Phase Transitions for Mutual Information

K. Raj Kumar*, Payam Pakzad†, Amir Hesam Salavati* and Amin Shokrollahi*
*Laboratoire d'algorithmique
Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
Email: {raj.kumar,hesam.salavati,amin.shokrollahi}@epfl.ch
†Qualcomm, Inc., 3195 Kifer Road, Santa Clara, CA 95051, USA
Email: payam@qualcomm.com

***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 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.**

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.

For this setting, the primary question that we would like to answer is the following: under the assumption of long block-lengths, 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 few important and practical classes of channels, including the binary symmetric channel (BSC) 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.

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

978-1-4244-6746-4/10/$26.00 ©2010 IEEE

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

G. Hence in this case, we have that $\Pi_{d,\mathcal{I}} = \Pr\{\mathrm{rk}(G) < n\}$. Using the union bound, one can show that

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

where $\mathrm{lker}(G)$ denotes the left kernel of the matrix G. 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].

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 $\mathrm{E}[|\mathrm{lker}(G)|] \to 1$, and if $\alpha > \alpha(d,\mathcal{I})$, then $\mathrm{E}[|\mathrm{lker}(G)|] \to \infty$.

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.

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.

II. THE CASE $\mathcal{C} = \mathrm{BSC}(p)$

In this section, we study the case of a BSC with crossover probability p, denoted by $\mathcal{C} = \mathrm{BSC}(p)$. 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

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

[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.

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. ∎

We omit the proof of the following corollary for lack of space.

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. 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.

The proof of Theorem 3 follows from the previous corollary. For lack of space, we confine ourselves to providing a sketch only.

Proof: (Of Theorem 3 - Sketch) 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: $\mathbb{E}[\log_2(U)] \leq \log_2(\mathbb{E}[U])$, hence it suffices to prove that

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

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 inequality of Theorem 4 cannot be improved since it is tight for flat distributions. In other words, if D is a uniform distribution over a k-dimensional subspace of $\mathbb{F}_2^n$, where $1 \leq k \leq n$, then equality is achieved.

Using the above results, we obtain the following theorem whose proof is omitted for lack of space.

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

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

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 $u(\phi) < p$. 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

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

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

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

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 $\mathrm{Cap}(\mathrm{BSC}(p))\alpha(d,\mathrm{BSC}(p)) = (1-h(p))\alpha(d,\mathrm{BSC}(p))$ 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 $1/2$

$d \backslash p$	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 $(1-h(p))\alpha(d, \text{BSC}(p))$ FOR VARIOUS VALUES OF d AND p.

when p converges to $1/2$. This suggests that there is room for improvement in the bounds of Theorem 5.

It can be shown that the results for the BSC also extend to results for the convex combination of BSCs. Details are omitted for brevity.

III. 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 $W \sim \text{i.i.d. } \mathcal{N}\left(0, \frac{1}{\rho} I\right)$. 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. As before, we first develop a lower bound on the mutual information between the input and output.

$$\begin{aligned} I(X;Z) &= H(Z) - H(Z|X) \\ &= H(Z) - \frac{1}{2}\log\frac{(2\pi e)^n}{\rho^n}. \end{aligned} \tag{3}$$

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

$$H(Z) \geq -\log\left[\int p^2(Z)\, dZ\right], \tag{4}$$

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

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

A simple manipulation yields

$$\begin{aligned} \int p^2(Z)\, dZ &= \frac{\rho^n}{(2\pi)^n 2^{2k}} \sum_{i,j=1}^{2^k} \int \exp\left\{-\frac{\rho}{2}\left[\frac{|Y_i - Y_j|^2}{2}\right.\right. \\ &\quad \left.\left. +2\left|Z - \frac{Y_i+Y_j}{2}\right|^2\right]\right\} dZ \\ &= \frac{(\rho\pi)^{n/2}}{(2\pi)^n 2^{2k}} \sum_{i,j=1}^{2^k} \exp\left\{-\frac{\rho}{4}|Y_i - Y_j|^2\right\}. \end{aligned} \tag{5}$$

From (3), (4) and (5), 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:

$$I(X;Z) \geq -\log\left[\left(\frac{e}{2}\right)^{n/2} \frac{1}{2^k} \sum_{w=0}^{n} C_w e^{-w\rho}\right]. \tag{6}$$

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

$$\begin{aligned} &\Pr\{I(X;Z) < n(\text{Cap}(\mathcal{C}) - \epsilon)\} \leq \\ &\Pr\left\{-\log\left[\left(\frac{e}{2}\right)^{n/2} \frac{1}{2^k} \sum_{w=0}^{n} C_w e^{-w\rho}\right] < n(\text{Cap}(\mathcal{C}) - \epsilon)\right\} \\ &= \Pr\left\{\left(\sqrt{\frac{e}{2}} 2^{\text{Cap}(\mathcal{C}) - R}\right)^n \sum_{w=0}^{n} C_w e^{-w\rho} > 2^{n\epsilon}\right\}, \end{aligned} \tag{7}$$

where ϵ is a constant independent of n, k. An analysis of the term $\mathcal{S} \triangleq b^n \sum_{w=1}^{n} C_w e^{-\rho w}$, where $b = \sqrt{\frac{e}{2}} 2^{\text{Cap}(\mathcal{C}) - R}$ is a constant that is independent of n leads to the following theorem.

Theorem 6: Let $d \geq 3$ be a fixed constant, and $\mathcal{C} = \text{AWGN}(\rho)$. Define

$$\begin{aligned} f(\lambda, \theta) = \ln\left[b^{\frac{\theta}{\theta-1}} \left(\frac{ed\theta}{(\theta-1)\lambda\tanh\lambda}\right)^{\frac{\lambda\tanh\lambda}{d}}\right] - \frac{\rho\lambda\tanh\lambda}{d} \\ + \ln\cosh\lambda + \lambda\tanh\lambda\ln\left(\frac{\tanh\lambda}{e}\right) \end{aligned} \tag{8}$$

and

$$g(\lambda,\theta) \triangleq \frac{d}{\lambda\tanh\lambda}\ln\left[\left(\frac{\tanh\lambda}{e}\right)^{\lambda\tanh\lambda} b^{\frac{\theta}{\theta-1}}\cdot \frac{(\theta d)^{\frac{\theta}{\theta-1}}\left(d\frac{\theta}{\theta-1}-\lambda\tanh\lambda\right)^{\frac{\lambda\tanh\lambda}{d}-\frac{\theta}{\theta-1}}\cosh\lambda}{(\theta-1)^{\frac{\theta}{\theta-1}}(\lambda\tanh\lambda)^{\frac{\lambda\tanh\lambda}{d}}}\right]-\rho. \quad (9)$$

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

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

Proof: (Sketch) In order to analyze C_w, we consider two scenarios. When $n \ge k$, 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 $n \ge k$.

1) Channel coding when $n \ge k$: We define our channel coding ensemble in the following manner. Choose the parity check matrix $H \in \mathbb{F}_2^{(n-k)\times n}$ from the ensemble $\mathcal{E}_d$. The quantity C_w is equal to the number of vectors with weight w in the right kernel of H, or in the left kernel of H^{T}. Along the lines of the proof of Theorem 3.5.1 in Sec. 3.6 of [3], we analyze $\mathcal{S}$ by splitting the sum into three regions, $\mathcal{S}_1 = b^n\sum_{1\le w<\delta n} C_w e^{-\rho w}$, $\mathcal{S}_2 = b^n\sum_{\delta n\le w<(1-\delta)n} C_w e^{-\rho w}$ and $\mathcal{S}_3 = b^n\sum_{(1-\delta)n\le w\le n} C_w e^{-\rho w}$, where $\delta\to 0$. We outline the analysis of these terms in the sequel, omitting details for lack of space.

a) Analysis of $\mathcal{S}_1$: From the proof of Lemma 3.5.1 in [3], we can show that $\mathcal{S}_1 \le \epsilon'$, for some $\delta>0$, where $\epsilon'>0$ is an arbitrary constant.

b) Analysis of $\mathcal{S}_2$: Define $\theta = n/k$. Along the lines of the analysis in [3], it can be shown that $\mathcal{S}_2$ vanishes if for all non-negative λ such that $\lambda\tanh\lambda \le \frac{d\theta}{\theta-1}$, either $f(\lambda,\theta)<0$ or $g(\lambda,\theta)<0$, where the functions f and g are as defined in (8), (9).

c) Analysis of $\mathcal{S}_3$: Along the lines of the proof of Lemma 3.5.2 in [3], we can show that $\mathcal{S}_3\to 0$ when we set $\epsilon \ge \log_2\left(\sqrt{\frac{e}{2}}\right)$ bits.

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 (7) as

$$\Pr\{I(X;Z) < n(\mathrm{Cap}(C)-\epsilon)\} \le \Pr\left\{b2^{\frac{k}{n}-1}(1+e^{-\rho}) > 2^{\epsilon}\right\}.$$

$d \setminus \rho$(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.

It can be show that if one sets $\epsilon \ge \log_2\left(\sqrt{\frac{e}{2}}\right)$, the right-hand side converges to zero. ■

Shown in Table III are the values of $\mathrm{Cap}(\mathrm{AWGN}(\rho))\theta(d,\mathrm{AWGN}(\rho))$ 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 6 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.

IV. ACKNOWLEDGEMENTS

This work was supported by Grant 228021-ECCSciEng of the European Research Council. 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.

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, Cambridge 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)*, pp. 411-415, June 2007.

[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.