \mathcal{C}$ such that the modified word belongs to C' if and only if the original word belongs to C. The code C' is constructed by modifying the labels on the edges adjacent to the corresponding variable node using $g' = a \cdot g/a'$ (evaluated over GF(q)).

We have now obtained that the probability of a word belonging to a randomly selected code is dependent on its weight alone, and not on its type. We use this fact in the following theorem, to produce a convenient expression for the normalized spectrum of LDPC codes over GF(q).

Theorem 8 *The asymptotic normalized ensemble spectrum of a (c,d)-regular LDPC code over $GF(q)$ is given by:*

$$B(\boldsymbol{\theta}) = H(\boldsymbol{\theta}) - cH(\lambda) - c\lambda\log(q-1) + (1-R)\log \inf_{sgn(x)=sgn(\lambda)} \frac{A(x)}{x^{d\lambda}} \tag{25}$$

where $\lambda = 1-\theta_0$ and $A(x)$ is given by

$$A(x) = \frac{1}{q}\left\{[1+(q-1)x]^d + (q-1)[1-x]^d\right\} \tag{26}$$

Note that in (25) the parameter λ is implied by the relation $\lambda = 1-\theta_0$ and therefore the right hand side of the equation is in fact a function of $\boldsymbol{\theta}$ alone. The proof of the theorem is provided in Appendix C.1.

Theorems 3, 5, 6 and Lemma 1 carry over from MQC-LDPC to GQC-LDPC codes, with minor modifications. In Theorem 3, modulo-q addition is replaced by addition over GF(q). In Theorem 5, the requirement $q > 2$ is added and the definition of J_δ is replaced by

$$J_\delta = \{(\theta_0,...,\theta_{q-1}): \quad 0 \le \theta_0 \le 1-\delta, \quad \theta_i \ge 0 \quad i = 1,...,q-1, \quad \sum_{i=0}^{q-1}\theta_i = 1\} \tag{27}$$

The proof is similar to the case of MQC-LDPC codes, setting $x = 1/(q-1)\cdot\lambda/(1-\lambda)$ in (25) to upper bound $B(\boldsymbol{\theta})$. In Lemma 1, (20) is replaced by

$$\Pr[\mathbf{w} \in C] \le (cN+1)q^{N(1-R)\log\tilde{A}(\lambda)}$$

where $\lambda \triangleq l/N$ and $\tilde{A}(\lambda)$ is given by

$$\tilde{A}(\lambda) \triangleq \frac{1}{q}\left\{1+(q-1)\left[1-\frac{q}{q-1}\lambda\right]^d\right\}$$

Finally, Theorem 6 carries over unchanged from MQC-LDPC codes.

The definition of expurgated GQC-LDPC ensembles is identical to the equivalent MQC-LDPC definition. We now state the main theorem of this section (which is similar to Theorem 7).

Theorem 9 *Let $p[y|u]$ be the transition probability assignment of a discrete memoryless channel with an input alphabet $\{0,...,A-1\}$. Suppose $q > 2$. Let R be some rational positive number, and let $Q(a)$, $\delta(x)$ and N be defined as in Theorem 3. Let $\epsilon > 0$ be an arbitrary number. Then for d and N large enough there exists a (c,d)-regular expurgated GQC-LDPC ensemble $\hat{\mathcal{D}}^x$ (containing all but a diminishingly small proportion of the (c,d)-regular GQC-LDPC codes, as shown in Theorem 6) of length N and design rate R, satisfying $R = 1-c/d$, such that the ML decoding error probability $\overline{P}_e^x$ over $\hat{\mathcal{D}}^x$ satisfies:*

$$\overline{P}_e^x \le q^{-NE_Q(R+\epsilon)} \tag{28}$$

The proof of this theorem follows in the lines of Theorem 7, replacing (78) with

$$U \triangleq \{\mathbf{t} \in \mathcal{T} : 0 < \mathrm{wt}(\mathbf{t}) \leq \delta N\} \tag{29}$$

The difference between (28) and (24) results from the larger span of the set J_δ defined in (27) in comparison with the one defined in Theorem 5. Thus we are able to define U such that the first term in (12) disappears.

Applying the same arguments as with BQC-LDPC codes and MQC-LDPC codes, we obtain that GQC-LDPC codes can be designed for reliable transmission (under ML decoding) at any rate below capacity. Furthermore, the above theorem guarantees that at any rate, the decoding error exponent for GQC-LDPC codes asymptotically approaches the random-coding exponent, thus outperforming our bounds for both the BQC-LDPC and MQC-LDPC ensembles.

Figure 5 compares several GQC-LDPC code ensembles over the AWGN channel. The quantizations used are the same as those used for Figure 2 and are given by (10) (using the representation of elements of GF(2^m) by m-dimensional binary vectors [3]). To obtain these bounds, we again use methods similar to the ones discussed above for BQC-LDPC codes (here we seek the maximum δ such that $\overline{S}_{\boldsymbol{\theta}_N}\mathbf{D}^{\boldsymbol{\theta}_N} \to 0$ for all $\boldsymbol{\theta}$ having $\mathrm{wt}(\boldsymbol{\theta}N) \leq \delta N$ and then define U as in (29)).

V Iterative decoding

Our analysis so far has focused on the desirable properties of the proposed codes under optimal maximum-likelihood (ML) decoding. In this section we demonstrate that the various codes show favorable performance also under practical iterative decoding.

The BQC-LDPC belief propagation decoder is based on the well-known LDPC decoder, with differences involving the addition of *symbol* nodes alongside variable and check nodes, derived from a factor graph representation by McEliece [20].

The decoding process attempts to recover $\mathbf{c}$, the codeword of the underlying LDPC code. Figure 6 presents an example of the bipartite graph for a $(2, 3)$-regular BQC-LDPC code with $T = 3$. Variable nodes and check nodes are defined in a manner similar to that of standard LDPC codes. Symbol nodes correspond to the symbols of the quantized codeword (1). Each symbol node is connected to the T variable nodes that make up the sub-vector that is mapped to the symbol. Decoding consists of alternating *rightbound* and *leftbound* iterations. In a rightbound iteration, messages are sent from variable nodes to symbol nodes and check nodes. In a leftbound iteration, the opposite occurs. Unlike the standard LDPC belief-propagation decoder, the channel output

"resides" at the symbol nodes, rather than at the variable nodes. The use of the coset vector is easily accounted for at the symbol nodes.

The MQC-LDPC and GQC-LDPC belief propagation decoders are modified versions of the belief propagation decoder introduced by Gallager in [12] for arbitrary alphabet LDPC codes, employing q-dimensional vector messages. The modifications, easily implemented at the variable nodes, account for the addition of the coset vector $\mathbf{v}$ at the transmitter and for the use of quantization. Efficient implementation of belief propagation decoding of arbitrary-alphabet LDPC codes is discussed by Davey and MacKay [8] and by Richardson and Urbanke [23]. The ideas in [8] are suggested in the context of LDPC codes over GF(q) but also apply to codes employing modulo-q arithmetic. The method discussed in [23] suggests using the DFT (the m dimensional DFT for codes over GF(p^m)) to reduce complexity. This is particularly useful for codes defined over GF(2^m), since then the multiplications are eliminated.

An important property of quantized coset LDPC codes (BQC, MQC and GQC) is that for any given bipartite graph, the probability of decoding error, averaged over all possible values of the coset vector $\mathbf{v}$, is independent of the transmitted codeword $\mathbf{c}$ of the underlying LDPC code (as observed by Kavčić *et al.* [15] in the context of coset LDPC codes for ISI channels). This facilitates the extension of the density evolution analysis method (see [23]) to quantized coset codes.

BQC-LDPC density evolution evaluates the density of messages originating from symbol nodes using Monte-Carlo simulations. Density evolution for MQC and GQC-LDPC codes is complicated by to the exponential amount of memory required to store the probability densities of q-dimensional messages. A possible solution to this problem is to rely on Monte-Carlo simulations to evolve the densities.

V.1 Simulation Results for BQC-LDPC Codes

In this paper, we have generally confined our discussion to LDPC codes based on regular graphs. Nevertheless, as with standard LDPC codes, the best BQC-LDPC codes under belief-propagation decoding are based on *irregular* graphs as introduced by Luby *et al.* [18]. In this section we therefore focus on irregular codes.

A major element in generating good BQC-LDPC codes is the design of good quantizations. In our analysis of ML decoding, we focused on the probability assignment $Q(a)$ associated with the quantizations. A good probability assignment for memoryless AWGN channels is generally designed to approximate a Gaussian distribution (see [30] and [31]). Iterative decoding, however,

is sensitive to the particular mapping $\delta(\mathbf{x})$.

A key observation in the design of BQC-LDPC quantizations is that the degree of error protection provided to different bits that are mapped to the same channel symbol, is not identical. Useful figures of merit are the following values (adapted from [24][Section III.E], replacing the log-likelihood messages of [24] with plain likelihood messages):

$$\Delta_l \;=\; \frac{1}{2}\mathrm{E}\sqrt{\frac{X^{(l)}}{1-X^{(l)}}} \quad l=1,...,T \tag{30}$$

where $X^{(l)}$ is a random variable corresponding to the leftbound message from the lth position of a symbol node, assuming the transmitted symbol was zero, in the first iteration of belief propagation. It is easy to verify that the following relation holds

$$\begin{aligned}\Delta_l &= \frac{1}{2^T}\sum_y \sqrt{\Gamma_1^l \Gamma_0^l} \\ \Gamma_1^l &= \sum_{\mathbf{v}:v_l=1} \Pr[y \mid \delta(\mathbf{v})], \quad \Gamma_0^l = \sum_{\mathbf{v}:v_l=0} \Pr[y \mid \delta(\mathbf{v})]\end{aligned}$$

Quantizations rendering a low Δ_l for a particular l produce a stronger leftbound message at the corresponding position, at the first iteration. Given a particular distribution $Q(a)$, different quantizations $\delta(\mathbf{x})$ associated with $Q(a)$ produce different sets $\{\Delta_l\}_{l=1}^T$. In keeping with the experience available for standard LDPC codes, simulation results indicate that good quantizations favor an *irregular* set of $\{\Delta_l\}_{l=1}^T$, meaning that some values of the set are allowed to be larger than others. In the design of edge distributions, to further increase irregularity, it is useful to consider not only the fraction λ_i of edges of given left-degree i but rather $\lambda_{i,l}$, accounting for the left variable node's symbol node position.

In this paper we have employed only rudimentary methods of designing BQC-LDPC edge distributions. Table 1 presents the edge distributions of a rate 2 (measured in bits per channel use) BQC-LDPC code for the AWGN channel. The following length $T=4$ quantization was used with 8-PAM signaling ($A=8$). Applying the notation of (10), we obtain that a simple assignment of values in an ascending order renders a good quantization.

$$\begin{aligned}\delta \;=\; &[-1.94, -1.39, -0.832, -0.832, -0.832, -0.277, -0.277, -0.277, \\ &0.277, 0.277, 0.277, 0.832, 0.832, 0.832, 1.39, 1.94]\end{aligned}$$

The SNR threshold (determined by simulations) is 1.1 dB away from the Shannon limit (which is 11.76 dB at rate 2) at a block length of $N_b = 200,000$ bits or $N = 50,000$ symbols and a bit error rate of about 10^{-5}. Decoding typically requires 100-150 iterations to converge. Note that

the ML decoding threshold of a random code, constructed using the distribution $Q(a)$ associated with $\delta(\mathbf{x})$, is 11.99 dB (evaluated using the mutual information as described in [13][Section 5.6]). Thus the gap to the random-coding limit is 0.87 dB. The edge distributions were designed based on a rate 0.5 binary LDPC code suggested by [24], using trial-and-error to design $\{\lambda_{i,l}\}$ while keeping the marginal distribution $\{\lambda_i\}$ fixed.

V.2 Simulation Results for MQC-LDPC Codes

The MQC equivalent of (30) are the values $\{D_i\}_{i=0}^{q-1}$ given by equations (11). As with BQC-LDPC codes, MQC-LDPC codes favor quantizations rendering an irregular set $\{D_i\}_{i=0}^{q-1}$ (as indicated by simulation results).

In [24] and [6], methods for the design of edge distributions are presented that use singleton error probabilities produced by density evolution and iterative application of linear programming. In this work we have used a similar method, replacing the probabilities produced by density evolution with results of Monte-Carlo simulations. An additional improvement was obtained by replacing singleton error probabilities by the functional $Eh(\mathbf{X})$ (h denoting q-ary entropy).

Table 2 presents the edge distributions of a rate 2 (measured again in bits per channel use) MQC-LDPC code for the AWGN channel. The code alphabet size is $q = 17$, and 9-PAM signaling was used ($A = 9$), with the quantization listed below. The values of $\delta(x)$ are listed in ascending order, i.e. $[\delta(0), \delta(1), ..., \delta(16)]$. The quantization values were selected to approximate a Gaussian distribution, and their order was determined by trial-and-error.

$$\begin{aligned}\delta \;=\; & [0, 0.972, 1.46, -0.486, 0.486, 0, -0.972, 0.486, -1.94, 0, -0.972, -0.486, 0.972, \\ & -0.486, 0.486, 1.94, -1.46]\end{aligned}$$

At an SNR 0.9 dB away from the Shannon limit, the symbol error rate was $5.5 \cdot 10^{-5}$ (50 Monte Carlo simulations). The gap to the ML decoding threshold of a random code, constructed using the distribution $Q(a)$ associated with $\delta(\mathbf{x})$, is 0.7 dB. The block length used was $N = 50,000$ symbols or $N_b = 204,373$ bits. Decoding typically required 100-150 iterations to converge.

V.3 Simulation Results for GQC-LDPC Codes

In contrast to BQC-LDPC and MQC-LDPC codes, GQC-LDPC appear resilient to the ordering of values in the quantizations used. This is the result of the use of random labels, which infer a permutation on the rightbound and leftbound messages at check nodes.

Table 3 presents the edge distributions of a rate 2 (measured again in bits per channel use) GQC-LDPC code for the AWGN channel. The methods used to design the codes are the same as those described above for MQC-LDPC codes. The code alphabet size is $q = 16$, and 9-PAM signaling was used ($A = 9$), with the quantization listed below. The values of $\delta(x)$ are listed in ascending order (using the representation of elements of GF(2^m) by m-dimensional binary vectors [3]). The quantization values were selected to approximate a Gaussian distribution.

$$\begin{aligned}\delta \;=\; & [-1.9, -1.42, -0.949, -0.949, -0.475, -0.475, -0.475, 0, 0, 0, \\ & 0.475, 0.475, 0.949, 0.949, 1.42, 1.9]\end{aligned}$$

At an SNR 0.65 dB away from the Shannon limit, the symbol error rate was less than 10^{-5} (100 Monte Carlo simulations). The gap to the ML decoding threshold of a random code, constructed using the distribution $Q(a)$ associated with $\delta(\mathbf{x})$, is 0.4 dB. The block length used was $N = 50,000$ symbols or $N_b = 200,000$ bits. Decoding typically required 100-150 iterations to converge.

V.4 Comparison with existing methods

Wachsmann *et al.* [31], present reliable transmission using multilevel codes with Turbo Code constituent codes, at approximately 1 dB of the Shannon limit at rate 1 bit per dimension. Multilevel coding is similar to BQC quantization mapping, in that multiple bits (each from a separate binary code) are combined to produce channel symbols. However, the division of the code into independent subcodes and the hard decisions performed during multistage decoding, are potentially suboptimal.

Robertson and Wörts [25], using Turbo Codes with TCM constituent codes, report several results that include reliable transmission within 0.8 dB of the Shannon limit at rate 1 bit per dimension.

Kavčić, Ma, Mitzenmacher and Varnica [16], [19] and [29] present reliable transmission at 0.74 dB of the Shannon limit at rate 2.5 bits per dimension. Their method has several similarities to BQC-LDPC codes. As in multilevel schemes, multiple bits from the outer codes are combined and mapped into channel symbols in a manner similar to BQC quantizations. Quantization mapping can be viewed as a special case of a one-state MIR trellis code, having 2^T parallel branches connecting every two subsequent states of the trellis. Furthermore, the irregular LDPC construction method of [19] and [29] is similar to the method of Section V.1. In [19] and [29], the edge distributions for the outer LDPC codes are optimized separately, and the resulting codes are interleaved into one LDPC code. Similarly, BQC-LDPC construction distinguishes

between variable nodes of different symbol node position. However, during the construction of the BQC-LDPC bipartite graph, no distinction is made between variable nodes in the sense that all variable nodes are allowed to connect to the same check nodes (see Section II). This contrasts the interleaved LDPC code of [19] and [29], where variable nodes originating from different subcodes are connected to distinct sets of check nodes.

VI Conclusion

The codes presented in this paper provide a simple approach to the problem of applying LDPC codes to arbitrary discrete memoryless channels. Quantization mapping (based on ideas by Gallager [13] and McEliece [20]) has enabled the adaptation of LDPC codes to channels where the capacity-achieving source distribution is nonuniform. It is thus a valuable method of overcoming the shaping gap to capacity. The addition of a random coset vector (based on ideas by Gallager [13] and Kavčić *et al.* [15]) is a crucial ingredient for rigorous analysis.

Our focus in this paper has been on ML decoding. We have shown that all three code configurations presented, under ML decoding, are capable of approaching capacity arbitrarily close. Our analysis of MQC and GQC codes has relied on generalizations of concepts useful with binary codes, like code spectrum and expurgated ensembles.

We have also demonstrated the performance of practical iterative decoding. The simple structure of the codes presented lends itself to the design of conceptually simple decoders, which are based entirely on the well-established concepts of iterative belief-propagation (e.g. [20], [22], [12]). We have presented simulation results of promising performance within 0.65 dB of the Shannon Limit at rate 2 bits per channel use. We are currently working on an analysis of iterative decoding of quantized coset LDPC codes [1].

Appendix

A Proofs for Section II

A.1 Proof of $D_{\mathbf{x}} < 1$ for $\mathbf{x} \neq \mathbf{0}$ for non-degenerate quantizations and channels

We define a quantization to be non-degenerate if there exists no integer $n > 1$ such that $Q(a) \cdot 2^T$ is an integer multiple of n for all a (such a quantization could be replaced by a simpler quantization over an alphabet of size $2^T/n$ that would equally attain input distribution $Q(a)$). A channel is non-degenerate if there exist no values $a_1, a_2 \in \{0, ..., A-1\}$ such that $\Pr[y \mid a_1] = \Pr[y \mid a_2]$

for all y. By inspecting when the Cauchy-Schwartz inequality becomes an equality, we see that $D_{\mathbf{x}} = 1$ ($\mathbf{x} \neq \mathbf{0}$) if and only if $\Pr[y|\delta(\boldsymbol{\xi})] = \Pr[y|\delta(\boldsymbol{\xi} \oplus \mathbf{x})]$ for all y and $\boldsymbol{\xi}$. Hence, by the channel non-degeneracy assumption, $\delta(\boldsymbol{\xi}) = \delta(\boldsymbol{\xi} \oplus \mathbf{x})$, for all $\boldsymbol{\xi}$. This contradicts the quantization non-degeneracy assumption, since the number of elements in $\{0,1\}^T$ that are mapped to each channel symbol is some integer multiple of the additive order of $\mathbf{x}$ (which is two).

A.2 Sketch of Proof for Theorem 1

The proofs of Theorems 1 and 3 are similar. In this paper we have selected to elaborate on Theorem 3, as the gap between its proof and proofs provided in [26] and [21] is greater. Thus we concentrate here only on the differences between the two.

As in Appendix B.1, we define $\overline{P}^{U}_{e|m,\hat{C}}$ and $\overline{P}^{U}_{e}$ (and equivalently $\overline{P}^{U^c}_{e|m,\hat{C}}$ and $\overline{P}^{U^c}_{e}$) as

$$\begin{aligned}
\overline{P}^{U}_{e|m,\hat{C}} &= \Pr\{\mathbf{y} : \exists m' \neq m : \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_{m'})] \geq \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_{m})], \mathrm{wt}(\hat{\mathbf{c}}_{m'} \oplus \hat{\mathbf{c}}_{m}) \in U \mid m \text{ was transmitted}, \hat{C}\} \\
\overline{P}^{U}_{e} &= \sum_{m} \sum_{\hat{C}} \frac{1}{M} \Pr[\hat{C}] \cdot \overline{P}^{U}_{e|m,\hat{C}}
\end{aligned}$$

where $\hat{C}$ is a code of ensemble $\hat{\mathcal{C}}$, defined as in Definition 1. We obtain (using the union bound):

$$\overline{P}_e \leq \overline{P}^{U}_{e} + \overline{P}^{U^c}_{e} \tag{31}$$

We proceed to bound both elements of the sum:

1. As in Appendix B.1 we obtain:

$$P^{U}_{e|m,\hat{C}} \leq \sum_{m':\mathrm{wt}(\hat{\mathbf{c}}_{m'} \oplus \hat{\mathbf{c}}_{m}) \in U} \sum_{\mathbf{y}} \sqrt{\Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_{m'})] \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_{m})]}$$

   Recalling (1) and defining $\hat{\mathbf{c}}_{m,i} \triangleq < \hat{c}_{m,(i-1)T+1}, ..., \hat{c}_{m,iT} >$ we obtain the following identity:

$$\delta(\hat{\mathbf{c}}_m) = < \delta(\hat{\mathbf{c}}_{m,1}), ..., \delta(\hat{\mathbf{c}}_{m,N}) >$$

   Each element y_i of the vector $\mathbf{y}$ is now dependent only on the transmitted symbol $\delta(\hat{\mathbf{c}}_{m,i})$ at position i. We therefore obtain

$$\begin{aligned}
P^{U}_{e|m,\hat{C}} &\leq \sum_{m':\mathrm{wt}(\hat{\mathbf{c}}_{m'} \oplus \hat{\mathbf{c}}_{m}) \in U} \sum_{\mathbf{y}} \prod_{i=1}^{N} \sqrt{\Pr[y_i|\delta(\hat{\mathbf{c}}_{m',i})] \Pr[y_i|\delta(\hat{\mathbf{c}}_{m,i})]} \\
&= \sum_{m':\mathrm{wt}(\hat{\mathbf{c}}_{m'} \oplus \hat{\mathbf{c}}_{m}) \in U} \prod_{i=1}^{N} \sum_{y} \sqrt{\Pr[y|\delta(\hat{\mathbf{c}}_{m',i})] \Pr[y|\delta(\hat{\mathbf{c}}_{m,i})]}
\end{aligned}$$

Using the above result, we now examine $\overline{P}^U_{e|m,C'}$, the probability of error for a fixed parent code C' (obtained by step 1 of Definition 1), averaged over all possible values of $\mathbf{v}$.

$$\overline{P}^U_{e|m,C'} \leq \sum_{m':\mathrm{wt}(\mathbf{c}'_{m'} \oplus \mathbf{c}'_m) \in U} \prod_{i=1}^{N} E_{\mathbf{v}_i} \sum_y \sqrt{\Pr[y|\delta(\mathbf{c}'_{m',i} \oplus \mathbf{v}_i)] \Pr[y|\delta(\mathbf{c}'_{m,i} \oplus \mathbf{v}_i)]}$$

$\mathbf{v}_i$ and $\mathbf{c}'_{m,i}$ are defined in a manner similar to the definition of $\hat{\mathbf{c}}_{m,i}$. Letting $\tilde{\mathbf{c}}_{m',i} = \mathbf{c}'_{m',i} \oplus \mathbf{c}'_{m,i}$ and $\tilde{\mathbf{v}}_i = \mathbf{v}_i \oplus \mathbf{c}'_{m,i}$ we obtain:

$$\begin{aligned} \overline{P}^U_{e|m,C'} &\leq \sum_{m':\mathrm{wt}(\tilde{\mathbf{c}}_{m'}) \in U} \prod_{i=1}^{N} E_{\tilde{\mathbf{v}}_i} \sum_y \sqrt{\Pr[y|\delta(\tilde{\mathbf{v}}_i)] \Pr[y|\delta(\tilde{\mathbf{c}}_{m',i} \oplus \tilde{\mathbf{v}}_i)]} \\ &= \sum_{m':\mathrm{wt}(\tilde{\mathbf{c}}_{m'}) \in U} \prod_{i=1}^{N} D_{\tilde{\mathbf{c}}_{m',i}} \end{aligned} \tag{32}$$

where $D_{\tilde{\mathbf{c}}_{m',i}}$ is defined as in (2). Defining $l_{m'} = \mathrm{wt}(\tilde{\mathbf{c}}_{m'})$ we observe that at least $\lceil l_{m'}/T \rceil$ of the components $\{\tilde{\mathbf{c}}_{m',i}\}$ are not equal to $\mathbf{0}$, and hence, recalling (3),

$$\begin{aligned} \overline{P}^U_{e|m,C'} &\leq \sum_{m':l_{m'} \in U} D^{\lceil l_{m'}/T \rceil} \leq \sum_{m':l_{m'} \in U} D^{l_{m'}/T} \\ &= \sum_{l \in U} S'_l D^{l/T} \end{aligned}$$

where S'_l is the spectrum of code C'. Clearly this bound can equally be applied to $P_{e|C'}$, abandoning the assumption of a fixed m. The average spectrum of ensemble $\mathcal{C}'$ is equal to that of our original ensemble $\mathcal{C}$, and therefore we obtain

$$\overline{P}^U_e = E_{C'}\{P^U_{e|C'}\} \leq \sum_{l \in U} \overline{S}_l D^{l/T} \tag{33}$$

2. The methods used to bound the second element of (31) are similar to those used in Appendix B.1. The bound obtained is

$$\overline{P}^{U^c}_e \leq 2^{-NE_Q(R + \log \alpha/N)} \tag{34}$$

Combining (31), (33) and (34) we obtain our desired result of (4). □

A.3 Proof of Theorem 2

Let γ be defined as in (7). Let δ be some arbitrary number smaller than γ that will be determined later. Let $\mathcal{C}^x$ be the underlying expurgated LDPC ensemble of $\hat{\mathcal{D}}^x$, obtained by expurgating codes with minimum distance less than or equal to δN. We use Theorem 1 to bound $\overline{P}^x_e$.

Defining

$$\begin{aligned} U_0 &= \{1, ..., \delta NT\} \\ U_1 &= \{(1-\delta/2)NT, ..., NT\} \\ U &= U_0 \cup U_1 \end{aligned}$$

we have (using (4))

$$\overline{P}_e^x \leq \sum_{l \in U} \overline{S}_l^x D^{l/T} + 2^{-NE_Q(R+(\log \alpha)/N)} \tag{35}$$

We now examine both elements of the above sum.

1. Given that all codes in the expurgated ensemble have minimum distance greater than δNT, we obtain that $\overline{S}_l^x = 0$ for all $l \in U_0$. We therefore turn to examine U_1.

   A desirable effect of expurgation is that it also reduces the number of words having large l. Assume $\mathbf{c}_1$ and $\mathbf{c}_2$ are two codewords of a code $C^x \in \mathcal{C}^x$, such that their weight satisfies $l \geq (1-\delta/2)NT$. The weight of the codeword $\mathbf{c}_0 \stackrel{\Delta}{=} \mathbf{c}_1 \oplus \mathbf{c}_2$ satisfies $l \leq \delta NT$, contradicting the construction of the expurgated ensemble. Hence, C^x cannot contain more than one codeword having $l \in U_1$. Letting $\{S_l^x\}_{l=1}^{NT}$ denote the spectrum of code C^x, we have

   $$\sum_{l \in U} S_l^x D^{l/T} \leq 1 \cdot D^{(1-\delta/2)NT/T} = D^{(1-\delta/2)N}$$

   Clearly this average carries over to the average spectrum. Selecting δ so that $\delta/2 \leq \epsilon_2$ we obtain:

   $$\sum_{l \in U} \overline{S}_l^x D^{l/T} \leq D^{(1-\epsilon_2)N} \tag{36}$$

2. Examining the proof by [21][Appendix E], we obtain

   $$\frac{\log \alpha}{NT} \leq (1 - \frac{R}{T}) \log \left(1 + (1-2\delta/2)^d\right) + \frac{o(NT)}{NT}$$

   Therefore, there exist positive integers N_0 and d_0 such that for c, d, N satisfying $R = 1 - c/d$, $d > d_0$ and $N > N_0$ the following inequality holds:

   $$\frac{1}{N} \log \alpha < \epsilon_1 \tag{37}$$

Combining (35), (36) and (37) we obtain our desired result of (9). □

B Proofs for Section III

B.1 Proof of Theorem 3

The proof of this theorem is a generalization of the proofs available in [21] and [26], from binary-input symmetric-output to arbitrary channels.

All codes $\hat{D}$ in $\hat{\mathcal{D}}$ have the form $\{\delta(\hat{\mathbf{c}}_m) : \hat{\mathbf{c}}_m \in \hat{C}\}$ for some code $\hat{C}$ of $\hat{\mathcal{C}}$. We can therefore write:

$$\overline{P}_e = \sum_m \sum_{\hat{C}} \frac{1}{M} \Pr[\hat{C}] \cdot \overline{P}_{e|m,\hat{C}}$$

where

$$\overline{P}_{e|m,\hat{C}} = \Pr\{\mathbf{y} : \exists m' \neq m : \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_{m'})] \geq \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_m)] \mid m \text{ was transmitted}, \hat{C}\}$$

We define $\overline{P}^U_{e|m,\hat{C}}$ and $\overline{P}^U_e$ (and equivalently $\overline{P}^{U^c}_{e|m,\hat{C}}$ and $\overline{P}^{U^c}_e$) as

$$\begin{aligned} \overline{P}^U_{e|m,\hat{C}} &= \Pr\{\mathbf{y} : \exists m' \neq m : \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_{m'})] \geq \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_m)], \text{type}(\hat{\mathbf{c}}_{m'} - \hat{\mathbf{c}}_m) \in U \mid m \text{ was transmitted}, \hat{C}\} \\ \overline{P}^U_e &= \sum_m \sum_{\hat{C}} \frac{1}{M} \Pr[\hat{C}] \cdot \overline{P}^U_{e|m,\hat{C}} \end{aligned} \tag{38}$$

and obtain (using the union bound):

$$\overline{P}_e \leq \overline{P}^U_e + \overline{P}^{U^c}_e \tag{39}$$

We now proceed to bound both elements of the sum:

1. We first bound $\overline{P}^U_{e|m,\hat{C}}$ (defined above), the error probability for a fixed index m and code $\hat{C}$. This is the probability that some codeword $\hat{\mathbf{c}}_{m'}$ such that $\text{type}(\hat{\mathbf{c}}_{m'} - \hat{\mathbf{c}}_m) \in U$ is more likely to have produced the observation $\mathbf{y}$ than the true $\hat{\mathbf{c}}_m$.

   $$\begin{aligned} \overline{P}^U_{e|m,\hat{C}} &= \Pr\{\mathbf{y} : \exists m' \neq m : \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_{m'})] \geq \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_m)], \text{type}(\hat{\mathbf{c}}_{m'} - \hat{\mathbf{c}}_m) \in U \mid m \text{ was transmitted}, \hat{C}\} \\ &= \sum_{\mathbf{y}} I(\mathbf{y}) \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_m)] \end{aligned}$$

   where $I(\mathbf{y})$ is defined as follows:

   $$I(\mathbf{y}) = \begin{cases} 1 & \exists m' \neq m : Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_{m'})] \geq \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_m)], \text{type}(\hat{\mathbf{c}}_{m'} - \hat{\mathbf{c}}_m) \in U \\ 0 & \text{otherwise} \end{cases}$$

   Therefore:

   $$\overline{P}^U_{e|m,\hat{C}} \leq \sum_{\mathbf{y}} I(\mathbf{y}) \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_m)] \sum_{m' : \text{type}(\hat{\mathbf{c}}_{m'} - \hat{\mathbf{c}}_m) \in U} \sqrt{\frac{\Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_{m'})]}{\Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_m)]}}$$

$$
\begin{aligned}
&\leq \sum_{\mathbf{y}} \Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_m)] \sum_{m':\text{type}(\hat{\mathbf{c}}_{m'}-\hat{\mathbf{c}}_m)\in U} \sqrt{\frac{\Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_{m'})]}{\Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_m)]}} \\
&= \sum_{m':\text{type}(\hat{\mathbf{c}}_{m'}-\hat{\mathbf{c}}_m)\in U} \sum_{\mathbf{y}} \sqrt{\Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_{m'})]\Pr[\mathbf{y}|\delta(\hat{\mathbf{c}}_m)]} \\
&= \sum_{m':\text{type}(\hat{\mathbf{c}}_{m'}-\hat{\mathbf{c}}_m)\in U} \sum_{\mathbf{y}} \prod_{i=1}^{N} \sqrt{\Pr[y_i|\delta(\hat{c}_{m',i})]\Pr[y_i|\delta(\hat{c}_{m,i})]} \\
&= \sum_{m':\text{type}(\hat{\mathbf{c}}_{m'}-\hat{\mathbf{c}}_m)\in U} \prod_{i=1}^{N} \sum_{y} \sqrt{\Pr[y|\delta(\hat{c}_{m',i})]\Pr[y|\delta(\hat{c}_{m,i})]}
\end{aligned}
$$

Each code $\hat{C}$ has the form $\{\mathbf{c}+\mathbf{v} : \mathbf{c} \in C\}$ for some parent code $C \in \mathcal{C}$ and a vector $\mathbf{v}$. We now use the above result to bound $\overline{P}^U_{e|m,C}$, the probability of error for a fixed parent code C, averaged over all possible values of $\mathbf{v}$.

$$
\begin{aligned}
\overline{P}^U_{e|m,C} &= \sum_{\mathbf{v}} P^U_{e|m,C,\mathbf{v}} \Pr[\mathbf{v}] = E_{\mathbf{v}}\{P^U_{e|m,C,\mathbf{v}}\} \\
&\leq E_{\mathbf{v}}\{ \sum_{m':\text{type}(\mathbf{c}_{m'}-\mathbf{c}_m)\in U} \prod_{i=1}^{N} \sum_{y} \sqrt{\Pr[y|\delta(c_{m',i}+v_i)]\Pr[y|\delta(c_{m,i}+v_i)]} \}
\end{aligned}
$$

The elements of the random vector $\mathbf{v}$ are N i.i.d. random variables $\{v_i\}_{i=1}^N$. Therefore

$$
\overline{P}^U_{e|m,C} \leq \sum_{m':\text{type}(\mathbf{c}_{m'}-\mathbf{c}_m)\in U} \prod_{i=1}^{N} E_{v_i} \sum_{y} \sqrt{\Pr[y|\delta(c_{m',i}+v_i)]\Pr[y|\delta(c_{m,i}+v_i)]}
$$

Letting $\tilde{c}_{m',i} = c_{m',i} - c_{m,i}$ and $\tilde{v}_i = v_i + c_{m,i}$ we obtain:

$$
\begin{aligned}
\overline{P}^U_{e|m,C} &\leq \sum_{m':\text{type}(\tilde{\mathbf{c}}_{m'})\in U} \prod_{i=1}^{N} E_{\tilde{v}_i} \sum_{y} \sqrt{\Pr[y|\delta(\tilde{v}_i)]\Pr[y|\delta(\tilde{c}_{m',i}+\tilde{v}_i)]} \\
&= \sum_{m':\text{type}(\tilde{\mathbf{c}}_{m'})\in U} \prod_{i=1}^{N} D_{\tilde{c}_{m',i}}
\end{aligned}
$$

where $D_{\tilde{c}_{m',i}}$ is defined as in (11). Letting $\mathbf{t}_{m'} = \text{type}(\tilde{\mathbf{c}}_{m'})$, we have:

$$
\begin{aligned}
\overline{P}^U_{e|m,C} &\leq \sum_{m':\mathbf{t}_{m'}\in U} \mathbf{D}^{\mathbf{t}_{m'}} \\
&= \sum_{\mathbf{t}\in U} S_{\mathbf{t}}\mathbf{D}^{\mathbf{t}}
\end{aligned}
$$

Clearly this bound can equally be applied to $P_{e|C}$, abandoning the assumption of a fixed m. Finally, we bound $\overline{P}^U_e$:

$$
\overline{P}^U_e = E_C\{P^U_{e|C}\} \leq \sum_{\mathbf{t}\in U} \overline{S}_{\mathbf{t}}\mathbf{D}^{\mathbf{t}} \tag{40}
$$

2. We now bound $\overline{P}^{U^c}_e$. For this we define a set of auxiliary ensembles.

(a) Ensemble $\mathcal{C}'$ is generated from $\mathcal{C}$ by adding, for each code C in $\mathcal{C}$, the codes generated by all possible permutations on the order of codewords within the code.

(b) Ensemble $\mathcal{C}''$ is generated from $\mathcal{C}'$ by adding, for each code C' in $\mathcal{C}'$, the codes generated by all possible permutations σ on the order of symbols within the code.

(c) Ensemble $\tilde{\mathcal{C}}$ is generated from $\mathcal{C}''$ by adding, for each code C'' in $\mathcal{C}''$, codes $C''_{\mathbf{v}}$ of the form $\{\mathbf{c}'' + \mathbf{v} | \mathbf{c}'' \in C''\}$ for all possible vectors $\mathbf{v} \in \{0, ..., q-1\}^N$.

Examining the above construction, it is easy to verify that any reordering of the steps has no impact on the final result. Therefore ensemble $\tilde{\mathcal{C}}$ can be obtained from $\hat{\mathcal{C}}$ by employing steps 2a and 2b.

We also define ensemble $\tilde{\mathcal{D}}$ as the quantization of $\tilde{\mathcal{C}}$. $\tilde{\mathcal{D}}$ can equivalently be obtained employing the above steps 2a and 2b on ensemble $\hat{\mathcal{D}}$. Therefore $\overline{P}_e^{U^c}$, defined based on (38), is identical to a similarly defined $\tilde{P}_e^{U^c}$, evaluated over ensemble $\tilde{\mathcal{C}}$.

We proceed to examine $\tilde{P}_{e|m}^{U^c}$

$$\begin{aligned}\tilde{P}_{e|m}^{U^c} &= \sum_{\tilde{C}} \Pr[\tilde{C}] \cdot \tilde{P}_{e|m,\tilde{C}}^{U^c} \\ &= \sum_{\mathbf{y}} \sum_{\mathbf{x}_0} \Pr[\tilde{\mathbf{c}}_m = \mathbf{x}_0] \cdot \Pr[\mathbf{y} \mid \delta(\mathbf{x}_0)] \cdot \tilde{P}_{e|m,\mathbf{x}_0,\mathbf{y}}^{U^c} \end{aligned} \tag{41}$$

where $\tilde{P}_{e|m,\mathbf{x}_0,\mathbf{y}}^{U^c}$ is defined as the probability, for a fixed $\mathbf{y}$ and index m, assuming $\tilde{\mathbf{c}}_m = \mathbf{x}_0$, that a codeword of the form $\delta(\tilde{\mathbf{c}}_{m'})$, where $\text{type}(\tilde{\mathbf{c}}_{m'} - \tilde{\mathbf{c}}_m) \in U^c$ should be more likely than the transmitted $\delta(\tilde{\mathbf{c}}_m)$. The random space here consists of a random selection of the code $\tilde{C}$ from $\tilde{\mathcal{C}}$. Therefore:

$$\begin{aligned}\tilde{P}_{e|m,\mathbf{x}_0,\mathbf{y}}^{U^c} &= \Pr\{\exists m' \neq m : \Pr[\mathbf{y}|\delta(\tilde{\mathbf{c}}_{m'})] \geq \Pr[\mathbf{y}|\delta(\tilde{\mathbf{c}}_m)],\ \text{type}(\tilde{\mathbf{c}}_{m'} - \tilde{\mathbf{c}}_m) \in U^c \mid \tilde{\mathbf{c}}_m = \mathbf{x}_0\} \\ &= \Pr\{\exists m' \neq m,\ \exists \mathbf{x}_1 \ :\ \Pr[\mathbf{y}|\delta(\mathbf{x}_1)] \geq \Pr[\mathbf{y}|\delta(\mathbf{x}_0)], \text{type}(\mathbf{x}_1 - \mathbf{x}_0) \in U^c,\ \tilde{\mathbf{c}}_{m'} = \mathbf{x}_1 \mid \tilde{\mathbf{c}}_m = \mathbf{x}_0\}\end{aligned}$$

Letting $0 \leq \rho \leq 1$ be an arbitrary number, we have by the union bound (extended as in [13])

$$\tilde{P}_{e|m,\mathbf{x}_0,\mathbf{y}}^{U^c} \leq \left[\sum_{m' \neq m} \sum_{\substack{\mathbf{x}_1 : \Pr[\mathbf{y}|\delta(\mathbf{x}_1)] \geq \Pr[\mathbf{y}|\delta(\mathbf{x}_0)], \\ \text{type}(\mathbf{x}_1 - \mathbf{x}_0) \in U^c}} \Pr[\tilde{\mathbf{c}}_{m'} = \mathbf{x}_1 \mid \tilde{\mathbf{c}}_m = \mathbf{x}_0] \right]^\rho$$

Employing Lemma 3, which is provided in Appendix B.2, we obtain

$$\tilde{P}_{e|m,\mathbf{x}_0,\mathbf{y}}^{U^c} \leq \left[(M-1) \sum_{\substack{\mathbf{x}_1 : \Pr[\mathbf{y}|\delta(\mathbf{x}_1)] \geq \Pr[\mathbf{y}|\delta(\mathbf{x}_0)], \\ \text{type}(\mathbf{x}_1 - \mathbf{x}_0) \in U^c}} \alpha q^{-N} \right]^\rho$$

$$\leq \left[(M-1)\alpha \sum_{\mathbf{x}_1:\Pr[\mathbf{y}|\delta(\mathbf{x}_1)]\geq\Pr[\mathbf{y}|\delta(\mathbf{x}_0)]} q^{-N} \right]^{\rho}$$
$$= \left[(M-1)\alpha \sum_{\mathbf{u}_1:\Pr[\mathbf{y}|\mathbf{u}_1]\geq\Pr[\mathbf{y}|\delta(\mathbf{x}_0)]} \sum_{\mathbf{x}_1:\delta(\mathbf{x}_1)=\mathbf{u}_1} q^{-N} \right]^{\rho}$$

Letting $Q(\mathbf{u}) \triangleq \prod_{i=1}^{N} Q(u_i)$, we have:

$$\begin{aligned}
\tilde{P}^{U^c}_{e|m,\mathbf{x}_0,\mathbf{y}} &\leq \left[(M-1)\alpha \sum_{\mathbf{u}_1:\Pr[\mathbf{y}|\mathbf{u}_1]\geq\Pr[\mathbf{y}|\delta(\mathbf{x}_0)]} Q(\mathbf{u}_1) \right]^{\rho} \\
&\leq \left[(M-1)\alpha \sum_{\mathbf{u}_1:\Pr[\mathbf{y}|\mathbf{u}_1]\geq\Pr[\mathbf{y}|\delta(\mathbf{x}_0)]} Q(\mathbf{u}_1) \left(\frac{\Pr[\mathbf{y}|\mathbf{u}_1]}{\Pr[\mathbf{y}|\delta(\mathbf{x}_0)]} \right)^{s} \right]^{\rho} \\
&\leq \left[(M-1)\alpha \sum_{\mathbf{u}_1} Q(\mathbf{u}_1) \left(\frac{\Pr[\mathbf{y}|\mathbf{u}_1]}{\Pr[\mathbf{y}|\delta(\mathbf{x}_0)]} \right)^{s} \right]^{\rho}
\end{aligned}$$

where $s > 0$ is some arbitrary number. Letting $\mathbf{u}_0 = \delta(\mathbf{x}_0)$ we have:

$$\tilde{P}^{U^c}_{e|m,\mathbf{x}_0,\mathbf{y}} \leq \left[(M-1)\alpha \sum_{\mathbf{u}_1} Q(\mathbf{u}_1) \left(\frac{\Pr[\mathbf{y}|\mathbf{u}_1]}{\Pr[\mathbf{y}|\mathbf{u}_0]} \right)^{s} \right]^{\rho} \tag{42}$$

Using (41) and (42) we obtain:

$$\tilde{P}^{U^c}_{e|m} \leq \sum_{\mathbf{y}} \sum_{\mathbf{u}_0} \sum_{\mathbf{x}_0:\delta(\mathbf{x}_0)=\mathbf{u}_0} \Pr[\tilde{\mathbf{c}}_m = \mathbf{x}_0] \cdot \Pr[\mathbf{y} \mid \mathbf{u}_0] \cdot \left[(M-1)\alpha \sum_{\mathbf{u}_1} Q(\mathbf{u}_1) \left(\frac{\Pr[\mathbf{y}|\mathbf{u}_1]}{\Pr[\mathbf{y}|\mathbf{u}_0]} \right)^{s} \right]^{\rho}$$

Employing Lemma 3 once more, we obtain

$$\begin{aligned}
\tilde{P}^{U^c}_{e|m} &\leq \alpha^{\rho} \sum_{\mathbf{y}} \sum_{\mathbf{u}_0} \left[\sum_{\mathbf{x}_0:\delta(\mathbf{x}_0)=\mathbf{u}_0} q^{-N} \right] \cdot \Pr[\mathbf{y} \mid \mathbf{u}_0] \cdot \left[(M-1) \sum_{\mathbf{u}_1} Q(\mathbf{u}_1) \left(\frac{\Pr[\mathbf{y}|\mathbf{u}_1]}{\Pr[\mathbf{y}|\mathbf{u}_0]} \right)^{s} \right]^{\rho} \\
&= \alpha^{\rho} \sum_{\mathbf{y}} \sum_{\mathbf{u}_0} Q(\mathbf{u}_0) \cdot \Pr[\mathbf{y} \mid \mathbf{u}_0] \cdot \left[(M-1) \sum_{\mathbf{u}_1} Q(\mathbf{u}_1) \left(\frac{\Pr[\mathbf{y}|\mathbf{u}_1]}{\Pr[\mathbf{y}|\mathbf{u}_0]} \right)^{s} \right]^{\rho}
\end{aligned}$$

The remainder of the proof follows in direct lines as in [13][Theorem 5.6.1]. Abandoning the assumption of a fixed m, and recalling that $\overline{P}^{U^c}_{e} = \tilde{P}^{U^c}_{e}$, we obtain:

$$\overline{P}^{U^c}_{e} \leq q^{-NE_Q(R+\log\alpha/N)} \tag{43}$$

Combining (39), (40) and (43) we obtain our desired result of (12). □

B.2 Statement and proof of Lemma 3

The following lemma is used in the proof of Theorem 3 and relies on the notation introduced there.

Lemma 3 *Let $\mathbf{x}_0, \mathbf{x}_1 \in \{0, ..., q-1\}^N$ and let $1 \leq i, j \leq M$ be two distinct indexes. Let $\tilde{\mathbf{c}}_i$ and $\tilde{\mathbf{c}}_j$ be the respective codewords of a randomly selected code $\tilde{C}$ of $\tilde{\mathcal{C}}$. Then*

1. $\Pr[\tilde{\mathbf{c}}_i = \mathbf{x}_0] = q^{-N}$
2. *If* $\text{type}(\mathbf{x}_1 - \mathbf{x}_0) \in U^c$ *then*

$$\Pr[\tilde{\mathbf{c}}_j = \mathbf{x}_1 | \tilde{\mathbf{c}}_i = \mathbf{x}_0] \leq \alpha q^{-N} \tag{44}$$

where α is defined by (13).

Proof. Let $\tilde{C}$ be a randomly selected code from $\tilde{\mathcal{C}}$. We first fix the ancestor code C and investigate the relevant probabilities. We define the following random variables: C'' denotes the "parent" code from which $\tilde{C}$ was generated. Likewise, C' denotes the parent of C'', and C denotes the parent of C'.

Let $\mathbf{w}$ be an arbitrary nonzero word. We now examine the codewords of C'

$$\Pr[\mathbf{c}'_i - \mathbf{c}'_j = \mathbf{w} \mid C] = \begin{cases} \frac{1}{M_C - 1} & \mathbf{w} \in C \\ 0 & \text{otherwise} \end{cases}$$

where M_C is the number of codewords in C. Recalling that C is of rate at least R, we have that $M_C \geq M$ and hence $1/(M_C - 1) \leq 1/(M-1)$.

We now examine C'':

$$\begin{aligned} \Pr\left[\mathbf{c}''_i - \mathbf{c}''_j = \mathbf{w} \mid C\right] &= \sum_{\sigma} \Pr\left[\mathbf{c}'_i - \mathbf{c}'_j = \sigma^{-1}(\mathbf{w}) \mid C\right] \Pr(\sigma) \\ &\leq \frac{1}{M-1} \Pr\left[\sigma^{-1}(\mathbf{w}) \in C\right] \\ &= \frac{1}{M-1} \cdot \frac{S_{\mathbf{t}}}{\binom{N}{t_0, ..., t_{q-1}}} \end{aligned}$$

where $\mathbf{t}$ is the type of $\mathbf{w}$. The last equation holds because, given a uniform random selection of the symbol-permutation σ, the probability of $\sigma^{-1}(\mathbf{w})$ being a codeword of C is equal to the fraction of $\mathbf{t}$-type C codewords within the entire set of $\mathbf{t}$-type words.

Examining $\tilde{C}$ we have:

$$\Pr[\tilde{\mathbf{c}}_i = \mathbf{x}_0 \mid C] = \sum_{\mathbf{v}} \Pr[\tilde{\mathbf{c}}_i = \mathbf{x}_0 \mid \mathbf{v}, C] \cdot \Pr[\mathbf{v}] = q^{-N} \sum_{\mathbf{v}} \Pr[\mathbf{c}''_i = \mathbf{x}_0 - \mathbf{v} \mid C] = q^{-N}$$

$$\begin{aligned}
\Pr[\tilde{\mathbf{c}}_j = \mathbf{x}_1 \mid \tilde{\mathbf{c}}_i = \mathbf{x}_0, C] &= \frac{\Pr[\tilde{\mathbf{c}}_j = \mathbf{x}_1, \tilde{\mathbf{c}}_i = \mathbf{x}_0 \mid C]}{\Pr[\tilde{\mathbf{c}}_i = \mathbf{x}_0 \mid C]} \\
&= q^N \sum_{\mathbf{v}} q^{-N} \Pr[\tilde{\mathbf{c}}_j = \mathbf{x}_1, \tilde{\mathbf{c}}_i = \mathbf{x}_0 \mid C, \mathbf{v}] \\
&= \sum_{\mathbf{v}} \Pr[\mathbf{c}''_j = \mathbf{x}_1 - \mathbf{v}, \mathbf{c}''_i = \mathbf{x}_0 - \mathbf{v} \mid C] \\
&= \sum_{\mathbf{v}} \Pr[\mathbf{c}''_j - \mathbf{c}''_i = \mathbf{x}_1 - \mathbf{x}_0, \mathbf{c}''_i = \mathbf{x}_0 - \mathbf{v} \mid C] \\
&= \Pr[\mathbf{c}''_j - \mathbf{c}''_i = \mathbf{x}_1 - \mathbf{x}_0 \mid C] \leq \frac{1}{M-1} \cdot \frac{S_{\mathbf{t}}}{\binom{N}{t_0,\ldots,t_{q-1}}}
\end{aligned}$$

where $\mathbf{t}$ is now the type of $\mathbf{x}_1 - \mathbf{x}_0$.

We finally abandon the assumption of a fixed C, and obtain:

$$\begin{aligned}
&\Pr[\tilde{\mathbf{c}}_i = \mathbf{x}_0] = q^{-N} \\
&\Pr[\tilde{\mathbf{c}}_j = \mathbf{x}_1 \mid \tilde{\mathbf{c}}_i = \mathbf{x}_0] \leq \frac{1}{M-1} \cdot \frac{\overline{S}_{\mathbf{t}}}{\binom{N}{t_0,\ldots,t_{q-1}}}
\end{aligned}$$

If $\mathbf{t}$ is in U^c, we obtain from (13) the desired result of (44), and thus complete the proof of the lemma. □

B.3 Proof of Theorem 4

To prove this theorem, we first quote the following notation and theorem, by Burshtein and Miller [4] (see also [14]). Given a multinomial $p(x_1, ..., x_m)$, the coefficient of $\prod_i x_i^{n_i}$ is denoted[3] $\lfloor p(x_1, ..., x_m)_{n_1,...,n_m} \rfloor$. The theorem is stated for the two dimensional case but is equally valid for higher dimensions (we present a truncated version of the theorem actually discussed in [4]).

Theorem 10 *Let $\gamma > 0$ be some rational number and let $p(x, y)$ be a function such that $p(x, y)^\gamma$ is a multinomial with non-negative coefficients. Let $\alpha > 0$ and $\beta > 0$ be some rational numbers and let n_i be the series of all indices j such that j/γ is an integer and $\lfloor p(x, y)^j \rfloor_{\alpha j, \beta j} \neq 0$. Then*

$$\lfloor p(x, y)^{n_i} \rfloor_{\alpha n_i, \beta n_i} \leq \inf_{x>0, y>0} \frac{p(x, y)^{n_i}}{x^{\alpha n_i} y^{\beta n_i}} \tag{45}$$

and

$$\lim_{i \to \infty} \frac{1}{n_i} \log \lfloor p(x, y)^{n_i} \rfloor_{\alpha n_i, \beta n_i} = \log \inf_{x>0, y>0} \frac{p(x, y)}{x^\alpha y^\beta} \tag{46}$$

We now use this theorem to prove Theorem 4. The proof follows in the lines of a similar proof for binary codes in [4].

[3] The same notation $\lfloor x \rfloor$ is used to denote the largest integer smaller than or equal to x. The distinction between the two meanings is to be made based on the context of the discussion.

We first calculate the asymptotic spectrum for $\boldsymbol{\theta}$ satisfying $\theta_i > 0$ for all i.

Given a type $\mathbf{t}$ and $1 \leq j \leq \binom{N}{t_0,\ldots,t_{q-1}}$, let $X_{\mathbf{t}}^j$ be an indicator r.v. equal to 1 if the j-th word of type $\mathbf{t}$ (by some arbitrary ordering of type-$\mathbf{t}$ words) is a codeword of a drawn code, and 0 otherwise. Then

$$S_{\mathbf{t}} = \sum_{j=1}^{\binom{N}{t_0,\ldots,t_{q-1}}} X_{\mathbf{t}}^j$$

Therefore,

$$\overline{S}_{\mathbf{t}} = E \sum_{j=1}^{\binom{N}{t_0,\ldots,t_{q-1}}} X_{\mathbf{t}}^j = \sum_{j=1}^{\binom{N}{t_0,\ldots,t_{q-1}}} EX_{\mathbf{t}}^j = \sum_{j=1}^{\binom{N}{t_0,\ldots,t_{q-1}}} \Pr[X_{\mathbf{t}}^j = 1] = \binom{N}{t_0,\ldots,t_{q-1}} \Pr[X_{\mathbf{t}}^1 = 1] \quad (47)$$

the final equality resulting from the symmetry of the ensemble construction. Thus, for all $\boldsymbol{\theta}$

$$\lim_{N\to\infty} \frac{1}{N} \log \overline{S}_{\boldsymbol{\theta}N} = H(\boldsymbol{\theta}) + \lim_{N\to\infty} \frac{1}{N} \log \Pr[X_{\boldsymbol{\theta}N}^1 = 1] \quad (48)$$

We now examine $Pr[X_{\boldsymbol{\theta}N}^1 = 1]$. Let $\mathbf{w}$ be a word of type $\boldsymbol{\theta}N$. Each assignment of edges by a random permutation infers a symbol "coloring" on each of the check node sockets, using $\theta_i Nc$ colors of type i ($i = 0, \ldots, q-1$). The total number of colorings is:

$$t(\boldsymbol{\theta}; N) = \binom{Nc}{\theta_0 Nc, ..., \theta_{q-1} Nc} \quad (49)$$

Given an assignment of graph edges, $\mathbf{w}$ is a codeword if at each check node, the modulo-q sum of all adjacent variable nodes is zero. The number of valid assignments (assignments rendering $\mathbf{w}$ a codeword), is determined using an enumerating function:

$$e(\boldsymbol{\theta}; N) = \lfloor [A(x_0, ..., x_{q-1})]^{(1-R)N} \rfloor_{\boldsymbol{\theta}Nc} \quad (50)$$

where $A(\mathbf{x})$ is given by

$$A(\mathbf{x}) = \sum_{(k_0,\ldots,k_{q-1}) \,:\, 0\leq k_i \leq d, \sum k_i = d, \sum i k_i (\mathrm{mod}\ \ q) = 0} \binom{d}{k_0, ..., k_{q-1}} x_0^{k_0} \cdot \ldots \cdot x_{q-1}^{k_{q-1}}$$

we now have

$$\lim_{N\to\infty} \frac{1}{N} \log \Pr[X_{\boldsymbol{\theta}N}^1 = 1] = \lim_{N\to\infty} \frac{1}{N} \log \frac{e(\boldsymbol{\theta}; N)}{t(\boldsymbol{\theta}; N)} \quad (51)$$

Using (49) we have

$$\lim_{N\to\infty} \frac{1}{N} \log t(\boldsymbol{\theta}; N) = cH(\boldsymbol{\theta}) \quad (52)$$

Using (50) and Theorem 10 we have:

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\log e(\boldsymbol{\theta};N) &= (1-R)\lim_{N\to\infty}\frac{1}{(1-R)N}\log\lfloor[A(x_0,...,x_{q-1})]^{(1-R)N}\rfloor_{\boldsymbol{\theta}cN}\\ &= (1-R)\lim_{N\to\infty}\frac{1}{N}\log\lfloor[A(x_0,...,x_{q-1})]^{N}\rfloor_{\boldsymbol{\theta}dN}\\ &= (1-R)\log\inf_{\mathbf{x}>\mathbf{0}}\frac{A(x_0,...,x_{q-1})}{x_0^{d\theta_0}\cdot...\cdot x_{q-1}^{d\theta_{q-1}}} \qquad (53)\end{aligned}$$

Combining (48), (51), (52) and (53) we obtain:

$$\lim_{N\to\infty}\frac{1}{N}\log\overline{S}_{\boldsymbol{\theta}N}=(1-c)H(\boldsymbol{\theta})+(1-R)\log\inf_{\mathbf{x}>0}\frac{A(\mathbf{x})}{\mathbf{x}^{d\boldsymbol{\theta}}}$$

To adapt the proof to $\boldsymbol{\theta}$ having $\theta_i=0$ for some i, we modify the expression for $e(\boldsymbol{\theta};N)$ as follows. Assuming (without loss of generality) that $\theta_i\neq 0$ for all $i\leq i_0$, and $\theta_i=0$ for all $i>i_0$, we obtain:

$$e(\boldsymbol{\theta};N)=\lfloor[\tilde{A}(x_0,...,x_{i_0})]^{(1-R)N}\rfloor_{(\theta_0,...,\theta_{i_0})\cdot cN}$$

where $\tilde{A}(x_0,...,x_{i_0})$ is given by

$$\begin{aligned}\tilde{A}(x_0,...,x_{i_0}) &= \sum_{(k_0...k_{i_0})\,:\,0\leq k_i\leq d,\sum k_i=d,\sum ik_i(\mathrm{mod}\ \ q)=0}\binom{d}{k_0,...,k_{i_0}}x_0^{k_0}\cdot...\cdot x_{i_0}^{k_{i_0}}\\ &= A(x_0,...,x_{i_0},0...0)\end{aligned}$$

Following in the line of previous development, we obtain:

$$\begin{aligned}B(\boldsymbol{\theta})=\lim_{N\to\infty}\frac{1}{N}\log\overline{S}_{\boldsymbol{\theta}N} &= (1-c)H(\boldsymbol{\theta})+(1-R)\log\inf_{x_o,...,x_{i_0}>0,\ \ x_{i_0+1}=...=x_{q-1}=0}\frac{A(\mathbf{x})}{\mathbf{x}^{d\boldsymbol{\theta}}}\\ &= (1-c)H(\boldsymbol{\theta})+(1-R)\log\inf_{sgn(\mathbf{x})=sgn(\boldsymbol{\theta})}\frac{A(\mathbf{x})}{\mathbf{x}^{d\boldsymbol{\theta}}}\end{aligned}$$

We now obtain (16). The development follows in the lines of a similar development in [12][Theorem 5.1]. Given $\mathbf{x}$, we define for $l=0,...,q-1$

$$\mathbf{x}^l=<x_0e^{j\frac{2\pi l}{q}\cdot 0},...,x_{q-1}e^{j\frac{2\pi l}{q}\cdot(q-1)}>$$

and

$$\begin{aligned}B(\mathbf{x}) &= \left(\sum_{i=0}^{q-1}x_i\right)^d\\ &= \sum_{(k_0,...,k_{q-1})\,:\,0\leq k_i\leq d,\sum k_i=d}\binom{d}{k_0,...,k_{q-1}}x_0^{k_0}\cdot...\cdot x_{q-1}^{k_{q-1}} \qquad (54)\end{aligned}$$

therefore

$$B(\mathbf{x}^l) = \sum_{(k_0...,k_{q-1})\ :\ 0\le k_i\le d, \sum k_i = d} \binom{d}{k_0,...,k_{q-1}} x_0^{k_0} \cdot ... \cdot x_{q-1}^{k_{q-1}} e^{2\pi j \frac{\sum i k_i}{q} \cdot l}$$

and

$$\begin{aligned}
\frac{1}{q}\sum_{l=0}^{q-1} B(\mathbf{x}^l) &= \sum_{(k_0,...,k_{q-1})\ :\ 0\le k_i\le d, \sum k_i = d} \binom{d}{k_0,...,k_{q-1}} x_0^{k_0} \cdot ... \cdot x_{q-1}^{k_{q-1}} \left(\frac{1}{q}\sum_{l=0}^{q-1} e^{2\pi j \frac{\sum i k_i}{q} \cdot l} \right) \\
&= \sum_{(k_0,...,k_{q-1})\ :\ 0\le k_i\le d, \sum k_i = d, \sum i k_i (\mathrm{mod}\ q) = 0} \binom{d}{k_0,...,k_{q-1}} x_0^{k_0} \cdot ... \cdot x_{q-1}^{k_{q-1}} \\
&= A(\mathbf{x}) \qquad (55)
\end{aligned}$$

Combining (54) with (55) we obtain (16). □

B.4 Proof of Theorem 5

From (15) we obtain, selecting $\mathbf{x} = \boldsymbol{\theta}$,

$$\begin{aligned}
B(\boldsymbol{\theta}) &\le (1-c)H(\boldsymbol{\theta}) + (1-R)\log \frac{A(\boldsymbol{\theta})}{\boldsymbol{\theta}^d \boldsymbol{\theta}} \\
&= H(\boldsymbol{\theta}) + (1-R)\log A(\boldsymbol{\theta}) \qquad (56)
\end{aligned}$$

We now bound $A(\mathbf{x})$ for $\mathbf{x} \in J_\delta$ (J_δ being defined in (17)). We introduce the following standard notation, borrowed from literature covering the DFT.

$$W_q \triangleq e^{j\frac{2\pi}{q}} \qquad (57)$$

$$X_k \triangleq \sum_{l=0}^{q-1} x_l W_q^{lk} \qquad (58)$$

Incorporating this notation into (16), we now have:

$$A(\mathbf{x}) = \frac{1}{q}\sum_{k=0}^{q-1} X_k^d \qquad (59)$$

Let $\mathbf{x} \in J_\delta$.

- Using $\sum x_i = 1$ we have that there exists at least one $s \in \{0, ..., q-1\}$ satisfying $x_s \ge \frac{1}{q}$.
- Using $x_s \le 1 - \delta$, we obtain that there exists at least one $t \ne s, t \in \{0, ..., q-1\}$ satisfying $x_t \ge \frac{\delta}{q}$.

We now bound X_k for all $k = 0, ..., q-1$.

$$X_0 = \sum_{l=0}^{q-1} x_l W_q^{l0} = \sum_{l=0}^{q-1} x_l = 1 \tag{60}$$

For $k > 0$:

$$\begin{aligned} |X_k| &= |\sum_{l=0}^{q-1} x_l W_q^{lk}| \\ &\leq \sum_{l=0,\ l\neq s,t}^{q-1} |x_l W_q^{lk}| + |x_s W_q^{sk} + x_t W_q^{tk}| \end{aligned} \tag{61}$$

We separately treat both elements of the above sum:

$$\sum_{l=0,\ l\neq s,t}^{q-1} |x_l W_q^{lk}| = \sum_{l=0,\ l\neq s,t}^{q-1} x_l = 1 - x_s - x_t \tag{62}$$

$$|x_s W_q^{sk} + x_t W_q^{tk}|^2 = |x_s W_q^{sk}|^2 + |x_t W_q^{tk}|^2 + 2x_s x_t \mathrm{Re}\, W_q^{(t-s)k}$$

Using $0 < k < q$, $-q < t - s < q$ and $t - s \neq 0$ and using the fact that q is prime, we obtain that q and $(t-s)k$ have no common divisors. Therefore, $W_q^{(t-s)k} \neq 1$. Defining ρ as

$$\rho = \max_{l=1,...,q-1} \mathrm{Re}\, W_q^l \tag{63}$$

we have:

$$\mathrm{Re}\, W_q^{(t-s)k} \leq \rho$$

therefore

$$\begin{aligned} |x_s W_q^{sk} + x_t W_q^{tk}|^2 &\leq x_s^2 + x_t^2 + 2x_s x_t \rho \\ &= (x_s + x_t)^2 - 2(1-\rho)x_s x_t \\ &= (x_s + x_t)^2 [1 - \frac{2(1-\rho)x_s x_t}{(x_s + x_t)^2}] \\ &\leq (x_s + x_t)^2 [1 - \frac{2(1-\rho)\frac{1}{q}\frac{\delta}{q}}{1}] \end{aligned}$$

Taking the square root of both sides of the above equation, we obtain:

$$|x_s W_q^{sk} + x_t W_q^{tk}| \leq (x_s + x_t)\varphi \tag{64}$$

where φ is some positive constant smaller than 1, dependent on q and δ but independent of $\mathbf{x}$ and d.

Combining (61), (62), (64) we obtain:

$$\begin{aligned} |X_k| &\leq 1 - x_s - x_t + \varphi(x_s + x_t) \\ &= 1 - (1-\varphi)(x_s + x_t) \\ &\leq 1 - (1-\varphi)(\frac{1}{q} + \frac{\delta}{q}) \\ &\leq \psi \end{aligned} \tag{65}$$

where ψ, like φ, is some positive constant smaller than 1, dependent on q and δ but independent of $\mathbf{x}$ and d.

Recalling (59)

$$|A(\mathbf{x}) - \frac{X_0^d}{q}| = |\frac{1}{q}\sum_{k=1}^{q-1} X_k^d|$$

using (60) and (65) we obtain:

$$|A(\mathbf{x}) - \frac{1}{q}| \leq \frac{1}{q}\sum_{k=1}^{q-1} \psi^d$$

and therefore $A(\mathbf{x})$ approaches $\frac{1}{q}$ with d uniformly on J_δ. We now obtain:

$$\lim_{d\to\infty}[H(\boldsymbol{\theta}) + (1-R)\log A(\boldsymbol{\theta})] = H(\boldsymbol{\theta}) - (1-R) = \mathcal{R}(\boldsymbol{\theta})$$

where the above bound is obtained uniformly over $\boldsymbol{\theta} \in J_\delta$. Finally, combining this result with (56) we obtain our desired result of (18). □

B.5 Proof of Lemma 1

1. The proof of (19) is similar to the proof of Lemma 2 of [21]. We concentrate here only on the differences between the two, resulting from the enlargement of the alphabet size. The mapping between a bipartite graph and a matrix $B_{d\times L}$, defined in [21], is extended so that an element $B_{i,j}$ of the matrix is set to 1 if the corresponding left-vertex (variable node) is *nonzero*, and to 0 otherwise. Therefore, although our alphabet size q is in general larger than 2, we still restrict the ensemble of matrices $\{B\}$ to binary values.

   Unlike [21], applying the above mapping to a valid q-ary codeword does not necessarily produce a matrix B whose columns are all of an even weight. Therefore, the requirement that lc be even does not hold. However, the matrix B must still satisfy the condition that

each of its populated columns have a weight of at least 2. Hence the number of populated columns cannot exceed $\lfloor lc/2 \rfloor$. The remainder of the proof follows in direct lines as in [21].

2. We now turn to the proof of (20). Let $\boldsymbol{\theta}N$ be the type of $\mathbf{w}$. As we have seen in Appendix B.3, equations (49) and (50)

$$\frac{1}{N}\log \Pr[\mathbf{w} \in C] = \frac{1}{N}\log \frac{\lfloor [A(x_0,...,x_{q-1})]^{(1-R)N} \rfloor_{\boldsymbol{\theta}cN}}{\binom{Nc}{\theta_0 Nc,...,\theta_{q-1}Nc}} \tag{66}$$

To bound the numerator, we use methods similar to those used to obtain (53), applying the bound (45) rather than the limit of equation (46). We obtain

$$\frac{1}{N}\log \lfloor [A(x_0,...,x_{q-1})]^{(1-R)N} \rfloor_{\boldsymbol{\theta}cN} \quad \leq \quad (1-R)\log \inf_{sgn(\mathbf{x})=sgn(\boldsymbol{\theta})} \frac{A(\mathbf{x})}{\mathbf{x}^{d\boldsymbol{\theta}}} \tag{67}$$

To lower bound the denominator, we use the well known bound (e.g. [7][Theorem 12.1.3]),

$$\frac{1}{(N+1)^q}\cdot q^{NH(\boldsymbol{\theta})} \leq \binom{N}{\theta_0 N,...,\theta_{q-1}N} \leq q^{NH(\boldsymbol{\theta})} \tag{68}$$

Combining (66), (67) and (68) we obtain:

$$\Pr[\mathbf{w} \in C] \leq (cN+1)^q \cdot q^{N[-cH(\boldsymbol{\theta})+(1-R)\log \inf_{sgn(\mathbf{x})=sgn(\boldsymbol{\theta})} A(\mathbf{x})/\mathbf{x}^{d\boldsymbol{\theta}}]} \tag{69}$$

Using arguments similar to those used to obtain (56), we obtain:

$$\Pr[\mathbf{w} \in C] \leq (cN+1)^q \cdot q^{N[(1-R)\log A(\boldsymbol{\theta})]} \tag{70}$$

Recalling that $\boldsymbol{\theta}N$ is the type of word $\mathbf{w}$ and given that $l = \lambda N$ is the number of nonzero elements in $\mathbf{w}$, we obtain $\theta_0 = 1-\lambda$. We now bound $A(\mathbf{x})$ for values of $\mathbf{x}$ satisfying $x_0 = 1-\lambda$, $x_i \geq 0$, $i = 1,...,q-1$ and $\sum_{i=0}^{q-1} x_i = 1$. To do this, we use methods similar to those used in Appendix B.4. Employing the notation of (57) and (58), we bound X_k for $k = 1,\ldots,q$:

$$\begin{aligned} |X_k|^2 &= |\sum_{m=0}^{q-1} x_m W_q^{mk}|^2 \\ &= \sum_{m=0}^{q-1} x_m^2 + 2\cdot \left(\sum_{m,n=0,...,q-1, n>m} x_m x_n \mathrm{Re}\, W_q^{(n-m)k} \right) \end{aligned}$$

We first examine the elements of the second sum on the right hand of the above equation. As in the proof of Theorem 5 we have $W_q^{(n-m)k} \neq 1$ for all $n > m$. Defining ρ as in (63), we obtain:

$$\mathrm{Re}\, W_q^{(n-m)k} \leq \rho$$

We now have

$$\begin{aligned} |X_k|^2 &\leq \sum_{m=0}^{q-1} x_m^2 + 2\rho \cdot \sum_{m,n=0,\ldots,q-1,n>m} x_m x_n \\ &= \left(\sum_{m=0}^{q-1} x_m\right)^2 - 2(1-\rho) \cdot \sum_{m,n=0,\ldots,q-1,n>m} x_m x_n \end{aligned}$$

The first sum in the last equation is 1. Dropping elements from the second sum can only increase the overall result, and hence we obtain

$$\begin{aligned} |X_k|^2 &\leq 1 - 2(1-\rho) \cdot \sum_{n=1,\ldots,q-1} x_0 x_n \\ &= 1 - 2(1-\rho) x_0 (1 - x_0) \\ &= 1 - 2(1-\rho)(1-\lambda)\lambda \end{aligned} \tag{71}$$

Combining (59), (60) and (71), we obtain

$$A(\boldsymbol{\theta}) \leq \frac{1}{q}\left\{1 + (q-1)[1 - 2(1-\rho)\lambda(1-\lambda)]^{\frac{d}{2}}\right\} \tag{72}$$

Combining (72) with (70), recalling (21), we obtain our desired result of (20).

□

B.6 Proof of Theorem 6

The proof of this theorem follows in lines similar to those used in the proofs of Theorems 2 and 3 of [21].

Using a union bound, we obtain

$$\begin{aligned} \Pr[d_{min} \leq \gamma N] &= \Pr[\exists \mathbf{w} \in C : \mathrm{wt}(\mathbf{w}) \leq \gamma N] \\ &\leq \Pr[\exists \mathbf{w} \in C : \mathrm{wt}(\mathbf{w}) \leq \frac{\beta}{d} N] + \Pr[\exists \mathbf{w} \in C : \frac{\beta}{d} N \leq \mathrm{wt}(\mathbf{w}) \leq \gamma N] \end{aligned} \tag{73}$$

where β and γ are determined later. We now proceed to bound both elements of the above sum.

1. Requiring $\beta < 2$ and invoking (19) we obtain

$$\Pr[\exists \mathbf{w} \in C : \mathrm{wt}(\mathbf{w}) \leq \frac{\beta}{d} N] \leq \sum_{l=1}^{(\beta/d)N} (q-1)^l \binom{N}{l} \binom{L}{\lfloor \frac{lc}{2} \rfloor} \left(\frac{lc}{2L}\right)^{lc} \tag{74}$$

As in the proof of Theorem 2 of [21], we define:

$$f(l) = (q-1)^l \binom{N}{l} \binom{L}{\lfloor \frac{lc}{2} \rfloor} \left(\frac{lc}{2L}\right)^{lc}$$

Using methods similar to those in [21] we obtain:

$$\frac{f(l+2)}{f(l)} \leq \left\{ \frac{c(q-1)}{2(1-R)} e^{2c} \left(\frac{lc}{2L} \right)^{c/2-1} \right\}^2$$

Recalling $L = (c/d)N$ and $l < (\beta/d)N$ we obtain:

$$\frac{lc}{2L} \leq \frac{(\beta/d)Nc}{2(c/d)N} = \frac{\beta}{2}$$

We now define

$$\beta = 2 \cdot \inf_{c \geq 3} \left\{ e^{-(12 + \frac{6 \ln(c(q-1)/(1-R))}{c})} \right\}$$

The inner contents of the braces are positive for all c and approach e^{-12} as $c \to \infty$. Therefore, the above value is positive. Recalling $c \geq 3$ and $\beta \leq 2$ we have:

$$\begin{aligned} \frac{f(l+2)}{f(l)} &\leq \left\{ \frac{c(q-1)}{2(1-R)} e^{2c} \left(e^{-(12 + \frac{6 \ln(c(q-1)/(1-R))}{c})} \right)^{c/6} \right\}^2 \\ &= \left\{ \frac{c(q-1)}{2(1-R)} e^{2c - 2c - \ln[c(q-1)/(1-R)]} \right\}^2 = \left(\frac{1}{2} \right)^2 \end{aligned}$$

We now return to (74) and obtain:

$$\begin{aligned} \Pr[\exists \mathbf{w} \in C : \mathrm{wt}(\mathbf{w}) \leq \frac{\beta}{d} N] &\leq f(1) \sum_{l=1,3,\ldots} \left(\frac{1}{2} \right)^{l-1} + f(2) \sum_{l=2,4,\ldots} \left(\frac{1}{2} \right)^{l-2} \\ &\leq (q-1)^1 \binom{N}{1} \binom{L}{\lfloor \frac{1 \cdot c}{2} \rfloor} \left(\frac{1 \cdot c}{2L} \right)^{1c} \cdot 2 + (q-1)^2 \binom{N}{2} \binom{L}{\lfloor \frac{2 \cdot c}{2} \rfloor} \left(\frac{2 \cdot c}{2L} \right)^{2c} \cdot 2 \\ &= O(N^{1+c/2-c}) + O(N^{2+c-2c}) = O(N^{1-c/2}) \end{aligned} \tag{75}$$

2. We now examine values of λ satisfying $\beta/d \leq \lambda \leq \gamma$. From (20) we have

$$\begin{aligned} \Pr[\exists \mathbf{w} \in C : \mathrm{wt}(\mathbf{w}) = \lambda N] &\leq (q-1)^{\lambda N} \binom{N}{\lambda N} (cN+1)^q q^{N(1-R) \log A(\lambda)} \\ &\leq (cN+1)^q q^{N[H(\lambda) + \lambda \log(q-1) + (1-R) \log A(\lambda)]} \end{aligned} \tag{76}$$

We now restrict $\gamma \leq \frac{1}{2}$, and obtain, recalling (21) and using $\beta/d \leq \lambda \leq \gamma$

$$A(\lambda) \leq \frac{1}{q} \left\{ 1 + (q-1)[1 - 2(1-\rho) \frac{\beta}{d} \frac{1}{2}]^{\frac{d}{2}} \right\}$$

Using $1 - x \leq e^{-x}$ for all x, we have

$$\begin{aligned} A(\lambda) &\leq \frac{1}{q} \left[1 + (q-1)(e^{-(1-\rho)\frac{\beta}{d}})^{\frac{d}{2}} \right] \\ &= \frac{1}{q} \left[1 + (q-1) e^{-(1-\rho)\frac{\beta}{2}} \right] \end{aligned}$$

The function $H(\lambda) + \lambda \log(q-1)$ ascends from zero in the range $0 \leq \lambda \leq 1/2$. Defining

$$M = -\left\{ H(\gamma) + \gamma \log(q-1) + (1-R) \log[\frac{1}{q}(1 + (q-1)e^{-(1-\rho)\frac{\beta}{2}})] \right\}$$

we obtain that for all λ in the range $\beta/d \leq \lambda \leq \gamma$

$$H(\lambda) + \lambda \log(q-1) + (1-R) \log A(\lambda) \leq -M$$

We now define γ as some value in the range $0 < \gamma \leq 1/2$ yielding $M > 0$. Using (76) we obtain

$$\Pr[\exists \mathbf{w} \in C : \mathrm{wt}(\mathbf{w}) = \lambda N] \leq (cN+1)^q q^{-NM}$$

Therefore

$$\begin{aligned} \Pr[\exists \mathbf{w} \in C : \frac{\beta}{d} N \leq \mathrm{wt}(\mathbf{w}) \leq \gamma N] &\leq \sum_{l=(\beta/d)N}^{\gamma N} (cN+1)^q q^{-NM} \\ &\leq \gamma N (cN+1)^q q^{-NM} \end{aligned} \tag{77}$$

Combining (73) with (75) and (77) we obtain our desired result of (22). □

B.7 Proof of Theorem 7

Let γ be defined as in Theorem 6. Let δ be some number smaller than γ that will be determined later. Let $\mathcal{C}^x$ be the underlying expurgated LDPC ensemble of $\hat{\mathcal{D}}^x$, obtained by expurgating codes with minimum distance less than or equal to δN. We use Theorem 3 to bound $\overline{P}_e^x$.

Assigning

$$\begin{aligned} U_k &= \{\mathbf{t} \in \mathcal{T} : t_0 \neq N, t_k > (1 - \frac{\delta}{2})N\} \\ U &= \bigcup_{k=0}^{q-1} U_k \end{aligned} \tag{78}$$

we have (using (12))

$$\overline{P}_e^x \leq \sum_{\mathbf{t} \in U} \overline{S}_{\mathbf{t}}^x \mathbf{D}^{\mathbf{t}} + q^{-NE_Q(R + (\log \alpha)/N)} \tag{79}$$

We now examine both elements of the above sum.

1. As in the proof of Theorem 2 we obtain that $\overline{S}_t^x = 0$ for all $\mathbf{t} \in U_0$. We also obtain that for each $k \in \{1, ..., q-1\}$, C^x cannot contain more than one codeword having $\mathbf{t} \in U_k$.

Examining $\mathbf{D}^{\mathbf{t}}$ for $\mathbf{t} \in U_k$ we have

$$\mathbf{D}^{\mathbf{t}} = \prod_{i=0}^{q-1} D_i^{t_i} = D_k^{t_k} \cdot \prod_{i=0, i\neq k}^{q-1} D_i \leq D_k^{(1-\delta/2)N} \cdot 1$$

Letting $\{S_{\mathbf{t}}^x\}_{\mathbf{t}\in\mathcal{T}}$ denote the spectrum of code C^x, we have

$$\sum_{\mathbf{t}\in U_k} S_{\mathbf{t}}^x \mathbf{D}^{\mathbf{t}} \leq 1 \cdot D_k^{(1-\delta/2)N}$$

Summing over all k and taking the average spectrum we obtain:

$$\sum_{\mathbf{t}\in U} \overline{S}_{\mathbf{t}}^x \mathbf{D}^{\mathbf{t}} \leq \sum_{k=1}^{q-1} D_k^{(1-\delta/2)N}$$

Finally, selecting δ so that $\delta/2 \leq \epsilon_2$ we obtain

$$\sum_{\mathbf{t}\in U} \overline{S}_{\mathbf{t}}^x \mathbf{D}^{\mathbf{t}} \leq \sum_{k=1}^{q-1} D_k^{(1-\epsilon_2)N} \tag{80}$$

2. U^c can be written as

$$U^c = \{\boldsymbol{\theta} \cdot N : \boldsymbol{\theta} \in C_{\delta/2}\}$$

Where $C_{\delta/2}$ is defined as in (17). We now examine

$$\begin{aligned}
\frac{1}{N}\log\alpha &= \frac{1}{N}\log \max_{\boldsymbol{\theta}\in C_{\delta/2}} \frac{\overline{S}_{\boldsymbol{\theta}N}^x}{(M-1)\binom{N}{\theta_0 N,\ldots,\theta_{q-1}N} q^{-N}} \\
&= \max_{\boldsymbol{\theta}\in C_{\delta/2}} \left[\frac{1}{N}\log \overline{S}_{\boldsymbol{\theta}N}^x - \frac{1}{N}\log\left((M-1)\binom{N}{\theta_0 N,\ldots,\theta_{q-1}N} q^{-N}\right)\right] \\
&\leq \max_{\boldsymbol{\theta}\in C_{\delta/2}} \left[\frac{1}{N}\log \overline{S}_{\boldsymbol{\theta}N}^x - \frac{1}{N}\log \overline{S}_{\boldsymbol{\theta}N}\right] + \max_{\boldsymbol{\theta}\in C_{\delta/2}} \left[\frac{1}{N}\log \overline{S}_{\boldsymbol{\theta}N} - B(\boldsymbol{\theta})\right] \qquad (81) \\
&\quad + \max_{\boldsymbol{\theta}\in C_{\delta/2}} \left[B(\boldsymbol{\theta}) - \mathcal{R}(\boldsymbol{\theta})\right] + \max_{\boldsymbol{\theta}\in C_{\delta/2}} \left[\mathcal{R}(\boldsymbol{\theta}) - \frac{1}{N}\log\left((M-1)\binom{N}{\theta_0 N,\ldots,\theta_{q-1}N} q^{-N}\right)\right]
\end{aligned}$$

Examining (23) it is easy to verify that the first element of the above sum approaches zero as $N \to \infty$. To bound the second element, we first rely on (47), (68) and (69), and obtain a finite-N bound on $1/N \log \overline{S}_{\boldsymbol{\theta}N}$:

$$\begin{aligned}
\frac{1}{N}\log \overline{S}_{\boldsymbol{\theta}N} &= \frac{1}{N}\log\left[\binom{N}{\theta_0 N,\ldots,\theta_{q-1}N} \Pr[\mathbf{w}\in C]\right] \\
&\leq H(\boldsymbol{\theta}) + \left[\frac{1}{N}\log(cN+1)^q - cH(\boldsymbol{\theta}) + (1-R)\log \inf_{sgn(\mathbf{x})=sgn(\boldsymbol{\theta})} \frac{A(\mathbf{x})}{\mathbf{x}^d \boldsymbol{\theta}}\right] \\
&= B(\boldsymbol{\theta}) + \frac{1}{N}\log(cN+1)^q
\end{aligned}$$

Thus, the second element in (81), evaluated over all valid $\boldsymbol{\theta}$, is bounded arbitrarily close to zero as $N \to \infty$. The third element is upper bounded using Theorem 5, for c, d satisfying

$R = 1 - c/d$ and $d > d_0$. The fourth element approaches zero as $N \to \infty$ by (68) (recalling $M = q^{NR}$).

Summarizing, there exist positive integers N_0, d_0 such that for c, d, N satisfying $R = 1 - c/d$, $d > d_0$ and $N > N_0$ the following inequality holds:

$$\frac{1}{N} \log \alpha < \epsilon_1 \tag{82}$$

Combining (79), (80) and (82) we obtain our desired result of (24). □

B.8 Restriction to Prime Values of q

Let $r > 1$ be a positive integer, and let $q = 2 \cdot r$ be a nonprime number. We now show that there exist values of $\boldsymbol{\theta}$ such that no d_0 exists as in Theorem 5. Moreover, expurgation is impossible for those values of $\boldsymbol{\theta}$.

To show this, we examine the subgroup of the modulo-q group, formed by $\{0, r\}$. This subgroup is isomorphic to the group modulo-2 (the binary field). Given a modulo-q LDPC code C, we denote by C_2 the subcode produced by codewords containing symbols from $\{0, r\}$ alone. We now examine the ensemble of subcodes C_2.

The binary asymptotic normalized spectrum is defined, for $\theta \in (0, 1)$ as

$$B_2(\theta) \triangleq \lim_{N \to \infty} \frac{1}{N} \log_2 \overline{S}_{\theta N}$$

The limiting properties of this spectrum are well known (see [17]). Fixing the ratio $c/d = 1 - R$ and letting $c, d \to \infty$, $B_2(\theta)$ approaches the value $H_2(\theta) - (1 - R)$ where $H_2(\theta) = -\theta \log_2 \theta - (1 - \theta) \log_2(1 - \theta)$ is the binary entropy function.

We now examine values of $\boldsymbol{\theta}$ belonging to the set A_2, defined by

$$A_2 = \{\boldsymbol{\theta} = (\theta_0, ..., \theta_{q-1}) : \quad \theta_i = 0 \ \forall i \notin \{0, r\}, \quad \sum \theta_i = 1, \quad \theta_i \geq 0\}$$

Given the above discussion, the normalized spectrum of modulo-q LDPC, defined by (14), evaluated over A_2, corresponds to $\log_q 2 \cdot B_2(\theta_r)$. We therefore obtain that for $\boldsymbol{\theta} \in A_2$

$$\begin{aligned}
\lim_{c,d \to \infty} B(\boldsymbol{\theta}) &= \lim_{c,d \to \infty} \log_q 2 \cdot B_2(\theta_r) \\
&= \log_q 2 \cdot H_2(\theta_r) - \log_q 2 \cdot (1 - R) \\
&= H(\boldsymbol{\theta}) - \log_q 2 \cdot (1 - R) \\
&> H(\boldsymbol{\theta}) - (1 - R) \\
&= \mathcal{R}(\boldsymbol{\theta})
\end{aligned}$$

We now show that expurgation is impossible. The probability of the existence of words of normalized-type in A_2, in a randomly selected code, corresponds to the probability that *any* binary word satisfies the constraints imposed by the LDPC parity-check matrix. This results from the above discussed isomorphism between $\{0, r\}$ and the modulo-2 group. This probability clearly does not approach zero (in fact it is one), and hence expurgation of codes containing such words is impossible.

C Proofs for Section IV

C.1 Proof of Theorem 8

The proof of this theorem follows in the lines of the proof of Theorem 4.

Equation (48) is carried over from the modulo-q case. Let $\mathbf{w}$ be a word of type $\boldsymbol{\theta}N$. Recalling Lemma 2, we can assume without loss of generality that $\mathbf{w}$ is a binary word of weight $\lambda N = (1-\theta_0)N$.

Each assignment of edges and labels infers a symbol "coloring" on each of the check node sockets (the ith color corresponding to the ith element of GF(q)), according to the adjacent variable node's value multiplied by the label of the connecting edge. Exactly λNc of the sockets are assigned colors of the set $\{1, ..., q-1\}$. All color assignments satisfying the above requirement are equally probable. The total number of colorings is:

$$t(\lambda; N) = \binom{Nc}{\lambda Nc}(q-1)^{\lambda Nc} \tag{83}$$

Given an assignment of graph edges, $\mathbf{w}$ is a codeword if at each check node, the $GF(q)$ sum of the values of its sockets is zero. The number of valid assignments is determined using the enumerating function:

$$e(\lambda; N) = \lfloor A(x)^{(1-R)N} \rfloor_{\lambda Nc} \tag{84}$$

where $A(x)$ is the weight enumerating function of d-length words over GF(q) satisfying the requirement that the sum of all symbols must equal 0 over GF(q).

Using (48), (51) (with $t(\boldsymbol{\theta}; N)$ and $e(\boldsymbol{\theta}; N)$ replaced by $t(\lambda; N)$ and $e(\lambda; N)$), (83) and (84), and applying Theorem 10 as in the proof of Theorem 4 we arrive at (25). We now show that $A(x)$ is given by (26).

We first examine $\hat{A}(x_0, ..., x_{q-1})$, defined such that the coefficient of $x_0^{k_0}, ..., x_{q-1}^{k_{q-1}}$ is the number of words of type $\mathbf{t} =< k_0, ..., k_{q-1} >$ (the index i corresponds to the ith element of GF(q)) whose

sum over GF(q) is zero. A useful expression for $\hat{A}(x_0, ..., x_{q-1})$ is

$$\hat{A}(x_0, ..., x_{q-1}) = \sum_{h_1,...,h_d \in \mathrm{GF}(q) \,:\, \sum_{i=1}^{d} h_i = 0 \ (\mathrm{GF}(q))} x_{h_1}, ..., x_{h_d} \tag{85}$$

Elements of GF(p^m) can be modelled as m-dimensional vectors over $0, ..., p-1$ (see [3]). The sum of two GF(p^m) elements corresponds to the sum of the corresponding vectors, evaluated as the modulo-p sum of each of the vector components. Thus, adding d elements over GF(p^m) is equivalent to adding d m-dimensional vectors. Letting $\mathbf{n}^{(i)} \in \{0, ..., p-1\}^m, i = 1, ..., d$, we have the following expression for $\hat{A}(x_0, ..., x_{q-1})$:

$$\hat{A}(x_0, ..., x_{q-1}) = \sum_{\mathbf{n}^{(1)},...,\mathbf{n}^{(d)} : \sum_{i=1}^{d} n_l^{(i)} = 0 \bmod p, \quad l=1,...,m} x_{\mathbf{n}^{(1)}}, ..., x_{\mathbf{n}^{(d)}}$$

Consider, for example, the simple case of $d = 2$ and $m = 2$. To simplify our notations, we denote $\mathbf{n}^{(1)}$ by (i, j) and $\mathbf{n}^{(2)}$ by (k, l)

$$\begin{aligned} \hat{A}(x_0, ..., x_{q-1}) &= \sum_{i,j,k,l=0,...,p-1 \,:\, i+k=0,\, j+l=0 \ (\mathrm{mod}\ p)} x_{i,j} \cdot x_{k,l} \\ &= \sum_{i,j=0,...,p-1} x_{i,j} \cdot x_{0-i,0-j} \end{aligned}$$

where $0 - i$ and $0 - j$ are evaluated modulo-p. This last equation is clearly the output of the 2-dimensional cyclic convolution of $\{x_{i,j}\}$ with itself, evaluated at zero. In the general case (85), we have the m-dimensional cyclic-convolution of the function $f(n_1, ..., n_m) = x_{\mathbf{n}}$ with itself d times, evaluated at zero. Using the m dimensional DFT of size p to evaluate the convolution, we obtain

$$\begin{aligned} \hat{A}(x_0, ..., x_{q-1}) &= \mathrm{IDFT}\left[\mathrm{DFT}(x_0, ..., x_{q-1})^d\right]_{\mathbf{n}=0} \\ &= \frac{1}{p^m} \sum_{\mathbf{k} \in \{0,...,p-1\}^m} \left[\sum_{\mathbf{n} \in \{0,...,p-1\}^m} x_{\mathbf{n}} e^{j\frac{2\pi}{p} \sum_{i=1}^{m} k_i n_i} \right]^d \end{aligned}$$

We now examine $A(x) = \hat{A}(1, x, ..., x)$.

$$\begin{aligned} A(x) &= \frac{1}{q} \sum_{\mathbf{k} \in \{0,...,p-1\}^m} \left[1 \cdot e^{j\frac{2\pi}{p} \sum k_i \cdot 0} + \sum_{\mathbf{n} \in \{0,...,p-1\}^m, \mathbf{n} \neq \mathbf{0}} x e^{j\frac{2\pi}{p} \sum_{i=1}^{m} k_i n_i} \right]^d \\ &= \frac{1}{q} \sum_{\mathbf{k} \in \{0,...,p-1\}^m} \left[1 + \left(\sum_{\mathbf{n} \in \{0,...,p-1\}^m} e^{j\frac{2\pi}{p} \sum_{i=1}^{m} k_i n_i} - 1 \right) \cdot x \right]^d \end{aligned} \tag{86}$$

Consider the elements $\left\{ \sum_{\mathbf{n} \in \{0,...,p-1\}^m} e^{j\frac{2\pi}{p} \sum k_i n_i} \right\}_{\mathbf{k} \in \{0,...,p-1\}^m}$. These elements correspond to the DFT of the function $f(n_1, ..., n_m) \equiv 1$, which is a delta function. Thus we obtain:

$$\sum_{\mathbf{n} \in \{0,...,p-1\}^m} e^{j\frac{2\pi}{p} \sum k_i n_i} = \begin{cases} q & \mathbf{k} = \mathbf{0} \\ 0 & \mathbf{k} \neq \mathbf{0} \end{cases}$$

Combining the above with (86) we obtain our desired result of (26). □

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments.

References

[1] A. Bennatan and D. Burshtein, "Iterative decoding of LDPC codes over arbitrary discrete-memoryless channels", in preparation.

[2] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error correcting coding and decoding: turbo codes", *Proceedings 1993 IEEE International Conference on Communications*, Geneva, Switzerland, pp. 1064–1070, 1993.

[3] R.E. Blahut, *Theory and Practice of Error Control Codes*, Addison-Wesley Publishing Company, 1984.

[4] D. Burshtein and G. Miller, "Asymptotic enumeration methods for analyzing LDPC codes", submitted for publication *IEEE Trans. on Inform. Theory*, available in http://www.eng.tau.ac.il/ burstyn/AsymptEnum.ps.

[5] G. Caire, D. Burshtein and S. Shamai, "LDPC Coding for Interference Mitigation at the Transmitter", in *40th Annual Allerton Conf. on Commun., Cont. and Comp.*, Monticello, IL, October 2002.

[6] S.-Y. Chung, J. G. D. Forney, T. Richardson and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit", *IEEE Commun. Lett.*, vol. 5, pp. 58–60, February 2001.

[7] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.

[8] M. C. Davey and D. MacKay, "Low-Density Parity Check Codes over GF(q)",*IEEE Comm. Letters*, vol. 2, pp. 165–167, June 1998.

[9] D. Divsalar and F. Pollara, "Turbo trellis coded modulation with iterative decoding for mobile satellite communications", *Proc. International Mobile Satellite Conference*, June 1997.

[10] U. Erez and G. Miller, "The ML decoding performance of LDPC Ensembles over Z_q", submitted for publication *IEEE Trans. on Inform. Theory*

[11] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and Coding for Linear Gaussian Channels", *IEEE Trans. on Inform. Theory*, vol. 44, pp. 2384–2415, October 1998.

[12] R. G. Gallager, *Low Density Parity Check Codes*, M.I.T Press, Cambridge, Massachusetts, 1963.

[13] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons, 1968.

[14] I. J. Good, "Saddle point methods for the multinomial distribution", *Ann. Math. Statist.*, vol. 28, pp. 861 – 881, 1957.

[15] A. Kavčić, X. Ma and M. Mitzenmacher, "Binary Intersymbol Interference Channels: Gallager Codes, Density Evolution and Code Performance Bounds", *Accepted to IEEE Trans. on Inform. Theory.*

[16] A. Kavčić, X. Ma, M. Mitzenmacher and N. Varnica, "Capacity approaching signal constellations for channels with memory", *Proc. Allerton Conf.*, pp. 311–320 (Allerton IL.), October 2001.

[17] S. Litsyn and V. Shevelev, "On Ensembles of Low-Density Parity-Check Codes: asymptotic distance distributions", *IEEE Trans. on Inform. Theory*, vol. 48, pp. 887–908, April 2002.

[18] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Improved Low-Density Parity-Check Codes Using Irregular Graphs", *IEEE Trans. on Inform. Theory* vol. 47, pp. 585–598, Februray 2001.

[19] X. Ma, N. Varnica and A. Kavčić, "Matched information rate codes for binary ISI channels ", *Proc. 2002 IEEE International Symposium on Information Theory*, (Lausanne, Switzerland), July 2002.

[20] R.J. McEliece, "Are Turbo-like Codes Effective on Nonstandard Channels?" IEEE Information Theory Society Newsletter, vol. 51, pp. 1-8, December 2001.

[21] G. Miller and D. Burshtein, "Bounds on the Maximum-Likelihood Decoding Error Probability of Low-Density Parity-Check Codes", *IEEE Trans. on Inform. Theory* vol. 47, pp. 2696–2710, November 2001.

[22] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, 1988.

[23] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding", *IEEE Trans. on Inform. Theory*, vol. 47, pp. 599–618, February 2001.

[24] T. Richardson, A. Shokrollahi and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes", *IEEE Trans. on Inform. Theory*, vol. 47, pp. 619–637, February 2001.

[25] P. Robertson and T. Wörts, "Bandwidth-efficient Turbo trellis-coded modulation using punctured component codes", *IEEE J. Select. Areas Commun.* , vol. 16, pp. 206–218, February 1998.

[26] N. Shulman, and M. Feder, "Random Coding Techniques for Nonrandom Codes",*IEEE Trans. on Inform. Theory*, vol. 45, pp. 2101–2104, September 1999.

[27] R. M. Tanner, "A recursive approach to low complexity codes", *IEEE Trans. on Inform. Theory*, vol. IT-27, pp. 533-547, September 1981.

[28] S. ten Brink, G. Kramer and A. Ashikhmin, "Design of Low-Density Parity-Check Codes for Multi-Antenna Modulation and Detection", submitted for publication *IEEE Trans. on Inform. Theory.*

[29] N. Varnica, X. Ma and A. Kavčić, "Iteratively Decodable Codes for Bridging the Shaping Gap in Communications Channels", *Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), November 2002.

[30] N. Varnica, X. Ma and A. Kavčić, "Capacity of power constrained memoryless AWGN channels with fixed input constellations", *Proc. IEEE Global Telecomm. Conf (GLOBECOM)* (Taipei, Taiwan), November 2002.

[31] U. Wachsmann, R. F. Fischer and J. B. Huber, "Multilevel Codes: Theoretical Concepts and Practical Design Rules", *IEEE Trans. on Inform. Theory*, vol. 45, pp. 1361–1391, July 1999.

List of Tables

List of Figures

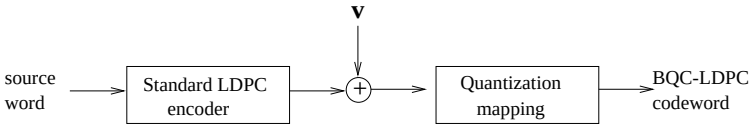

Figure 1: Encoding of BQC-LDPC codes

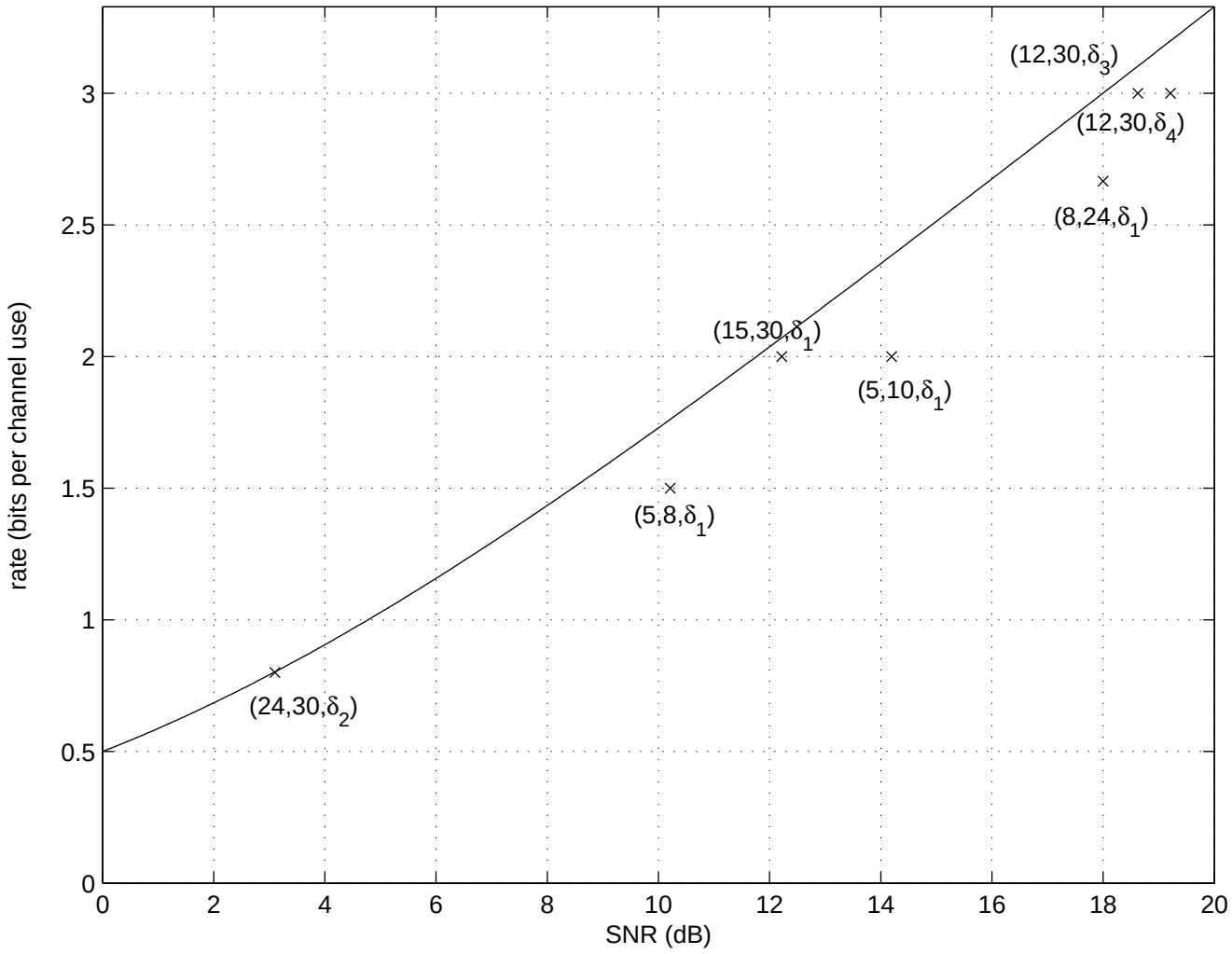

Figure 2: Upper bounds on the minimum required SNR for successful ML decoding of some BQC-LDPC codes over an AWGN channel. The solid line is the Shannon limit.

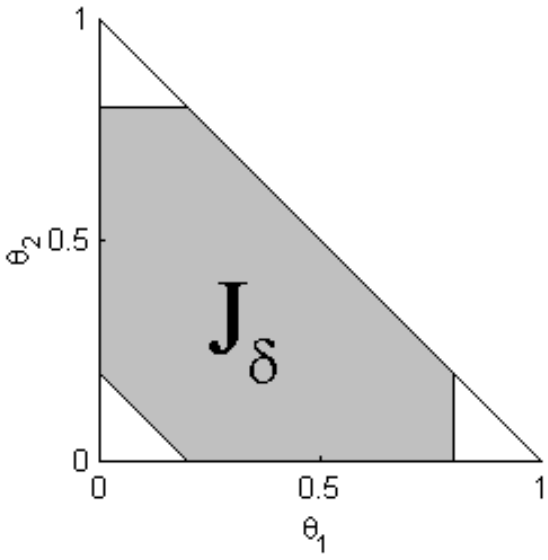

Figure 3: J_δ for $\delta = 0.2$

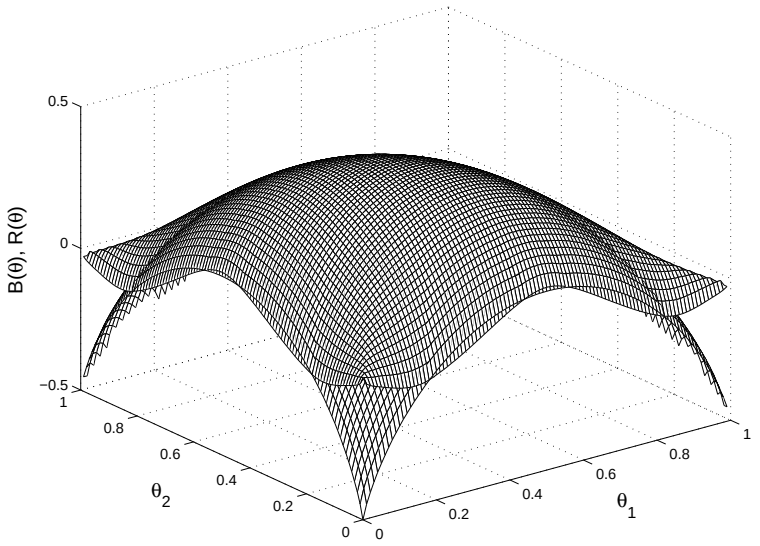

Figure 4: Comparison of the normalized spectrums of the ternary $(3, 6)$-regular LDPC code ensemble and the ternary rate $1/2$ random-coding ensemble.

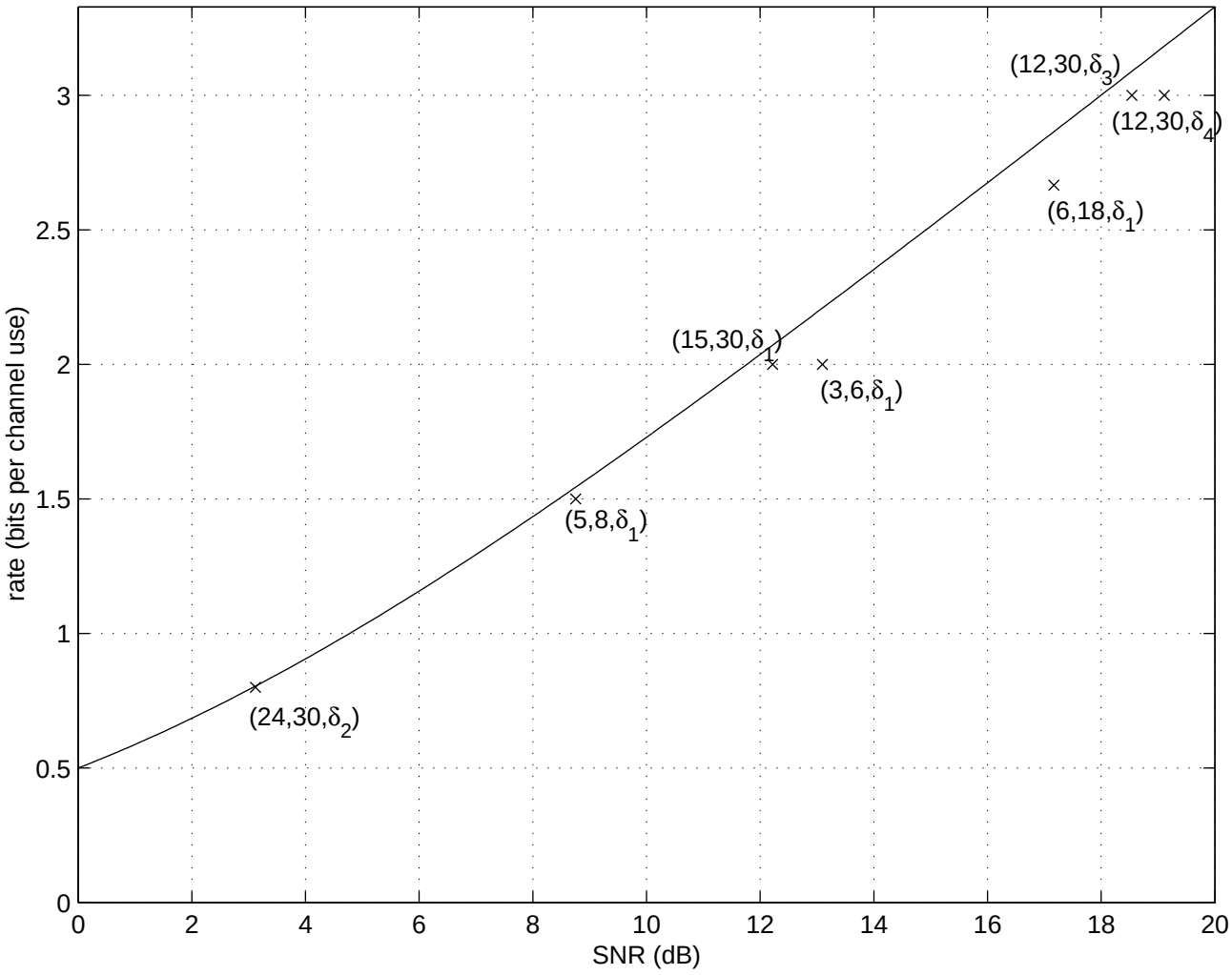

Figure 5: Upper bounds on the minimum required SNR for successful ML decoding of some GQC-LDPC codes over an AWGN channel. The solid line is the Shannon Limit.

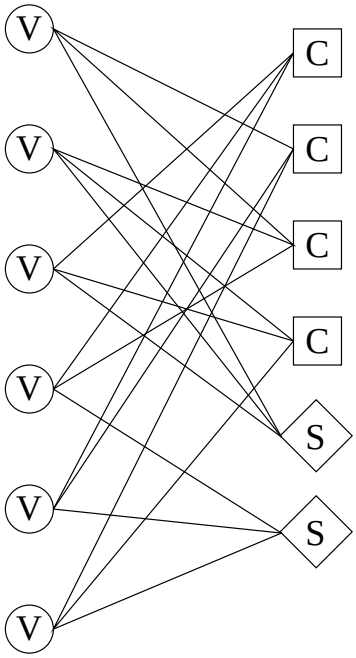

Figure 6: Diagram of the bipartite graph of a $(2, 3)$-regular BQC-LDPC code with $T = 3$

i \ l	1	2	3	4
2	0.07333	0.07875	0.05958	0.00917
3	0.06170	0.04210	0.06875	0.01375
5	0	0.01205	0.01375	0.08021
6	0	0.00583	0.00688	0.08250
20	0.00117	0.01007	0.01375	0.36667

Table 1: Left edge distribution for a rate 2 BQC-LDPC code within 1.1 dB of the Shannon limit. The right edge distribution is given by $\rho_8 = 0.25, \rho_9 = 0.75$.

i	2	3	4	6	10	12
λ_i	0.57077	0.040604	0.088118	0.15039	0.020119	0.13

Table 2: Left edge distribution for a rate 2 MQC-LDPC code within 0.9 dB of the Shannon limit. The right edge distribution is given by $\rho_5 = 0.5, \rho_6 = 0.5$.

i	2	4	6	12
λ_i	0.62258	0.08344	0.12024	0.17374

Table 3: Left edge distribution for a rate 2 GQC-LDPC code within 0.65 dB of the Shannon limit. The right edge distribution is given by $\rho_5 = 0.5, \rho_6 = 0.5$.