Upper Bounds on the Rate of LDPC Codes *

David Burshtein †, Michael Krivelevich ‡, Simon Litsyn * and Gadi Miller *

Abstract

We derive upper bounds on the rate of low density parity check (LDPC) codes for which reliable communication is achievable. We first generalize Gallager's bound to a general binary-input symmetric-output channel. We then proceed to derive tighter bounds. We also derive upper bounds on the rate as a function of the minimum distance of the code. We consider both individual codes and ensembles of codes.

Index Terms - Low density parity check (LDPC) codes, iterative decoding, maximum-likelihood decoding, error probability, minimum distance.

I Introduction

Low density parity check (LDPC) codes were proposed by Gallager [9] in 1963. Gallager demonstrated that these codes possess some very desirable properties, and can be used to transmit information at rates close to channel capacity. In addition to that, Gallager proposed a practical iterative algorithm that can be used for decoding these codes. Gallager's pioneering work was relatively ignored for about three decades, until the recent introduction of turbo codes [3] in 1993. Both turbo codes and LDPC codes have recently attracted significant commercial and academic interest.

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 [9, 11, 13, 17]. The second relates to the properties of practical iterative decoding algorithms for these codes [4, 5, 9, 12, 18, 19, 20].

*This research was supported by the Israel science foundation, grants No. 553–00 and 22/01–1.

†Dept. of Elect. Eng. – Systems, Tel-Aviv University, Tel-Aviv, Israel.

‡Dept. of Mathematics, Tel-Aviv University, Tel-Aviv, Israel.

In this paper we consider the first problem and present upper bounds on the rate of both individual LDPC codes and ensembles of codes when operated over an arbitrary binary-input symmetric-output channel. Most of the research on the properties of ML decoding is concerned with lower bounds on the performance. To the best of our knowledge, only Gallager [9] derived upper bounds on the performance of ML decoding of LDPC codes. Unlike the lower bounds, the upper bounds on the performance of ML decoding also apply to any other decoding algorithm, such as iterative decoding. In Section II we provide some background on LDPC codes and their graph representation. In Section III we present a lower bound on the rate required for reliable communication. This bound is a generalization of a result first obtained by Gallager for the BSC [9]. In Section IV we provide some graph theoretic results on row matching in sparse matrices. In Section V we use these results to tighten the bound on the rate. In Section VI we present upper bounds on the rate of LDPC codes as a function of the minimum distance. In Section VII we conclude the paper and discuss further possibilities for improving the obtained bounds.

II Background

A code can be represented by its parity check matrix. Alternatively, we can use a bipartite graph representation, in which there is one set of N variable (left) nodes and another set of L parity check (right) nodes. 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 node on the right and j-th node on the left, is set to '1' if there is an odd number of edges between the two nodes, and to '0' otherwise. In this paper we consider both regular and irregular bipartite graphs. The (c, d)-regular ensemble that we consider in this paper is constructed as follows. For each variable node we assign c variable sockets. Similarly, for each check node we assign d check sockets. The total number of variable sockets, Nc, is equal to the number of check sockets, Ld. The ensemble of bipartite graphs is obtained by choosing a permutation π with uniform probability from the space of all permutations of size Nc. For each $1 \leq i \leq Nc$ the i-th variable socket is matched with the $\pi(i)$-th check socket by an edge. Note that in this way multiple edges may link a pair of nodes. Figure 1 demonstrates this construction. The irregular ensemble is constructed in a similar way [12], where the degrees of the nodes are chosen according to some given profile.

Let R denote the rate of the code. We define R' by $L = (1 - R')N$. Note that $R \geq R'$ due to a possible degeneracy in the parity check equations.

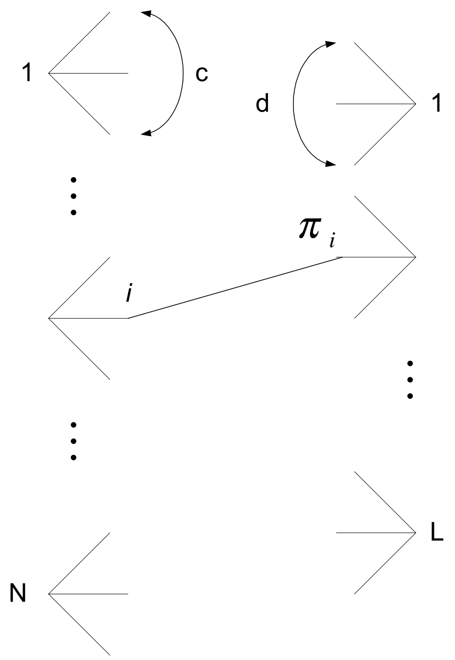

Figure 1: Construction of a random $c-d$ regular graph

III A Generalization of Gallager's bound

In [9], Gallager proposed an upper bound on the rate required for reliable communication when using an LDPC code over the BSC. In this section we show how Gallager's bound can be generalized to an arbitrary memoryless binary-input symmetric-output channel.

Consider a binary-input symmetric-output channel with input symbol $X \in \{0,1\}$ and output symbol $Y \in \mathcal{R}$ i.e. $P(Y=y|X=1) = P(Y=-y|X=0) = f(y)$. We define the crossover probability of the channel, ϵ, as

$$\epsilon = \frac{1}{2}\int_{-\infty}^{\infty} \min\left(f(y), f(-y)\right) dy \tag{1}$$

Note that our notation assumes a continuous channel. However, our results also apply to the case of discrete or mixed continuous-discrete channels.

We say that a sequence of codes can be used for reliable communication over some given channel if the maximum likelihood (ML) decoding error probability approaches 0 as the block length approaches infinity.

Theorem 1 *Consider a binary code with parity check matrix $A_{L\times N}$ and rate R over a memoryless binary-input symmetric-output channel with crossover probability ϵ defined in (1). Suppose that A has the property that all its rows have a constant weight d. Then a necessary condition for reliable communication is:*

$$R \leq 1 - \frac{1-C}{h(\epsilon_d)} \tag{2}$$

where

$$\epsilon_d = \frac{1}{2}\left(1-(1-2\epsilon)^d\right) \tag{3}$$

$h(x) = -x\log_2 x - (1-x)\log_2(1-x)$ *is the binary entropy function and C is the channel capacity.*

Note: For the BSC case, Theorem 1 reduces to

$$R \leq 1 - \frac{h(\epsilon)}{h(\epsilon_d)}$$

which is [9][Equation (3.71)].

Proof. To prove the theorem we show that if (2) is not satisfied, the ML decoding error probability is bounded away from 0 by a quantity independent of N.

Denote the transmitted codeword by $\mathbf{X} = (X_1,\ldots,X_N)$ and the received word by $\mathbf{Y} = (Y_1,\ldots,Y_N)$. In order to keep our notation simple, we use the same notation $H(\mathbf{Z})$ to denote both the entropy of the discrete random variable (r.v.) $\mathbf{Z}$ and the differential entropy of the continuous r.v. $\mathbf{Z}$. $I(\mathbf{U};\mathbf{V})$ denotes the mutual information between the r.v.-s $\mathbf{U}$ and $\mathbf{V}$. Using

$$I(\mathbf{X};\mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}\mid\mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y}\mid\mathbf{X})$$

we have:

$$H(\mathbf{X}\mid\mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{Y}) + H(\mathbf{Y}\mid\mathbf{X}) \tag{4}$$

Now,

$$H(\mathbf{X}) = RN \tag{5}$$

and

$$\begin{aligned} H(\mathbf{Y}\mid\mathbf{X}) &= \sum_{l=1}^{N} H(Y_l\mid Y_1,\ldots,Y_{l-1},\mathbf{X}) \\ &= NH(Y\mid X) \end{aligned} \tag{6}$$

where $X = X_1$ and $Y = Y_1$. The last equality is due to the fact that given X_l, Y_l is independent of $Y_1,\ldots,Y_{l-1},X_1,\ldots,X_{l-1},X_{l+1},\ldots,X_N$. Thus, to lower bound $H(\mathbf{X}\mid\mathbf{Y})$ using (4) we need to upper bound $H(\mathbf{Y})$.

To this end, we define a binary $\{0,1\}$ r.v. Z_l, $l = 1, \ldots, N$, as follows:

$$
\begin{aligned}
P\left(Z_l = 1 \mid f(Y_l) > f(-Y_l)\right) &= 1 \\
P\left(Z_l = 1 \mid f(Y_l) < f(-Y_l)\right) &= 0 \\
P\left(Z_l = 1 \mid f(Y_l) = f(-Y_l)\right) &= \frac{1}{2}
\end{aligned}
$$

Thus,

$$
P(Z=0) = P\left(f(Y) < f(-Y)\right) + \frac{1}{2} P\left(f(Y) = f(-Y)\right) = 1/2
$$

where $Y = Y_l$ and $Z = Z_l$. Thus,

$$
\begin{aligned}
H(Y \mid Z) &= H(Y) - I(Y;Z) \\
&= H(Y) - H(Z) + H(Z \mid Y) \\
&= H(Y) - 1 + H(Z \mid Y)
\end{aligned}
\tag{7}
$$

Let $\mathbf{Z} = (Z_1, \ldots, Z_N)$. We also have

$$
\begin{aligned}
H(\mathbf{Y}) - H(\mathbf{Y} \mid \mathbf{Z}) &= I(\mathbf{Y};\mathbf{Z}) \\
&= H(\mathbf{Z}) - H(\mathbf{Z} \mid \mathbf{Y}) \\
&= H(\mathbf{Z}) - NH(Z \mid Y)
\end{aligned}
\tag{8}
$$

The last equality follows from the same considerations used in (6). From (8) we get

$$
H(\mathbf{Y}) = H(\mathbf{Z}) - NH(Z \mid Y) + H(\mathbf{Y} \mid \mathbf{Z}) \tag{9}
$$

Now,

$$
\begin{aligned}
H(\mathbf{Y} \mid \mathbf{Z}) &= \sum_{l=1}^{N} H(Y_l \mid Y_1, \ldots, Y_{l-1}, \mathbf{Z}) \\
&\leq NH(Y \mid Z)
\end{aligned}
\tag{10}
$$

since conditioning reduces entropy [7][p. 27]. Equations (9), (10) and (7) yield

$$
H(\mathbf{Y}) \leq H(\mathbf{Z}) + NH(Y) - N \tag{11}
$$

Recalling (1) we have $P(X_l \neq Z_l) = \epsilon$. Hence, $X \to Z$ is a BSC with crossover parameter ϵ. It is shown in Appendix A that

$$
H(\mathbf{Z}) = RN + H(\mathbf{S}) \tag{12}
$$

where $\mathbf{S} = A\mathbf{Z}$ is the syndrome. Equations (4), (5), (6), (11) and (12) yield

$$
H(\mathbf{X} \mid \mathbf{Y}) \geq N(1 - C) - H(\mathbf{S}) \tag{13}
$$

where we used the fact that $I(X;Y) \leq C$ (C is the channel capacity).

We now turn to derive Gallager's bound on $H(\mathbf{S})$. Let S_l denote the l-th component of the syndrome, $\mathbf{S}$. By the chain rule for entropy,

$$H(\mathbf{S}) = \sum_{l=1}^{L} H(S_l \mid S_{l-1}, \ldots, S_1)$$

Now, $\operatorname{rank}(A) = (1-R)N$. Hence, without loss of generality we may assume that the first $(1-R)N$ rows of A are linearly independent and that all other rows are linear combinations of the first $(1-R)N$ rows. Let $\mathbf{A}_l$ denote the l-th row of A. If $\mathbf{A}_l = \sum_{i=1}^{l-1} \alpha_i \mathbf{A}_i$ for some $\alpha_i \in \{0,1\}$, $i = 1, \ldots, l-1$, then $S_l = \sum_{i=1}^{l-1} \alpha_i S_i$. Hence, $H(S_l \mid S_{l-1}, \ldots, S_1) = 0$ for $l > (1-R)N$. Therefore,

$$\begin{aligned} H(\mathbf{S}) &= \sum_{l=1}^{(1-R)N} H(S_l \mid S_{l-1}, \ldots, S_1) \\ &\leq \sum_{l=1}^{(1-R)N} H(S_l) \\ &= N(1-R) H(S_1) \\ &= N(1-R) h(\epsilon_d) \end{aligned} \tag{14}$$

where ϵ_d is given by (3). Substituting (14) in (13) we obtain

$$H(\mathbf{X} \mid \mathbf{Y}) \geq N\left[1 - C - (1-R)h(\epsilon_d)\right] \tag{15}$$

If $H(\mathbf{X} \mid \mathbf{Y})/N$ is bounded away from zero, then by Fano's inequality [7][p. 39], so is the decoding error probability. The claim of the theorem follows from this argument and (15). $\square$

The following generalization of Theorem 1 to the irregular case is immediate:

Theorem 2 *Consider a binary code with parity check matrix $A_{L\times N}$ and rate R over a memoryless binary-input symmetric-output channel with crossover probability ϵ. Assume, without loss of generality, that the first $(1-R)N$ rows of A are linearly independent, and suppose that A has the property that a p_d fraction of its first $(1-R)N$ rows has weight d. Then a necessary condition for reliable communication is:*

$$R \leq 1 - \frac{1-C}{\sum_d p_d h(\epsilon_d)} \tag{16}$$

where ϵ_d is defined in (3), $h(\cdot)$ is the binary entropy function and C is the channel capacity.

Note: If in Theorem 2, p_d is the fraction of all rows of A with weight d, then R in (16) should be replaced by R'.

The implication of the theorem is that a necessary condition for capacity approaching codes is $\sum_d p_d h(\epsilon_d) \to 1$. Hence it is necessary to have $\sum_{d=1}^{k} p_d \to 0$ for any fixed k.

IV Row matching

Before proceeding to tighten the bounds on the rate, we first derive some graph theoretic results that concern row matching in sparse matrices.

IV.1 Matching pairs of rows

Consider some matrix A over $GF(2)$ and suppose that A' is obtained by permuting the rows of A using a permutation $\pi \in S_L$, where S_L is the set of all length L permutations. Let the rows of A be denoted by $\mathbf{v}_1, ..., \mathbf{v}_L$. Then the i-th row of A' is $\mathbf{v}_{\pi(i)}$. Suppose that A' satisfies the following property for some even integer $M \leq L$. For all $1 \leq i \leq M/2$, $\mathbf{v}_{\pi(2i)} \times \mathbf{v}_{\pi(2i-1)} \neq \mathbf{0}$, where '$\times$' denotes bit-wise logical AND. That is, there is at least one position in which both $\mathbf{v}_{\pi(2i)}$ and $\mathbf{v}_{\pi(2i-1)}$ are 1. We say that $\mathbf{v}_{\pi(2i)}$ and $\mathbf{v}_{\pi(2i-1)}$ *overlap*. Denote by $\mu(A)$ the maximal number of matched pairs that can be obtained. Formally,

$$\mu(A) = \max_{\pi \in S_L} \min \left\{ 1 \leq i \leq \frac{L}{2} \;\middle|\; \mathbf{v}_{\pi(2i)} \times \mathbf{v}_{\pi(2i-1)} = \mathbf{0} \right\} - 1$$

In addition, if $\mathbf{v}_{\pi(2i)} \times \mathbf{v}_{\pi(2i-1)} \neq \mathbf{0}$ for all $1 \leq i \leq \frac{L}{2}$ then $\mu(A) = L/2$. Further let $\Delta(A)$ denote the fraction of rows that cannot be matched, i.e.

$$\Delta(A) = \frac{1}{N} \left(L - 2\mu(A) \right) \tag{17}$$

Lemma 1 *Suppose that the weight of each row in A is at least d. Then $\Delta(A) \leq 1/d$.*

Proof. Consider the following greedy matching algorithm. Suppose we have already matched $2n$ rows. We then look for some match in the remaining $L - 2n$ rows, and continue in this way until we can no longer find a match. Denote the number of pairs obtained in this way by $M/2$. When the algorithm terminates, we are left with a set $\mathcal{S}$ of $l = L - M$ rows which do not match. Denote the (length N) rows of this set by $\mathbf{u}_1, ..., \mathbf{u}_l$. The fact that a match cannot be found in $\mathcal{S}$ implies that for any component $1 \leq j \leq N$, this component is 1 in at most one row in $\mathcal{S}$, i.e.,

$$\sum_{i=1}^{l} u_i^j \leq 1$$

where u_i^j is the j-th component of $\mathbf{u}_i$. Thus,

$$\sum_{j=1}^{N} \sum_{i=1}^{l} u_i^j \leq N \tag{18}$$

On the other hand, for any $1 \leq i \leq l$ we have:

$$\sum_{j=1}^{N} u_i^j \geq d$$

so that

$$\sum_{i=1}^{l} \sum_{j=1}^{N} u_i^j \geq ld \tag{19}$$

(18) and (19) imply $l \leq N/d$, from which the claim of the lemma follows. $\square$

In order to improve the bound in Lemma 1 we now introduce the concept of a *secondary graph*. Let G be a (c, d)-regular bipartite graph that corresponds to the parity check matrix A. We call this representation the *primary graph*. Denote the *check* nodes of G by $v_1, ..., v_L$. The secondary graph of G (denoted by G') is a graph comprising L vertices $u_1, ..., u_L$ such that an edge connects u_i and u_j iff there exists a variable node w such that the two pairs (w, v_i) and (w, v_j) are connected by edges in G. Put differently, u_i and u_j are connected in G' iff the distance between v_i and v_j in the primary graph is exactly two. An example of the relation between the primary and secondary graphs is shown in Figure 2. Since there is a correspondence between u_i

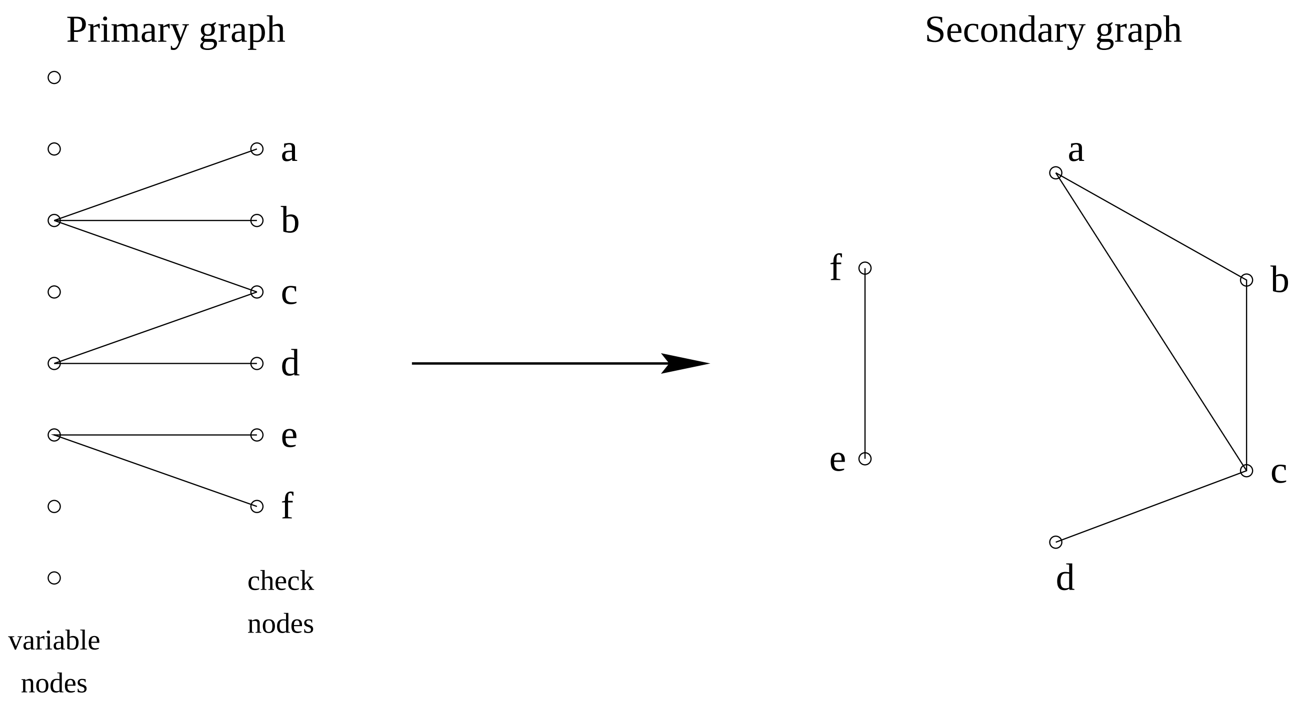

Figure 2: The relation between the primary and secondary graphs.

and v_i, we may refer to both by the same name (e.g. v_i) and no confusion results, as long as we keep in mind the graph to which these vertices belong.

It follows from the definition of the secondary graph, that vertices v_i and v_j in G' are connected by an edge iff the i-th and j-th rows of A overlap. Thus, $\mu(A)$ is equal to the size of the maximal matching in G' (a matching is a set of n edges that connects exactly $2n$ nodes).

Lemma 1 may now be improved as follows.

Lemma 2 *Suppose that the weight of each row in A is d and the weight of each column is c. Further suppose that each pair of rows overlap in at most one position. Then*

$$\Delta(A) \leq \frac{L}{N}\frac{1}{d(c-1)+1}$$

Proof. Since G is (c,d)-regular and has no length four cycles (recall that two rows of A overlap in at most one position), G' is $d(c-1)$-regular with L nodes. We now use the following auxiliary lemma.

Lemma 3 *Let G' be a graph with e edges, and with all vertex degrees at most k. Then G' contains a matching of size at least $e/(k+1)$.*

Proof. According to Vizing's theorem [23], the edge set of a graph G' of maximal degree k can be decomposed into $k+1$ matchings. Then clearly one of the matchings in such a decomposition contains at least $e/(k+1)$ edges. □

In our case $k = d(c-1)$ and $e = Lk/2$. The claim of Lemma 2 follows immediately. □

In Appendix B we prove the following.

Lemma 4 *Let A be drawn from the (c,d)-regular ensemble. Then for any $\delta > 0$,*

$$\lim_{N\to\infty} P(\Delta(A) < \delta) = 1$$

IV.2 Matching multiple rows

We now consider the problem of matching i-tuples of rows instead of pairs. More precisely, given a $0,1$-matrix A with L rows and N columns, a partition $[L] = \mathcal{S}_1 \cup \ldots \cup \mathcal{S}_k$ of its rows is called *feasible* if for every $1 \leq i \leq k$ there exists a column $1 \leq j \leq N$, where all rows of $\mathcal{S}_i$ have value 1 (they overlap at the same position). If A has at most c ones in each column, then clearly each block of a feasible partition has at most c elements. For such matrix A and a feasible partition $(\mathcal{S}_1, \ldots, \mathcal{S}_k)$, we characterize the partition by a vector $\bar{t} = (t_1, \ldots, t_c)$, where t_i is the number of blocks of size i in the partition ($\sum_i it_i = L$).

In Appendix C we prove the following.

Lemma 5 *Let $A_{L\times N}$ be a binary matrix. Assume that each row of A has a constant weight d, and each column of A has a constant weight c. Then there exists a feasible row partition, whose vector $(t_1,\ldots,t_c)$ satisfies:*

$$t_i \geq \frac{d(L-\sum_{j=i+1}^{c} jt_j)-(i-1)(N-\sum_{j=i+1}^{c} t_j)}{(d-1)i+1}\ ; \quad 2\leq i\leq c\ . \tag{20}$$

In Appendix D we prove the following.

Lemma 6 *Let $A_{L\times N}$ be a random binary matrix drawn from the (c,d)-regular ensemble. Assume that $c<1+\frac{\log d}{(d-1)\log\frac{d}{d-1}}$. Then almost surely there exists a feasible partition of the rows of A with $t_c=(1-o(1))L/c$.*

V A tighter bound on the rate

In this section we show that the bound in Theorem 1 is not tight, by improving the bound on $H(\mathbf{S})$. Recall that (14) follows by using the inequality

$$H(\mathbf{S})\leq \sum_{i=1}^{N(1-R)} H(S_i) \tag{21}$$

V.1 Matching pairs of rows

To improve the bound (21) we proceed as follows. Given a parity check matrix $A_{L\times N}$ and recalling the definition of $\Delta(A)$ in (17), there exists a row permuted matrix A' such that M, the number of rows that are matched, can be chosen as

$$M=L-N\Delta(A) \tag{22}$$

Using A', we now bound $H(\mathbf{S})$ as follows

$$H(\mathbf{S})\leq \sum_{i=1}^{M/2} H(S_{2i-1},S_{2i})+\sum_{i=M+1}^{L} H(S_i) \tag{23}$$

Now, assuming that the weight of each row of A is d, $P(S_i=1)=\epsilon_d$ for each i. Thus, $H(S_i)=h(\epsilon_d)$. Suppose that the overlap between the first and second rows of A' is exactly 1. We now calculate the joint distribution of (S_1,S_2). Denote the joint (overlap) noise bit by v. Then $P(v=1)=\epsilon$.

$$\begin{aligned} P(S_1=1,S_2=1) &= P(S_1=1,S_2=1|v=1)P(v=1)+P(S_1=1,S_2=1|v=0)P(v=0) \\ &= \epsilon(1-\epsilon_{d-1})^2+(1-\epsilon)\epsilon_{d-1}^2 \end{aligned} \tag{24}$$

Similarly,

$$P(S_1 = 0, S_2 = 0) = (1-\epsilon)(1-\epsilon_{d-1})^2 + \epsilon\epsilon_{d-1}^2$$

and

$$P(S_1 = 0, S_2 = 1) = P(S_1 = 1, S_2 = 0) = (1-\epsilon_{d-1})\epsilon_{d-1}$$

Thus, $H(S_1, S_2) = 2\psi_2(\epsilon, d)$ where

$$\begin{aligned} \psi_2(\epsilon, d) &= \frac{1}{2}\hat{h}\left(\epsilon(1-\epsilon_{d-1})^2 + (1-\epsilon)\epsilon_{d-1}^2,\ (1-\epsilon)(1-\epsilon_{d-1})^2 + \epsilon\epsilon_{d-1}^2,\right. \\ &\qquad \left.\epsilon_{d-1}(1-\epsilon_{d-1}),\ \epsilon_{d-1}(1-\epsilon_{d-1})\right) \end{aligned} \tag{25}$$

where $\hat{h}(\cdot,\cdot,\cdot,\cdot)$ is the entropy function of a four-valued random variable. $H(S_1, S_2) = 2\psi_2(\epsilon, d)$ was derived under the assumption that the overlap between the first and second rows is exactly one. In our case, we know that the overlap is at least one. We thus make use of the following lemma (sums of binary variables are taken modulo 2):

Lemma 7 *Let $\{X_1, ..., X_d, Y_1, ..., Y_d\}$ be $2d$ i.i.d. binary r.v.-s such that $P(X_1 = 1) = p < 1/2$. Denote $U = \sum_{i=1}^{d} X_i$ and $V = \sum_{i=1}^{k} X_i + \sum_{i=1}^{d-k} Y_i$ for $0 \le k \le d$. Then $H(U, V)$ is monotonically decreasing in k.*

The proof is provided in Appendix E. At this point we distinguish between two cases.

V.1.1 Bounds for individual codes

It follows from Lemma 7 that for $1 \le i \le M/2$, $H(S_{2i-1}, S_{2i}) \le 2\psi_2(\epsilon, d)$. Thus, from (23), $H(\mathbf{S}) \le M\psi_2(\epsilon, d) + (L - M)h(\epsilon_d)$. Using (13) and (22) this gives:

$$H(\mathbf{X} \mid \mathbf{Y}) \ge N\left[1 - C - (1 - R')\psi_2(\epsilon, d) - \Delta(A)\left(h(\epsilon_d) - \psi_2(\epsilon, d)\right)\right]$$

Again, if $H(\mathbf{X} \mid \mathbf{Y})/N$ is bounded away from 0 then by Fano's inequality, so is the decoding error probability. This yields the following:

Theorem 3 *Consider a binary code with parity check matrix $A_{L\times N}$ over a memoryless binary-input symmetric-output channel with crossover probability ϵ. Suppose that A has the property that all its rows have a constant weight d. Then a necessary condition for reliable communication is:*

$$R' \le 1 - \frac{1 - C}{\psi_2(\epsilon, d)} + \Delta(A)\left(\frac{h(\epsilon_d)}{\psi_2(\epsilon, d)} - 1\right)$$

Let $\tilde{A}$ be a matrix obtained by choosing some $(1-R)N$ linearly independent rows from A. Applying Theorem 3 to $\tilde{A}$ yields,

$$R \leq 1 - \frac{1-C}{\psi_2(\epsilon,d)} + \Delta(\tilde{A})\left(\frac{h(\epsilon_d)}{\psi_2(\epsilon,d)} - 1\right)$$

Now, $h(\epsilon_d) \geq \psi_2(\epsilon,d)$ (since $H(S_1,S_2) \leq H(S_1)+H(S_2)$). Hence this bound is monotonically increasing in $\Delta(A)$.

V.1.2 Bounds for ensembles of codes

Theorem 3 can readily be generalized to ensembles as follows:

Theorem 4 *Consider an ensemble of parity check matrices $A_{L\times N}$ over a memoryless binary-input symmetric-output channel with crossover probability ϵ. Denote $R' = 1 - L/N$. Let $A = (\mathbf{v}_1, ..., \mathbf{v}_L)$ be a matrix randomly drawn from the ensemble, where $\mathbf{v}_i$ denotes the i-th row of A. Suppose that the following two conditions are satisfied for some $D > 0$ and any $\delta > 0$:*

$$\lim_{N\to\infty} P\left(\frac{|\{\mathbf{v}_i | w(\mathbf{v}_i) \neq d\}|}{L} < \delta\right) = 1 \tag{26}$$

where $w(\mathbf{v}_i)$ denotes the weight of $\mathbf{v}_i$, and

$$\lim_{N\to\infty} P(\Delta(A) < D + \delta) = 1$$

Then a necessary condition for reliable communication is:

$$R' \leq 1 - \frac{1-C}{\psi_2(\epsilon,d)} + D\left(\frac{h(\epsilon_d)}{\psi_2(\epsilon,d)} - 1\right)$$

Note that by reliable communication for an ensemble, we mean that a typical code in the ensemble achieves reliable communication.

By applying Lemma 4 to the (c,d)-regular ensemble, we see that D in Theorem 4 may be taken as 0 to obtain the tightest bound. In addition, it is easy to verify that condition (26) is satisfied for the (c,d)-regular ensemble.

V.2 Matching multiple rows

To further improve the bound (21) we now match i-tuples of rows instead of pairs. We first generalize Theorem 3. We assume that the number of pairs of rows with an overlap larger than one is a negligible fraction of L. That is, we assume that the number of cycles of size four in the bipartite graph is negligible compared to L. In fact, by Lemma 9 in Appendix B, this property holds for a typical code in the (c,d)-regular ensemble. Following the derivation of Theorem 3 we have

$$H(\mathbf{S}) \leq \sum_{i=1}^{c} i t_i \psi_i(\epsilon, d) \tag{27}$$

where $\psi_i(\epsilon, d) = H(S_1, \ldots, S_i)/i$ under the assumption that the first i rows constitute a block. In particular, $\psi_1(\epsilon, d) = h(\epsilon_d)$ and $\psi_2(\epsilon, d)$ is consistent with (25). The generalization to an arbitrary i is

$$\psi_i(\epsilon, d) = -\frac{1}{i} \sum_{j=0}^{i} \binom{i}{j} \tilde{\epsilon} \log \tilde{\epsilon}$$

where

$$\tilde{\epsilon} = (\epsilon_{d-1})^j (1-\epsilon_{d-1})^{i-j} (1-\epsilon) + (1-\epsilon_{d-1})^j (\epsilon_{d-1})^{i-j} \epsilon$$

Note that $\psi_i(\epsilon, d)$ is monotonically decreasing in i, since

$$H(S_1, \ldots, S_i) = \sum_{j=1}^{i} H\left(S_j \mid S_1, \ldots, S_{j-1}\right)$$

and since $H(S_j \mid S_1, \ldots, S_{j-1})$ is monotonically decreasing in j. Hence i-tuples may offer an improvement over pair matching.

Thus, using (13) we obtain the following.

Theorem 5 *Consider a binary code with parity check matrix $A_{L \times N}$ over a memoryless binary-input symmetric-output channel with crossover probability ϵ, and assume that the number of pairs of rows with an overlap larger than one is a negligible fraction of L. Further suppose that A has the property that all its rows have a constant weight d. Then a necessary condition for reliable communication is:*

$$R' \leq 1 - \frac{1 - C}{\sum_i \tau_i \psi_i(\epsilon, d)}$$

where $\tau_i = i t_i / L$ and $(t_1, t_2, \ldots)$ corresponds to a feasible partition of A.

Theorem 5 may now be used in conjunction with Lemma 5 or Lemma 6 to derive tighter bounds on the rate.

In Table 1 we present lower and upper bounds on the threshold crossover probability of a BSC for reliable communication. The given upper bound on the ML decoding error probability is the tighter among the bound in [9] and the bound in [17]. We also provide the actual threshold of the belief propagation algorithm as evaluated using density evolution. Note that the threshold of belief propagation is a lower bound on the ML threshold. In Table 2 we consider the binary input additive white Gaussian noise (BIAWGN) channel. In this case the input to the channel is taken from $\{\pm 1\}$, and the variance of the additive white Gaussian noise is σ^2. Note that the bound in

(c,d)	(3,6)	(3,7)	(4,6)	(3,10)	(3,20)
BP	0.0840	0.0654	0.116	0.0374	0.0132
ML lower	0.0827	0.055	0.170	0.0275	0.0083
Gallager	0.10245	0.07964	0.17263	0.04553	0.01621
Pairs	0.10234	0.07954	0.17262	0.04548	0.01621
Clusters	–	0.07943	–	0.04542	0.01618
Shannon	0.11	0.08765	0.17395	0.05324	0.02154

Table 1: Lower and upper bounds on the threshold crossover probability of a BSC channel. The various bounds are abbreviated as follows. *BP* is the belief propagation threshold. *ML lower* is a lower bound on the threshold of ML decoding. *Gallager* is Gallager's upper bound on the threshold of ML decoding. *Pairs* is the pair matching bound given by Theorem 4 and Lemma 4. *Clusters* is the bound given by Theorem 5 and Lemma 6. *Shannon* is Shannon's bound on the threshold for the given code rate.

(c,d)	(3,6)	(3,7)	(4,6)	(3,10)	(3,20)
BP	0.881	0.801	1.011	0.677	0.540
ML lower	0.841	0.759	1.285	0.636	0.50
ML upper	0.97170	0.87328	1.29549	0.72403	0.56779
Shannon	0.9787	0.8801	1.2966	0.7300	0.5723

Table 2: Lower and upper bounds on the threshold standard deviation of a BIAWGN channel. The various bounds are abbreviated as follows. *BP* is the belief propagation threshold. *ML lower* is a lower bound on the threshold of ML decoding. *ML upper* is the upper bound given by Theorem 1 on the threshold of ML decoding. *Shannon* is Shannon's bound on the threshold for the given code rate.

Theorem 1 applies to an arbitrary right regular code which is not necessarily left regular. Thus the upper bound on a (c,d)-regular code applies also to any right regular code with right degree d and rate $R = 1 - c/d$.

VI Upper bounds on the minimum distance

Gallager [9] derived a lower bound on the typical minimum distance for an ensemble of LDPC codes. This result was recently generalized to other ensembles in [11]. Tanner [21] described a method for finding lower bounds on the minimum distance of individual LDPC codes. In this section we derive upper bounds on the minimum distance using a modified Hamming bound.

Theorem 6 *Consider a binary code with parity check matrix $A_{L\times N}$ and rate R. Suppose that A has the property that all its rows have a constant weight d. Let the minimum distance of the code be $N\delta$. Then for any integer k, such that $k < N\delta/2$, we have*

$$\frac{1}{N}\log\binom{N}{k} \le (1-R)h\left(\sum_{l=1,3,\ldots;l\le d}\binom{d}{l}\binom{N-d}{k-l}\Big/\binom{N}{k}\right) \tag{28}$$

Furthermore, for N sufficiently large,

$$h\left(\frac{\delta}{2}\right) \le (1-R)h\left(\frac{1}{2}\left[1-(1-\delta)^d\right]\right) \tag{29}$$

Note that the standard Hamming bound which applies to an arbitrary code asserts that

$$h\left(\frac{\delta}{2}\right) \le 1-R$$

Our improvement to the Hamming bound is realized by introducing an additional factor that multiplies the right hand side.

Proof. Let k be some integer such that $k < N\delta/2$. Consider the experiment where we choose at random k different columns of A with uniform probability. Let the resulting syndrome, which is the sum (bitwise xor) of the k columns be denoted by $\mathbf{S}$. We claim that each set of k columns produces a different syndrome. Otherwise two sets of k columns would have produced the same syndrome, hence a set of at most $2k$ columns produces an all zero syndrome, so that there exists a codeword of weight smaller then $N\delta$. The last conclusion contradicts the fact that the minimum distance is $N\delta$. Let the number of syndromes be denoted by N_s and let the entropy of $\mathbf{S}$ be denoted by $H(\mathbf{S})$. Then

$$N_s = \binom{N}{k}$$

In addition, since each syndrome occurs with uniform probability $(1/N_s)$, then

$$H(\mathbf{S}) = \log N_s \tag{30}$$

Let S_l denote the l-th component of the syndrome, $\mathbf{S}$. Then,

$$H(\mathbf{S}) \leq N(1-R)H(S_1) = N(1-R)h\left(P(S_1=1)\right) \tag{31}$$

Now,

$$P(S_1 = 1) = \sum_{l=1,3,\ldots;l\leq d} \binom{d}{l}\binom{N-d}{k-l}/\binom{N}{k} \tag{32}$$

Equation (28) follows from (30), (31) and (32).

Now setting k to the largest integer such that $k < N\delta/2$, and letting $N \to \infty$, we have

$$\begin{aligned}\binom{N-d}{k-l}/\binom{N}{k} &= \frac{(N-k)\ldots(N-k-(d-l-1))}{N\ldots(N-(d-1))}k(k-1)\ldots(k-l+1)\\ &\to \left(\frac{\delta}{2}\right)^l\left(1-\frac{\delta}{2}\right)^{d-l}\end{aligned} \tag{33}$$

Thus,

$$P(S_1=1) \to \sum_{l=1,3,\ldots;l\leq d}\binom{d}{l}\left(\frac{\delta}{2}\right)^l\left(1-\frac{\delta}{2}\right)^{d-l} = \frac{1}{2}\left[1-(1-\delta)^d\right] \tag{34}$$

Equation (29) follows from (30), (31) and (34). $\square$

In Table 3 we compare Theorem 6 to the best currently known asymptotic upper bound on the rate of an arbitrary binary code [16],

$$R(\delta) \leq \min_{0\leq u\leq 1-2\delta}\left\{1+g(u^2)-g(u^2+2\delta u+2\delta)\right\}+o(1), \tag{35}$$

where

$$g(x) = H((1-\sqrt{1-x})/2)$$

(c,d)	(3,6)	(3,7)	(4,6)	(3,10)	(3,20)
LP	0.182	0.151	0.26	0.097	0.041
New	0.205	0.159	0.345	0.091	0.032

Table 3: Upper bounds on the minimum distance. The various bounds are abbreviated as follows. *LP* is the linear programming bound (35). *New* is the bound given by Theorem 6.

To obtain tighter upper bounds on the minimum distance, we may use the same matching techniques as in Section V. We assume that N is sufficiently large. Let the weight of each row of the parity check matrix, A, be d. Suppose that the rows of A are permuted to obtain A' such that there is exactly one index for which both the first row, $\mathbf{v}_1$, and the second, $\mathbf{v}_2$, are 1. In this case, $H(S_1, S_2)$ is given by $2\psi_2(\epsilon, d)$ which is defined by (25) and (3) by setting $\epsilon = \delta/2$. To see that let us denote the joint (overlap) bit by v. Then,

$$P(S_1 = 1, S_2 = 1, v = 1) = \sum_{l_1=0,2,\ldots;l_1 \le d} \sum_{l_2=0,2,\ldots;l_2 \le d} \binom{d-1}{l_1}\binom{d-1}{l_2}\binom{N-2d+1}{k-l_1-l_2-1} \Big/ \binom{N}{k}$$

Now it is easy to verify, as in (33), that if we set k to the largest integer such that $k < N\delta/2$ and then set $N \to \infty$, then

$$\binom{N-2d+1}{k-l_1-l_2-1} \Big/ \binom{N}{k} \to \left(\frac{\delta}{2}\right)^{l_1+l_2+1} \left(1-\frac{\delta}{2}\right)^{2d-l_1-l_2-2}$$

Hence,

$$P(S_1 = 1, S_2 = 1, v = 1) \to \frac{\delta}{2} \cdot \frac{1}{2}\left[1 + (1-\delta)^{d-1}\right] \cdot \frac{1}{2}\left[1 + (1-\delta)^{d-1}\right]$$

which coincides with the corresponding expression in Section V (see (24)) by setting $\epsilon = \delta/2$. Similar analogy holds for $P(S_1 = 0, S_2 = 1)$, $P(S_1 = 1, S_2 = 0)$ and $P(S_1 = 0, S_2 = 0)$. Thus, asymptotically $H(S_1, S_2)$ is indeed given by $2\psi_2(\epsilon, d)$ with $\epsilon = \delta/2$. By the same technique it follows that $H(S_1, \ldots, S_i) = i\psi_i(\delta/2, d)$ where $S_1, \ldots, S_i$ constitute a block in a feasible partition.

Concluding the above discussion, we have the following generalization to Theorem 6.

Theorem 7 *Consider a binary code with parity check matrix $A_{L\times N}$. Suppose that A has the property that all its rows have a constant weight d. Let the minimum distance of the code be $N\delta$. Then for any integer k, such that $k < N\delta/2$, and N sufficiently large,*

$$h\left(\frac{\delta}{2}\right) \le (1-R') \sum_{i=1}^{c} \tau_i \psi_i\left(\frac{\delta}{2}, d\right)$$

for any feasible row partition, $(t_1, \ldots, t_c)$ and $\tau_i = it_i/L$.

VII Conclusion

We derived improved upper bounds on the rate of LDPC codes for which reliable communication is achievable.

Good LDPC codes under iterative decoding are often obtained by imposing a right regular structure. For a given right degree, one then optimizes the left degree profile using some optimization technique [19]. Theorem 1 can be used to lower bound the required right degree that is required for reliable communication at some given rate. Theorem 2 generalizes this bound to the right-irregular case.

The applicability of our results is not limited to LDPC ensembles per se, but also to other cases such as turbo codes which are in fact characterized by a low density parity check matrix [13]. In particular Theorem 2 implies that in order to achieve capacity, the constraint length of the turbo code must approach infinity.

Retracing the proof of Theorem 1, we see that the inequality in (13) follows from the inequality (10). All other transitions hold with equality. In particular, for the BSC (13) holds with equality. Hence for the BSC, a tight bound on $H(\mathbf{S})$ would provide a tight bound on $H(\mathbf{X} \mid \mathbf{Y})$.

In Section V we derived various bounds on $H(\mathbf{S})$. However, these bounds are not tight. To see this, consider a (c, d)-regular code and suppose that the conditions of Lemma 6 hold. We now show how Theorem 5 can be improved by tightening (27). Lemma 6 asserts that almost all rows can be partitioned into blocks of size c, such that all the rows within a block overlap in the same position. We now show that a certain fraction of the blocks can be paired up in the following manner. Two paired blocks, $\mathcal{B}_1$ and $\mathcal{B}_2$, have the property that there exists two rows $\mathbf{v}_1 \in \mathcal{B}_1$ and $\mathbf{v}_2 \in \mathcal{B}_2$, such that $\mathbf{v}_1$ and $\mathbf{v}_2$ overlap. To pair up the blocks, we first remove from the matrix the L/c columns that are all one within a block (i.e., columns that correspond to an overlap position). We then sum up all the rows within a block to obtain an $L/c \times (N - L/c)$ matrix. The weight of each row in this matrix is $c(d-1)$. Hence by Lemma 1, most of its rows can be paired up.

Let $\mathbf{S}_i$ be the vector obtained by taking the c components of $\mathbf{S}$ corresponding to $\mathcal{B}_i$, for $i = 1, 2$. Clearly, $H(\mathbf{S}_1, \mathbf{S}_2) < H(\mathbf{S}_1) + H(\mathbf{S}_2)$. Hence, this pairing up tightens (27). Instead of pairs of blocks, we may also use groups of size c. Moreover, this process can be repeated to an arbitrary nesting of blocks. However, the marginal improvement is expected to converge rapidly. This is due to the fact that the known lower bounds on the threshold approach capacity even for small values of c and d [9, 17].

Acknowledgment

The authors would like to thank the reviewers and the associate editor for their helpful comments.

Appendix

A Proof of Equation 12

Without loss of generality we assume that

$$A = [A_1 \ A_2]$$

where A_1 comprises $(1-R)N$ linearly independent columns and where A_2 comprises the complementary RN columns. Denote $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, where $\mathbf{X}_1$ and $\mathbf{X}_2$ are of lengths $(1-R)N$ and RN respectively. The BSC noise vector, $\mathbf{N} = (\mathbf{N}_1, \mathbf{N}_2)$, satisfies $\mathbf{Z}_i = \mathbf{X}_i + \mathbf{N}_i$ for $i = 1, 2$. Now,

$$H(\mathbf{Z}) = H(\mathbf{Z}_1, \mathbf{Z}_2) = H(\mathbf{Z}_2) + H(\mathbf{Z}_1|\mathbf{Z}_2) \tag{36}$$

$\mathbf{Z}_1$ and $\mathbf{Z}_2$ determine $\mathbf{S}$ (by $\mathbf{S} = A\mathbf{Z}$). Moreover, $\mathbf{Z}_2$ and $\mathbf{S}$ determine $\mathbf{Z}_1$. To see this, we write $A_1\mathbf{Z}_1 = \mathbf{S} + A_2\mathbf{Z}_2$. Since the columns of A_1 are linearly independent, this equation uniquely determines $\mathbf{Z}_1$. Thus

$$H(\mathbf{Z}_1|\mathbf{Z}_2) = H(\mathbf{S}|\mathbf{Z}_2) = H(\mathbf{S}) - I(\mathbf{S};\mathbf{Z}_2) \tag{37}$$

We now show that $I(\mathbf{S};\mathbf{Z}_2) = 0$. To this end we note that $\mathbf{S} = A_1\mathbf{N}_1 + A_2\mathbf{N}_2$, and show a stronger result, namely that $(A_1\mathbf{N}_1, A_2\mathbf{N}_2)$ and $\mathbf{Z}_2$ are statistically independent.

$$\begin{aligned} I(A_1\mathbf{N}_1, A_2\mathbf{N}_2; \mathbf{Z}_2) &= I(A_1\mathbf{N}_1, A_2\mathbf{N}_2; \mathbf{X}_2 + \mathbf{N}_2) \\ &= I(A_1\mathbf{N}_1; \mathbf{X}_2 + \mathbf{N}_2) + I(A_2\mathbf{N}_2; \mathbf{X}_2 + \mathbf{N}_2 | A_1\mathbf{N}_1) \end{aligned} \tag{38}$$

$I(A_1\mathbf{N}_1; \mathbf{X}_2 + \mathbf{N}_2) = 0$ since $\mathbf{N}_1$ and $(\mathbf{X}_2, \mathbf{N}_2)$ are independent. For the same reason, $I(A_2\mathbf{N}_2; \mathbf{X}_2 + \mathbf{N}_2 | A_1\mathbf{N}_1) = I(A_2\mathbf{N}_2; \mathbf{X}_2 + \mathbf{N}_2)$. Now, since $\mathbf{X}_2$ is uniformly distributed over all (2^{RN}) length RN binary vectors, and since $\mathbf{X}_2$ and $\mathbf{N}_2$ are independent, $\mathbf{X}_2 + \mathbf{N}_2$ and $\mathbf{N}_2$ are also independent. Thus $I(A_2\mathbf{N}_2; \mathbf{X}_2 + \mathbf{N}_2) = 0$. We see that both terms in (38) are 0, proving that indeed

$$I(\mathbf{S};\mathbf{Z}_2) = 0 \tag{39}$$

Now, $\mathbf{Z}_2 = \mathbf{X}_2 + \mathbf{N}_2$. Thus, from the discussion above, $\mathbf{Z}_2$ is also uniformly distributed over all length RN binary vectors. Thus

$$H(\mathbf{Z}_2) = RN \tag{40}$$

Equations (36), (37), (39) and (40) yield (12). $\square$

B Proof of Lemma 4

To prove the lemma we use the concept of a secondary graph, introduced in Section IV with the same notations. A statement equivalent to that of Lemma 4 is

$$\lim_{N\to\infty} P\left(\frac{\mu(G')}{L} < \frac{1}{2} - \epsilon\right) = 0 \tag{41}$$

for any $\epsilon > 0$, where G' is a random secondary graph from the ensemble, and $\mu(G')$ is the size of maximal matching in G'. Thus is remains to prove (41).

We now use three lemmas that are proved in the sequel. First note that the degree of any node in any secondary graph in the ensemble cannot exceed $d(c-1)$. Moreover, the following lemma tells us that almost all vertices have exactly this degree.

Lemma 8 *Consider the (c,d)-regular ensemble with N variable nodes and $L = Nc/d$ check nodes. Let G be some bipartite graph from this ensemble. For each check node v in G, denote by $\mathcal{N}(v)$ the set of all vertices in G that are at distance 2 from v. Finally denote by $\mathcal{B}(G)$ the set*

$$\mathcal{B} = \{v \text{ is a check node} : |\mathcal{N}(v)| \neq d(c-1)\}$$

Then for any $\epsilon > 0$,

$$\lim_{N\to\infty} P\left(\frac{|\mathcal{B}|}{N} > \epsilon\right) = 0$$

Next we state the following lemma, which bounds the probability of having many short cycles in a randomly chosen secondary graph.

Lemma 9 *Consider the (c,d)-regular ensemble with N variable nodes and $L = Nc/d$ check nodes. Let S_i be a r.v. representing the number of simple cycles (i.e. cycles that are not union of cycles) of length i in a randomly chosen secondary graph from the ensemble. Then for any $\epsilon > 0$ and n, there exists a $K = K_{n,\epsilon}$ such that $P(\sum_{i=1}^{n} S_i > K) < \epsilon$ for N large enough.*

We now state a third lemma, which will be used to prove (41), together with Lemmas 8 and 9.

Lemma 10 *Let $n \geq 1$ be an integer. Let $G' = (V,E)$ be a graph on L vertices, without parallel edges, such that $L - C$ vertices have degree k, and C vertices have a degree less than k. Let S_i denote the number of simple cycles of length i in G'. Then G' contains a matching M with $|M| \geq \frac{L}{2} - \frac{L}{4n+6} - C - \frac{1}{2}\sum_{i=1}^{n} S_{2i+1}$ edges.*

We are now ready to prove (41). Given $\delta > 0$ and $\epsilon > 0$ choose $n = 1/\epsilon$. From Lemma 9 we know that there exists a K such that $P(\sum_{i=1}^{2n+1} S_i > K) < \delta/2$. From Lemma 8 we have $P(C > L\epsilon/2) < \delta/2$ for N sufficiently large, where C is the number of vertices with degree less than $d(c-1)$ in a random secondary graph. Thus, using a union bound,

$$P\left(\sum_{i=1}^{2n+1} S_i > K \text{ or } C > \frac{L\epsilon}{2}\right) < \delta \tag{42}$$

for N sufficiently large. Hence, choosing $n = 1/\epsilon$, Lemma 10 and (42) imply

$$P\left(\frac{\mu(G')}{L} < \frac{1}{2} - \epsilon\right) < \delta$$

for N sufficiently large, from which (41) follows. $\square$

B.1 Proof of Lemma 8

Denote by X_i an indicator r.v. equal to 0 if $|\mathcal{N}(v_i)| = d(c-1)$, where v_i is the i-th check node, and to 1 otherwise. Then $|\mathcal{B}| = \sum_{i=1}^{L} X_i$. Since the X_i-s are identically distributed we have

$$\mathrm{E}|\mathcal{B}| = L\mathrm{E}X_i = P(X_1 = 1)\frac{Nc}{d} \tag{43}$$

$P(X_1 = 1)$ is the probability that the first check node has less then $d(c-1)$ neighbors at distance 2. We shall now upper bound this probability. To this end, consider performing the socket-matching by which the ensemble is defined in the following order (this does not change the ensemble since each of the $(Nc)!$ matchings is equiprobable). We first match the d sockets of v_1 to some d sockets of the Nc variable sockets. Let us denote the variable nodes to which these sockets belong by $u_1, ..., u_d$. Note that they are not necessarily all distinct. We then match the remaining sockets of u_1 with the respective number of check sockets, then match the sockets of u_2 and so on until we match the sockets of u_d. We denote the check nodes corresponding to the sockets matched to u_i by $w_{i,1}$, $w_{i,2}$, ... Finally, we randomly match the remaining check sockets with the remaining variable sockets. This process is depicted in Figure 3.

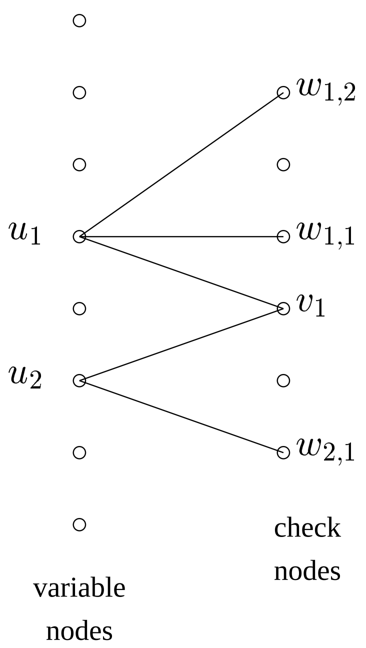

Figure 3: The relation between v_1, u_i and $w_{i,j}$.

Denote by F the event that v_1 has exactly d neighbors at distance 1.

$$P(F) = \frac{cN\, c(N-1)...c(N-d+1)}{cN\,(cN-1)...(cN-d+1)} \geq \left(\frac{c(N-d)}{cN}\right)^d \geq 1 - \frac{d^2}{N} \tag{44}$$

Now suppose that F holds. Denote by $e_i = (u_i, v_1)$, $1 \le i \le d$, the i-th edge emanating from v_1. Similarly, denote by $e_{i,j} = (u_i, w_{i,j})$, $1 \le j \le c-1$, the j-th edge emanating from u_i without re-counting the edge e_i. Denote by $\mathbf{y}$ the following vector of length $d(c-1)$:

$$\mathbf{y} = (w_{1,1}, w_{1,2}, ...w_{1,c-1}, w_{2,1}, ...w_{2,c-1}, ...w_{d,c-1})$$

Finally denote by E_k the event that the first k components of $\mathbf{y}$ are distinct and different from v_1.

Suppose that E_{k-1} and F hold and let us bound the probability that the k-th component of $\mathbf{y}$ is different than the first $k-1$ components and than v_1. This corresponds to the respective edge reaching a check node different than the check nodes reached by the preceding $k-1$ edges and v_1 itself. This edge is chosen by matching a given variable socket with some free check socket. The total number of check sockets is Ld. If E_{k-1} and F hold then exactly k check nodes are present in the vector $(v_1, y_1, y_2, ..., y_{k-1})$. Since these check nodes have a total of kd check sockets, each of the remaining $(L-k)d$ right sockets will be chosen with an equal probability not smaller than $1/Ld$. since $k < cd$ we get:

$$P(E_k | E_{k-1}, F) \ge \frac{(L-cd)d}{Ld} \tag{45}$$

We also have

$$P(X_1 = 1) = P(\overline{E}_{d(c-1)}) \le P(\overline{F}) + P(\overline{E}_{d(c-1)} \mid F) \le P(\overline{F}) + \sum_{k=1}^{d(c-1)} P(\overline{E}_k \mid E_{k-1}, F) \tag{46}$$

From (44), (45) and (46) we have

$$P(X_1 = 1) \le \frac{d^2}{N} + d(c-1)\frac{cd}{L} \le \frac{cd^3}{N} \tag{47}$$

Using (47) and (43) we have $\mathrm{E}|\mathcal{B}| \le c^2 d^2$, from which the claim of the lemma follows, using Markov's inequality. $\Box$

B.2 Proof of Lemma 9

Label the nodes in the secondary graph (or, equivalently, the check nodes in the primary graph) from 1 to L. Denote by ν_i the total number of possible (simple) cycles of length i in any graph in the ensemble. Then,

$$\nu_i \le \frac{L^i}{2i} \le L^i \tag{48}$$

For $1 \le j \le \nu_i$, denote by X_j^i a r.v. equal to 1 if the j-th cycle of length i is present in the graph (for some ordering of the cycles), and 0 otherwise. We wish to upper bound $P(X_j^i = 1)$. To this end, denote the vertices of the j-th cycle by v_1^j,v_2^j,...,v_i^j. To upper bound $P(X_j^i = 1)$ we again describe a specific order in which the random matching of sockets is performed, as was done in the previous proof. We first match the d sockets corresponding to v_1^j, then we proceed to match the d sockets of v_2^j, and so on, until we match the sockets of v_i^j. After this, we randomly match the remaining check sockets with variable sockets.

We now define several events. Denote by E_k^j, $1 \le k \le i$, the event that nodes v_k^j and v_{k+1}^j are connected in the secondary graph by an edge. Here v_{i+1}^j is defined as v_1^j. Further denote $F_k^j = \bigcap_{l=1}^{k-1} E_l^j$. Now, if nodes v_k^j and v_{k+1}^j are connected in the secondary graph, then in the primary graph the check nodes corresponding to these nodes are both connected to at least one common variable node.

Let us bound $P(E_k^j|F_k^j)$ for $1 \le k \le i-2$. This corresponds to the case where at least one of the d sockets of v_{k+1}^j is matched to a *marked socket*, where a marked socket is a free socket of a variable node that has another socket which was matched to some socket of v_k^j. Since v_k^j has d sockets, no more then dc variable sockets are marked. On the other hand, when we match the sockets of v_{k+1}^j, there are $Ld - kc$ free sockets. Thus, for each socket of v_{k+1}^j, the probability that it will be matched to a marked socket is upper bounded by $dc/(Ld-kc)$. Thus by the union bound,

$$P(E_k^j|F_k^j) \le d\frac{dc}{Ld-ic} \le \frac{2dc}{L} \qquad 1 \le k \le i-2 \tag{49}$$

for L sufficiently large. Similarly, when matching the sockets of v_i^j we wish to bound the probability that this matching will connect v_i^j both to v_{i-1}^j and to v_1^j (in the secondary graph). Using arguments analogous to the ones that led to (49) we now have

$$P(E_{i-1}^j, E_i^j|F_{i-1}^j) \le d(d-1)\left(\frac{dc}{Ld-ic}\right)^2 \le \left(\frac{2dc}{L}\right)^2 \tag{50}$$

for L sufficiently large. We are now ready to bound $P(X_j^i = 1)$.

$$P(X_j^i = 1) = P(E_1^j, ..., E_i^j) = \prod_{k=1}^{i-2} P(E_k^j|F_k^j)\cdot P(E_{i-1}^j, E_i^j|F_{i-1}^j) \le \left(\frac{2dc}{L}\right)^i \tag{51}$$

for L sufficiently large, where in the last transition we have used (49) and (50). Thus,

$$\mathrm{E}\sum_{i=1}^n S_i = \mathrm{E}\sum_{i=1}^n\sum_{j=1}^{\nu_i} X_j^i = \sum_{i=1}^n\sum_{j=1}^{\nu_i}\mathrm{E}X_j^i = \sum_{i=1}^n\sum_{j=1}^{\nu_i} P(X_j^i = 1)$$

Using (51) and (48) we thus have

$$\mathrm{E}\sum_{i=1}^n S_i \le \sum_{i=1}^n\sum_{j=1}^{\nu_i}\left(\frac{2dc}{L}\right)^i = \sum_{i=1}^n \nu_i\left(\frac{2dc}{L}\right)^i \le \sum_{i=1}^n L^i\left(\frac{2dc}{L}\right)^i \le (2dc)^{n+1} \tag{52}$$

for L large enough. Finally, given ϵ, choose $K_{n,\epsilon} = (2dc)^{n+1}/\epsilon$. Then (52) and Markov's inequality yield $P(\sum_{i=1}^n S_i > K_{n,\epsilon}) < \epsilon$ for N sufficiently large. $\square$

B.3 Proof of Lemma 10

G' is a graph on L vertices, with all degrees not exceeding k, and with at least $Lk/2 - Ck$ edges. We will estimate from below the size of a maximum matching in G'. For this purpose, we invoke the powerful machinery of Linear Programming, applied to fractional matchings in graphs (see, e.g. [8]).

Recall that a non-negative real-valued function $f : E(G') \to R^+$ is called a *fractional matching* of G' if for every $v \in V(G')$, $\sum_{e \ni v} f(e) \le 1$. The *value* of f is $|f| = \sum_{e \in E(G')} f(e)$. The maximal value of a fractional matching of G' is called the *fractional matching number* of G' and is denoted by $\nu^*(G')$. A fractional matching f with a maximal possible value $|f| = \nu^*(G)$ is an *optimal fractional matching.* Recall that the maximal degree of G' is bounded by k. Assigning the value $f(e) = 1/k$ to each edge of G' produces a fractional matching f of value $|f| = |E(G')|/k \ge \frac{L}{2} - C$. This implies that $\nu^*(G') \ge |f| \ge \frac{L}{2} - C$. According to the results of Balinski and Spielberg [1, 2], there exists an optimal fractional matching f^* of G' with the following properties:

1. $f(e) \in \{0, 0.5, 1\}$ (the so-called half-integrality of an optimal solution);

2. Let $E_1 = \{e \in E(G') : f(e) = 1/2)\}$; $E_2 = \{e \in E(G') : f(e) = 1\}$. Then the edges of E_1 form a collection of vertex disjoint odd cycles, and the edges of E_2 form a matching disjoint from E_1.

Let f^*, E_1, E_2 be as above. For each $i \ge 1$ denote by t_i the number of cycles of length i, formed by E_1. Then

$$\nu^*(G') = |E_2| + \sum_{i \ge 1} t_{2i+1} \frac{2i+1}{2} \; = |E_2| + \sum_{i \ge 1} i t_{2i+1} + \frac{1}{2} \sum_{i \ge 1} t_{2i+1} \; .$$

Recalling the above obtained bound on $\nu^*(G')$, we get:

$$\begin{aligned} |E_2| + \sum_{i \ge 1} i t_{2i+1} \; &= \; \nu^*(G') - \frac{1}{2} \sum_{i \ge 1} t_{2i+1} \\ &\ge \; \frac{L}{2} - C - \frac{1}{2} \sum_{i \ge 1} t_{2i+1} \; . \end{aligned}$$

For each odd cycle in E_1, delete its first, third and so on (i.e. every odd) edge. Clearly, the remaining edges united with E_2 form a matching M in G'. The cardinality of M can be estimated as follows:

$$|M| = |E_2| + \sum_{i \ge 1} i t_{2i+1} \ge \frac{L}{2} - C - \frac{1}{2} \sum_{i \ge 1} t_{2i+1} \; .$$

It remains only to estimate from above the last sum. Clearly, $t_{2i+1} \leq S_{2i+1}$ for all $1 \leq i \leq n$. Also, the number of cycles of length at least $2n+3$ in E_1 does not exceed $L/(2n+3)$. We thus get:

$$\sum_{i\geq 1} t_{2i+1} = \sum_{i=1}^{n} t_{2i+1} + \sum_{i>n} t_{2i+1} \leq \sum_{i=1}^{n} S_{2i+1} + \frac{L}{2n+3} \; .$$

Then it follows that

$$|M| \geq \frac{L}{2} - C - \frac{L}{4n+6} - \frac{1}{2}\sum_{i=1}^{n} S_{2i+1} \; ,$$

as promised. □

C Proof of Lemma 5

We will show that a desired partition can be produced by the following greedy algorithm: start with A, find a column with a maximal number of 1's, take the rows corresponding to 1's in this column to be the next class of the partition, delete them from the matrix, and delete also the corresponding column. The algorithm stops when all rows of A are packed.

Let $(t_1, \ldots, t_c)$ be the vector of the obtained partition. Clearly, the algorithm produces first blocks of size c, then blocks of size $c-1$, etc., ending with blocks of size one. Accordingly, we partition the execution of the algorithm into rounds $c, c-1, \ldots, 1$, where at round i the algorithm puts blocks of size i into the formed partition. Fix i, $1 \leq i \leq c$, and look at the current matrix A_i before the beginning of the i-th round. This matrix has $L - \sum_{j=i+1}^{c} jt_j$ rows and $N - \sum_{j=i+1}^{c} t_j$ columns. Each row of A_i has d ones, while each column of A_i has at most i ones (otherwise we would have formed another block of size larger than i). The total number of ones in A_i is thus $d(L - \sum_{j=i+1}^{c} jt_j)$. On the other hand, counting by the columns, the total number of ones in A_i is at most $i(N - \sum_{j=i+1}^{c} t_j)$. Thus

$$d\left(L - \sum_{j=i+1}^{c} jt_j\right) \leq i\left(N - \sum_{j=i+1}^{c} t_j\right)$$

Rearranging terms and changing i to $i-1$ yields (20). □

D Proof of Lemma 6

Let $\mathcal{A}_{L,c,d}$ denote the set of all 0,1-matrices with L rows, all columns of constant weight $c \geq 3$, and all rows of constant weight d. Let furthermore $\mathcal{A}^0 \subset \mathcal{A}_{L,c,d}$ denote the subset of matrices without two identical columns. Using known results on the number of rectangular matrices with given row and column sums [10, 22] and the assumption $c \geq 3$, one can show that almost all matrices in $\mathcal{A}_{L,c,d}$ have no identical columns. This implies: $|\mathcal{A}^0|/|\mathcal{A}_{L,c,d}| = 1 - o(1)$.

Given a matrix $A \in \mathcal{A}^0$, define a hypergraph $H = H(A)$ as follows. The vertex set of H is L, and for every $1 \leq j \leq Ld/c$, the set of positions, where the j-th column of A has ones, forms an edge of H. Then H is a c-uniform, d-regular hypergraph on L vertices. Moreover, as $A \in \mathcal{A}^0$ has no identical columns, the corresponding hypergraph H has no parallel edges. Denote by $\mathcal{H}_{L,c,d}$ the set of all c-uniform d-regular hypergraphs on L vertices. It is easy to see that the above described mapping $\phi : \mathcal{A}^0 \to \mathcal{H}_{L,c,d}$ is $(Ld/c)!$-to-one, i.e. for every $H \in \mathcal{H}_{L,c,d}$, the set $\phi^{-1}(H)$ contains exactly $(Ld/c)!$ matrices. Then ϕ is measure-preserving. Now, according to a result of Cooper, Frieze, Molloy and Reed [6], if $c < 1 + \log d/((d-1)\log(d/(d-1)))$, then almost all hypergraphs H in $\mathcal{H}_{L,c,d}$ have a perfect matching. Clearly, a perfect matching in H translates immediately to a partition $[L] = \mathcal{S}_1 \cup \ldots \cup \mathcal{S}_{L/c}$, described in the formulation of the theorem. Therefore, almost all matrices $A \in \mathcal{A}^0$ possess a desired partition. As the measure of $\mathcal{A}^0$ in $\mathcal{A}_{L,c,d}$ is $1 - o(1)$, we have proven that almost all matrices in $\mathcal{A}_{L,c,d}$ admit a feasible partition with $t_c = L/c$.

Let us return now to the (c,d)-regular ensemble from the formulation of the lemma. Invoking the above mentioned estimates on the cardinality of $\mathcal{A}_0$ and thus of $\mathcal{A}_{L,c,d}$, we can easily obtain that the probability that a random matrix A from the (c,d)-regular ensemble belongs to $\mathcal{A}_{L,c,d}$ can be asymptotically bounded from below by a function $\alpha(c,d) > 0$, depending on c,d only and thus independent of L. Applying results on the concentration of measure in the symmetric group [14, 15] we derive that almost every matrix A from our ensemble can be obtained from a matrix $A' \in \mathcal{A}_{L,c,d}$ by at most $\sqrt{n}\omega(n)$ transpositions of the permutation creating A', where $n = Nc$ and $\omega(n)$ is any function tending to infinity arbitrarily slowly with n. As every transposition changes the value of t_c in a feasible partition by at most 2, it follows that almost all matrices A from our ensemble admit a feasible partition with $t_c \geq L/c - 2\sqrt{n}\omega(n) = L/c(1 - o(1))$. $\square$

E Proof of Lemma 7

$H(U) = H(V)$ are independent of k, and $H(U,V) = H(U) + H(V) - I(U;V)$. Thus we have to show that $I(U;V)$ is monotonically increasing in k.

$$\begin{aligned} P(U=0,V=0) &= P\left(\sum_{i=1}^{k} X_i = 0\right) P\left(U=V=0 \middle| \sum_{i=1}^{k} X_i = 0\right) \\ &+ P\left(\sum_{i=1}^{k} X_i = 1\right) P\left(U=V=0 \middle| \sum_{i=1}^{k} X_i = 1\right) \end{aligned}$$

$$
\begin{aligned}
&= 2^{-3}\left(1+(1-2p)^k\right)\left(1+(1-2p)^{d-k}\right)^2 \\
&+ 2^{-3}\left(1-(1-2p)^k\right)\left(1-(1-2p)^{d-k}\right)^2 \\
&= 2^{-3}\left(2+4(1-2p)^d+2(1-2p)^{2(d-k)}\right)
\end{aligned}
$$

Thus, if $0 < p < 1/2$, $\alpha_k = P(U=0, V=0)$ is monotonically increasing in k for $0 \le k \le d$. In particular, $\alpha_d = (1+(1-2p)^d)/2$ and $\alpha_0 = \alpha_d^2$. Note also that $P(U=0) = P(V=0) = \alpha_d$. Thus, $P(U=0, V=1) = P(U=1, V=0) = \alpha_d - \alpha_k$ and $P(U=1, V=1) = 1 - 2\alpha_d + \alpha_k$.

Denote by a vector $\mathbf{s}_k = (P(U=0,V=0), P(U=0,V=1), P(U=1,V=0), P(U=1,V=1))$. Then

$$
\mathbf{s}_k = (\alpha_k, \alpha_d - \alpha_k, \alpha_d - \alpha_k, 1 - 2\alpha_d + \alpha_k) = (0, \alpha_d, \alpha_d, 1 - 2\alpha_d) + \alpha_k(1, -1, -1, 1)
$$

Since α_k is increasing in k, for $0 \le k_1 < k_2 \le d$ there exists $0 \le \lambda \le 1$ such that $\alpha_{k_1} = \lambda\alpha_0 + (1-\lambda)\alpha_{k_2}$. Thus, this λ also satisfies: $\mathbf{s}_{k_1} = \lambda\mathbf{s}_0 + (1-\lambda)\mathbf{s}_{k_2}$. Denote $I(U;V)$ by I_k. Then $I_k = D(P(U,V)||P(U)P(V)) = D(\mathbf{s}_k||\mathbf{s}_0)$. We thus have

$$
\begin{aligned}
I_{k_1} &= D(\mathbf{s}_{k_1}||\mathbf{s}_0) = D(\lambda\mathbf{s}_0 + (1-\lambda)\mathbf{s}_{k_2}||\mathbf{s}_0) \\
&\le \lambda D(\mathbf{s}_0||\mathbf{s}_0) + (1-\lambda)D(\mathbf{s}_{k_2}||\mathbf{s}_0) = (1-\lambda)D(\mathbf{s}_{k_2}||\mathbf{s}_0) \le I_{k_2}
\end{aligned}
$$

where the first inequality follows from the convexity property of $D(\cdot||\cdot)$. □

References

[1] M. L. Balinski, "Establishing the matching polytope", *J. Comb. Th. Ser. B*, vol 13, pp. 1–13, 1972.

[2] M. L. Balinski and K. Spielberg, "Methods for integer programming: algebraic, combinatorial and enumerative", In: *Progress in operations research, Vol. III, Relationship between operations research and the computer*, Wiley, New York, pp. 195–292, 1969.

[3] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error correcting coding and decoding: turbo codes", *Proceedings 1993 IEEE International Conference on Communications*, Geneva, Switzerland, pp. 1064–1070, 1993.

[4] D. Burshtein and G. Miller, "Expander graph arguments for message passing algorithms", *IEEE Trans. Inform. Theory*, vol. 47, pp. 782–790, February 2001.

[5] D. Burshtein and G. Miller, "Bounds on the Performance of Belief Propagation Decoding", *IEEE Trans. Inform. Theory*, vol. 48, pp. 112–122, January 2002.

[6] C. Cooper, A. Frieze, M. Molloy and B. Reed, "Perfect matchings in random r-regular, s-uniform hypergraphs", *Combin. Probab. Comput.*, vol. 5, pp. 1–14, 1996.

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

[8] Z. Füredi, "Matchings and covers in hypergraphs", *Graphs and Combinatorics*, vol. 4, pp. 115–206, 1988.

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

[10] I. J. Good and J. F. Crook, "The enumeration of arrays and a generalization related to contingency tables", *Discrete Math.*, vol. 19, pp. 23–45, 1977.

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

[12] M. Luby, M. Mitzenmacher, 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.

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

[14] B. Maurey, "Construction de suites symétriques", *Compt. Rend. Acad. Sci. Paris*, vol. 288, pp. 679–681, 1979.

[15] V. Milman and G. Schechtman, "Asymptotic theory of finite dimensional normed spaces", *Lecture Notes in Math.* 1200, Springer-Verlag, Berlin, 1986.

[16] R. J. McEliece, E. R. Rodemich, H. Rumsey and L. R. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities", *IEEE Trans. Inform. Theory*, vol.23, no. 2, pp.157–166, 1977.

[17] G. Miller and D. Burshtein, "Bounds on the maximum likelihood decoding error probability of low density parity check codes", *IEEE Trans. Inform. Theory*, vol. 47, pp. 2696–2710, November 2001.

[18] 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, February 2001.

[19] T. Richardson, A. Shokrollahi and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes", *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, February 2001.

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

[21] R. M. Tanner, "Minimum distance bounds by graph analysis", *IEEE Trans. Inform. Theory*, vol. 47, pp. 808–821, February 2001.

[22] O. E. O'Neil, "Asymptotics and random matrices with row-sum and column-sum restrictions", *Bull. Amer. Math. Soc.*, vol. 75, pp. 1276–1282, 1969.

[23] V. G. Vizing, "On an estimate of the chromatic class of a p-graph", *Diskret. Analiz.*, vol. 3, pp. 25–30, 1964.