Capacity-Achieving Codes for Channels with Memory with Maximum-Likelihood Decoding

Jung Hyun Bae and Achilleas Anastasopoulos
Electrical Engineering and Computer Science Department
University of Michigan
Ann Arbor, MI, 48109-2122
email: {jungbae, anastas}@umich.edu

Abstract

Codes on sparse graphs have been shown to achieve remarkable performance in point-to-point channels with low decoding complexity. Most of the results in this area are based on experimental evidence and/or approximate analysis. The question of whether codes on sparse graphs can achieve the capacity of noisy channels with iterative decoding is still open, and has only been conclusively and positively answered for the binary erasure channel. On the other hand, codes on sparse graphs have been proven to achieve the capacity of memoryless, binary-input, output-symmetric channels with finite graphical complexity per information bit when maximum likelihood (ML) decoding is performed.

In this paper, we consider transmission over finite-state channels (FSCs). We derive upper bounds on the average error probability of code ensembles with ML decoding. Based on these bounds we show that codes on sparse graphs can achieve the symmetric information rate (SIR) of FSCs, which is the maximum achievable rate with independently and uniformly distributed input sequences. In order to achieve rates beyond the SIR, we consider a simple quantization scheme that when applied to ensembles of codes on sparse graphs induces a Markov distribution on the transmitted sequence. By deriving average error probability bounds for these quantized code ensembles, we prove that they can achieve the information rates corresponding to the induced Markov distribution, and thus approach the FSC capacity.

I. Introduction

The capacity and the capacity-achieving input distribution of memoryless, binary-input, output-symmetric(MBIOS) channels, is well known [1, p. 94]. It is also well known that linear codes can achieve the capacity of MBIOS channels [2]. However, capacity achieving codes with low complexity were not known until the remarkable discovery of turbo codes in 1993 [3] and the subsequent rediscovery of low-density, parity-check (LDPC) codes [4]–[7]. Since then, codes defined on sparse graphs, collectively referred to as "turbo-like" or "LDPC-like" codes have attracted a large amount of attention in the quest to achieve channel capacity with small complexity.

For the special case of the binary erasure channel (BEC), it is known that capacity can be achieved with bounded decoding complexity per information bit. In particular, based on the density evolution (DE) technique, it was shown in [8]–[10] that LDPC-like codes achieve the BEC capacity with iterative decoding, while [11] showed that this can be done with bounded (and small) decoding complexity per information bit.

For general MBIOS channels, however, there is no conclusive answer regarding capacity achievability of turbo-like codes with iterative decoding. This is so due to the fact that DE is not amenable to theoretical analysis for MBIOS channels, since code performance can only be evaluated through an uncountably-infinite-dimensional non-linear recursive equation (there is a line of research studying iterative decoding performance [12]–[14] based on results from statistical physics). Thus, only numerical results exist that show how turbo-like codes can approach capacity using iterative decoding [15]. On the other hand, when maximum-likelihood (ML) decoding is assumed, the performance of turbo-like codes can be analyzed. For instance, it was shown that LDPC-like codes achieve the capacity of MBIOS channels with ML decoding in [16]–[19]. Moreover, it was shown in [20] that a family of turbo-like codes consisting of a serial concatenation of an outer LDPC code and an inner rate-one low-density, generator matrix (LDGM) code, can achieve the capacity of MBIOS channels with bounded graphical complexity (i.e., number of edges in the graph representing the code) per information bit when ML decoding is performed.

The aforementioned results were developed for memoryless channels. For channels with memory, few results exist on their capacity and the corresponding capacity-achieving input distribution, and even fewer on capacity-achieving codes. Regarding the former, the capacity of Gilbert-Elliott (GE) channels was studied in [21] and the capacity of more general finite-state Markov channels (FSMCs) was studied in [22], [23]. An efficient method for computing the information rate of a finite-state channel (FSC) whose input is a Markov process was proposed independently in [24]–[26] and extended in [27]. In [28] an optimization algorithm was proposed to numerically compute tight lower bounds on the capacity of FSCs, and the techniques were generalized in [29]. Upper bounds on the capacity of FSCs were developed in [30], [31]. The above results, demonstrated that a Markov input sequence provides larger information rate than the information rate achievable with independent and uniformly distributed (i.u.d.) input sequences, also called the symmetric information rate (SIR) of the channel. Furthermore, it was recently shown in [32] that a sequence of Markov sources asymptotically achieves the capacity of FSCs as the Markov order approaches infinity.

Regarding capacity-achieving codes for channels with memory, methods of constructing codes which induce Markov distribution on the transmitted sequence were proposed in [33]–[36], and their performance on partial response channels was evaluated using simulation. Pfister and Siegel [37] introduced a generalized erasure channel (GEC) as a simple model of a channel with memory and proved that LDPC-like codes achieve its SIR. Several authors have investigated the performance of LDPC-like codes with iterative decoding using numerical tools based on DE for some specific channels with memory. In particular, LDPC-like codes were analyzed using numerical tools based on DE for binary ISI channels in [38], for the GE channel in [39], and for FSMCs in [40]. Finally, the results derived for MBIOS channels in [17], were extended to establish an upper bound on the rate of LDPC codes for GE channels [41] and FSMCs [42].

In this paper, capacity-achieving codes are constructed for FSCs (with or without ISI) with ML decoding. The codes are derived from the corresponding capacity-achieving codes for MBIOS channels through simple modifications. In particular, it is shown that several LDPC-like coset codes which are capacity-achieving on MBIOS channels achieve the SIR of FSCs. Next, a family of *quantized coset codes* is constructed by using block-wise Markov quantization of LDPC-like codes. This technique generalizes the quantization technique presented in [43] for memoryless channels and results in a simple encoding and iterative decoding algorithm. The constructed quantized codes induce a k-th order Markov distribution on the channel input sequence and are shown to achieve the corresponding information rate. The basic analytical tool used to prove these results is an upper bound on the ensemble average of the corresponding modified LDPC-like ensembles. This bound is a non-trivial generalization of the "union plus Shulman and Feder" bound [16], [43], [44] which relates the average code error probability with the asymptotic growth rate of the average code weight enumerator, the capacity of channels with memory (as opposed to the corresponding channel error exponent) and an appropriately defined Battacharrya-like parameter.

The remainder of the paper is organized as follows. Definitions, the channel model, and some preliminary facts are presented in section II. In section III an upper bound on the average block error probability of LDPC-like coset codes transmitted over an FSC is derived. This bound is used to show how specific LDPC-like coset ensembles achieve the SIR of FSCs. In section IV construction of Markov-quantized coset codes is presented and an upper bound on their average block error probability over FSCs is derived. This bound is then applied to specific Markov-quantized LDPC-like coset ensembles. Section V concludes the paper.

II. Channel model and preliminaries

Let $\{S_n\}_{n=1}^{\infty}$ with $s_n \in \mathcal{S} = \{1, 2, ..., K\}$ be a sequence representing the channel states at time n. It is assumed that the state space $\mathcal{S}$ corresponds to K different MBIOS channels. Let $\{X_n\}_{n=1}^{\infty}$ be the random process representing the channel input sequence, where $x_n \in \mathcal{X} = \{0, 1\}$. Let $\{Y_n\}_{n=1}^{\infty}$ be the random process representing the channel output sequence, where $y_n \in \mathcal{Y}$ and $\mathcal{Y}$ is assumed to be a discrete subset of the real line (a continuous subset can also be assumed throughout the paper by changing summations over $\mathcal{Y}$ to integrations) symmetric around zero. Since for each state the channels are symmetric, it is true that $Q(y_n|x_n, s_n) = Q(-y_n|x_n \oplus 1, s_n)$, where $\oplus$ denotes modulo-2 addition. Let $P(\mathbf{x}^N, \mathbf{y}^N)$ be the joint probability mass function (pmf) of $\mathbf{X}^N$ and $\mathbf{Y}^N$, where $\mathbf{x}^N$ denotes the length-N vector $(x_1, \ldots, x_N)$. Then,

$$P(\mathbf{x}^N, \mathbf{s}^N, \mathbf{y}^N) = Q(\mathbf{y}^N|\mathbf{x}^N, \mathbf{s}^N) P(\mathbf{s}^N|\mathbf{x}^N) P(\mathbf{x}^N) \tag{1a}$$

$$= P(\mathbf{x}^N) P(s_1) \prod_{n=1}^{N} Q(y_n|x_n, s_n) \prod_{n=1}^{N-1} P(s_{n+1}|s_n, x_n). \tag{1b}$$

Implicit in (1), is the fact that the considered channel states are affected by both nature and ISI. In this paper, we also consider non-inverting channels, that is, channels with the property

$$\forall s \in \mathcal{S}, \qquad \sum_{y>0} Q(y|1, s) > \sum_{y<0} Q(y|1, s). \tag{2}$$

For any sequence of real-valued random variables $(Z_1, Z_2, Z_3, ...)$, define the *limit inferior in probability* $p - \liminf_{N\to\infty} Z_N$ as

$$p - \liminf_{N\to\infty} Z_N \triangleq \sup\{\alpha| \lim_{N\to\infty} Pr[Z_N < \alpha] = 0\}. \tag{3}$$

Then, the capacity of the aforementioned FSC when no channel state information is available at the transmitter and the receiver is defined as [45, sec. 3.2]

$$C \triangleq \sup_{\mathbf{X}} \underline{I}(\mathbf{X};\mathbf{Y}), \tag{4}$$

where

$$\underline{I}(\mathbf{X};\mathbf{Y}) = p - \liminf_{N\to\infty} \frac{1}{N} \log_2 \frac{Q(\mathbf{y}^N|\mathbf{x}^N)}{P(\mathbf{y}^N)}. \tag{5}$$

Since the focus of this paper is not in finding this maximizing input distribution, in the following we will be interested in the maximum achievable rate for a certain sequence of input pmfs $P \triangleq \{P(\mathbf{x}^N)\}_N$, defined as

$$C_P \triangleq \underline{I}(\mathbf{X};\mathbf{Y}), \tag{6}$$

where $I(\mathbf{X};\mathbf{Y})$ is computed under $\{P(\mathbf{x}^N)\}_N$ using (1). The maximum achievable rate relative to independent and uniformly distributed (i.u.d.) inputs is also known as SIR (for binary input) and will be denoted as

$$SIR \triangleq C_{IUD}. \tag{7}$$

The achievability of C_P can be shown by the following coding theorem which is a modification of channel coding theorem using information spectrum methods [45, Th. 3.2.1] according to the idea in [44].

Proposition 1. *Consider an arbitrary discrete channel. Let $Q(\mathbf{y}^N|\mathbf{x}^N)$ be the conditional pmf for sequences of length $N \geq 1$ on this channel. Let $P(\mathbf{x}^N)$ be an arbitrary input pmf. Consider an ensemble of codes of size M, length N, and rate R, whose codewords $\mathbf{c}_m$, $1 \leq m \leq M$ satisfy the following properties*

$$Pr[\mathbf{c}_i = \mathbf{x}^N] = P(\mathbf{x}^N) \qquad \forall i \in \{1,\ldots,M\}, \tag{8}$$

$$Pr[\mathbf{c}_i = \mathbf{x}^N|\mathbf{c}_j = \mathbf{x}'^N] \leq \alpha P(\mathbf{x}^N) \qquad \forall i,j \in \{1,\ldots,M\} \text{ with } i \neq j \text{ and } d(\mathbf{c}_i,\mathbf{c}_j) \in U^c, \tag{9}$$

where $U \subseteq \{1,2,...,N\}$. Note that $\alpha = 1$ for an ensemble of codes whose codewords are selected independently with pmf $P(\mathbf{x}^N)$. Suppose that an arbitrary message m, $1 \leq m \leq M$ enters the encoder and that ML decoding is employed. Then for any input pmf $P = P(\mathbf{x}^N)$ and for any $\epsilon > 0$, the average probability of decoding error over this ensemble of codes, is bounded *as*

$$\overline{P}_{e|m} \leq \overline{P}^{U}_{e|m} + Pr[\mathbf{X}^N\mathbf{Y}^N \notin T_N] + 2^{-N(C_P - R - \frac{\log_2 \alpha}{N} - \epsilon)}, \tag{10a}$$

where

$$\overline{P}^{U}_{e|m} = \sum_{C} Pr[C] P^{U}_{e|m}, \tag{10b}$$

$P^{U}_{e|m}$ *denote the probability that there exists some codeword* $\mathbf{c}_{m'}$ *such that* $Q(\mathbf{y}|\mathbf{c}_{m'}) \geq Q(\mathbf{y}|\mathbf{c}_m)$ *and* $d(\mathbf{c}_m, \mathbf{c}_{m'}) \in U$,

$$T_N = \left\{ (\mathbf{x}^N, \mathbf{y}^N) \in \mathcal{X}^N \times \mathcal{Y}^N | \frac{1}{N} \log_2 \frac{Q(\mathbf{y}^N|\mathbf{x}^N)}{P(\mathbf{y}^N)} > C_P - \epsilon \right\}. \tag{10c}$$

Proof: See Appendix A. ■

Note that the presence of the factor α in (9) allows one to deal with structured ensembles for which strict independence of the codeword selection cannot be guaranteed, as long as one shows that the rate loss due to α as manifested in (10a) approaches zero asymptotically for long codes. In the following section we give specific expressions for α for several structured ensembles of interest.

We now summarize the bounding techniques developed in the literature for showing that certain turbo-like ensembles are capacity achieving for MBIOS channels. As it will become evident in the following sections, the bounds derived for the FSCs are similar in structure to their MBIOS counterparts. An upper-bound on the error probability of linear codes transmitted over MBIOS channels was derived in [16]. The bound involves two terms; the first term is based on the Shulman and Feder bound [44] for MBIOS channels and the second term is based on a union bound which involves the Battacharrya parameter for MBIOS channels. For a given ensemble $\mathcal{C}$ of codes with length N, we denote the average weight enumerator of the ensemble by by $\overline{A}_l$, $l = 0, 1, \ldots, N$. The error probability bound is presented in the following fact.

Fact 1. *Consider an ensemble* $\mathcal{C}$ *of linear codes with average weight enumerator* $\overline{A}_l$*, where each code is comprised of* M *codewords of length* N*. Let* $Q(y|x)$ *be an MBIOS channel, and let*

$$D_0 \triangleq \sum_{y} \sqrt{Q(y|0)Q(y|1)} \tag{11}$$

be the Battacharrya channel parameter. Denote the ensemble averaged ML decoding error *probability by* $\overline{P_e}$. *Let* $U \subseteq \{1, 2, ..., N\}$. *Then*

$$\overline{P}_e \leq \sum_{l \in U} \overline{A}_l D_0^l + 2^{-NE_r(R + \frac{\log_2 \alpha}{N})}, \tag{12}$$

where

$$\alpha \triangleq \max_{l \in U^c} \frac{\overline{A}_l}{M-1} \frac{2^N}{\binom{N}{l}}, \tag{13}$$

and where$E_r(\cdot)$ *is the random coding error exponent*

$$E_r(R) \triangleq \max_P \max_{0 \leq s \leq 1} \left\{ -sR + E_0(s, P) \right\}, \tag{14a}$$

$$E_0(s, P) \triangleq -\log_2 \left\{ \sum_y \left\{ \sum_x P(x) Q(y|x)^{\frac{1}{1+s}} \right\}^{1+s} \right\}. \tag{14b}$$

As discussed in [16], the second part of (12) approaches $2^{-NE_r(R)}$ provided that the asymptotic growth rate of $\overline{A}_l$ approaches the asymptotic growth rate, $H(l/N)+R-1$, of the random ensemble for $l \in U^c$. Therefore, in order to prove that a specific sequence of ensembles approaches capacity, we have to choose U such that the above statement is true and the first part of (12) approaches zero. Consider the following three ensembles with increasing degree of sophistication.

Ensemble A is the regular Gallager (N, d_v, d_c) LDPC ensemble [46] with column and row degrees d_v and d_c, respectively (we also assume an even d_c).

Ensemble B is the punctured LDPC ensemble introduced in [19] resulting from puncturing a sufficiently low-rate (N, d_v, d_c) Gallager ensemble to achieve the specified rate.

Ensemble C is the LDPC-GM ensemble introduced in [20] consisting of a a serial concatenation of an outer (N, d_v, d_c) Gallager LDPC code and an inner rate-one regular LDGM code with row and column degrees equal to d_c.

For all these three ensembles, we can find an appropriate set U and prove capacity achievability of MBIOS channels [19], [20] as can be seen in the following fact.

Fact 2. *For any* $\epsilon > 0$*, the sequence of the aforementioned ensembles A, B, and C, have vanishing average block error probability under ML decoding on MBIOS channels, and limiting (with respect to* N*) rate* $(1-\epsilon)C$*, where* C *is the channel capacity, when the following conditions are satisfied.*

Ensemble A: $d_c \geq d_c(\epsilon)$.

Ensemble B: *$d_c > d_v \geq 5$, d_c even, and $R_0 \leq R_0(\epsilon)$, where R_0 is the rate of the original code before puncturing.*

Ensemble C: *$d_c \geq d_{c,min}$, d_c even, and $d_v \geq 4$, where $d_{c,min}$ is independent of ϵ and bounded.*

Proof: The dependence of the upper bound in Fact 1 on the code ensemble is only through $\overline{A}_l$ and in particular only through the corresponding asymptotic growth rate. In [19], [20] upper bounds on these asymptotic growth rates have been developed for the ensembles A, B, and C. Combining Fact 1 with these upper bounds, and with an appropriate choice of U as in [16], [19], [20], proves the fact under the conditions mentioned above.

We present the choice of U, for ensemble A as an example. The choices for ensembles B and C are similar to that of the ensemble A. It was shown in [19] that $\overline{A}_l$ of the ensemble A satisfies

$$\begin{aligned} w_0(a) \leq{}& (1-R)\log_2[1+(1-2a)^{d_c}] \\ &+ [H(a)-(1-R)], \qquad 0 \leq a \leq 1, \end{aligned} \tag{15}$$

where $w_0(a) \triangleq \lim_{N\to\infty} \frac{1}{N} \log_2 \overline{A}_{\lceil Na \rceil}$, and $R = 1 - d_v/d_c$. Furthermore, there exists a $\delta_0 \in (0, 1/2)$ such that

$$\sum_{l \in (0, N\delta_0)} \overline{A}_l = O(N^{-d_v+2}). \tag{16}$$

Finally, when d_c is even, $\overline{A}_l = \overline{A}_{N-l}$, for all $l \in \{0, 1, 2, ..., N\}$.

Let

$$U \triangleq \left\{ l : \frac{l}{N} \in (0, \delta_0) \cup (1-\delta_0, 1] \right\}. \tag{17}$$

With this choice of U, the union bound term becomes

$$\begin{aligned} \sum_{l \in U} \overline{A}_l D_0^l &= \sum_{l \in U \setminus \{N\}} \overline{A}_l D_0^l + \overline{A}_N D_0^N \\ &\leq \sum_{l \in U \setminus \{N\}} \overline{A}_l + 1 \times D_0^N = O(N^{-d_v+2}), \end{aligned} \tag{18}$$

and the rate loss in the error exponent is bounded above by $\log_2[1+(1-2\delta_0)^{d_c}]$. As a result, for $d_v \geq 3$ and $\log_2[1+(1-2\delta_0)^{d_c}] \leq \epsilon C$, vanishing error probability is ensured for all rates up to $(1-\epsilon)C$. The last inequality is guaranteed by selecting a large enough $d_c \geq d_c(\epsilon) \stackrel{\text{def}}{=} \log_2(2^{\epsilon C} - 1)/\log_2(1-2\delta_0)$. ■

III. SIR-ACHIEVING CODES FOR FSCs

In order to analyze the performance of code ensembles on FSCs, we first derive an upper-bound on the error probability of linear coset codes transmitted over FSCs, which is similar to the bound in Fact 1. Since we are interested in turbo-like ensembles that achieve the SIR of FSCs, and since linear codes do not necessarily induce an i.u.d. input, we need to enlarge linear code ensembles in a way similar to [38]. This motivates the following definition.

Definition 1. *Consider a code ensemble $\mathcal{C}$. The coset ensemble $\mathcal{C}'$ generated by $\mathcal{C}$ is defined as the ensemble generated from $\mathcal{C}$ by including, for each code $C \in \mathcal{C}$, codes of the form $C' = \{\mathbf{c} \oplus \mathbf{v} | \mathbf{c} \in C\}$ for all possible vectors $\mathbf{v}$. The measure on $\mathcal{C}'$ is the product of the uniform measure over all vectors $\mathbf{v}$ and the measure of the original ensemble $\mathcal{C}$.*

In the following lemma, we establish an upper-bound for the pairwise error probability between two sequences. This will be needed for the union-bound term of the aforementioned bounding technique. We denote by $P_{e|\mathbf{x}^N,\mathbf{x}'^N}$ the pairwise error probability of decoding the sequence $\mathbf{x}'^N$ with ML decoding on a FSC conditioned on $\mathbf{x}^N$ being transmitted.

Lemma 1. *The pairwise error probability $P_{e|\mathbf{x}^N,\mathbf{x}'^N}$ is upper-bounded as*

$$P_{e|\mathbf{x}^N,\mathbf{x}'^N} \le D^{d(\mathbf{x}^N,\mathbf{x}'^N)}, \tag{19}$$

where $d(\mathbf{x}^N,\mathbf{x}'^N)$ denotes the Hamming distance between $\mathbf{x}^N$ and $\mathbf{x}'^N$, and

$$D \triangleq \min_{s_0} \max_{s} \left[Q_s^+ \sqrt{\frac{Q_{s_0}^-}{Q_{s_0}^+}} + Q_s^- \sqrt{\frac{Q_{s_0}^+}{Q_{s_0}^-}} \right], \tag{20}$$

where

$$Q_s^+ = \frac{1}{2} Q(0|1,s) + \sum_{y>0} Q(y|1,s), \tag{21a}$$

$$Q_s^- = \frac{1}{2} Q(0|1,s) + \sum_{y<0} Q(y|1,s). \tag{21b}$$

Furthermore, $D < 1$.

Proof: See Appendix B. ■

In order to utilize the techniques in [44] and [16] for bounding the error probability, a special kind of linear code ensemble is considered. Let Π_N denote the set of all permutations of N numbers. We define a *permutation-invariant* code ensemble as follows.

Definition 2. *Let $\mathcal{C}$ be an ensemble of length-N block codes. We say that $\mathcal{C}$ is a* permutation-invariant *ensemble if for all permutations $\pi \in \Pi_N$ and for all codes $C \in \mathcal{C}$, it is true that $\pi(C) \in \mathcal{C}$ and the codes $\pi(C)$ are selected with the same probability. Here $\pi(C)$ denotes the codebook constructed by permuting the order of the symbols in all the codewords of C according to π.*

We are now ready to establish an upper bound on the average error probability of permutation-invariant coset ensembles on FSCs.

Proposition 2. *Consider a permutation-invariant ensemble $\mathcal{C}$ of binary linear codes with M codewords of length N, rate R, and average weight enumerator $\overline{A}_l$. Consider the coset ensemble $\mathcal{C}'$ generated by $\mathcal{C}$. Let $U \subseteq \{1, 2, ..., N\}$. Then, for any $\epsilon > 0$, the average (over $\mathcal{C}'$) error probability with ML decoding given the mth message is transmitted is upper-bounded as*

$$\overline{P}_{e|m} \leq \sum_{l \in U} \overline{A}_l D^l + Pr[\mathbf{X}^N \mathbf{Y}^N \notin T_N] + 2^{-N(C_{IUD} - R - \frac{\log_2 \alpha}{N} - \epsilon)}, \tag{22}$$

where D is as defined in Lemma 1 and T_N is as defined in Proposition 1,

$$\alpha \triangleq \max_{l \in U^c} \frac{\overline{A}_l}{M-1} \frac{2^N}{\binom{N}{l}}, \tag{23}$$

Proof: Due to space limitations, the proof of this proposition is omitted. This is a simplified version of the proof of Proposition 3 that follows. ∎

Since the aforementioned ensembles A, B and C are permutation-invariant, Proposition 2 can be applied for those ensembles to prove the SIR-achievability which is stated as the following corollary.

Corollary 1. *For any $\epsilon > 0$, there exists a sequence of the coset ensembles generated by the ensembles A, B, and C, have limiting (with respect to N) rate $(1-\epsilon)C_{IUD}$ and vanishing average block error probability under ML decoding on FSCs when the conditions stated in Fact 2 are satisfied.*

Proof: From the definition of T_N, we have $Pr[\mathbf{X}^N \mathbf{Y}^N \notin T_N] \to 0$ as $N \to \infty$. Since the bound in Fact 1 and the first and the third term of the bound in Proposition 2 have a similar form, the same technique used in Fact 2 can also be used to prove SIR-achievability for the sequence of ensembles A, B, and C satisfying Proposition 2. ■

Some observations regarding the complexity of the aforementioned SIR-achieving ensembles are in order. Regarding ensemble A, the quantity $d_c(\epsilon)$ is diverging to ∞ as $\log(1/\epsilon)$ when ϵ approaches zero. Therefore, the density of the parity check matrix, and thus the number of edges in the graph per information bit approaches ∞, which implies an infinite complexity per information bit per iteration even when iterative decoding is applied. A similar conclusion can be reached for ensemble B, as detailed in [19]. The advantage of B over A is the universality of the former, i.e., a single mother code of low rate can be used to approach a wide range of channel SIRs for different FSCs. Finally, ensemble C achieves the SIR with bounded number of edges per information bit in the graph representing the code and thus it can be decoded with finite complexity per information bit per iteration when iterative decoding is performed. It should be noted however, that SIR-achievability is only guaranteed for ML decoding of these ensembles.

We also note that if we concentrate only on ensemble C, neither the symmetric channel assumption nor the non-inverting channel assumption is required for the proof of SIR achievability. This is so because these two assumptions only enter our arguments through Lemma 1. However, the first term of (22) can always be substituted by $\sum_{l\in U} \bar{A}_l$, which makes the overall upper bound independent of the Battacharrya parameter. Moreover, for ensemble C, for all $l \in U$ the asymptotic growth rate of the number of codewords with weight l is negative [20], and thus the above sum is converging to zero.

We finally point out that the three examples mentioned above are only examples of possible SIR-achieving ensembles. Other ensembles may also possess this property and may also possess additional properties that make them more desirable in practical applications.

IV. CAPACITY-ACHIEVING CODES ON FSCS

In section III, we proposed several LDPC-like coset code ensembles that achieve the SIR of FSCs. For general FSCs, however, the capacity could be greater than the SIR. Motivated by the need to achieve rates above the SIR, in this section we propose a simple quantization technique that induces a Markov distribution on the transmitted sequence and analyze its performance.

A. Construction of quantized coset code ensembles

Bennatan and Burshtein [43] presented a method of constructing codes for transmission over arbitrary memoryless channels by using a linear code followed by a simple memoryless quantization technique. Since memoryless quantization can only induce an i.i.d. (not necessarily uniform) on the input sequence, and since the capacity achieving input might not be i.i.d. for FSCs, we present a modified quantization technique that can induce a k-th order (stationary) Markov distribution on the input sequence. The block diagram of the proposed scheme is shown in Fig. 1. Other methods of constructing codes which induce Markov distribution on the transmitted sequence can be found in [34], [35].

Definition 3. *Consider a sequence $\mathbf{x}^{NT}$ and some arbitrary function $f:\{0,1\}^T\times\{0,1\}^k\rightarrow\{0,1\}$. An order-$k$ Markov quantizer (denoted by MQ-k) is a mapping $\delta:\{0,1\}^{NT}\rightarrow\{0,1\}^N$ with $\delta(\mathbf{x}^{NT})=\mathbf{w}^N$, with the following structure*[1]

$$w_n=f(\mathbf{x}_{(n-1)T+1}^{nT},\mathbf{w}_{n-k}^{n-1}),\qquad n=1,2,\ldots,N. \tag{24}$$

Consider now a pmf of a k-th order stationary Markov process $P(\mathbf{w}^N)=\prod_{n=1}^N P(w_n|\mathbf{w}_{n-k}^{n-1})$ for a binary sequence of length N. An order-k Markov quantizer with respect to P (denoted by MQ-k-P) is an MQ-k satisfying

$$\frac{|\{\mathbf{x}_{(n-1)T+1}^{nT}|f(\mathbf{x}_{(n-1)T+1}^{nT},\mathbf{w}_{n-k}^{n-1})=0\}|}{2^T}=P\left(0\mid\mathbf{w}_{n-k}^{n-1}\right),\qquad \forall\mathbf{w}_{n-k}^{n-1}\text{ and }n=1,2,\ldots,N \tag{25}$$

More descriptively, an MQ$-k$ partitions a length-NT binary sequence into N blocks of length T each, and then quantizes each block into a bit using a mapping that depends also on the k previously produced bits, thus producing a length-N binary sequence. This is shown in Fig. 2. Furthermore, if the input sequence to an MQ-k-P is i.u.d., then the quantizer performs quantization in a way that induces the pmf P on the transmitted sequence. Note that MQ-k-P does not exist for some pmfs, since there is a granularity of 2^{-T} in the quantization process.

In the following lemma we establish that if a codeword drawn from a coset ensemble is the input to an MQ-k-P, then the induced pmf on the transmitted sequence $\mathbf{w}$ is indeed P.

[1] For notational simplicity we do not specify precisely the quantizer for the first k symbols. These edge effects will be negligible for large N. In the following we use $\mathbf{x}_i^j$ to denote the subsequence $(x_i,x_{i+1}\ldots,x_j)$.

Lemma 2. *Consider a length-NT coset ensemble $\mathcal{C}'$. A code C' is picked from the ensemble and a codeword $\mathbf{c}' \in C'$ is picked uniformly from this code. Let P be a pmf of an arbitrary kth order Markov process for a binary sequence of length N for which an MQ-k-P exists. The codeword $\mathbf{c}'$ is quantized into $\mathbf{w}^N = \delta(\mathbf{c}')$ where $\delta(\cdot)$ is an MQ-k-P. Let $\hat{P}$ be the induced pmf on the transmitted sequence $\mathbf{w}^N$. Then, $P = \hat{P}$.*

Proof: Since $\mathcal{C}'$ is generated by adding a uniformly selected random vector to all codewords of all codes of $\mathcal{C}$ (the original ensemble from which $\mathcal{C}'$ is generated), we have

$$Pr[\mathbf{c}' = \mathbf{a}^{NT}] = 2^{-NT} \qquad \text{for all } \mathbf{a}^{NT} \tag{26a}$$

$$Pr[\mathbf{c}'^{iT}_{(i-1)T+1} = \mathbf{a}^T | \mathbf{w}^{i-1}_{i-j} = \mathbf{b}^T] = 2^{-T} \qquad \text{for all } \mathbf{a}^T, \mathbf{b}^T, \text{ and } 1 \leq j \leq i \leq N. \tag{26b}$$

Therefore,

$$\begin{aligned}
\hat{P}(\mathbf{w}^N) &= \prod_{i=1}^{N} \hat{P}(w_i | \mathbf{w}^{i-1}) && (27a)\\
&= \prod_{i=1}^{N} \sum_{\mathbf{a}^T} Pr[w_i, \mathbf{c}'^{iT}_{(i-1)T+1} = \mathbf{a}^T | \mathbf{w}^{i-1}] && (27b)\\
&= \prod_{i=1}^{N} \sum_{\mathbf{a}^T} Pr[w_i | \mathbf{c}'^{iT}_{(i-1)T+1} = \mathbf{a}^T, \mathbf{w}^{i-1}] Pr[\mathbf{c}'^{iT}_{(i-1)T+1} = \mathbf{a}^T | \mathbf{w}^{i-1}] && (27c)\\
&= \prod_{i=1}^{N} \sum_{\mathbf{a}^T} Pr[w_i | \mathbf{c}'^{iT}_{(i-1)T+1} = \mathbf{a}^T, \mathbf{w}^{i-1}] 2^{-T} && (27d)\\
&= \prod_{i=1}^{N} \sum_{\mathbf{a}^T} Pr[w_i | \mathbf{c}'^{iT}_{(i-1)T+1} = \mathbf{a}^T, \mathbf{w}^{i-1}_{i-k}] 2^{-T} && (27e)\\
&= \prod_{i=1}^{N} \sum_{\mathbf{a}^T} I\left\{ f(\mathbf{a}^T, \mathbf{w}^{i-1}_{i-k}) = w_i \right\} 2^{-T} && (27f)\\
&= \prod_{i=1}^{N} P(w_i | \mathbf{w}^{i-1}_{i-k}) && (27g)\\
&= P(\mathbf{w}^N), && (27h)
\end{aligned}$$

where (27d) is due to (26b), (27e) and (27f) is due to the definition of a quantizer (Definition 3), and (27g) is due to (25). Here $I\{\cdot\}$ denotes the indicator function of its argument. ■

B. Analysis of Markov-quantized coset code ensembles

In this section, we derive an upper-bound on the error probability of quantized linear coset codes over FSCs. First, we establish an upper-bound on the pairwise error probability between two sequences which is similar to Lemma 1.

Definition 4. *An MQ-k is called "robust" if the all-zeros block of length T, $\mathbf{0}^T$, and the all-ones block of length T, $\mathbf{1}^T$, are quantized to different values regardless of the quantizer memory, i.e.,*

$$\forall \mathbf{w}^k \in \{0,1\}^k : \quad f(\mathbf{0}^T, \mathbf{w}^k) \neq f(\mathbf{1}^T, \mathbf{w}^k). \tag{28}$$

This property will be used below to establish a pairwise error probability upper bound between two codewords having Hamming distance NT before quantization.

Lemma 3. *Consider a code C. Let $\mathcal{C}'$ be an ensemble which consists of codes of the form $C' = \{\mathbf{c} \oplus \mathbf{v} | \mathbf{c} \in C\}$ for all uniformly selected vectors $\mathbf{v} \in \{0,1\}^{NT}$. A code $C' \in \mathcal{C}'$ is quantized using a "robust" MQ-k-P before transmission. Then, the ensemble-averaged pairwise ML decoding error probability $\overline{P}_{e|m,m'}$ of decoding message m' when m is transmitted on FSC when $d(\mathbf{c}_m, \mathbf{c}_{m'}) = NT$ is upper-bounded as*

$$\overline{P}_{e|m,m'} \leq D_1^N, \tag{29}$$

with

$$D_1 \triangleq \frac{2^T - 1 + D}{2^T}, \tag{30}$$

and D as defined in Lemma 1.

Furthermore, $D_1 < 1$.

Proof: See Appendix C. ■

It is noted that the above Lemma is more specialized than Lemma 1 in that it only establishes a bound for a pair of maximally separated codewords. As shown below, this is sufficient for our purpose of establishing capacity-achievability for the quantized ensembles A, B, C defined earlier. However, it might not be sufficient to prove capacity achievability for other quantized ensembles. As it turns out, this bound can be generalized to non-maximally separated codewords. However, due to space limitations we do not present this more general result.

We now state an error probability upper bound for Markov-quantized ensembles.

Proposition 3. *Consider a permutation-invariant ensemble $\mathcal{C}$ of binary linear codes with M codewords of length NT, rate R/T, and average weight enumerator $\overline{A}_l$. Let $\mathcal{C}'$ be the coset ensemble generated by $\mathcal{C}$. A code from $\mathcal{C}'$ is quantized using a "robust" MQ-k-P, and transmitted over an FSC. Let $U \subseteq \{1, 2, ..., NT\}$ and $NT \in U$. Then, for any $\epsilon > 0$, the average error probability with ML decoding given the $m$the message is transmitted, is upper-bounded as*

$$\overline{P}_{e|m} \leq \sum_{l \in U \setminus \{NT\}} \overline{A}_l + \overline{A}_{NT} D_1^N + Pr[\mathbf{X}^N \mathbf{Y}^N \notin T_N] + 2^{-N(C_P - R - \frac{\log_2 \alpha}{N} - \epsilon)}, \tag{31}$$

where

$$\alpha = \max_{l \in U^c} \frac{\overline{A}_l}{M-1} \frac{2^{NT}}{\binom{NT}{l}} \tag{32}$$

Proof: See Appendix D. ∎

The following corollary proves that the coset ensembles generated by the ensembles A, B and C mentioned in Section III in conjunction with an MQ-k-P achieves C_P by using Proposition 3.

Corollary 2. *Consider a sequence of coset ensembles generated by the ensembles A, B, and C mentioned in Section III with length NT. For a given pmf P of a stationary ergodic Markov process, a "robust" MQ-k-P is used to quantize these ensembles before transmission. Then, for any $\epsilon > 0$, there exists a sequence of quantized coset ensembles generated by A, B, and C, have limiting (with respect to N) rate $(1-\epsilon)C_P$ and vanishing average block error probability under ML decoding on FSCs when conditions stated in Fact 2 are satisfied.*

Proof: The form of the bound in Proposition 3 is slightly different from that of the bound in Proposition 2. The union bound part with Battacharrya-like parameter in Proposition 3 is established only for $l = NT$ instead of being established for the whole set U as in Proposition 2. However, this is not problematic, since $\sum_{l \in U \setminus \{NT\}} \overline{A}_l$ approaches 0 as $N \to \infty$ for all three ensembles. Hence we can proceed again as in the proof of Fact 2. ∎

Since a sequence of stationary ergodic Markov sources asymptotically achieves the capacity of FSCs as the order k goes to infinity as in [32], a sequence of quantized coset code ensembles asymptotically achieves the capacity of FSCs for a large enough T. We note however, that [32] does not provide any useful bounds on how fast (with respect to k) a k-th order Markov process approaches the capacity of a FSC, and thus, we cannot make more accurate predictions about the requited order of the MQ-k-P.

As mentioned in the previous Section, if we concentrate on ensemble C, neither the symmetric assumption nor the non-inverting assumption about the channel is required to prove capacity achievability, since these two assumptions only enter our arguments through Lemma 3 which is not needed when proving capacity achievability for ensemble C. Thus, for a large enough k and T, the quantized coset ensemble generated by ensemble C can achieve the capacity of any binary-input FSC.

V. Conclusion

We presented SIR-achieving and capacity-achieving code ensembles for FSCs. We first established an upper bound on the average block error probability of coset code ensembles transmitted through FSCs. We used this bound to show that coset ensembles generated by regular LDPC, punctured LDPC, and LDPC-GM ensembles which achieve the capacity of MBIOS channels also achieve the SIR of FSCs. Next, we presented a method of quantization that enables the construction of code ensembles inducing a Markov distribution. We established an upper bound on the average block error probability of quantized coset code ensembles transmitted through FSCs. Using this bound, we showed that the sequences of quantized regular LDPC, punctured LDPC, and LDPC-GM coset ensembles can achieve the capacity of FSCs as the order of the induced Markov distribution approaches infinity.

Appendix A
Proof of Proposition 1

For a specific code, let $P_{e|m}^{U^c}$ denote the probability that there exists some codeword $\mathbf{c}_{m'}$ such that $Q(\mathbf{y}|\mathbf{c}_{m'}) \geq Q(\mathbf{y}|\mathbf{c}_m)$ and $d(\mathbf{c}_m, \mathbf{c}_{m'}) \in U^c$. Let $\overline{P}_{e|m}^{U^c}$ denote the corresponding ensemble averaged probabilities. Then,

$$\overline{P}_{e|m} = \overline{P}_{e|m}^{U} + \overline{P}_{e|m}^{U^c}. \tag{33}$$

Consider now a new ensemble of codes generated by removing all codewords which satisfy $d(\mathbf{c}_m, \mathbf{c}_{m'}) \in U$. Let $\overline{P}'_{e|m}$ be the average error probability of the new ensemble when the mth codeword of the original ensemble is transmitted. Then, $\overline{P}_{e|m}^{U^c} \leq \overline{P}'_{e|m}$.

From now on, we will upper-bound $\overline{P}'_{e|m}$. Consider a decoder which declares that the ith message is transmitted if there exists a *unique* $\mathbf{c}_i$ satisfying

$$(\mathbf{c}_i, \mathbf{y}^N) \in T_N. \tag{34}$$

Otherwise, it declares an error. Then, the probability of error of the ML decoder is upper-bounded by the probability of error of this decoder. Let

$$E_i = \left\{(\mathbf{c}_i, \mathbf{y}^N) \in T_N\right\}. \tag{35}$$

Then,

$$P'_{e|m} \leq Pr\left[E_m^C \cup \left(\cup_{m' \neq m} E_{m'}\right)\right] \tag{36a}$$

$$\leq Pr\left[E_m^C\right] + \sum_{m' \neq m : d(\mathbf{c}_m, \mathbf{c}_{m'}) \in U^c} Pr\left[E_{m'}\right]. \tag{36b}$$

where $\mathbf{y}^N$ is interpreted as an output corresponding to $\mathbf{c}_m$. We have

$$Pr[E_m^C] = \sum_{\mathbf{y}^N} Pr[(\mathbf{c}_m, \mathbf{y}^N) \notin T_N], \tag{37}$$

and for $m' \neq m$,

$$Pr[E_{m'}] = \sum_{\mathbf{y}^N} Pr[(\mathbf{c}_{m'}, \mathbf{y}^N) \in T_N]. \tag{38}$$

Note that

$$\overline{P}'_{e|m} \leq \sum_C Pr[C] Pr\left[E_m^C\right] + \sum_C Pr[C] \sum_{m' \neq m : d(\mathbf{c}_m, \mathbf{c}_{m'}) \in U^c} Pr\left[E_{m'}\right]. \tag{39}$$

We have

$$\sum_C Pr[C] Pr\left[E_m^C\right] = \sum_{\mathbf{y}^N} \sum_{\mathbf{x}^N} Pr[\mathbf{c}_m = \mathbf{x}^N] Pr[(\mathbf{x}^N, \mathbf{y}^N) \notin T_N] \tag{40a}$$

$$= \sum_{\mathbf{y}^N} \sum_{\mathbf{x}^N} P(\mathbf{x}^N) Pr[(\mathbf{x}^N, \mathbf{y}^N) \notin T_N] \tag{40b}$$

$$= Pr[\mathbf{X}^N \mathbf{Y}^N \notin T_N], \tag{40c}$$

and

$$\sum_{C} Pr[C] \sum_{m'\neq m: d(\mathbf{c}_m,\mathbf{c}_{m'})\in U^c} Pr\left[E_{m'}\right] \tag{41a}$$

$$= \sum_{m'\neq m} \sum_{\mathbf{x}^N,\mathbf{x}'^N: d(\mathbf{x}^N,\mathbf{x}'^N)\in U^c} Pr[\mathbf{c}_m=\mathbf{x}^N]Pr[\mathbf{c}_{m'}=\mathbf{x}'^N|\mathbf{c}_m=\mathbf{x}^N]\sum_{\mathbf{y}^N} Pr[(\mathbf{x}'^N,\mathbf{y}^N)\in T_N)] \tag{41b}$$

$$\leq \sum_{m'\neq m}\sum_{\mathbf{y}^N}\sum_{\mathbf{x}^N,\mathbf{x}'^N: d(\mathbf{x}^N,\mathbf{x}'^N)\in U^c} P(\mathbf{x}^N)\alpha P(\mathbf{x}'^N)Pr[(\mathbf{x}'^N,\mathbf{y}^N)\in T_N)] \tag{41c}$$

$$\leq \alpha \sum_{m'\neq m}\sum_{\mathbf{y}^N}\sum_{\mathbf{x}^N,\mathbf{x}'^N} P(\mathbf{x}^N)P(\mathbf{x}'^N)Pr[(\mathbf{x}'^N,\mathbf{y}^N)\in T_N)] \tag{41d}$$

$$\leq \alpha \sum_{m'\neq m}\sum_{\mathbf{x}^N,(\mathbf{x}'^N,\mathbf{y}^N)\in T_N} P(\mathbf{x}^N)P(\mathbf{x}'^N)Q(\mathbf{y}^N|\mathbf{x}^N) \tag{41e}$$

$$= \alpha \sum_{m'\neq m}\sum_{\mathbf{x}^N,(\mathbf{x}'^N,\mathbf{y}^N)\in T_N} P(\mathbf{x}^N,\mathbf{y}^N)P(\mathbf{x}'^N) \tag{41f}$$

$$= \alpha \sum_{m'\neq m}\sum_{(\mathbf{x}'^N,\mathbf{y}^N)\in T_N} P(\mathbf{x}'^N)P(\mathbf{y}^N), \tag{41g}$$

Note that we have, for $(\mathbf{x}'^N,\mathbf{y}^N)\in T_N$,

$$P(\mathbf{y}^N) \leq Q(\mathbf{y}^N|\mathbf{x}'^N)2^{-N(C_P-\epsilon)}. \tag{42}$$

Hence,

$$\sum_{C} Pr[C] \sum_{m'\neq m: d(\mathbf{c}_m,\mathbf{c}_{m'})\in U^c} Pr\left[E_{m'}\right] \leq \alpha \sum_{m'\neq m}\sum_{(\mathbf{x}'^N,\mathbf{y}^N)\in T_N} P(\mathbf{x}'^N)Q(\mathbf{y}^N|\mathbf{x}'^N)2^{-N(C_P-\epsilon)} \tag{43a}$$

$$\leq \alpha \sum_{m'\neq m} 2^{-N(C_P-\epsilon)}. \tag{43b}$$

Consequently,

$$\overline{P}'_{e|m} \leq Pr[\mathbf{X}^N\mathbf{Y}^N \notin T_N] + (M-1)\alpha 2^{-N(C_P-\epsilon)} \tag{44a}$$

$$\leq Pr[\mathbf{X}^N\mathbf{Y}^N \notin T_N] + 2^{-N(C_P-R-\frac{\log_2\alpha}{N}-\epsilon)}. \tag{44b}$$

Finally,

$$\overline{P}_{e|m} \leq \overline{P}^U_{e|m} + Pr[\mathbf{X}^N\mathbf{Y}^N \notin T_N] + 2^{-N(C_P-R-\frac{\log_2\alpha}{N}-\epsilon)}. \tag{45}$$

APPENDIX B
PROOF OF LEMMA 1

Consider a decoder with the following properties. First, the decoder assumes that the FSC stays at the state s_0 for the whole transmission of $\mathbf{x}$. Second, it quantizes every received value y to z in the following way

$$z = \begin{cases} 1, & \text{when } y > 0 \\ -1, & \text{when } y < 0 \\ \pm 1 \text{ w.p. } 1/2, & \text{when } y = 0. \end{cases} \tag{46}$$

Third, it decides that $\mathbf{x}^N$ is transmitted instead of $\mathbf{x}'^N$ if and only if $P(\mathbf{z}^N|\mathbf{x}^N, \mathbf{s}_0^N) \geq P(\mathbf{z}^N|\mathbf{x}'^N, \mathbf{s}_0^N)$, where $\mathbf{s}_0^N = (s_0, s_0, \ldots, s_0)$. Then, the pairwise error probability between $\mathbf{x}^N$ and $\mathbf{x}'^N$ with ML decoding is no greater than that with this decoder. Therefore,

$$P_{e|\mathbf{x}^N,\mathbf{x}'^N} \leq \sum_{\mathbf{z}^N} P(\mathbf{z}^N|\mathbf{x}^N) I\left\{P(\mathbf{z}^N|\mathbf{x}^N, \mathbf{s}_0^N) < P(\mathbf{z}^N|\mathbf{x}'^N, \mathbf{s}_0^N)\right\} \tag{47a}$$

$$\leq \sum_{\mathbf{z}^N} P(\mathbf{z}^N|\mathbf{x}^N) \sqrt{\frac{P(\mathbf{z}^N|\mathbf{x}'^N, \mathbf{s}_0^N)}{P(\mathbf{z}^N|\mathbf{x}^N, \mathbf{s}_0^N)}} \tag{47b}$$

$$= \sum_{\mathbf{z}^N} \sum_{\mathbf{s}^N} P(\mathbf{s}^N|\mathbf{x}^N) \prod_{i=1}^{N} P(z_i|x_i, s_i) \sqrt{\frac{P(z_i|x_i', s_0)}{P(z_i|x_i, s_0)}} \tag{47c}$$

$$= \sum_{\mathbf{s}^N} P(\mathbf{s}^N|\mathbf{x}^N) \prod_{i=1}^{N} \left\{ \sum_{z} P(z|x_i, s_i) \sqrt{\frac{P(z|x_i', s_0)}{P(z|x_i, s_0)}} \right\} \tag{47d}$$

$$= \sum_{\mathbf{s}^N} P(\mathbf{s}^N|\mathbf{x}^N) \prod_{i:x_i \neq x_i'} \left\{ \sum_{z} P(z|x_i, s_i) \sqrt{\frac{P(z|x_i', s_0)}{P(z|x_i, s_0)}} \right\} \tag{47e}$$

$$= \sum_{\mathbf{s}^N} P(\mathbf{s}^N|\mathbf{x}^N) \times \prod_{i:x_i=1} \left\{ \sum_{z} P(z|1, s_i) \sqrt{\frac{P(z|0, s_0)}{P(z|1, s_0)}} \right\} \prod_{i:x_i=0} \left\{ \sum_{z} P(z|0, s_i) \sqrt{\frac{P(z|1, s_0)}{P(z|0, s_0)}} \right\} \tag{47f}$$

$$\overset{(a)}{=} \sum_{\mathbf{s}^N} P(\mathbf{s}^N|\mathbf{x}^N) \prod_{i:x_i \neq x_i'} \left\{ \sum_{z} P(z|1, s_i) \sqrt{\frac{P(z|0, s_0)}{P(z|1, s_0)}} \right\} \tag{47g}$$

$$\leq \sum_{\mathbf{s}^N} P(\mathbf{s}^N|\mathbf{x}^N) \prod_{i:x_i \neq x'_i} \max_s \left\{ \sum_z P(z|1,s) \sqrt{\frac{P(z|0,s_0)}{P(z|1,s_0)}} \right\} \tag{47h}$$

$$= \left[\max_s \left\{ \sum_z P(z|1,s) \sqrt{\frac{P(z|0,s_0)}{P(z|1,s_0)}} \right\} \right]^{d(\mathbf{x}^N,\mathbf{x}'^N)}, \tag{47i}$$

where $I\{\cdot\}$ denotes the indicator function of its argument, and the equality in (a) is due to the fact that the channel at each state is symmetric. Since the above is true for any choice of the hypothesized state s_0, we have

$$P_{e|\mathbf{x}^N,\mathbf{x}'^N}$$

$$\leq \left[\min_{s_0} \max_s \left\{ \sum_z P(z|1,s) \sqrt{\frac{P(z|0,s_0)}{P(z|1,s_0)}} \right\} \right]^{d(\mathbf{x}^N,\mathbf{x}'^N)} \tag{48a}$$

$$= D^{d(\mathbf{x}^N,\mathbf{x}'^N)}. \tag{48b}$$

In order to show that $D < 1$ we argue as follows. Since the channel at each state is non-inverting, we have

$$\sqrt{\frac{Q^+_{s_0}}{Q^-_{s_0}}} > 1 > \sqrt{\frac{Q^-_{s_0}}{Q^+_{s_0}}}. \tag{49}$$

Hence, for any s_0 we have

$$s^* \triangleq \arg\max_s \left[Q^+_s \sqrt{\frac{Q^-_{s_0}}{Q^+_{s_0}}} + Q^-_s \sqrt{\frac{Q^+_{s_0}}{Q^-_{s_0}}} \right] \tag{50a}$$

$$= \arg\max_s Q^-_s, \tag{50b}$$

which means that the maximizer of (50) is independent of s_0. Thus choosing $s_0 = s^*$ we have

$$D \leq Q^+_{s^*} \sqrt{\frac{Q^-_{s^*}}{Q^+_{s^*}}} + Q^-_{s^*} \sqrt{\frac{Q^+_{s^*}}{Q^-_{s^*}}} = 2\sqrt{Q^+_{s^*} Q^-_{s^*}} < 1. \tag{51a}$$

APPENDIX C
PROOF OF LEMMA 3

Consider a decoder with the following properties. First, the decoder assumes that the FSC stays at the state s_0 for the whole codeword transmission. Second, it quantizes every received value y to z according to (46). Third, it decides that message m is transmitted instead of m'

if and only if $P(\mathbf{z}^N|\delta(\mathbf{c}'_m), \mathbf{s}_0^N) \geq P(\mathbf{z}^N|\delta(\mathbf{c}'_{m'}), \mathbf{s}_0^N)$, where $\mathbf{c}'_m,\ \mathbf{c}'_{m'} \in C'$, $\delta(\cdot)$ is the Markov quantizer mapping, and $\mathbf{s}_0^N = (s_0, s_0, \ldots, s_0)$. Then, the pairwise error probability between m and m' with ML decoding is no greater than that with this decoder. For a binary sequence $\mathbf{x}^{NT}$, let

$$\delta(\mathbf{x}^{NT}) \triangleq \left(\delta(\mathbf{x}^{NT})_1, \delta(\mathbf{x}^{NT})_2, ..., \delta(\mathbf{x}^{NT})_N\right). \tag{52}$$

Then,

$$\begin{aligned}
P_{e|m,m'} &\leq \sum_{\mathbf{z}^N} P(\mathbf{z}^N|\delta(\mathbf{c}'_m)) I\left\{P(\mathbf{z}^N|\delta(\mathbf{c}'_m), \mathbf{s}_0^N) < P(\mathbf{z}^N|\delta(\mathbf{c}'_{m'}), \mathbf{s}_0^N)\right\} && \text{(53a)}\\
&\leq \sum_{\mathbf{z}^N} P(\mathbf{z}^N|\delta(\mathbf{c}'_m)) \sqrt{\frac{P(\mathbf{z}^N|\delta(\mathbf{c}'_{m'}), \mathbf{s}_0^N)}{P(\mathbf{z}^N|\delta(\mathbf{c}'_m), \mathbf{s}_0^N)}} && \text{(53b)}\\
&= \sum_{\mathbf{z}^N} \sum_{\mathbf{s}^N} P(\mathbf{s}^N|\delta(\mathbf{c}'_m)) \prod_{i=1}^{N} \left\{P(z_i|\delta(\mathbf{c}'_m)_i, s_i) \sqrt{\frac{P(z_i|\delta(\mathbf{c}'_{m'})_i, s_0)}{P(z_i|\delta(\mathbf{c}'_m)_i, s_0)}}\right\} && \text{(53c)}\\
&\leq \prod_{i=1}^{N} \max_s \left\{\sum_z P(z|\delta(\mathbf{c}'_m)_i, s) \sqrt{\frac{P(z|\delta(\mathbf{c}'_{m'})_i, s_0)}{P(z|\delta(\mathbf{c}'_m)_i, s_0)}}\right\}. && \text{(53d)}
\end{aligned}$$

We can now average over the ensemble $\mathcal{C}'$, which is equivalent to averaging over all possible translation vectors, as follows

$$\begin{aligned}
&\overline{P}_{e|m,m'}\\
&\leq \mathbf{E}_{\mathbf{v}}\left[\prod_{i=1}^{N} \max_s \left\{\sum_z P(z|\delta(\mathbf{c}_m \oplus \mathbf{v})_i, s) \sqrt{\frac{P(z|\delta(\mathbf{c}_{m'} \oplus \mathbf{v})_i, s_0)}{P(z|\delta(\mathbf{c}_m \oplus \mathbf{v})_i, s_0)}}\right\}\right] && \text{(54a)}\\
&= \sum_{\mathbf{v}} \frac{1}{2^{NT}} \prod_{i=1}^{N} \max_s \left\{\sum_z P(z|\delta(\mathbf{c}_{m'} \oplus \mathbf{c}_m \oplus \mathbf{v})_i, s) \sqrt{\frac{P(z|\delta(\mathbf{v})_i, s_0)}{P(z|\delta(\mathbf{c}_{m'} \oplus \mathbf{c}_m \oplus \mathbf{v})_i, s_0)}}\right\} && \text{(54b)}\\
&\leq \sum_{\mathbf{v}} \frac{1}{2^{NT}} \prod_{i=1}^{N} \max_s \left\{\sum_z P(z|\delta(\mathbf{1}^{NT} \oplus \mathbf{v})_i, s) \sqrt{\frac{P(z|\delta(\mathbf{v})_i, s_0)}{P(z|\delta(\mathbf{1}^{NT} \oplus \mathbf{v})_i, s_0)}}\right\} && \text{(54c)}\\
&\leq \sum_{\mathbf{v}} \frac{1}{2^{NT}} \prod_{i=1, \delta(\mathbf{v})_i \neq \delta(\mathbf{1}^{NT} \oplus \mathbf{v})_i}^{N} \max_s \left\{\sum_z P(z|\delta(\mathbf{1}^{NT} \oplus \mathbf{v})_i, s) \sqrt{\frac{P(z|\delta(\mathbf{v})_i, s_0)}{P(z|\delta(\mathbf{1}^{NT} \oplus \mathbf{v})_i, s_0)}}\right\} && \text{(54d)}\\
&\leq \sum_{\mathbf{v}} \frac{1}{2^{NT}} \prod_{i=1, \delta(\mathbf{v})_i \neq \delta(\mathbf{1}^{NT} \oplus \mathbf{v})_i}^{N} \max_s \left\{\sum_z P(z|0, s) \sqrt{\frac{P(z|1, s_0)}{P(z|0, s_0)}}\right\} && \text{(54e)}
\end{aligned}$$

where (54b) is due to a change of variables in the summation over $\mathbf{v}$, (54c) is due to the fact that the two codewords in consideration are distance NT apart, (54e) is due to the state conditioned channel symmetry. Since the above is true for any choice of the hypothesized state s_0. we have

$$\overline{P}_{e|m,m'} \leq \sum_{\mathbf{v}} \frac{1}{2^{NT}} \prod_{i=1,\delta(\mathbf{v})_i \neq \delta(\mathbf{1}^{NT}\oplus\mathbf{v})_i}^{N} \min_{s_0} \max_{s} \left\{ \sum_{z} P(z|0,s) \sqrt{\frac{P(z|1,s_0)}{P(z|0,s_0)}} \right\} \tag{55a}$$

$$\leq \sum_{\mathbf{v}} \frac{1}{2^{NT}} \prod_{i=1,\delta(\mathbf{v})_i \neq \delta(\mathbf{1}^{NT}\oplus\mathbf{v})_i}^{N} D \tag{55b}$$

$$\leq \sum_{\mathbf{v}} \frac{1}{2^{NT}} \left[\prod_{\substack{i=1,\delta(\mathbf{v})_i \neq \delta(\mathbf{1}^{NT}\oplus\mathbf{v})_i \\ \mathbf{v}_{(i-1)T+1}^{iT} = \mathbf{0}^T}}^{N} D \right] \left[\prod_{\substack{i=1,\delta(\mathbf{v})_i \neq \delta(\mathbf{1}^{NT}\oplus\mathbf{v})_i, \\ \mathbf{v}_{(i-1)T+1}^{iT} \neq \mathbf{0}^T}}^{N} D \right]. \tag{55c}$$

Since $D < 1$ (due to Lemma 1), the last factor is upper bounded by one. Furthermore, since the quantizer is “robust”, for every i for which $\mathbf{v}_{(i-1)T+1}^{iT} = \mathbf{0}^T$, it is also true that $\delta(\mathbf{1}^{NT} \oplus \mathbf{v})_i \neq \delta(\mathbf{v})_i$. Thus we can write

$$\overline{P}_{e|m,m'} \leq \sum_{\mathbf{v}} \frac{1}{2^{NT}} \prod_{i=1}^{N} D^{I\{\mathbf{v}_{(i-1)T+1}^{iT} = \mathbf{0}^T\}} \tag{56a}$$

$$= \left(\mathrm{E}\{ D^{I\{\mathbf{v}_{(i-1)T+1}^{iT} = \mathbf{0}^T\}} \} \right)^N \tag{56b}$$

$$= \left(\frac{2^T - 1}{2^T} + \frac{1}{2^T} D \right)^N \tag{56c}$$

$$= D_1^N, \tag{56d}$$

where $\mathrm{E}\{\cdot\}$ denotes expectation. Again, since $D < 1$, we deduce that $D_1 < 1$.

APPENDIX D
PROOF OF PROPOSITION 3

We will use the bound in Proposition 1 which is stated as follows.

$$\overline{P}_{e|m} \leq \overline{P}_{e|m}^{U} + Pr[\mathbf{X}^N \mathbf{Y}^N \notin T_N] + 2^{-N(C_P - R - \frac{\log_2 \alpha}{N} - \epsilon)}, \tag{57}$$

First, we will bound $\overline{P}^U_{e|m}$ by using the union bound. Consider a code in $\mathcal{C}'$ which results from a code C in $\mathcal{C}$ having weight distribution A_l by adding a constant vector to all codewords. Let $P_{e|m,m'}$ be the pairwise error probability between messages m and m'. Then,

$$P^U_{e|m} \leq \sum_{m' \neq m: d(\mathbf{c}'_m, \mathbf{c}'_{m'}) \in U} P_{e|m,m'} \tag{58a}$$

$$= \sum_{m' \neq m: d(\mathbf{c}'_m, \mathbf{c}'_{m'}) \in U \setminus \{NT\}} P_{e|m,m'} + \sum_{m' \neq m: d(\mathbf{c}'_m, \mathbf{c}'_{m'}) = NT} P_{e|m,m'} \tag{58b}$$

$$\leq \sum_{m' \neq m: d(\mathbf{c}'_m, \mathbf{c}'_{m'}) \in U \setminus \{NT\}} 1 + \sum_{m' \neq m: d(\mathbf{c}'_m, \mathbf{c}'_{m'}) = NT} P_{e|m,m'} \tag{58c}$$

$$= \sum_{l \in U \setminus \{NT\}} A_l + A_{NT} P_{e|m,m'}. \tag{58d}$$

Let $\tilde{P}^U_{e|m,C}$ be the average of $P^U_{e|m}$ over the coset ensemble generated by C. Then, from Lemma 3,

$$\tilde{P}^U_{e|m,C} \leq \sum_{l \in U \setminus \{NT\}} A_l + A_{NT} D_1^N. \tag{59}$$

Then, the average error probability (over the ensemble $\mathcal{C}'$) is

$$\overline{P}^U_{e|m} = \sum_{C \in \mathcal{C}} Pr(C) \tilde{P}^U_{e|m,C} \tag{60a}$$

$$\leq \sum_{l \in U \setminus \{NT\}} \overline{A}_l + \overline{A}_{NT} D_1^N. \tag{60b}$$

To apply Proposition 1, the code ensemble must satisfy (8) and (9). In [44], a code ensemble which satisfy (8) and (9) is generated from a certain linear code for MBIOS channel. In the following we point out how the derivation is different from the one in [44]. In [44, Lemma 1 and Th. 1], starting from an original code C, three ensembles are generated with increasing degree and randomness. The first ensemble $\mathcal{C}^1$ is generated by including $\sigma(C)$ in $\mathcal{C}^1$, for all $\sigma \in \mathcal{S}_M$, where $\sigma(C)$ denotes the code resulting by permuting the order of codewords of C according to σ. The second ensemble $\mathcal{C}^2$ is generated by including $\pi(C^1)$ in $\mathcal{C}^2$, for all $\pi \in \mathcal{S}_N$, and for all $C^1 \in \mathcal{C}^1$, where $\pi(C^1)$ denotes the code resulting by permuting the order of the symbols of all codewords in C^1 according to π. The third ensemble $\mathcal{C}^3$ is generated by including codes of the form $\{\mathbf{c}^2 \oplus \mathbf{v} | \mathbf{c}^2 \in C^2\}$ in $\mathcal{C}^3$, for all $\mathbf{v} \in \{0,1\}^N$, and for all $C^2 \in \mathcal{C}^2$. Although in [44], $\mathcal{C}^3$ is generated starting from a specific code C, it is straightforward to see that the conclusions drawn in that paper still hold if $\mathcal{C}^3$ is generated starting from an ensemble of codes, with the only difference being that the average weight distribution over the original ensemble is used in place of the weight distribution of the original code. Furthermore, it is true that the order of the three operations which generate $\mathcal{C}^1$, $\mathcal{C}^2$, $\mathcal{C}^3$ is irrelevant. Therefore, if we apply σ to the coset permutation-invariant ensemble $\mathcal{C}'$ to generate a new ensemble $\mathcal{C}''$, then $\mathcal{C}''$ also satisfies

$$Pr[\mathbf{c}_i'' = \mathbf{x}^{NT}] = 2^{-NT} \qquad \forall i \in \{1, \ldots, M\} \tag{61a}$$

$$Pr[\mathbf{c}_i'' = \mathbf{x}^{NT} | \mathbf{c}_j'' = \mathbf{x}'^{NT}] \leq \alpha Pr[\mathbf{c}_i'' = \mathbf{x}^{NT}] \; \forall i, j \in \{1, \ldots, M\} \text{ with } i \neq j \text{ and } d(\mathbf{c}_i'', \mathbf{c}_j'') \in U^c, \tag{61b}$$

with $\alpha = \max_{l \in U^c} \frac{\overline{A_l}}{M-1} \frac{2^{NT}}{\binom{NT}{l}}$ (recall that $\mathcal{C}'$ already includes all symbol-permuted and vector-translated codes. In other words, $\mathcal{C}''$ can be thought as generated from $\mathcal{C}_0$ by applying symbol permutation, vector translation, codeword permutation as depicted in Figure 3).

Now consider quantization of the ensemble $\mathcal{C}''$ that produces the Markov quantized ensemble $\mathcal{W}''$. We have,

$$\begin{aligned} Pr[\delta(\mathbf{c}_i'') = \mathbf{w}] &= \sum_{\mathbf{x}:\delta(\mathbf{x}) = \mathbf{w}} Pr[\mathbf{c}_i'' = \mathbf{x}] && \text{(62a)} \\ &\overset{(a)}{=} \sum_{\mathbf{x}:\delta(\mathbf{x}) = \mathbf{w}} 2^{-NT} && \text{(62b)} \\ &= P(\mathbf{w}), && \text{(62c)} \end{aligned}$$

where the equality in (a) is due to (61a). Furthermore, for $\mathbf{c}_i'', \mathbf{c}_j''$ such that $d(\mathbf{c}_i'', \mathbf{c}_j'') \in U^c$,

$$\begin{aligned} & Pr[\delta(\mathbf{c}_i'') = \mathbf{w} | \delta(\mathbf{c}_j'') = \mathbf{w}'] \frac{Pr[\delta(\mathbf{c}_i'') = \mathbf{w}, \delta(\mathbf{c}_j'') = \mathbf{w}']}{Pr[\delta(\mathbf{c}_j'') = \mathbf{w}']} && \text{(63a)} \\ &= \frac{\sum_{\mathbf{x}:\delta(\mathbf{x}) = \mathbf{w}} \sum_{\mathbf{x}':\delta(\mathbf{x}') = \mathbf{w}'} Pr[\mathbf{c}_i'' = \mathbf{x}, \mathbf{c}_j'' = \mathbf{x}']}{\sum_{\mathbf{x}':\delta(\mathbf{x}') = \mathbf{w}'} Pr[\mathbf{c}_j'' = \mathbf{x}']} && \text{(63b)} \\ &\overset{(a)}{\leq} \frac{\sum_{\mathbf{x}:\delta(\mathbf{x}) = \mathbf{w}} \sum_{\mathbf{x}':\delta(\mathbf{x}') = \mathbf{w}'} \alpha 2^{-NT} Pr[\mathbf{c}_j'' = \mathbf{x}']}{\sum_{\mathbf{x}':\delta(\mathbf{x}') = \mathbf{w}'} Pr[\mathbf{c}_j'' = \mathbf{x}']} && \text{(63c)} \\ &= \sum_{\mathbf{x}:\delta(\mathbf{x}) = \mathbf{w}} \alpha 2^{-NT} && \text{(63d)} \\ &= \alpha P(\mathbf{w}), && \text{(63e)} \end{aligned}$$

where the inequality in (a) is due to (61b). Note that the average error probability of the quantized ensemble $\mathcal{W}''$ of $\mathcal{C}''$ is the same as the average error probability of the quantized ensemble $\mathcal{W}'$ of the original coset permutation-invariant ensemble $\mathcal{C}'$ (see Fig. 3).

Even though (62) has a slightly different form than that of (9), we can still apply Proposition 1 for this quantized ensemble. Then, for any $\epsilon > 0$,

$$\overline{P}_{e|m} \leq \sum_{l \in U \backslash \{NT\}} \overline{A}_l + \overline{A}_{NT} D_1^N + Pr[\mathbf{X}^N \mathbf{Y}^N \notin T_N] + 2^{-N(C_P - R - \frac{\log_2 \alpha}{N} - \epsilon)}. \tag{64}$$

References

[1] R. G. Gallager, *Information theory and Reliable Communications*. New York: John Wiley & Sons, 1968.

[2] P. Elias, "Coding for noisy channels," in *IRE Conv. Rec.*, 1955, pp. 37–46.

[3] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo-codes," in *Proc. International Conf. Communications*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[4] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1963.

[5] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *IEE Electronics Letters*, vol. 32, no. 18, pp. 1645–1646, Aug. 1996.

[6] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, D. A. Spielman, and V. Stemann, "Practical loss-resilient codes," in *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, El Paso, Texas, May 1997, pp. 150–159.

[7] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Information Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[8] M. A. Shokrollahi, "New sequences of linear time erasure codes approaching channel capacity," in *Proc. International Symposium on Information Theory and its Applications*, Honolulu, Hawaii, Nov. 1999, pp. 65–76.

[9] A. Shokrollahi, "Capacity-achieving sequences," in *Codes, Systems, and Graphical Models*, ser. IMA Volumes in Mathematics and its Applications, B. Marcus and J. Rosenthal, Eds. Springer-Verlag, 2000, vol. 123, pp. 153–166.

[10] P. Oswald and A. Shokrollahi, "Capacity-achieving sequences for the erasure channel," *IEEE Trans. Information Theory*, vol. 48, no. 12, pp. 3017–3028, Dec. 2002.

[11] H. D. Pfister, I. Sason, and R. Urbanke, "Capacity-achieving ensembles for the binary erasure channel with bounded complexity," *IEEE Trans. Information Theory*, vol. 51, no. 7, pp. 2352–2379, July 2005.

[12] A. Montanari, "Tight bounds for LDPC and LDGM codes under MAP decoding," *IEEE Trans. Information Theory*, vol. 51, no. 9, pp. 3221–3246, Sept. 2005.

[13] C. Measson, A. Montanari, and R. Urbanke, "The generalized area theorem and some of its consequences," *IEEE Trans. Information Theory*, (Submitted).

[14] N. Macris, "Sharp bounds on generalized EXIT functions," *IEEE Trans. Information Theory*, vol. 53, no. 7, pp. 2365–2375, July 2007.

[15] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2007.

[16] G. Miller and D. Burshtein, "Bounds on the maximum-likelihood decoding error probability of low-density parity-check codes," *IEEE Trans. Information Theory*, vol. 47, no. 7, pp. 2696–2710, Nov. 2001.

[17] I. Sason and R. Urbanke, "Parity-check density versus performance of binary linear block codes over memoryless symmetric channels," *IEEE Trans. Information Theory*, vol. 49, no. 7, pp. 1611–1635, July 2003.

[18] G. Wiechman and I. Sason, "Parity-check density versus performance of binary linear block codes over memoryless symmetric channels: New bounds and applications," *IEEE Trans. Information Theory*, vol. 53, no. 2, pp. 550–579, Feb. 2007.

[19] C.-H. Hsu and A. Anastasopoulos, "Capacity achieving LDPC codes through puncturing," *IEEE Trans. Information Theory*, vol. 54, pp. 4698–4706, Oct. 2008.

[20] ——, "Capacity-achieving codes for noisy channels with bounded graphical complexity and maximum likelihood decoding," *IEEE Trans. Information Theory*, Feb. 2006, accepted after revision.

[21] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channels," *IEEE Trans. Information Theory*, vol. 35, no. 6, pp. 1277–1290, Nov. 1989.

[22] A. Goldsmith and P. Varaiya, "Capacity, mutual information and coding for finfite-state Markov channels," *IEEE Trans. Information Theory*, vol. 42, pp. 868–886, May 1996.

[23] T. Holliday, A. Goldsmith, and P. Glynn, "Capacity of finite state channels based on Lyapunov exponents of random matrices," *IEEE Trans. Information Theory*, vol. 52, no. 8, pp. 3509–3532, Aug. 2006.

[24] D. M. Arnold and H.-A. Loeliger, "On the information rate of binary-input channels with memory," in *Proc. International Conf. Communications*, Helsinki, Finland, June 2001, pp. 2692–2695.

[25] V. Sharma and S. K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels," in *Proc. International Symposium on Information Theory*, Washington DC, USA, June 2001, p. 283.

[26] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," in *Proc. Globecom Conf.*, San Antonio, USA, Nov. 2001, pp. 2992–2996.

[27] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Information Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[28] A. Kavcic, "On the capacity of Markov sources over noisy channels," in *Proc. Globecom Conf.*, San Antonio, USA, Nov. 2001, pp. 2997–3001.

[29] P. O. Vontobel, A. Kavcic, D. M. Arnold, and H.-A. Loeliger, "A generalization of the Blahut-Arimoto algorithm to finite-state channels," *IEEE Trans. Information Theory*, vol. 54, no. 5, pp. 1887–1918, May 2008.

[30] P. O. Vontobel and D. M. Arnold, "An upper bound on the capacity of channels with memory and constraint input," in *Proc. Information Theory Workshop*, Cairns, Australia, Sept. 2001, pp. 147–149.

[31] S. Yang, A. Kavcic, and S. Tatikonda, "Feedback capacity of finite-state machine channels," *IEEE Trans. Information Theory*, vol. 51, no. 3, pp. 799–810, Mar. 2005.

[32] J. Chen and P. H. Siegel, "Markov processes asymptotically achieve the capacity of finite-state intersymbol interference channels," *IEEE Trans. Information Theory*, vol. 54, no. 3, pp. 1295–1303, Mar. 2008.

[33] X. Ma, N. Varnica, and A. Kavcic, "Matched information rate codes for binary ISI channels," in *Proc. International Symposium on Information Theory*, Lausanne, Switzerland, June 2002, p. 269.

[34] J. B. Soriaga and P. H. Siegel, "On near-capacity coding systems for partial-response channels," in *Proc. International Symposium on Information Theory*, Chicago, USA, June 2004, p. 263.

[35] A. Kavcic, X. Ma, and N. Varnica, “Matched information rate codes for partial response channels,” *IEEE Trans. Information Theory*, vol. 51, no. 3, pp. 973–989, Mar. 2005.

[36] J. B. Soriaga, H. D. Pfister, and P. H. Siegel, “Determining and approaching achievable rates of binary intersymbol interference channels using multistage decoding,” *IEEE Trans. Information Theory*, vol. 53, no. 4, pp. 1416–1429, Apr. 2007.

[37] H. D. Pfister and P. H. Siegel, “Joint iterative decoding of ldpc codes for channels with memory and erasure noise,” *IEEE J. Select. Areas Commun.*, vol. 26, no. 2, pp. 320–327, Feb. 2008.

[38] A. Kavcic, X. Ma, and M. Mitzenmacher, “Binary intersymbol interference channels: Gallager codes, density evolution, and code performance bounds,” *IEEE Trans. Information Theory*, vol. 49, no. 7, pp. 1636–1652, July 2003.

[39] A. W. Eckford, F. R. Kschischang, and S. Pasupathy, “Analysis of low-density parity-check codes for the Gilbert-Elliott channel,” *IEEE Trans. Information Theory*, vol. 51, no. 11, pp. 3872–3889, Nov. 2005.

[40] ——, “On designing good LDPC codes for Markov channels,” *IEEE Trans. Information Theory*, vol. 53, no. 1, pp. 5–21, Jan. 2007.

[41] P. Grover and A. K. Chaturvedi, “Upper bounds on the rate of LDPC codes for Gilbert-Elliott channels,” in *Proc. Information Theory Workshop*, 2004, pp. 230–235.

[42] ——, “Upper bounds on the rate of LDPC codes for a class of finite-state Markov channels,” *IEEE Trans. Information Theory*, vol. 53, no. 2, pp. 794–804, Feb. 2007.

[43] A. Bennatan and D. Burshtein, “On the application of LDPC codes to arbitrary discrete-memoryless channels,” *IEEE Trans. Information Theory*, vol. 50, no. 3, pp. 417–438, Mar. 2004.

[44] N. Shulman and M. Feder, “Random coding techniques for nonrandom codes,” *IEEE Trans. Information Theory*, vol. 45, no. 6, pp. 2101–2104, Sept. 1999.

[45] T. S. Han, *Information-spectrum methods in information theory*. Springer-Verlag, 2003.

[46] R. G. Gallager, “Low density parity check codes,” *IEEE Trans. Information Theory*, vol. 8, pp. 21–28, Jan. 1962.

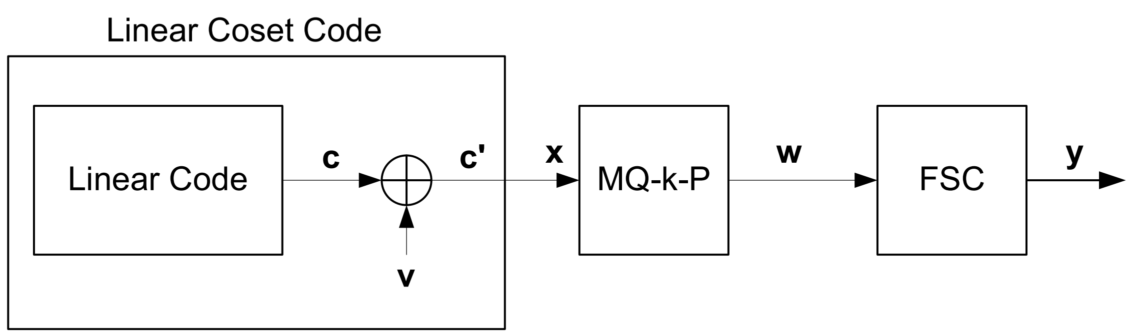

Fig. 1. Capacity achieving transmission scheme with coset codes and Markov quantization.

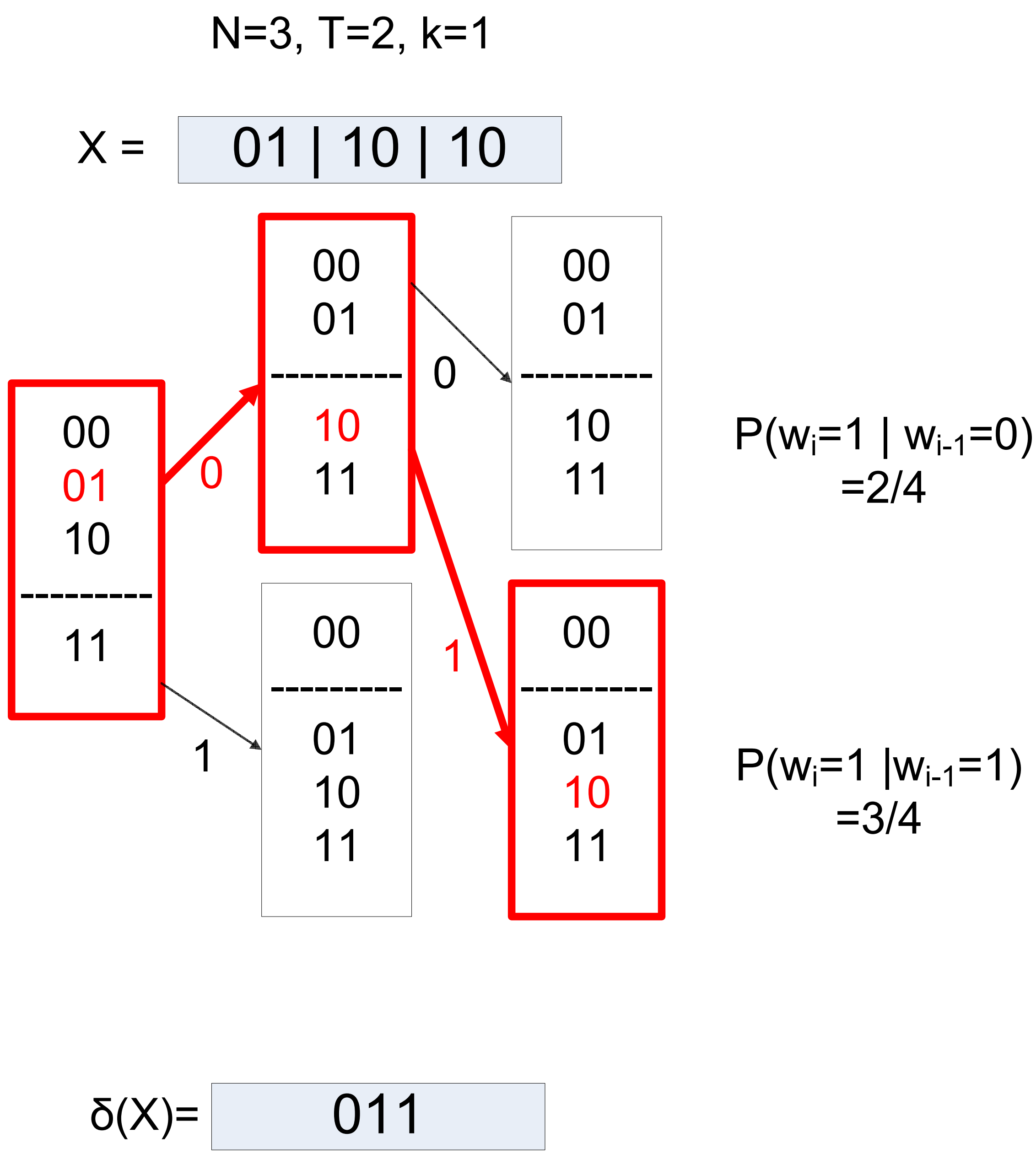

Fig. 2. An example of Markov quantization ($k = 1$).

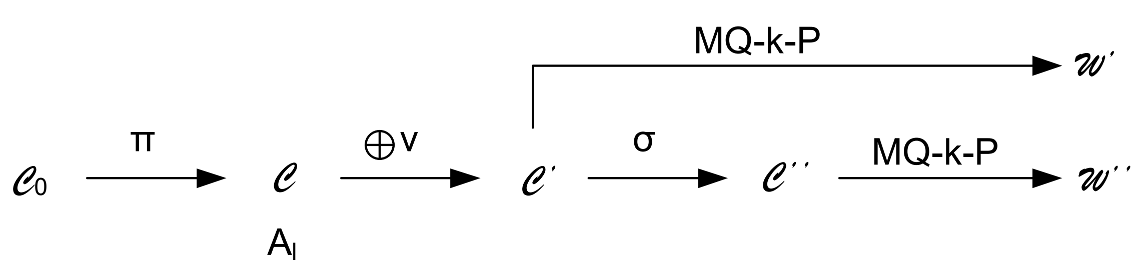

Fig. 3. Relation between different code ensembles used in the proof of Proposition 3.