Phase Transitions for Mutual Information

K. Raj Kumar, Payam Pakzad, Amir Hesam Salavati, and Amin Shokrollahi

Abstract

We consider ensembles of binary linear error correcting codes, obtained by sampling each column of the generator matrix G or parity check matrix H independently from the set of all binary vectors of weight d (of appropriate dimension). We investigate the circumstances under which the mutual information between a randomly chosen codeword and the vector obtained after its transmission over a binary input memoryless output symmetric channel (BIMSC) $\mathcal{C}$ is exactly n times the capacity of $\mathcal{C}$, where n is the length of the code. For several channels such as the binary symmetric channel (BSC) and the binary-input additive white Gaussian noise (AWGN) channel, we prove that the probability of this event has a threshold behaviour, depending on whether n/k is smaller than a certain quantity (that depends on the particular channel $\mathcal{C}$ and d), where k is the number of source bits. To show this, we prove a generalization of the following well-known theorem: the expectation of the size of the right kernel of G has a phase transition from 1 to infinity, depending on whether or not n/k is smaller than a certain quantity depending on the chosen ensemble. We subsequently generalize the results to arbitrary discrete BIMSCs.

Index Terms

I. Introduction

The development of the theory of modern codes over the past two decades has resulted in the construction of practical error correcting codes that operate extremely close to the performance limits dictated by information theory. These modern codes admit low complexity decoding techniques based on the idea of belief propagation. It has been shown that the ensemble of low density parity check (LDPC) codes and Raptor codes are capacity achieving over the binary erasure channel (BEC). Roughly speaking, the BEC is an idealized channel in which information symbols (bits) are either "lost", or recovered with no error, which may be used to model for example a packetized communication system with perfect error detection. However, most practical channels also involve errors, causing information symbols to be confused with one another. When one moves away from the framework of the BEC to more general memoryless symmetric channels, very few analytical results exist regarding the performance of modern coding ensembles. The goal of this paper is to investigate such results from an information theoretic point of view.

We consider the transmission of a vector $X \in \mathbb{F}_2^k$ over a general binary input memoryless symmetric channel (BIMSC) $\mathcal{C}$. The vector X is transformed by a linear encoder into a vector $Y \in \mathbb{F}_2^n$ through the mapping $Y = XG$, with the "generator matrix" $G \in \mathbb{F}_2^{k\times n}$. (Note that we are not assuming that $n \geq k$, so G is not a generator matrix in traditional sense.) The vector Y is then transmitted over the channel $\mathcal{C}$, to obtain Z. We will consider the case that the encoder G corresponds to a family of "LT/Raptor-like" or "LDPC-like" codes. In order to make this notion more precise, we define $\mathcal{E}_d(\ell, m)$ to be the ensemble of all $\ell \times m$ matrices whose columns are sampled independently from the set of vectors $v \in \mathbb{F}_2^\ell$ of weight d. For brevity, we will use the notation $\mathcal{E}_d$ for the ensemble $\mathcal{E}_d(\ell, m)$, the dimensions will be clear from context. Further, we will call an ensemble of codes to be *d-uniform* if either the generator matrix of the code or the parity check matrix of the code is drawn from $\mathcal{E}_d$. We will be interested in evaluating the performance of such d-uniform ensembles of codes.

K. R. Kumar, A. H. Salavati and A. Shokrollahi are with the Laboratoire d'algorithmique, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland ({raj.kumar,hesam.salavati,amin.shokrollahi}@epfl.ch). P. Pakzad is with Qualcomm, Inc., 3195 Kifer Road, Santa Clara, CA 95051, USA (payam@qualcomm.com)

Part of the material in this paper has been submitted to the 6th Intl. Symp. on Turbo Codes & Iterative Information Processing (ISTC 2010).

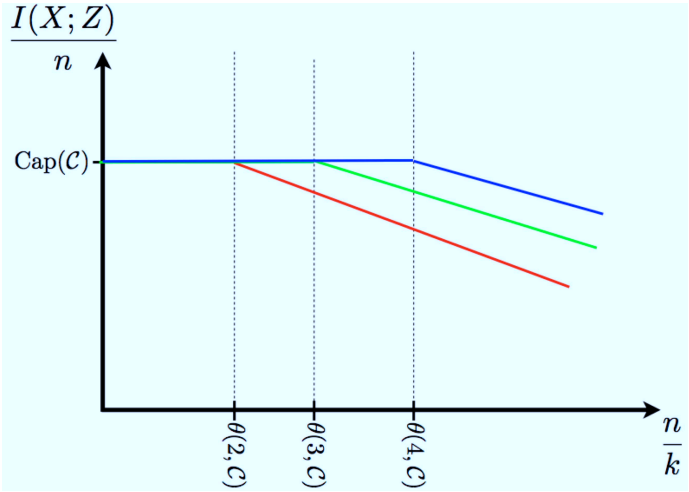

Fig. 1. Illustrating the practical significance of Conjecture 1

For this setting, the primary question that we would like to answer is the following: under the assumption of long blocklengths, what is the probability that the mutual information $I(X;Z)$ is close to n times the capacity of the channel $\mathcal{C}$? For a particular d-uniform ensemble of codes, we define the probability

$$\Pi_{d,\mathcal{C}} = \Pr\{I(X;Z) < n\text{Cap}(\mathcal{C})\}.$$

For reasons of analytical tractability, we also define the following probability

$$\hat{\Pi}_{d,\mathcal{C}} = \Pr\{I(X;Z) < n\text{Cap}(\mathcal{C}) - o(n)\}.$$

We conjecture that the mutual information exhibits the following phase transition.

Conjecture 1: For any BIMSC $\mathcal{C}$ and any integer d, there exists a positive real number $\theta(d,\mathcal{C})$ such that if n/k converges to a value η as $n \to \infty$, then

$$\hat{\Pi}_{d,\mathcal{C}} \to \begin{cases} 0 \text{ if } \eta < \theta(d,\mathcal{C}) \\ 1 \text{ if } \eta > \theta(d,\mathcal{C}) \end{cases}.$$

In the current work, we attempt to prove this conjecture for a broad class of channels, that includes any discrete BIMSC, and the additive white Gaussian noise (AWGN) channel. In certain cases, we will prove a weaker version of Conjecture 1, by showing that $\Pi_{d,\mathcal{C}}$ converges to 0 if n/k is below a certain threshold. In the most general case of an arbitrary BIMSC, the proof of Conjecture 1 remains an open problem.

Fig. 1 illustrates the behaviour of the mutual information versus n/k, as conjectured above. When using a particular d-uniform ensemble of codes, the mutual information per channel use is at capacity till n/k exceeds $\theta(d,\mathcal{C})$, after which it fails to achieve capacity.

It is very informative to look first at the case where the channel $\mathcal{C} = \mathcal{I}$ is the trivial (error-free) channel, such that $Z = Y$. In this case, it is easy to see that the mutual information $I(X;Z) = \text{rk}(G)$, where $\text{rk}(G)$ denotes the rank of the matrix G. Hence in this case, we have that $\Pi_{d,\mathcal{I}} = \Pr\{\text{rk}(G) < n\}$. Using the union bound, one can show that

$$\Pr\{\text{rk}(G) < n\} \leq \mathbb{E}[|\text{lker}(G)|] - 1, \tag{1}$$

where $\text{lker}(G)$ denotes the left kernel of the matrix G, and $\mathbb{E}[\cdot]$ denotes the expectation operator. We have the following well-known result on the phase transition behavior of the size of the left-kernel, see [3, Theorem 3.5.1].

d	$\alpha(d,\mathcal{I})$	$\theta(d,\mathcal{I})$
3	0.8894928741	0.9179352769
4	0.9671474457	0.9767701646
5	0.9891624451	0.9924383911
6	0.9962283325	0.9973795526
7	0.9986504364	0.9990637586

TABLE I
VALUES OF $\alpha(d,\mathcal{I})$ AND $\theta(d,\mathcal{I})$ FOR VARIOUS d

Theorem 1: Let the generator matrix G be drawn from the ensemble $\mathcal{E}_d$, with $d \geq 3$. Further, let $\alpha(d,\mathcal{I})$ be defined as the first component of the vector (a,x,λ) that is the unique solution of the system of equations

$$\begin{aligned} \mathrm{e}^{-x}\cosh(\lambda)\left(\frac{ad}{ad-x}\right)^{a} &= 1, \\ \frac{x}{\lambda}\left(\frac{ad-x}{x}\right)^{1/d} &= 1, \\ \lambda\tanh(\lambda) &= x. \end{aligned}$$

Suppose that $k,n \to \infty$ such that $n/k \to \alpha$. Then, if $\alpha < \alpha(d,\mathcal{I})$, then $\mathbb{E}[|\ker(G)|] \to 1$, and if $\alpha > \alpha(d,\mathcal{I})$, then $\mathbb{E}[|\ker(G)|] \to \infty$.

Notice that this immediately yields that $\Pi_{d,\mathcal{I}} \to 0$ if $\alpha < \alpha(d,\mathcal{I})$, but only yields a trivial bound on $\Pi_{d,\mathcal{I}}$ if $\alpha > \alpha(d,\mathcal{I})$. The following theorem from [5] proves Conjecture 1 for the case $\mathcal{C} = \mathcal{I}$.

Theorem 2: Let

$$\gamma_d := -\frac{\ln\zeta_d}{d(1-\zeta_d)^{d-1}},$$

where ζ_d is the smallest root of $z(1-\ln z) - \frac{1-z}{d}\ln z - 1 = 0$ for $z \in [0,1]$. Then Conjecture 1 is true for $\mathcal{C} = \mathcal{I}$ and $\theta(d,\mathcal{I}) := \gamma_d$. In other words, if $n,k \to \infty$ such that $n/k \to \alpha$ and $\alpha < \theta(d,\mathcal{I})$, then $\hat{\Pi}_{d,\mathcal{I}} \to 0$, whereas $\hat{\Pi}_{d,\mathcal{I}} \to 1$ if $\alpha > \theta(d,\mathcal{I})$.

Table I gives values of $\theta(d,\mathcal{I})$ and $\alpha(d,\mathcal{I})$ for various d.

Finally, we would like to comment on literature related to the topic of this paper. The results that are closest to the spirit of those in this paper are the ones in [1]–[3], [5]; one can think of these results as special cases of our results when the channel $\mathcal{C}$ is error-free.

There is a whole set of other papers that discuss under which conditions $I(X;Z) = k$, so that ML-decoding is successful[1]. The most general among such results (but with a limited range of applicability) are those of MacMullan and Collins [6] which analyze the inherent gap of certain families of binary linear codes such as the Hamming and Golay codes to the capacity of the BSC. For ensembles of sparse matrices the question of achievability of capacity is not new, of course. Already in his thesis, Gallager [7, pp. 37–38] showed that the rate of a right (or check-) regular LDPC code that achieves reliable communication over a BSC using ML decoding is bounded away from the capacity of the channel by a function depending on the right degree of the underlying graph. In particular, the right degree has to go to infinity if the code is to approach capacity. Richardson et al. [8] proved that the same conclusion holds for the maximum right degree, if the graph is not right-regular; this implies that the result also applies when taking the average right degree, instead. Burshtein et al. [9] generalized these results to general BIMSC. These results were themselves generalized and optimized by Sason and Urbanke [10] who gave rather close gaps to capacity for LDPC codes with given average right degree.

[1]For G to achieve capacity, we need to have that $I(X;Z) = k$; however, we are interested in this paper in the case when $I(X;Z) = n\mathrm{Cap}(\mathcal{C})$. These two quantities are equal only if $k = n\mathrm{Cap}(\mathcal{C})$, so that the rate of the code is equal to the capacity.

Though it may seem to a reader that this paper is investigating a similar problem as those of the above papers, this is not entirely the case. In all the above cases, either $k - I(X;Z)$ is calculated directly (e.g., in [6]), or an *upper* bound is obtained on the entropy $H(Z)$ to show that $I(X;Z)$ is bounded away from k (as is the case in [7]–[10]). For us a direct calculation of $k - I(X;Z)$ is very difficult, so that the results of [6] are not directly applicable. Moreover, we are interested in *lower* bounds for $H(Z)$ (or rather, its expectation), rather than upper bounds, so the mentioned results are not applicable either.

The rest of the paper is organized as follows. In Section II, we examine the case of a BSC, and prove results relating to the threshold behaviour of the MI for this channel. Subsequently, in Section III, we generalize these results to the case where the channel is a convex combination of BSCs, which can be used to model any discrete BIMSC. Section IV treats the case of the binary-input, continuous-output AWGN channel. The lengthy proofs are relegated to the appendices.

II. The Binary Symmetric Channel $\mathcal{C} = \mathrm{BSC}(p)$

In this section, we study the case of a BSC with crossover probability p, denoted by $\mathcal{C} = \mathrm{BSC}(p)$. To generalize the results for $\mathcal{C} = \mathcal{I}$ to this channel, we proceed as follows.

Lemma 1: Suppose that $\mathcal{C} = \mathrm{BSC}(p)$. Then, with the notation introduced in the introduction, we have

$$\Pr[I(X;Z) < n\mathrm{Cap}(\mathcal{C})] \leq n - \mathbb{E}[H(Z)],$$

where $H(Z)$ is the entropy of Z.

Proof: Let u be a random variable taking values in the set $\{0, 1, \ldots, t\}$, and let p_i denote the probability that the value of u is i. Then $\Pr[u < t] = p_0 + \cdots + p_{t-1} = 1 - p_t \leq p_{t-1} + 2p_{t-2} + \cdots + tp_0 = t - \mathbb{E}[u]$. We apply this result to $u = I(X;Z)$ and $t = n\mathrm{Cap}(\mathcal{C})$ to obtain that the probability in question is upper bounded by $n\mathrm{Cap}(\mathcal{C}) - E[I(X;Z)]$. Noting that $I(X;Z) = H(Z) - H(Z|X) = H(Z) - n\,h(p)$, and that $\mathrm{Cap}(\mathcal{C}) = 1 - h(p)$, the result follows. ∎

The main theorem of this section is the following.

Theorem 3: Let B_w denote the number of words of weight w in the right kernel of the matrix G. Then

$$\Pr[I(X;Z) < n\mathrm{Cap}(\mathcal{C})] \leq \log_2 \left(\sum_{w=0}^{n} \mathbb{E}[B_w](1-2p)^{2w} \right).$$

Note that if we assume $\mathcal{C} = \mathcal{I}$, so that $p = 0$, then $I(X;Z)$ is the rank of the matrix G, and $\sum_w B_w$ is the size of the right kernel of G. Hence, the statement of the theorem says that $\Pr[\mathrm{rk}(G) < n] \leq \log_2 (\mathbb{E}[|\mathrm{lker}(G)|]) \leq \mathbb{E}[|\mathrm{lker}(G)|] - 1$, and we have retrieved (1).

To prove Theorem 3, we need an auxiliary result, which may be interesting in its own right.

Theorem 4: Let D be a distribution on $\mathbb{F}_2^n$, with entropy $H(D)$, and let $p_u := \Pr_D[x = u]$. For $v \in \mathbb{F}_2^n$ let $q_v := \sum_{u, \langle u|v \rangle = 1} p_u$, where $\langle u \mid v \rangle$ is the scalar product of u and v (over $\mathbb{F}_2^n$). Then we have

$$n - H(D) \leq \log_2 \left(\sum_{v \in \mathbb{F}_2^n} (1-2q_v)^2 \right).$$

Proof: We will first remark some general facts. First, note that

$$1 - 2q_v = 1 - q_v - q_v = \sum_{\substack{u \\ \langle u|v \rangle = 0}} p_u - \sum_{\substack{u \\ \langle u|v \rangle = 1}} p_u = \sum_u (-1)^{\langle u|v \rangle} p_u,$$

so that the vector $(1 - 2q_v \mid v \in \mathbb{F}_2^n)$ is the Hadamard transform of the vector $(p_u \mid u \in \mathbb{F}_2^n)$. Let H be the $2^n \times 2^n$-Hadamard matrix. Since $H/\sqrt{2^n}$ is a unitary matrix, we have

$$\sum_{u \in \mathbb{F}_2^n} p_u^2 = \frac{1}{2^n} \sum_v (1-2q_v)^2. \tag{2}$$

Note that by the concavity of the logarithm function, we have for all $x_1, \ldots, x_m \geq 0$ and all $a_1, \ldots, a_m \geq 0$ with $\sum_i a_i = 1$:

$$\log_2(\sum_i a_i x_i) \geq \sum_i a_i \log_2(x_i).$$

Specializing to $m = 2^k$, and $a_u = x_u = p_u$, we see that $\sum_u p_u^2 \geq \prod_u p_u^{p_u} = 2^{-H(D)}$, so that

$$-H(D) \leq \log_2 \left(\sum_{u \in \mathbb{F}_2^n} p_u^2 \right) = -n + \log_2 \left(\sum_{v \in \mathbb{F}_2^n} (1-2q_v)^2 \right),$$

which is the statement of the theorem. ∎

Corollary 1: Suppose that C is an $[n,k]$-code, and that $y = (y_1, \ldots, y_n)$ is a vector chosen from C uniformly at random. Denote the dual code of C by $C^\perp$. Moreover, suppose that for $i = 1, \ldots, n$ the random variables ξ_i are independent binary Bernoulli random variables with $\Pr[\xi_i = 1] = p_i$, and suppose that $y, \xi_1, \ldots, \xi_n$ are independent. Let $z = (z_1, \ldots, z_n)$ be the vector with $z_i = y_i + \xi_i$. Then we have

$$n - H(z) \leq \log_2 \left(\sum_{c \in C^\perp} \prod_{\substack{i \\ c_i = 1}} (1-2p_i)^2 \right),$$

where $H(z)$ is the entropy of the probability distribution of the random variable z. In particular, if all the p_i are equal to p, then

$$n - H(z) \leq \log_2 \left(\sum_{w=0}^{n} B_w (1-2p)^{2w} \right),$$

where B_w is the number of words of weight w in the right kernel of G (i.e., the number of vectors of weight w in $C^\perp$).

Proof: The second assertion follows from the first, so it suffices to prove the first one. This assertion follows from Theorem 4 if we can show the following:

$$\begin{aligned} q_v &= \frac{1}{2}, \quad \text{if } v \notin C^\perp, \\ q_v &= \frac{1 - \prod_{i, v_i = 1} (1-2p_i)}{2}, \quad \text{if } v \in C^\perp. \end{aligned}$$

Note that $q_v = \Pr[vz^\top = 1]$. We have $v^\top z = vy^\top + v\xi^\top$, where $\xi = (\xi_1, \ldots, \xi_n)$. If $v \notin C^\perp$, then $vy^\top$ is a uniform binary random variable, and since y and ξ are independent, we see that $v(y+\xi)^\top$ is a uniform binary random variable, and hence $q_v = 1/2$. If $v \in C^\perp$, then $vz^\top = v\xi^\top = \sum_{i, v_i = 1} \xi_i$. Since the ξ_i's are independent, it follows that

$$\Pr[vz^\top = 1] = \Pr\big[\sum_{i, v_i = 1} \xi_i = 1 \big] = \frac{1 - \prod_{i, v_i = 1} (1-2p_i)}{2},$$

and we are done. ∎

The proof of Theorem 3 follows from the previous corollary.

Proof: (Of Theorem 3) Let C be the code generated by the rows of the matrix G. Then, by the previous corollary, we have

$$n - \mathbb{E}[H(Z)] \leq \mathbb{E} \left[\log_2 \left(\sum_{w=0}^{n} B_w (1-2p)^{2w} \right) \right].$$

Since $\log_2(x)$ is a concave function, we have for any random variable U over the positive real numbers that $\mathbb{E}[\log_2(U)] \leq \log_2(\mathbb{E}[U])$. Hence, it suffices to prove that

$$\Pr[I(X;Z) < n\text{Cap}(\mathcal{C})] \leq n - \mathbb{E}[H(Z)],$$

which was proven in lemma 1. ■

Using the above results, we obtain the following theorem whose proof is given in Appendix A.

Theorem 5: Let $d \geq 3$ be fixed and $\mathcal{C} = \text{BSC}(p)$. Define

$$f(\lambda) := \frac{ed}{\lambda \tanh(\lambda)} \cosh(\lambda)^{d/\lambda \tanh(\lambda)} \left(\frac{\tanh(\lambda)}{\mathrm{e}}\right)^d (1-2p)^2,$$

$$g(\lambda,\phi) := \cosh(\lambda) \left(\frac{\tanh(\lambda)}{e}\right)^{\lambda \tanh(\lambda)} \left(\frac{d\,\phi}{d\,\phi - \lambda \tanh(\lambda)}\right)^{\phi} \cdot \left(\frac{d\,\phi - \lambda \tanh(\lambda)}{\lambda \tanh(\lambda)}(1-2p)^2\right)^{\lambda \tanh(\lambda)/d},$$

and

$$u(\phi) = \left\{ \left(\frac{1+e^{-2\phi d}}{2}\right)^{\left(1-\frac{1}{\phi}\right)} \right\}.$$

Let $1/\theta_0$ be the maximum of $f(\lambda)$ in the interval $(0,\infty)$, and let θ_1 be the largest positive value of ϕ such that $g(\lambda,\phi) \leq 1$ for all λ with $\lambda \tanh(\lambda) \leq d\,\phi$. Also, let θ_2 be the maximum value of $\phi \geq 0$ such that $u(\phi) > (1-2p)^2$. Set $\alpha(d,\mathcal{C}) := \min(\max(\theta_0,\theta_1),\theta_2)$. Suppose that n,k go to infinity such that $n/k \to \alpha$. Then

$$\Pi_{d,\mathcal{C}} \to 0 \text{ if } \alpha < \alpha(d,\mathcal{C}).$$

Proof: See Appendix A. ■

Theorem 6: The largest value of the function

$$f(\lambda) = \frac{2\mathrm{e}}{\lambda \tanh(\lambda)} \cosh(\lambda)^{2/\lambda \tanh(\lambda)} \left(\frac{\tanh(\lambda)}{\mathrm{e}}\right)^2 (1-2p)^2$$

in $[0,\infty]$ is $\lim_{\lambda \to 0} f(\lambda) = \frac{1}{2(1-2p)^2}$.

Hence, even though we cannot show Theorem 5 for $d = 2$, an informal application of this theorem for $d = 2$ gives the correct result for $\theta(2,\mathcal{C})$.

Proof: See appendix B. ■

The case $d = 2$ is more involved. In fact, we cannot show that $\alpha(2,\text{BSC}(p))$ exists. However, it was proved in [4] that in this case the analogous threshold for $\hat{\Pi}_{2,\text{BSC}(p)}$, viz., $\theta(2,\text{BSC}(p))$ exists and

$$\theta(2,\text{BSC}(p)) = \frac{1}{2(1-2p)^2}.$$

Moreover, when $d = 2$, one can show that the largest value of function $f(\lambda)$ is equal to $\frac{1}{2(1-2p)^2}$ and happens when $\lambda \to 0$.

Table II gives the value of $\text{Cap}(\text{BSC}(p))\alpha(d,\text{BSC}(p)) = (1-h(p))\alpha(d,\text{BSC}(p))$ for various d and p. One would expect these values to converge to 1 as d grows. While this is seen to happen for $p \ll 1$ (see also Table I that corresponds to the limiting case of $p = 0$), the values converge to around $1/2$ when p converges to $1/2$. This suggests that there is room for improvement in the bounds of Theorem 5.

As an auxiliary result, we now show that the inequality of Theorem 4 is tight for flat distributions, i.e., for distributions D that are uniform over a k-dimensional subspace of $\mathbb{F}_2^n$, where $1 \leq k \leq n$.

Theorem 7: We have the following:

$d \setminus p$	10^{-4}	10^{-3}	0.01	0.1	0.2	0.4	0.45
3	0.889	0.881	0.837	0.680	0.590	0.496	0.488
4	0.959	0.959	0.910	0.728	0.617	0.510	0.500
5	0.979	0.979	0.928	0.738	0.623	0.512	0.503
6	0.989	0.979	0.938	0.738	0.626	0.513	0.503
7	0.989	0.989	0.938	0.743	0.626	0.513	0.503
8	0.989	0.989	0.938	0.743	0.626	0.513	0.503
9	0.999	0.989	0.938	0.743	0.626	0.513	0.503
10	0.999	0.989	0.938	0.743	0.626	0.513	0.503

TABLE II
THE VALUES OF $(1-h(p))\alpha(d, \mathrm{BSC}(p))$ FOR VARIOUS VALUES OF d AND p.

(1) Suppose that D is a distribution which is uniform on a k-dimension subspace of $\mathbb{F}_2^n$, where $1 \leq k \leq n$. Then

$$n - H(D) = n - k = \log_2 \left(\sum_{v \in \mathbb{F}_2^n} (1-2q_v)^2 \right).$$

(2) Let f be a Bent function on $\mathbb{F}_2^n$, and let D be a probability distribution that is uniform on its support $S = \{x \mid f(x) = 1\}$. Then we have

$$\begin{aligned} n - H(D) &= n - \log_2(|S|) \\ &= \log_2 \left(\sum_{v \in \mathbb{F}_2^n} (1-2q_v)^2 \right) - O(1/\sqrt{2^n}) \end{aligned}$$

as $n \to \infty$.

Proof: See Appendix C. ■

III. CONVEX COMBINATION OF BSCS

Assuming that the all zero codeword was transmitted, the probability density function (pdf) of the log-likelihood ratio (LLR) of the output of a BSC with crossover probability p is given by

$$\phi_p = p\Delta_{\log \frac{p}{1-p}} + (1-p)\Delta_{-\log \frac{p}{1-p}}, \tag{3}$$

where Δ_x denotes a Delta function at the value x. Define $P = (p_1, \ldots, p_N)$ and $\beta = (\beta_1, \ldots, \beta_N)$. Let $\sum_{i=1}^N \beta_i = 1$. We define a new channel CBSC(P, β) to be the channel whose LLR ϕ corresponds to the convex combination of the BSC LLRs:

$$\phi = \sum_{i=1}^N \beta_i \phi_{p_i}. \tag{4}$$

Consider the binary input, $2N-$ary output channel with the particular transition probabilities show in Fig. 2. Let the outputs of the channel be denoted by $\{0_i, 1_i\}_{i=1}^N$. It can be verified that the channel shown in Fig. 2 gives rise to the above mentioned LLR ϕ.

Further, notice that any *discrete* BIMSC may be modeled as a convex combination of BSCs, since the LLRs constitute a sufficient statistic for the binary input channel. Hence by proving results for arbitrary CBSC(P, β) channels, we are in effect proving results for any discrete BIMSC.

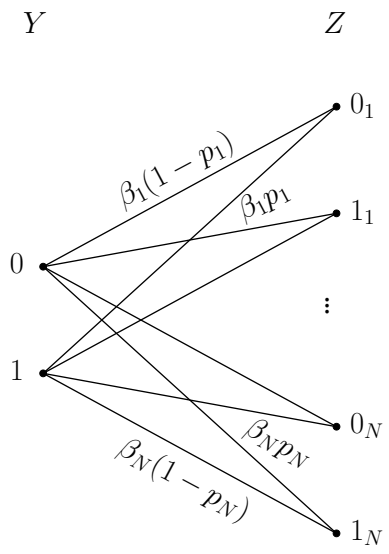

Fig. 2. Convex combination of BSCs

A. Output entropy and the capacity of CBSC(P, β)

We now compute the entropy of the output Z of the channel in Fig. 2, resulting in the following lemma.

Lemma 2: The output entropy of CBSC(P, β) is given by

$$H(Z) = nH(\beta) + \sum_{i=1}^{N} \beta_i H(Z_i), \tag{5}$$

where Z_i denotes the output of BSC(p_i) in response to the same input as that of CBSC(P, β).

Proof: We first consider the case of scalar inputs, i.e., X, Y and Z are all scalars. Let the input probability distribution $P(X = 0) = \alpha, P(X = 1) = 1 - \alpha$. For every $i = 1, \ldots, N$, we have that

$$\begin{aligned} P(Z = 0_i) &= \alpha\beta_i(1 - 2p_i) + \beta_i p_i \\ &\triangleq \beta_i(c_i\alpha + d_i), \end{aligned}$$

where $c_i = (1 - 2p_i)$ and $d_i = p_i$. Also,

$$\begin{aligned} P(Z = 1_i) &= \alpha\beta_i(2p_i - 1) + \beta_i(1 - p_i) \\ &\triangleq \beta_i(e_i\alpha + f_i), \end{aligned}$$

where $e_i = 2p_i - 1$ and $d_i = 1 - p_i$. Hence the output entropy

$$\begin{aligned} H(Z) = -\sum_{i=1}^{N} \beta_i \left[(c_i'\alpha + d_i') \log(c_i'\alpha + d_i') + (e_i'\alpha + f_i') \right. \\ \left. \cdot \log(e_i'\alpha + f_i') + (c_i'\alpha + d_i') \log \beta_i + (e_i'\alpha + f_i') \log \beta_i \right]. \end{aligned} \tag{6}$$

Notice that $c_i'\alpha + d_i' = \alpha(1 - 2p_i) + p_i$ and $e_i'\alpha + f_i' = \alpha(2p_i - 1) + (1 - p_i)$ are exactly equal to the output probabilities of the channel BSC(p_i), when the input probabilities are α and $1 - \alpha$. Defining the random variable Z_i as the output of a BSC with transition probability p_i and simplifying the above, we have shown Lemma 2 for scalar inputs. Generalizing the statement to the general vector input case is straightforward, and is omitted for brevity. ■

From the above expression, it is also clear that uniform inputs achieve the capacity of CBSC(P, β), owing to them being optimal for the BSC.

Further, we may compute the conditional entropy of the output of CBSC(P,β) given the input as

$$\begin{aligned} H(Z|X) &= H(Z|X=0) \\ &= -n\sum_{i=1}^{N}\left[\beta_i(1-p_i)\log\beta_i(1-p_i)+\beta_i p_i\log\beta_i p_i\right] \\ &= n\left[H(\beta)+\sum_{i=1}^{N}\beta_i h(p_i)\right], \end{aligned} \tag{7}$$

where $h(\cdot)$ denotes the binary entropy function. Hence, the capacity of the CBSC is given by

$$\mathrm{Cap}(\text{CBSC}(P,\beta)) = 1-\sum_{i=1}^{N}\beta_i h(p_i). \tag{8}$$

B. An Equivalent Compound Channel

Consider the compound channel [12] shown in Fig. 3. The channel is a BSC, whose crossover probability is determined by the random variable $\mathcal{B}$. More specifically, let $\mathcal{B}$ be an $N-$ary random variable taking on the value i with probability β_i, for all $i=1,2,\ldots,N$. Each time the compound channel is used, a realization i of $\mathcal{B}$ is chosen i.i.d., and according to this value i, the compound channel behaves as BSC(p_i). Let Z' denote the output of this BSC. We consider both $\mathcal{B}$ and Z' to be the outputs of the compound

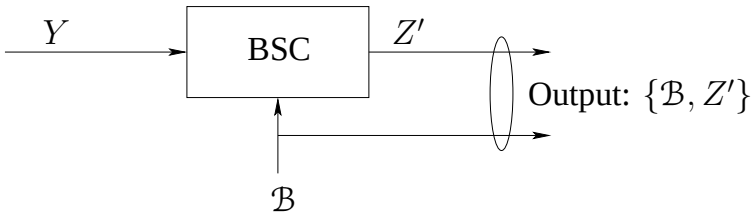

Fig. 3. Equivalent compound channel

channel. The output entropy of the compound channel is given by

$$\begin{aligned} H(\mathcal{B},Z') &= H(\mathcal{B})+H(Z'|\mathcal{B}) \\ &= nH(\beta)+\sum_{i=1}^{N}\beta_i H(Z_i), \end{aligned}$$

where Z_i is the corresponding output of BSC(p_i) as defined in the previous section. From (5), it is clear that the channel CBSC(P,β) is indistinguishable from the compound channel in Fig. 3 in terms of output entropy. We will hence focus on the compound channel, in order to compute the entropy of the output of CBSC(P,β).

Lemma 3: Let $[N]$ denote the set $\{1,2,\ldots,N\}$. Then we have the following lower bound on the output entropy of $n-$uses of the compound channel:

$$H(\mathcal{B},Z') \geq n(1+H(\beta))-\log_2\left[\sum_{w=0}^{n}B_w\mathcal{P}^w\right], \tag{9}$$

where B_w is the number of words of weight w in the right kernel of G, and $\mathcal{P}\triangleq\sum_{i=1}^{N}\beta_i(1-2p_i)^2$.

Proof: See Appendix D. ■

C. Thresholds of the MI for CBSC(P,β)

From Lemma 3, (7) and the established equivalence between CBSC(P,β) and the compound channel, we obtain a lower bound on the mutual information for CBSC(P,β) as

$$I(X;Z) \geq n - \log_2\left[\sum_{w=0}^{n} B_w \mathcal{P}^w\right] - n\sum_{i=1}^{N}\beta_i h(p_i).$$

In conjunction with (8), the above lower bound gives us that

$$\Pr\{I(X;Z) < n\text{Cap}(\mathcal{C})\} \leq \Pr\left\{\sum_{w=1}^{n} B_w \mathcal{P}^w > 0\right\}.$$

Evaluating the above, we obtain the following theorem:

Theorem 8: Let $d \geq 3$ be fixed and $\mathcal{C} = CBSC(P,\beta)$. Define

$$f(\lambda,\phi) \triangleq \frac{\mathcal{P}\phi e d}{\lambda\tanh\lambda}\cosh(\lambda)^{\frac{d}{\lambda\tanh(\lambda)}}\left(\frac{\tanh(\lambda)}{e}\right)^{d},$$

$$g(\lambda,\phi) = \cosh(\lambda)\left(\frac{\tanh(\lambda)}{e}\right)^{\lambda\tanh(\lambda)}\left(\frac{\phi d}{\phi d - \lambda\tanh(\lambda)}\right)^{\phi}\left(\frac{\phi d - \lambda\tanh(\lambda)}{\lambda\tanh(\lambda)}\mathcal{P}\right)^{\frac{\lambda\tanh(\lambda)}{d}}$$

and

$$u(\phi) \triangleq \left\{\left(\frac{1+e^{-2\phi d}}{2}\right)^{\left(1-\frac{1}{\phi}\right)}\right\}.$$

Let θ_0 (θ_1) be the largest positive value of ϕ such that $f(\lambda,\phi) \leq 1$ (respectively, $g(\lambda,\phi) \leq 1$) for all $\lambda > 0$ with $\lambda\tanh(\lambda) \leq d\,\phi$. Also, let θ_2 be the maximum value of $\phi \geq 0$ such that $u(\phi) > \mathcal{P}$. Set $\alpha(d,\mathcal{C}) := \min\left(\max(\theta_0,\theta_1),\theta_2\right)$. Suppose that n,k go infinity such that $n/k \to \alpha$. Then

$$\Pi_{d,\mathcal{C}} \to 0 \text{ if } \alpha < \alpha(d,\mathcal{C}).$$

Proof: See Appendix A. ∎

Define the threshold for the rate $R(d,\mathcal{C}) = \frac{1}{\alpha(d,\mathcal{C})}$. We now bound the rate threshold $R(d,\text{CBSC}(P,\beta))$ in terms of the rate thresholds $R(d,\text{BSC}(p_i))$ of the constituent BSCs.

Theorem 9: The rate threshold of the convex combination of BSCs is bounded in terms of the rate thresholds of the constituent BSCs by $R(d,\text{CBSC}(P,\beta)) \leq \max\{R_0, e\overline{R}\}$, where

$$R_0 = \frac{d}{\ln\left(\frac{2}{1+\sum_{i=1}^{N}\beta_i e^{-2d/R(d,\text{BSC}(p_i))}}\right)},$$

and

$$\overline{R} = \sum_{i=1}^{N}\beta_i R(d,\text{BSC}(p_i)).$$

Proof: See Appendix E. ∎

IV. The AWGN Channel

We now turn our attention to the case when $\mathcal{C} = \text{AWGN}(\rho)$ is a binary input (real) AWGN channel, whose output $Z \in \mathbb{R}^n$ may be written as

$$Z = Y + W,$$

where $W \sim$ i.i.d. $\mathcal{N}\left(0, \frac{1}{\rho}I\right)$. We assume standard binary phase shift keying (BPSK) modulation for transmission over the AWGN channel, i.e., we map component-wise the binary codeword $Y \mapsto (-1)^Y$ prior to transmission over the channel. With a slight abuse of notation, we refer to both the binary codeword and the modulated symbols with the same notation Y; the one being referred to will be clear from the context. Hence ρ denotes the signal to noise ratio (SNR) of the AWGN channel.

A. A lower bound on the mutual information

As before, we first develop an upper bound on the mutual information between the input and output.

$$\begin{aligned} I(X;Z) &= H(Z) - H(Z|X) \\ &= H(Z) - \frac{1}{2}\log\frac{(2\pi e)^n}{\rho^n}. \end{aligned} \tag{10}$$

Using Jensen's inequality, we lower bound the entropy as

$$H(Z) \geq -\log\left[\int p^2(Z)\ dZ\right], \tag{11}$$

where $p(Z)$ denotes the pdf of the output Z. Define the code $C_Y = \left\{Y = XG \,\middle|\, X \in \mathbb{F}_2^k\right\} \triangleq \{Y_1, Y_2, \ldots, Y_{2^k}\}$ (note that if G is not full-rank, then not all Y_i are distinct). Then, we may write

$$p(Z) = \frac{1}{2^k}\sum_{i=1}^{2^k} \frac{\rho^{n/2}}{(\sqrt{2\pi})^n} e^{-\frac{\rho}{2}|Z-Y_i|^2}.$$

We may hence evaluate

$$\int p^2(Z)\ dZ = \frac{\rho^n}{2^{2k}(2\pi)^n}\int \sum_{i,j=1}^{2^k} \exp\left\{-\frac{\rho}{2}\left[|Z-Y_i|^2 + |Z-Y_j|^2\right]\right\}\ dZ$$

A simple manipulation yields

$$\begin{aligned} \int p^2(Z)\ dZ &= \frac{\rho^n}{(2\pi)^n 2^{2k}}\sum_{i,j=1}^{2^k}\int \exp\left\{-\frac{\rho}{2}\left[\frac{|Y_i-Y_j|^2}{2} + 2\left|Z - \frac{Y_i+Y_j}{2}\right|^2\right]\right\} dZ \\ &= \frac{(\rho\pi)^{n/2}}{(2\pi)^n 2^{2k}}\sum_{i,j=1}^{2^k}\exp\left\{-\frac{\rho}{4}|Y_i-Y_j|^2\right\}. \end{aligned} \tag{12}$$

From (10), (11) and (12), we obtain

$$I(X;Z) \geq -\log\left[\left(\frac{e}{2}\right)^{n/2}\frac{1}{2^{2k}}\sum_{i,j=1}^{2^k}\exp\left\{-\frac{\rho}{4}|Y_j-Y_i|^2\right\}\right].$$

Since BPSK modulation is used, $|Y_j - Y_i|^2 = 4d_H(Y_j, Y_i)$, where $d_H(A,B)$ denotes the Hamming distance between A and B. Since we employ a linear code, we may further simplify the above inequality to obtain

$$I(X;Z) \geq -\log\left[\left(\frac{e}{2}\right)^{n/2}\frac{1}{2^k}\sum_{i=1}^{2^k}\exp\left\{-\rho\ w_H(Y_i)\right\}\right],$$

where $w_H(X)$ denotes the Hamming weight of X. If we define C_w to be the number of codewords with Hamming weight w, we may rewrite the above as:

$$I(X;Z) \geq -\log\left[\left(\frac{e}{2}\right)^{n/2}\frac{1}{2^k}\sum_{w=0}^{n} C_w e^{-w\rho}\right]. \tag{13}$$

In order to examine Conjecture 1 for $\mathcal{C} = \text{AWGN}(\rho)$, we evaluate

$$\begin{aligned}\Pr\{I(X;Z) < n(\text{Cap}(\mathcal{C}) - \epsilon)\} &\leq \Pr\left\{-\log\left[\left(\frac{e}{2}\right)^{n/2}\frac{1}{2^k}\sum_{w=0}^{n} C_w e^{-w\rho}\right] < n(\text{Cap}(\mathcal{C}) - \epsilon)\right\} \\ &= \Pr\left\{\left(\sqrt{\frac{e}{2}}2^{\text{Cap}(\mathcal{C})-R}\right)^n \sum_{w=0}^{n} C_w e^{-w\rho} > 2^{n\epsilon}\right\}, \end{aligned} \tag{14}$$

where ϵ is a constant independent of n, k. For $\epsilon \geq \log_2\left(\sqrt{\frac{e}{2}}\right)$, the above reduces to

$$\Pr\{I(X;Z) < n(\text{Cap}(\mathcal{C}) - \epsilon)\} \leq \Pr\left\{\left(2^{\text{Cap}(\mathcal{C})-R}\right)^n \sum_{w=0}^{n} C_w e^{-w\rho} > 1\right\}.$$

Notice that we are solely interested in the case that $R \geq \text{Cap}(\mathcal{C})$, since $I(X;Z)$ is trivially lesser than $\text{Cap}(\mathcal{C})$ for $R < \text{Cap}(\mathcal{C})$. In this case,

$$\Pr\{I(X;Z) < n(\text{Cap}(\mathcal{C}) - \epsilon)\} \leq \Pr\left\{\left(2^{\text{Cap}(\mathcal{C})-R}\right)^n \sum_{w=1}^{n} C_w e^{-w\rho} > 0\right\}. \tag{15}$$

B. Thresholds of the MI for the AWGN channel

It will be shown in the sequel that the probability in (15) converges to zero for all $\epsilon \geq \log_2\left(\sqrt{\frac{e}{2}}\right)$ under certain conditions on the rate and SNR, resulting in the following theorem.

Theorem 10: Let $d \geq 3$ be a fixed constant, and $\mathcal{C} = \text{AWGN}(\rho)$. Define $b = 2^{\text{Cap}(\mathcal{C})-R}$,

$$f(\lambda, \phi) \triangleq \frac{e^{-\rho}\phi e d}{\lambda \tanh\lambda} b^{\frac{d\phi}{\lambda\tanh(\lambda)}} \cosh(\lambda)^{\frac{d}{\lambda\tanh(\lambda)}} \left(\frac{\tanh(\lambda)}{e}\right)^d, \tag{16}$$

and

$$g(\lambda, \phi) = \cosh(\lambda)\left(\frac{\tanh(\lambda)}{e}\right)^{\lambda\tanh(\lambda)} \left(\frac{b\phi d}{\phi d - \lambda\tanh(\lambda)}\right)^{\phi} \left(\frac{\phi d - \lambda\tanh(\lambda)}{\lambda\tanh(\lambda)} e^{-\rho}\right)^{\frac{\lambda\tanh(\lambda)}{d}}. \tag{17}$$

Let θ_1 (θ_2) denote the maximum value of ϕ for which the function $f(\lambda, \phi)$ (respectively, $g(\lambda, \phi)$) is less than one for all non-negative λ such that $\lambda\tanh\lambda \leq d\phi$. Set $\theta(d, \mathcal{C}) = 1 + \min\{1, \max\{\theta_1, \theta_2\}\}$. If $n, k \to \infty$ such that $n/k \to \alpha$, then

$$\hat{\Pi}_{d,\mathcal{C}} \to 0 \quad \text{if } \alpha < \theta(d, \mathcal{C}).$$

Proof:

In order to analyze C_w, we consider two scenarios. When $n \geq k$, we use a good channel code to transmit information across the channel. On the other hand, when $n < k$, we need to compress (quantize) the information to be sent over the channel. We first examine the case when $n \geq k$.

1) Channel coding when $n \geq k$: This is a special case of a general proof that we have given in Appendix A.

$d \setminus \rho$(dB)	-8	-7	-5	-4	-2	0	3	10
3	0.221	0.288	0.485	0.618	0.923	1	1	1
4	0.313	0.425	0.775	1	1	1	1	1
5	0.527	0.784	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1

TABLE III
THE VALUES OF CAP(AWGN(ρ))$\theta(d,$ AWGN(ρ)) FOR VARIOUS VALUES OF d AND ρ.

2) Compression (Quantization) for $n < k$: We fix G to be any $k \times n$ binary matrix that is of rank n. Hence, as X varies over all k-tuples, Y varies over all n-tuples, with each n-tuple appearing 2^{k-n} times in the quantizer codebook. This results in $C_w = 2^{k-n}\binom{n}{w}$. Consider

$$\begin{aligned} b^n \sum_{w=0}^{n} C_w e^{-\rho w} &= b^n 2^{k-n} \sum_{w=0}^{n} \binom{n}{w} e^{-\rho w} \\ &= b^n 2^{k-n} \left(1 + e^{-\rho}\right)^n. \end{aligned}$$

We may now evaluate (14) as

$$\Pr\{I(X;Z) < n(\text{Cap}(C) - \epsilon)\} \leq \Pr\left\{ b 2^{\frac{k}{n}-1}(1 + e^{-\rho}) > 2^{\epsilon} \right\}.$$

For $\epsilon \geq \log_2\left(\sqrt{\frac{e}{2}}\right)$, the above is equivalent to

$$\Pr\{I(X;Z) < n(\text{Cap}(C) - \epsilon)\} \leq \Pr\left\{ 2^{\text{Cap}(\mathcal{C})\left(\frac{1+e^{-\rho}}{2}\right)} > 1 \right\}.$$

Notice that $\text{Cap}(\mathcal{C})$ is not known in closed form for the binary constrained AWGN channel, which makes it difficult to analytically evaluate the above. It can however be verified numerically that the right-hand side converges to zero. ■

Shown in Table III are the values of Cap(AWGN(ρ))$\theta(d,$ AWGN(ρ)) for several values of ρ and d. Notice that the bounds for the case of the AWGN are very tight in terms of the threshold values for the rate being almost at capacity for even moderate values of d.

However, our main theorem for the AWGN, Theorem 10 is in some sense weaker than the corresponding result for the BSC in Theorem 5, since the former proves a statement about $\hat{\Pi}_{d,\mathcal{C}}$ involving a linear back-off from capacity, while the latter shows a result relating to $\Pi_{d,\mathcal{C}}$ with no back-off from capacity.

In Theorem 10, we fixed the SNR ρ of the channel, and examined the rate of transmission required for the mutual information to be close to the capacity. Analogously, in the following corollary, we fix the rate of transmission, and provide lower bounds on the SNR such that the ensemble is capacity achieving (on average).

Corollary 2: Let the rate of transmission be R bits/channel use. In order for the d-uniform parity check ensemble to be capacity achieving over the AWGN channel, the channel SNR must be greater than a threshold ρ_0 given by:

$$\rho_0 = \max\left\{\rho_{0_1}, \min\{\rho_{0_2}, \rho_{0_3}\}\right\}, \tag{18}$$

where

$$\rho_{0_1} = \text{Cap}(\mathcal{C})\ln(2) - R\ln\left(\frac{1 + e^{-\frac{2d}{1-R}}}{2}\right), \tag{19}$$

$$\rho_{0_2} = \max_{\lambda > 0} h_1(\lambda, R), \tag{20}$$

$$\rho_{0_3} = \max_{\lambda>0} h_2(\lambda, R), \tag{21}$$

and

$$h_1(\lambda, R) = \frac{d}{\lambda\tanh(\lambda)}\left[\frac{\ln(b)}{1-R} + \frac{\lambda\tanh\lambda}{d}\ln\left(\frac{ed}{(1-R)\lambda\tanh\lambda}\right) + \ln(\cosh(\lambda)) + \lambda\tanh(\lambda)\ln\left(\frac{\tanh(\lambda)}{e}\right)\right],$$

$$h_2(\lambda, R) = d\ln\left(\frac{\tanh(\lambda)}{e}\right) + \left(\frac{d}{\lambda\tanh(\lambda)}\right)\ln(\cosh(\lambda)) + \left(\frac{d}{(1-R)\lambda\tanh(\lambda)}\right)\ln\left(\frac{bd}{d-(1-R)\lambda\tanh(\lambda)}\right) + \ln\left(\frac{d-(1-R)\lambda\tanh(\lambda)}{(1-R)\lambda\tanh(\lambda)}\right) \tag{22}$$

Proof: Follows from the proof of Theorem 10, in particular from (34), (36) and (37). ■

C. Simulating the AWGN channel as a convex combination of infinite BSCs

As mentioned previously, Theorem 10 involves a linear backoff from capacity, which was not present in the corresponding results for the BSC and CBSC. In order to get an idea of the effect that the linear backoff from capacity entails on our results, we model the AWGN channel as a convex combination of an infinite number of BSCs, and use the results of Theorem 8 to analyze this channel. Notice that our analysis in this subsection is not rigorous, since we are assuming that the results that we have proved in Section III for convex combinations of a finite number of BSCs extend naturally to the case of an infinite number of constituent channels (i.e., all sums converge to their corresponding integrals). However, we believe that this analysis, albeit not rigorous, will give us valuable insight regarding the tightness of our bounding techniques, and identifying areas that can be improved upon.

Define the SNR $\rho \triangleq \frac{1}{\sigma^2}$. The pdf of the LLR at the output of the AWGN channel, assuming that the all-zero codeword has been transmitted is given by [11]

$$\phi_{\text{AWGN}} = \sqrt{\frac{\sigma^2}{8\pi}} e^{-\frac{\left(y-\frac{2}{\sigma^2}\right)^2\sigma^2}{8}}.$$

In order to express the above pdf as a convex combination of the pdfs of the BSC LLRs ϕ_p given in (3), we proceed as follows. Notice that choosing a BSC with a crossover probability of $\frac{1}{1+e^y}$ for some $y > 0$ would result in the corresponding pdf of the LLRs $\phi_{\frac{1}{1+e^y}}$ being equal to two delta functions at y and $-y$, with amplitudes $\frac{e^y}{1+e^y}$ and $\frac{1}{1+e^y}$ respectively. Using the symmetry property of the pdf of the LLRs for the BIMSC, we may write down ϕ_{AWGN} as a convex combination of the pdfs of BSC LLRs,

$$\phi_{\text{AWGN}} = \int_0^\infty \left(\frac{1+e^y}{e^y}\right)\sqrt{\frac{\sigma^2}{8\pi}} e^{-\left(y-\frac{2}{\sigma^2}\right)^2\frac{\sigma^2}{8}} \phi_{\frac{1}{1+e^y}}\, dy.$$

Comparing this representation with (4), we may write down the equivalent parameter $\mathcal{P}$ (that appears in the analysis of the CBSC in Theorem 8) for the AWGN channel as

$$\mathcal{P}_{\text{AWGN}} = \sqrt{\frac{\sigma^2}{8\pi}} \int_0^\infty \left(\frac{e^y-1}{e^y}\right) e^{-\left(y-\frac{2}{\sigma^2}\right)^2\frac{\sigma^2}{8}} dy.$$

We now compute the values of $\mathcal{P}_{\text{AWGN}}$ numerically, and use Theorem 8 to calculate $\alpha(d, \mathcal{C})$ for $\mathcal{C} = \text{AWGN}(\rho)$. Using these, we plot in Table IV the values of $\text{Cap}(\text{AWGN}(\rho))\alpha(d, \text{AWGN}(\rho))$. Comparing this with Table III allows us to get an idea of what the effect of a linear backoff from capacity has on the threshold values.

$d \setminus \rho$(dB)	-8	-7	-5	-4	-2	0	3	10
3	0.2578	0.2885	0.3599	0.4020	0.4919	0.5880	0.7350	0.8871
4	0.2684	0.3016	0.3796	0.4237	0.5198	0.6317	0.7927	0.9569
5	0.2716	0.3055	0.3836	0.4285	0.5303	0.6414	0.8071	0.9868
6	0.2716	0.3055	0.3856	0.4309	0.5303	0.6414	0.8143	0.9868
8	0.2726	0.3055	0.3856	0.4309	0.5303	0.6463	0.8143	0.9968
10	0.2726	0.3055	0.3856	0.4309	0.5338	0.6463	0.8143	0.9968
15	0.2726	0.3055	0.3856	0.4309	0.5338	0.6463	0.8143	0.9968
20	0.2726	0.3055	0.3856	0.4309	0.5338	0.6463	0.8143	0.9968

TABLE IV
THE VALUES OF CAP(AWGN(ρ))$\alpha(d, \text{AWGN}(\rho))$ FOR VARIOUS VALUES OF d AND ρ.

V. ACKNOWLEDGEMENTS

The authors would like to thank Mahdi Cheraghchi, Venkat Guruswami, Shlomo Shamai, Emre Telatar, and David Tse for helpful discussions and comments during the research on this paper.

APPENDIX A
ASYMPTOTIC BEHAVIOR OF THE MAIN SUM

Consider a parity check matrix A with dimensions $\nu n \times n$ where ν is a constant (we assume that νn is an integer). Let A be drawn from the ensemble $\mathcal{E}_d$. More specifically, we generate an ensemble that is asymptotically identical to $\mathcal{E}_d$ (along the lines of [3]), as follows.

Consider the following linear equation in $\mathbb{F}_2$:

$$E_{i_1} + \cdots + E_{i_d} = 0, \tag{23}$$

where $i_1, \ldots, i_d$ are i.i.d. random variables taking on the values $\{1, 2, \ldots, \nu n\}$ with equal probability. If an even number of these random variables take on the same value j, then E_j does not appear in the equation, owing to the mod-two addition over $\mathbb{F}_2$. We may associate with the equation in (23) a column vector $E \in \mathbb{F}_2^{\nu n}$, defined as follows. The j^{th} coordinate of E is equal to one if E_j appears (i.e., does not cancel out) in (23), else it is equal to zero. We generate an ensemble of matrices in the following way: each matrix $A \in \mathbb{F}_2^{\nu n \times n}$ is such that every column of A is drawn independently from the column vector ensemble defined above. Hence each column of A can have a maximum of d ones. As $n, k \to \infty$, we have that this ensemble of matrices tends towards the ensemble $\mathcal{E}_d$, where each column has *exactly* d ones.

Let B_w denote the number of vectors in the right null-space of A. We are interested in characterizing the behavior of the following sum, in the limit of large n:

$$\mathcal{S} = b^n \sum_{w=1}^{n} \mathbb{E}[B_w] \mathcal{P}^w, \tag{24}$$

where $b, \mathcal{P}$ are fixed constants that are independent of n. In particular, we will be interested in determining the range of values for ν for which $\mathcal{S}$ converges to zero asymptotically. Using this analysis, we will provide proofs to Theorems 5, 8 and 10. In the following, we use the same approach as in the proof of Theorem 3.5.1 in [3].

We first divide the sum (24) into three parts and show that for any fixed $\varepsilon > 0$, there exists $\delta > 0$ and a range of values for ν such that for large n:

$$\mathcal{S}_1 \triangleq b^n \sum_{1\le w\le \delta n} \mathbb{E}[B_w]\mathcal{P}^w < \varepsilon, \tag{25}$$

$$\mathcal{S}_2 \triangleq b^n \sum_{(1-\delta)n\le w\le n} \mathbb{E}[B_w]\mathcal{P}^w < \varepsilon\mathcal{P}^{(1-\delta)n}, \tag{26}$$

$$\forall \delta n < w < (1-\delta)n: \quad b^n\mathbb{E}[B_w]\mathcal{P}^w < \mathrm{e}^{-c_w n}, \tag{27}$$

for some constant $c_w > 0$ that depends only on w/n and ν. Given that (25), (26) and (27) are satisfied for a particular range of values of ν, it is obvious that $\mathcal{S}$ converges to zero in this region.

Let p_w be the probability that a vector $v \in \mathbb{F}_2^n$ with weight w satisfies $Av = 0$. Then, $E[B_w] = \binom{n}{w} p_w$. To show (25), we follow the proof of Lemma 3.5.1 of [3], from which we obtain:

$$\begin{aligned}\mathcal{S}_1 &= b^n \sum_{1\le w\le \delta n} \binom{n}{w} p_w \mathcal{P}^w \\ &\le b^n \sum_{1\le w\le \delta n} \left(\left(\frac{1}{\nu}\right)^{d/2} d^{d/2-1} \mathrm{e}^{4d} \delta^{d/2-1} \mathcal{P} \right)^w \end{aligned} \tag{28}$$

Hence, by choosing δ small enough, we can make sure that this sum is smaller than ε.

To prove (26), we use Equation (3.5.5) of [3] which states that for arbitrary $\lambda > 0$, we have that

$$p_w \le [\cosh(\lambda)]^{\nu n} \frac{(dw)!}{(\lambda\nu n)^{dw}}. \tag{29}$$

Proceeding along the lines of the proof of Lemma 3.5.2 of [3], we set $\lambda = \frac{dw}{\nu n}$ to obtain:

$$\begin{aligned}\mathcal{S}_2 &= b^n \sum_{(1-\delta)n\le w\le n} \binom{n}{w} p_w \mathcal{P}^w \\ &\le b^n \sum_{(1-\delta)n\le w\le n} \binom{n}{w} \mathcal{P}^w [\cosh(\lambda)]^{\nu n} \frac{(dw)!}{(dw)^{dw}} \end{aligned} \tag{30}$$

Since in the domain of summation, λ is greater than some positive constant, there exists a $q < 1$ such that $e^{-\lambda}\cosh(\lambda) \le q$. Moreover, we have that $(dw)! \le c(dw)^{dw} e^{-dw} (dn)^{1/2}$, for some constant c. Combining these relationships with (30), we get:

$$\begin{aligned}\mathcal{S}_2 &\le cb^n q^{\nu n} (dn)^{1/2} \sum_{(1-\delta)n\le w\le n} \binom{n}{w} \mathcal{P}^w \frac{q^w (1-q)^{n-w}}{q^n (1-q)^{n\delta}} \\ &\le cb^n q^{\nu n} (dn)^{1/2} \mathcal{P}^{n(1-\delta)} \sum_{(1-\delta)n\le w\le n} \binom{n}{w} \frac{q^w (1-q)^{n-w}}{q^n (1-q)^{n\delta}} \\ &\le c(dn)^{1/2} b^n \mathcal{P}^{n(1-\delta)} \left[\frac{q}{(1-q)^{\delta/(\nu-1)}} \right]^{n(\nu-1)} \end{aligned} \tag{31}$$

For the case of $\nu > 1$, since $q < 1$, the value $\delta/(\nu-1)$ can be made arbitrarily small by choosing a sufficiently small δ, and therefore the value $q/(1-q)^{\delta/(\nu-1)}$ can be made arbitrarily small. Hence for $\nu > 1$ and sufficiently small δ, $\mathcal{S}_2 \to 0$ as n grows.

For the case of $\nu < 1$, $\mathcal{S}_2$ tends to zero under certain conditions, derived in the following. Simplifying (31) leads to the following equation:

$$\mathcal{S}_2 \le c(dn)^{1/2} ((1-q)\mathcal{P})^{-n_0} [b\mathcal{P}q^{\nu-1}]^n, \tag{32}$$

where n_0 is some integer such that $n_0 \leq n\delta$. Hence, in order to prove that $\mathcal{S}_2$ goes to zero as $n \to \infty$, all we need is to have the following inequality satisfied:

$$b\mathcal{P}q^{\nu-1} < 1. \tag{33}$$

Since $q \geq \cosh(\lambda)e^{-\lambda}$, to have a feasible inequality we must have $\cosh(\lambda)e^{-\lambda} = (1+e^{-2\lambda})/2 \leq (1/b\mathcal{P})^{1/(\nu-1)}$. Moreover, since $\lambda = \frac{dw}{\nu n}$ for $n(1-\delta) \leq w \leq n$, as n becomes large and δ small enough, we have that $\lambda \to \frac{dn}{\nu n} = \frac{d}{\nu}$. Therefore, in order for inequality (33) to be feasible we must have that

$$b\mathcal{P} \leq \left[\frac{1+e^{-2d/v}}{2}\right]^{1-\nu}. \tag{34}$$

Equation (34) gives us a relationship between channel parameters, namely b and $\mathcal{P}$, and the code rate. If we denote by $1/\theta_0$ the minimum value of ν for which inequality (34) is satisfied, then for all $\nu > 1/\theta_0$ the last part of the sum goes to zero as n tends to infinity.

Now all that remains to do is to show that (27) holds. From (29), we have that for any $\lambda > 0$,

$$\mathbb{E}[B_w] \;\leq\; \binom{n}{w}\cosh(\lambda)^{\nu n}\frac{(wd)!}{(\lambda\nu n)^{wd}} \leq 2\binom{n}{w}\cosh(\lambda)^{\nu n}\left(\frac{wd}{\lambda\nu n\mathrm{e}}\right)^{wd} wd, \tag{35}$$

for any $\lambda > 0$, and uniformly in w, where the second inequality follows from the estimate $n! \leq 2n(n/\mathrm{e})^n$.

Consider the following two different ways of estimating the above binomial coefficient. The first estimate we use is the well-known one:

$$\binom{n}{w} \leq \left(\frac{n\mathrm{e}}{w}\right)^w.$$

Using this, we obtain

$$b^n\mathbb{E}[B_w]\mathcal{P}^w \leq \left(b\left(\frac{n\mathrm{e}}{w}\right)^{w/n}\cosh(\lambda)^{\nu}\left(\frac{wd}{\lambda\nu n\mathrm{e}}\right)^{\frac{wd}{n}}\sqrt[n]{2wd}\;\mathcal{P}^{\frac{w}{n}}\right)^n.$$

Since $\sqrt[n]{2wd} \to 1$ as n goes to infinity, it suffices to show that the rest of the expression in the paranthesis is less than 1 to prove (27). To do so, set $x = \frac{w}{\nu n}$ and $\phi = 1/\nu$. By taking logarithms, and multiplying by ϕ, we need to show that

$$\phi\ln(b) + x\ln\left(\frac{e\phi}{x}\right) + \ln(\cosh(\lambda)) + xd\ln\left(\frac{xd}{\lambda e}\right) + x\ln\mathcal{P} < 0$$

for $\delta\phi \leq x \leq (1-\delta)\phi$. Since we can choose λ freely, we select λ such that $x = \lambda\tanh(\lambda)/d$. Doing so, we reduce the above to

$$\phi\ln(b) + \frac{\lambda\tanh\lambda}{d}\ln\left(\frac{\phi ed}{\lambda\tanh\lambda}\right) + \ln(\cosh(\lambda)) + \lambda\tanh(\lambda)\ln\left(\frac{\tanh(\lambda)}{e}\right) + \frac{\lambda\tanh(\lambda)}{d}\ln\mathcal{P} < 0.$$

Rewriting the above results in the following sufficient condition for (27) to hold:

$$f(\lambda,\phi) \triangleq \frac{\mathcal{P}\phi ed}{\lambda\tanh\lambda}b^{\frac{d\phi}{\lambda\tanh(\lambda)}}\cosh(\lambda)^{\frac{d}{\lambda\tanh(\lambda)}}\left(\frac{\tanh(\lambda)}{e}\right)^d < 1, \tag{36}$$

for all values of $\lambda > 0$ such that $\lambda\tanh\lambda < d\phi$.

Let θ_1 be the maximum value of ϕ for which $f(\lambda,\phi) < 1$ for all $\lambda > 0$ such that $\lambda\tanh\lambda < d\phi$. Then, for all $\phi < \theta_1$, we have that (27) holds.

Now, in order to derive another sufficient condition for (27) to converge, we apply a different estimate for the binomial coefficient $\binom{n}{w}$, namely, the one in equation (3.5.7) of [3, page 163]. This gives us:

$$\binom{n}{w} = \frac{\varphi}{\sqrt{2\pi x(\varphi-x)\varphi k}}\left(\frac{\varphi^{\varphi}(\varphi-x)^x}{x^x(\varphi-x)^{\varphi}}\right)^k(1+o(1)),$$

where $x = \frac{w}{\nu n}$ and $\phi = 1/\nu$. Choosing λ such that $x = (\lambda \tanh \lambda)/d$, we can write:

$$\begin{aligned} b^n \mathbb{E}[B_w] \mathcal{P}^w &\leq 2cwdb^n \binom{n}{w} \cosh \lambda^{\nu n} \left(\frac{wd}{\lambda e \nu n}\right)^{wd} \mathcal{P}^w \\ &= \frac{2cwd\phi b^n \mathcal{P}^w}{\sqrt{2\pi x(\phi - x)\phi \nu n}} \cosh \lambda^{\nu n} \left(\frac{wd}{\lambda e \nu n}\right)^{wd} \left[\frac{\phi^\phi (\phi - x)^x}{x^x (\phi - x)^\phi}\right]^{\nu n} \\ &\leq \frac{2cwd\phi}{\sqrt{2\pi x(\phi - x)\phi \nu n}} g(\lambda, \phi)^{\nu n} \end{aligned} \tag{37}$$

where c is some constant and

$$g(\lambda, \phi) = \cosh(\lambda) \left(\frac{\tanh(\lambda)}{e}\right)^{\lambda \tanh(\lambda)} \left(\frac{b\phi d}{\phi d - \lambda \tanh(\lambda)}\right)^{\phi} \left(\frac{\phi d - \lambda \tanh(\lambda)}{\lambda \tanh(\lambda)} \mathcal{P}\right)^{\frac{\lambda \tanh(\lambda)}{d}}$$

Hence, in order to show (27), it is sufficient to ensure that $g(\lambda, \phi) < 1$ for all $\lambda > 0$ and ϕ such that $\phi d = d/\nu \geq \lambda \tanh(\lambda) = xd = w/(\nu n)$. If we denote the maximum value of ϕ for which $g(\lambda, \phi) < 1$ and $\phi d \geq \lambda \tanh(\lambda)$ by θ_2, then for all $\phi < \theta_2$, $g(\lambda, \phi)$ is less than 1 in the range of interest of (27).

To sum up, if we choose ν such that $\nu \geq \max[1/\theta_0, \min\{1/\theta_1, 1/\theta_2\}]$, then the entire sum $\mathcal{S}$ in (24) go to zero as n becomes large enough.

In the following, we will use the above analysis to prove Theorem 8 (relating to the convex combination of BSCs), and to complete the proof of Theorem 10 (relating to the AWGN channel). The proof of Theorem 5 (relating to the BSC) will follow as a special case of the proof of Theorem 8.

A. Proof of Theorem 8

Proof: Theorem 8 follows directly from the above analysis by setting $b = 1$ and $\mathcal{P} = \sum_{i=1}^{N} \beta_i (1 - 2p_i)^2$. ■

B. Proof of Theorem 10

Proof: Theorem 10 follows directly from the above analysis with $b = 2^{\mathrm{Cap}(\mathcal{C}) - R}$ and $\mathcal{P} = e^{-\rho}$. ■

APPENDIX B
PROOF OF THEOREM 6

The derivative of $f(\lambda)$ with respect to λ is

$$\frac{\mathrm{d}f}{\mathrm{d}\lambda} = 2 \cosh(\lambda)^{-2 \frac{-\cosh(\lambda) + \sinh(\lambda)\lambda}{\sinh(\lambda)\lambda}} (1 - 2p)^2 g(\lambda),$$

where $g(\lambda)$ equals

$$\frac{-2 \ln(\cosh(\lambda)) \sinh(\lambda) \cosh(\lambda)^2 - 2\lambda \ln(\cosh(\lambda)) \cosh(\lambda) + \cosh(\lambda)^3 \lambda - \cosh(\lambda)\lambda + \lambda^2 \sinh(\lambda)}{\mathrm{e} \sinh(\lambda) \lambda^3}.$$

The sign of the derivative of $f(\lambda)$ is the same as the sign of the numerator of $g(\lambda)$, which is why we are going to concentrate on this qunatity. This numerator equals

$$-2 \ln(\cosh(\lambda)) \cosh(\lambda) (\sinh(\lambda) \cosh(\lambda) + \lambda) + \lambda \cosh(\lambda) \sinh(\lambda)^2 + \lambda^2 \sinh(\lambda)$$

which in turn equals

$$(\sinh(\lambda) \cosh(\lambda) + \lambda) (\lambda \sinh(\lambda) - 2 \cosh(\lambda) \ln(\cosh(\lambda))).$$

We therefore need to show that

$$\lambda \tanh(\lambda) < 2 \ln \cosh(\lambda),$$

for $\lambda > 0$. This is done in two steps. First, using Taylor expansions, it can be shown that

$$\lambda \tanh(\lambda) - 2 \ln \cosh(\lambda) \leq -\frac{\lambda^4}{6} + \frac{4}{45}\lambda^6 - \frac{17}{420}\lambda^8 + \frac{248}{14175}\lambda^{10} =: p(\lambda).$$

The smallest positive root of $p(\lambda)$ is smaller than $r := 1.4409$, and $p(\lambda) < 0$ for $0 < \lambda \leq r$. For $\lambda \geq r$, we use the equality

$$\begin{aligned}
\lambda \tanh(\lambda) - 2 \ln \cosh(\lambda) &= \lambda - \frac{2\lambda}{1+\mathrm{e}^{2\lambda}} - 2 \ln \left(\mathrm{e}^{\lambda} \frac{1-\mathrm{e}^{-2\lambda}}{2} \right) \\
&= -\lambda + 2\ln(2) + \frac{2\lambda}{1+\mathrm{e}^{2\lambda}} - 2\ln(1+\mathrm{e}^{-2\lambda}) \\
&< -\lambda + 2\ln(2) + \frac{2\lambda}{1+\mathrm{e}^{2r}} - 2\ln(1+\mathrm{e}^{-2r}) \\
&< 1.1682 - \lambda \\
&< 0,
\end{aligned}$$

for $\lambda > r$. Altogether, this shows that $g(\lambda) < 0$ for $\lambda > 0$. Therefore, the maximum of $f(\lambda)$ is $\lim_{\lambda \to 0} f(\lambda)$, i.e.,

$$\theta_0 = \frac{1}{\lim_{\lambda \to 0} f(\lambda)}.$$

To calculate the limit, we use Taylor expansions. We have

$$\begin{aligned}
\lambda \tanh(\lambda) &= \lambda^2 + O(\lambda^4) \\
\left(\frac{\tanh(\lambda)}{\mathrm{e}} \right)^2 &= \frac{\lambda^2}{\mathrm{e}^2} + O(\lambda^4) \\
\cosh(\lambda)^{d/\lambda \tanh(\lambda)} &= \mathrm{e} + O(\lambda^2).
\end{aligned}$$

Therefore,

$$\begin{aligned}
f(\lambda) &= \frac{2\mathrm{e}}{\lambda^2}(1+O(\lambda^2))(\mathrm{e}+O(\lambda^2)) \left(\frac{\lambda^2}{\mathrm{e}^2} \left(1+O(\lambda^4)\right) \right) (1-2p)^2 \\
&= 2(1-2p)^2 + O(\lambda^2),
\end{aligned}$$

so that $\lim_{\lambda \to 0} f(\lambda) = 2(1-2p)^2$.

APPENDIX C
PROOF OF THEOREM 7

(1) In this case, $H(D)$ is the dimension k of the subspace, which we call U. We denote the dual space of U by $U^{\perp}$. Then, for $v \notin U^{\perp}$, we have $q_v = 1/2$: in this case, exactly half the vectors in U are orthogonal to v. If $v \in U^{\perp}$, then $q_v = 0$: all the vector in U are orthogonal to v. Therefore, we have

$$\sum_{v \in \mathbb{F}_2^n} (1-2q_v)^2 = \sum_{v \in U^{\perp}} 1 = |U^{\perp}| = 2^{n-k}.$$

The result follows.

(2) We first recall the definition of a Bent function. f is Bent iff

$$\forall u \in \mathbb{F}_2^n : \quad \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+\langle u|x \rangle} = \pm\sqrt{2^n}.$$

Let $N := |S|$. Then, by choosing $u = 0$ in the above equation, we see that $N = 2^{n-1} \pm \sqrt{2^{n-2}}$. If D is the uniform distribution on S, then U assigns the probability

$$p_u := \frac{1}{2N}(1 - (-1)^{f(u)})$$

to $u \in \mathbb{F}_2^n$. Then

$$1 - 2q_v = \frac{1}{2N}\left(\sum_{x\in\mathbb{F}_2^n}(-1)^{\langle v|x\rangle} - \sum_{x\in\mathbb{F}_2^n}(-1)^{f(x)+\langle v|x\rangle}\right).$$

By the definition of Bent functions, we have for $u \neq 0$:

$$(1 - 2q_v) = \frac{\pm\sqrt{2^n}}{2N}.$$

Therefore,

$$\sum_{v\in\mathbb{F}_2^n}(1-2q_v)^2 = \frac{2^{2n}}{4N^2} = \frac{2^{2n}}{2^{2n} + O(\sqrt{2^n})} = 1 + O(1/\sqrt{8^n}).$$

Hence,

$$\log_2\left(\sum_{v\in\mathbb{F}_2^n}(1-2q_v)^2\right) = 1 + O(1/\sqrt{8^n}).$$

On the other hand, $H(D) = \log_2(|S|)$ as D is uniform on S, so

$$n - \log_2(|S|) = n - \log_2(N) = 1 + O(1/\sqrt{2^n}).$$

The result follows now.

APPENDIX D
PROOF OF LEMMA 3

We simplify the output entropy of $n-$uses of the compound channel as

$$\begin{aligned} H(\mathcal{B}, Z') &= H(\mathcal{B}) + H(Z'|\mathcal{B}) \\ &= nH(\beta) + \sum_{S\in[N]^n} \Pr(\mathcal{B} = S) \cdot H(Z'|\mathcal{B} = S). \end{aligned}$$

Notice that conditioned on a particular realization of $\mathcal{B}$, we are back in the framework of a BSC, albeit with different transition probabilities for different channel uses. Using Corollary 1, we bound the above as

$$\begin{aligned} H(\mathcal{B}, Z') &\geq nH(\beta) + \sum_{S\in[N]^n}\left(\prod_{i=1}^n \beta_{s_i}\right)\left[n - \log_2\left(\sum_{V\in C^\perp}\prod_{\{i|v_i=1\}}(1-2p_{s_i})^2\right)\right] \\ &= n(1+H(\beta)) - \sum_{S\in[N]^n}\left(\prod_{i=1}^n \beta_{s_i}\right)\log_2\left(\sum_{V\in C^\perp}\prod_{\{i|v_i=1\}}(1-2p_{s_i})^2\right), \end{aligned}$$

since $\sum_{S\in[N]^n}\left(\prod_{i=1}^n \beta_{s_i}\right)=\left(\sum_{i=1}^N \beta_i\right)^n=1$. We define the notations $S_{\mathfrak{I}}\triangleq\{s_i|i\in\mathfrak{I}\}$, and $\mathfrak{I}_V=\{i|v_i=1\}$. Notice that $|\mathfrak{I}_V|=w_H(V)$. Using Jensen's inequality, we obtain

$$\begin{aligned}
H(\Lambda,Z') &\geq n(1+H(\lambda))-\log_2\left[\sum_{S\in[N]^n}\left(\prod_{i=1}^n\lambda_{s_i}\right)\sum_{V\in C^\perp}\prod_{\mathfrak{I}_V}(1-2p_{s_i})^2\right]\\
&= n(1+H(\lambda))-\log_2\left[\sum_{V\in C^\perp}\sum_{S\in[N]^n}\left(\prod_{i=1}^n\lambda_{s_i}\right)\prod_{\mathfrak{I}_V}(1-2p_{s_i})^2\right]\\
&= n(1+H(\lambda))-\log_2\left[\sum_{V\in C^\perp}\sum_{S_{[N]\setminus\mathfrak{I}_V}\in[N]^{n-w_H(V)}}\left(\prod_{i\in[N]\setminus\mathfrak{I}_V}\lambda_{s_i}\right)\right.\\
&\qquad\left.\cdot\sum_{S_{\mathfrak{I}_V}\in[N]^{w_H(V)}}\left(\prod_{i\in\mathfrak{I}_V}\lambda_i(1-2p_{s_i})^2\right)\right]\\
&= n(1+H(\lambda))-\log_2\left[\sum_{V\in C^\perp}\left[\sum_{i=1}^N\lambda_i(1-2p_i)^2\right]^{w_H(V)}\underbrace{\sum_{S_{[N]\setminus\mathfrak{I}_V}\in[N]^{|S_{[N]\setminus\mathfrak{I}_V}|}}\left(\prod_{i\in[N]\setminus\mathfrak{I}_V}\lambda_{s_i}\right)}_{=\left[\sum_{i=1}^N\lambda_i\right]^{n-w_H(V)}=1}\right].
\end{aligned}$$

Defining $\mathcal{P}\triangleq\sum_{i=1}^N\beta_i(1-2p_i)^2$, we obtain that

$$H(\mathcal{B},Z')\geq n(1+H(\beta))-\log_2\left[\sum_{w=0}^n B_w\mathcal{P}^w\right],$$

which concludes the proof of Lemma 3.

APPENDIX E
RELATIONSHIP BETWEEN THE THRESHOLDS OF INDIVIDUAL BSCS AND THEIR CONVEX COMBINATION (PROOF OF THEOREM 9)

In this appendix, we investigate the relationship between the thresholds of individual BSCs and that of their convex combinations. Denote the threshold $\alpha(d,\mathrm{BSC}(p_i))$ of the BSC channel with cross-over probability of p_i to be α_i. From theorem 5 we know that for all $\phi<\alpha_i$ we have:

$$f_i(\lambda,\phi)=\frac{(1-2p_i)^2\phi ed}{\lambda\tanh\lambda}\cosh(\lambda)^{\frac{d}{\lambda\tanh(\lambda)}}\left(\frac{\tanh(\lambda)}{e}\right)^d<1$$

or

$$g_i(\lambda,\phi) = \cosh(\lambda)\left(\frac{\phi d}{\phi d-\lambda\tanh(\lambda)}\right)^{\phi}\left(\frac{\tanh\lambda}{e}\right)^{\lambda\tanh\lambda}\left(\frac{\phi d-\lambda\tanh(\lambda)}{\lambda\tanh(\lambda)}(1-2p_i)^2\right)^{\frac{\lambda\tanh(\lambda)}{d}}<1.$$

Rewriting the above equations results in the following inequalities:

$$(1-2p_i)^2<\frac{1}{\alpha_i}\ \frac{\lambda\tanh\lambda}{ed(\cosh\lambda)^{\frac{d}{\lambda\tanh(\lambda)}}}\left(\frac{e}{\tanh\lambda}\right)^d$$

or

$$(1-2p_i)^2 \;<\; \frac{1}{\alpha_i d}\frac{\lambda\tanh(\lambda)}{1-\lambda\tanh(\lambda)/(\alpha_i d)}(\cosh\lambda)^{-\frac{d}{\lambda\tanh(\lambda)}}\left(\frac{e}{\tanh(\lambda)}\right)^d\left(1-\frac{\lambda\tanh(\lambda)}{\alpha_i d}\right)^{\frac{\alpha_i d}{\lambda\tanh(\lambda)}},$$

for all $\lambda\tanh(\lambda)\leq d\,\phi$ and $\lambda\in(0,\infty)$.

Consider a probability distribution β and define $\mathcal{P}=\sum_{i=1}^N\beta_i(1-2p_i)^2$. From the above inequalities we have that

$$\mathcal{P}<\frac{\lambda\tanh\lambda}{ed\cosh(\lambda)^{\frac{d}{\lambda\tanh(\lambda)}}}\left(\frac{e}{\tanh(\lambda)}\right)^d\sum_{i=1}^N\frac{\beta_i}{\alpha_i} \tag{38}$$

or

$$\mathcal{P} \;<\; \sum_{i=1}^N\frac{\beta_i}{\alpha_i d}\;\frac{\lambda\tanh(\lambda)}{1-\lambda\tanh(\lambda)/(\alpha_i d)}(\cosh(\lambda))^{-\frac{d}{\lambda\tanh(\lambda)}}\left(\frac{e}{\tanh(\lambda)}\right)^d\left(1-\frac{\lambda\tanh(\lambda)}{\alpha_i d}\right)^{\frac{\alpha_i d}{\lambda\tanh(\lambda)}}. \tag{39}$$

Now consider the convex combination of the mentioned BSC channels, CBSC(P,β), as defined in Section III. From Theorem 8, we know that for all $\phi<\alpha_C$, where α_C is the threshold for the convex combination of BSC channels, we have the following:

$$f_C(\lambda,\phi)\triangleq\frac{\mathcal{P}\phi ed}{\lambda\tanh\lambda}\cosh(\lambda)^{\frac{d}{\lambda\tanh(\lambda)}}\left(\frac{\tanh(\lambda)}{e}\right)^d<1$$

and

$$g_C(\lambda,\phi)=\cosh(\lambda)\left(\frac{\phi d}{\phi d-\lambda\tanh(\lambda)}\right)^{\phi}\left(\frac{\tanh\lambda}{e}\right)^{\lambda\tanh\lambda}\left(\frac{\phi d-\lambda\tanh(\lambda)}{\lambda\tanh(\lambda)}\mathcal{P}\right)^{\frac{\lambda\tanh(\lambda)}{d}}<1,$$

with $\mathcal{P}$ defined as before. Rewriting the above inequalities we find that:

$$\mathcal{P}<\frac{1}{\alpha_C}\frac{\lambda\tanh\lambda}{ed}\cosh(\lambda)^{-\frac{d}{\lambda\tanh(\lambda)}}\left(\frac{e}{\tanh(\lambda)}\right)^d \tag{40}$$

and

$$\mathcal{P}<\frac{1}{\alpha_C d}\;\frac{\lambda\tanh(\lambda)}{1-\lambda\tanh(\lambda)/(\alpha_C d)}\left(\frac{e}{\tanh(\lambda)}\right)^d(\cosh(\lambda))^{-\frac{d}{\lambda\tanh(\lambda)}}\left(1-\frac{\lambda\tanh(\lambda)}{\alpha_C d}\right)^{\frac{\alpha_C d}{\lambda\tanh(\lambda)}}. \tag{41}$$

From equation (38) and (39), we see that in order to satisfy equations (40) and (41), it is sufficient to choose α_C such that

$$\frac{\lambda\tanh\lambda}{ed}\cosh(\lambda)^{-\frac{d}{\lambda\tanh(\lambda)}}\left(\frac{e}{\tanh(\lambda)}\right)^d\sum_{i=1}^N\frac{\beta_i}{\alpha_i}\leq\frac{1}{\alpha_C}\frac{\lambda\tanh\lambda}{ed}\cosh(\lambda)^{-\frac{d}{\lambda\tanh(\lambda)}}\left(\frac{e}{\tanh(\lambda)}\right)^d,$$

and

$$\begin{aligned}&\sum_{i=1}^N\frac{\beta_i}{\alpha_i d}\frac{\lambda\tanh(\lambda)}{1-\lambda\tanh(\lambda)/(\alpha_i d)}\left(\frac{e}{\lambda\tanh(\lambda)}\right)^d(\cosh(\lambda))^{-\frac{d}{\lambda\tanh(\lambda)}}\left(1-\frac{\lambda\tanh(\lambda)}{\alpha_i d}\right)^{\frac{\alpha_i d}{\lambda\tanh(\lambda)}}\\ \leq\;&\frac{1}{\alpha_C d}\frac{\lambda\tanh(\lambda)}{1-\lambda\tanh(\lambda)/(\alpha_C d)}\left(\frac{e}{\lambda\tanh(\lambda)}\right)^d(\cosh(\lambda))^{-\frac{d}{\lambda\tanh(\lambda)}}\left(1-\frac{\lambda\tanh(\lambda)}{\alpha_C d}\right)^{\frac{\alpha_C d}{\lambda\tanh(\lambda)}}.\end{aligned}$$

The first inequality leads to:

$$\overline{R}<R_C \tag{42}$$

In which $R_i=1/\alpha_i$ is the code rate threshold for the i^{th} BSC channel, R_C is the same threshold for the convex combination of BSC channels and $\overline{R}=\sum_{i=1}^N\beta_i R_i$.

The second inequality simplifies to

$$\sum_{i=1}^{N} \frac{\beta_i}{\alpha_i} \left(1 - \frac{\lambda \tanh(\lambda)}{\alpha_i d}\right)^{\frac{\alpha_i d}{\lambda \tanh(\lambda)} - 1} \leq \frac{1}{\alpha_C} \left(1 - \frac{\lambda \tanh(\lambda)}{\alpha_C d}\right)^{\frac{\alpha_C d}{\lambda \tanh(\lambda)} - 1}. \tag{43}$$

Denoting $\frac{\alpha_i d}{\lambda \tanh(\lambda)}$ by u_i and $\frac{\alpha_c d}{\lambda \tanh(\lambda)}$ by u_C, we note that u_i and u_C are all greater than 1, since the range of interest in Theorem 8 is restricted to all $\lambda > 0$ such that $\lambda \tanh \lambda \leq d\phi$. Therefore, the function $f(u) = (1 - \frac{1}{u})^{u-1}$ is always less than 1 for $u > 1$. Moreover, $f(u)$ is always greater than e^{-1}. Therefore, to satisfy (43), it is sufficient to have the following inequality:

$$\sum_{i=1}^{N} \frac{\beta_i}{\alpha_i} \leq \frac{1}{e\alpha_C}$$

Equivalently, it is sufficient to have that

$$\overline{R} \leq \frac{R_C}{e}. \tag{44}$$

Since R_C should satisfy both of the inequalities (42) and (44), we may choose $R_C = e\overline{R}$.

In addition to the above relationship, from Theorem 8, we see that one must choose α_C such that

$$\left(\frac{1 + e^{-2\alpha_C d}}{2}\right)^{\left(1 - \frac{1}{\alpha_C}\right)} > \mathcal{P}. \tag{45}$$

On the other hand, from Theorem 5, we observe that the following relationship holds for a BSC channel with cross-over probability p_i:

$$\left(\frac{1 + e^{-2\alpha_i d}}{2}\right)^{\left(1 - \frac{1}{\alpha_i}\right)} > (1 - 2p_i)^2.$$

By multiplying the above inequality with β_i and summing up from $i = 1$ to N, we find out that in order to satisfy (45), it is sufficient to choose α_C such that

$$\left(\frac{1 + e^{-2\alpha_C d}}{2}\right)^{\left(1 - \frac{1}{\alpha_C}\right)} > \sum_{i=1}^{N} \beta_i \left(\frac{1 + e^{-2\alpha_i d}}{2}\right)^{\left(1 - \frac{1}{\alpha_i}\right)}$$

$$\Leftrightarrow e^{-\alpha_C d} \left(\cosh(\alpha_C d)\right)^{\left(1 - \frac{1}{\alpha_C}\right)} > \sum_{i=1}^{N} \beta_i e^{-\alpha_i d} \left(\cosh(\alpha_i d)\right)^{\left(1 - \frac{1}{\alpha_i}\right)}.$$

Since $\cosh(x) \geq 1$ for all x, it is sufficient to have:

$$\begin{aligned} e^{-\alpha_C d} &> \sum_{i=1}^{N} \beta_i e^{-\alpha_i d} \cosh(\alpha_i d) \\ &= \sum_{i=1}^{N} \beta_i \frac{1 + e^{-2\alpha_i d}}{2}, \end{aligned}$$

which simplifies to $R_C > R_0$, where

$$R_0 = \frac{d}{\ln\left(\frac{2}{1 + \sum_{i=1}^{N} \beta_i e^{-2d/R_i}}\right)} \tag{46}$$

Therefore, the relationship between the thresholds of BSCs and their convex combination is given by the inequality $R_C = \max\{R_0, e\overline{R}\}$.

REFERENCES

[1] J. Blömer, R. Karp, and E. Welzl, "The rank of sparse random matrices over finite fields," *Random Structures and Algorithms*, vol. 10, no. 4, pp. 407-419,1997.

[2] N. Calkin, "Dependent sets of constant weight binary vectors," *Combinatorics, Probability, and Computing*, vol. 6, pp. 49-53, 2003.

[3] V.F. Kolchin, *Random Graphs*, Number 53 in Encyclopedia of Mathematics and its Applications. Cambrdige University Press, 1999.

[4] O. Etesami and A. Shokrollahi, "Raptor codes on binary memoryless symmetric channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2033-2051, 2006.

[5] P. Pakzad and A. Shokrollahi, "EXIT functions for LT and raptor codes, and asymptotic ranks of random matrices," in *Proc. of the IEEE International Symposium on Information Theory (ISIT)*, June 2007, pp. 411-415.

[6] S.J. MacMullan and O.M. Collins, "The capacity of binary channels that use linear codes and decoders," *IEEE Trans. Inform. Theory*, vol. 44, pp. 197-214, 1998.

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

[8] T. Richardson, A. Shokrollahi, and R. Urbanke, "Finite-length analysis of various low-density parity-check ensembles for the binary erasure channel," in *Proc. of the IEEE International Symposium on Information Theory (ISIT)*, pp. 1, 2002.

[9] D. Burshtein, M. Krivelevich, S. Litsyn, and G. Miller, "Upper bounds on the rate of ldpc codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2437-2449, 2002.

[10] I. Sason and R. Urbanke, "Parity-check density versus performance of binary linear block codes over memoryless symmetric channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1611-1635, 2003.

[11] T. J. Richardson, M. A. Shokrollahi and R. L. Urbanke, "Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619-637, 2001.

[12] J. Wolfowitz, *Coding Theorems of Information Theory*, Number 31 in a Series of Modern Surveys in Mathematics, Springer Verlag, 1978.