Analytical Framework of LDGM-based Iterative Quantization with Decimation

Qingchuan Wang, Chen He, Lingge Jiang

Abstract—While iterative quantizers based on low-density generator-matrix (LDGM) codes have been shown to be able to achieve near-ideal distortion performance with comparatively moderate block length and computational complexity requirements, their analysis remains difficult due to the presence of decimation steps. In this paper, considering the use of LDGM-based quantizers in a class of symmetric source coding problems, with the alphabet being either binary or non-binary, it is proved rigorously that, as long as the degree distribution satisfies certain conditions that can be evaluated with density evolution (DE), the belief propagation (BP) marginals used in the decimation step have vanishing mean-square error compared to the exact marginals when the block length and iteration count goes to infinity, which potentially allows near-ideal distortion performances to be achieved. This provides a sound theoretical basis for the degree distribution optimization methods previously proposed in the literature and already found to be effective in practice.

Index Terms—LDGM, sparse-graph codes, belief propagation, decimation, source coding, density evolution

I. INTRODUCTION

Near-ideal quantization is important not only in source coding, but also in many channel coding problems due to e.g. signal shaping [1] or compress-and-forward [2] concerns; in particular, in many low-rate source or channel coding applications, such as dirty-paper coding, small gaps from ideal performance in the quantizer can translate to a significant percentage loss of the overall code rate [3]. For the symmetric cases considered in this paper, where the shaping gain [4] is to be maximized and the boundary gain is not an issue, practical near-ideal quantization methods include structured trellis-coded quantization (TCQ) [5] and polar codes [6,7], as well as sparse-graph constructions mostly based on low-density generator matrix (LDGM) codes [8–10]. Although all three methods are able to achieve near-ideal distortion performance, as the gap closes, TCQ requires a large memory length and thus exponential computational complexity, while polar codes are more severely hampered by the finite block lengths available in practice [11], [12], making LDGM-based codes the only choice if performance extremely close to the theoretical limit, e.g. 0.012 dB for MSE (mean-square error) quantization [13] obtained in [12], is to be achieved with reasonable computational complexity and block lengths. Such advantage in performance, combined with the high flexibility and wide applicability of sparse-graph codes in a variety of source and channel coding problems, makes the analysis and design of LDGM-based constructions for quantization highly important both theoretically and in practice.

In terms of implementation, LDGM-based quantizers require a practical encoding algorithm as well as optimized degree distributions, and good ones have now been obtained in the literature. In particular, the encoding algorithm can be either belief propagation (BP) [14] or survey propagation (SP) [9] combined with decimation and preferably also a recovery procedure [12], and other variations such as [15] have also been proposed for specific cases. The degree distribution optimization problem has also been tackled in [16], although the duals of optimized low-density parity-check (LDPC) degree distributions used in earlier works, e.g. [9], can often give adequate performance as well.

On the other hand, theoretical analysis of the quantization algorithm remains difficult due to its iterative nature and use of decimation. While distortion performance under optimal (MAP) encoding has been analyzed in [9,10] for specific degree distributions using codeword-counting arguments, good performance under MAP encoding is far insufficient for guaranteeing good performance under practical BP or SP-based encoding algorithms. An effective approach to BP analysis is density evolution (DE), which has been successfully applied to LDPC decoding [17]; however, while the BP process in LDGM decoding will converge by itself as long as the decoding threshold is reached, in the LDGM quantizer BP will not converge without additional decimation steps, and there is no obvious method to make DE work across decimation steps due to its requirement on the independence of BP messages. Analysis of similar decimation steps has been attempted in [18] for the solution of boolean satisfiability problems, and [7] for quantization based on polar codes, and although both papers offer insights that are valuable to our work, the methods there are not sufficient for use in LDGM quantization. Specifically, the successful analysis in [7] relies on the availability of exact marginals (or extrinsic information) during decimation when polar codes are used, allowing them to be viewed as conditional probabilities corresponding to a known joint probability distribution, but in LDGM quantization only BP approximations of these marginals are available, whose accuracy remains to be evaluated; when confronting a more difficult problem where the available marginals are limited to BP approximations as well, [18] provides some insights on the application of DE in such situations, but it still has difficulty accounting for the impact of loops in the factor graph. Inspired by the works [19,20] attempting to characterize the accuracy...
of BP marginals using extrinsic information transfer (EXIT) for LDPC decoding. Our previous paper [16] applies the same method to LDGM quantization, and conjectures that the BP marginals can be asymptotically accurate when the degree distribution satisfies certain monotonicity conditions that can be evaluated using DE, in which case the distortion performance can then be approximated using methods similar to that used for polar codes in [17]; although this rough analysis allows the degree distribution to be optimized to yield good performance, the arguments there are largely heuristic and lack mathematical rigor, particularly for cases other than binary erasure quantization (BEQ).

Building upon the aforementioned results, this paper attempts to extend the analytical approach of [16] to a class of “symmetric” source coding problems, both binary and non-binary. With the introduction of a reference codeword in DE, the properties regarding the symmetry and degradation relationships among message densities, previously used in LDPC analysis in [17], are generalized, and they are then used to relate the actual densities of BP messages to those obtainable with DE, and to bound the difference between BP and exact marginals used in decimation with the difference in their mutual information characterized by EXIT curves. In this way, we are able to show rigorously that the monotonicity condition used as the optimization criteria in [16] can indeed lead to good distortion performance in a certain asymptotic sense. The difficulty in applying DE across decimation steps is side-stepped by considering each decimation step separately, assuming that exact marginals have been used in all previous decimation steps. Even though the actual quantizer can only use BP marginals in all decimation steps, and errors in the earlier BP marginals can affect subsequent BP marginals in a manner that is difficult to analyze, we believe that the present results are still able to provide important insights to BP-based quantization algorithms; in any case, the recovery algorithm in [12] can greatly alleviate this problem in practice.

The rest of this paper is organized as follows. Section II starts from the MSE quantization problem and introduces a more general class of symmetric lossy compression problems to be considered in the rest of the paper. Section III reviews the LDGM code construction and quantization algorithm that are used to solve such problems, and gives an outline of the analytical approach. Our main analytical results are presented in Section IV. Starting from some basic properties of message densities in the presence of an explicit reference codeword, the error bounds of BP marginals expressed in terms of DE results are used to justify the monotonicity conditions for degree distribution optimization, and some more intuitive results are then given for the special case of BEQ. Subsequently, Section V briefly shows how to extend this analytical approach to non-binary constructions, and finally Section VI concludes the paper.

Notational conventions: $\mathbb{Z}$ and $\mathbb{R}$ are respectively the set of integers and real numbers. $\mathbb{Z}_q \triangleq \mathbb{Z}/q\mathbb{Z}$ is the modulo-$q$ additive group. $\mathcal{A} \setminus \mathcal{B}$ is the difference set containing the elements of set $\mathcal{A}$ that are not in set $\mathcal{B}$. $E[\cdot]$ is the expectation operator. $\|\cdot\|$ is the Euclidean norm. $|\mathcal{A}|$ is the cardinality of set $\mathcal{A}$. $1[\cdot]$ is 1 if the condition $\mathcal{A}$ is true, 0 otherwise. $\log(\cdot)$, entropy and mutual information are computed in base-2, while $\ln(\cdot)$ and $\exp(\cdot)$ are base-e. Bold letters denote sequences or vectors whose elements are indicated by subscripts, e.g. $y = (y_1, \ldots, y_n)$, $y_{-i}$ is the sub-sequence $(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$, and a sub-sequence with index set $\mathcal{S}$ can be denoted by $y_{\mathcal{S}} = (y_i)_{i \in \mathcal{S}}$: note that $y$ itself can denote a scalar variable unrelated to $y$. Addition and multiplication on sets are element-wise, e.g. $U + 2\mathbb{Z}^n = \{u + 2d_1, \ldots, 2d_n\} | \ u \in U, d_i \in \mathbb{Z}\}$, $\oplus$ and $\ominus$ denote addition and subtraction in a specific additive abelian group $\mathbb{G}$, but can also denote variants of the check-node operation when applied to probability tuples and densities, as will be explained in Sections III IV-A and IV-A. $x \mod (a, b)$, or simply $(x)_{[a,b]}$, is defined as the unique element of $((x-(b-a)\mathbb{Z}) \cap [a,b])$, and similarly $x \mod (a, b)^n$ or $(x)_{[a,b]^n}$ is the unique element of $(x-(b-a)\mathbb{Z}) \cap [a,b]^n$. The unit “b/s” means “bits per symbol”. For convenience, we do not distinguish in notation between random variables and their possible values, or between the pmfs of discrete random variables and pdfs of continuous ones, which should be clear from context; for example, $p(b = b')$ or $p\rho(b')$ denotes the probability (density) that random variable $b$ takes the value $b'$, while we simply write $p(b)$ if both the random variable and the value are denoted by $b$, or if it is clear from context what the random variable is.

II. PROBLEM FORMULATION AND PERFORMANCE BOUNDS

A. MSE Quantization

The mean-squared error (MSE) quantization problem of $\mathbb{R}^n$ [13] Sec. II-C can be formulated as follows. Let $\Lambda$ be a non-empty discrete subset of $\mathbb{R}^n$ (the quantization codebook, or simply code), and $Q\Lambda : \mathbb{R}^n \to \Lambda$ be a quantizer that maps each $y \in \mathbb{R}^n$ to a nearby codeword $Q\Lambda(y) \in \Lambda$. The mean-squared quantization error, averaged over $y$, is given by

$$\sigma^2 = \limsup_{M \to \infty} \frac{1}{(2M)^n} \frac{1}{n} \int_{[-M,M]^n} \|y - Q\Lambda(y)\|^2 \, dy.$$  \hspace{1cm} (1)

The objective is to design $\Lambda$ and a practical quantizer $Q\Lambda(\cdot)$ such that the scale-normalized MSE $G(\Lambda) \triangleq \sigma^2 \rho^{2/n}$ is minimized, where $\rho$ is the codeword density

$$\rho = \limsup_{M \to \infty} \frac{1}{(2M)^n} |\Lambda \cap [-M,M]^n|.$$  \hspace{1cm} (2)

It should be noted that [13] assumes that $\Lambda$ is a lattice, which ensures that the Voronoi regions corresponding to different codewords in $\Lambda$ differ only by a translation, and since lattices are closed under addition, such codebooks can often achieve better performance than unstructured ones in e.g. network coding problems involving channels with similar additive structures [21]. On the other hand, in plain quantization problems, the lattice structure is fairly unimportant, and indeed trellis codebooks or those generated with a modulation mapping often lack such a structure and yet still achieve good performance. Therefore, in this problem formulation we do not constrain $\Lambda$ to be a lattice, and the definitions in [13] have been generalized accordingly.
In this paper we consider asymptotically large dimensionality $n$. By a volume argument, it is easy to find an asymptotic lower bound $G^* = \frac{1}{2\pi n}$ for $G(A)$ as $n \to \infty$. This bound can be approached by the nearest-neighbor quantizer with a suitable random codebook [13], whose codewords' Voronoi regions are asymptotically spherical, but such a quantizer has exponential computational complexity in $n$ and is thus impractical. The simplest scalar quantizer $A_1 = \mathbb{Z}^n$, on the other hand, has the 1.5329-dB larger $G_1 = G(A_1) = \frac{1}{12}$, corresponding to the well-known 1.532-dB loss of scalar quantization. In general, we call $10\log_{10}(G(A)/G^*)$ the shaping loss of a quantizer, and it is also the gap from the granular gain and shaping gain defined in [4], for source and channel coding respectively, toward the 1.53-dB limit.

In order to design a practical quantization codebook with a finite alphabet, we consider $\Lambda$ with a periodic structure $\Lambda = \mathcal{U} + M\mathbb{Z}^n$, where $\mathcal{U}$ is a set of $2^n R$ codewords from $\mathbb{Z}_M^n$ with each $\mathbf{u} = \mathbf{u}(b) \in \mathcal{U}$ labeled by a binary sequence $b \in \mathbb{Z}_2^n$. Such a $\Lambda$ is called an $M$-ary rate-$R$ quantization code, and is also used by TCQ. Constrained by this $M$-ary structure, the MSE quantization problem is then equivalent to the lossy compression of an i.i.d. uniform source over $\mathcal{Y} \equiv [0, M]$ using codebook $\mathcal{U}$ and the modulo-$I$ $(I \equiv \{ -\frac{M}{2}, \frac{M}{2}\})$ distortion measure $d(u, y) = (y - u)^2$, and $\sigma^2$ in (1) is simply the average distortion and $\rho = \frac{2^{nR}}{M^n}$; this equivalent problem is henceforth called $M$-ary MSE quantization. At a given $R$, $\sigma^2$ corresponding to the bound $G^*$ is

$$\sigma^2(\mathcal{I}) \equiv G^* \rho^{-2/n} = (2\pi e(2^R/M)^2)^{-1}. \quad (3)$$

While $\sigma^2(\mathcal{I})$ is not exactly achievable at any finite $M$, leaving a gap called the random-coding loss in Section II-C, this gap can become extremely small as $M$ increases.

B. Symmetric Source Coding Problems over a Finite Abelian Group

$M$-ary MSE quantization is now generalized as follows for uniformity of presentation.

**Definition 1**: Consider the source coding problem involving i.i.d. source $y$ taking values in $\mathcal{Y}$ with pmf or pdf $p(y)$, under distortion measure $d(u, y)$; that is, given any block size $n$ and rate $R > 0$, we design a codebook $\mathcal{U}$ of size $2^n R$ along with encoding and decoding functions, which map each possible source sequence $y$ into a reconstructed sequence $u(y) \in \mathcal{U}$ with distortion $d(u(y), y) \equiv \frac{1}{n} \sum_{j=1}^{n} d(u_j(y), y_j)$, and the objective is to minimize the average distortion $D \equiv E[d(u(y), y)]$ with the expectation taken over $p(y) = p_0(y_1) \cdots p_{n}(y_n)$. This is called a symmetric source coding problem over $\mathcal{G}$, if the reconstruction alphabet is a finite abelian group $\mathcal{G}$ (i.e. $\mathcal{U} \subseteq \mathbb{G}^n$), and if a measure-preserving group action $\psi$ of $\mathcal{G}$ exists on $\mathcal{Y}$, such that

$$p(y) = p(\psi_u(y)) \quad d(u, y) = d(0, \psi_u(y)) \quad (4)$$

for any $y \in \mathcal{Y}$ and $u \in \mathcal{G}$.\footnote{When $p(y)$ is a pdf, we require the group action $\psi$ to be measure-preserving w.r.t. the measure over $\mathcal{Y}$ used to define that pdf, so that the symmetry $p(y) = p(\psi_u(y))$ in probability density implies the symmetry in the probability itself.}

Below are some examples with $\mathcal{G} = \mathbb{Z}_M$, which may be called $M$-ary symmetric source coding problems\footnote{Not to be confused with source coding of $M$-ary symmetric sources, i.e. Example 8 below, which is only a special case.}

**Example 1**: In $M$-ary MSE quantization, $p(y)$ is uniform over $\mathcal{Y} = [0, M)$, $d(u, y) = (y - u)^2$ (the $\mathcal{I}$ in the subscript denotes modulo operation like above), and $\psi_u(y) = (y - u) y$.

**Example 2**: In quantization of an $M$-ary discrete source with Hamming distortion, $p(y)$ is uniform over $\mathcal{Y} = \mathbb{Z}_M$, $d(u, y) = 1[y \neq u]$, and $\psi_u(y) = (y - u) \mod M$.

**Example 3**: Another well-known example is $M$-ary erasure quantization, where $\mathcal{Y} = \mathbb{Z}_M \cup \{\ast\}$ (denotes an erased symbol), $p_\ast(y) = \epsilon$ with $0 < \epsilon < 1$, $p(y) = (1 - \epsilon)/M$ for $y \in \mathbb{Z}_M$, $d(u, y) = 1[y \neq u$ and $y \neq \ast]$, while $\psi_u(y) = (y - u) \mod M$ for $y \in \mathbb{Z}_M$ and $\psi_\ast(y) = \ast$. This is usually considered in the zero-distortion limit, particularly when $M = 2$ (known as binary erasure quantization (BEQ) [8]), due to its simplicity.

There are also noteworthy symmetric lossy quantization problems with other reconstruction alphabets $\mathcal{G}$:

**Example 4**: MSE quantization can be generalized to $N$ real dimensions per source symbol as follows. Given $N$, let $\mathcal{C}_1$ be a lattice in $\mathbb{R}^N$, i.e. a discrete additive subgroup of $\mathbb{R}^N$, and $\mathcal{C}_2$ be $\mathcal{C}_1$’s subgroup, which forms a coarser lattice. Now we make the source alphabet $\mathcal{Y} = \mathbb{R}^N / \mathcal{C}_2$, and the reconstruction alphabet $\mathcal{G} = \mathcal{C}_1 / \mathcal{C}_2$ quotient groups w.r.t. $\mathcal{C}_2$, such that each source symbol $y$ and reconstruction symbol $u$ can be viewed as an $N$-dimensional vector modulo $\mathcal{C}_2$, and $p(y)$ is then the uniform distribution over $\mathcal{Y}$, $d(u, y) = \| (y - u) \mod \mathcal{C}_2 \|^2$ is the squared modulo-$\mathcal{C}_2$ Euclidean distance, and $\psi_u(y) = (y - u) \mod \mathcal{C}_2$ is simply subtraction in the group $\mathcal{Y}$, of which $\mathcal{G}$ is a subgroup. In particular, Example 1 is the case that $N = 1$, $\mathcal{C}_1 = \mathbb{Z}$, and $\mathcal{C}_2 = \mathbb{Z}_M$. This is related to vector precoding often performed in MIMO systems, especially MIMO broadcast channels, that performs spatial signal shaping in order to approach capacity more closely; for example, $\mathcal{C}_1$ and $\mathcal{C}_2$ can be chosen as respectively the lattices $\mathbb{Z}_N$ and $\mathbb{M}_2$ in the receiver-side signal space, transformed to the transmitter side using the inverted channel matrix.

**Example 5**: BEQ can be generalized to $K$ dimensions per source symbol as follows. Given $K$, we let $\mathcal{G} = \mathbb{Z}_K^n$ be the $K$-dimensional linear space over $\mathbb{Z}_2$, and $\mathcal{Y}$ be the set of all affine subspaces of $\mathcal{G}$, which can be partitioned by the corresponding vector subspace $x$ into $\mathcal{Y}_x = \{y + d | d \in \mathcal{G}\}$ being the set of affine subspaces from each $x$. Now let $d(u, y) = 1[u \neq y]$ for $u \in \mathcal{G}$ and $y \in \mathcal{Y}$, and constrain $p(y)$ to be uniform over each $\mathcal{Y}_x$, so that (4) holds with $\psi_u(y) = y + u$, where $\oplus$ is bitwise subtraction in $\mathbb{Z}_2^K$ applied element-wise to $y$. When $K = 1$, this reduces to BEQ if the affine subspaces $\{0\}, \{1\}$ and $\{0, 1\}$ of $\mathbb{Z}_2$ are identified with $0, 1$ and $\ast$ in $\mathcal{Y}$.

According to rate-distortion theory [22] Sec. 10.4–10.5, in the limit of large $n$, each possible test channel $p(u | y)$ corresponds to an average distortion $D = E[d(u, y)]$ achievable at rate $R = I(u; y)$ with a random codebook and a quantizer based on joint typicality, and conversely, any achievable rate
can be achieved in this way with some test channel; here \(u\) and \(y\) are viewed as random variables and \(D\) and \(R\) are computed according to joint distribution \(p(y)p(u | y)\). The optimal test channel that minimizes \(D\) at a given \(R\) (or vice versa) is straightforward to compute:

**Proposition 1:** The optimal test channel for symmetric source coding over \(\mathbb{G}\) is

\[
p(u | y) = e^{-td(u, y)}/Q(y), \quad u \in \mathbb{G}
\]

where \(Q(y) \triangleq \sum_u e^{-td(u, y)}\) is the normalization factor, and \(t\) is the value that makes \(D_0(t) \triangleq \mathbb{E}[d(u, y)]\) or \(R_0(t) \triangleq I(u; y)\) equal to the desired \(D\) or \(R\); in the latter case this \(t\) is denoted by \(\tau_0(R)\).

**Proof:** See Appendix [LA].

In general, for any \(t > 0\) (not necessarily equal to \(\tau_0(R)\)), we call \(p(u | y) = e^{-td(u, y)}/Q(y)\) of the above form, or the corresponding \(p(y | u)\), a test channel of the symmetry source coding problem. It is trivial to verify the following symmetry properties of such a test channel:

**Proposition 2:** Given the \(p(y)\) and \(d(u, y)\) from a symmetric source coding problem over \(\mathbb{G}\), let \(p(u | y) = e^{-td(u, y)}/Q(y)\) with \(Q(y) \triangleq \sum_u e^{-td(u, y)}\) for some arbitrary \(t > 0\), then \(p(u) \triangleq \sum_y p(u | y)p(y)\) is a uniform distribution, and \(p(y | u) \triangleq p(y)p(u | y)/p(u)\) satisfies \(p(y | u)(u | y) = p(y | u)\).

It is also possible to prove that \(R_0(t)\) is an increasing function of \(t\) while \(D_0(t)\) is decreasing. Intuitively, given \(t\) and the corresponding \(p(u | y)\), for each “typical” \(y\) w.r.t. \(p(y)\), the probability that an independent \(u\) typical w.r.t. \(p(u)\) is jointly typical with \(y\) is approximately \(2^{-nI(u; y)}\) so on average there are \(2^{n(\tau_0(t) - R_0(t))}\) jointly typical sequences \(u\) in a random codebook \(\mathcal{U}\), and as long as \(R > R_0(t)\) one such \(u\) likely exists that will yield an average distortion close to \(D_0(t)\). In practice, the quantization algorithm is necessarily non-ideal, and the actual rate \(R\) and average distortion \(D\) could be slightly larger than resp. \(R_0(t)\) and \(D_0(t)\).

\[
C. \text{ The Random-Coding Loss of M-ary MSE Quantization}
\]

Proposition [1] gives the minimum \(G(\Lambda) = \sigma^2 \rho^2/n = (2^R/M)^2 D\) achievable with \(M\)-ary MSE quantization at each rate \(R\). This is less than the optimal \(G^*\) and we call the corresponding shaping loss \(10 \log_{10}(G(\Lambda)/G^*)\) the random-coding loss as random coding is one way to achieve it. The random-coding loss measures the suboptimality of the period- \(M\) structure of \(\Lambda\); as shown in Fig. [1] for \(M = 2\) and \(M = 4\), it is very small for large \(M\) and moderate \(R\), meaning that \(M\)-ary MSE quantization is near-optimal in such cases.

III. THE BINARY LDGM QUANTIZER

Previous works such as [3], [14], [24–26] suggest that LDGM-based code constructions are good candidates for approaching the performance limit in Proposition [1] for symmetric source coding problems and, in particular, achieve near-zero shaping loss in MSE quantization. In this and the next section, we will carry out a deeper analysis on the use of LDGM codes with BP in the simpler binary case (i.e. \(M = 2\) and \(\mathbb{G} = \mathbb{Z}_2\)), while in Section [V] we will consider non-binary constructions that can be applied to more general symmetric source coding problems and achieve lower random-coding loss in MSE quantization.

In the quantization algorithm for binary codes, the \(a \text{ priori}\) information (priors), extrinsic information and BP messages are likewise binary and can be viewed as probability distributions of binary random variables. In this paper, they are mainly represented by probability tuples, e.g., \(\mu = (\mu(0), \mu(1))\), \(\mu(b)\) being the probability that the variable equals \(b \in \mathbb{Z}_2\); the corresponding log-likelihood ratio (LLR) is \(\ell(\mu) \triangleq \ln(\mu(0)/\mu(1))\). For convenience, these tuples are implicitly normalized; that is, when we write \(\mu(b) = q_b\), we actually make \(\mu(b) = q_b/(q_0 + q_1)\) so that \(\mu(0) + \mu(1) = 1\), and later appearances of \(\mu(b)\) refer to this normalized value. \(\mu \circ \mu' \triangleq (\mu(0)\mu'(0), \mu(1)\mu'(1))\) (implicitly normalized) and \(\mu \circ \mu' \triangleq (\mu(0)\mu'(0)+\mu(1)\mu'(1), \mu(0)\mu'(1)+\mu(1)\mu'(0))\) are the variable-node and check-node operations in LDPC literature, which are associative and thus immediately applicable to more than two probability tuples. More generally, if we view \(\mathbb{Z}_2^m\) as a vector space over field \(\mathbb{Z}_2\) and let \(\mathcal{C}\) be an affine subspace of it, then given \(m-1\) probability tuples \(\lambda_{-1} \triangleq \lambda_1, \ldots, \lambda_{m-1}, \lambda_{m+1}, \ldots, \lambda_m\), we may define \(\nu(\mathcal{C}; \lambda_{-1})\) as the probability tuple \(\nu\) with \(\nu(b) = \sum_{b \in \mathcal{C}, b_0 = b} \prod_{j \neq i} \lambda_j(b_j)\), \(b \in \mathbb{Z}_2^m\); \(\otimes\) and \(\oplus\) are then its special cases with \(\mathcal{C}\) being respectively the (3, 1) repetition code and the (3, 2) single-parity-check code. Here \(\mathcal{P} \triangleq (1, 0), \mathcal{T} \triangleq (0, 1)\) and \(\mathfrak{F} \triangleq 0.5 + \mathfrak{F} \triangleq 0.5 + \mathfrak{F}\) are respectively the “sure-0”, “sure-1” and “unknown” probability tuples. We also define \(H(\mu) \triangleq H_2(\mu(0))\) and \(I(\mu) \triangleq 1 - H(\mu)\), where \(H_2(p) \triangleq -p \log p - (1-p) \log(1-p)\) is the binary entropy function.

A. Outline of the Quantizer and Its Analysis

When \(\mathcal{G} = \mathbb{Z}_2\), we use the binary LDGM codebook

\[
\mathcal{U} = \mathcal{U}(a) = \{u = u(b, a) \triangleq c \triangleq bG \oplus a | b \in \mathbb{Z}_2^n\}, \quad (6)
\]

where \(G = (g_{ij})_{n_c \times n_c}\) is the sparse generator matrix randomly generated according to the degree distributions optimized below, the matrix multiplication in \(bG\) as well as \(\oplus\) are modulo-2, \(n_c \triangleq n, n_b \triangleq nR\), and \(R\) is the rate of the LDGM code. A fixed scrambling sequence \(\alpha\) randomly chosen from \(\mathbb{Z}_2^n\) has been introduced in [6], which ensures that every point

![Fig. 1. Random-coding losses of binary and 4-ary MSE quantization. In the binary case, the minimum loss is approximately 0.0945 dB at \(t = 3.7114\) and \(R = R_0(t) = 0.4143\) b/s. In the 4-ary case, the minimum loss is only 0.0010 dB at approximately \(t = 2.0053\) and \(R = R_0(t) = 0.9550\) b/s.](image-url)
\( \mathbb{Z}_2^n \) is covered by \( 2^{nR} \) of the \( \mathcal{U}(a) \)'s, even though each \( \mathcal{U}(a) \) may be “clumped” around certain points in \( \mathbb{Z}_2^n \). This will be essential in results such as Proposition 3 below.

The quantization algorithm is based on belief propagation, with a *decimation* step that makes hard decisions in order to help the algorithm converge \([14, 26]\). Proper analysis of the decimation steps is essential to a good understanding of the algorithm and its performance characteristics, so before presenting the algorithm in detail, we first outline our analytical approach. We consider a fixed \( \mathcal{G} \) for the rest of this section; that is, all probabilities are implicitly conditioned on \( \mathcal{G} \). Given the source sequence \( y \), we assign a probability to each \( \nu \) according to the test channel \( p(u | y) = e^{-td(u,y)}/Q(y) \), which has the same form as the optimal one in Proposition 1 and makes Proposition 2 applicable; here \( R_0(t) = I(u;y) \) is generally close, but not equal, to \( R \) (although we will still assume that \( R_0(t) > 0 \)), and its choice will be briefly covered in Section IV-B. Ignoring normalization factors depending only on \( y \), the probability thus assigned is

\[
q(u | y) = \prod_{j=1}^n e^{-td(u_j,y_j)} = e^{-ntd(u,y)}. \tag{7}
\]

As any \( u \in \mathbb{Z}_2^n \) is equal to \( u(b,a) \) for \( 2^nR \) distinct \( (b,a) \)'s, (7) also gives a joint distribution of \( u \) and \( a \), which is \( q(b,a | y) = e^{-ntd(u(b,a),y)} \) without normalization. If \( b \) and \( a \) were sampled from this distribution, all \( 2^n \) possible values of \( u \) would be obtained with probabilities proportional to (7), and the expected distortion would simply be the \( D_0(t) \) from Proposition 1. In reality, \( a \) is fixed first, independently from \( y \), and given \( y \) the quantizer has to choose a \( b \), or equivalently a \( u \) from \( \mathcal{U}(a) \), but under certain conditions this will, in a sense, yield the same result as random sampling of \( b \) and \( a \) and thus the same distortion \( D_0(t) \).

To make this notion of “same result” rigorous, prior to the determination of \( a \) and actual quantization, we first generate two sequences of respectively \( n_c = n \) and \( n_b = n \) i.i.d. uniform samples over \( \{0, 1\} \), \( \omega^a \) and \( \omega^b \), as the source of randomness. The determination of \( a \) and \( b \) in quantization are then divided respectively into \( n_c \) a-steps that determine \( a_1, a_2, \ldots, a_{n_c} \) successively, followed by \( n_b \) b-steps determining \( b_1, \ldots, b_{n_b} \). In an a-step \( j \), we compute a binary probability tuple \( \nu^a_j \) and set \( a_j = 1 \) \( \nu^a_j \geq \bar{\nu}^a_j(0) \), and similarly in b-step \( \nu^b_j \) is used to compute \( b_j = 1 \) \( \nu^b_j \geq \bar{\nu}^b_j(0) \). The two processes can then be described by the way \( \nu^a_j \) and \( \nu^b_j \) are computed.

Definition 2: The above quantization process is called the *true probabilistic quantizer* (TPQ), if \( \bar{\nu}^a \) and \( \bar{\nu}^b \) are set to the conditional probabilities \( \nu^a_j \) and \( \nu^b_j \) corresponding to \( q(b,a | y) \), that is,

\[
\nu^a_j(a) = \sum_{a \in A_j(a)} q(b,a | y), \tag{8}
\]

where \( A_j(a) \) contains those \( a \) with \( a_j = a \) and \( a_{1}, \ldots, a_{j−1} \) matching the values determined in a-steps \( 1, \ldots, j-1 \), and

\[
\nu^b_j(b) = \sum_{b \in B_j(b)} q(b,a | y), \tag{9}
\]

where \( a \) has been determined in the a-steps and \( B_j(b) \) contains those \( b \) with \( b_j = b \) and \( b_{1}, \ldots, b_{j−1} \) matching the values determined in the previous b-steps.

Definition 3: The quantization process is called the *BP probabilistic quantizer* (BPPQ), if it sets each \( \bar{\nu}^a_j \) to \( \bar{\nu} \) and \( \bar{\nu}^b_j \) to \( \nu^b_j \), the BP approximation of \( \nu^b_j \) above. These \( \nu^b_j \)'s, unlike those used by TPQ, do not depend on \( y \), so \( a \) can be determined before quantization with a given \( y \), which is necessary for a useful scheme.

Clearly, the TPQ yields each possible \( b \) and \( a \) with probability proportional to \( q(b,a | y) \), so the average distortion is \( D_0(t) \) as stated above. For each TPQ instance associated with some \( y \), \( \omega^a \) and \( \omega^b \), if the *synchronization conditions*

- \( \nu^a_j = \bar{\nu} \) for all \( j \), and
- \( \nu^b_j \) is precisely computed by BP for all \( i \),

are met in every step, then the corresponding BPPQ instance will also yield the same \( a \) and \( b \); if this is true for all TPQ instances, the BPPQ’s average distortion will be \( D_0(t) \) as well. Consequently, we can base our quantization algorithm on the BPPQ, and optimize the degree distributions so that the synchronization conditions are met asymptotically for large block sizes \( n \) and BP iteration counts \( L \), under as high a \( t \) (and thus low \( D_0(t) \)) as possible. These conditions cannot be met precisely at finite \( n \) and \( L \), and the BPPQ will lose synchronization with the TPQ and yield higher distortion, but a *recovery algorithm* has been proposed in \([12]\) that can minimize the impact of such synchronization loss.

\[ B \text{. The Quantization Algorithm} \]

Fig. 2(a) shows the factor graph that can be used to estimate each \( \nu^a_j \) and \( \nu^b_j \) given by (8) and (9). The *a priori* information of each variable \( u_j \), denoted \( \lambda^u_j \), is given by

\[
\lambda^u_j(u) = e^{-td(u_j,y_j)}, \tag{10}
\]

which corresponds to a factor in \( q(b,a | y) \). The priors of the \( a_j \)'s and \( b_j \)'s, denoted \( \lambda^a_j \) and \( \lambda^b_j \) respectively, are set according to the ranges of summation in (8) and (9). That is, when estimating \( \nu^a_j \), we know from (8) that \( \lambda^a_j = \frac{1}{\bar{\nu}^a_j} \) for \( j' < j \) with \( a_1, \ldots, a_{j−1} \) taking the previously determined values, while the remaining \( \lambda^a_j \)'s and all the \( \lambda^b_j \)'s are \( \bar{\nu} \); similarly, when estimating \( \nu^b_j \) in (9) we let all \( \lambda^a_j = \frac{1}{\bar{\nu}^a_j} \), while \( \lambda^b_j \) is \( \frac{1}{\bar{\nu}^b_j} \) if \( b_j \) has been determined (decimated), and \( \bar{\nu} \) otherwise.

The function nodes, shown as black squares in Fig. 2(a) represent the relationship \( u = \mathcal{G} \oplus a \), so similar to LDPC we call them *check nodes*. In this way, \( \nu^a_j \) and \( \nu^b_j \) are simply the exact marginals (true extrinsic information) of variable \( a_j \) and \( b_j \) on the factor graph when using those priors, and they can be approximated by respectively \( \nu^a_j \) and \( \nu^b_j \), the margins (BP extrinsic information) computed with the BP (a.k.a. sum-product) algorithm.

The quantization algorithm is essentially an implementation of BPPQ: \( a \) is chosen randomly, and then in each b-step, \( \nu^b_j \) is
computed with a number of BP iterations as an approximation of \( \nu_{b_j}^2 \), and \( b_i \) is decimated to 1 \([\omega^1_i \geq \nu^2_i(0)]\). In practice, to reduce the number of iterations needed in the entire quantization process, BP message values from earlier b-steps are reused, and multiple b-steps are carried out after each BP iteration, but this has little impact on the theoretical analysis below. The algorithm can thus be outlined in Fig. 2, where, apart from the priors \( \lambda^j_1 \) and \( \lambda^j_2 \), extrinsic information \( \nu^j_i \), we also use \( \mu^b_{ji} \) to denote a BP message from variable node \( b_j \) toward check node \( j \) (the check node to the left of \( u_j \)), and \( \mu^b_{ji} \) for the BP message in the inverse direction, as indicated by the arrows in Fig. 2 and these BP messages are binary probability tuples here as well: \( N^{bc} = N^{cb} \) is the set of indices \( i \) for which there exists an edge between check node \( j \) and variable node \( b_i \), and \( N^{bc} = N^{cb} \) is defined similarly. To follow BPPQ exactly, in each decimation step, the bit index \( i^* \) is chosen sequentially\(^3\) and the decimated value is \( b^* \in \mathbb{Z}_2 \) with probability \( \nu^b_i(b^*) \), which is equivalent to letting \( b^* = 1 \left[ \omega^1_i \geq \nu^2_i(0) \right] \); this is called the probabilistic decimator (PD) and is more amenable to analysis.\(^4\) An intuitive alternative is the greedy decimator (GD) which always decimates the “most certain” bit, among the set \( E \) of undecided bit indices, to its most likely value, i.e.

\[
(i^*, b^*) = \arg \max_{(i,b) \in E \times \mathbb{Z}_2} \nu^b_i(b).
\] (11)

As expected, the GD yields better performance than the PD, but it is more useful in practice, although we will not attempt to analyze it.

In practice, it is important to control the amount of decimation in each iteration (which we call the pace of decimation), so that distortion performance can be optimized under a limited number of iterations. Moreover, the recovery algorithm mentioned at the end of Section III-A is also necessary for good performance. However, these issues can safely be ignored in the theoretical analysis in this paper, and thus will not be considered in detail here; practical algorithms for them have been proposed in [16] and [12].

IV. ASYMPTOTIC ANALYSIS OF THE SYNCHRONIZATION CONDITIONS

Compared to the analysis of LDPC decoding via density evolution, the analysis of the LDGM quantizer is complicated by its use of decimation based on extrinsic information, as well as the lack of a natural reference codeword corresponding to the all-zero codeword in LDPC analysis. To solve these problems, we have introduced the TPQ, the BPPQ and the synchronization conditions, and in this section we will show that TPQ gives a reference codeword that allows the synchronization conditions to be analyzed with density evolution methods, for asymptotically large block length \( n \) and iteration count \( L \).

We use LDGM codes that are regular at variable nodes \( b_i \) and irregular at the check nodes for quantization, as suggested by the LDGM-LDPC duality in [8]. The degree distribution is described by \( d_b \geq 2 \), the number of 1’s in each of the \( n_b \) rows of \( G \), as well as the \( w_d \)’s, each of which representing the fraction of columns in \( G \) with \( d \) 1’s; we also use \( v_d \equiv dw_d/\left(Rd_b \right) \) to denote the fraction of 1’s residing in these columns among the \( nRd_b \) 1’s in the entire \( G \). All degrees are assumed to be at least 1. These degree distribution parameters satisfy the constraints

\[
\sum_d w_d = 1, \quad \sum_d v_d = 1, \quad w_d \geq 0 \text{ for } d = 1, 2, \ldots. \quad (12)
\]
Strictly speaking, a given degree distribution cannot be followed exactly at arbitrary block lengths \( n \) since \( nR \) and the \( nwd's \) are not necessarily integers. To avoid this problem, for each \( n \) we pick \( R(n) \) and \( w_d(n) \) such that \( nR(n) \) and all \( nwd(n) \)'s are integers, and at the same time \( R(n) \to R \) and \( w_d(n) \to w_d \) as \( n \to \infty \). Denoting by \( w \) and \( w(n) \) the vector comprised of respectively \( w_d's \) and the \( w_d(n) \)'s, we can now redefine \( n_b = nR(n) \) and let \( G(n)(db, w) \) be the set of \( G' \)s with rate \( R(n) \) and degree distribution given by \( \{db, w(n)\} \).

At each \( n \), let \( G \) be uniformly distributed in \( G(n)(db, w) \), and we then have an ensemble of TPQ and corresponding BPQ instances, with one for each \((G, y, \omega^a, \omega^b)\) tuple; when \( G, y, \omega^a \) and \( \omega^b \) are viewed as random variables, so are the resulting \( a \) and \( b \) from either quantizer, as well as the BP priors, messages and extrinsic information. During the analysis of the synchronization conditions below, all random variables will be defined over the TPQ ensemble. In other words, the bits in \( a \) used as input for the quantization algorithm are chosen sequentially as \( a_j = 1 \) \( \omega^a_j \geq \omega^a_j(0) \), \( j = 1, \ldots, n_a \), and the BP priors, messages and extrinsic information in each iteration are then defined by following the algorithm in Fig. 3 except that the sequential decimation of each \( b_i \) in \( b \) uses \( \nu^b \) from the TPQ formula \( b_i = 1 \) \( \nu^b_i \geq \nu^b_i(0) \)) instead of the BP extrinsic information \( \nu^b_i \), thus yielding the \( b \) from TPQ at the end, and we then say the quantization algorithm follows TPQ. In this way, we can investigate the synchronization conditions when all previous a- and b-steps have yielded TPQ’s decimation result, i.e. whether the BPQ will remain synchronized with the TPQ if it is previously so. We denote the \( b \) and \( a \) from TPQ by \( b^* \) and \( a^* \) respectively, and use them or the corresponding \( u^* \triangleq c^* \triangleq b^*G \oplus a^* \) as the reference codeword for density evolution. Conditioned on a fixed \( G \), the joint distribution of \( b^*, a^* \) and \( u^* \) can be obtained following the discussion in Section II-A as follows:

**Proposition 3:** Conditioned on a fixed \( G \) (omitted in the conditional probabilities below), \((b^*, a^*)-u^*-y\) as well as \((b^*, a^*)-u^*-y\) for any \( j \) form Markov chains, 
\[
p(b^*, a^*) \cdot u^* = 2^{-n_b} \text{ (i.e. uniform) for any } (b^*, a^*) \text{ satisfying } b^*G \oplus a^* = u^* \text{, while } p(u^* | y) = \prod_j p_{u_j}(u_j^*) \text{ and } p(y) = \prod_j p_{y_j}(y_j) \text{ with } p(u_j = 1) = c^{-d(u,G)}/Q(y) \text{ being the test channel chosen in Section II-A and } p(y) \text{ being the source pdf. Consequently, } p(u^*) = \prod_j p_{u_j}(u_j^*) \text{ is uniform because } p(u) \text{ is so according to Proposition 2 while } p(y | u^*) = \prod_j p_{y_j | u_j}(y_j | u_j^*) \text{ and } p(b^*, a^*) = 2^{-n+\nu} \text{ is uniform as well.}
\]

**Proof:** See Appendix [1B]

The need to have an explicit reference codeword in density evolution necessitates the use of some new notations; first of all, we will introduce these notations and express some known results in terms of them.

### A. Review of Binary Message Densities and Their Properties

Given the reference codeword, each variable node \( b_i, u_j \) or \( a_j \) then corresponds to a bit in the reference codeword, namely \( b_i^*, u_j^* \) or \( a_j^* \), which is a binary random variable. Consequently, each probability tuple involved in BP can also be assigned such a bit from the reference codeword as its reference bit according to the associated variable node. In particular, for binary LDGM quantization, the reference bit of each \( \lambda_1^b, \nu_1^b, \nu_1^b, \mu_{ij}^b \) and \( \mu_{ij}^b \) is \( b_i^* \), while that of \( \lambda_1^a \) and \( \lambda_2^a \) are \( u_j^* \) and \( a_j^* \) respectively.

A message density (or simply density) is a conditional probability distribution of a probability tuple (itself a random variable) given its reference bit, and is usually shown in bold; for example, the density of \( \mu_{ij}^b \) (with reference bit \( b_i^* \)) can be denoted by \( \mu^b \), and we then write \( \mu_{ij}^b \sim \mu^b \). Such a density \( \mu \) can be concretely represented by the conditional pdf or pmf of \( \mu(0) \) or the LLR \( l(\mu) \) given \( b \) when we let \( \mu | b \sim \mu \), and they are respectively denoted \( \mu_{ij}(p(b)) \) and \( \mu_{ij}(1 | b) \). We also formally write \( \mu(b | \mu) \) as the conditional pdf if the actual representation of the probability tuple is not of concern, so that \( \mu(b | \mu) \) implies \( \mu | b \sim \mu \).

Unless otherwise noted, the distributions of all the random variables here, particularly the densities of probability tuples, are defined with respect to the entire ensemble of TPQ and BPQ instances involving all \( G \in G(n)(db, w) \). Sometimes we will also limit our consideration to those instances involving a specific \( G \) or subset of \( G' \)s (e.g. those with certain loop-free neighborhoods), and obtain the conditional distributions and message densities over this sub-ensemble denoted by e.g. \( \mathcal{E} \); for example, if the conditional probability density \( p(\mu(b | \mathcal{E})) \) can be represented by message density \( \mu \), then we may write \( \mu(b | \mathcal{E}) \sim \mu \). The properties of message densities given below are clearly applicable to such conditional densities as well.

The symmetry condition of message densities plays an important role in both LDPC analysis [17] and here. Based on the above definitions, symmetry can be defined as follows:

**Definition 4:** A message density \( \mu \) is said to be symmetric if
\[
(13) \quad \mu_{ij}(p | 0) = \mu_{ij}(1 - p | 1),
\]
\[
(14) \quad (1 - p) \cdot \mu_{ij}(p | 0) = p \cdot \mu_{ij}(1 - p | 0),
\]
for all \( p \in [0,1] \). If \( \mu | b \sim \mu \), we then say the random probability tuple \( \mu \) has a symmetry density (is symmetric) with respect to (w.r.t.) \( b \); if not stated explicitly, the reference bit \( b \) refers to that of \( \mu \) defined above.

A message density \( \mu \) can be viewed as a binary-input channel \( \mu(b | \mu) \) with the reference bit \( b \) as input and the probability tuple \( \mu \) as output. Under this view, [13] is simply a kind of input symmetry of this channel, commonly used in LDPC literature when they assume that the correct codeword used as reference is all-zero. Condition [14] is about the “consistency” of the density, i.e. whether each possible channel output \( \mu(p(1 - p)) \) has its likelihood ratio \( \mu_{ij}(p | 0) / \mu_{ij}(p | 1) \) equal to \( p/(1 - p) \), which can also be formally expressed as \( \mu_{ij}(0) / \mu_{ij}(1) = \mu(0)/\mu(1) \) for any \( \mu \). In this paper, \( p(b) \) is often uniform over \( \mathbb{Z}_2 \); if so, then when \( \mu \) has a symmetric density w.r.t. \( b \), i.e. \( p(\mu(b = 0))/p(\mu(b = 1) = \mu(0)/\mu(1) \), we have
\[
(15) \quad p(b | \mu) \propto p(\mu | b) \propto \mu(b), \text{ i.e. } p(\mu | b) = \mu(b),
\]
where \( \propto \) denotes equality up to a factor not containing \( b \). In LLR form [14] becomes \( \mu_{ij}(1) / \mu_{ij}(1) = e^i \), which is exactly the symmetry condition in LDPC literature.
Naturally, for any symmetric binary-input channel, its likelihood function has a symmetric density:

**Proposition 4:** Let \( b \) be a binary random variable, \( y \) be another random variable taking values in \( \mathcal{Y} \) and with conditional pmf or-pdf \( p(y|b) \), and \( \mu \) be the probability tuple giving the likelihood of \( y \), i.e. \( \mu(b') = p(y|b') \). If there exists an measure-preserving group action \( \psi_b(\cdot) \) of \( \mathbb{Z}_2 \) on \( \mathcal{Y} \), such that \( p_{y|b}(y|b) = p_{y|\psi_b(y)}(\psi_b(y)|0) \), then \( \mu|b \sim \mu \) is a symmetric density.

**Proof:** Theorem 4.27 in [27] is a proof for the case \( \mathcal{Y} = \mathbb{R} \) and \( \psi_b(\cdot) \) being \( \psi_1(y) = -y \). This generalization is proved similarly; see Appendix I-C.

Given a symmetric density \( \mu \), if we let \( b \) be an equiprobable binary random variable and \( \mu \) satisfying \( \mu|b \sim \mu \), then for any possible value \( \mu_b' \) of \( \mu \), we have \( p_{\mu|b}(b|\mu_b') = \mu(b') \), so the entropy \( H(b|\mu = \mu_b') = H(\mu_b') \); taking the expectation over \( \mu \) we get \( H(b|\mu) = E[H(\mu)] \) and \( I(b;\mu) = E[I(\mu)] \). We thus define \( H(\mu) = E[H(\mu)] \) and \( I(\mu) = E[I(\mu)] \), and call them respectively the entropy and mutual information (MI) of the symmetric density \( \mu \).

Given densities \( \mu_1, \ldots, \mu_m \) and weights \( \alpha_1, \ldots, \alpha_m \in [0,1] \) with \( \sum \alpha_i = 1 \), we can straightforwardly define the convex combination \( \mu = \sum \alpha_i \mu_i \), e.g. by making \( \mu_i(p|b) = \sum \alpha_i \mu_i(p|b) \). This definition can naturally be extended to an arbitrary random vector \( \mu_1 \) may be called the density of \( \mu \) conditioned on \( I \), while the message density over the entire ensemble \( \mu|b \sim \mu \) is called \( \mu \)'s density (averaged) over all \( I \in \mathcal{X} \); in this case, \( \mu \) is a convex combination of \((\mu_1)_{I \in \mathcal{X}}\) weighted by the pmf or pdf of \( I \).

Convex combinations of symmetric densities remain symmetric (a more general result, Proposition 26 will be proved in detail). Conversely, for any \( q \in [0,1] \), we may let \( q(0) \equiv q \) and \( q(1) \equiv 1 - q \), and define \( b \) and \( \mu \) such that given \( b \in \mathbb{Z}_2 \), \( \mu = (q,1-q) \) with probability \( q(0) \) and is \((1-q,q) \) otherwise, i.e. the conditional pmf

\[
p(\mu|b) = \sum_{e \in \mathbb{Z}_2} q(b \oplus e) \cdot 1 \left[ \mu(b') = q(b' \oplus e), b' = 0,1 \right].
\]

then the density \( \mu|b \sim D_q = D_{1-q} \) is symmetric, and any symmetric density can be expressed as a convex combination of the family \( (D_q)_{q \in [0,1/2]} \). In this way, many results need only to be proved for \( D_q \), and they can then be applied to symmetric densities by linearity.

The \( \nu(\cdot;\cdot) \) operator for probability tuples defined in Section III which includes \( \oplus \) and \( \otimes \) as special cases, can naturally be applied to densities using the following definition:

**Definition 5:** Given a deterministic affine subspace \( C \) of \( \mathbb{Z}_2^m \) and \((m - 1)\) message densities denoted by \( \lambda_{-i} \equiv (\lambda_1, \ldots, \lambda_{i-1}, \lambda_{i+1}, \ldots, \lambda_m) \), we let \( b = (b_1, \ldots, b_m) \) be uniformly distributed over \( C \), construct \( m - 1 \) random binary probability tuples \( \lambda_{-i} \) such that for any \( j \neq i \),

- \( \lambda_j \) and \( \lambda_j' \) depend only on bit \( b_j \) in \( b \), with \( \lambda_j | b_j \sim \lambda_j \) and \( \lambda_j' | b_j \sim \lambda_j' \) both being symmetric densities,

\( \lambda_j' \preceq \lambda_j \) w.r.t. \( b_j \). Now let \( \nu_i = \nu(C; \lambda_{-i}) \), then \( \nu_i' \preceq \nu_i \).

**Proof:** Similar to [27] Lemma 4.82; we will give the proof of the more general Proposition 30 in Appendix I-K. Note that \( \nu_i \) being a sufficient statistic is important; the result would not hold if \( \nu_i \) loses too much information from \( \lambda_{-i} \) that \( \nu_i' \) happens to retain.
Proposition 9: Let $C$ be a deterministic affine subspace of $\mathbb{Z}_2^n$, and $\Lambda_{x_{-i}}$ and $\Lambda'_{x_{-i}}$ each be $m-1$ symmetric densities with $\Lambda'_{x_{-i}} \leq \Lambda_i$ for all $j \neq i$, then $\nu(C; \Lambda'_{x_{-i}}) \leq \nu(C; \Lambda_{x_{-i}})$.

Proof: This is an obvious corollary to Proposition 8.

Physical degradation relationships enable us to prove the closeness of individual probability tuples from the synchronization conditions by comparing the average MIs:

Proposition 10: Given an equiprobable binary random variable $b$ and two random probability tuples $\mu_1$ and $\mu_2$ such that $\mu_2 \leq \mu_1$ w.r.t. $b$. If $\mu_1 \left| b \sim \mu_1$ and $\mu_2 \left| b \sim \mu_2$ are both symmetric densities, then

$$E \left[ (\mu_1(0) - \mu_2(0))^2 \right] \leq \frac{\ln 2}{2} (I(\mu_1) - I(\mu_2)).$$

This implies that $I(\mu_2) \leq I(\mu_1)$, which is also obvious from the data processing inequality.

Proof: Similar to [24] Lemma 15]; see Appendix I.

Conversely, we have the following result:

Proposition 11: For any $\epsilon > 0$ there exists a $\delta > 0$ such that, given an equiprobable binary random variable $b$ and two random probability tuples $\mu_1$ and $\mu_2$ with $\mu_1 \left| b \sim \mu_1$ and $\mu_2 \left| b \sim \mu_2$ being symmetric densities, if $|I(\mu_1) - I(\mu_2)| \geq \epsilon$, then $E \left[ (\mu_1(0) - \mu_2(0))^2 \right] \geq \delta$.

Proof: Since $E \left[ |I(\mu_1) - I(\mu_2)| \right] \geq |I(\mu_1) - I(\mu_2)| \geq \epsilon$, and $|I(\mu_1) - I(\mu_2)| \leq 1$ with probability 1, we have

$$\Pr \left[ |I(\mu_1) - I(\mu_2)| \geq \epsilon/2 \right] \geq \epsilon/2.$$

Now $I(\mu_1)$ is a continuous function of $\mu_1(0)$ over $[0, 1]$ and thus uniformly continuous, so there exists a $\delta' > 0$ such that $|I(\mu_1) - I(\mu_2)| \geq \epsilon/2$ implies that $|\mu_1(0) - \mu_2(0)| \geq \delta'$. Therefore, letting $\delta = (\delta')^2$, $\epsilon/2$ leads to the desired result.

An important class of symmetric densities is the erasure-like densities defined as follows:

Definition 6: For $x \in [0, 1]$, let $b$ be a binary random variable, and $\mu$ be a random probability tuple that equals $\mathbb{E}$ with probability $x$ and $\mathbb{W}$ with probability $1-x$, then we define the resulting density $\mu \left| b \sim \mathbb{E}$ and call such densities erasure-like. In particular, $E_0$ and $E_1$ are respectively the always-unknown and always-sure densities.

Erasure-like densities are thus similar to binary erasure channels (BECs) and have the following simple properties, whose proofs are omitted here:

Proposition 12: For any $x, x_1, x_2 \in [0, 1]$,

$\bullet$ $E_x$ is symmetric with $I(E_x) = x$;

$\bullet$ $E_{x_1} \otimes E_{x_2} = E_{x}$ where $x = 1 - (1 - x_1)(1 - x_2)$;

$\bullet$ $E_{x_1} \otimes E_{x_2} = E_{x_1 \cdot x_2}$;

$\bullet$ $\sum_i \alpha_i E_{x_i} = E_x$, where $x = \sum_i \alpha_i x_i$;

$\bullet$ If $x_1 \leq x_2$, then $E_{x_1} \leq E_{x_2}$.

Moreover, the $\nu(C; \cdot)$ operator preserves erasure-like densities:

Proposition 13: If $C$ is a deterministic affine subspace of $\mathbb{Z}_2^n$ and $\Lambda_{\cdot \cdot \cdot \cdot}$ are $m-1$ erasure-like message densities, then $\nu \triangleq \nu(C; \Lambda_{\cdot \cdot \cdot \cdot})$ is also erasure-like.

Proof: See Appendix I.

B. Synchronization at b-steps

We now analyze the synchronization condition at the $i$-th b-step of the TPQ: namely, assuming that all a-steps and the previous b-steps have followed TPQ to yield $a = a^*$ and $b_i = b_i^*$ for all $i < i$, whether the $\nu_i$ obtained in b-step $i$ approaches $\nu_i^{\star}$ after a large number of BP iterations, so that BPPQ can maintain synchronization with the TPQ after this b-step.

BPPQ in the actual quantization algorithm starts with the $\mu_{\beta_i}$'s being all-$\mathbb{E}$ and updates them with BP across all b-steps. To simplify the analysis of one specific b-step here, we instead assume that the $\mu_{\beta_i}$'s are reinitialized to all-$\mathbb{E}$ at the beginning of this b-step. BP is carried out for $L$ iterations, and the resulting $\nu_i$ is used as the $\nu_i^0$ in decimation. While such treatment is inefficient in practice, it is straightforward to prove via physical degradation arguments that, in terms of whether the synchronization condition is asymptotically satisfied (in the sense of Proposition 18 below), it is equivalent to the actual algorithm. This $\nu_i$ obtained from $L$ BP iterations starting from all-$\mathbb{E}$ $\mu_{\beta_i}$'s is henceforth denoted by $\nu_i^{\beta_i(L)}$; on the other hand, if every $\nu_i^{\beta_j}$ is hypothetically initialized to hard decision $b_i^\star$ before the $L$ BP iterations, the resulting $\nu_i$ is denoted by $\nu_i^{\beta_i(L)^{\star}}$.

In the factor graph Fig. 2(a), these $L$ BP iterations involve a neighborhood $N_i^\star = N_{b_i}^\star$ of the variable node $b_i$, which can be further divided into the interior part $N_i^\#_\star$ and the border part $N_i^\star_\star$. Fig. 4(b) illustrates the structure of the factor graph around variable node $b_i$, with $N_i$ being the entire unshaded region, in which each layer shown in Fig. 4(a) corresponds to one BP iteration. Only the priors of the variable nodes in $N_{i+1}^\#_\star$, and the initial BP messages from variable nodes in $N_i^\star_\star$ to check nodes in $N_i^\#_\star$ (labeled with $\mu_{i+1}^{\beta_i(L)}$ in Fig. 4(b)), affect $\nu_i^{\beta_i(L)}$ and $\nu_i^{\beta_i(L)^{\star}}$. Below we will use $b_i \in N_i^\#_\star$ to express that the variable node $b_i$ (denoted by $b_i$ to avoid confusion with the value of $b_i$) is in the neighborhood $N_i^\#_\star$.

Analysis of the BP process frequently requires $N_i^\#_\star$ to be loop-free. Given the degree distributions, $L$, $n$, and $i$, we use $G_n(L)$ to denote the sub-ensemble of $G_n(d_b, u)$ with a loop-free $N_i^\#_\star$. If $G$ is uniformly distributed over $G_n(L)$, the probability that $G \not\in G_n(L)^{\#_\star}$ obviously does not vary with $i$, and is thus denoted by $P_n^{\#_\star}$. Using the methods employed in LDPC analysis (e.g. the proof of Theorem 1 in [28]), it is possible to prove that

$$\lim_{n \to \infty} P_n^{\#_\star} = 0,$$

for any degree distribution and $L$.

Now consider a fixed $G \in G_n(L)$. Define

$$C \triangleq \{(b, a, u) \mid u = bG + a\},$$

then $C$ is a linear (and thus affine) subspace of $\mathbb{Z}_2^n + n + m^n$, and by Proposition 3 (b, a, u) is uniformly distributed over it when conditioned on $G$. Given any priors $\lambda_i^b \triangleq (\lambda_1^b, ..., \lambda_n^b)$, $\lambda_i^a \triangleq (\lambda_1^a, ..., \lambda_n^a)$, $\lambda_i^u \triangleq (\lambda_1^u, ..., \lambda_n^u)$, the result of $\nu(C; \lambda_i^b, \lambda_i^a, \lambda_i^u)$ is then

$$\nu(b) = \sum_{(b, a, u) \in C} \prod_{i \neq \lambda_i^b(b_i)} \prod_j \lambda_i^a(a_j) \prod_j \lambda_i^u(u_j),$$

which is also the true extrinsic information at $b_i$ on the factor graph in Fig. 2(a) given those priors. In particular, if the priors are those used in the quantization algorithm, i.e.
We need to prove now that whether the priors at the variable nodes correspond to a factor

\[ f_j(u_j, a_j, b) \triangleq 1 \{ u_j \oplus a_j \oplus (bG) \}_{j} = 0 \],

and variable \( a_j \) only occurs in this factor and the prior \( \tilde{\lambda}_j^a \) (correspondingly, the variable node \( a_j \) is only connected to check node \( j \)). By definition, the true extrinsic information at \( b_i \) over the complete factor graph is given by the product of the factors corresponding to the function nodes and to the priors at variable nodes other than \( b_i \), then summed over all variables other than \( b_i \). Here the summation over \( a_j \) involves just the two factors \( f_j(u_j, a_j, b) \tilde{\lambda}_j^a(a_j) \) it appears in, and when we let \( \tilde{\lambda}_j^a = \pi \), since \( \tilde{\lambda}_j^a(a_j) \) is always \( \frac{1}{2} \) while \( f_j(u_j, a_j, b) \) is once 0 and once 1 as \( a_j \) varies over \( \{0, 1\} \), this summation over \( a_j \) also gives a constant \( \frac{1}{2} \), thus eliminating the factor \( f_j(u_j, a_j, b) \); in other words, the check node \( j \) and variable \( a_j \) in the factor graph can be removed without affecting the true extrinsic information at \( b_i \), and what remains is \( \mathcal{N}_i \) along with a subgraph disconnected from it, so the true extrinsic information at \( b_i \) on the entire factor graph can equivalently be computed on just \( \mathcal{N}_i \), giving \( \mu_{b_i}^{(L)} \).

For \( \mathcal{P}_{b_i}^{(L)} \), we note that any variable node \( b_i \) in \( \mathcal{N}_i^- \) now has prior \( \tilde{\lambda}_i^b = b_i \), so in the summation formula yielding the true extrinsic information at \( b_i \), any non-zero term has \( b_i = b_i \). For any check node \( j \) in \( \mathcal{N}_i^+ \) connected to \( b_i \), the corresponding factor \( f_j(u_j, a_j, b) \) can then have \( b_i \) substituted for \( b_i \), thus breaking the edge between check node \( j \) and variable node \( b_i \). In this way \( \mathcal{N}_i \) also gets separated from the rest of the factor graph.

Remark 1: The scrambling sequence \( \alpha \) plays an important role in eliminating the impact of the possibly loopy part of the factor graph beyond \( \mathcal{N}_i^{(L)} \), thus allowing us to relate \( \nu^{b_i} \), which involves the entire factor graph, to its BP counterparts involving \( \mathcal{N}_i^{(L)} \) only. Incidentally, the closely related result about the relationship between exact and BP.
extrinsic information in LDPC decoding, [20] Theorem 9, apparently requires similar treatment as well: the BP estimate of a transmitted bit \( X_i \) there is not simply \( E[X_i | Y^{(l)}_{v_i}] \), as stated in that paper, and instead the parity constraints beyond the \( l \)-iteration neighborhood must be ignored when taking that expectation. Introducing a scrambling sequence there and making its bits at those parity constraints have \( \pi \)-priors seems to be an effective way to achieve this, as demonstrated in the above proof regarding \( b_i^{(L)} \).

Given \( i' \) and \( j \) and conditioned on a fixed \( G \), all the \( \lambda^{b_i'} \)’s and \( \lambda^{a_j} \)’s in Table [1] are deterministic given their respective reference bits, \( b_i' \) and \( a_j \), and their densities are either \( \mathcal{E}_0 \) or \( \mathcal{E}_1 \), which are also symmetric. As for \( \lambda^{a_j} \) (and thus \( \lambda^{b_i'} \)), since it is a function of \( y_j \), and \( y_j = u_j^* - (b^*, a^*) \) is a Markov chain by Proposition [3] we see that it depends only on \( u_j^* \) among \((b^*, a^*, u^*)\). It is easy to prove that the density of \( \lambda^{a_j} \) w.r.t. \( u_j^* \) conditioned on \( G \) is symmetric as well:

Proposition 15: Given a binary symmetric source coding problem, generator matrix \( G \) and parameter \( t > 0 \), the \( \lambda^{a_j} \) as defined in [10] has a symmetric density w.r.t. \( u_j^* \) conditioned on \( G \), and this density does not vary with \( G \) or \( j \).

Proof: From Proposition [3] that the conditional pdf \( p(y_j | u_j^*) = p_{y_j | u_j}(y_j | u_j^*) \) is given by the test channel \( p(y | u_j) \), which by Proposition [2] satisfies \( p_{y_j | u_j}(y_j | 0) = p_{y_j | u_j}(y_j | 0) \), so Proposition [4] can be applied to prove that the likelihood function of \( y = y_j \) has a symmetric density w.r.t. \( u = u_j^* \). Now \( \lambda^{a_j}(u) = e^{-t d(u, y_j)} \), so by the definition of the test channel, when viewed as a function of \( u \) it is proportional to \( p_{y_j | u_j}(y_j | u_j) \) and thus \( p_{y_j | u_j}(y_j | u_j) \), i.e. \( \lambda^{a_j} \) is exactly the said likelihood function of \( y_j \) on the test channel. Therefore, \( \lambda^{a_j} \) has a symmetric density w.r.t. \( u_j^* \), and this density is determined by the test channel only, so it does not vary with \( G \) and \( j \).

Combining the results of Proposition [4] and Proposition [15] we can immediately apply Proposition [5] and Proposition [8] to see that, when conditioned on a fixed \( G \in \mathcal{G}_n^{(L)} \) (which is also in \( \mathcal{G}_n^{(l)} \) for any \( l \leq L \)) and using \( b_i' \) as the reference bit (which is 0 or 1 with equal probability independent from \( G \) by Proposition [5]), we have

\[
\nu^{b_i^*}_{\mathcal{L}} \leq \cdots \leq \nu^{b_i^*}_{\mathcal{L}_1(L)} \leq \nu^{b_i^*}_{\mathcal{L}_1(L)} \leq \cdots \leq \nu^{b_i^*}_{\mathcal{L}_1(L)},
\]

and all these probability tuples have symmetric densities. By Proposition [10] the mean-square differences among these probability tuples can be upper-bounded with the MI differences of their densities.

Averaging over all \( G \) with loop-free \( \mathcal{N}^{(L)}_i \), we obtain the densities defined over \( \mathcal{G}_n^{(L)} \) for iteration counts \( l = 1, 2, \ldots, L, \)

\[
\nu^{b_i^*}_{\mathcal{L}}(G) = \mathbb{E}_{Y^{(L)}}[\nu^{b_i^*}_{Y^{(L)}}], \quad \nu^{b_i^*}_{\mathcal{L}(l)}(G) = \mathbb{E}_{Y^{(L)}}[\nu^{b_i^*}_{Y^{(L)}}],
\]

and all these probability tuples have symmetric densities. By Proposition [10] the mean-square differences among these probability tuples can be upper-bounded with the MI differences of their densities.

These densities, being convex combinations of the densities conditioned on individual \( G \)’s (note that we need \( b_i^* \) and \( G \) to be independent when taking the convex combination, which is true since \( p(b_i^* | G) = 1/2 \) for any \( b_i^* \) and \( G \)) due to Proposition [3], clearly remain symmetric. Using Proposition [7] the physical degradation relationships are also preserved, thus

\[
\nu^{b_{i'}^*}_{\mathcal{L}_1(L)}(G) \leq \cdots \leq \nu^{b_{i'}^*}_{\mathcal{L}_1(L)} \leq \nu^{b_{i'}^*}_{\mathcal{L}_1(L)} \leq \cdots \leq \nu^{b_{i'}^*}_{\mathcal{L}_1(L)},
\]

and the bound from Proposition [10] can likewise be averaged to yield

\[
\frac{2}{\ln 2} \mathbb{E} \left[ (\nu^{b_{i'}^*}_{c(1)}(0) - \nu^{b_{i'}^*}_{c(1)}(0))^2 \right] \leq I(\nu^{b_{i'}^*}_{c(1)}(L)) - I(\nu^{b_{i'}^*}_{c(1)}(L)) \leq I(\nu^{b_{i'}^*}_{c(1)}(L)) - I(\nu^{b_{i'}^*}_{c(1)}(L))
\]

for any \( l \leq L \), which bounds the amount of synchronization error at the \( i \)-th b-step. The hard-to-compute \( \nu^{b_{i'}^*}_{c(1)} \) has been eliminated from this bound, leaving only \( \nu^{b_{i'}^*}_{c(1)} \) and \( \nu^{b_{i'}^*}_{c(1)} \), which in the \( n \rightarrow \infty \) limit can be obtained via DE.

C. Synchronization at a-steps

Now we analyze the synchronization condition at the \( j \)-th a-step of the TPQ, namely whether \( \nu^{a_j} \) is close to \( \pi \). To make analysis feasible, analogous to the \( \nu^{a_j} \) above, we define an “upper bound” of \( \nu^{a_j} \) (denoted by \( \nu^{a_j} \)) by hypothetically running BP for \( L \) iterations starting with all \( \mu^{b_{i'}}_{\mathcal{L}_1} = \overline{a_i} \), as shown in Fig. 5.

Again, the computation of \( \nu^{a_j} \) only involves a neighborhood of variable node \( a_j \) in the factor graph, as shown in Fig. 4(c) and denoted by \( N_j \), and it can be further divided into the interior part \( N_j^\circ \) and the border part \( N_j^b \), with each repetition unit in Fig. 4(a) corresponding to one iteration. Given the degree distribution, \( L \) and \( j \), the set of \( G \in \mathcal{G}_n^{(L)} \) with a loop-free \( \mathcal{N}_j^{(L)} \) is denoted by \( \mathcal{G}_n^{(L)} \), the probability that a uniformly distributed \( G \) over \( \mathcal{G}_n^{(L)} \) lies outside \( \mathcal{G}_n^{(L)} \) is again independent of \( j \), and can be denoted by \( P_{\text{loop}} \) that satisfies \( 1 - P_{\text{loop}} = 0 \), etc.

For any priors \( \lambda_{b_{i'}}^a, \lambda_{a_j}^a, \lambda_{b_{i'}}^a, \lambda_{a_j}^a \), \( \nu \equiv \nu(G; \lambda_{b_{i'}}^a, \lambda_{a_j}^a, \lambda_{b_{i'}}^a) \) is now the true extrinsic information corresponding to these priors at variable node \( a_j \) in the factor graph in Fig. 2(a). In particular, \( \nu^{a_j} \) as defined in [3] is equal to \( \nu(G; \lambda_{b_{i'}}^a, \lambda_{a_j}^a, \lambda_{b_{i'}}^a) \), where \( \lambda_{b_{i'}}^a = \overline{a_i} \) for \( i < j \) (i.e. at the positions decimated in previous a-steps) and is \( \overline{a_j} \) for \( j' > j \), and \( \lambda_{b_{i'}}^a(u_j^*) = e^{-t d(u_j, y_j)} \) for all \( j' \). When \( \lambda_{b_{i'}}^a \) is loop-free,
similar to Proposition 14, we can prove that $\text{pr}_j(L)$ can be expressed in this form as well:

**Proposition 16:** If $N_j^L$ is loop-free, then $\text{pr}_j(L) = \nu(C; \lambda_b^*, \lambda_{-j}^*, \lambda_u)$, with the priors $\lambda_b^*$, $\lambda_{-j}^*$, and $\lambda_u^*$ given by Table II.

**Proof:** Let $\nu = \nu(C; \lambda_b^*, \lambda_{-j}^*, \lambda_u)$, where the priors are those for $\text{pr}_j(L)$ in the table. $\nu$ is then the extrinsic information at $a_j$ in the factor graph corresponding to these priors. Similar to the treatment of $\text{pr}_i(L)$ in the proof of Proposition 14, since the variable nodes $b_i$ in $N_j^+$ have prior $\lambda_b = \overline{\nu}$, they can be disconnected from the check nodes in $N_j^+$ and have $b_i$ substituted into the corresponding factors. After such a transformation, the $N_j^-$ part of the factor graph becomes disconnected from the rest, and the true extrinsic information at $a_j$ on this tree-like part of the factor graph, which is still equal to $\nu$, can now be exactly computed with BP using the algorithm in Fig. 5.

Combining Proposition 16 and Proposition 15 with Propositions 5 and 8, we again find that, conditioned on a fixed $G \in G_n^{(L)}$ (which is thus also in $G_n^{(l)}$ for any $l \leq L$) and using $a_j^*$ as the reference bit, we have the physical degradation relationships

\[ \bar{\nu} \leq \nu_j^{a_j^*} \leq \text{pr}_j(L) \leq \cdots \leq \text{pr}_1(L), \tag{29} \]

with all these probability tuples having symmetric densities, so Proposition 10 can still be applied to bound the mean-square difference between $\nu_j^{a_j^*}(0)$ and $1/2$. Now we define for $l = 1, 2, \ldots, L$ the average densities over $G_n^{(L)}$, namely

\[ \nu_j^{a_j^*} | a_j^* \in G_n^{(L)} \sim \nu_j^{a_j^*}, \]

\[ \text{pr}_j(L) | a_j^* \in G_n^{(L)} \sim \text{pr}_j(L); \tag{30} \]

then they remain symmetric and satisfy

\[ \bar{\nu} \leq \nu_j^{a_j^*} \leq \text{pr}_j(L; L) \leq \cdots \leq \text{pr}_j(1; L), \tag{31} \]

and the bound from Proposition 10 can also be averaged to yield, for any $l \leq L$,

\[ \frac{2}{ln2} E \left[ (\nu_j^{a_j^*}(0) - 1/2)^2 \, | \, G \in G_n^{(L)} \right] \leq I(\nu_j^{a_j^*}(L)) - I(\nu_j^{a_j^*}(0)) \tag{32} \]

\[ \leq I(\text{pr}_j(L; L)). \tag{33} \]

Eq. (33) now bounds the amount of synchronization error at the $j$-th $a$-step in terms of $I(\text{pr}_j(L; L))$, a quantity computable with DE in the $n \to \infty$ limit.

D. The Asymptotic Synchronization Conditions in terms of DE Results

We now introduce some notations for DE results. We use $\mu_{\pm d} \triangleq \mu \oplus \cdots \oplus \mu$ to denote the result of the $\oplus$ operation on $d$ independent message densities (with $\mu_{\pm (0)} \triangleq E_0, \mu_{\pm (d)}$ for the $\ominus$ operation with $\mu_{\pm (0)} \triangleq E_0$, and $\sum$ to denote the convex combination operation in Section IV-A. The density $\lambda^u$ is that of each $\lambda^u_j$ w.r.t. $u^*_j$, which does not vary with $G$ or $j$ due to Proposition 15 and its MI is $I(\lambda^u) = I(u; y) = H_0(t) > 0$. We also let $\lambda^b \triangleq E_b$, where $I_b \in [0, 1]$ can be understood as the fraction of bits in $b$ decimated in previous $b$-steps. Now, corresponding to the $L$ BP iterations that yield $\nu(L)$, we can let $\mu_{b(L)} \triangleq E_0$ and define iteratively

\[ \mu_{b(L)} \triangleq \lambda^u \oplus \left( \sum_d v_d \cdot (\mu_{b(L-1)})^{\oplus(d-1)} \right), \tag{34} \]

\[ \mu_{b(L)} \triangleq \lambda^b \ominus (\mu_{b(L)})^{\ominus(d-1)}, \tag{35} \]

which finally yields

\[ \nu(L) = (\mu_{b(L)})^{\ominus(d)}. \tag{36} \]

If the above process instead starts from $\mu_{b(0)} \triangleq E_1$, the result is then denoted by $\nu_{b(L)}$. Since $\text{pr} = \overline{\nu}$ used in BP has the “always sure” density $E_1$, the $\oplus$ operation with it has no effect and has been omitted from (34).

Similarly, during the $a$-steps, if we let $I_a \in [0, 1]$ be the fraction of bits in $a$ decimated in the previous steps and let $\lambda^a \triangleq E_{a},$ then the density $\nu_{a(l)}(L)$ corresponding to the process in Fig. 5 can be defined as follows:

\[ \mu_{b(L)} \triangleq E_1, \tag{37} \]

\[ \mu_{b(L)} = (\lambda^u \oplus \lambda^a) \ominus \left( \sum_d v_d \cdot (\mu_{b(L-1)})^{\ominus(d-1)} \right), \tag{38} \]

\[ \mu_{b(L)} \triangleq (\mu_{b(L)})^{\ominus(d-1)}, \tag{39} \]

\[ \text{pr}_{a(L)} \triangleq \lambda^u \ominus \left( \sum_d w_d \cdot (\mu_{b(L)})^{\ominus(d)} \right). \tag{40} \]

Now compare the DE result $\nu_{b(L)}$ defined above to the $\nu_{b(L)}$ defined in Section II-B for a given $l \leq L$. As $n \to \infty$, we make $i$ a function of $n$ that causes the fraction of decimated bits in $b_{-i}$, $(i-1)/(n-1)$, to converge to some $I_b$. $\nu_{b(L)}$ is an average over $G \in G_n^{(L)}$, and over this ensemble, the degrees of different nodes in $N_i^{(L)}$, as well as their degeneratedness (i.e. whether the node’s index $i$ is above or below $i$), are asymptotically independent as $n \to \infty$ and the probability that a given $b_j$ has been decimated is $(i-1)/(n-1)$, which approaches $I_b$ as well. Comparing the definitions of $\nu_{b(L)}$ and its density $\nu_{b(L)}$ to the above DE result $\nu_{b(L)}$ and noting that each DE step in (35), (36) and (37) is obviously continuous with respect to convergence.

6 Technically, they are not exactly independent because the total number of nodes of some degree $d$ and the number of decimated nodes in the entire factor graph are fixed, so one node in the neighborhood having a certain degree makes another node less likely to have the same degree, but this has negligible impact when $n$ is large enough that $N_i^{(L)}$ is only a vanishing fraction of it, and can be dealt with using conventional bounding techniques.
in distribution, we can conclude that \( \mathcal{L}^{b}_{i(L)} \) converges in distribution to \( \mathcal{L}^{b}_{i} \) as \( n \to \infty \). Similarly, \( \mathcal{T}^{b}_{j(L)} \) converges in distribution to \( \mathcal{T}^{b}_{j} \), and if \( j \) is made to vary with \( n \) such that \( \lim_{n \to \infty} (j - 1)/(n_{c} - 1) = I_{s} \in [0, 1] \), then \( \mathcal{T}^{b}_{j(L)} \) also converges in distribution to \( \mathcal{T}^{b}_{j} \) as \( n \to \infty \).

The above discussion involves the densities of the BP extrinsic information \( \nu_{i}^{b} \) and \( \nu_{j}^{b} \) and the corresponding DE results. For the true extrinsic information \( \nu_{i}^{b} \), we have defined in Section IV-B its density over \( G \in G_{n} \) as \( \nu_{i}^{b} \). The density of \( \nu^{b} \) over all \( G \in G_{n}(d_{b}, w) \), including those with loopy neighborhoods, will be denoted by \( \nu^{a} \) (if \( b = b^{a} \)) and its symmetry can still be established with Proposition 5. Similarly, \( \nu^{a} \) is defined as the density of \( \nu^{a} \) over all \( G \in G_{n}(d_{b}, w) \).

The aforementioned densities can all be characterized by their MIs. For the DE results, we define \( I_{b, ext}^{(h,n,l,L)} \) and \( T_{b, ext}^{(h,n,l,L)} \) respectively as the MI and TMI between any given node and the entire codebook, i.e.,

\[
I_{b, ext}^{(h,n,l,L)}(a_{b}, b) = \text{MI between } a_{b} \text{ and } b \text{ over all } G \in G_{n}(d_{b}, w), \quad T_{b, ext}^{(h,n,l,L)}(a_{b}, b) = \text{TMI between } a_{b} \text{ and } b \text{ over all } G \in G_{n}(d_{b}, w).
\]

For the densities of the true extrinsic information, namely \( \nu_{i}^{b} \), \( \nu_{j}^{b} \), and \( \nu_{a}^{b} \), their MIs are likewise denoted by \( I_{b, ext}^{(h,n,l,L)} \), \( I_{a, ext}^{(h,n,l,L)} \), and \( T_{b, ext}^{(h,n,l,L)} \) respectively, where \( I_{b} = (i - 1)/(n_{b} - 1) \), \( I_{s} = (j - 1)/(n_{c} - 1) \), and linear interpolation is performed to extend their definitions to all \( I_{b} \) and \( I_{s} \). Since the MI of a probability tuple is a bounded and continuous function, the above convergence in distribution results immediately lead to the convergence of MI due to the portmanteau theorem; specifically, for any \( I_{b}, I_{s} \in [0, 1] \) and \( l \leq L \), we have

\[
\lim_{n \to \infty} I_{b, ext}^{(h,n,l,L)}(a_{b}, b) = I_{b, ext}(a_{b}, b), \quad \lim_{n \to \infty} T_{b, ext}^{(h,n,l,L)}(a_{b}, b) = T_{b, ext}(a_{b}, b).
\]

Note that the limits depend only on \( l \) but not \( n \), as long as \( L \geq l \).

For the densities of the true extrinsic information, namely \( \nu^{b} \), \( \nu^{b} \), and \( \nu^{a} \), their MIs are likewise denoted by \( I_{b, ext}^{(h,n,l,L)} \), \( I_{a, ext}^{(h,n,l,L)} \), and \( T_{b, ext}^{(h,n,l,L)} \) respectively, where \( I_{b} = (i - 1)/(n_{b} - 1) \), \( I_{s} = (j - 1)/(n_{c} - 1) \), and linear interpolation is performed to extend their definitions to all \( I_{b} \) and \( I_{s} \). Since the MI of a probability tuple is a bounded and continuous function, the above convergence in distribution results immediately lead to the convergence of MI due to the portmanteau theorem; specifically, for any \( I_{b}, I_{s} \in [0, 1] \) and \( l \leq L \), we have

\[
\lim_{n \to \infty} I_{b, ext}^{(h,n,l,L)}(a_{b}, b) = I_{b, ext}(a_{b}, b), \quad \lim_{n \to \infty} T_{b, ext}^{(h,n,l,L)}(a_{b}, b) = T_{b, ext}(a_{b}, b).
\]

Fig. 6. The relationship among the MIs involved in b-steps

are continuous functions of each \( I_{b} \) and the degree distribution \( w \). However, their \( L \to \infty \) limits \( I_{b, ext}(a_{b}, b) \) and \( T_{b, ext}(a_{b}, b) \) defined below are not necessarily so, and neither are the \( n \to \infty \) MIs of the true extrinsic information, \( I_{b, ext}^{(h,n,l,L)}(a_{b}, b) \) and \( T_{b, ext}^{(h,n,l,L)}(a_{b}, b) \). On the other hand, the finite-\( n \) MIs such as \( I_{b, ext}(a_{b}, b) \) are trivially continuous w.r.t. \( I_{b} \) due to them being linear interpolations.

The relationships among the above MIs are given by the following result. For the MIs involved in b-steps, these relationships can be visualized by Fig. 6 and the relationships among the MIs in a-steps are similar.

**Proposition 17**: The MIs above satisfy the following results:

1. Given \( I_{b} \in [0, 1] \) and \( n > 0 \), then as long as \( l \leq L \), \( I_{b, ext}(a_{b}, b) \) is increasing (not necessarily strictly so; same below) and \( T_{b, ext}(a_{b}, b) \) is decreasing with \( l \), with

\[
I_{b, ext}(a_{b}, b) \leq I_{b, ext}^{(h,n,l,L)}(a_{b}, b) \leq \lim_{n \to \infty} I_{b, ext}^{(h,n,l,L)}(a_{b}, b) \leq T_{b, ext}(a_{b}, b) \leq \lim_{n \to \infty} T_{b, ext}^{(h,n,l,L)}(a_{b}, b).
\]

Consequently, the \( n \to \infty \) limits \( I_{b, ext}(a_{b}, b) \) and \( T_{b, ext}(a_{b}, b) \) are likewise respectively increasing and decreasing functions of \( l \), whose \( l \to \infty \) limits \( I_{b, ext}(a_{b}, b) \) and \( T_{b, ext}(a_{b}, b) \) exist and satisfy

\[
I_{b, ext}(a_{b}, b) \leq I_{b, ext}(a_{b}, b) \leq \lim_{n \to \infty} I_{b, ext}^{(h,n,l,L)}(a_{b}, b) \leq \lim_{n \to \infty} T_{b, ext}^{(h,n,l,L)}(a_{b}, b) \leq T_{b, ext}(a_{b}, b) \leq \lim_{n \to \infty} T_{b, ext}^{(h,n,l,L)}(a_{b}, b), \quad \forall l.
\]

2. Given \( I_{s} \in [0, 1] \) and \( n > 0 \), \( L > 0 \), then as long as \( l \leq L \), \( I_{a, ext}(a_{b}, b) \) is a decreasing function of \( l \) and satisfies

\[
I_{a, ext}^{(h,n,l,L)}(a_{b}, b) \leq I_{a, ext}(a_{b}, b) \leq \lim_{n \to \infty} I_{a, ext}^{(h,n,l,L)}(a_{b}, b) \leq \lim_{n \to \infty} T_{a, ext}^{(h,n,l,L)}(a_{b}, b), \quad \forall l.
\]

3. For any \( n \) and \( l \leq L \), \( I_{b, ext}(a_{b}, b) \), \( I_{b, ext}(a_{b}, b) \), \( I_{b, ext}(a_{b}, b) \) and \( T_{b, ext}(a_{b}, b) \) are increasing functions of \( I_{b} \); consequently, so are \( I_{b, ext}(a_{b}, b) \), \( I_{b, ext}(a_{b}, b) \), \( I_{b, ext}(a_{b}, b) \), as well as \( I_{b, ext}(a_{b}, b) \) and \( T_{b, ext}(a_{b}, b) \).

4. For any \( n \) and \( l \leq L \), \( I_{a, ext}(a_{b}, b) \), \( I_{a, ext}(a_{b}, b) \), and \( I_{a, ext}(a_{b}, b) \) are increasing functions of \( I_{s} \), and consequently \( I_{a, ext}(a_{b}, b) \), \( I_{a, ext}(a_{b}, b) \), and \( I_{a, ext}(a_{b}, b) \) are likewise.

**Proof**: See Appendix B.

By Proposition 10 and Proposition 11 the synchronization conditions should hold in an asymptotic sense if and only if
\( \bar{T}^{(I_a)}_{a,\text{ext}} = 0 \) for all \( I_a \in [0, 1] \) (actually the \( I_a = 1 \) case is sufficient due to monotonicity) and \( \bar{T}^{(I_b)}_{b,\text{ext}} = \bar{T}_{b,\text{ext}} \) for all \( I_b \in [0, 1] \). This is expressed formally with the following proposition:

**Proposition 18:** Given a degree distribution, if
\[
\bar{T}^{(I_a)}_{b,\text{ext}} = \bar{T}^{(I_b)}_{b,\text{ext}}, \quad \forall I_b \in [0, 1],
\]
(45)
\[
\bar{T}^{(I_a)}_{a,\text{ext}} = 0, \quad \forall I_a \in [0, 1],
\]
(46)
then

1. For any sequence of \( i = i(n) \in \{1, \ldots, n_b\} \) indexed by \( n \), as long as \( I^{(n)}_b \triangleq (i - 1)/(n_b - 1) \) has an \( n \to \infty \) limit \( I^c_b \) at which \( \bar{T}^{(I_b)}_{b,\text{ext}} \) is continuous w.r.t. \( I_b \), then
\[
\lim_{n \to \infty} \limsup_{l \to \infty} \mathbb{E} \left[ (\bar{\nu}^{I_b}_{(i)}(0) - \nu^{I_b}_{c}(0))^2 \right] = 0;
\]
(47)
2. For any sequence of \( j = j(n) \in \{1, \ldots, n_c\} \) indexed by \( n \), as long as \( I^{(n)}_c \triangleq (j - 1)/(n_c - 1) \) has an \( n \to \infty \) limit \( I^c_c \), then
\[
\lim_{n \to \infty} \mathbb{E} \left[ (I^{I_c}_{j}(0) - \frac{1}{2})^2 \right] = 0.
\]
(48)
When these two results hold, we say the synchronization conditions are asymptotically satisfied. Conversely,

1. If (45) fails to hold, then there exists \( \epsilon > 0 \) such that, for any \( l \) and \( n_0 \), there always exist \( n \geq n_0 \) and \( i \in \{1, \ldots, n_b\} \) (where \( n_b = nR^{(n)} \) as explained at the beginning of this section) that satisfy
\[
\mathbb{E} \left[ (\bar{\nu}^{I_b}_{(i)}(0) - \nu^{I_b}_{c}(0))^2 \right] \geq \epsilon;
\]
(49)
2. If (46) fails to hold, then there exists \( \epsilon > 0 \) such that, for any \( n_0 \), there always exist \( n \geq n_0 \) and \( j \in \{1, \ldots, n_c\} \) \((n_c = n)\) that satisfy
\[
\mathbb{E} \left[ (I^{I_c}_{j}(0) - \frac{1}{2})^2 \right] \geq \epsilon.
\]
(50)
In either case, we say the synchronization conditions are asymptotically unsatisfied.

**Proof:** See Appendix C.\( \square \) As the convergence of \( \bar{T}^{(I_b,l)}_{b,\text{ext}} \) to \( \bar{T}^{(I_b)}_{b,\text{ext}} \) as \( l \to \infty \) may not be uniform w.r.t. \( I_b \), it seems necessary to introduce the continuity condition at \( I_b^{c} \) in the direct part; however, since \( \bar{T}^{(I_b)}_{b,\text{ext}} \) is a monotonc function of \( I_b \), it has at most countably many discontinuities, and its continuity can be checked numerically anyway. The a-step result (43) does not require such a continuity condition because the counterpart of \( \bar{T}^{(I_b)}_{b,\text{ext}} \) is constant zero, which is always continuous. \( \square \)

Similar to (19), we may plot \( \bar{T}^{(I_b)}_{b,\text{ext}} \) against \( I_b \) and call the resulting curves the lower and upper BP EXIT curves, which can be obtained with DE methods. On the other hand, the curves of \( \bar{T}^{(I_b)}_{b,\text{ext}} \) versus \( I_b \) can be called the MAP (maximum a posteriori) EXIT curves, which are difficult to obtain directly, but by (43), they always lie between the BP EXIT curves, and an example will be given in Fig. 8 below.

We will now present a sufficient condition for the synchronization conditions to be asymptotically satisfied, in terms of the BP curves only. For this purpose we need the following lemma:

**Lemma 19:** Let \( \bar{T}^{(0)}_{b,\text{ext}} \) be the value of \( \bar{T}^{(I_b)}_{b,\text{ext}} \) at \( I_b = 0 \), and \( \bar{T}^{(0)}_{a,\text{ext}} \) be the value of \( \bar{T}^{(I_a)}_{a,\text{ext}} \) at \( I_a = 1 \), then for any degree distribution, \( \bar{T}^{(0)}_{b,\text{ext}} \neq 0 \) if and only if \( \bar{T}^{(0)}_{a,\text{ext}} \neq 0 \).

**Proof:** Comparing (44) and (55) with (38) and (39), we note that \( \bar{T}^{(1,1)}_{(L)} \) and \( \bar{T}^{(0,0)}_{(L)} \) have the same \( \mu_{cb}^{(i)} \)'s and \( \mu_{bc}^{(i)} \)'s in their iterative definitions, and they can respectively be expressed as
\[
\bar{T}^{(1,1)}_{(L)} = \lambda^u \oplus \left( \sum_d w_d \cdot (\mu_{cb}^{(L)})^{(d)} \circ (d) \right),
\]
(51)
\[
\bar{T}^{(0,0)}_{(L)} = (\mu_{bc}^{(L)})^{(d)}.
\]
(52)
Since we have assumed that \( I(\lambda^u) = I_u = R_0(t) \) is strictly positive, \( d_b \geq 2 \) and all degrees are non-zero, we can use the results in (29) regarding the MI combining behavior of the \( \oplus \) and \( \circ \) operators to show that \( I(\bar{T}^{(1,1)}_{(L)}) \) and \( I(\bar{T}^{(0,0)}_{(L)}) \) go to zero as \( L \to \infty \) if and only if \( I(\mu_{cb}^{(L)}) \) does. Consequently, \( \bar{T}^{(1)}_{a,\text{ext}} = 0 \) if and only if \( \bar{T}^{(0)}_{b,\text{ext}} = 0 \).

Using Lemma 19 and the monotonicity of \( \bar{T}^{(I_a)}_{a,\text{ext}} \) w.r.t. \( I_a \) in Proposition 17 we can immediately obtain the following sufficient condition from Proposition 18:

**Theorem 20:** Given a degree distribution, if
\[
\bar{T}^{(0)}_{b,\text{ext}} = 0,
\]
(53)
\[
\bar{T}^{(I_a)}_{b,\text{ext}} = \bar{T}^{(I_b)}_{b,\text{ext}}, \quad \forall I_b \in [0, 1],
\]
(54)
then the synchronization conditions are asymptotically satisfied.

Although the MAP curves themselves are difficult to compute, they are known to satisfy the following area theorem:

**Proposition 21:** For any degree distribution and \( n \), we have
\[
\sum_{i=1}^{n} I^{(I_a,i)}_{a,\text{ext}} + \sum_{i=1}^{n} I^{(I_b,i)}_{b,\text{ext}} = nI_u,
\]
(55)
where \( I_{a,i} \triangleq (i - 1)/(n_c - 1), I_{b,i} \triangleq (i - 1)/(n_b - 1) \). Note that (55) uses the average MI \( I^{(I_a,i)}_{a,\text{ext}} \) over all \( G \), including those with loopy neighborhoods.

Consequently, as \( n \to \infty \), any degree distribution satisfies the area theorem
\[
\int_0^1 \bar{T}^{(I_b)}_{b,\text{ext}}(I_a) dI_a + R \int_0^1 \bar{T}^{(I_a)}_{a,\text{ext}}(I_b) dI_b \leq I_u
\]
\[
\leq \int_0^1 \bar{T}^{(I_a)}_{a,\text{ext}}(I_b) dI_a + R \int_0^1 \bar{T}^{(I_b)}_{b,\text{ext}}(I_b) dI_b.
\]
(56)

**Proof:** See Appendix C.\( \square \) This can be regarded as a special case of (19) Theorem 1, where the reference codeword \((b^*, a^*)\) corresponds to \( X \) there, the \( \lambda_{cb}^{(i)} \)'s and \( \lambda_{bc}^{(i)} \)'s are the \( Y \), and the \( \lambda_{cb}^{(i)} \)'s (or \( y \)) are the additional observation \( \Omega \).

This immediately leads to the following necessary condition for the synchronization conditions to be satisfied:

**Theorem 22:** For any degree distribution,
\[
\int_0^1 \bar{T}^{(I_b)}_{b,\text{ext}} dI_b \leq I_u/R.
\]
(57)
Moreover, equality holds in (57) when the synchronization conditions are asymptotically satisfied.

Proof: Application of (43) and (44) in the first inequality of (56) gives

$$R \int_0^1 I_b^{(l)} \, dI_b \leq \int_0^1 I_a^{(l)} \, dI_a + R \int_0^1 I_b^{(l)} \, dI_b \leq I_a,$$

which leads to (57).

When the synchronization conditions are asymptotically satisfied, i.e. (45) and (46) hold, the second inequality of (56) becomes

$$I_a \leq \int_0^1 T_a^{(l)} \, dI_a + R \int_0^1 T_b^{(l)} \, dI_b = R \int_0^1 I_b^{(l)} \, dI_b,$$

so equality holds in (57). ■

E. The Case of Binary Erasure Quantization

As an important and intuitive special case, we consider the BEQ problem as defined in Example 3 at $t \to \infty$. Given $G$ and a source sequence $y$, we say a certain $(b, a)$ or the corresponding $u = u(b, a)$ is consistent with it if $d(y, u) = 0$ (i.e. $y_j = *$ or $y_j = u_j$ for all $j$), and the set of such $(b, a)$'s, which is non-empty due to the freedom in the choice of $a$, is denoted $C_y$. For a BEQ problem with a definite $a$, it is said to have a solution if there is some $(b, a) \in C_y$. According to the discussion in Section III-A, the reference codeword $(b^*, a^*)$ yielded by the TPQ is uniformly distributed over $C_y$, and the joint distribution of $b^*, a^*, u^*$ and $y$ is given by Proposition 3 with

$$p(u_j^* \mid y_j) = p_a \mid y(u_j^* \mid y_j) = \begin{cases} 1/2, & y_j = *; \\ 1 \left[ y_j = u_j^* \right], & y_j = 0, 1; \end{cases}$$

$$p(y_j) = p_b(y_j) = \begin{cases} \epsilon, & y_j = *; \\ (1 - \epsilon)/2, & y_j = 0, 1; \end{cases}$$

for all $j$. We thus have

$$p(y_j \mid u_j^*) = \begin{cases} \epsilon, & y_j = *; \\ (1 - \epsilon) \cdot 1 \left[ y_j = u_j^* \right], & y_j = 0, 1. \end{cases}$$

Each $\lambda^u_j$ is a function of $y_j$; according to (10), it is $\mathbb{P}$ when $y_j = *$ and $\overline{y_j}$ when $y_j$ is 0 or 1. Combined with (62), we have

$$p(\lambda^u_j \mid u_j^*) = \begin{cases} \epsilon, & \lambda^u_j = \overline{u}_j; \\ (1 - \epsilon), & \lambda^u_j = u_j^*. \end{cases}$$

In other words, $\lambda^u_j \mid u_j^* \sim \mathcal{X}$ is simply $\mathbb{E}_1 - \epsilon$, a symmetric density independent of $j$, and $I_b = I(\mathcal{X}) = 1 - \epsilon$. These properties are consistent with the above discussion such as Proposition 3.

By Proposition 12 and Proposition 13 all the densities involved in the DE steps in Section IV-D as well as those of the true extrinsic information, $\nu_{bc}^x$ and $\nu_{bc}^x$, are erasure-like, and can thus be uniquely determined by their MIs, so the conditions (53) and (54) can be evaluated as follows. Let $I_{bc}^{(l)} \triangleq I(\mu_{bc}^{(l)})$ and $I_{bc}^{(l)} \triangleq I(\mu_{bc}^{(l)})$, then (54) and (55) can respectively be expressed as

$$I_{bc}^{(l)} = I_u \sum_d v_d \cdot (I_{bc}^{(l-1)})^{d-1},$$

$$I_{bc}^{(l)} = 1 - (1 - I_b)(1 - I_{bc}^{(l)})^{d_b-1},$$

while (56) becomes

$$I_{bc}^{(l, L)} = 1 - (1 - I_{bc}^{(L)})^{d_b},$$

where the resulting $I_{bc}^{(l, L)}$ is $I_{bc}^{(l, L)} = I(y_b^{(l)} - y_b^{(L)})$ when starting with $I_{bc}^{(0)} = 0$, and $I_{bc}^{(l, L)} = I(y_b^{(l)} - y_b^{(L)})$ when $I_{bc}^{(0)} = 1$.

For conciseness of presentation, we introduce functions $f(\cdot)$, $g(\cdot)$ and $h(\cdot)$ which, in the case of BEQ, are defined as

$$f(x) \triangleq \sum_d v_d x^{d-1},$$

$$g(y) \triangleq \alpha y_b^{-1},$$

$$h(y) \triangleq \beta y_b,$$

so that we can write

$$I_{bc}^{(l)} = I_u \cdot f(I_{bc}^{(l-1)}),$$

$$I_{bc}^{(l)} = 1 - (1 - f(I_{bc}^{(l)})) \cdot g(1 - I_{bc}^{(l)}),$$

$$I_{bc}^{(l)} = 1 - h(1 - I_{bc}^{(l)}).$$

When all the $I_{bc}^{(l)}$'s are equal to the same $I_b$, the resulting $I_{bc}^{(L)}$ is the $I_{bc}^{(l, L)}$ above (53) and (54) depending on the initial $I_{bc}^{(0)}$.

Now we combine (70) and (71) to yield a mapping $I_{bc}^+$ such that $I_{bc}^{(l)} = I_{bc}^+(I_{bc}^{(l-1)}; I_u, I_b)$. $I_{bc}^+(\cdot; \cdot, \cdot)$ is an increasing function of all three variables in $[0, 1]$ and its result is also in $[0, 1]$; therefore, given fixed $I_u$ and $I_b$ and starting with $I_{bc}^{(0)} = 0$ (resp. 1), iterative application of $I_{bc}^+(\cdot; I_u, I_b)$ gives an increasing (resp. decreasing) sequence $(I_{bc}^{(l)})_{l=0}^\infty$ whose limit as $l \to \infty$ always exists and can be denoted $I_{bc}^{(\infty)}$. Taking the $l \to \infty$ limit of (70) and (71) and using continuity, we can finally express $I_{bc}^{(L)}$ and $T_{bc}^{(L)}$ in (53) and (54) in terms of $I_{bc}^{(\infty)}$ and $T_{bc}^{(\infty)}$, as

$$I_{bc}^{(l)} = I_{bc}^+(I_{bc}^{(\infty)}),$$

$$T_{bc}^{(l)} = I_{bc}^{(\infty)},$$

where

$$I_{bc}^{(\infty)} \triangleq 1 - h(1 - I_u \cdot f(x))$$

is a strictly increasing function of $x$.

Both $I_{bc}^{(\infty)}$ and $T_{bc}^{(\infty)}$ are clearly fixed points of $I_{bc}^+$ at the given $I_u$ and $I_b$; indeed, due to monotonicity of $I_{bc}^+(\cdot)$, to prove that they are the minimum and the maximum fixed points among the possibly multiple ones at such $I_u$ and $I_b$.

In the case of BEQ, it is actually straightforward to obtain all the fixed points by equating $I_{bc}^{(l)}$ and $I_{bc}^{(l-1)}$ in (70) and (71). Denoting $x \triangleq I_{bc}^{(l-1)} - I_{bc}^{(l)}$, we get

$$I_b = 1 - \frac{1 - x}{g(1 - I_u \cdot f(x))}. $$
therefore as we vary \( x \) over \([0, 1]\), if the \( I_b \) given by (75) is also within \([0, 1]\), then \( x \) is a fixed point of \( I_b^{\text{bc}}(\cdot) \) at this \( I_b \), and all fixed points can be obtained in this way (note that the denominator in (75) cannot be zero as long as \( I_b < 1 \)). Each fixed point \( x \) can equivalently be expressed in terms of \( I_{b, \text{ext}} \triangleq I_{b, \text{ext}}(x) \). We can now define the EBP EXIT curve (or simply the EBP curve), original proposed in [19] for LDPC decoding over BEC, as the parametric \( I_b \) vs. \( I_{b, \text{ext}} \) curve given by (75) and (74) for \( x \in [0, 1] \).

While \( I_{b, \text{ext}} \) is a strictly increasing function of \( x \), \( I_b \) is not necessarily so. However, with simple algebra we can still obtain the following properties of the EBP curve:

**Proposition 23:** The EBP curve for any degree distribution under BEQ satisfies

\[
I_b|x=0 \leq 0, \quad I_b|x=1 = 1, \quad I_{b, \text{ext}}|x=0 = 1 - (1 - I_b v_1)^{d_b}, \quad d_b I_b v_1 + \int_0^1 (1 - I_b) \frac{dI_{b, \text{ext}}}{dx} \, dx = \frac{I_u}{R}. \tag{78}
\]

In (76) equality holds if and only if \( v_1 = 0 \).

**Proof:** See Appendix [13].

Eq. (78) can be visualized as an area result in Fig. 7 if we define the total shaded area as the area under the EBP curve

\[
A_{\text{ebp}} \triangleq I_{b, \text{ext}}|x=0 + \int_0^1 (1 - I_b) \frac{dI_{b, \text{ext}}}{dx} \, dx, \tag{79}
\]

then since

\[
I_{b, \text{ext}}|x=0 = 1 - (1 - I_b v_1)^{d_b} \leq d_b I_b v_1, \tag{80}
\]

by (78) we have

\[
A_{\text{ebp}} \leq \frac{I_u}{R}, \tag{81}
\]

where equality holds if and only if \( v_1 = 0 \); actually, from (80) we see that the difference is only \( O(v_1^2) \).

Every crossing the EBP EXIT curve makes with a constant-\( I_b \) vertical line corresponds to a fixed point at this \( I_b \). As stated above, the minimum and maximum fixed points at each \( I_b \) are at \( x = I_b^{(I_b, \infty)} \) and \( x = I_b^{(I_b, \infty)} \), or equivalently at \( I_{b, \text{ext}} = I_b^{(I_b, \infty)} \) and \( I_{b, \text{ext}} = I_b^{(I_b, \infty)} \), respectively, so the lower and upper BP EXIT curves defined in Section [IV-D] are simply the lower and upper envelopes of the EBP EXIT curve. The conditions (54) and (53) can now be expressed in terms of the monotonicity of the EBP EXIT curve, which can be easily computed from the degree distribution; this is formally expressed by the following theorem:

**Theorem 24:** For BEQ, if the EBP EXIT curve given by (75) satisfies the following monotonicity condition\(^7\)

\[
I_b|x=0 = 0, \quad \frac{dI_b}{dx} > 0, \quad x \in [0, 1],
\]

then the synchronization conditions are asymptotically satisfied. Conversely, if \( I_b|x=0 < 0 \), or \( dI_b/dx < 0 \) for any \( x \in [0, 1] \), then the synchronization conditions are asymptotically unsatisfied.

**Proof:** Direct part: Condition (83) implies that \( I_b \) is a strictly increasing function of \( x \), so and thus \( I_{b, \text{ext}} \) are also uniquely defined and strictly increasing functions of \( I_b \), and by (76) they are defined for all \( I_b \in [0, 1] \). Therefore, at each \( I_b \), \( I_b^{(I_b, \infty)} \) has a unique fixed point corresponding to this \( I_{b, \text{ext}} \), so \( \mathcal{I}_{b, \text{ext}} \) and \( \mathcal{I}_{b, \text{ext}} \) will both be equal to this value, thus (54) holds. Condition (82) implies that the fixed point is at \( I_{b, \text{ext}} = 0 \) when \( I_b = 0 \), so (53) holds as well. Theorem 20 can thus be applied to obtain the desired result.

Converse part: Since the lower-BP curve is the lower envelope of the EBP curve, the area under it never exceeds \( A_{\text{ebp}} \) in Fig. 7 and a finite difference will exist if \( dI_b/dx < 0 \) for any \( x \in [0, 1] \) (note that \( I_{b, \text{ext}} \) is strictly increasing with respect to \( x \)). On the other hand, we have found that \( A_{\text{ebp}} \leq I_u/R \) and is strictly smaller when \( v_1 > 0 \) or equivalently \( I_b|x=0 < 0 \). Combining the two results, we can see from Theorem 22 that the synchronization conditions will be asymptotically unsatisfied in either case.

Finally, we give as examples in Fig. 8 the EXIT curves under BEQ of some \((d_b, d_c)\) regular LDGM codes at different values of \( t \), or equivalently, \( I_u = R_0(t) \).

Fig. 8(a) shows the EBP curves of the \((4, 2)\) regular code with rate \( R = 1/2 \). When \( 1/2 > I_u > t^{\text{thr}} \triangleq 1/3 \), part of the EBP curve lies in the \( I_b < 0 \) half-plane, but once \( I_b \) becomes positive, it is monotonically increasing. Therefore, for any \( I_b > 0 \), \( I_b^{(I_b, \infty)} \) has a unique fixed point with the corresponding \( I_{b, \text{ext}} \) equal to \( \mathcal{I}_{b, \text{ext}}^{(I_b)} = \mathcal{I}_{b, \text{ext}}^{(I_b)} = \mathcal{I}_{b, \text{ext}}^{(I_b)} \), thus the synchronization conditions are asymptotically satisfied in b-steps. On the other hand, corresponding to the fixed point with the largest MI at \( I_b = 0 \), we have \( \mathcal{I}_{b, \text{ext}}^{(0)} > 0 \) and consequently \( \mathcal{I}_{a, \text{ext}}^{(0)} > 0 \); in fact, since \( A_{\text{ebp}} = I_u/R \) and the left-hand side of (57) corresponds to a strictly smaller area, the necessary condition (57) is unsatisfied, so the synchronization conditions must fail to hold \( \mathcal{I}_{a, \text{ext}}^{(0)} > 0 \) at some \( I_a \), so a non-vanishing mean-square difference will exist between some \( v_a^2 \) and \( \bar{v}^2 \) as \( n \to \infty \), implying that the corresponding BEQ problems usually have no solutions. When \( I_u < 1/3 \), the synchronization conditions are asymptotically satisfied in both b- and a-steps due to Theorem 24.

\(^7\)Note that this has nothing to do with monotonicity with respect to a class of channels, as discussed in LDPC literature [30].
If the results in [19] remain true, we may conjecture that (19), as both involve a system of linear equations over quite similar to LDPC decoding over BEC considered in Proposition 23, and the analysis of the similar EXIT curves in Proposition 17, the area results in Proposition 21 and postulated MAP curves based on the monotonicity results are asymptotically satisfied. However, unless $I_u$ is further reduced below $\nu^{th} = 7/16 = 0.4375$, the EBP curve is still not monotonic, therefore the BP fixed points are not unique at some values of $I_b$, where $\overline{T}_{b,ext}^{(h)}$ and $\overline{T}_{b,ext}^{(h)}$ lie between them. Indeed, (57) is again unsatisfied because its left-hand side is strictly smaller than $A_{ebp} = I_u/R$, so the synchronization conditions also fail to hold at some $I_b$, i.e. the BP result $v_b^R$ will fail to converge to $v_b^{\star}$ in a mean-square sense as $n \rightarrow \infty$; in other words, the solutions of the BEQ problems can usually not be obtained with BP. Only when $I_u < 0.4375$ will the EBP curve become monotonic, allowing the synchronization conditions to be asymptotically satisfied.

Fig. 8(b) is for the (5, 3) regular code with rate $R = 0.6$. When $I_u$ is reduced below 0.5176, the EBP curve no longer extends into the $I_b \leq 0$ half-plane, so both $\overline{T}_{b,ext}^{(0)}$ and $\overline{T}_{b,ext}^{(1)}$ are zero, and consequently all $\overline{T}_{a,ext}^{(h)}$ are zero as well, implying that the BEQ problems have solutions in an asymptotic sense. However, unless $I_u$ is further reduced below $\nu^{th} = 7/16 = 0.4375$, the EBP curve is still not monotonic, therefore the BP fixed points are not unique at some values of $I_b$, where $\overline{T}_{b,ext}^{(h)}$ and $\overline{T}_{b,ext}^{(h)}$ lie between them. Indeed, (57) is again unsatisfied because its left-hand side is strictly smaller than $A_{ebp} = I_u/R$, so the synchronization conditions also fail to hold at some $I_b$, i.e. the BP result $v_b^R$ will fail to converge to $v_b^{\star}$ in a mean-square sense as $n \rightarrow \infty$; in other words, the solutions of the BEQ problems can usually not be obtained with BP. Only when $I_u < 0.4375$ will the EBP curve become monotonic, allowing the synchronization conditions to be asymptotically satisfied.

Fig. 8(c) is a comparison of the EBP and the lower-BP curves of the (5, 3) regular code at $I_u = 0.5$, as well as postulated MAP curves based on the monotonicity results in Proposition 17, the area results in Proposition 21 and Proposition 23 and the analysis of the similar EXIT curves arising in LDPC decoding over BEC in [19]. BEQ is actually quite similar to LDPC decoding over BEC considered in [19], as both involve a system of linear equations over $\mathbb{Z}_2$. If the results in [19] remain true, we may conjecture that $\overline{L}_{b,ext}^{(h)} = \overline{T}_{b,ext}^{(h)}$ for all $I_b$, and the MAP curve formed by it looks like the dashed line in Fig. 8(c). Note that the area under this MAP curve is $I_u/R$ according to (56), which is also equal to $A_{ebp}$, so the two regions between the EBP and MAP curves necessarily have the same area $A_1 = A_2$. The area $A_2$ to the right of the MAP curve represents the $b_i$'s whose $v_b^R = 0$ but $v_b^R = 0$ and thus violate the synchronization condition; that is, the values of these bits are determined by previous decimation results but not available from BP at the time, and they are apparently “guesses” until they are “confirmed” by an equal number of equations encountered later represented by $A_2$. That $A_2 = A_1$ intuitively means that confirmations constrain earlier guesses rather than $\alpha$, so the BEQ problem does have a solution in an asymptotic sense. This is not the case for e.g. the (4, 2) regular code at $I_u = 0.5$ in Fig. 8(a) there the MAP and the lower-BP curves overlap with the EBP curve in the $I_b \geq 0$ half-plane but does not extend to the left, and the area between the EBP curve and the $I_b = 0$ axis represent “confirmations” that, having no earlier guesses, become constraints on $\alpha$.

F. Application in Degree Distribution Optimization

We may summarize the above analysis as follows:

- The quantization algorithm using PD, being an implementation of BPPQ, can reach the distortion $D_0(t)$ of the TPQ if the synchronization condition is satisfied exactly.
- The synchronization condition is satisfied asymptotically, as the block length $n$ and the iteration count $L$ goes to infinity, if the degree distribution satisfies the conditions in Theorem 20 or (in case of BEQ) Theorem 24 at the chosen $t$ (or $I_u$).

These results suggest that the asymptotic synchronization condition, which can be evaluated numerically with DE for any specific degree distribution, can be used as the constraint for LDGM degree distribution optimization. For ordinary symmetric source coding problems, we want to maximize $t$ such that $D_0(t)$ is minimized, while for BEQ, $t$ is fixed at infinity with $D_0(t) = 0$, and we want to find the source with the minimum $\epsilon$ that can still be encoded at a given $R$. This is thus equivalent to the maximization of $I_u$, which is $R_0(t)$ in the former case and $1 - \epsilon$ in the latter. Alternatively, the optimization problem can also be formulated as the minimization of $R$ at a given $t$ or $I_u$.

The details of this optimization have been tackled in [16]. The method starts from the degree distribution optimized for BEQ using Theorem 24 i.e. the erasure approximation (EA) result, due to the availability of an explicit formula for the EBP curve in this case; numerical DE is then performed on this degree distribution and the results are used to derive a correction factor $r(x)$ for use in the next iteration of the optimization process. As the degree distribution resulting from this iterative process can be numerically verified to satisfy the asymptotic synchronization condition, our analysis above suffices as a theoretical justification for this approach.

It should be noted that asymptotic satisfaction of the synchronization condition does not imply its exact satisfaction, particularly since both the block length $n$ and the iteration count $L$ are necessarily finite in practice. While this residual
synchronization error can be effectively tackled with the recovery algorithm in [16] and [12], this also suggests that making the synchronization condition asymptotically satisfied might not be optimal, as allowing for a small asymptotic synchronization error might lower $D_0(t)$ at the same $R$ by a larger amount than the extra distortion caused by the synchronization error; indeed, an improved optimization method for finite $L$ has been proposed in [16]. However, these improvements can still be regarded as variations of the method based on the asymptotic synchronization condition.

V. EXTENSION TO NON-BINARY CONSTRUCTIONS

We now consider non-binary LDGM-based code constructions that are necessary in many source coding problems. For example, it has been shown in Section II-C that the shaping loss of binary MSE quantization is lower-bounded by 0.0945 dB due to the random-coding loss, and this loss can be greatly reduced if a larger alphabet is used; this issue has also been noted in e.g. [31] in the context of shaping for dirty-paper coding. In general, a symmetric source coding problem over a finite abelian group $G$ ($G = \mathbb{Z}_M$ in $M$-ary MSE quantization) can be solved using LDGM codes in either of the following two ways:

- When $|G| = 2^K$, binary LDGM codes may be used, with every $K$ bits from an LDGM codeword modulated into a reconstructed symbol, similar to bit-interleaved coded modulation (BICM) in channel coding [32, 33];
- Use an $|G|$-ary LDGM code directly, similar to the use of trellis-coded modulation (TCM) [34] and non-binary LDPC codes in channel coding, or TCQ in source coding.

The latter approach has been attempted in e.g. [35], but degree distribution optimization and convergence issues have not been tackled there and will be more difficult than the binary case; a notable issue is that many possible $G$’s, such as $G = \mathbb{Z}_M$ with $M = 2^K > 2$ used in $M$-ary MSE quantization, cannot be given a field structure, so the LDGM code has to be defined on a field, usually $GF(M)$, with a different additive group structure, which is no more natural than the simpler former approach. Therefore, in the previous work [26] as well as this paper we adopt the former BICM-like approach, which allows near-ideal codes to be designed with relative ease; such an approach has also been used in other works such as [36]. Of course, if linearity is a concern, e.g. in some problems involving network coding, it would be necessary to adopt the TCM-like approach, usually with $G$ possessing a field structure (e.g. $G = \mathbb{Z}_p$ with $p$ being a prime number) and with the LDGM code defined on it; such code constructions will not be considered in this paper, but can be analyzed with largely the same method.

A. Probability Tuples over a Finite Abelian Group

For symmetric source coding over a finite abelian group $G$, the proposed non-binary LDGM quantizer will make use of probability distributions over either $G$ or $\mathbb{Z}_p^K$; as $\mathbb{Z}_p^K$ is itself a finite abelian group under component-wise addition and can thus be regarded as a special case, it suffices to consider distributions over $G$, which can be viewed as non-negative-valued functions defined on $G$ and represented by probability tuples over finite abelian group $G$. Similar to the binary case, each component of such a probability tuple $\lambda$ is denoted by $\lambda(u)$ ($u \in G$), whose sum is implicitly normalized to 1, and various definitions can also be extended in a straightforward manner as follows:

- Given $u \in G$, $\pi$ is the sure-$u$ probability tuple with $\pi(u) = 1$ and all other components being zero, while $\pi$ is the “unknown” probability tuple with all components being $1/|G|$;
- The entropy of a probability tuple $\lambda$ over $G$ is $H(\lambda) \triangleq -\sum_{u \in G} \lambda(u) \log \lambda(u)$, while its MI $I(\lambda) \triangleq \log |G| - H(\lambda)$.
- The $\odot$ operation on two probability tuples does pairwise multiplication of the $|G|$ components and then normalizes the result;
- The $\oplus$ operation on two probability tuples are defined according to the addition operator on $G$, also denoted by $\oplus$ (specifically, given two probability tuples $\lambda_1$ and $\lambda_2$ over $G$, $\lambda = \lambda_1 \oplus \lambda_2$ is defined as $\lambda(u) = \sum_{u_1, u_2 \in G} \lambda_1(u_1) \lambda_2(u_2)$, $u \in G$. (84)

The $\ominus$ operator is defined similarly for subtraction over $G$. In particular, $\lambda \ominus \pi$ is simply $\lambda$ with its components permuted, and $\pi_1 \ominus \pi_2 = u_1 \ominus u_2$.
- More generally, let $(Z_i)_{i=1}^m$ be $m$ finite abelian groups, and $Z \triangleq Z_1 \times Z_2 \times \cdots \times Z_m$ (85) be their direct product (thus also an abelian group under element-wise $\oplus$ addition). Now let $C$ be a subset (usually a subgroup or its coset) of $Z$, $i \in \{1, \ldots, m\}$, $\lambda_{z_i}$ be $m-1$ probability tuples with each $\lambda_i$ defined over $Z_i$, we then define $\nu(C; \lambda_{z_i})$ as the probability tuple $\nu$ over $Z$, with $\nu(u_i) = \sum_{u \in C, u'_i = u_i} \prod_{j \neq i} \lambda_j(u'_j)$, (86) where $u' = (u_1, \ldots, u_m)$. $\odot$, $\oplus$ and $\ominus$ then refer to the case with $i = m = 3$, $Z_1 = Z_2 = Z_3 = G$, and $C$ being respectively $\{(u, u, u) | u \in G\}$, $\{(u_1, u_2, u_1 \oplus u_2) | u_1, u_2 \in G\}$ and $\{(u_1, u_2, u_1 \oplus u_2) | u_1, u_2 \in G\}$, which are all subgroups of $Z$.

If $\lambda$ is a random probability tuple over $G$, i.e. each possible value of $\lambda$ is a (deterministic) probability tuple over $G$, we can assign to it a random variable $u$ whose value lies in $G$ as its reference variable, and the conditional distribution of $\lambda$ given $u$ is again called its density. Since the set of possible values of $\lambda$ is the unit ($|G| - 1$)-simplex (which is no longer one-dimensional when $|G| > 2$), the probability distribution of $\lambda$ is a probability measure over this simplex, and can be represented by its pdf w.r.t. the Hausdorff measure with a suitable dimensionality depending on the discreteness of the distribution ($|G| - 1$ in the fully continuous case and zero in the fully discrete case). When we write e.g. $p(\lambda)$, it will refer to such a pdf. In the binary case, we have used bold
greek letters to represent the densities themselves; while such notations remain usable here, e.g. \( \lambda \mid u \sim \lambda \), we usually prefer to talk about “the density of \( \lambda \) w.r.t. \( u \)” directly.

Like the binary case, given a random variable \( u \) in \( G \) and random probability tuples \( \lambda_1 \) and \( \lambda_2 \) over \( G \), if \( u \longrightarrow \lambda_1 \longrightarrow \lambda_2 \) forms a Markov chain, we say \( \lambda_2 \) is a physically degraded version of \( \lambda_1 \) w.r.t. \( u \), and write \( \lambda_2 \preceq \lambda_1 \).

Based on the binary case in Definition 4 and the discussion that follows, the notion of symmetric message densities can likewise be extended as follows:

**Definition 7:** Let \( \lambda \) be a random probability tuple over finite abelian group \( G \) and \( u \) be a random variable in \( G \), then we say \( \lambda \) has a symmetric density w.r.t. \( u \) if, for any deterministic \( u', u'' \in G \) and probability tuple \( \lambda' \) over \( G \),

\[
p_{\lambda \mid u}(\lambda' \mid u') = p_{\lambda \mid u}(\lambda' \oplus \overline{u} \mid u' \oplus u''),
\]

and where the normalization factor \( C(\lambda') \) does not vary with \( u' \).

To facilitate further discussion involving symmetric densities, we now define for any probability tuple \( \lambda, u \in G \),

\[
\langle \lambda \rangle \triangleq \{ \lambda \oplus \pi \mid u \in G \}
\]

as a kind of orbit containing \( \lambda \), and view its \( |G| \) elements as distinct for convenience; we then use e.g. \( p(\langle \lambda \rangle) \) to denote the probability that \( \lambda \) (as a random variable) lies in a certain (deterministic) \( \langle \lambda \rangle \). This allows each symmetric density to be reduced to a probability distribution of \( \langle \lambda \rangle \) through the following proposition:

**Proposition 25:** Let \( u \) be a random variable over a finite abelian group \( G \) and \( \lambda \) be a random probability tuple over it, then \( \lambda \) has a symmetric density w.r.t. \( u \) if and only if \( p(\lambda \mid u) \) satisfies

\[
p(\lambda \mid u) = p(\langle \lambda \rangle) \cdot \lambda(u).
\]

**Proof:** For any random probability tuple \( \lambda \) with a symmetric density w.r.t. \( u \), we have \( p(\langle \lambda \rangle \mid u) = p(\langle \lambda \rangle) \) due to (87), and

\[
p(\lambda \mid u) \cdot \lambda(u) = \frac{p(\lambda \mid u)}{p(\langle \lambda \rangle \mid u)} = \frac{p(\lambda \mid u)}{\sum_{u' \in G} p_{\lambda \mid u}(\lambda \oplus \overline{u} \mid u')} = \frac{\lambda(u)}{\sum_{u' \in G} \lambda(u' \oplus u')} = \lambda(u).
\]

Consequently,

\[
p(\lambda \mid u) = p(\langle \lambda \rangle \mid u)p(\langle \lambda \rangle \mid u) = p(\langle \lambda \rangle) \cdot \lambda(u).
\]

Conversely, any \( p(\lambda \mid u) \) in the form of (90) obviously satisfies (87) and (88) and thus makes \( \lambda \) symmetric w.r.t. \( u \).

Convex combinations of symmetric densities can be defined just like the binary case: let the index variable \( I \) be an arbitrary random variable and the reference variable \( u \) be a random variable over \( G \) that is independent from \( I \), then the density of a random probability tuple \( \lambda \) over \( G \) w.r.t. \( u \), represented by \( p(\lambda \mid u) \), is regarded as a convex combination of densities conditioned on \( I \) represented by \( p(\lambda \mid u, I) \). In particular, using the independence of \( u \) from \( I \), we have

\[
p(\lambda \mid u) = \sum_{I} p(I \mid u)p(\lambda \mid u, I) = \sum_{I} p(I)p(\lambda \mid u, I).
\]

From Proposition 25 it is easy to obtain the following results regarding symmetric densities and their convex combinations. Firstly, convex combinations of symmetric densities remain symmetric:

**Proposition 26:** Let \( I \) be an arbitrary random variable, \( u \) be a random variable over a finite abelian group \( G \) that is independent from \( I \), and \( \lambda \) be a random probability tuple over \( G \). If \( \lambda \) has a symmetric density w.r.t. \( u \) conditioned on each possible \( I \), then it is symmetric w.r.t. \( u \) unconditionally (i.e. when averaged over \( I \)).

**Proof:** By Proposition 25 each \( p(\lambda \mid u, I) \) has a corresponding \( p(\langle \lambda \rangle \mid I) \) that satisfies

\[
p(\lambda \mid u, I) = p(\langle \lambda \rangle \mid I)\lambda(u)
\]

Substitution into (92) gives \( p(\lambda \mid u) = p(\langle \lambda \rangle)\lambda(u) \) with \( p(\langle \lambda \rangle) = \sum_{I} p(I)p(\langle \lambda \rangle \mid I) \), so \( \lambda \) has a symmetric density w.r.t. \( u \).

Secondly, similar to \( D_2 \) in the binary case, we can construct a set of “minimal” symmetric densities such that all symmetric densities are convex combinations of them, allowing many properties satisfied by such densities to be applicable to all symmetric densities by linearity.

**Proposition 27:** Given any deterministic probability tuple \( \lambda^* \) over finite abelian group \( G \), we define conditional pmf (which can be regarded as a pdf w.r.t. the zero-dimensional Hausdorff measure)

\[
p(\lambda \mid u) = \sum_{u' \in G} \lambda^*(u \oplus u') \cdot 1\{\lambda = \lambda^* \oplus \overline{u}\},
\]

then \( \lambda \) has a symmetric density w.r.t. \( u \). Moreover, all symmetric densities are convex combinations of such densities with various values of \( \lambda^* \).

**Proof:** It is easy to verify that the \( p(\lambda \mid u) \) in (95) can be expressed in the form of (90) with pmf \( p(\langle \lambda \rangle) = 1\{\langle \lambda \rangle = \langle \lambda^* \rangle\} \), so \( \lambda \) is symmetric w.r.t. \( u \). Convex combinations of such densities can then yield any possible \( p(\langle \lambda \rangle) \) and thus any symmetric density.

For symmetric densities, physical degradation relationships are still preserved after taking convex combinations:

**Proposition 28:** Let \( I \) be an arbitrary random variable, \( u \) be uniformly distributed over a finite abelian group \( G \) and independent from \( I \), and \( \mu \) and \( \nu \) be random probability tuples over \( G \) that, when conditioned on \( I \), are symmetric w.r.t. \( u \) and satisfy \( \nu \preceq \mu \), then after averaging over all \( I \), we still have \( \nu \preceq \mu \) w.r.t. \( u \).

**Proof:** We need to prove that

\[
p(\nu \mid \mu, u) = \sum_{I} p(\nu \mid \mu, u, I)p(I \mid \mu, u)
\]

does not vary with \( u \). Given that \( \nu \preceq \mu \) conditioned on \( I \), \( p(\nu \mid \mu, u, I) \) is already independent from \( u \), so it suffices to prove that \( p(I \mid \mu, u) \) does not vary with \( u \) either. Using the independence between \( u \) and \( I \) as well as the symmetry of \( \mu \).
w.r.t. \( u \) conditioned \( I \) (and consequently, when averaged over \( I \)), we can find that

\[
p(I | \mu, u) = \frac{p(I | u)p(\mu | u, I)}{p(\mu | u)} = \frac{p(I)p(\mu | I)p(\mu | u)}{p(\mu | u)},
\]

which indeed does not vary with \( u \).

\( \nu(C; \cdot) \) can be applied to densities over finite abelian groups, like Definition 5 as follows:

**Definition 8:** Let \( (Z_i)_{i=1}^m \) be \( m \) finite abelian groups, \( Z \) be their direct product, \( C \) be a deterministic subset of \( Z \), \( i \in \{1, \ldots, m\} \), and \( \lambda_{-i} \triangleq (\lambda_1, \ldots, \lambda_{i-1}, \lambda_{i+1}, \ldots, \lambda_m) \) be \((m-1)\) message densities, with each \( \lambda_j \) defined over \( Z_j \). Now make \( u = (u_1, \ldots, u_m) \) uniformly distributed over \( C \), construct \((m-1)\) random probability tuples \( \lambda_{-i} \) such that for any \( j \neq i \), \( \lambda_j \) is over \( Z_j \), depends only on \( u_j \), and has \( \lambda_j | u_j \sim \lambda_j \), then the distribution of the probability tuple \( \nu(C; \lambda_{-i}) \) conditioned on the reference \( u_i \) is the message density denoted by \( \nu(C; \lambda_{-i}) \).

Similar to the binary case (Proposition 5 and Proposition 8), we can prove that \( \nu(C; \cdot) \) on symmetric densities preserves symmetry and physical degradation relationships, provided that \( C \) is a subgroup of \( Z \) or a coset thereof. When \( Z_i = Z_2^k \), \( i = 1, \ldots, m \), the direct product \( Z = Z_2^k (K = \sum_i K_i) \) can also be regarded as a vector space over \( Z_2 \), and it is then equivalent to require that \( C \) be an affine subspace of \( Z \).

**Proposition 29:** Let \( (Z_i)_{i=1}^m \) be \( m \) finite abelian groups, \( Z \) be their direct product, \( C \) be a deterministic subgroup or coset of \( Z \), and \( u = (u_1, \ldots, u_m) \) be uniformly distributed over \( C \). Now given \((m-1)\) random probability tuples \( \lambda_{-i} \), each \( \lambda_j \) defined over \( Z_j \), depending only on \( u_j \) and having a symmetric density with respect to it, the probability tuple \( \nu \triangleq \nu(C; \lambda_{-i}) \) over \( Z \), then satisfies the follows:

- \( \nu \) has a symmetric density w.r.t. \( u_i \);
- \( \nu \) depends only on \( u_i \), and is also a sufficient statistic for \( u_i \) given \( \lambda_{-i} \), i.e. \( u - u_i - \nu - \lambda_{-i} \) forms a Markov chain.

**Proof:** See Appendix I.

**Proposition 30:** Let \( (Z_i)_{i=1}^m \) be \( m \) finite abelian groups, \( Z \) be their direct product, \( C \) be a deterministic subgroup or coset of \( Z \), and \( u = (u_1, \ldots, u_m) \) be uniformly distributed over \( C \), and \( \lambda_{-i} \) and \( \lambda_{-i}' \) each be \((m-1)\) random probability tuples such that for each \( j \neq i \),

- \( \lambda_j \) and \( \lambda_j' \) are probability tuples over \( Z_j \), depend only on \( u_j \) in \( u \), and have symmetric densities w.r.t. \( u_j \);
- \( \lambda_j' \leq \lambda_j \) w.r.t. \( u_j \).

Now let \( u_i = \nu(C; \lambda_{-i}) \) and \( u_i' = \nu(C; \lambda_{-i}') \), then \( u_i' \leq u_i \) w.r.t. \( u_i \).

**Proof:** See Appendix I.

When the test channel has the form of 5 in Proposition 8, analogous to Proposition 4 the likelihood function used as the BP priors has a symmetric density over \( G \):

**Proposition 31:** Let \( u \) be a random variable in \( G \), \( y \in Y \) be another random variable with conditional pmf or pdf \( p(y | u) \), and \( \lambda \) be a probability tuple over \( G \) determined by \( y \) with \( \lambda(u) = p(y | u) \) before normalization. If there exists an measure-preserving group action \( \psi_u(\cdot) \) of \( G \) on \( Y \), such that \( p_y | u(y | u) = p(y | u(\psi_u(y) | 0)) \),

\[
p_y | u(y | u) = p(y | u(\psi_u(y) | 0)),
\]

then \( \lambda \) has a symmetric density w.r.t. \( u \).

**Proof:** See Appendix I.

However, what we actually need is symmetry over \( Z_2^k \) after the priors pass through a modulation mapping, so such mappings are investigated in detail below.

Given \( G \) with \(|G| = 2^K\), we define a modulation mapping \( \phi(\cdot) \) as a possibly random bijection from \( Z_2^k \) to \( G \), which can thus map between probability tuples over \( Z_2^k \) and those over \( G \) as well. In particular, since a probability tuple \( \lambda \) over \( G \) is a real-valued function over \( G \), the corresponding probability tuple over \( Z_2^k \) is simply a function composition \( \lambda \circ \phi \). In general, a random probability tuple \( \lambda \)'s symmetry w.r.t. random variable \( u \in G \) does not necessarily imply \( \lambda \circ \phi \)'s symmetry w.r.t. \( \phi^{-1}(u) \), nor vice versa; similar to the case of non-binary LDPC coding 37, dithering is necessary to maintain symmetry.

**Proposition 32:** Let \( \phi(\cdot) \) be a deterministic modulation mapping from \( Z_2^k \) to \( G \), \( u \) be a random variable uniformly distributed in \( G \) and \( \lambda \) be a random probability tuple over \( G \) with a symmetric density w.r.t. \( u \). Now define a random modulation mapping \( \phi_1(\cdot) \) with \( \phi_1(\tilde{c}) \triangleq \phi(\tilde{c} \oplus \varepsilon) \) for any vector \( \tilde{c} \in Z_2^k \), where \( \varepsilon \) is uniformly distributed over \( Z_2^k \) and independent from \( \lambda \) and \( u \), then \( \lambda \circ \phi_1 \) has a symmetric density w.r.t. \( \phi_1^{-1}(u) \).

**Proof:** See Appendix I.

Conversely, if we want to preserve symmetry when converting a density over \( Z_2^k \) into one over \( G \), dither should be introduced on the \( G \)-side:

**Proposition 33:** Let \( \phi(\cdot) \) be a deterministic modulation mapping from \( Z_2^k \) to \( G \), \( \tilde{c} \) be a random vector uniformly distributed in \( Z_2^k \) and \( \mu \) be a random probability tuple over \( Z_2^k \) with a symmetric density w.r.t. \( \tilde{c} \). Now define a random modulation mapping \( \phi_1(\cdot) \) with \( \phi_1(\tilde{c}) \triangleq \phi(\tilde{c} \oplus \delta) \), where \( \delta \) is uniformly distributed over \( \mu \) and independent from \( \mu \) and \( \tilde{c} \), then \( \mu \circ \phi_1^{-1} \) has a symmetric density w.r.t. \( \phi_1^{-1}(\tilde{c}) \).

**Proof:** Similar to the proof of Proposition 32 see Appendix I.

In light of these results, our code construction below will perform dithering over both \( Z_2^k \) and \( G \) by using the modulation mapping \( \phi_1(\tilde{c}) = \phi(\tilde{c} \oplus \varepsilon) \oplus \delta \). In this way, the symmetry of the priors over \( G \) from Proposition 31 can be promoted to symmetry over \( Z_2^k \), and through straightforward generalizations to Proposition 13 and Proposition 16 the BP messages and extrinsic information are also appropriately symmetric when a loop-free neighborhood is available, allowing their errors to be bounded using physical degradation relationships just like the binary case. At the same time, the extrinsic information of \( u_j \), denoted by \( \nu_j' \) in 12, will also have a symmetric density, enabling the recovery algorithm there to be used.

Finally, for probability tuples over \( Z_2^k \), the definition of the entropy \( H(\cdot) \) can be extended as follows for use in the analysis below. Given a deterministic probability tuple \( \mu \) over \( Z_2^k \), it can be viewed as the probability distribution of some random vector \( \tilde{c} \in Z_2^k \), i.e. \( \Pr[\tilde{c} = \tilde{c}'] = \mu(\tilde{c}') \) for any \( \tilde{c}' \in Z_2^k \). Now for any \( S \subseteq \{1, \ldots, K\} \), we can define \( H_S(\mu) \triangleq H(\tilde{c}_S) \triangleq H((c_k)_{k \in S}) \) as the joint entropy of the corresponding subset of bits in \( \tilde{c} \), and over the \((2^K - 1)\) possible non-empty choices of
S, the \((2^K - 1)\)-dimensional vector of \(H_S(\mu)\)'s can be called the entropy function \(^{38}\) of \(\mu\). For convenience, for any non-intersecting subsets \(S\) and \(S'\) of \(\{1, \ldots, K\}\), we also define the conditional entropy \(H_{\big| S'}(\mu) \triangleq H(\tilde{c}_S | \tilde{c}_{S'}) = H_{S,S'}(\mu) - H_{S'}(\mu)\). By averaging the components of the entropy function with the same \(|S|\), we obtain the average entropy function \(^{39}\)

\[
 h_k(\mu) = \left(\frac{K}{k}\right)^{-1} \sum_{S \subseteq \{1, \ldots, K\} \atop |S| = k} H_S(\mu). \tag{99}
\]

When \(\mu\) is a random probability tuple, we can take the expectation and obtain the (average) entropy function of its density. It is obvious that \(H_{\{1, \ldots, K\}}(\mu) = h_K(\mu) = H(\mu)\) and \(H_{\emptyset}(\mu) = h_0(\mu) = 0\). Moreover, if \(\mu\) has a symmetric density w.r.t. some uniformly distributed random vector \(\tilde{c}^* \in \mathbb{Z}_P\), then \(H_S(\mu)\) is simply the conditional entropy \(H(\tilde{c}_S | \mu)\), and the (average) entropy function then gives the amount of correlation among the bits in \(\tilde{c}^*\) in the conditional distribution \(p(\tilde{c}^* | \mu)\).

\[\text{B. Code Construction and the Quantization Algorithm}\]

Given a symmetric source coding problem with \(|G| = 2^K\), we thus construct the codebook

\[\mathcal{U} = \mathcal{U}(a) = \{u = u(b, a) \triangleq \phi(c) \triangleq \phi(bG \oplus a) \mid b \in \mathbb{Z}_{2^n}\}, \tag{100}\]

where \(G = (g_{ij})_{m \times n}\) is a binary sparse generator matrix, now with \(n_c \triangleq nK\) and \(n_b \triangleq nR\), and the scrambling sequence \(a \in \mathbb{Z}_{2^n}\). For each \(c = bG \oplus a \in \mathbb{Z}_{2^n}\), a codeword \(u = \phi(c) \in \mathbb{G}^m\) is obtained by mapping every \(K\) consecutive bits \(\tilde{c}_j \triangleq (c_{jK}, \ldots, c_{jK})\), where \(j_k \triangleq K(j - 1) + k\), into \(u_j \triangleq \phi_j(\tilde{c}_j)\) for \(j = 1, \ldots, n\). Each \(\phi_j\) is an independently dithered version of a fixed modulation mapping \(\phi\), that is, \(\phi_j(\tilde{c}_j) \triangleq \phi(\tilde{c}_j \oplus \epsilon_j) \oplus \delta_j\), with each \(\epsilon_j\) and \(\delta_j\) chosen i.i.d. uniform from resp. \(\mathbb{Z}_P\) and \(G\) and known to both the encoder and the decoder, and the combined dithering sequences are denoted by \(\epsilon = (\epsilon_j)_{j=1}^n \in \mathbb{Z}_P^n\) and \(\delta = (\delta_j)_{j=1}^n \in \mathbb{G}^n\). In particular, when \(G = Z_{2^K}\) with \(M = 2^K\), \(\phi(c)\) can (but not forced to) be the Gray mapping, and the resulting \(\mathcal{U}\) can be periodically extended into \(\Lambda = \mathcal{U} + M\mathbb{G}^n\) for use in M-ary MSE quantization.

Since every possible \(u \in \mathbb{G}^m\) occurs \(2^n\) times over the \(2^n\) \(\mathcal{U}(a)\)'s (each for one \(a\)), the discussion in Section III-A remains applicable. Specifically, given \(t > 0\) and under a fixed \(G\), each \(y \in \mathcal{Y}^n\) still gives a probability distribution \(q(b, a) \mid y) = e^{-ntd(u(b,a),y)}\) over all \((b,a)\)'s, and the quantization algorithm can still be regarded as an implementation of BPPQ as defined in Section III-A which gives the same average distortion \(D_{b}(t)\) as TPQ when the synchronization conditions are satisfied. Each \(\nu_q^{\mu_{\nu}}\) in (9) can be expressed by the factor graph in Fig. 9 where the variable nodes corresponding to \(a\), \(\epsilon\) and \(\delta\) have been omitted due to them being constant during the algorithm; the factor nodes between variable node \(b_{\nu}\)'s and \(c_{s}\)’s give the relationship \(c = bG \oplus a\), and they are thus called check nodes like the binary case, while each factor node between variable nodes \(c_1, \ldots, c_K\) and \(a)\) corresponds to \(u_j = \phi_j(\tilde{c}_j)\). The priors are also the same as the binary case: for any \(i', s\) and \(j\), \(\lambda_{ij}^{s} = \overline{b_{i'}}\) if \(b_{i'}\) has been determined (decimated) and \(\overline{\nu}\) otherwise, \(\lambda_{s}^{2} = \overline{\nu}\), while \(\lambda_{d}^{s}\), now a probability tuple over \(G\), is still given by (10) according to \(y\) (possibly adjusted by the recovery algorithm). By following the BP rules on this factor graph, we thus yield the quantization algorithm in Fig. 10 which is essentially the same as that used in \(^{11}\) except that the computation of the \(\mu_{\nu}^{\mu_{\nu}}\)'s has been moved to the beginning of each iteration in order to simplify the presentation of the analysis below. Like the binary case, there remains the choice between GD and PD in decimation as well as the decimation algorithm, which are dealt with in \(^{12}\) and will not be discussed in detail here.

\[\text{C. The Asymptotic Synchronization Conditions}\]

The synchronization conditions for BPPQ to yield the same distortion performance of TPQ at asymptotically large \(n\) can now be analyzed in essentially the same way as the binary case. \(G\) is still chosen to be the generator matrix of a variable-regular check-irregular LDGM code, with all \(n_b\) rows of \(G\) having \(d_b \geq 2\)'s, i.e. every variable node \(b_i\) in the factor graph has the same degree \(d_b\). To simplify analysis, for each \(j = 1, \ldots, n\), the columns \(j_1, \ldots, j_K\) corresponding to the bits mapped to the same \(u_j\) are made to possess the same number \(d\) of 1’s each, and we use \(w_d\) to denote the fraction of columns with this \(d\), and \(v_d \triangleq Kdw_d/(Rd_b)\) to denote the fraction of 1’s in such columns, which satisfy the constraints

\[
 \sum_d w_d = 1, \quad \sum_d v_d = 1, \quad w_d \geq 0, \quad d = 1, 2, \ldots. \tag{101}
\]

This \(d\) is henceforth called the check-degree of variable node \(u_j\). At each \(n\), the set of \(G\)'s with some given \(R\), \(d_b\) and \(w_d\) \((w_1, w_2, \ldots)\) (rounded so that \(nR\) and \(nw\) contain only integers) is the LDGM code ensemble with this degree.
distribution, and is denoted $\mathcal{G}^K(d_b, w)$, over which $G$ is uniformly distributed. TPQ (or BPPQ) instances having different values of $G$, $e$, $\delta$, $y$, as well as random sources $\omega^a$ and $\omega^b$ in the decoding steps of TPQ and BPPQ, thus form an ensemble over which probabilities can be defined. The analysis of the synchronization condition is performed again over the TPQ ensemble, and the reference codeword $(b^*, a^*)$ or the corresponding $c^*$ or $u^*$ remain defined as the TPQ result. The reference variables for the BP priors, messages and extrinsic information are the same as the binary case in Section IV-A with the addition of $c^*$ for $\mu^a_{\text{bc}}$ and $\mu^b_{\text{bc}}$.

It is easy to prove that the non-binary version of Proposition 2 still holds when conditioned on any fixed $G$, $e$ and $\delta$; in particular, $p(u^* | G, e, \delta)$ is uniform and $p(y^* | u^*; G, e, \delta) = \prod_{j} p_{j|u}(y_{j^*} | u_{j^*})$ is determined by the test channel, so both $y^*$ and $u^*$ remain independent from $e$ and $\delta$. The test channel's symmetry (Proposition 2) then ensures via Proposition 51 that each $\lambda^u_j$ has a symmetric density over $\mathbb{G}$ w.r.t. $u^*_j$, and by Proposition 52 after averaging over $e$ and $\delta$ (i.e. over all TPQ instances in the ensemble with the given $G$), $\lambda^u_j \circ \phi_j$ has a symmetric density over $\mathbb{Z}_2^K$ w.r.t. $\tilde{c}^*_j \triangleq (c^*_{j,k})_{k=1 \ldots N^u}$. Similar to the $I_u$ in the binary case, we therefore define

$$I_u \triangleq K - E\left[H(\lambda^u_j)\right] = K - H(u^*_j | \lambda^u_j) = K - H(u^*_j; y_j) = I(u; y),$$

(102)

where we have used the symmetry of $\lambda^u_j$, as well as the fact that $\lambda^u_j$ is a function of $y_j$ and a sufficient statistic for $u^*$ given $y_j$, and $I(u; y)$ is defined for the test channel. Since a modulation mapping only permutes the components of a probability tuple without changing its entropy, we have

$$I_u = K - E\left[H(\lambda^u_j \circ \phi_j)\right]$$

(103)

as well.

When analyzing the synchronization between TPQ and BPPQ in $b$-steps, similar to the binary case discussed in Section V-B, we can define $L_{i(L)}^b$ and $L_{i(L)}^b$ as BP approximations to each $v_i^b$ using $L$ iterations, and thus affected by a depth-$L$ neighborhood $N^u(L)$ of variable node $b_i$ in the factor graph. Like the binary case depicted in Fig. 4(b) $N^u(L)$ consists of repeated layers, but each layer is now as shown in Fig. 11(a) for clarity, the variable nodes $a_j$ previously omitted in Fig. 9 are also included here. If we consider a fixed $G$ whose $N^u(L)$ is loop-free, but still average the message densities over all $e$ and $\delta$ (i.e. conditioned on $G$ only), and define $C \triangleq \{(b, a, c) : c = bG + a\} \triangleq \{b\}$, then it is straightforward to show that Proposition 12 remains true (with $u$ replaced by $e$ and each $\lambda^u_j$ in the theorem replaced by $\lambda^u_j \circ \phi_j$ over $\mathbb{Z}_2^L$ and collectively denoted $\lambda^u \circ \phi$, i.e. $v_i^b$, $L_{i(L)}^b$ and $L_{i(L)}^b$ still possess the form $\nu(C; \lambda^u_j, \lambda^u_j \circ \phi_j)$; therefore, using the symmetry and the physical degradation relationships among the priors $\lambda^u_j$, $\lambda^u_j \circ \phi_j$, as well as the fact that $C$ is a linear subspace (and thus a abelian subgroup) of $\mathbb{Z}_2^{K+n}$, we can apply Proposition 29 and Proposition 30 to obtain the symmetry and physical degradation relationships among $v_i^b$, $L_{i(L)}^b$ and $L_{i(L)}^b$ w.r.t. $b^*_i$, and these properties remain true when averaged over all $G$ with a loop-free $N^u(L)$, which occur at high probability as $n \rightarrow \infty$. The synchronization error is thus still bounded by (28). Similarly, in $a$-steps the synchronization error can be bounded by (33). Using these results, it is straightforward to prove that conditions analogous to those in Proposition 18 and Theorem 20 are still sufficient for the synchronization conditions to be asymptotically satisfied, and the MI values used by these conditions can be evaluated for a given degree distribution via density evolution, just like the binary case. In particular, if we adopt the notations in Section IV-D, e.g. $L_{i(L)}^b$, to represent the densities of various binary message densities arising in DE, the DE rules at each variable node $b_i$ remains the same, i.e. (35) and (36) in b-steps and (39) and (40) in a-steps, while the check-node rules (34) and (36) are now different.

For concreteness, we now take a look at the computation of $L_{i(L)}^b$, $L_{i(L)}^b$, and $L_{i(L)}^b$ can be obtained in an analogous manner. Similar to Proposition 18 in the binary case, we define a sequence $i = i(n) \in \{1, \ldots, n_b\}$ that varies with $n$ with $\lim_{n \rightarrow \infty} (i - 1)/(n_b - 1)$ equal to some $I_b$. Given $n$, $i = i(n)$ and $G$ whose factor graph has a loop-free neighborhood $N_{i} = N^u(L)$ that can be further divided into $N^u_{i'}$ and $N^u_{i}$, we initialize the BP messages $\mu_{bc}^i$ from $N^u_{i'}$ to $N^u_{i}$ to be all-\$,$ and define the priors $\lambda^u_j$ for $i' \neq i$ to be $\tilde{b}^u_j$ when $i' < i$ and $\$ otherwise, and $L$ BP iterations then yield $\nu^b_i$. Now consider the density of this $\nu^b_i$ w.r.t. $b^*_i$ over the entire TPQ ensemble with block length $n$; as $n \rightarrow \infty$, the
discretizing the possible values of the probability tuples. As one-dimensional space, is amenable to practical discretization, only binary probability tuples, whose possible values lie in a

Fig. 11. Repetition units in each layer of the neighborhood $N_i^{(L)}$ in $2^K$-ary LDGM quantization.

D. The Case of Erasure-Like Problems

In the binary case, BEQ is important due to its comparative simplicity of analysis, and the optimized degree distributions of BEQ can serve as the starting point of degree distribution optimization in more general problems. For the $2^K$-ary LDGM code construction discussed here, this role is played by quantization problems in the form of Example 5, henceforth called erasure-like problems, and again in the limit of $t \to \infty$ with the modulation mapping $\phi : Z_2^K \to \mathcal{G} = Z_2^K$ chosen to be identity. Like BEQ, when TPQ is run on such a problem, we will prove that all probability tuples encountered in BP are erasure-like, making analysis of the DE process substantially easier.

The analysis is similar to that in Section V-C Recalling that the source alphabet $\mathcal{Y}$ is now the set of all affine subspaces of $\mathcal{G} = Z_2^K$, the test channel $p(u | y)$ is, for any $y \in \mathcal{Y}$, a uniform distribution over those $u \in y$. Consequently, using the generalization version of Proposition 5 in Section V-C when conditioned on $G$, $e$ and $\delta$ (not explicitly shown as conditions), $p(y) = \prod_j p_y(y_j)$, $p(u^* | y) = \prod_j p_{u_j}(u_j^* | y_j)$ is a uniform distribution over those $u^*$ with $u_j^* \in y_j$ for all $j$ (the set of those $u^*$, i.e. the Cartesian product of all $y_j$s, is henceforth denoted $U_y$), while $p(b^* \cap a^* | u^*)$ is a uniform distribution over those $(b^*, a^*)$ with $\phi(b^* G \oplus a^*) = u^*$. In other words, given $G$, $e$, $\delta$ and $y$, TPQ yields any $(b^*, a^*)$ in

\[
C_y \triangleq \{(b, a) | \phi(bG \oplus a) \in U_y\},
\]

which is non-empty due to the freedom in choosing $a$, with equal probability. Since $\phi(\cdot)$ is the identity map, we have $\phi(bG \oplus a) = bG \oplus a \oplus e \oplus \delta$, where both $e$ and $\delta$ are in $Z_2^K$ since $G = Z_2^K$. Now as $U_y$ is a Cartesian product of affine subspaces, it is itself an affine subspace of $Z_2^K$, so $C_y$ is an affine subspace of $Z_2^n + \phi$ as well.

As a generalization to Definition 6 we adopt the following definition for erasure-like probability tuples over $Z_2^K$:

Definition 9: A deterministic probability tuple $\mu$ over $Z_2^K$ is said to be erasure-like w.r.t. some deterministic $\tilde{c} \in Z_2^K$, if there exists a (non-empty) affine subspace $\tilde{C}$ of $Z_2^K$ such that $\tilde{c} \in \tilde{C}$ and $\mu(\tilde{c'} - \tilde{c}) = 1/|\tilde{C}| - 1$ $\tilde{c'} \in \tilde{C'}$. A random probability tuple $\mu$ over $Z_2^K$ is said to have an erasure-like density w.r.t. a random variable $\tilde{c}$ in $Z_2^K$ if $\tilde{c}$ is erasure-like w.r.t. $\tilde{c}$ with probability 1.

Each $\lambda_y^k$ from (10) is given by $\lambda_y^k(u) = (1/|y_j|) \cdot 1[u \in y_j]$, where $|y_j|$ is the cardinality of affine space $y_j$. Now that $\phi$ is identity and thus $\phi_j(\tilde{c}_j) = \tilde{c}_j \oplus e_j \oplus \delta_j$ (the addition is over $G = Z_2^K$), the probability tuple $\lambda_y^k \circ \phi_j$ over $Z_2^K$ is given by $\lambda_y^k \circ \phi_j(\tilde{c}) = 1/|y_j| \cdot 1[c_j \in y_j \oplus (e_j \oplus \delta_j)]$. Since $y_j \oplus (e_j \oplus \delta_j)$ is an affine subspace of $Z_2^K$, and under TPQ $u \in U_y$ implies that $\tilde{c}_j^* \triangleq (c_{j_1} \ldots, c_{j_k})$ is its member, we see that $\lambda_y^k \circ \phi_j$ is erasure-like w.r.t. $\tilde{c}_j^*$.

The computation of $\mu_{jk}^{uc}$ in BP preserves erasure-likeness:

Proposition 34: Let $(b^*, a^*)$ and the corresponding $c^* = b^* G \oplus a^*$ be random variables serving as the reference codeword. For each $j$ and $k$, if $\lambda_y^k \circ \phi_j$ is erasure-like w.r.t. $\tilde{c}_j^*$, then $\lambda_{jk}^{uc}(k' \neq k)$ is erasure-like w.r.t. $\tilde{c}_{jk'}^* \triangleq c_{jk'}^* \neq c_{jk}^*$, and each $\mu_{jk}^{uc} (k' \neq k)$ is erasure-like w.r.t. $\tilde{c}_{jk}^* \triangleq c_{jk}^*$, then $\mu_{jk}^{uc}$ given by

\[
\mu_{jk}^{uc} (c) = \sum_{\tilde{c}_{jk} = c} \lambda_{jk}^{uc}(\phi_j(\tilde{c}_j)) \prod_{k' \neq k} \mu_{jk'}^{uc}(\tilde{c}_{k'}), \quad c = 0, 1,
\]

is erasure-like w.r.t. $\tilde{c}_{jk}^* \triangleq c_{jk}^*$.

\[
\mu_{jk}^{uc} (c) = \sum_{\tilde{c}_{jk} = c} \lambda_{jk}^{uc}(\phi_j(\tilde{c}_j)) \prod_{k' \neq k} \mu_{jk'}^{uc}(\tilde{c}_{k'}), \quad c = 0, 1,
\]

is erasure-like w.r.t. $\tilde{c}_{jk}^* \triangleq c_{jk}^*$.

Proof: Each possible erasure-like value of $\lambda_y^k \circ \phi_j$ has a corresponding affine subspace $\tilde{C}_j$ of $Z_2^K$ such that $\tilde{c}_j^* \in \tilde{C}_j$, and for any $\tilde{c} \in Z_2^K$, $(\lambda_y^k \circ \phi_j) \tilde{c} = 1/|\tilde{C}_j| \cdot 1[\tilde{c} \in \tilde{C}_j]$, if $\tilde{c} \in \tilde{C}_j$ and 0 otherwise. Likewise, for every $k' \neq k$, each possible erasure-like value of $\mu_{jk'}^{uc}$ corresponds to an affine subspace $\tilde{C}_{k'}$, given by $\{\tilde{c} \in Z_2^K | \tilde{c}_{k'} = c\}$ if $\mu_{jk'}^{uc} \neq c_{jk'}$, then $\mu_{jk}^{uc}$ given by

\[
\mu_{jk}^{uc} (c) = \sum_{\tilde{c}_{jk} = c} \lambda_{jk}^{uc}(\phi_j(\tilde{c}_j)) \prod_{k' \neq k} \mu_{jk'}^{uc}(\tilde{c}_{k'}), \quad c = 0, 1,
\]

is erasure-like w.r.t. $\tilde{c}_{jk}^* \triangleq c_{jk}^*$.

Proof: Each possible erasure-like value of $\lambda_y^k \circ \phi_j$ has a corresponding affine subspace $\tilde{C}_j$ of $Z_2^K$ such that $\tilde{c}_j^* \in \tilde{C}_j$, and for any $\tilde{c} \in Z_2^K$, $(\lambda_y^k \circ \phi_j) \tilde{c} = 1/|\tilde{C}_j| \cdot 1[\tilde{c} \in \tilde{C}_j]$, if $\tilde{c} \in \tilde{C}_j$ and 0 otherwise. Likewise, for every $k' \neq k$, each possible erasure-like value of $\mu_{jk'}^{uc}$ corresponds to an affine subspace $\tilde{C}_{k'}$, given by $\{\tilde{c} \in Z_2^K | \tilde{c}_{k'} = c\}$ if $\mu_{jk'}^{uc} \neq c_{jk'}$, then $\mu_{jk}^{uc}$ given by

\[
\mu_{jk}^{uc} (c) = \sum_{\tilde{c}_{jk} = c} \lambda_{jk}^{uc}(\phi_j(\tilde{c}_j)) \prod_{k' \neq k} \mu_{jk'}^{uc}(\tilde{c}_{k'}), \quad c = 0, 1,
\]

is erasure-like w.r.t. $\tilde{c}_{jk}^* \triangleq c_{jk}^*$.
\( \tau \); obviously, \( \tilde{c}_k \) also contains \( \tilde{c}^* \) and, over \( \tilde{c} \in \mathbb{Z}_K^{\mathrm{f}}, \mu_{ji}^\mathrm{uc} (\tilde{c}_k) \) is likewise equal to a positive constant normalization factor if \( \tilde{c} \in \tilde{c}_k \) and 0 otherwise. Consequently, the intersection of \( \tilde{c}_0 \) and all \( \tilde{c}_k \) for \( k' \neq k \), denoted by \( \tilde{c} \), is still an affine subspace containing \( \tilde{c}_j \), and \( \lambda_j^\mathrm{f}(\phi_j(\tilde{c})) \prod_{k' \neq k} \mu_{ji}^\mathrm{uc} (\tilde{c}_k) \) is a constant for \( \tilde{c} \in \tilde{c} \) and zero otherwise. If all \( \tilde{c} \in \tilde{c} \) have the same \( \tilde{c}_k \) (which must be \( \tilde{c}_j^* \)), then \( \mu_{ji}^\mathrm{uc} = \tilde{c}_j^* \); otherwise, \( \tilde{c} \) being an affine subspace implies that the number of \( \tilde{c} \in \tilde{c} \) with \( \tilde{c}_k = 0 \) must be the same as those with \( \tilde{c}_k = 1 \), and \( \mu_{ji}^\mathrm{uc} = \tau \). Therefore, \( \mu_{ji}^\mathrm{uc} \) is always erasure-like w.r.t. \( c_j \).

We thus conclude that, under TPQ, all probability tuples involved in BP are indeed erasure-like, so the densities of the binary ones can be characterized solely in terms of their MI. Since (35) and (36) are unchanged from the binary case, now the corresponding MIs \( I^{(l)}_{\text{be}} \) and \( I^{(l)}_{\text{e}} \) still satisfy (65) and (66), and only the relationship between \( I^{(l)}_{\text{be}} \) and \( I^{(l)}_{\text{e}} \) remains to be derived.

Following the discussion in Section V.C, we consider the factor graph fragment in Fig. 11(b) with \( t \) having check-degree \( d \). Given \( I_{\text{be}} \triangleq I^{(l)}_{\text{be}} \), we let each incoming \( \mu_{ji}^\mathrm{be} \) and \( \mu_{ii}^\mathrm{ib} \) from the bottom of Fig. 11(b) be independently \( \beta_j \) (with probability \( I_{\text{be}} \)) or \( \tau \), then each message \( \mu_{ji}^\mathrm{ub} \) in the figure, conditioned on the reference codeword, is also independent and erasure-like, being \( c_j^* \) with probability \( I_{\text{ub},d} \triangleq (I_{\text{be}})^d \) and \( \tau \) otherwise.

We now know from Proposition 53 that the \( \mu_{ji}^\mathrm{uc} \) obtained from the \( \mu_{ji}^\mathrm{ub} \) is erasure-like as well. The probability that \( \mu_{ji}^\mathrm{uc} = c_j^* \) depends on \( k \), but for the purpose of computing \( I^{(l)}_{\text{be}} \) only its average value over \( k = 1, \ldots, K \) is needed, which is denoted by \( I_{\text{uc},d} \) and can be obtained from the entropy function of the density of \( \lambda_j^\mathrm{f}(\phi_j) \) w.r.t. \( \tilde{c}_j \). Specifically, let

\[
S \triangleq \{k' \in \{1, \ldots, K\} \mid \mu_{ji}^\mathrm{ub} = c_j^* \} \tag{106}
\]

be the set of “known” incoming messages, then each \( S \) with \( |S| = l \) occurs with probability \( \prod_{k \in S} (\lambda_j^\mathrm{f}(\phi_j)) \prod_{k \notin S} (1 - \lambda_j^\mathrm{f}(\phi_j)) \) and given \( S \) and \( \lambda_j^\mathrm{f}(\phi_j) \), the probability that \( \mu_{ji}^\mathrm{uc} = \tilde{c}_j^* \) is simply \( 1 - H_k(\lambda_j^\mathrm{f}(\phi_j)) \). Taking expectations over \( S \) and \( \lambda_j^\mathrm{f}(\phi_j) \), and denoting the \( (\lambda_j^\mathrm{f}(\phi_j))^\text{-}\text{expectation of e.g. } H_k(\lambda_j^\mathrm{f}(\phi_j)) \) by just \( H_k(\lambda_j^\mathrm{f}(\phi_j)) \), we get

\[
I_{\text{uc},d} = 1 - \frac{1}{K} \sum_{k=1}^{K} E_S \left[ H_k(\lambda_j^\mathrm{f}(\phi_j)) \right] \tag{107}
\]

\[
= 1 - \frac{1}{K} \sum_{k=1}^{K} \sum_{S \subseteq \{1, \ldots, K\} \setminus \{k\}} p_{d,S} \cdot H_k(\lambda_j^\mathrm{f}(\phi_j)) \tag{107}
\]

\[
= 1 - \frac{1}{K} \sum_{l=0}^{K-1} (\prod_{k=1}^{K} \sum_{S \subseteq \{1, \ldots, K\} \setminus \{k\}} H_k(\lambda_j^\mathrm{f}(\phi_j))) \tag{107}
\]

\[
= 1 - \frac{1}{K} \sum_{l=0}^{K-1} \left( \frac{K - 1}{l} \right) \cdot p_{d,l} \cdot (h_{l+1} - h_l), \tag{107}
\]

where \( h_l \) is the average entropy function \( h_l(\lambda_j^\mathrm{f}(\phi_j)) \) with expectation taken over \( \lambda_j^\mathrm{f}(\phi_j) \).

Finally, according to the BP rule computing \( \mu_{ji}^\mathrm{be} \), in Fig. 11(b), its MI averaged over \( k \) and \( d \) should be

\[
I^{(l)}_{\text{be}} = \sum_{d} v_d I_{\text{uc},d} \alpha_{d,l}(I^{(l-1)}_{\text{be}}), \tag{108}
\]

where we have defined for brevity

\[
\alpha_{d,l}(x) \triangleq \left( \frac{K - 1}{l} \right) \cdot x^{d-1} p_{d,l} \tag{109}
\]

and

\[
I_{e,l} \triangleq 1 - (h_{l+1} - h_l). \tag{110}
\]

In other words, when expressed in the form of (70), we now have

\[
f(x) = \sum_{l=0}^{K-1} \frac{I_{e,l}}{I_u} \sum_{d} v_d \alpha_{d,l}(x), \tag{111}
\]

where the \( I_{e,l} \)'s can be shown using (103) to satisfy

\[
\sum_{l=0}^{K-1} I_{e,l} = K - h_K = K - E \left[ H(\lambda_j^\mathrm{f}(\phi_j)) \right] = I_u. \tag{112}
\]

Having obtained the \( f(x) \) of erasure-like problems, the optimization of degree distribution can proceed using Theorem 24 just like the binary case. For general symmetric source coding problems, erasure approximation using the same entropy function (and thus the same \( I_{e,l} \)'s) and correction with DE results also allow degree distribution optimization to proceed iteratively, which is essentially the optimization method in [16] and has been shown to give good results in [12]. We have thus obtained a sound theoretical basis for this optimization method.

VI. CONCLUSION AND FUTURE WORK

In this paper, considering the LDGM-based quantization codes for symmetric source coding problems previously analyzed in [16] and [13], we have introduced the synchronization conditions that allow the distortion performance of TPQ, namely \( D_0(t) \), to be achieved by the practically possible BPPQ, and then proved that degree distributions satisfying certain criteria allow these synchronization conditions to be satisfied in an asymptotic sense as block length \( n \) and iteration count \( L \) go to infinity. By making use of the properties of symmetric message densities, both binary ones and those over an abelian group, these results have been obtained not only for binary code constructions but for \( 2^K \)-ary BICM-like constructions as well. In this way, a firm theoretical basis for the optimization methods in [16] has been established.

On the other hand, the asymptotic synchronization conditions are not able to analyze the impact of a loss of synchronization between BPPQ and TPQ, sometimes called a decimation error, which is inevitable in practice due to finite \( n \) and \( L \). Such decimation errors can be tackled in practice with the recovery algorithm proposed in [16] and [12], and some ideas, including the introduction of an idealized recovery algorithm in [12], have been proposed to analyze the resulting performance. However, except for the sometimes
simpler BEQ case, the analysis has yet to be made rigorous and should be improved accordingly. Moreover, most analysis work so far consider only the probabilistic decimator rather than the greedy decimator used in practice, and all optimization methods are also based on them. Some analysis of the characteristics of GD, even empirical ones, would likely allow better optimization and a more thorough understanding of the quantization process.

**APPENDIX I PROOFS**

**A. Proof of Proposition 7**

Due to the symmetry of $p(y)$ and $d(u, y)$, the optimal test channel $p(u | y)$ can be assumed to give a uniform $p(u)$; otherwise the test channel $p'(u | y) = \frac{1}{|\mathcal{G}|} \sum_{\mathbf{g} \in \mathcal{G}} p_{\mathbf{g}}(u \oplus v | \psi_{\nu}(y))$ would be better as it gives the same $D$, the same or lower $R = I(u, y)$ (mutual information is convex w.r.t. the channel transfer probabilities), and the corresponding $p'(u) = \sum_{y} p(y)p'(u | y)$ is uniform. Now $H(u)$ is a constant, so given $D$ the minimization of $I(u; y)$ is equivalent to the maximization of $H(u | y)$, which is easily done with Lagrange multipliers and yields the results in Proposition 1. It can be verified that the corresponding $p(u)$ is indeed uniform.

**B. Proof of Proposition 8**

Conditioned on a fixed $G$, we have found in Section III-A that the TPQ yields any $(b^*, a^*, u^*)$ satisfying $b^*G \oplus a^* = u^*$ with probability proportional to $q(b^*, a^*) = e^{-ntd(u^*, y)/P}$, in other words, $p(b^*, a^*, u^* | y) = e^{-ntd(u^*, y)}/Q(y)$, where the normalization factor

$$Q(y) = \sum_{b^*, a^*} e^{-ntd(u^*, y)}/P = 2^n b_y \sum_{u^* \in \mathbb{Z}_2^n} e^{-ntd(u^*, y)}/Q(y)$$

$$= 2^n \prod_{j=1}^{n} 1 + u_{j} y_j = 2^n \prod_{j=1}^{n} Q(y)$$

(113)

(114)

due to each $u^* \in \mathbb{Z}_2^n$ having $2^n$ combinations of $(b^*, a^*)$ with $u^* = b^*G \oplus a^*$. We thus have

$$p(b^*, a^*, u^* | y) = 2^{-n} 1 [b^*G \oplus a^* = u^*] \prod_{j=1}^{n} e^{-ntd(u^*, y)} / Q(y)$$

(115)

so the joint distribution of $(b^*, a^*, u^*, y)$ given $G$ is known, and the desired results immediately follow.

**C. Proof of Proposition 9**

Here we only consider the case where $y$ and thus

$$z = f(y) \triangleq \mu(0) = \frac{p_{y}b(y | 0)}{p_{y}b(y | 0) + p_{y}b(y | 1)}$$

are continuous-valued. Since $\psi_{\nu}(\cdot)$ is a group action of $\mathbb{Z}_2$, $\psi_1$ must be a bijection with $\psi_1^{-1} = \psi_1$, and using $p_{y}b(y | 1) = p_{y}b(\psi_1(y) | 0)$, we see that $f(\psi_1(y)) = 1 - f(y)$.

According to Definition 3 we need to prove that, for any $z \in [0, 1]$,

$$p_{z \mid b}(z | 1) = p_{z \mid b}(1 - z | 0),$$

(117)

$$\left(1 - z\right) \cdot p_{z \mid b}(z | 0) = z \cdot p_{z \mid b}(1 - z | 0).$$

(118)

Let $\Delta z$ be a small positive number. Given an arbitrary $z_0 \in [0, 1 - \Delta z]$, we can define $Z_0 \triangleq [z_0, z_0 + \Delta z]$, $Z_1 \triangleq 1 - Z_0 = [1 - z_0 - \Delta z, 1 - z_0]$, $Y_0 \triangleq f^{-1}(Z_0)$ and $Y_1 \triangleq f^{-1}(Z_1)$. Now for any $z = f(y)$, the events $z \in Z_0$, $1 - z \in Z_1$, $y \in Y_0$ and $\psi_{\nu}(y) \in Y_1$ are equivalent, and we can define $P_{z}$ as their probability conditioned on a fixed $b = 0, 1$ and $P \triangleq P_0 + P_1$.

It can be observed that

$$P_1 = P_{z | b}(z | 1) = P_{y | b}(y | 0) = Pr[y \in Y_0 | b = 1]$$

= $Pr[\psi_1(y) \in Y_0 | b = 0] = Pr[1 - z \in Z_0 | b = 0],$

(119)

and by letting $\Delta z \to 0$, we get (117).

To prove (118), first note from (116) that, for any $y \in Y_0$, $z_0 P(y) \leq f(y)P(y) = y_{b}(y | 0)\leq (z_0 + \Delta z)P(y)$, where $P(y) \triangleq p_{y | b}(y | 0) + p_{y | b}(y | 1)$. Integrating over $y \in Y_0$ we obtain

$$z_0 P \leq P_0 \leq (z_0 + \Delta z)P, \quad i.e. \quad z_0 \leq P_0 / P \leq z_0 + \Delta z.$$

(121)

As $\Delta z \to 0$, this becomes

$$p_{z \mid b}(z | 0) / p_{z \mid b}(z | 0) + p_{z \mid b}(z | 1)) = z_0,$$

which is equivalent to (118).

**D. Proof of Proposition 10**

We use $q_i \equiv \mu_i(0)$ $(i = 1, 2)$ to uniquely represent each $\mu_i$. $b - q_1 - q_2$ thus forms a Markov chain. As $b$ is equiprobable and $\mu_1$ and $\mu_2$ are symmetric, we have

$$q_2 = p_{b \mid q_2}(0 | q_2) = \int_{0}^{1} p_{b \mid q_1}(0 | q_1)p(q_1 | q_2) dq_1$$

$$= \int_{0}^{1} q_1 p(q_1 | q_2) dq_1 = E[q_1 | q_2].$$

(123)

Now let $f(q) = H_2(q) \cdot \ln 2 + 2(q - q_{2}^2)$, which is concave in the interval $[0, 1]$ as $f''(q) = 4 - 1/(q + 1/(1 - q)) \leq 0$ for $0 < q < 1$, so by Jensen’s inequality $f(q) \geq E[f(q_1 | q_2)$, i.e.

$$E[H_2(q_1 | q_2) + \frac{2}{\ln 2} E[(q_1 - q_{2}^2) | q_2] \leq H_2(q_2).$$

(124)

As $I(\mu_i) = 1 - E[H(\mu_i)] = 1 - E[H_2(q_1)], i = 1, 2$, taking the expectation of (124) over $q_2$ yields the desired result.

**E. Proof of Proposition 11**

By definition it suffices to prove that, given $b \in C$, if each $\lambda_i$ ($i' \neq i$) is either $b_i$ or $\overline{b_i}$, then $\nu \triangleq \nu(C; \lambda_{i' \neq i})$ is either $b_i$ or $\overline{b_i}$. To prove this, we note that

$$C' \triangleq \{b \in C | b_{i'} = b_i \text{ for all } i' \neq i \text{ with } \lambda_i = \overline{b_i}\}$$

(125)

is a non-empty (since $b \in C'$) affine subspace of $C$, so either all vectors in $C'$ have the same value at the $i$-th position (which is necessarily $b_i$), or exactly half is 0 (or 1) at that position. From the definition of $\nu$, it is $b_i$ in the former case and $\overline{b_i}$ in the latter.
F. Proof of Proposition 17

1 and 2 follow immediately from respectively 25 and 31 using Proposition 9 and the symmetry of the densities. Alternatively, properties of the MIs of DE results $I_{b,\text{ext}}(I_n,L)$ and $\mathcal{T}_{b,\text{ext}}(I_n,L)$ can also be obtained by noting that degradation relationships are preserved in every DE step.

In order to obtain properties 3 and 4, it is necessary to prove for a fixed $n$ and $l \leq L$ that $\nu^{b,s}$, $\nu^{b,s}(L)$, $\mathcal{L}^{b,s}(L)$, and $\mathcal{T}^{b,s}(L)$ are respectively ordered by degradation in $i$, while $\nu^{b,s}$, $\nu^{b,s}(L)$, and $\mathcal{L}^{b,s}(L)$ are respectively ordered by degradation in $j$. As the methods are essentially the same, we give the proof for $\nu^{b,s}(L)$.

Recall that for any $i$, $\nu^{b,s}(L) = \nu(C; \lambda^b_{x,y}, \lambda^s_{y,y})$, with $C$ defined in (20), all $\lambda^b_x = \lambda^s_x$, and $\lambda^s_y = \lambda^b_{y,y}$ if $i'' < i$ and $\pi$ otherwise, while $\nu^{b,s}(L)$ is the density of this $\nu^{b,s}$ w.r.t. $b^s_L$ over $G$ uniformly distributed in $G^s_n(L)$. In other words, $(b^s_L, \nu^{b,s}(L))$ can be viewed as random variables defined on the probability space

$$\Omega \triangleq \{(G,y,\omega^b,\omega^s)| G \in G^s_n(L), y \in \mathcal{Y}^n, \omega^b \in [0,1]^n_b, \omega^s \in [0,1]^n_s\}$$

containing TPQ instances with $G$ having loop-free neighborhoods, and $\nu^{b,s}(L)$ is their conditional distribution.

Now for any $i' > i$, $\nu^{b,s}(L)$ is the density of $\nu^{b,s}$ w.r.t. $b^s_L$ over uniform $G \in G^s_n(L)$, and the probability space $\Omega'$, over which the random variables $b^s_L$ and $\nu^{b,s}(L)$ are defined, is given by (126) with $G^s_n(L)$ replaced by $G^s_n(L')$. As $\nu^{b,s}(L)$ and $\nu^{b,s}$ are conditional distributions of random variables defined on respectively $\Omega$ and $\Omega'$, for the purpose of comparison we define a permutation $\pi$ of $\{1, \ldots, n_b\}$ as $\pi(i'') = (i'' + (i'' - i)) \bmod n_b$ (where the modulo operation is onto $\{1, \ldots, n_b\}$), which then gives a probability-preserving bijection from $\Omega$ to $\Omega'$ that renumbers every variable node $b^s_L$ in each TPQ instance in $\Omega$ into $b^s_L(\pi(i''))$ specifically, the TPQ instance $(G = (g^{s}_{(i'')(y,n,x,y,\omega^b,\omega^s)})) \in \Omega$ is mapped to $(G' = (g^{s}_{(i'')(y,n,x,y,\omega^b,\omega^s)}) \in \Omega'$, where $g^{s}_{(i'')(y,n,x,y,\omega^b,\omega^s)} = g^{s}_{(i'')(y,n,x,y,\omega^b,\omega^s)}$ so that the factor graph remains unchanged apart from the renumbering, and $\omega^b$ can be transformed into $\omega^b$ in a probability-preserving manner such that the pre-transformation $b^s_L(\pi(i''))$ is equal to the post-transformation $b^s_L(\pi(i''))$.

As $G \in G^s_n(L)$ and only if $G' \in G^s_n(L')$, we have indeed obtained a probability-preserving bijection from $\Omega$ to $\Omega'$.

With this bijection, the random variable $b^s_L$ on $\Omega'$ becomes $b^s_L$ on $\Omega$, and $\nu^{b,s}$ on $\Omega'$ becomes $\nu^{b,s} = \nu(C; \lambda^b_{x,y}, \lambda^s_{y,y})$ defined on $\Omega$, where each $\lambda^b_x$ is $\lambda^s_x$ when $i'' < i$ or $i'' > n_b - (i'' - i)$ and is $\pi$ otherwise, i.e. $\lambda^b_x$ contains $\{i'' - i\}$ extra “known” probability tuples compared to $\lambda^s_x$. Consequently, the density of $\nu^{b,s}$ w.r.t. $b^s_L$ on $\Omega$ is the same as that of $\nu^{b,s}$ w.r.t. $b^s_L$, i.e. $\nu^{b,s}(L)$ and by Proposition 8 we also have $\nu^{b,s} \leq \nu^{b,s}(L)$, hence $\nu^{b,s}(L)$ is a degraded version of $\nu^{b,s}(L)$.

G. Proof of Proposition 18

Direct part: To prove (47), we start from (27). For any $l \leq L$ and sufficiently large $n$ (such that $G^s_n(L)$ is non-empty), (27) can be reexpressed as

$$E \left[ (\omega^b_{(i')(0)} - \nu^{b,s}(0))^2 \mid G \in G^s_n(L) \right] \leq \frac{\ln 2}{2} \left( (I^s_{b,\text{ext}}(n,L) - I^s_{b,\text{ext}}(0,L)) \right) ; \quad (127)$$

As $\mathcal{P}[G \notin G^s_n(L)] = \mathcal{P}_{n,L}$, the unconditional expectation (over all $G \in G^s_n(s,n,L)$) can also be bounded as

$$E \left[ (\omega^b_{(i')(0)} - \nu^{b,s}(0))^2 \right] \leq \frac{\ln 2}{2} \left( (I^s_{b,\text{ext}}(n,L) - I^s_{b,\text{ext}}(0,L)) + I_{n,L} \right) ; \quad (128)$$

For any $\epsilon > 0$, let $I^n_{b,\text{ext}} \triangleq \max(0, I^n_{b,\text{ext}} - \epsilon)$, $I^n_{b,\text{ext}} \triangleq \min(1, I^n_{b,\text{ext}} + \epsilon)$, then $I^n_{b,\text{ext}} \leq I^n_{b,\text{ext}} \leq I^n_{b,\text{ext}}$ for all $n$ larger than some threshold $n_0(\epsilon)$, so we can use the monotonicity of $I^s_{b,\text{ext}}(n,L)$ and $I^s_{b,\text{ext}}(n,L)$ w.r.t. $I_b$ to transform (128) into

$$E \left[ (\omega^b_{(i')(0)} - \nu^{b,s}(0))^2 \right] \leq \frac{\ln 2}{2} \left( (I^s_{b,\text{ext}}(n,L) - I^s_{b,\text{ext}}(0,L)) + I_{n,L} \right) ; \quad (129)$$

and taking the $n \to \infty$ limit then yields, for any $\epsilon > 0$,

$$\limsup_{n \to \infty} E \left[ (\omega^b_{(i')(0)} - \nu^{b,s}(0))^2 \right] \leq \frac{\ln 2}{2} \left( (I^s_{b,\text{ext}} - I^s_{b,\text{ext}}) \right). \quad (130)$$

Now $\mathcal{T}_{b,\text{ext}} = I^s_{b,\text{ext}}$, so $\mathcal{T}_{b,\text{ext}} - I^s_{b,\text{ext}}$ is the sum of $I^s_{b,\text{ext}} - I^s_{b,\text{ext}}$ and $I^s_{b,\text{ext}} - I^s_{b,\text{ext}}$. Since we have assumed that $I^s_{b,\text{ext}}$ is continuous at $I^s_{b,\text{ext}}$, the former can be made arbitrarily small by choosing a sufficiently small $\epsilon$, and the latter then vanishes as well when $l \to \infty$. We have thus proved (47) as desired, and (48) can be proved similarly.

Converse part: Assume that $\mathcal{T}_{b,\text{ext}} - I^s_{b,\text{ext}} \leq \delta > 0$. By Proposition 11 there exists $\epsilon > 0$ such that for any $n, l$, and $i$ satisfying $l \leq L$, $F_{n,L} = 1/2$ and $I^s_{b,\text{ext}} - I^s_{b,\text{ext}} \geq \delta/4$ (where $I_b = (i - 1)/(n_b - 1)$), we have

$$E \left[ (\omega^b_{(i')(0)} - \nu^{b,s}(0))^2 \mid G \in G^s_n(L) \right] \geq 2 \epsilon, \quad (131)$$

and (49) thus holds. Now just as before, we have to find, for any given $l$ and $n_0$, some $n \geq n_0$, $L \leq 1$ and $i$ with $F_{n,L} = 1/2$ and $I^s_{b,\text{ext}} - I^s_{b,\text{ext}} \geq \delta/4$. Firstly, since $I^s_{b,\text{ext}} \leq I^s_{b,\text{ext}}$ for any $I_b$ and in particular $I^s_{b,\text{ext}}$, we have

$$\mathcal{T}_{b,\text{ext}} - I^s_{b,\text{ext}} \geq \delta. \quad (132)$$
Using the continuity of $I_{b}^{(t_{b,ext})}$ w.r.t. $I_{b}$, an $I_{b}^{t_{b,ext}} \in (I_{b}^{+}, 1]$ (except that $I_{b}^{+}$ is allowed to be 1 when $I_{b}^{-} = 1$) can be found that makes

$$I_{b,ext}^{(t_{b,ext})} \leq I_{b,ext}^{(t_{b,ext} + \delta/4)}. \quad (133)$$

Now let $n_{h,1} \triangleq 1 + 1/(I_{b}^{+} - I_{b}^{-})$ (or 2 when $I_{b}^{-} = 1$), and choose $n_{1}$ such that any $n \geq n_{1}$ has the corresponding $n_{b} \geq n_{h,1}$. Then there exists, for any $n \geq n_{1}$, integer $i \in \{1, \ldots, n_{b}\}$ such that $(i - 1)/(n_{h,1} - 1) \in [I_{b}^{-}, I_{b}^{+}]$. If we further choose any $L \geq l$, then by $I_{b,ext}^{(t_{b,ext})} = \lim_{n \to \infty} I_{b,ext}^{(t_{b,ext}, n, L)}$ and (129), there also exists $n_{2}$ such that for any $n \geq n_{2},$

$$I_{b,ext}^{(t_{b,ext}, n, L)} \leq I_{b,ext}^{(t_{b,ext})} + \delta/4, \quad I_{b,ext}^{(t_{b,ext}, n, L)} \leq \frac{1}{2}. \quad (134)$$

On the other hand, since $I_{b,ext}^{(t_{b,ext})} = \lim sup_{n \to \infty} I_{b,ext}^{(t_{b,ext}, n, L)}$, for the given $n_{0}$ we can find $n \geq \max(n_{0}, n_{1}, n_{2})$ such that

$$I_{b,ext}^{(t_{b,ext}, n, L)} \geq I_{b,ext}^{(t_{b,ext}, n, L)} - I_{b,ext}^{(t_{b,ext}, n, L)} \geq \delta/4. \quad (135)$$

As the $n$ chosen above satisfies $n \geq n_{1}$, an $i \in \{1, \ldots, n_{b}\}$ can be found such that $(i - 1)/(n_{h,1} - 1) \in [I_{b}^{-}, I_{b}^{+}]$, and (136) is then satisfied at $I_{b} = (i - 1)/(n_{h,1} - 1)$, in which case $I_{b,ext}^{(t_{b,ext}, n, L)} = I_{b,ext}^{(t_{b,ext}, n, L)} \geq \delta/4$, making (49) satisfied.

The part of the result when (46) fails to hold can be proved similarly. ■

H. Proof of Proposition 21

Since the probabilities here are defined over the TPQ ensemble, by the arguments in Section III-A, given $y$ and $G$ each reference codeword $(b^{*}, a^{*})$ occurs with probability $p(b^{*}, a^{*} | y, G) = C \cdot q(b^{*}, a^{*} | y)$ with $C$ being a normalization factor. Substituting this into (8) and (9), we see that

$$I_{b}^{(t_{b,ext})} = I_{b}^{(t_{b,ext})} + I_{b}^{(t_{b,ext})} \geq I_{b,ext}^{(t_{b,ext})} + I_{b,ext}^{(t_{b,ext})} \geq \delta/4. \quad (137)$$

Therefore, in each TPQ instance, we have

$$H(I_{b}^{(t_{b,ext})}) = H(a_{b}^{*} | a_{1}^{*}, \ldots, a_{k-1}^{*}, y, G), \quad (138)$$

$$H(I_{b}^{(t_{b,ext})}) = H(b_{1}^{*} | a_{b}^{*}, a_{2}^{*}, b_{1}^{*}, \ldots, b_{1}^{*} \ldots, y, G). \quad (140)$$

where no expectation has been taken over the conditions in the entropy. Now take the expectation over all TPQ instances (i.e. over $y, G, b^{*}$ and $a^{*}$) and sum over $i$ and $j$, and we get

$$\sum_{j=1}^{n_{y}} H(I_{b}^{(t_{b,ext})}) + \sum_{i=1}^{n_{y}} H(I_{b}^{(t_{b,ext})}) = H(b^{*}, a^{*} | y, G). \quad (141)$$

On the other hand, from Proposition 3 we have

$$H(b^{*}, a^{*} | y, G) = H(b^{*}, a^{*} | u^{*}, y, G) + H(u^{*} | y, G) = n_{b} + n_{y} H(u | y), \quad (142)$$

Recall that we have defined $n_{b} = nR^{(n)}$ for each $n$ with $\lim_{n \to \infty} R^{(n)} = R$. The part of the result when (46) fails to hold can be proved similarly. ■

I. Proof of Proposition 23

Eqs. (76) and (77) follow immediately from respectively (75) and (74). For (78), first note from (67)–(69) that, under BEQ,

$$\int_{0}^{1} f(x) dx = \sum_{d} \frac{\nu_{d}}{d} x^{d} |_{x=0} = \frac{1}{R_{b}^{*}}, \quad (146)$$

and $h'(y)/g(y) = d_{b}$. Therefore, letting $y = 1 - I_{b} f(x)$, we have

$$\int_{0}^{1} (1 - I_{b}) d_{b}^{(t_{b,ext})} dx \quad (147)$$

$$= \int_{0}^{1} (1 - x) I_{b} f'(x) dx$$

$$= d_{b} I_{b} \int_{0}^{1} (1 - x) f'(x) dx$$

$$= d_{b} I_{b} \left( (1 - x) f(x) |_{x=0} + \int_{0}^{1} f(x) dx \right)$$

$$= d_{b} I_{b} (-v_{1} + 1/(R_{b}^{*})) = \frac{I_{b}}{R_{b}^{*}} - d_{b} I_{b} v_{1}.$$
I. Proof of Proposition 29

In the proof we will use \( u'_{-i} \) to denote \( (u'_1, \ldots, u'_{i-1}, u'_{i+1}, \ldots, u'_m) \), and \( Z_{-i} \) to denote the direct product of \( (Z_j)_{j \neq i} \).

As \( C \) is a coset, we have \( C = X' \oplus u_0 \), where \( X' \) is the corresponding subgroup of \( Z \) and \( u_0 \in Z \).
For each \( d \in D_i \), we define \( X_d \overset{\text{def}}{=} \{ d \in X' \mid d_i = d \} \), then \( X_0 \) is in turn a subgroup of \( X \), and any other \( X_d \) is either empty or equal to \( X_0 \oplus d \) where \( d \) is any element in \( X_d \). We can also define \( D_i \overset{\text{def}}{=} \{ d \in Z_i \mid X_d \neq \emptyset \} \), which is a subgroup of \( Z_i \), and it is easy to prove that \( X_d \oplus X_{d'} = X_{d+d'} \) for all \( d, d' \in D_i \).

Note that since \( u \) is distributed over \( C \), for \( p(\nu \mid u) \) it is only necessary to consider \( u \in C \), and similarly for \( p(\nu \mid u_i) \) only \( u_i \in C_i \) is relevant, as \( u_i \) never takes other values.

By linearity, we may first assume that each \( \lambda_j \) is discrete and has a conditional pmf in the form of (95), i.e.

\[
p(\lambda_j \mid u_j) = \sum_{u'_j \in Z_j} \lambda^*_j(u_j \oplus u'_j) \cdot 1 \left[ \lambda_j = \lambda^*_j \oplus u'_j \right]
\]

for some deterministic probability tuple \( \lambda_j^* \) over \( Z_j \), as a result, given \( u \), the probability that \( \lambda_j = \lambda_j^* \oplus u'_j \) (here \( \lambda_j^* \oplus u'_j \) for different values of \( u'_j \) are safely viewed as distinct) for all \( j \neq i \) is \( \lambda^*_{-i}(u_{-i} \oplus u'_{-i}) \overset{\text{def}}{=} \prod_{j \neq i} \lambda^*_j(u_j \oplus u'_j) \), and the corresponding value of \( \nu \overset{\text{def}}{=} \nu(C; \lambda_{-i}) \) is denoted by \( \nu_{u'_{-i}} \), which is given by (without normalization)

\[
\nu_{u'_{-i}}(u) = \sum_{u' \in C: u'_{-i} = u_{-i}} \prod_{j \neq i} \lambda_j(u'_j) \cdot 1 \left[ \nu = \nu_{u'_{-i}} \right]
\]

(149) whose r.h.s. is simply (from \( 149 \)) \( \nu_{u'_{-i}}(u) \) with \( u = d \oplus u_0 \in C \) if \( u \in X_d \oplus u_0 \) or equivalently \( u_i = u \), hence the other symmetry condition (83) is satisfied as well. We have thus proved that \( \nu \) has a symmetric density w.r.t. \( u \) and \( u_i - u \) forms a Markov chain when the \( \lambda_j \)'s have densities in the form of (95). As both properties are preserved in convex combinations, they remain true when the \( \lambda_j \)'s have general symmetric densities.

Finally we prove that \( \nu \) is a sufficient statistic for \( u_i \), i.e. \( p(u_i \mid \nu) = p(u_i \mid \lambda_{-i}, u_{-i}) \) (note that the r.h.s. is equal to \( p(u_i \mid \nu, \lambda_{-i}, u_{-i}) \) because \( \nu \) is a function of \( \lambda_{-i} \)). This is where we need to use the uniformity of \( p(u) \) over \( C \), which implies that \( p(u_i) \) is also uniform over \( C_i \); under this condition, for any \( u_i \in C_i \),

\[
p(u_i \mid \nu) \propto p(u_i) \propto \nu(u_i)
\]

(155) and (156) use the symmetry of resp. \( \nu \) and \( \lambda_{-i} \)'s density, while (155) and (159) use the uniformity of \( u_i \) and \( u \overline{u} \) over respectively \( C_i \) and \( C \).

K. Proof of Proposition 30

The known Markov-chain relationships among the random variables can be expressed as

\[
\begin{array}{c c c}
\mathbf{u} & \rightarrow & \lambda_{-i} \\
| & | & | \\
\mathbf{u}_i & \rightarrow & \mathbf{u}_{-i}
\end{array}
\]

(162)

where every simple path in the graph forms a Markov chain. Therefore, we can formally write (the summations over \( \lambda_{-i} \) may represent integrals)

\[
p(\nu_i \mid \nu_i, u_i) = \sum_{\lambda_{-i}} p(\nu_i \mid \lambda_{-i}, \nu_i, u_i) = \sum_{\lambda_{-i}} p(\nu_i \mid \lambda_{-i}, \nu_i, u_i)p(\lambda_{-i} \mid \nu_i, u_i),
\]

(163)
where \( p(\nu_i' | \lambda_{-i}, \nu_i, u_i) = p(\nu_i' | \lambda_{-i}) \) is evident from the figure above, while \( p(\lambda_{-i} | \nu_i, u_i) = p(\lambda_{-i} | \nu_i) \) comes from Proposition \( 29 \). We have thus shown that \( p(\nu_i' | \nu_i, u_i) \) does not depend on the value of \( u_i \), making \( u_i - \nu_i - \nu_i' \) a Markov chain.

L. Proof of Proposition 31

As \( \psi_u(\cdot) \) is a group action, it is a bijection for any \( u \in \mathbb{G} \) and partitions \( \mathcal{Y} \) into orbits \( \mathcal{Y}' = \cup_{\alpha} \mathcal{Y}_{\alpha} \), where each orbit \( \mathcal{Y}_{\alpha} \) is a discrete set \( \{ \psi_u^{-1}(y_{\alpha}) | u \in \mathbb{G} \} \) for some deterministic \( y_{\alpha} \in \mathcal{Y} \).

We can first consider the case where \( \mathcal{Y} \) is a discrete set containing a single orbit \( \{ y_{\alpha} \triangleq \psi_u^{-1}(y_{\alpha}) | u \in \mathbb{G} \} \) for some \( y_{\alpha} \), such that the conditional pmf has the form \( p_{\gamma|u}(y | 0) = \sum_{u} p_{u'} \cdot 1 [y = y_{u'}] \) (with \( \sum_{u} p_{u'} = 1 \)), and by \( 95 \), \( p_{y|u} = \sum_{u} p_{u'} u_{\alpha} | 1 [y = y_{u'}] \). The use of \( 1 \) here allows for duplications among the \( y_{\alpha} \)'s; such duplications can be characterized by the stabilizer subgroup \( \mathbb{G} \) of the group action, which is the same over the entire orbit since \( \mathbb{G} \) is abelian. The normalized \( \lambda \) corresponding to a given \( y \) is then \( \lambda(u) = (1/|\mathbb{G}|) \sum_{u} p_{u'} \cdot 1 [y = y_{u'}] \), for each \( u \in \mathbb{G} \), when \( y = y_{u'} \) this \( \lambda \) is denoted \( \lambda_{u'} \). It is easy to see that \( \lambda_{u'}(y) = (1/|\mathbb{G}|) \sum_{u'} p_{u'} u_{\alpha} | 1 [y = y_{u'}] \), for any \( u' \in \mathbb{G} \) we have \( y_{u'} = y_{u'\oplus u} \) and thus \( \lambda_{u'} = \lambda_{u''} \).

Consequently,

\[
p(\lambda | u) = \sum_{u'} p_{u'} \cdot 1 [\lambda = \lambda_{u'}] = (1/|\mathbb{G}|) \sum_{u' \in \mathbb{G}} \sum_{u \in \mathbb{G}} p_{u'} u_{\alpha} | 1 [\lambda = \lambda_{u'}] = \sum_{u' \in \mathbb{G}} \lambda_{0}(u \oplus u') \cdot 1 [\lambda = \lambda_{u'}],
\]

which has the form of \( 95 \), so \( \lambda \) has a symmetric density w.r.t. \( u \).

For more general \( \mathcal{Y} \) and channel \( p(y | u) \) satisfying \( 98 \), we can let \( E_{\alpha} \) be the event that \( y \in \mathcal{Y}_{\alpha} \), and define \( p_{\alpha}(y | u) \triangleq p_{y|u}(y_{\alpha} | u) \) as the pmf conditioned on each \( E_{\alpha} \), so that \( p(y | u) \) can be viewed as a convex combination (or time-sharing) of channels \( p_{\alpha}(y | u) \), each with a discrete output alphabet \( \mathcal{Y}_{\alpha} \); here summation of \( 98 \) over \( y \in \mathcal{Y}_{\alpha} \) gives \( p(E_{\alpha} | u) = p(E_{\alpha} | 0) \) (both viewed as pdfs), so the required independence between \( E_{\alpha} \) and \( u \) is satisfied. For any \( y \in \mathcal{Y}_{\alpha} \), \( p_{\alpha}(y | u) = p(y | u)/p(E_{\alpha} | u) \) with \( p(E_{\alpha} | u) \) not varying with \( u \), so the \( \lambda \) computed from \( p(y | u) \) and from \( p_{\alpha}(y | u) \) are identical. By the above argument, each \( p_{\alpha}(\cdot | \cdot) \) yields a symmetric density for \( \lambda \), while the overall density of \( \lambda \) is a convex combination of these densities and thus symmetric as well.

M. Proof of Proposition 22

We only need to consider the case that \( p(\lambda | u) \) has the form of \( 95 \), i.e.

\[
p(\lambda | u) = \sum_{u'} \lambda^* (u \oplus u') \cdot 1 [\lambda = \lambda^* \oplus \overline{u'}].
\]

Transforming \( \lambda \) and \( u \) into \( \mu \triangleq \lambda \circ \phi \) and \( \overline{\psi} \triangleq \phi^{-1}(u) \), then they are still independent from \( e \), and

\[
p(\mu | \overline{\psi}) = \sum_{u'} \lambda^* (\phi(\overline{\psi}) \oplus u') \cdot 1 [\mu = (\lambda^* \oplus \overline{u'}) \circ \phi] = \sum_{u'} \mu_{u'}^* (\overline{\psi}) \cdot 1 [\mu = \mu_{u'}^*],
\]

where we have defined \( \mu_{u'}^* \triangleq (\lambda^* \oplus \overline{u'}) \circ \phi \). Eq. (166) shows that \( \mu \) does not necessarily have a symmetric density w.r.t. \( \overline{\psi} \), thus the necessity of \( e \). On the other hand, now \( \mu_{1} \triangleq \lambda \circ \phi_{1} = \mu \circ \overline{\psi} \) (here \( \mu_{1}, \mu \) and \( \overline{\psi} \) are probability tuples over \( \mathbb{G}_u \) and \( \overline{\psi}_{1} \triangleq \phi_{1}^{-1}(u) = \overline{\psi} \circ e \), and since \( \overline{\psi} \) is uniformly distributed over \( \mathbb{G}_u \), we also have \( p(\overline{\psi}) = p(\overline{\psi}_{1}) = 1/|\mathbb{G}_u| \), so

\[
p(\mu_{1}, \overline{\psi}_{1} | e) = p_{\mu}(\vec{u}, \overline{\psi}, \overline{\psi}_{1} \circ e) = \frac{1}{|\mathbb{G}_u|} \sum_{u'} \mu_{u'} (\overline{\psi}_{1} \circ e) \cdot 1 [\mu_{1} = \mu_{u'} \circ \overline{\psi}],
\]

and marginalizing over \( e \) yields

\[
p(\mu_{1} | \overline{\psi}_{1}) = \frac{1}{|\mathbb{G}_u|} \sum_{u'} \mu_{u'} (\overline{\psi}_{1}) \cdot 1 [\mu_{1} = \mu_{u'} \circ \overline{\psi}].
\]

The symmetry of \( \mu_{1} \) w.r.t. \( \overline{\psi}_{1} \) is now obvious, as each term in the summation over \( u' \) corresponds to a symmetric density in the form of 95, and the summation creates a convex combination of these densities.

N. Proof of Proposition 23

It is only necessary to consider the case that

\[
p(\mu | \overline{\psi}) = \sum_{\overline{\psi}'} \mu^* (\overline{\psi} \oplus \overline{\psi}') \cdot 1 [\mu = \mu^* \oplus \overline{\psi}].
\]

Now transform \( \mu \) and \( \overline{\psi} \) into respectively \( \lambda \triangleq \mu \circ \phi^{-1} \) and \( u \triangleq \phi(\overline{\psi}) \) such that \( \lambda_{1} \triangleq \mu \circ \phi_{1}^{-1} = \lambda \circ \overline{\psi} \) and \( u_{1} \triangleq \phi(\overline{\psi}_{1}) = u \circ \delta \). \( \lambda \) and \( u \) thus remain independent from \( \delta \), with

\[
p(\lambda | u) = \sum_{\overline{\psi}'} \mu^* (\phi^{-1}(u) \oplus \overline{\psi}') \cdot 1 [\lambda = (\mu^* \oplus \overline{\psi}') \circ \phi_{1}^{-1}] = \sum_{\overline{\psi}'} \lambda_{\overline{\psi}'}^* (u) \cdot 1 [\lambda = \lambda_{\overline{\psi}'}],
\]

where \( \lambda_{\overline{\psi}'}^* \triangleq (\mu^* \oplus \overline{\psi}') \circ \phi_{1}^{-1} \). Since \( \overline{\psi} \) is uniformly distributed over \( \mathbb{G}_u \), we have \( p(u) = p(u_{1}) = 1/|\mathbb{G}_u| \), so

\[
p(\lambda_{1}, u_{1} | \delta) = p_{\lambda_{1}} (\lambda_{1} \oplus \overline{\psi}_{1}, u_{1} \circ \delta) = \frac{1}{|\mathbb{G}_u|} \sum_{\overline{\psi}'} \lambda_{\overline{\psi}'}^* (u_{1} \circ \delta) \cdot 1 [\lambda_{1} = \lambda_{\overline{\psi}'}],
\]

and marginalizing over \( \delta \) yields

\[
p(\lambda_{1} | u_{1}) = \frac{1}{|\mathbb{G}_u|} \sum_{\overline{\psi}'} \sum_{\delta} \lambda_{\overline{\psi}'}^* (u_{1} \circ \delta) \cdot 1 [\lambda_{1} = \lambda_{\overline{\psi}'}],
\]

which is a convex combination of symmetric densities and thus symmetric.
