Bounds on the Maximum Likelihood Decoding Error Probability of Low Density Parity Check Codes[*]

Gadi Miller and David Burshtein

Dept. of Electrical Engineering Systems

Tel-Aviv University

Tel-Aviv 69978, Israel

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

Abstract

We derive both upper and lower bounds on the decoding error probability of ML decoded LDPC codes. The results hold for any binary-input symmetric-output channel. Our results indicate that for various appropriately chosen ensembles of LDPC codes, reliable communication is possible up to channel capacity. However, the ensemble averaged decoding error probability decreases polynomially, and not exponentially. The lower and upper bounds coincide asymptotically, thus showing the tightness of the bounds. However, for ensembles with suitably chosen parameters, the error probability of almost all codes is exponentially decreasing, with an error exponent that can be set arbitrarily close to the standard random coding exponent.

Index Terms - Code ensembles, Error exponent, Low density parity check (LDPC) codes.

I Introduction

Low density parity check (LDPC) codes have attracted a great deal of interest recently. LDPC codes were originally suggested by Gallager [3] in 1963. However, they became popular only after the introduction of turbo-codes [1].

There are two types of results regarding the performance of LDPC codes. The first concerns the properties of these codes under the assumption of optimal (Maximum Likelihood, ML) decoding [3], [7], [10]. The second relates to the properties of practical iterative decoding algorithms for these codes [3], [13], [12], [6], [9].

[*] IEEE Transactions on Information Theory, volume 47, no. 7, pp. 2696-2710, November 2001

In this paper we consider the first problem. Analysis of the code under the assumption of ML decoding is important from several aspects. The analysis of optimal decoding sets upper limits on the performance of any decoding algorithm. Hence the performance of optimal (ML) decoding may be compared to the performance of the practical iterative decoding algorithms, thus making it possible to assess the gap between the two. Optimal decoding analysis is also expected to improve our intuition and understanding regarding the code and the setting of its parameters.

In [3][Theorem 2.4] Gallager obtained an upper bound on the probability that the minimum distance, $d_{\min}$, of a code in the ensemble is smaller then some value, x ($\Pr(d_{\min} < x)$). For small values of x this bound is dominated by a polynomial N^{2-j} (where j is a parameter of the ensemble and N is the block length). Motivated by this result, Gallager then proceeded to eliminate possible bad codes by expurgation, considering only this enhanced ensemble for decoding error calculation. In this paper we consider two other ensembles and give conditions under which the decoding error probability of the unexpurgated ensembles is proportional to N^{2-j}, by showing that an upper bound on the decoding error probability asymptotically coincides with a lower bound. This result is then contrasted with a third ensemble, for which the asymptotic averaged probability of error is one. The results hold for any binary-input symmetric-output channel. To derive our upper bounds we prove a new upper bound on the probability of error of an ensemble of codes in terms of the average spectrum of the ensemble. We then apply this bound to the two ensembles. The first ensemble was proposed by MacKay [7]. The second ensemble is based on bipartite regular graphs, and was used by several researchers, e.g. [9], [12]. We conclude that for appropriately chosen LDPC ensembles, reliable communication is possible up to channel capacity. However, the ensemble averaged decoding error probability decreases polynomially, and not exponentially. For the expurgated ensembles, when the ensemble parameters are suitably chosen, the error exponent can be set arbitrarily close to the standard random coding exponent. For certain ensemble parameters Gallager's bound on the expurgated ensemble is tighter while for others our bound is better. To obtain our results we derive bounds on the average spectra of various ensembles. Recently, Litsyn and Shevelev [5] derived the exact asymptotic spectra for these and other ensembles.

The paper is organized as follows. In Section II we derive a general upper bound on the probability of decoding error of an ensemble of linear codes, formulated in terms of the average spectrum of the ensemble. In Section III we use this result to derive upper bounds on the probability of error of two LDPC ensembles. In Section IV we derive lower bounds for these ensembles. In Section V we consider the expurgated ensembles and compare our bound to Gallager's bound.

Section VI concludes the paper.

II An upper bound on the decoding error probability

Consider a discrete memoryless channel (DMC) which is characterized by the probability distribution function $P(y|x)$. Here $x \in \mathcal{X}$ is the input to the channel, and $y \in \mathcal{Y}$ is the output of the channel. We assume a binary-input symmetric-output channel. Thus $\mathcal{X} = \{0, 1\}$, and without loss of generality we assume that $\mathcal{Y} \subseteq \mathcal{R}$ such that $P(y|0) = P(-y|1)$ for any $y \in \mathcal{Y}$.

$P_N(\mathbf{y}|\mathbf{x})$ is the probability distribution function for sequences of length N. Since the channel is memoryless, we have

$$P_N(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^{N} P(y_n|x_n). \tag{1}$$

In fact all our results hold also for the continuous output case. In that case $P(y|0) = P(-y|1)$ is a probability density function and $\sum_y(\cdot) \to \int_y(\cdot)dy$ everywhere. However for convenience we formulate the results for the discrete output case.

Let $\mathcal{C}$ denote an ensemble of parity check codes, each represented by its parity check matrix, $A_{L\times N}$, with L rows and N columns, $N \geq L$, such that $\mathbf{x}$ (of dimension N) is a codeword if and only if it satisfies $A\mathbf{x} = 0$. Let M denote the number of different codewords in the code, and let $R \stackrel{\triangle}{=} 1 - L/N$. The rate of the code is $R' \stackrel{\triangle}{=} (\log M)/N$ (throughout the paper we use the convention $\log \equiv \log_2$ and $\ln \equiv \log_e$). Note that $R' \geq R$. The last inequality accounts for possible redundancy in the constraints defined by the parity check matrix. Given some ensemble $\mathcal{C}$ of codes, we define the average weight distribution (average number of words of weight l, or average spectrum) by $\overline{S}_l = \sum_{C\in\mathcal{C}} \Pr(C)S_l(C)$ where $\Pr(C)$ is the probability of the code C in the ensemble, and $S_l(C)$ is the number of words of weight l in code C. Finally, $w(\mathbf{x})$ denotes the weight of the binary vector $\mathbf{x}$ and $d(\mathbf{x}, \mathbf{y})$ is the Hamming distance between the binary vectors $\mathbf{x}$ and $\mathbf{y}$.

In Appendix A we prove the following theorem.

Theorem 1 *Consider an ensemble $\mathcal{C}$ of linear codes, where each code is comprised of M codewords of length N. Denote by $\overline{S}_l$ the average number of words of weight l in a code in the ensemble. Let $P(y|x)$ be a discrete memoryless binary-input symmetric-output channel, and let $D \stackrel{\triangle}{=} \sum_y \sqrt{P(y|0)P(y|1)}$. Denote the ensemble averaged maximum likelihood (ML) decoding error probability by $\overline{P}_e$. Let $U \subseteq \{1, 2, ..., N\}$ and denote the complementary set $\{1, 2, ..., N\} \setminus U$ by U^c. Then:*

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

where

$$\alpha = \max_{l \in U^c} \frac{\overline{S}_l}{M-1} \frac{2^N}{\binom{N}{l}} \tag{3}$$

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

$$\begin{aligned} E_r(R) &\triangleq \max_{Q} \max_{0 \le \rho \le 1} \{E_0(\rho, Q) - \rho R\}, \\ E_0(\rho, Q) &\triangleq -\log \sum_{y \in \mathcal{Y}} \left[\sum_{x=0}^{1} Q(x) P(y|x)^{1/(1+\rho)} \right]^{1+\rho} \end{aligned}$$

($Q = (Q(0), Q(1))$ *is some binary probability distribution).*

Notes: Our bound (2) is comprised of two parts. The first is a union bound that utilizes the Bhattacharyya distance. Typically U is chosen as $\{1, 2, \ldots, \Gamma\}$ such that the set of codewords $\{\mathbf{x}|w(\mathbf{x}) \in U\}$ constitutes a Hamming sphere. The second part is obtained by applying the method recently proposed by [11]. Essentially, this method enlarges the given ensemble to a bigger one by using a randomization technique. The new ensemble has the same probability of error as the original one. However the probability of error of the new ensemble can be bounded using a derivation which is similar to that used by Gallager [4] for deriving the random coding exponent.

Note that the bound proposed by Poltyrev [8] is also comprised of two parts, where the first corresponds to a Hamming sphere in the noise space. Sason and Shamai [10] applied Poltyrev's bound to LDPC codes in order to obtain improved bounds on the probability of decoding error. However, they did not obtain explicit properties for LDPC codes, such as the dependence of the probability of error on the block length, or the maximal rate for which reliable communication is possible.

Finally note that the second part of (2) approaches $2^{-NE_r(R)}$ provided that the average spectrum $\overline{S}_l$ of the ensemble approaches the average spectrum of the random code used by Gallager [4] in his proof, for $l \in U^c$ (more precisely, the ratio of these spectra is required to approach one uniformly in that range).

III Upper bounds for low density parity check codes

We now focus on ensembles of low density parity check (LDPC) codes. A sequence of LDPC ensembles, $\{\mathcal{C}_n\}_{n=1,2,\ldots}$, with block length $N = n$ and rate R' has the property that each code in $\mathcal{C}_n$ has a parity check matrix such that the expected weight of each column and each row is bounded by two constants independent of n.

We consider two LDPC ensembles. In the first ensemble, the columns of the parity check matrix are independent, identically distributed. This ensemble was suggested by MacKay [7]. The second ensemble is based on bipartite regular graphs and was used e.g. by [9], [12].

III.1 Independent matrix columns

In this section we consider the ensemble of parity check matrices $A_{L\times N}$ defined by applying the following procedure to each column of A. First set the entire column to 0's. Then, repeat the following procedure t times. Draw an index uniformly and independently from $\{1, 2, ..., L\}$ and flip the corresponding bit. That is, if an index is chosen an odd number of times then the corresponding element will be one at the end of the procedure. Otherwise the element will be zero. Note that the resulting column weight is even if t is even and odd otherwise. In the following, $h(x)$ and $h_2(x)$ are the entropy functions

$$\begin{aligned} h(x) &= -x\ln x - (1-x)\ln(1-x) \qquad 0 \le x \le 1, \\ h_2(x) &= -x\log x - (1-x)\log(1-x) \qquad 0 \le x \le 1. \end{aligned}$$

The notation $a_N = O(N)$ implies that there exists some K such that $0 \le a_N \le KN$ for N large enough. We use the notation $a_N = o(N)/N$ if $0 \le a_N \to 0$ as $N \to \infty$. Due to the correspondence between codes and their respective parity-check matrices, we use the term "an ensemble of parity check matrices" interchangeably with the ensemble of codes corresponding to these matrices.

Theorem 2 *Consider the ensemble of binary parity check matrices $A_{L\times N}$ described above. Let $\overline{P}_e$ be the ensemble averaged probability of decoding error. Let $R \stackrel{\triangle}{=} 1 - L/N$, $t \ge 3$ and*

$$\beta = (1-R)\frac{2}{t}e^{-12-K} \quad , \tag{4}$$

where

$$K = \frac{6\ln\left(t/(1-R)\right)}{t} \quad . \tag{5}$$

If there exists $0 < \gamma < 1/2$ such that

$$\max_{\beta \le y \le \gamma}\left\{h_2(y) - (1-R) + \max_{0\le x\le 1/2}\{x\log D + (1-R)h_2(x) + yt\log(1-2x)\}\right\} < 0 \tag{6}$$

and

$$R + G(R, \gamma t) < C, \tag{7}$$

where

$$G(R,\gamma t) \stackrel{\triangle}{=} \max_{0\le x\le 1/2}\left\{(1-R)h_2(x) + \gamma t\log(1-2x)\right\}, \tag{8}$$

then

$$\overline{P}_e \leq \begin{cases} N^{1-\frac{t}{2}} \frac{D}{1-D} \frac{(\frac{t}{2})^t}{\frac{t}{2}!} (1-R)^{-\frac{t}{2}} \left(1 + \frac{o(N)}{N}\right) & t \text{ even} \\ \\ N^{2-t} \frac{D^2}{2(1-D^2)} \frac{t^{2t}}{t!} (1-R)^{-t} \left(1 + \frac{o(N)}{N}\right) & t \text{ odd} \end{cases} \tag{9}$$

The channel parameter D is defined in Theorem 1 and C is the channel capacity. Furthermore, (6) and (7) are satisfied whenever one of the following is satisfied:

1. *Given $R < C$, t is sufficiently large.*

2. *Given $t \geq 3$ and $R < 1$, $D > 0$ is sufficiently small.*

Note that the average spectrum of the ensemble is related to the probability that a certain length N sequence of weight l is a codeword, as follows:

$$\overline{S}_l = \binom{N}{l} \Pr\left(A\mathbf{x} = \mathbf{0} | w(\mathbf{x}) = l\right) \tag{10}$$

To prove the theorem we use the following lemma, which provides upper bounds to the average spectrum of the code ensemble.

Lemma 1 *If lt is odd then:*

$$\Pr(A\mathbf{x} = \mathbf{0} | w(\mathbf{x}) = l) = 0. \tag{11}$$

If lt is even, the following two upper bounds hold:

$$\begin{aligned} \Pr(A\mathbf{x} = \mathbf{0} | w(\mathbf{x}) = l) &\leq \binom{L}{\frac{lt}{2}} \left(\frac{lt}{2L}\right)^{lt} \quad \text{for } l \leq \tfrac{2L}{t}, & (12) \\ \Pr(A\mathbf{x} = \mathbf{0} | w(\mathbf{x}) = l) &\leq 2 \times 2^{-L} \sum_{j=0}^{L/2} \binom{L}{j} \left(1 - \frac{2j}{L}\right)^{lt}. & (13) \end{aligned}$$

The proof of Lemma 1 is provided in Appendix B and the proof of Theorem 2 is provided in Appendix C.

Recall that the rate of the code, R', satisfies $R' \geq R$. Hence the theorem asserts that for t sufficiently large, reliable communication is possible up to channel capacity. In the terminology used in [7], whenever t is sufficiently large the ensemble is "very good". The ensemble averaged probability of decoding error decreases at least polynomially in N. The theorem also asserts (item *2*) that the ensemble is "good" for any $t \geq 3$. This means that for any $t \geq 3$ and $R < 1$, there are codes (with parameter t, and rate R) that achieve reliable communication, provided that $D > 0$ is small enough.

MacKay [7] considered the BSC case and showed that this ensemble is "good" for $t \geq 3$ and "very good" for t sufficiently large. Theorem 2 generalizes this result to any binary-input symmetric-output channel. In addition to that Theorem 2 provides a polynomial upper bound on the probability of decoding error.

III.2 Codes derived from bipartite regular graphs

A popular method for obtaining an ensemble of sparse parity-check codes is defined in terms of a bipartite graph. This is done by constructing a $c-d$ regular bipartite graph in which there are N vertices on the left side of the graph, each of degree c, and L vertices on the right, each of degree d, so that $Nc = Ld$. The vertices on the left side are associated with the codeword bits and the vertices on the right are associated with the parity-check equations (constraints). The mapping from the bipartite graph space to the parity-check matrix space is such that an element $A_{i,j}$ in the matrix, corresponding to the i'th vertex on the right and j'th vertex on the left, is set to '1' if there is an odd number of arcs between the two vertices, and to '0' otherwise. An ensemble of $c-d$ regular graphs is defined as follows. The Nc arcs originating from left vertices are labeled from 1 to Nc. The same procedure is applied for the Nc arcs originating from right vertices. A permutation π is then uniformly drawn from the space of all permutations of $\{1, 2, \ldots, Nc\}$. For each i the arc labeled i on the left side is associated with the arc labeled π_i on the right side. Note that in this way multiple arcs may link a pair of vertices.

Theorem 3 *Consider the ensemble of binary parity check matrices $A_{L\times N}$ described above. Let $\overline{P}_e$ be the ensemble averaged probability of decoding error. Let c and d be integers such that $3 \le c < d$, let $R \stackrel{\Delta}{=} 1 - L/N = 1 - c/d$ and let*

$$\beta = (1-R)\frac{2}{c}e^{-12-K} \quad . \tag{14}$$

where

$$K = \frac{6\ln\left(c/(1-R)\right)}{c} \quad , \tag{15}$$

If there exists $0 < \gamma < 1/2$ such that

$$\max_{\beta \le x \le \gamma} \left\{ x\log D + h_2(x) - (1-R) + (1-R)\log\left(1 + (1-2x)^d\right)\right\} < 0 \tag{16}$$

and

$$R + (1-R)\log(1 + (1-2\gamma)^d) < C \tag{17}$$

then:

$$\overline{P}_e \le \begin{cases} N^{1-\frac{c}{2}} \frac{D}{1-D} \frac{(\frac{c}{2})^c}{\frac{c}{2}!}(1-R)^{-\frac{c}{2}}\left(1 + \frac{o(N)}{N}\right) & c \textit{ even} \\ \\ N^{2-c}\frac{D^2}{2(1-D^2)}\frac{c^{2c}}{c!}(1-R)^{-c}\left(1+\frac{o(N)}{N}\right) & c \textit{ odd} \end{cases} \tag{18}$$

The channel parameter D is defined in Theorem 1 and C is the channel capacity. Furthermore, (16) and (17) are satisfied whenever one of the following is satisfied:

1. *Given $R_0 < C$, c and d are sufficiently large and satisfy $R_0 \leq R = 1 - c/d < C$.*
2. *Given c and d $(3 \leq c < d)$, $D > 0$ is sufficiently small.*

To prove the theorem we use the following lemma, which provides upper bounds to the average spectrum of the code ensemble.

Lemma 2 *Let A be a parity-check matrix drawn from an ensemble of $c - d$ regular graphs (with the mapping from bipartite graph to parity-check matrix as defined above). Then the following upper bounds hold: If lc is odd:*

$$\Pr(A\mathbf{x} = \mathbf{0} | w(\mathbf{x}) = l) = 0. \tag{19}$$

If lc is even:

$$\Pr(A\mathbf{x} = \mathbf{0} | w(\mathbf{x}) = l) \leq \binom{L}{\frac{lc}{2}} \left(\frac{lc}{2L}\right)^{lc} \quad \textit{for } l \leq \tfrac{2L}{c}. \tag{20}$$

$$\Pr(A\mathbf{x} = \mathbf{0} | w(\mathbf{x}) = l) \leq (Ld+1)2^{-L}\left(1 + \left(1 - \frac{2l}{N}\right)^d\right)^L. \tag{21}$$

The proof of Lemma 2 is provided in Appendix D and the proof of Theorem 3 is provided in Appendix E.

As was noted in Section III.1, whenever c is sufficiently large the ensemble is "very good". The Theorem also asserts (item *2*) that the ensemble is "good" for any $c \geq 3$.

IV A lower bound on the decoding error probability

In this section we develop lower bounds to the ensemble averaged decoding error probability of LDPC codes. The results of this section apply to any binary-input channel, not necessarily a symmetric-output one. Denote by ϵ the probability of error in ML decoding of a single bit transmitted through this channel (i.e., P_e for $N = 1$) assuming a uniform a-priori probability, $Q(0) = Q(1) = 1/2$. That is, ϵ is the probability of error when having to make a hard decision about a single bit. Using the notation of Theorem 1,

$$\epsilon \triangleq \frac{1}{2}\sum_y \min\{P(y|0), P(y|1)\}.$$

We term this parameter the crossover probability of the channel. In the special case of a binary symmetric channel (BSC), this definition coincides with the standard crossover probability.

IV.1 Independent matrix elements

In this section we consider an ensemble of LDPC matrices $A_{L\times N}$ in which the elements $A_{i,j}$ are i.i.d. with $\Pr(A_{i,j}=1)=t/L$, i.e., the expected weight of each column is t. We shall see that in this case the probability of decoding error, $\overline{P}_e$, cannot be made arbitrarily small as $N\to\infty$, regardless of the rate. In fact, it tends to 1.

Since the probability of a 1 in each matrix element is t/L, the probability that a specific column is all 0's is $(1-\frac{t}{L})^L$, which tends to e^{-t} as $L\to\infty$. Thus for any $\delta>0$, the probability that a specific column is all 0's is greater than $(e^{-t}-\delta)$ for L large enough.

Now consider a transmitted bit x_i that corresponds to an all zero i-th column of A. In that case y_i is a sufficient statistic for decoding x_i. In fact, x_i is decoded as 0 or 1, depending on whether $P(y_i|0)$ or $P(y_i|1)$ is larger. Thus x_i is decoded with error with probability ϵ.

Combining the two preceding arguments we get that for each x_i the probability of error is at least $\epsilon(e^{-t}-\delta)$ for N large enough, regardless of the value of the rate, R. Moreover, because the above described events are independent for different values of i, the probability that all N x_i's are correctly decoded is not greater than $[1-\epsilon(e^{-t}-\delta)]^N$, which tends to 0 as $N\to\infty$, indicating that $\overline{P}_e\to 1$ as $N\to\infty$.

IV.2 Independent matrix columns

In Appendix F we prove the following theorem.

Theorem 4 *Consider the ensemble of binary parity check matrices $A_{L\times N}$ described in Theorem 2, with $t\geq 3$, over a memoryless binary-input channel with crossover probability ϵ. Denote the ensemble averaged maximum likelihood decoding error probability by $\overline{P}_e$. Then*

$$\overline{P}_e > \begin{cases} N^{1-t/2}\epsilon\left(\frac{t}{2}!\right)(1-R)^{-t/2}\left(1-\frac{o(N)}{N}\right) & t \textit{ even} \\ \\ N^{2-t}\frac{\epsilon^2}{2}\left(t!\right)(1-R)^{-t}\left(1-\frac{o(N)}{N}\right) & t \textit{ odd}\end{cases} \tag{22}$$

Combining Theorems 2 and 4 yields,

Corollary 1 *Consider the ensemble of binary parity check matrices $A_{L\times N}$ described in Theorem 2, over a memoryless binary-input symmetric-output channel. Suppose further that (6) and (7) are satisfied. Denote the ensemble averaged maximum likelihood decoding error probability by $\overline{P}_e$. Then*

$$\lim_{N\to\infty}\frac{-\log\overline{P}_e}{\log N} = \begin{cases}\frac{t}{2}-1 & t \textit{ even}\\ t-2 & t \textit{ odd}\end{cases} \tag{23}$$

Perhaps the most striking feature of the corollary is that the right hand side of (23) is independent of both R and C. This behavior stands in contrast to the various bounds on the probability of error when using random coding, where the bound is monotonically increasing with increasing R or decreasing C.

By Theorem 2, (6) and (7) hold when either $R < C$ and t is large enough, or when for given t, $D > 0$ is small enough.

IV.3 Codes derived from bipartite regular graphs

In this section we derive a lower bound $\overline{P}_e$ for the ensemble of Section III.2. The results are very similar to the results of the preceding section.

Theorem 5 *Consider the ensemble of binary parity check matrices $A_{L\times N}$ described in Theorem 3, with $c \geq 3$, over a memoryless binary-input channel with crossover probability ϵ. Denote the ensemble averaged maximum likelihood decoding error probability by $\overline{P}_e$. Then*

$$\overline{P}_e > \begin{cases} \epsilon\left(1-\frac{1}{d}\right)^{c/2}\left(\frac{c}{2}!\right)(1-R)^{-c/2}N^{1-c/2}\left(1-\frac{o(N)}{N}\right) & c \textit{ even} \\ \\ \frac{\epsilon^2}{2}\left(1-\frac{1}{d}\right)^{c}(c!)(1-R)^{-c}N^{2-c}\left(1-\frac{o(N)}{N}\right) & c \textit{ odd} \end{cases} \tag{24}$$

The proof is provided in Appendix G.

As in the previous section, Theorems 3 and 5 yield the following corollary:

Corollary 2 *Consider the ensemble of binary parity check matrices $A_{L\times N}$ described in Theorem 3, with $c \geq 3$, over a memoryless binary-input symmetric-output channel. Suppose further that (16) and (17) are satisfied. Denote the ensemble averaged maximum likelihood decoding error probability by $\overline{P}_e$. Then*

$$\lim_{N\to\infty}\frac{-\log\overline{P}_e}{\log N} = \begin{cases} \frac{c}{2}-1 & c \textit{ even} \\ c-2 & c \textit{ odd} \end{cases}$$

The discussion after Corollary 1 applies here as well.

V Expurgated ensembles

Gallager [3] expurgated the ensemble by throwing away all codes that have a small enough relative minimum distance. We follow the same expurgation technique and show that as a result of the expurgation, the behavior of the averaged decoding error probability changes from polynomial to exponential. Thus, the original ensemble averaged error probability is dominated by a polynomially small fraction of bad codes. In fact, it turns out that when the ensemble parameter t (c and d) is sufficiently large, the error exponent of the expurgated ensemble gets arbitrarily close to the random-coding exponent.

V.1 Independent matrix columns

We begin by investigating the distribution of $d_{\min}$ - the minimal weight of a non-zero codeword. Let $0 < \beta < \gamma < 1/2$ and $\Gamma = \gamma N$ as in Theorem 2. Then, using a union bound we have:

$$\begin{aligned}\Pr(d_{\min} \le \Gamma) &= \Pr(\exists \mathbf{x} \neq \mathbf{0} : A\mathbf{x} = \mathbf{0}, w(\mathbf{x}) \le \Gamma) \le \sum_{l=1}^{\Gamma} \Pr(\exists \mathbf{x} : A\mathbf{x} = \mathbf{0}, w(\mathbf{x}) = l) \\ &\le \sum_{l=1}^{\Gamma} \binom{N}{l} \Pr(A\mathbf{x} = \mathbf{0} | w(\mathbf{x}) = l) = \sum_{l=1}^{\Gamma} \overline{S}_l \le \hat{\sigma}_1 + \hat{\sigma}_2, \end{aligned} \tag{25}$$

where $\hat{\sigma}_1 = \sum_{l=1}^{\beta N} \overline{S}_l$ and $\hat{\sigma}_2 = \sum_{l=\beta N}^{\gamma N} \overline{S}_l$. $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are very similar to σ_1 and σ_2 that are defined in the proof of Theorem 2 (see (42)). Thus their bounding is performed in an analogous manner. It follows that if (6) is satisfied for $D = 1$ then $\hat{\sigma}_1 \to 0$ and $\hat{\sigma_2} \to 0$ as $N \to \infty$. Hence $\Pr(d_{\min} \le \Gamma) \to 0$ as $N \to \infty$. Therefore, for N large enough, expurgating all codes with $d_{\min} \le \Gamma$ from the ensemble does not reduce the size of the ensemble by a factor greater than 2. Denoting quantities referring to this expurgated ensemble by a superscript 'x', we then have for the average spectrum elements:

$$\begin{aligned} \overline{S}_l^x &= 0, \qquad && l \le \Gamma \\ \overline{S}_l^x &\le 2\overline{S}_l, \qquad && l > \Gamma \quad . \end{aligned}$$

Thus by the same arguments used in the proof of Theorem 2 (setting $\sigma_1^x = \sigma_2^x = 0$ in (42) we have $\overline{P}_e^x \le \sigma_3^x$, where σ_3^x is bounded by the right hand side of (52)),

$$\frac{-\log \overline{P}_e^x}{N} \ge E_r(R + G(R, \gamma t)) - \frac{o(N)}{N}. \tag{26}$$

In particular, if (7) holds then $\overline{P}_e^x$ decreases exponentially. Furthermore, as $t, N \to \infty$, the right hand side of (26) approaches $E_r(R)$. Denote the LDPC ensemble by $\mathcal{C}$ and the expurgated ensemble by $\mathcal{C}^x$. Let C^0 be some code drawn at random from $\mathcal{C}$. Further denote the probability of error of C^0 by P_e^0. Now, for any E we have

$$\Pr\left(\frac{-\log P_e^0}{N} < E\right) \le \Pr\left(\frac{-\log P_e^0}{N} < E \mid C^0 \in \mathcal{C}^x\right) + \Pr\left(C^0 \notin \mathcal{C}^x\right) \tag{27}$$

Next we consider the case where $E < E_r(R + G(R, \gamma t))$ and show that if (6) is satisfied for $D = 1$ then both terms on the right hand side of (27) go to zero as $N \to \infty$. The second term satisfies this property by (25) and the discussion that follows. We bound the first term as follows. For N sufficiently large,

$$\begin{aligned} \Pr\left(\frac{-\log P_e^0}{N} < E \mid C^0 \in \mathcal{C}^x\right) &< \Pr\left(\frac{-\log P_e^0}{N} < \frac{-\log(N\overline{P}_e^x)}{N} \mid C^0 \in \mathcal{C}^x\right) \\ &= \Pr\left(P_e^0 > N\overline{P}_e^x \mid C^0 \in \mathcal{C}^x\right) < \frac{1}{N} \to 0 \end{aligned}$$

The first inequality follows from (26), and the constraint we set on E. The last inequality follows by the Markov inequality.

We summarize our results by the following theorem.

Theorem 6 *Let P_e^0 denote the probability of decoding error of a code which is randomly chosen from the LDPC ensemble described in Section III.1. If $t \geq 3$, and in addition (6) holds for $D = 1$, then for any $E < E_r(R + G(R, \gamma t))$,*

$$\lim_{N\to\infty} \Pr\left(\frac{-\log P_e^0}{N} \geq E\right) = 1. \tag{28}$$

In particular, if (7) holds then $E_r(R + G(R, \gamma t)) > 0$. Furthermore, as $t \to \infty$, (28) holds for any $E < E_r(R)$.

V.2 Codes derived from bipartite regular graphs

The derivations for this ensemble are completely analogous to those of the previous section. In this case we have:

Theorem 7 *Let P_e^0 denote the probability of decoding error of a code which is randomly chosen from the LDPC ensemble described in Section III.2. If $c \geq 3$, and in addition (16) holds for $D = 1$, then for any $E < E_r\left(R + (1 - R)\log\left(1 + (1 - 2\gamma)^d\right)\right)$,*

$$\lim_{N\to\infty} \Pr\left(\frac{-\log P_e^0}{N} \geq E\right) = 1. \tag{29}$$

In particular, if (17) holds then $E_r\left(R + (1 - R)\log\left(1 + (1 - 2\gamma)^d\right)\right) > 0$. Furthermore, as $c \to \infty$, (29) holds for any $E < E_r(R)$.

V.3 A comparison with Gallager's bound

In this section we compare the bounds on the threshold parameter of the channel, below which reliable communication is guaranteed, when using Theorem 1 to the bounds obtained by Gallager in [3].

Gallager defined an ensemble of LDPC codes parameterized by j and k, such that the weight of each column of the parity check matrix is j and the weight of each row in this matrix is k. Gallager obtained the following bound on the average spectrum, $\overline{S}_l$ of the ensemble,

$$\overline{S}_l \le C(\lambda, N)e^{-NB_{j,k}(\lambda)} \tag{30}$$

where $\lambda = l/n$,

$$B_{j,k}(\lambda) = (j-1)h(\lambda) - \frac{j}{k}\left[\mu(s) + (k-1)\ln 2\right] + js\lambda \quad ,$$
$$C(\lambda, N) = \left[2\pi N\lambda(1-\lambda)\right]^{(j-1)/2} e^{(j-1)/[12\lambda(1-\lambda)N]}$$

and s is the solution to $\lambda = \mu'(s)/k$ for

$$\mu(s) = \ln\left(2^{-k}\left[(1+e^s)^k + (1-e^s)^k\right]\right)$$

Gallager obtained bounds on the error probability of the expurgated ensemble. The expurgated ensemble satisfies $\overline{S}_l = 0$ for $l < \delta_{j,k}N$, where $\delta_{j,k}$ is the first positive zero of $B_{j,k}(\lambda)$.

In order to apply Theorem 1 to Gallager's ensemble suppose first that k is odd. We set $U = \{\delta_{j,k}N, \ldots, \gamma N\}$, and seek for the largest possible γ for which

$$\sum_{l=\delta_{j,k}N}^{\gamma N} \overline{S}_l D^l \to 0$$

as $N \to \infty$. Let this limiting value of γ be denoted by γ_0. Then γ_0 is the smallest value of $\lambda > 0$ for which

$$-B_{j,k}(\lambda) + \lambda \ln D = 0 \tag{31}$$

The average error probability of the ensemble approaches zero as $N \to \infty$ if

$$R + \frac{\log \alpha}{N} < C \tag{32}$$

for

$$\alpha = \max_{l > \gamma_0 N} \frac{\overline{S}_l}{M-1} \frac{2^N}{\binom{N}{l}} \tag{33}$$

By (33) and (30), a sufficient condition for (32) is

$$\max_{\lambda > \gamma_0} -B_{j,k}(\lambda)\log(e) + 1 - h_2(\lambda) < C \tag{34}$$

So far we assumed that k was odd. For even k we use the fact that the average spectrum in Gallager's ensemble is symmetric, i.e., $\overline{S}_l = \overline{S}_{N-l}$ (since in this case the all-one word is a codeword). As a result, for even k we modify (33) so that the maximization is applied only in the range $\gamma_0 N < l \le N/2$.

We have compared our bound, which is summarized by (31) and (34), to Gallager's bound for the expurgated ensemble. In general, Gallager's bound involves an optimization problem that might not be solvable (over the function f). In order to overcome this difficulty, Gallager suggested using some suboptimal solution, which is optimal for the BSC. In Figure 1 we compare the threshold crossover below which reliable communication is guaranteed for various LDPC ensembles. In Figure 2 we do the same for an additive white Gaussian noise (AWGN) channel with standard deviation σ. In this case

$$P(y|0) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x+1)^2}{2\sigma^2}\right\}$$

We have used equations (3.42)-(3.45) in [3] for the AWGN case and equations (3.60)-(3.63) for the BSC case (note the typo in equation (3.62): the denominator should be squared). It appears from these results that Gallager's bound is tighter for lower rate codes, while our bound is tighter for higher rate codes.

Our bounds, summarized in Figures 1 and 2 apply both to the original ensemble and to the expurgated one. As was noted above, the difference between the original and expurgated LDPC ensembles is in the rate of decrease of the average decoding error probability.

VI Discussion

In the following, the ensembles of Sections III.1, III.2 and IV.1 are denoted by $\mathcal{M}$ (MacKay ensemble), $\mathcal{C}$ ($c-d$ regular ensemble) and $\mathcal{B}$ (Bernoulli ensemble) respectively.

In Theorems 2 and 3 we derive polynomial upper bounds on the probability of decoding error. One might wonder whether (6) and (7) ((16) and (17)) are indeed necessary. These conditions may be summarized as $R < R'(C, D, t)$ ($R < R'(C, D, d)$) (where $R'(\cdot)$ is some function of its arguments). The derivations which led to these conditions contained approximations that can possibly be improved, in the sense that an enhanced function $R''(C, D, t) \geq R'(C, D, t)$ exists such that (9) holds for any $R < R''(C, D, t)$. It is thus natural to ask whether these conditions may ultimately be replaced simply by $R < C$. This is answered in the negative by Gallager [3](Theorem 3.3), at least for the special case of the BSC and ensemble $\mathcal{C}$. Denoting the BSC crossover probability by ϵ and defining $\epsilon_k \triangleq 1/2(1 + (1-2\epsilon)^k)$, Galleger's result asserts that if each row of the parity check matrix of some LDPC code has weight k and $R > 1 - h_2(\epsilon)/h_2(\epsilon_k)$, then the probability of decoding error, P_e, is bounded below by a constant independent of N. Since this is true for all codes in the ensemble, $\overline{P}_e$ is also bounded away from 0 for $R > 1 - h_2(\epsilon)/h_2(\epsilon_k)$.

In fact, the proof in [3] holds for codes where the weight of each row in the parity check matrix is bounded from above by k. Thus, for ensemble $\mathcal{C}$, $R > 1 - h_2(\epsilon)/h_2(\epsilon_d)$ implies that $\overline{P}_e$ is bounded away from 0. This means that any $R''(C, D, d)$ must satisfy:

$$R''(C, D, d) \leq 1 - \frac{h_2(\epsilon)}{h_2(\epsilon_d)} < 1 - h_2(\epsilon) = C.$$

As $d \to \infty$, $1 - h_2(\epsilon)/h_2(\epsilon_d) \to C$, which is consistent with the fact that (16) and (17) were shown to hold for c, d large enough.

Throughout Sections III - V, the similarity between the behavior of $\overline{P}_e$ in $\mathcal{M}$ and $\mathcal{C}$ is clearly evident, when setting $t = c$. Informally, this indicates that the statistical dependencies among the parity check matrix columns of ensemble $\mathcal{C}$ are sufficiently small, so that the ensemble behaves almost as if the columns were independent. For both ensembles, $\overline{P}_e$ is dominated by the probability that $d_{\min}$ is equal to 1 for t (c) even, or 2 for t (c) odd. This is apparent from the following. First, the upper bound (Section III) is dominated by the value of the term corresponding to codewords with weight 1 (2 for t or c odd). Second, although only the case where $d_{\min} = 1$ (2) was considered when calculating the lower bound, it still turned out to be asymptotically tight to the upper bound. Hence, the other terms are negligible.

Unlike $\mathcal{M}$ and $\mathcal{C}$, for which $\overline{P}_e$ behaves very similarly, $\mathcal{B}$ is strikingly different. For $\mathcal{B}$ we showed that $\overline{P}_e \to 1$ as $N \to \infty$. Hence for any $\theta < 1$, $\Pr(P_e > \theta) \to 1$ as $N \to \infty$, where P_e is the probability of error of some code drawn at random from the ensemble. This should be compared with our result for $\mathcal{M}$ and $\mathcal{C}$, that there exists $E > 0$ such that $\Pr(P_e < e^{-NE}) \to 1$ as $N \to \infty$. The difference stems from the behavior of $\Pr(d_{\min} \leq x)$. Whereas for $\mathcal{B}$ we have $\Pr(d_{\min} = 1) \to 1$ as $N \to \infty$, for $\mathcal{M}$ and $\mathcal{C}$ we have $\Pr(d_{\min} \leq \gamma N) \to 0$.

In fact, for high enough rates (for which the random coding, expurgated random coding and sphere packing exponents coincide) we have shown (Theorems 6 and 7) that there exist LDPC codes in $\mathcal{M}$ and $\mathcal{C}$ which are exponentially as good, up to a penalty which goes to 0 as $t \to \infty$, (or $c, d \to \infty$) as any other code. Thus, at least for these rates, arbitrarily small penalty is paid for using such a LDPC code. Moreover, this phenomenal performance of the LDPC codes is achieved by most codes in the respective ensemble. Thus, a *random* selection of a code from $\mathcal{M}$ or $\mathcal{C}$ will almost certainly produce a code that is exponentially as good as the best possible code for that rate (for t, N sufficiently large).

Appendix

A Proof of Theorem 1

First,

$$\overline{P}_e = \frac{1}{M}\sum_{m=0}^{M-1} \overline{P}_{e,m} \quad ,$$

where $\overline{P}_{e,m}$ is the ensemble averaged probability of error given that the m-th codeword has been transmitted. Now the code is linear and the channel is binary-input symmetric-output. Hence $\overline{P}_{e,m}$ is independent of m. In particular $\overline{P}_{e,m} = \overline{P}_{e,0}$. Hence $\overline{P}_e = \overline{P}_{e,0}$. Denote the codewords by $\mathbf{x}_i$ (each of length N), and let $\mathbf{y}$ be the output vector of the channel. For a specific code in the ensemble, let $P^1_{e,0}$ ($P^2_{e,0}$, respectively) denote the probability that there exists some codeword $\mathbf{x}$ such that $P_N(\mathbf{y}|\mathbf{x}) \geq P_N(\mathbf{y}|\mathbf{x}_0)$ for some codeword $\mathbf{x}$ with $d(\mathbf{x},\mathbf{x}_0) \in U$ ($d(\mathbf{x},\mathbf{x}_0) \in U^c$, respectively), given that codeword $\mathbf{x}_0$ has been transmitted. $\overline{P}^1_{e,0}$ and $\overline{P}^2_{e,0}$ denote the corresponding ensemble averaged probabilities. Hence, by the union bound

$$\overline{P}_e \leq \overline{P}^1_{e,0} + \overline{P}^2_{e,0}.$$

We use separate bounds for $\overline{P}^1_{e,0}$ and $\overline{P}^2_{e,0}$.

A union bound is used for $\overline{P}^1_{e,0}$. Without loss of generality we assume here $\mathbf{x}_0 = \mathbf{0}$, i.e. codeword number 0 is the all-zero codeword (which is always present in a linear code). First consider any specific code in the ensemble:

$$\begin{aligned} P^1_{e,0} &= \Pr\left(\mathbf{y} : \exists i : P_N(\mathbf{y}|\mathbf{x}_i) \geq P_N(\mathbf{y}|\mathbf{0}), w(\mathbf{x}_i) \in U \mid \mathbf{0} \text{ was transmitted}\right) \\ &= \sum_{\mathbf{y}} I(\mathbf{y}) P_N(\mathbf{y}|\mathbf{0}) \leq \sum_{\mathbf{y}} I(\mathbf{y}) P_N(\mathbf{y}|\mathbf{0}) \sum_{i:\, w(\mathbf{x}_i)\in U} \sqrt{\frac{P_N(\mathbf{y}|\mathbf{x}_i)}{P_N(\mathbf{y}|\mathbf{0})}}. \end{aligned}$$

where

$$I(\mathbf{y}) = \begin{cases} 1 & \exists i :\ w(\mathbf{x}_i) \in U, P_N(\mathbf{y}|\mathbf{x}_i) \geq P_N(\mathbf{y}|\mathbf{0}) \\ 0 & \text{otherwise} \end{cases}$$

Adding non-negative terms may only increase the sum, hence:

$$P^1_{e,0} \leq \sum_{\mathbf{y}} P_N(\mathbf{y}|\mathbf{0}) \sum_{i:\, w(\mathbf{x}_i)\in U} \sqrt{\frac{P_N(\mathbf{y}|\mathbf{x}_i)}{P_N(\mathbf{y}|\mathbf{0})}} = \sum_{i:\, w(\mathbf{x}_i)\in U} \sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}|\mathbf{x}_i) P_N(\mathbf{y}|\mathbf{0})}.$$

Now, using (1) we have:

$$\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}|\mathbf{0}) P_N(\mathbf{y}|\mathbf{x})} = \prod_{j=1}^{N} \sum_{y_j} \sqrt{P(y_j|0) P(y_j|x_j)} = D^{w(\mathbf{x})}.$$

Hence,

$$P^1_{e,0} \leq \sum_{l\in U} S_l D^l \quad ,$$

where S_l is the spectrum of the code, i.e. the number of words with weight l. Now averaging the last inequality over the ensemble of codes yields,

$$\overline{P}^1_{e,0} \leq \sum_{l \in U} \overline{S}_l D^l$$

Next we obtain a bound for $\overline{P}^2_{e,0}$. For that purpose we use the technique that was developed in [11]. In the following, a code is an ordered collection of codewords. For each code $C_l \quad l = 1, \ldots, |\mathcal{C}|$ in the given ensemble $\mathcal{C}$, we construct an ensemble $\mathcal{C}_l$ of codes as follows. First generate an ensemble $\mathcal{C}'_l$ from the given code by including all possible permutations, π, of the order of the codewords, where the permutations are uniformly distributed. Now generate an ensemble $\mathcal{C}''_l$ by including all possible permutations, σ, of the order of the symbols in the codewords (for each code in $\mathcal{C}'_l$), where the permutations are again uniformly distributed. Finally, from $\mathcal{C}''_l$ we create an ensemble $\tilde{\mathcal{C}}_l$ by including all possible binary offset vectors with equal probability. That is,

$$\tilde{\mathcal{C}}_l = \left\{ \{\mathbf{c}_0 \oplus \mathbf{v}, \ldots, \mathbf{c}_{M-1} \oplus \mathbf{v}\} \ \mid \ \{\mathbf{c}_0, \ldots, \mathbf{c}_{M-1}\} \in \mathcal{C}''_l, \mathbf{v} \in \{0,1\}^N \right\}.$$

where $\mathbf{v} \in \{0,1\}^N$ is a uniformly distributed random vector. The probability of decoding error of each code in $\tilde{\mathcal{C}}_l$ is obviously equal to the probability of decoding error of the code C_l (the probability of error is invariant to codeword and symbol permutation and also to displacement by a fixed vector $\mathbf{v}$). Now consider the ensemble $\tilde{\mathcal{C}}$ which consists of the union of the ensembles $\tilde{\mathcal{C}}_l$, i.e. $\tilde{\mathcal{C}} = \bigcup_{l=1,\ldots,|\mathcal{C}|} \tilde{\mathcal{C}}_l$. The averaged probability of error of $\tilde{\mathcal{C}}$ is equal to the averaged probability of error of the original ensemble, $\mathcal{C}$. We now show, using the techniques of [11], that for $(\tilde{\mathbf{c}}_0, \ldots, \tilde{\mathbf{c}}_{M-1}) \in \tilde{\mathcal{C}}$

$$\Pr(\tilde{\mathbf{c}}_i = \mathbf{x}) = 2^{-N} \tag{35}$$

$$\Pr(\tilde{\mathbf{c}}_i = \mathbf{x} \mid \tilde{\mathbf{c}}_j = \mathbf{y}) \leq \alpha 2^{-N} \qquad \text{whenever } d(\mathbf{x}, \mathbf{y}) \in U^c \tag{36}$$

where α is defined by (3).

To see that, note that for $\left(\mathbf{c}'_0, \ldots, \mathbf{c}'_{M-1}\right) \in \mathcal{C}'_l$, $i \neq j$ and $\mathbf{x} \neq \mathbf{0}$ we have:

$$\Pr(\mathbf{c}'_i \oplus \mathbf{c}'_j = \mathbf{x}) = \begin{cases} \frac{1}{M-1}, & \text{if } \mathbf{x} \in C_l \\ 0, & \text{if } \mathbf{x} \notin C_l \end{cases}$$

For $\left(\mathbf{c}''_0, \ldots, \mathbf{c}''_{M-1}\right) \in \mathcal{C}''_l$ and $i \neq j$ we have:

$$\Pr\left(\mathbf{c}''_i \oplus \mathbf{c}''_j = \mathbf{x}\right) = \sum_{\sigma} \frac{1}{N!} \Pr\left(\mathbf{c}'_i \oplus \mathbf{c}'_j = \sigma(\mathbf{x})\right) = \frac{S_m}{(M-1)\binom{N}{m}}$$

where $m = w(\mathbf{x}) > 0$ (Since the codes are linear, we may assume without loss of generality that $\mathbf{c}'_i = \mathbf{0}$. In addition $\Pr(w(\mathbf{c}'_j) = m) = S_m/(M-1)$). For $(\tilde{\mathbf{c}}_0, \ldots, \tilde{\mathbf{c}}_{M-1}) \in \tilde{\mathcal{C}}_l$ we have:

$$\begin{aligned}
\Pr(\tilde{\mathbf{c}}_i = \mathbf{x}) &= \sum_{\mathbf{v}} 2^{-N} \Pr(\mathbf{c}''_i = \mathbf{x} \oplus \mathbf{v}) = 2^{-N} && (37) \\
\Pr(\tilde{\mathbf{c}}_i = \mathbf{x} \mid \tilde{\mathbf{c}}_j = \mathbf{y}) &= 2^N \Pr(\tilde{\mathbf{c}}_i = \mathbf{x}, \tilde{\mathbf{c}}_j = \mathbf{y}) \\
&= 2^N \sum_{\mathbf{v}} 2^{-N} \Pr\left(\mathbf{c}''_i \oplus \mathbf{c}''_j = \mathbf{x} \oplus \mathbf{y}, \mathbf{c}''_j = \mathbf{y} \oplus \mathbf{v}\right) \\
&= \Pr\left(\mathbf{c}''_i \oplus \mathbf{c}''_j = \mathbf{x} \oplus \mathbf{y}\right) = \frac{S_k}{(M-1)\binom{N}{k}} && (38)
\end{aligned}$$

where $k = d(\mathbf{x}, \mathbf{y}) > 0$. The first equality in (38) is due to (37). The third equality is due to the total probability formula.

Averaging the last two equations yields (35)-(36) for the ensemble $\tilde{\mathcal{C}}$.

Now set $Q_N(\mathbf{x}) = 2^{-N}$ (the uniform distribution). Using a modified version of Gallager's random coding exponent derivation [4][Theorem 5.6.1], we have for the ensemble $\tilde{\mathcal{C}}$:

$$\overline{P}^2_{e,0} = \sum_{\mathbf{x}_0} \sum_{\mathbf{y}} Q_N(\mathbf{x}_0) P_N(\mathbf{y}|\mathbf{x}_0) \Pr[\text{error2}|\mathbf{x}_0, \mathbf{y}] \tag{39}$$

where $\Pr[\text{error2}|\mathbf{x}_0, \mathbf{y}]$ is the conditional probability (given that the transmitted codeword was $\mathbf{x}_0$ and that the received sequence was $\mathbf{y}$) that there exist some codeword $\mathbf{x}_m \quad m \neq 0$, such that $P_N(\mathbf{y}|\mathbf{x}_m) \geq P_N(\mathbf{y}|\mathbf{x}_0)$ and $d(\mathbf{x}_m, \mathbf{x}_0) \in U^c$. Using Gallager's enhanced union bound [4][p. 126] we have,

$$\begin{aligned}
\Pr[\text{error2}|\mathbf{x}_0, \mathbf{y}] &\leq \left[\sum_{m \neq 0} \sum_{\mathbf{x}_m \,:\, P_N(\mathbf{y}|\mathbf{x}_m) \geq P_N(\mathbf{y}|\mathbf{x}_0)\,,\, d(\mathbf{x}_m, \mathbf{x}_0) \in U^c} \Pr(\tilde{\mathbf{c}}_m = \mathbf{x}_m | \tilde{\mathbf{c}}_0 = \mathbf{x}_0)\right]^\rho \\
&\leq \left[(M-1) \sum_{\mathbf{x} \,:\, d(\mathbf{x}, \mathbf{x}_0) \in U^c} \alpha Q_N(\mathbf{x}) \left(\frac{P_N(\mathbf{y}|\mathbf{x})}{P_N(\mathbf{y}|\mathbf{x}_0)}\right)^s\right]^\rho && (40)
\end{aligned}$$

for any $0 \leq \rho \leq 1$ and $s > 0$. Note that in the second inequality we used (36). Substituting (40) in (39) yields,

$$\overline{P}^2_{e,0} \leq \alpha^\rho (M-1)^\rho \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}_0} Q_N(\mathbf{x}_0) P_N(\mathbf{y}|\mathbf{x}_0)^{1-s\rho}\right] \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y}|\mathbf{x})^s\right]^\rho \quad . \tag{41}$$

Now using (1) and

$$Q_N(\mathbf{x}) = \prod_{n=1}^{N} Q(x_n)$$

in (41) (as used in [4] to derive (5.6.13) from (5.6.10)) yields

$$\overline{P}^2_{e,0} \leq 2^{-N \max_{0 \leq \rho \leq 1} \{E_0(\rho, Q) - \rho(R + (\log \alpha)/N)\}}$$

where $Q(0) = Q(1) = 1/2$. The required result follows provided that $Q(0) = Q(1) = 1/2$ is the input distribution that yields capacity for a binary-input symmetric-output channel. However the last claim follows by a direct application of [4][Theorem 4.5.1].

B Proof of Lemma 1

The probability considered ($\Pr(A\mathbf{x} = \mathbf{0}|w(\mathbf{x}) = l)$) is the probability to return to the origin after performing lt random-walk steps on the L-dimensional unit cube [7]. Equation (13) is by [7][Eq. (87)].

To derive the first upper bound, consider the following experiment. Set a vector $\mathbf{v} = (v_1, v_2, \ldots, v_L)$ of length L to $\mathbf{0}$. Then repeat the following process lt times. Draw a number n uniformly from $1, 2, ..., L$, and increase v_n by one. This experiment describes the random walk process such that at the end of the experiment, v_i is the number of random walk steps in the i-th dimension. Thus $\Pr(A\mathbf{x} = \mathbf{0}|w(\mathbf{x}) = l)$ is equal to the probability that for each $1 \leq i \leq L$, v_i is even. When lt is odd, this probability is obviously 0 (hence (11)).

Now consider the case where lt is even. Define an outcome of the experiment above by the length lt vector of elements from $1, 2, ..., L$. If, for example, $lt = 2$ and in the experiment, the first draw yields 5 and the second yields 8, then the outcome would be $(5, 8)$. In this case, $v_5 = v_8 = 1$ and the rest of the v_i-s are zero.

The number of different possible outcomes of the experiment is L^{lt}. However, if all v_i's are even, there cannot be more than $lt/2$ non-zero v_i's (since each such v_i satisfies $v_i \geq 2$, and $\sum v_i = lt$). Hence the number of different outcomes that correspond to a vector $\mathbf{v}$ with all elements even is not greater than $\binom{L}{lt/2}(lt/2)^{lt}$. This yields (12). □

C Proof of Theorem 2

Let $\beta < \gamma < 1/2$ be some real number (independent of N) which will be determined later. Defining $\Gamma = \gamma N$ and choosing $U = \{1, 2, ..., \Gamma\}$, (2) may be written as:

$$\overline{P}_e \leq \sigma_1 + \sigma_2 + \sigma_3, \tag{42}$$

where $\sigma_1 = \sum_{l=1}^{\beta N} \overline{S}_l D^l$, $\sigma_2 = \sum_{l=\beta N}^{\gamma N} \overline{S}_l D^l$ and $\sigma_3 = 2^{-NE_r(R+(\log\alpha)/N)}$, and α is given by (3). We first bound σ_1. From (12), (11) and (10) we get:

$$\sigma_1 \leq \begin{cases} \sum_{l=1}^{\beta N} D^l \binom{N}{l} \binom{L}{\frac{lt}{2}} \left(\frac{lt}{2L}\right)^{lt} & t \text{ even} \\ \sum_{l=2,4,\ldots}^{\beta N} D^l \binom{N}{l} \binom{L}{\frac{lt}{2}} \left(\frac{lt}{2L}\right)^{lt} & t \text{ odd} \end{cases} \tag{43}$$

We now show that (43), (4) and (5) imply,

$$\sigma_1 \le \begin{cases} N^{1-\frac{t}{2}}\frac{D}{1-D}\frac{(\frac{t}{2})^t}{\frac{t}{2}!}(1-R)^{-\frac{t}{2}} & t \text{ even} \\ \\ N^{2-t}\frac{D^2}{2(1-D^2)}\frac{t^{2t}}{t!}(1-R)^{-t} & t \text{ odd} \end{cases} \tag{44}$$

Define

$$f(l) = \binom{N}{l}\binom{L}{\frac{lt}{2}}\left(\frac{lt}{2L}\right)^{lt}. \tag{45}$$

Suppose first that t is even. Then

$$\begin{aligned} \frac{f(l+1)}{f(l)} &= \frac{\binom{N}{l+1}}{\binom{N}{l}}\frac{\binom{L}{\frac{(l+1)t}{2}}}{\binom{L}{\frac{lt}{2}}}\frac{\left(\frac{(l+1)t}{2L}\right)^{(l+1)t}}{\left(\frac{lt}{2L}\right)^{lt}} \\ &= \frac{N-l}{l+1}\frac{(L-\frac{lt}{2})(L-\frac{lt}{2}-1)...(L-\frac{(l+1)t}{2}+1)}{(\frac{lt}{2}+1)(\frac{lt}{2}+2)...(\frac{(l+1)t}{2})}\frac{\left(\frac{(l+1)t}{2L}\right)^{(l+1)t}}{\left(\frac{lt}{2L}\right)^{(l+1-1)t}} \\ &< \frac{N}{l}\left(\frac{L-\frac{lt}{2}}{\frac{lt}{2}}\right)^{\frac{t}{2}}\left(1+\frac{1}{l}\right)^{(l+1)t}\left(\frac{lt}{2L}\right)^t \\ &\le \frac{2Lt}{2lt(1-R)}\left(\frac{2L-lt}{lt}\right)^{\frac{t}{2}}\left(1+\frac{1}{l}\right)^{2lt}\left(\frac{lt}{2L}\right)^t \\ &< \frac{t}{2(1-R)}e^{2t}\left(\frac{lt}{2L}\right)^{\frac{t}{2}-1}. \end{aligned} \tag{46}$$

If we require

$$\frac{lt}{2L} \le e^{-(12+\frac{6\ln(t/(1-R))}{t})}, \tag{47}$$

we have from (46) (recall that $t \ge 3$)

$$\frac{f(l+1)}{f(l)} < \frac{t}{2(1-R)}e^{2t}e^{-\left(12+\frac{6\ln(t/(1-R))}{t}\right)\frac{t}{6}} = \frac{t}{2(1-R)}e^{2t-2t-\ln(t/(1-R))} = \frac{1}{2} < 1,$$

indicating that as long as (47) holds, $f(l)$ is monotonically decreasing. (47) can also be written as $l \le \beta N$, where β is defined by (4). Hence $f(l)$ is monotonically decreasing for $l \le \beta N$. Thus using (43) we have,

$$\sigma_1 \le \frac{D}{1-D}\binom{N}{1}\binom{L}{\frac{t}{2}}\left(\frac{t}{2L}\right)^t \le \frac{ND}{1-D}\frac{[(1-R)N]^{t/2}}{\frac{t}{2}!}\left(\frac{t}{2(1-R)N}\right)^t. \tag{48}$$

(note that by the definition of D and the Cauchy Shwartz inequality, $D < 1$). If t is odd then it can be similarly shown that if (47) holds, $f(l+2)/f(l) < 1$ for l even. Thus $f(2), f(4), f(6), ...$ is monotonically decreasing, so σ_1 may be bounded by $D^2/(1-D^2) \times f(2)$. Substituting $f(2)$ by (45) yields (44).

Next we bound σ_2. From (13) and (10) we have:

$$\begin{aligned}\sigma_2 &\leq 2\sum_{l=\beta N}^{\gamma N}\binom{N}{l}D^l 2^{-L}\sum_{j=0}^{L/2}\binom{L}{j}\left(1-\frac{2j}{L}\right)^{lt} \\ &\leq 2\sum_{l=\beta N}^{\gamma N} 2^{Nh_2(l/N)}D^l 2^{-L}\sum_{j=0}^{L/2} e^{Lh(j/L)}\left(1-\frac{2j}{L}\right)^{lt} \\ &\leq (L+2)N \max_{\beta N\leq l\leq\gamma N}\left\{2^{Nh_2(l/N)}D^l 2^{-L}\max_{0\leq x\leq\frac{1}{2}} e^{Lh(x)}\left(1-2x\right)^{lt}\right\} \quad , \qquad (49)\end{aligned}$$

where in the second inequality we used the fact that $\binom{a}{b}\leq e^{ah(b/a)}$ (e.g. [2][Eq. (12.40)]). Hence

$$\frac{\log\sigma_2}{N}\leq\max_{\beta\leq y\leq\gamma}\left\{h_2(y)-(1-R)+\max_{0\leq x\leq 1/2}\{x\log D+(1-R)h_2(x)+yt\log(1-2x)\}\right\}+\frac{o(N)}{N} \qquad (50)$$

Thus if (6) is satisfied then σ_2 is exponentially decreasing with N.

Finally we bound σ_3. From (3), (13) and (10) we have:

$$\begin{aligned}\alpha &= \max_{l>\Gamma}\frac{2^N}{2^{RN}-1}\Pr(A\mathbf{x}=\mathbf{0}|w(\mathbf{x})=l)\leq\max_{l>\Gamma}\frac{2^N}{2^{RN}-1}2\times 2^{-L}\sum_{j=0}^{L/2}\binom{L}{j}\left(1-\frac{2j}{L}\right)^{lt} \\ &\leq (L+1)\max_{l>\Gamma}\max_{0\leq x\leq 1/2}2^{Lh_2(x)}(1-2x)^{lt}\leq(L+1)\max_{0\leq x\leq 1/2}2^{Lh_2(x)}(1-2x)^{\Gamma t}.\end{aligned}$$

Taking the logarithm of α and dividing by N we thus have:

$$\frac{\log\alpha}{N}\leq\max_{0\leq x\leq 1/2}(1-R)h_2(x)+\gamma t\log(1-2x)+\frac{o(N)}{N}=G(R,\gamma t)+\frac{o(N)}{N}. \qquad (51)$$

Substituting (51) in the expression for σ_3 we have:

$$\sigma_3\leq 2^{-NE_r(R+G(R,\gamma t)+o(N)/N)}. \qquad (52)$$

If (6) and (7) are satisfied, (50) and (52) indicate that σ_2 and σ_3 decrease exponentially with N, whereas σ_1 decreases polynomially. Hence (9) follows from (44).

To prove the other parts of the theorem we first show that a sufficient condition for (6) to hold is

$$h_2(\gamma)+(1-R)\left(\log\left(1+e^{-4e^{-12-K}}\right)-1\right)<0 \qquad (53)$$

We begin with

$$\begin{aligned}&\max_{\beta\leq y\leq\gamma}\left\{h_2(y)-(1-R)+\max_{0\leq x\leq 1/2}\{x\log D+(1-R)h_2(x)+yt\log(1-2x)\}\right\}< \\ &\qquad h_2(\gamma)-(1-R)+\max_{0\leq x\leq 1/2}\{(1-R)h_2(x)+\beta t\log(1-2x)\} \qquad (54)\end{aligned}$$

We now use the inequalities

$$1-2x\leq e^{-2x} \qquad (55)$$

and

$$\max_{0\leq x\leq 1} h(x)+ax=\ln(1+e^a), \tag{56}$$

which may be verified by differentiation. Using (55) and (56) yields

$$\begin{aligned}\max_{0\leq x\leq 1/2}\{(1-R)h_2(x)+\beta t\log(1-2x)\} &\leq \max_{0\leq x\leq 1/2}\{(1-R)h_2(x)-2x\beta t\log e\}\\ &\leq (1-R)\log\left(1+e^{-2\beta t/(1-R)}\right)\end{aligned} \tag{57}$$

Equations (54), (57) and (4) show that (53) is indeed sufficient for condition (6) to hold.

Now, given a rate $R<C$, first choose $\gamma>0$ small enough to satisfy (53) for $t=3$. Then (53) is satisfied for any $t\geq 3$, since the left hand side of (53) is monotonically decreasing in t for $t\geq 3$ (using (5) and the fact that $\ln t/t$ is monotonically decreasing in t for $t\geq 3$). Hence (6) is also satisfied. It is easily seen from (8) that as $t\to\infty$, $G(R,\gamma t)\to 0$. Hence, for t large enough (7) can be satisfied for any $R<C$.

Finally, given $t\geq 3$, choose $\gamma>0$ small enough to satisfy (53). Now, it is easily seen from (8) that $G(R,\gamma t)<1-R$. In addition, it is easy to show that as $D\to 0$, $C\to 1$. Hence given $R<1$, (7) can be satisfied by setting $D>0$ small enough. □

D Proof of Lemma 2

To evaluate $\Pr(A\mathbf{x}=\mathbf{0}|w(\mathbf{x})=l)$ consider the following experiment. Let $B_{d\times L}$ be a binary matrix of d rows and L columns. The matrix $B_{d\times L}$ is obtained by randomly choosing lc matrix elements set to '1', with the rest of the matrix set to '0'. Each such matrix is chosen with an equal probability of $1/\binom{Ld}{lc}$. We now claim that $\Pr(A\mathbf{x}=\mathbf{0}|w(\mathbf{x})=l)$ is equal to the probability that each column of B has even weight. To see that, consider the following mapping between B and the bipartite graph corresponding to A. Each column of B corresponds to a right vertex in the graph, and the i-th ($1\leq i\leq d$) element in each column is '1' if the i-th arc of the respective right vertex is connected to a left vertex that is equal to '1'. Thus, $A\mathbf{x}=\mathbf{0}$ if and only if the corresponding B has an even weight for each column. Finally, since $w(\mathbf{x})=l$, exactly lc arcs are connected to a left vertex equal to '1'.

It is clear that if lc is odd, this probability is 0 (Eq. (19)), so we now assume that lc is even. Denote the ensemble of all matrices B such that each column has even weight by $\mathcal{E}$. Let $B_0\in\mathcal{E}$. Then each of its populated columns (columns not identically zero) contains at least two ones. Since B_0 has lc ones, it has at most $\frac{lc}{2}$ populated columns. We conclude that

$$|\mathcal{E}|\leq\binom{L}{\frac{lc}{2}}\binom{\frac{lc}{2}d}{lc}, \tag{58}$$

since by first choosing the populated columns (or any set of columns containing the desired one, in case the desired number of populated columns is less than $\frac{lc}{2}$) and then arbitrarily populating only those columns, we can produce any desired matrix in $\mathcal{E}$. The number of ways of choosing lc positions in B subject to no constraint is $\binom{Ld}{lc}$. Thus from (58):

$$\Pr(A\mathbf{x}=\mathbf{0}|w(\mathbf{x})=l) \leq \binom{L}{\frac{lc}{2}} \frac{\binom{\frac{lc}{2}d}{lc}}{\binom{Ld}{lc}} = \binom{L}{\frac{lc}{2}} \frac{\frac{lc}{2}d}{Ld}\,\frac{\frac{lc}{2}d-1}{Ld-1}...\frac{\frac{lc}{2}d-lc+1}{Ld-lc+1} \leq \binom{L}{\frac{lc}{2}} \left(\frac{lc}{2L}\right)^{lc},$$

(the last inequality holds for $l \leq 2L/c$) which proves (20).

To prove (21), consider some $B_0 \in \mathcal{E}$ again. Denote by m_0 the number of columns of B_0 of weight 0, by m_2 the number of columns of B_0 of weight 2, and so on. We then have

$$\sum_{i=0,2,...} m_i = L \tag{59}$$

and

$$\sum_{i=0,2,...} i m_i = lc. \tag{60}$$

Conversely, for each set of integers $m_0, m_2, ...$ satisfying (59) and (60) there are

$$\binom{L}{m_0 \;\; m_2 \;\; ...} \prod_{i=0,2,...} \binom{d}{i}^{m_i} \tag{61}$$

different matrices in $\mathcal{E}$ satisfying these column population values. The multinomial coefficient enumerates all possibilities of dividing the L columns to subsets of size $m_0, m_2, \ldots$. The binomial coefficient $\binom{d}{i}$ corresponds to choosing the i bits which are set to one in each column. Since there are $\binom{Ld}{lc}$ ways of choosing lc bits out of Ld, we have:

$$\Pr(A\mathbf{x}=\mathbf{0}|w(\mathbf{x})=l) = \frac{1}{\binom{Ld}{lc}} \sum_{\substack{m_0,m_2,... \\ \text{satisfying (59) and (60)}}} \binom{L}{m_0 \;\; m_2 \;\; ...} \prod_{i=0,2,...} \binom{d}{i}^{m_i}. \tag{62}$$

Now consider a new ensemble $\mathcal{F}$ of matrices $B_{d\times L}$ defined as follows. The Ld matrix elements are i.i.d. with $\Pr(B_{i,j}=1)=\frac{l}{N}$. Denote by P_F the probability that a matrix from $\mathcal{F}$ has an even weight for all columns. For given column population values m_0, m_2, ... satisfying (59), the number of different matrices in $\mathcal{F}$ is given by (61). The probability of each of these matrices is

$$\prod_{i=0,2,...} \left[\left(\frac{l}{N}\right)^{i}\left(1-\frac{l}{N}\right)^{d-i}\right]^{m_i}. \tag{63}$$

Thus, from (61) and (63) we have:

$$P_F = \sum_{\substack{m_0,m_2,... \\ \text{satisfying (59)}}} \binom{L}{m_0 \;\; m_2 \;\; ...} \prod_{i=0,2,...} \left[\binom{d}{i}\left(\frac{l}{N}\right)^{i}\left(1-\frac{l}{N}\right)^{d-i}\right]^{m_i}$$

$$
\begin{aligned}
&\geq \sum_{\substack{m_0,m_2,\ldots\\ \text{satisfying (59) and (60)}}} \binom{L}{m_0\ m_2\ \ldots} \left[\prod_{i=0,2,\ldots} \binom{d}{i}^{m_i}\right] \left(\frac{l}{N}\right)^{\sum_i im_i} \left(1-\frac{l}{N}\right)^{d\sum_i m_i - \sum_i im_i} \\
&= \sum_{\substack{m_0,m_2,\ldots\\ \text{satisfying (59) and (60)}}} \binom{L}{m_0\ m_2\ \ldots} \left[\prod_{i=0,2,\ldots} \binom{d}{i}^{m_i}\right] \left(\frac{l}{N}\right)^{lc} \left(1-\frac{l}{N}\right)^{Ld-lc} \qquad (64)
\end{aligned}
$$

Comparing (64) and (62) we obtain:

$$
\begin{aligned}
\Pr(A\mathbf{x}=\mathbf{0}|w(\mathbf{x})=l) &\leq P_F \left[\binom{Ld}{lc}\left(\frac{l}{N}\right)^{lc}\left(1-\frac{l}{N}\right)^{Ld-lc}\right]^{-1} \\
&= P_F \left[\binom{Ld}{lc}\left(\frac{lc}{Ld}\right)^{lc}\left(1-\frac{lc}{Ld}\right)^{Ld-lc}\right]^{-1}. \qquad (65)
\end{aligned}
$$

Now, for any $a \geq b$ (e.g., [2][Eq. (12.40)]),

$$
\binom{a}{b} \geq \frac{1}{a+1} e^{ah(b/a)} = \frac{1}{a+1}\left(\frac{b}{a}\right)^{-b}\left(1-\frac{b}{a}\right)^{b-a}.
$$

In particular,

$$
\binom{Ld}{lc}\left(\frac{lc}{Ld}\right)^{lc}\left(1-\frac{lc}{Ld}\right)^{Ld-lc} \geq \frac{1}{Ld+1}. \qquad (66)
$$

On the other hand, by the definition of the ensemble $\mathcal{F}$, P_F can also be expressed as ([3] [p. 40, Eq. (3.76)])

$$
P_F = \left[\frac{1}{2}\left(1+\left(1-\frac{2l}{N}\right)^d\right)\right]^L \qquad (67)
$$

Substituting (67) and (66) into (65), yields (21). □

E Proof of Theorem 3

Let $\beta < \gamma < 1/2$ be some real number (independent of N) which will be determined later. Suppose first that d is odd. As in the previous section, we set $U = \{1, 2, ..., \Gamma\}$ in (2) ($\Gamma = \gamma N$) to obtain (42), with the same definitions of σ_1, σ_2 and σ_3. We first bound σ_1. From (20), (19) and (10) we have:

$$
\sigma_1 \leq \begin{cases} \sum_{l=1}^{\beta N} D^l \binom{N}{l}\binom{L}{\frac{lc}{2}}\left(\frac{lc}{2L}\right)^{lc} & c \text{ even} \\ \sum_{l=2,4,\ldots}^{\beta N} D^l \binom{N}{l}\binom{L}{\frac{lc}{2}}\left(\frac{lc}{2L}\right)^{lc} & c \text{ odd} \end{cases} \qquad (68)
$$

In Appendix C we showed that (43), (4) and (5) imply (44). The same technique can be used to show that (68), (14) and (15) imply

$$
\sigma_1 \leq \begin{cases} N^{1-\frac{c}{2}} \frac{D}{1-D} \frac{(\frac{c}{2})^c}{\frac{c}{2}!} (1-R)^{-\frac{c}{2}} & c \text{ even} \\ \\ N^{2-c} \frac{D^2}{2(1-D^2)} \frac{c^{2c}}{c!} (1-R)^{-c} & c \text{ odd} \end{cases} \qquad (69)
$$

We now bound σ_2. From (21) and (10) we obtain:

$$\begin{aligned}\sigma_2 &= \sum_{l=\beta N}^{\gamma N} D^l \binom{N}{l}(Ld+1)2^{-L}\left(1+\left(1-\frac{2l}{N}\right)^d\right)^L \\ &\leq N \max_{\beta N \leq l \leq \gamma N} D^l \binom{N}{l}(Ld+1)2^{-L}\left(1+\left(1-\frac{2l}{N}\right)^d\right)^L \qquad (70)\end{aligned}$$

Taking the logarithm of (70) we have:

$$\frac{\log \sigma_2}{N} \leq \max_{\beta \leq x \leq \gamma} x \log D + h_2(x) + (1-R)\left(\log\left(1+(1-2x)^d\right)-1\right) + \frac{o(N)}{N}. \qquad (71)$$

Finally we bound σ_3. From (3), (21) and (10) we have:

$$\alpha = \max_{l>\Gamma} \frac{2^N}{2^{RN}-1} \Pr(A\mathbf{x}=\mathbf{0}|w(\mathbf{x})=l) \leq \max_{l>\Gamma} \frac{2^N}{2^{RN}-1}(Ld+1)2^{-L}\left(1+\left(1-\frac{2l}{N}\right)^d\right)^L,$$

From which we have:

$$\frac{\log \alpha}{N} \leq (1-R)\max_{l>\Gamma} \log\left(1+\left(1-\frac{2l}{N}\right)^d\right) + \frac{o(N)}{N}.$$

Since d is odd, $1+(1-2l/N)^d$ is monotonically decreasing in l. Hence,

$$\frac{\log \alpha}{N} \leq (1-R)\log\left(1+(1-2\gamma)^d\right) + \frac{o(N)}{N}. \qquad (72)$$

Substituting (72) into the expression for σ_3 we have:

$$\sigma_3 \leq 2^{-NE_r(R+(1-R)\log\left(1+(1-2\gamma)^d\right)+o(N)/N)}. \qquad (73)$$

If (16) and (17) are satisfied, (71) and (73) indicate that σ_2 and σ_3 decrease exponentially in N. Since (69) is a polynomial bound for σ_1, (18) follows.

So far we assumed that d was odd. Now suppose d is even. We choose $U = \{1,2,...,\Gamma\} \cup \{N-\Gamma, N-\Gamma+1,...,N\}$ and write $P_e \leq \sum_{i=1}^{6}\sigma_i$, where $\sigma_1 = \sum_{l=1}^{\beta N}\overline{S}_l D^l$, $\sigma_2 = \sum_{l=\beta N}^{\gamma N}\overline{S}_l D^l$, $\sigma_3 = e^{-NE_r(R+(\log\alpha)/N)}$, $\sigma_4 = \sum_{l=(1-\gamma)N}^{(1-\beta)N}\overline{S}_l D^l$, $\sigma_5 = \sum_{l=(1-\beta)N}^{N-1}\overline{S}_l D^l$ and $\sigma_6 = \overline{S}_N D^N$. Now σ_1, σ_2 and σ_3 are still bounded by (69), (71) and (73) respectively (the fact that d was odd was only used in the derivation of (72)).

On the other hand, by the representation given at the beginning of the proof of Lemma 2, for d even we have:

$$\Pr(A\mathbf{x}=\mathbf{0}|w(\mathbf{x})=l) = \Pr(A\mathbf{x}=\mathbf{0}|w(\mathbf{x})=N-l).$$

Thus, by comparing the term with $l=r$ in σ_1 with the term with $l=N-r$ in σ_5, we see that $\sigma_5 \leq \sigma_1 D^{N(1-2\beta)} = \sigma_1 \cdot o(N)/N$. Similarly, $\sigma_4 \leq \sigma_2 D^{N(1-2\gamma)} = \sigma_2 \cdot o(N)/N$. $\sigma_6 = \overline{S}_N D^N \leq D^N$

which decreases exponentially with N. Thus, as $N \to \infty$, σ_1 remains the leading term in the expression for $\overline{P}_e$, so that once again, (18) follows.

To prove the other parts of the theorem we first note that a sufficient condition for (16) to hold is

$$h_2(\gamma) + (1-R)\left(\log\left(1 + e^{-4e^{-12-K}}\right) - 1\right) < 0 \tag{74}$$

This assertion follows from (55) and (14).

Now, given $R_0 < C$, choose some integers c' and d', such that $R_0 < R = 1 - c'/d' < C$. First set γ small enough so that (74) holds for c'. Then set $c = Ic'$, $d = Id'$, where I is an integer large enough so that (17) holds. The left hand side of (74) is monotonically decreasing in c for $c \geq 3$. Thus (74) is valid for any $c \geq c'$. Hence (16) is also satisfied.

Finally, given $c \geq 3$ and d, choose $\gamma > 0$ small enough to satisfy (74). It is easy to show that as $D \to 0$, $C \to 1$. Hence given $R < 1$, (17) can be satisfied by setting $D > 0$ small enough. □

F Proof of Theorem 4

Suppose first that t is even. Denote by P_s the probability that a specific column of A is identically zero. Denote the index of the i'th flipped bit by v_i, $1 \leq i \leq t$. The number of different ways of assigning values to $v_1, ..., v_t$ is L^t. On the other hand, there are $L(L-1)...(L-t/2+1)(t/2)!$ possibilities to have distinct $t/2$ first components followed by the same components, possibly permuted. However, each such assignment of v_i's corresponds to a bit-flipping sequence which results in an all-zero column. Thus:

$$P_s \geq \frac{L!\frac{t}{2}!}{(L-\frac{t}{2})!L^t} \geq \left(1 - \frac{t}{2L}\right)^{t/2} \frac{\frac{t}{2}!}{L^{t/2}} = \left(\frac{t}{2}!\right)[(1-R)N]^{-t/2}\left(1 - \frac{o(N)}{N}\right). \tag{75}$$

Denote by P_s^u the probability that at least one column of A is all-zeros. Since the columns of A are independent, we have:

$$\begin{aligned} P_s^u &= 1 - (1-P_s)^N \geq 1 - \left(1 - \left(\frac{t}{2}!\right)[(1-R)N]^{-t/2}\left(1 - \frac{o(N)}{N}\right)\right)^N \\ &= \left(\frac{t}{2}!\right)(1-R)^{-t/2}N^{1-t/2}\left(1 - \frac{o(N)}{N}\right). \end{aligned}$$

In the last equality we used the fact that $t \geq 3$. By the same argument as in Section IV.1, an all-zero column of A implies a decoding-error with probability ϵ, so that $\overline{P}_e \geq P_s^u \epsilon$, from which (22) follows.

Now suppose t is odd. Then $P_s = 0$. Denote by P_p the probability that two specific columns of A are identical. Then using arguments similar to those that led to (75), we have,

$$P_p \geq t! L^{-t} \left(1 - \frac{o(N)}{N}\right),$$

Let $1 \leq i < j \leq k < l \leq N$ be four indices. Denote by $A_{\cdot i}$ the i'th column of A. Let P'_f be the probability that $A_{\cdot i} = A_{\cdot j}$ and $A_{\cdot k} = A_{\cdot l}$, for the case $j < k$. Similarly, let P''_f be the corresponding probability for the case $j = k$. Since the columns are independent, we have:

$$P'_f = P_p^2 \leq \binom{L}{t}^2 \left(\frac{L}{t}\right)^{-4t} \leq L^{-2t} \frac{t^{4t}}{(t!)^2},$$

using (12) with $l = 2$.

P''_f is equal to the probability that three distinct specific columns of A are identical. Define three i.i.d. random vectors, $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$, chosen in the same manner used to generate vector $\mathbf{v}$ (with $l = 1$) in the proof of Lemma 1 in Appendix B. Further define $\mathbf{s} = \mathbf{u} + \mathbf{v} + \mathbf{w}$. P''_f is the probability that for each $1 \leq n \leq L$, u_n, v_n and w_n have the same parity. This implies $s_n \neq 1$. Since $\sum_n s_n = 3t$, there are at most $(3t-1)/2$ non-zero elements in $\mathbf{s}$. Thus, whereas L^{3t} equiprobable sequences of bit flips constitute the probability space by which $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ (and hence also $\mathbf{s}$) are defined, not more than $\binom{L}{(3t-1)/2}[(3t-1)/2]^{3t}$ yield an $\mathbf{s}$ which complies with this constraint. Thus:

$$P''_f \leq \frac{\binom{L}{\frac{3t-1}{2}} \left(\frac{3t-1}{2}\right)^{3t}}{L^{3t}} \leq L^{-\frac{3t+1}{2}} \frac{\left(\frac{3t-1}{2}\right)^{3t}}{\left(\frac{3t-1}{2}\right)!}.$$

Denote by P_p^u the probability that A has at least one pair of identical columns. Let us denote by E_i the event that the i'th pair of columns are identical, $1 \leq i \leq \binom{N}{2}$. Then

$$\begin{aligned} P_p^u &\geq \sum_i \Pr(E_i) - \sum_{i<j} \Pr(E_i \cap E_j) = \binom{N}{2} P_p - \binom{N}{2}(N-2) P''_f - \frac{1}{2}\binom{N}{2}\binom{N-2}{2} P'_f \\ &\geq \frac{N^2}{2} t! [N(1-R)]^{-t} \left(1 - \frac{o(N)}{N}\right) - N^3 N^{-\frac{3t+1}{2}} \frac{\left(\frac{3t-1}{2}\right)^{3t}}{\left(\frac{3t-1}{2}\right)!} - N^4 N^{-2t} \frac{t^{4t}}{(t!)^2} \\ &= N^{2-t} \frac{t!}{2} (1-R)^{-t} \left(1 - \frac{o(N)}{N}\right) \end{aligned}$$

for $t \geq 3$. A pair of identical columns corresponds to a weight 2 codeword. By an argument analogous to that in Section IV.1, a decoding-error for such codeword occurs with probability of at least ϵ^2. Thus, $\overline{P}_e \geq P_p^u \epsilon^2$, from which (22) follows. □

G Proof of Theorem 5

Only the case of c even is given here since the proof for odd c is completely analogous to that of even c and to the proof for odd t in Appendix F. Denote by P_s the probability that a specific column of A is zero. On the bipartite graph this corresponds to the probability that there is an even number of arcs between some specific left vertex (vertex on the left-hand side of the graph), u, and every right vertex. Denote by $v_i, 1 \leq i \leq c$, the right vertices which are directly connected to u by arcs.

The number of different ways of assigning values to $v_1, ..., v_c$ is $Ld(Ld-1)...(Ld-c+1)$. There are $Ld[(L-1)d]...[(L-c/2+1)d] \times [c/2(d-1)][(c/2-1)(d-1)]...[d-1]$ ways to have the first $c/2$ arcs go to distinct vertices on the right, and the last $c/2$ arcs go to the same $c/2$ vertices on the right, possibly in a permuted order. In each such assignment of v_i's the number of arcs connecting u and any right vertex is either 0 or 2, so that each such graph corresponds to an all-zero column in the corresponding parity-check matrix. Thus:

$$P_s \geq \frac{d^{c/2}(L-c/2)^{c/2}(d-1)^{c/2}\frac{c}{2}!}{(Ld)^c} = \left(1-\frac{1}{d}\right)^{c/2}\frac{c}{2}![N(1-R)]^{-c/2}\left(1-\frac{o(N)}{N}\right)$$

Now denote by P_d the probability that *two* specific columns of A are all zero. Such a matrix corresponds to a bipartite graph for which two specific left vertices are connected to each right vertex with an even (possibly zero) number of arcs. Thus, no more than c right vertices are connected to these two left vertices. Therefore, the number of different assignments of arcs to the two left vertices is not larger than $\binom{L}{c}(cd)^{2c}$, whereas the number of unconstrained arc assignments to the two vertices is $(Ld)(Ld-1)...(Ld-2c+1)$. Thus, we have

$$P_d \leq \frac{\binom{L}{c}(cd)^{2c}}{(Ld-2c)^{2c}} \leq \frac{L^c(cd)^{2c}}{(Ld-2c)^{2c}} = O(N^{-c}).$$

Finally, denote by P_s^u the probability that at least one column of A is all-zeros. Then:

$$P_s^u \geq NP_s - \binom{N}{2}P_d > N\left(1-\frac{1}{d}\right)^c\frac{c}{2}![(1-R)N]^{-c/2}\left(1-\frac{o(N)}{N}\right) - N^2O(N^{-c}),$$

which implies (24) for $c \geq 3$. □

Acknowledgment

The authors would like to thank Nadav Shulman for helpful discussions and in particular to his contributions that led to some of the results on lower bounds and expurgated ensembles.

References

[1] C. Berrou, A. Glavieux and P. Thitimajshima, “Near Shanon limit error correcting coding and decoding: turbo codes”, *Proceedings 1993 IEEE International Conference on Communications*, Geneva, Switzerland, pp. 1064-1070, 1993.

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

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

[4] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[5] S. Litsyn and V. Shevelev, “On Ensembles of Low-Density Parity-Check Codes: distance distributions”, submitted for publication, *IEEE Trans. Inform. Theory.*

[6] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi and D. A. Spielman, “Improved low-density parity-check codes using irregular graphs”, *IEEE Trans. Inform. Theory*, vol. 47, pp. 585–598, February 2001.

[7] D. J. C. MacKay, “Good error-correcting codes based on very sparse matrices”, *IEEE Trans. Inform. Theory*, vol. 45, pp. 399-431, March 1999.

[8] G. Poltyrev, “Bounds on the decoding error probability of binary linear codes via their spectra”, *IEEE Trans. Inform. Theory*, vol. 40, pp. 1284-1292, July 1994.

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

[10] I. Sason and S. Shamai, “Improved upper bounds on the ensemble performance of ML decoded low density parity check codes”, *IEEE Communications Letters*, vol. 4, pp. 89–91, March 2000.

[11] N. Shulman and M. Feder, “Random coding techniques for nonrandom codes”, *IEEE Trans. Inform. Theory*, vol. 45, pp. 2101-2104, September 1999.

[12] M. Sipser and D. Spielman, “Expander Codes”, *IEEE Trans. Inform. Theory*, vol. 42, pp. 1710–1723, November 1996.

[13] V. V. Zyablov and M. S. Pinsker, “Estimation of the error-correction complexity for Gallager low-density codes”, *Probl. Inform. Transm.*, vol. 11, no. 1, pp. 18-28, May 1976.

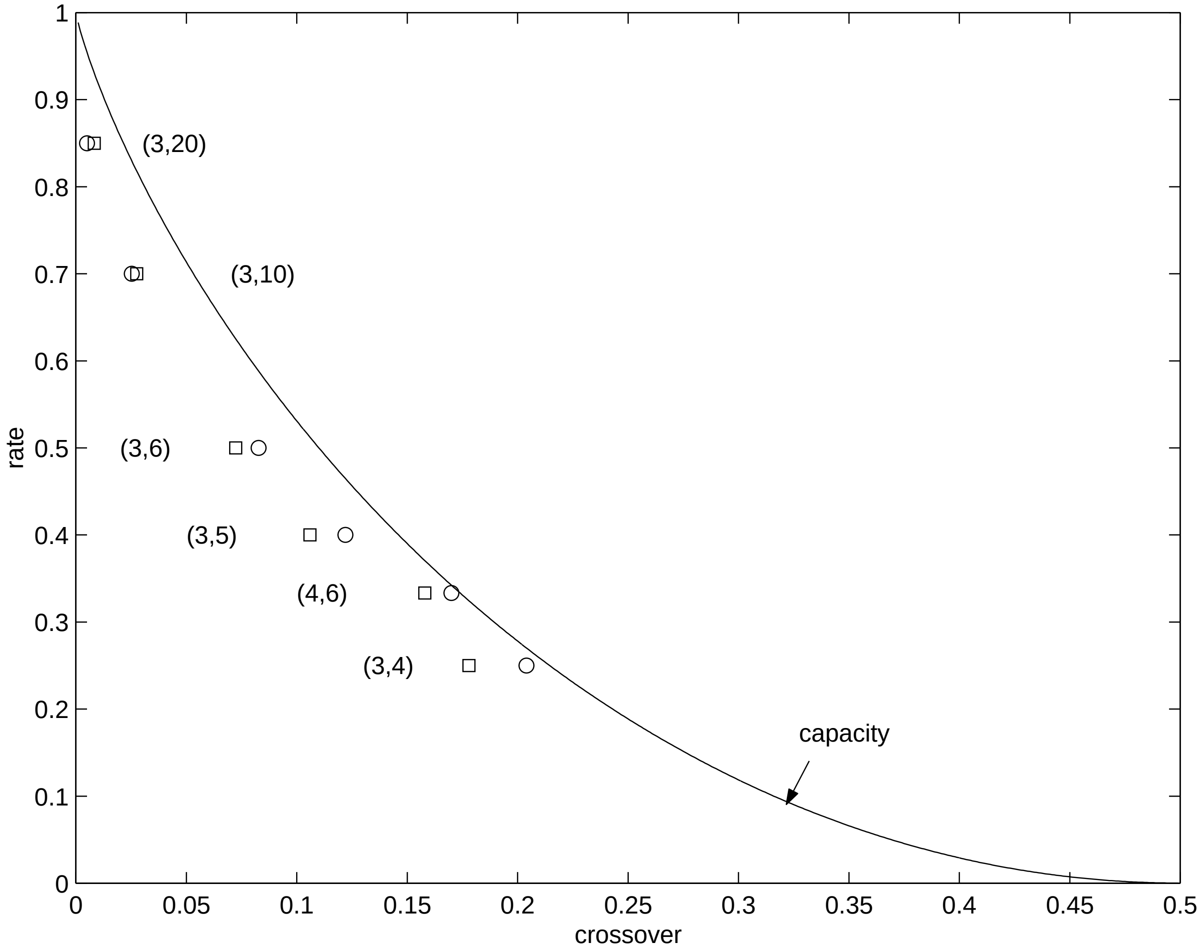

Figure 1: Lower bounds to the maximum correctable crossover on a BSC when using Gallager's bound (circles) and when using the new bound (squares)

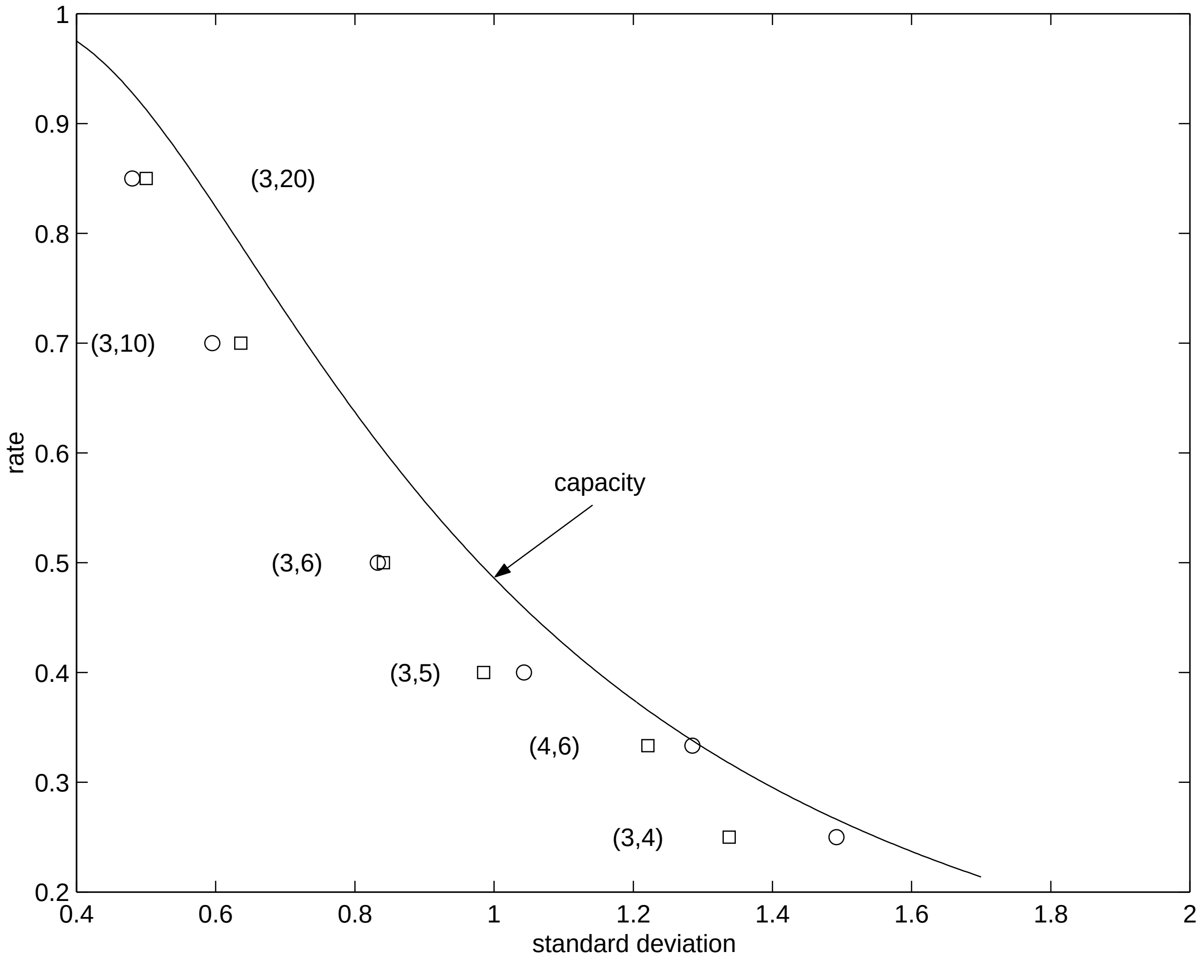

Figure 2: Lower bounds to the maximum correctable standard deviation on an AWGN channel when using Gallager's bound (circles) and when using the new bound (squares)