Parity-Check Density versus Performance of Binary Linear Block Codes over Memoryless Symmetric Channels*

Igal Sason Rüdiger Urbanke

EPFL – Swiss Federal Institute of Technology
Lausanne, CH–1015, Switzerland
E-mails: Igal.Sason@epfl.ch, Rudiger.Urbanke@epfl.ch.

August 29, 2003

Abstract

Low-density parity-check (LDPC) codes are efficiently encoded and decoded due to the sparseness of their parity-check matrices. Motivated by their remarkable performance and feasible complexity under iterative message-passing decoding, we derive lower bounds on the density of parity-check matrices of binary linear codes whose transmission takes place over a memoryless binary-input output-symmetric (MBIOS) channel. The bounds are expressed in terms of the gap between the rate of these codes for which reliable communications is achievable and the channel capacity; they are valid for *every* sequence of binary linear block codes. For every MBIOS channel, we construct a sequence of ensembles of regular LDPC codes, so that an upper bound on the asymptotic density of their parity-check matrices scales similarly to the lower bound. The tightness of the lower bound is demonstrated for the binary erasure channel by analyzing a sequence of ensembles of right-regular LDPC codes which was introduced by Shokrollahi, and which is known to achieve the capacity of this channel. Under iterative message-passing decoding, we show that this sequence of ensembles is asymptotically optimal (in a sense to be defined in this paper), strengthening a result of Shokrollahi. Finally, we derive lower bounds on the bit error probability and on the gap to capacity for binary linear block codes which are represented by bipartite graphs, and study their performance limitations over MBIOS channels. The latter bounds provide a quantitative measure for the number of cycles of bipartite graphs which represent good error-correction codes.

Index Terms – Bipartite graph, block codes, channel capacity, cycles in a graph, error probability, iterative decoding, linear codes, low-density parity-check (LDPC) codes, maximum likelihood (ML) decoding, memoryless channels, sparse matrices.

*The paper appears in the *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1611–1635, July 2003. The manuscript was submitted on May 9, 2002, and revised on February 17, 2003. The material in this paper was presented in part at the Fortieth Annual Allerton Conference on Communication, Control and Computing, Allerton House, Monticello, Illinois, USA, October 2–4, 2002. It was also presented in part at the 2003 IEEE International Symposium on Information Theory (ISIT 2003), Yokohama, Japan, June 29–July 4, 2003.
Communicated by Ralf Koetter, Associate Editor for Coding Theory.

1 Introduction

Low-density parity-check (LDPC) codes are well known capacity-approaching linear codes. Due to the sparseness of their parity-check matrices, these codes are efficiently encoded and decoded (see e.g. [14], [20] and [21]). We start our discussion by considering the following question: *How sparse can parity-check matrices of binary linear codes be, as a function of their gap to capacity ?* (where this gap depends in general on the channel and on the decoding algorithm). We derive information theoretic lower bounds on the density of parity-check matrices for binary linear codes which are used over memoryless binary-input output-symmetric (MBIOS) channels, and the bounds are expressed in terms of the gap of the codes to the channel capacity. In order to assess the tightness of the latter bounds, we construct sequences of ensembles of codes so that the asymptotic density of their parity-check matrices behaves similarly to these bounds. We continue our discussion with a derivation of information theoretic lower bounds on the bit error probability and on the gap to capacity of binary linear codes which are represented by bipartite graphs, and study their performance limitations over MBIOS channels. The latter bounds substantiate the statement that good error correction codes should have cycles, and in particular the bounds are exemplified for the binary erasure channel (BEC) and the binary symmetric channel (BSC). The discussion on the lower bounds here applies to *every* binary linear code which is used over an MBIOS channel and maximum-likelihood (ML) decoded (hence, the lower bounds are also valid under any sub-optimal decoding algorithm).

Using standard notation, an ensemble of (n, λ, ρ) LDPC codes is characterized by its length n, and the polynomials $\lambda(x) = \sum_{i=2}^{\infty} \lambda_i x^{i-1}$ and $\rho(x) = \sum_{i=2}^{\infty} \rho_i x^{i-1}$, where λ_i (ρ_i) is equal to the probability that a randomly chosen edge is connected to a variable (parity-check) node of degree i. The variables (parity-check sets) are represented by the left (right) nodes of bipartite graphs which represent LDPC codes. It is now well known (see e.g. [13], [18] and [25]) how to design ensembles of LDPC codes which asymptotically, as the block length tends to infinity, approach the capacity of the BEC within any desired gap. In [25], Shokrollahi proved that the growth rate of the average right degree is at least logarithmic in terms of the gap to capacity. The statement in [25] is a high probability result, and hence it is not necessarily satisfied for every particular code from this ensemble. Further, it assumes a sub-optimal (iterative) decoding algorithm. In [11, 16], Khandekar and McEliece have suggested to study the encoding/ decoding complexity of ensembles of turbo-like codes as a function of their gap to capacity. They conjectured that if the achievable rate under iterative message-passing decoding is a fraction $1 - \varepsilon$ of the channel capacity, then for a wide class of channels, the encoding complexity scales like $\ln \frac{1}{\varepsilon}$ and the decoding complexity scales like $\frac{1}{\varepsilon} \ln \frac{1}{\varepsilon}$. However, there is one exception: for a BEC, the decoding complexity behaves like $\ln \frac{1}{\varepsilon}$ (same as encoding complexity). This is true since for a BEC, the iterative message-passing decoding algorithm can be modified so that each edge is only used *once* (due to the absolute reliability of information which is not erased by the BEC). For a general MBIOS channel however, one has to consider the average number of iterations which are required for successful decoding; under iterative message-passing decoding, this number is conjectured to scale like $\frac{1}{\varepsilon}$.

The inherent gap of binary linear codes to the capacity of a BSC was analyzed in [5, 15], and it was based on the calculation of the composite capacity of a linear block encoder and the BSC. The analysis in [5, 15] requires the knowledge of the coset weight distribution of the linear code whose calculation is in general a hard task [3]. For ensembles of LDPC codes, it is possible (though not easy) to calculate the *average* asymptotic coset weight distribution, but it is currently unknown whether the coset weight distribution concentrates (as the block length tends to infinity). Therefore, the typical gap for these ensembles cannot be derived yet, and even if concentration will be proved in this case, it will only lead to a probabilistic statement which does not necessarily hold for *every* binary linear code from this ensemble.

Consider the number of ones in a parity-check matrix which represents a binary linear code, and normalize it per information bit (i.e., with respect to the dimension of the code). This quantity (which will be later defined as the *density* of the parity-check matrix) is equal to $\frac{1-R}{R}$ times the average right degree of the bipartite graph that represents the code, where R is the rate of the code in bits per channel use. In his thesis [8], Gallager proved that right-regular LDPC codes (i.e., LDPC codes with a constant degree (a_{R}) of the parity-check nodes) cannot achieve the channel capacity on a BSC, even under optimal ML decoding. This inherent gap to capacity is well approximated by an expression which decreases to zero exponentially fast in a_{R}. Richardson et al. [20] have extended this result, and proved that the same conclusion holds if a_{R} designates the *maximal right degree*. It is a simple observation, but has far reaching consequences, that the result still applies if we consider the *average right degree* instead. Gallager's bound [8, pp. 37–38] provides an upper bound on the rate of right-regular LDPC codes which achieve reliable communications over the BSC. Burshtein et al. have recently generalized Gallager's bound for a general MBIOS channel [4], and in this work we rely on their generalization. The bounds which are derived in this work provide an operational meaning to the density of parity-check matrices of binary linear codes used over an MBIOS channel, and relate the density with a lower bound on the gap of the code to the channel capacity under ML decoding (or any other decoding algorithm).

For every MBIOS channel and for *every* sequence of binary linear codes which achieves a fraction $1-\varepsilon$ (where $0 < \varepsilon < 1$) of the channel capacity with vanishing bit error probability, we prove that the asymptotic density of parity-check matrices which represent this sequence of codes is lower bounded by $\frac{K_1 + K_2 \ln \frac{1}{\varepsilon}}{1-\varepsilon}$, where K_1 and K_2 are constants which only depend on the channel. The tightness of this information theoretic lower bound is studied by suitable constructions of sequences of ensembles of LDPC codes, so that the asymptotic density of their parity-check matrices behaves similarly to the lower bound. For a general MBIOS channel, we construct a sequence of ensembles of regular LDPC codes, and show that under ML decoding, an upper bound on the asymptotic density of their parity-check matrices scales like the lower bound above. This indicates that the latter bound reflects the correct behavior of the growth rate of the asymptotic density of parity-check matrices (although there may be room to improve the coefficients of the lower bound). For the BEC, the tightness of the information theoretic lower bound is emphasized by constructing a sequence of ensembles of irregular LDPC codes which achieves this bound (up to a small constant) under iterative message-passing decoding. For the BEC, the optimality (in a weaker sense) of this sequence was proved by Shokrollahi [25] in the context of ensembles and iterative message-passing decoding; we strengthen this result by showing that this sequence is asymptotically optimal in a sense to be defined later. We note here that the requirement of achieving a certain fraction of capacity with vanishing *bit error probability* is the milder requirement (or alternatively, yields the stronger result) with respect to the information theoretic lower bound on the asymptotic density (this is true since vanishing block error probability implies also vanishing bit error probability, so proving a certain information theoretic bound on the asymptotic density under the assumption of vanishing bit error probability makes the bound also valid under the stronger condition of vanishing block error probability). On the other hand, the requirement of achieving the same fraction of capacity with vanishing *block error probability* is stronger than the one with vanishing bit error probability if one wishes to construct a sequence of ensembles which approaches the latter information theoretic lower bound. In each case, we prove our statement with respect to the requirement which yields the stronger result.

If a linear block code of length n can be represented by a factor graph without cycles (where it only includes variable nodes and parity-check nodes, but does not include state nodes), then it is known that ML soft-decision decoding can be achieved in time $O(n^2)$. However, the very poor minimum distance of cycle-free codes (see [7, Theorem 5]) indicates that cycle-free bipartite graphs cannot

support good error correction codes. The bounds in [7] refer to the minimum distance of cycle-free codes,[1] and in this work we derive lower bounds on their bit error probability and on their gap to capacity. The results are easily extended to linear codes whose bipartite graphs have cycles, and in fact, we first present the general results, and derive the results for cycle-free codes as a particular case. We present in this paper information theoretic lower bounds on the bit error probability of a binary linear code used over an MBIOS channel. The bounds are expressed in terms of the density of an arbitrary parity-check matrix of a binary linear code, and they are valid for codes which are represented by bipartite graphs with or without cycles. We introduce a quantitative measure for the cycles in a bipartite graph which represents a binary linear code. The latter bounds provide an information-theoretic interpretation for the tradeoff between the bit error probability or the gap to capacity of an LDPC code (i.e., its performance limitations), and the density of an arbitrary parity-check matrix which represents the code (where the latter affects the decoding complexity per information bit and per iteration, under iterative message-passing decoding). We present here quantitative results which indicate that in order to approach the channel capacity with vanishing bit error probability, LDPC codes should not have too sparse parity-check matrices, as otherwise their inherent gap to capacity becomes large. The latter lower bounds are tighter for a BEC.

The paper is organized as follows: the results are presented in Section 2 and proved in Section 3. Numerical results are exemplified and explained in Section 4. Finally, in Section 5, we present interesting open problems. Throughout this paper, $h(x) = -x\log(x) - (1-x)\log(1-x)$ designates the binary entropy function to the base 2. The rate of a code and the capacity of a channel are expressed in units of *bits per channel use*. Though throughout the paper we use the terms of bit error probability and block error probability for all MBIOS channels, we note that for the case of a BEC, the latter terms have the meaning of bit erasure probability and block erasure probability, respectively.

[1] According to Theorem 5 in [7], the minimum distance of a cycle-free code is upper bounded by $\frac{2}{R}$ (which does not depend on n).

2 Main Results

Definition 2.1. Let $\{\mathcal{C}_m\}$ be a sequence of codes of rate R_m, and assume that for every m, the codewords of the code $\mathcal{C}_m$ are transmitted with equal probability over a channel whose capacity is C. This sequence is said to *achieve a fraction* $1-\varepsilon$ *of the channel capacity with vanishing bit (block) error probability* if $\lim_{m\to\infty} R_m = (1-\varepsilon)C$, and if there exists a decoding algorithm under which the average bit (block) error probability of the code $\mathcal{C}_m$ tends to zero in the limit where $m \to \infty$.

Definition 2.2. Let $\mathcal{C}$ be a binary linear code of rate R and block length n, which is represented by a parity-check matrix H. We define the *density* of H, call it $\Delta = \Delta(H)$, as the normalized number of ones in H *per information bit*. The total number of ones in H is therefore equal to $nR\Delta$.

Theorem 2.1. Let $\{\mathcal{C}_m\}$ be a sequence of binary linear codes achieving a fraction $1-\varepsilon$ of the capacity of an MBIOS channel with vanishing *bit error probability*. Then, the asymptotic density (Δ_m) of their parity-check matrices satisfies

$$\liminf_{m\to\infty} \Delta_m > \frac{K_1 + K_2 \ln\frac{1}{\varepsilon}}{1-\varepsilon} \ , \tag{1}$$

where

$$K_1 = \frac{(1-C)\cdot \ln\left(\frac{1}{2\ln 2}\cdot\frac{1-C}{C}\right)}{2C\cdot\ln\left(\frac{1}{1-2w}\right)} \ , \ K_2 = \frac{1-C}{2C\cdot\ln\left(\frac{1}{1-2w}\right)} \ , \tag{2}$$

and where C is the channel capacity, $w \triangleq \frac{1}{2}\int_{-\infty}^{\infty} \min\left(f(y), f(-y)\right) dy$, and $f(y) \triangleq p(y|x=1)$ designates the conditional *pdf* of the output of the MBIOS channel.[2] For a BEC, the coefficients in (2) can be improved to

$$K_1 = \frac{p\cdot\ln\left(\frac{p}{1-p}\right)}{(1-p)\cdot\ln\left(\frac{1}{1-p}\right)} \ , \ K_2 = \frac{p}{(1-p)\cdot\ln\left(\frac{1}{1-p}\right)} \ , \tag{3}$$

where p designates the probability of erasure.[3]

Note that a-fortiori the same statement holds if we require that the block error probability tends asymptotically to zero.

Theorem 2.2. For any MBIOS channel, there exists a sequence of ensembles of regular LDPC codes which achieves under *ML decoding* a fraction $1-\varepsilon$ of the channel capacity with vanishing *block error probability*, and the asymptotic density of their parity-check matrices satisfies

$$\lim_{n\to\infty} \Delta_n \leq \frac{K_3 + K_4 \ln\frac{1}{\varepsilon}}{1-\varepsilon} \ , \tag{4}$$

where K_3 and K_4 are the following coefficients which only depend on the channel

$$K_3 = \max(\xi_1,\xi_2,\xi_3,\xi_4) + \frac{1}{(1-a)\cdot e\cdot\ln\left(\frac{1}{1-2\delta}\right)} + \frac{2}{C}, \quad K_4 = \frac{1-C}{(1-a)\cdot C\cdot\ln\left(\frac{1}{1-2\delta}\right)}. \tag{5}$$

[2] Under the mild condition that $f(y) > f(-y)$ for $y > 0$, then $w = \Pr(Y<0|X=1) + \frac{1}{2}\Pr(Y=0|X=1)$, where X and Y designate the input and the output, respectively, of an MBIOS channel. This condition is satisfied for e.g. a BEC, a BSC with crossover probability less than $\frac{1}{2}$, a binary-input AWGN channel etc.

[3] The improvement for the BEC doubles the coefficient of the logarithmic growth rate (i.e., K_2) of the lower bound (1) as compared to (2), and it also increases the coefficient K_1 by more than twice (since $w = \frac{p}{2}$ and $C = 1-p$ for a BEC with erasure probability p). This indicates that the coefficients in (2) are not tight in general, but as will be stated in Theorem 2.2, the logarithmic growth rate of the asymptotic density in the lower bound (1) reflects the real behavior for any MBIOS channel.

Here, C designates the channel capacity and

$$\xi_1 = \frac{\ln\left(\frac{2}{\delta(1-2\delta)\ln\left(\frac{1-\delta}{\delta}\right)}\right)}{C\ln\left(\frac{1}{1-2\delta}\right)} \tag{6}$$

$$\xi_2 = \frac{\ln\left(\frac{2}{\delta(1-2\delta)\ln 2}\left(\frac{1}{C\ln 2}\right)^{\frac{1}{1-a}}\right)}{C\ln\left(\frac{1}{1-2\delta}\right)} \tag{7}$$

$$\xi_3 = \frac{\ln\left(\frac{2}{\delta(1-2\delta)}\cdot\frac{1}{\ln\left[1+\ln 2\cdot\left(1-C-h(\delta)\right)^{\frac{1}{1-a}}\right]}\right)}{C\ln\left(\frac{1}{1-2\delta}\right)} \tag{8}$$

$$\xi_4 = \frac{1}{C}\cdot\left(\sqrt{\frac{\ln\left(\frac{2}{\delta(1-2\delta)\ln 2}\right)}{\ln\left(\frac{1}{1-2\delta}\right)}+\left(\frac{1}{2a\ln\left(\frac{1}{1-2\delta}\right)}\right)^2}+\frac{1}{2a\ln\left(\frac{1}{1-2\delta}\right)}\right)^2 \tag{9}$$

$$\delta = \eta\cdot h^{-1}(1-C)\,, \tag{10}$$

where $h^{-1}:[0,1]\to[0,\frac{1}{2}]$ is the inverse of the binary entropy function, and $0<\eta<1$, $0<a<1$ are arbitrary numbers.[4]

Based on Theorems 2.1 and 2.2, we derive bounds on the ratio between the logarithmic growth rates of the upper and lower bounds on the asymptotic density (i.e., $\frac{K_4}{K_2}$ where K_2 and K_4 are given in (2) and (5), respectively).

Corollary 2.1. For any MBIOS channel, there exists a sequence of ensembles of regular LDPC codes so that under *ML decoding*, the minimal value of $\frac{K_4}{K_2}$ (with respect to a and η in Theorem 2.2) satisfies

$$2 \le \frac{K_4}{K_2} \le \frac{2\ln\left(\frac{1}{C}\right)}{\ln\left(\frac{1}{1-2h^{-1}(1-C)}\right)}\,. \tag{11}$$

These lower and upper bounds on $\frac{K_4}{K_2}$ are achieved for the BSC and BEC, respectively.

However, we will see in Theorem 2.3 that for a BEC, the lower bound (1) with the *improved coefficients* in (3) is tight by showing that up to a small additive constant, there is a sequence of ensembles of right-regular LDPC codes which achieves this lower bound under *iterative decoding*.

From a point of view of code construction, the requirement of vanishing *block error probability* is stronger than the same requirement on the *bit error probability*. Theorems 2.1 and 2.2 motivate the following definition.

Definition 2.3. Let $\{\mathcal{C}_m\}$ be a sequence of binary linear codes which achieves for a corresponding sequence of decoders $\{\Pi_m\}$ a fraction $1-\varepsilon$ of the capacity of a BEC with vanishing *block error probability*. The combined sequence $\{(\mathcal{C}_m,\Pi_m)\}$ is said to be *asymptotically optimal* if the codes $\mathcal{C}_m$ can be represented by parity-check matrices whose asymptotic density fulfills the condition

$$\limsup_{m\to\infty}\Delta_m \le \frac{K_1^{'}+K_2\ln\frac{1}{\varepsilon}}{1-\varepsilon}\,,$$

[4]The parameter δ equals a fraction of the normalized Gilbert-Varshamov minimum distance of a code of rate $R=C$. If $\varepsilon\to 0$ (i.e., the gap to capacity tends to zero), then the tightest upper bound (4) is achieved in the limit where $\eta\to 1$ and $a\to 0$, since in the latter case the coefficient of $\ln\frac{1}{\varepsilon}$ (i.e., K_4) is minimized.

where K_2 is the same coefficient as in (3), and $K_1^{'}$ is a constant which does not depend on ε ($K_1^{'}$ may only depend on the channel, and from Theorem 2.1, $K_1^{'} > K_1$).

Consider the sequence of ensembles of $(n,\ \lambda_{\alpha,N},\ \rho_\alpha)$ LDPC codes, introduced in [25]:

$$\lambda_{\alpha,N}(x) = \frac{\alpha \sum_{k=1}^{N-1} \binom{\alpha}{k}(-1)^{k+1}x^k}{\alpha - N\binom{\alpha}{N}(-1)^{N+1}}\ ,\ \rho_\alpha(x) = x^{\frac{1}{\alpha}}\ ,\ \ 0<\alpha<1\ . \tag{12}$$

Based on the proofs in [25], it can be verified that under iterative message-passing decoding and a suitable choice of the parameters α and N, the sequence of ensembles of LDPC codes in (12) achieves a fraction $1-\varepsilon$ of the capacity of a BEC with vanishing bit erasure probability. For this purpose, it is possible to choose the parameters in (12) as follows

$$N \triangleq N_1(\varepsilon, p) = \max\left(\left\lceil \frac{1 - c_1 \cdot (1-p)(1-\varepsilon)}{\varepsilon}\right\rceil, \left\lceil \frac{1}{(1-p)^2}\right\rceil\right)\ ,\ \alpha = \frac{\ln\left(\frac{1}{1-p}\right)}{\ln N}\ , \tag{13}$$

where p designates the erasure probability of the BEC, $c_1 = \frac{1}{4}\cdot e^{1-\frac{\gamma}{2}} \approx 0.5092$, and $\gamma \approx 0.5772$ is Euler's constant. The following statement refines the analysis in the proofs of Proposition 1 and Theorem 2 in [25],[5] and it also demonstrates the tightness of the lower bound (1) for the BEC.

Theorem 2.3. For the sequence of ensembles of LDPC codes in (12) used over a BEC with erasure probability p, let

$$N \triangleq N_2(\varepsilon, p) = \max\left(\left\lceil \frac{1 - c_2(p) \cdot (1-p)(1-\varepsilon)}{\varepsilon}\right\rceil, \left\lceil (1-p)^{-\frac{1}{p}}\right\rceil\right)\ ,\ \alpha = \frac{\ln\left(\frac{1}{1-p}\right)}{\ln N}\ , \tag{14}$$

where $c_2(p) = (1-p)^{\frac{\pi^2}{6}} \cdot e^{(\frac{\pi^2}{6}-\gamma)p}\ \ (0<p<1)$.
This sequence of ensembles achieves a fraction $1-\varepsilon$ of the channel capacity with vanishing *bit error probability* (as the block length tends to infinity) under *iterative message-passing decoding*.[6] The asymptotic density of its parity-check matrices satisfies the inequality[7]

$$\lim_{n\to\infty} \Delta_n \leq \frac{K_1 + K_2 \ln\frac{1}{\varepsilon} + g(\varepsilon, p)}{1-\varepsilon}\ , \tag{15}$$

where K_1, K_2 are the coefficients in (3), and $g(\cdot,\cdot)$ is the positive function

$$g(\varepsilon, p) = \frac{\left(\frac{p}{1-p}+\varepsilon\right)\cdot \ln\left(\frac{\varepsilon \cdot N_2(\varepsilon,p)}{p}\right)}{\ln\left(\frac{1}{1-p}\right)} + \frac{\varepsilon(1-p)}{p}\cdot\left(K_1 + K_2 \ln\frac{1}{\varepsilon}\right)\ . \tag{16}$$

The function $g(\cdot, p)$ is upper bounded by a function which only depends on p, and the increase in (15) (as compared to the lower bound (1)) in the limit where capacity is achieved is

$$\lim_{\varepsilon\to 0^+} g(\varepsilon, p) = \frac{p\cdot \ln\left(\frac{1-c_2(p)\cdot(1-p)}{p}\right)}{(1-p)\cdot\ln\left(\frac{1}{1-p}\right)}\ , \tag{17}$$

[5] The inequality derived in the proof of Theorem 2 in [25] refers to the average right degree (which is the normalized number of ones in the parity-check matrix per *parity* bit). In our terminology, we normalize the number of ones in this matrix per *information* bit, which implies the term $1-\varepsilon$ in the denominators of (1) and (15), since the rate of the codes tends to $1-\varepsilon$ times the channel capacity.

[6] Based on the proof of Shokrollahi in [25], this property can be proved for other settings of parameters, e.g. (13), but the significance of (14) is connected with the properties of the function $g(\cdot,\cdot)$ in the continuation of this theorem.

[7] Inequality (15) becomes an equality when $\varepsilon \to 0$ (i.e. for the asymptotic case where capacity is achieved).

which is *upper bounded for* $0 < p < 1$ *by a constant (whose value is 0.5407).*[8]
There also exists a sequence of ensembles of LDPC codes[9] which under iterative message-passing decoding achieves a fraction $1-\varepsilon$ of the capacity of the BEC with vanishing *block error probability*, so that

$$\lim_{n\to\infty} \Delta_n \leq \frac{K_1 + K_2 \ln \frac{1}{\varepsilon'} + g(\varepsilon', p) + \frac{3}{1-p}}{1-\varepsilon} \,, \tag{18}$$

where $\varepsilon' < \varepsilon$, and ε' can be chosen to be arbitrarily close to ε.

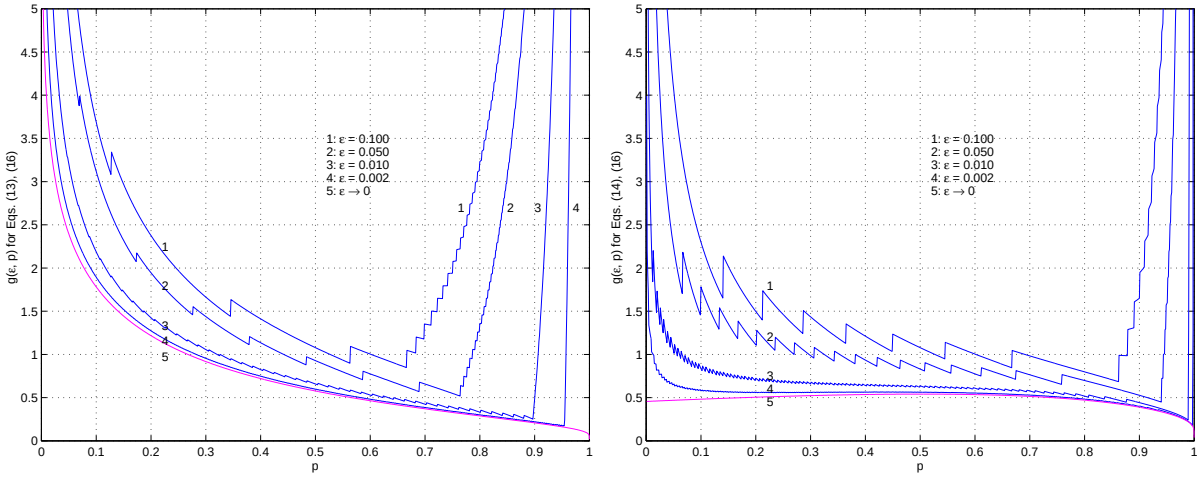

Figure 1: A comparison between the plots of the function $g(\varepsilon, \cdot)$ in (16) for the settings of the parameters in (13) and (14) (which refer to the left-hand side plot and the right-hand side plot, respectively). These plots are compared for the same values of ε, and are depicted as a function of the erasure probability (p) of the BEC (curves 1–4 are discontinuous because of the ceil operations in (13) and (14)).

Corollary 2.2. On the BEC, the sequence of ensembles satisfying inequality (18) under iterative message-passing decoding is asymptotically optimal in the sense of Definition 2.3.

Theorem 2.1 provides a lower bound on the asymptotic density of parity-check matrices for a sequence of codes $\{\mathcal{C}_m\}$ whose average bit error probability tends to zero. The following theorem can be used as a guidance for designing codes of a finite length over an MBIOS channel, where a pre-determined block or bit error probability is desired.

Theorem 2.4. Let $\mathcal{C}$ be a binary linear code of length n and rate R, used over an MBIOS channel. Let C designate the channel capacity, and assume that the codewords of the code $\mathcal{C}$ are transmitted with equal probability. Let ε be a positive number so that $R = (1-\varepsilon)C$, and let P_{B} (P_{b}) be the

[8] Note that for the choice of parameters in (13), the limit $\lim_{\varepsilon\to 0^+} g(\varepsilon, p)$ tends to infinity as $p \to 0$ (as reflected from curve 5 in the left-hand side plot of Fig. 1). This motivated the choice of the two parameters α and N in (14). We note that the small constant above can be further reduced in the limit where $\varepsilon \to 0$: by choosing the parameters in (12) to be $N = \max\left(\left\lceil \frac{1-c_2(p^m)\cdot(1-p)(1-\varepsilon)}{\varepsilon} \right\rceil, \left\lceil (1-p)^{-\frac{1}{p^m}} \right\rceil\right)$ (for an arbitrary positive value of m) and $\alpha = \frac{\ln\left(\frac{1}{1-p}\right)}{\ln N}$ (this selection of parameters yields that the sequence of ensembles (12) achieves a fraction $1-\varepsilon$ of the channel capacity with vanishing bit error probability), it can be verified that for large values of m, the limit $\lim_{\varepsilon\to 0^+} g(\varepsilon, p)$ can be made arbitrarily close to zero for all $0 < p < 1$. However, for large values of m, the curves which refer to positive values of ε become considerably worse as compared to those in the right-hand side plot of Fig. 1 (referring to $m = 1$).

[9] This sequence is specified in the proof of Theorem 2.3 (see Section 3.3.2).

average block (bit) error probability of the code $\mathcal{C}$ under an arbitrary decoding algorithm.[10] Then, the density of *every* parity-check matrix which represents this code satisfies

$$\Delta \geq \frac{1-(1-\varepsilon)C}{2(1-\varepsilon)C} \cdot \frac{\ln\left(\frac{1}{2\ln 2} \cdot \frac{1-C+\varepsilon C}{\delta_1 C+\delta_2}\right)}{\ln\left(\frac{1}{1-2w}\right)} \,, \tag{19}$$

where w is defined in Theorem 2.1.
For the BEC with an erasure probability p, this lower bound can be improved to[11]

$$\Delta \geq \frac{p+\varepsilon(1-p)}{(1-\varepsilon)(1-p)} \cdot \frac{\ln\left(\frac{p+\varepsilon(1-p)}{(1-p)\delta_1+\delta_2}\right)}{\ln\left(\frac{1}{1-p}\right)} \,. \tag{20}$$

In the lower bounds above, for a block error probability ($P_{\rm B}$)

$$\delta_1 = \varepsilon + (1-\varepsilon)P_{\rm B} \,,\; \delta_2 = \frac{h(P_{\rm B})}{n} \,, \tag{21}$$

and for a bit error probability ($P_{\rm b}$)

$$\delta_1 = \varepsilon \,,\; \delta_2 = h(P_{\rm b}) \,. \tag{22}$$

We note that Theorem 2.4 refers to *particular* codes (as opposed to Theorems 2.1, 2.2 and 2.3 which refer to *sequences* of codes or to sequences of ensembles of codes). The reason for this difference is that in the first three theorems we require vanishing bit/block error probability, whereas in Theorem 2.4 we allow a fixed positive decoding error probability. Clearly, Theorem 2.4 can be also stated for sequences of codes, and then it extends Theorem 2.1 to the case where achieving a fixed positive bit error probability is sufficient. We see that if $\varepsilon \ll P_{\rm b}$, then the logarithmic growth rate of the lower bound in Theorem 2.1 is replaced in Theorem 2.4 by a constant which is dominated by $P_{\rm b}$, the parameter w and the channel capacity (where the latter two parameters depend on the channel). This phenomenon is attributed to the fact that in the latter case, there is no need to increase the average right degree without limit (as opposed to the case where the bit error probability should be arbitrarily small for sufficiently large values of the block length n).

The following theorem relies on Theorem 2.4, and presents lower bounds on the bit error probability of binary linear codes which are represented by bipartite graphs with or without cycles. It refers to the performance limitations of these codes over MBIOS channels (in terms of their bit error probability and their gap to capacity). To this end, we define the *normalized density* of a parity-check matrix and then express the following results in terms of the normalized density.

Lemma 2.1. Let $\mathcal{C}$ be a binary linear code of block length n and rate R, and assume that its factor graph is a tree (where we only allow in this graph variable nodes and parity-check nodes). Then the density of the parity-check matrix which represents this cycle-free code is equal to $\Delta = \frac{2-R}{R} - \frac{1}{nR}$, and it is therefore equal to $\frac{2-R}{R}$ in the limit where $n \to \infty$.

Definition 2.4. Let $\mathcal{C}$ be a binary linear code of rate R, which is represented by a parity-check matrix H whose density is Δ. We define the *normalized density* of H, call it $t = t(H)$, to be $t = \frac{R\Delta}{2-R}$. This normalized density is therefore equal to the ratio of $\Delta = \Delta(H)$ and the density of a parity-check matrix which corresponds to a *cycle-free code* of asymptotically infinite block length and of the same rate R.

[10] $P_{\rm B}$ ($P_{\rm b}$) can also denote an upper bound on the block (bit) error probability.

[11] For the BEC, the improvement in the lower bound (20) as compared to the lower bound (19) (where the latter is valid for a general MBIOS channel) is at least by a factor of two.

Theorem 2.5. Let $\mathcal{C}$ be a binary linear block code of rate R which is used over an MBIOS channel whose capacity is C. Assume that the codewords of the code $\mathcal{C}$ are transmitted with equal probability. Let H be a parity-check matrix which represents the code $\mathcal{C}$, and t be its normalized density. Then under any decoding algorithm, the bit error probability (P_{b}) of the code $\mathcal{C}$ satisfies the inequality

$$h(P_{\mathrm{b}}) \geq R - C + \frac{1}{2\ln 2} \cdot (1-R) \cdot (1-2w)^{\frac{2(2-R)t}{1-R}} \ . \tag{23}$$

where w is the same as in Theorem 2.1.
For a BEC with erasure probability p, the lower bound on P_{b} is improved to

$$h(P_{\mathrm{b}}) \geq R - (1-p) + (1-R) \cdot (1-p)^{\frac{(2-R)t}{1-R}} \ . \tag{24}$$

Let $\varepsilon = 1 - \frac{R}{C}$ for $R < C$. Then, the lower bounds (23) and (24) are meaningful (i.e., the right-hand sides of (23) and (24) are positive) for $\varepsilon < \varepsilon_0$ where

$$\varepsilon_0 = 1 - \frac{1}{C} \left(1 - \frac{1}{\frac{1}{1-C} - \left(\frac{1}{2t\ln(1-2w)} \right) \cdot W\left(\frac{t \cdot \ln(1-2w) \cdot (1-2w)^{\frac{2t(2-C)}{1-C}}}{\ln 2 \cdot (1-C)} \right)} \right) \ , \tag{25}$$

and $W(\cdot)$ in (25) designates the Lambert W-function [28].
For a BEC with erasure probability p, the value of ε_0 is improved to

$$\varepsilon_0 = 1 - \frac{1}{1-p} \left(1 - \frac{1}{\frac{1}{p} - \left(\frac{1}{t\ln(1-p)} \right) \cdot W\left(\frac{t \cdot \ln(1-p) \cdot (1-p)^{t\left(1+\frac{1}{p}\right)}}{p} \right)} \right) \ . \tag{26}$$

Corollary 2.3. Let $\{\mathcal{C}_m\}$ be a sequence of binary linear codes, achieving a fraction $1-\varepsilon$ of the capacity of an MBIOS channel with vanishing *bit error probability*. Then $\varepsilon \geq \varepsilon_0$ under ML decoding (or any other decoding algorithm), where ε_0 is introduced in (25) and improved for a BEC in (26).

We note that a direct consequence of Lemma 2.1, Definition 2.4 and Theorem 2.5 yields the following result for cycle-free codes:

Corollary 2.4. By setting $t = 1 - \frac{1}{n(2-R)}$, the results in Theorem 2.5 are valid for cycle-free codes. Similarly, by setting $t = 1$, the conclusion in Corollary 2.3 is valid for a sequence of cycle-free codes with vanishing bit error probability (since the block length tends to infinity in the latter case).

Intuitively, the number of cycles in a bipartite graph is expected to increase with t (i.e., for a bipartite graph which is cycle-free and connected, we obtain from Lemma 2.1 and Definition 2.4 that $t = 1 - \frac{1}{n(2-R)}$. By increasing the value of t above this number, one would expect that the increasing number of edges which connect variable nodes and parity-check nodes in the graph will also increase its number of cycles). For a quantitative measure of this argument, we present here relevant definitions from graph theory (see e.g. [9]).

Definition 2.5. Let G be an arbitrary graph with $|V_G|$ vertices, $|E_G|$ edges and $C(G)$ components.[12] The *cycle rank* of G, denoted by $\beta(G)$, equals to the maximal number of edges which can be removed from the graph without increasing its number of components (so that it remains to be $C(G)$).

[12] If G is a connected graph then $C(G) = 1$.

From Definition 2.5, it is clear that the cycle rank is a measure of the edge redundancy with respect to the connectedness of the graph G. It is easy (see [9, p. 154]) to verify that

$$\beta(G) = |E_G| - |V_G| + C(G). \tag{27}$$

Definition 2.6. Let G be an arbitrary graph. The *full spanning forest* F of the graph G is the remaining part of G after removing the $\beta(G)$ edges in Definition 2.5. Clearly, the number of components of F and G is the same (i.e., $C(F) = C(G)$).

Definition 2.7. Let F be a full spanning forest of a graph G, and let e be any edge in the relative complement of F. The cycle in the subgraph $F + e$ (whose existence and uniqueness are guaranteed by Theorem 3.1.11 in [9]) is called a *fundamental cycle* of G (associated with F).

Definition 2.8. The *fundamental system of cycles* of a graph G which is associated with a full spanning forest F is the set of all fundamental cycles of G associated with F.

In our context, let G be a connected bipartite graph of a linear block code $\mathcal{C}$ of block length n and rate R. Let t be the normalized density of the parity-check matrix which corresponds to G. Since the number of edges in the graph G is $|E_G| = (2-R)tn$, the number of vertices is $|V_G| = (2-R)n$ (i.e., n variable nodes and $(1-R)n$ parity-check nodes), and $C(G) = 1$ (since we assume that G is a connected graph), then from Eq. (27), one calculates the cycle rank of G, which is also equal to the cardinality of the set of the fundamental cycles of G. This leads to the following result

Corollary 2.5. Let $\mathcal{C}$ be a linear block code of length n and rate R which is represented by a parity-check matrix whose normalized density is t. Then the cardinality of the fundamental cycles in the corresponding bipartite graph (i.e., the cycle rank) is given by

$$\beta(G) = (2-R)(t-1)n + 1. \tag{28}$$

We note that for a cycle-free code (so that from Lemma 2.1 and Definition 2.4, $t = 1 - \frac{1}{n(2-R)}$), one obtains from Eq. (28) that $\beta(G) = 0$ (as could be expected). From Eq. (28), the cardinality of the fundamental set of cycles increases linearly with the normalized density t, which agrees with the intuition that the larger is the normalized density of a parity-check matrix, one would expect an increased number of cycles in the corresponding bipartite graph which represents the code. In this context, we also refer the reader to [7, Section 5-B].

3 Proofs of the Theorems in Section 2

3.1 Proof of Theorem 2.1

The first part of this proof refers to a general MBIOS channel, and the second part refines the bound on the asymptotic density for a BEC.
Let $\mathbf{u}_m = (u_1, u_2, \cdots, u_{n_m})$ be a codeword of the code $\mathcal{C}_m$ which is transmitted over an MBIOS channel, and let $\mathbf{v}_m = (v_1, v_2, \cdots, v_{n_m})$ be the received sequence. Let k designate the degree of a specific parity-check node in the representation of the code $\mathcal{C}_m$ by a bipartite graph, let $d_{k,m}$ be the fraction of the parity-check nodes of degree k, and let R_m be the rate of the code $\mathcal{C}_m$. Based on [4, Theorems 1 and 2] (see [4, Eqs. (14) and (15)] for the right-regular case, where the transition to the irregular case is immediate), one obtains that

$$\frac{H(\mathbf{u}_m|\mathbf{v}_m)}{n_m} \geq 1 - C - (1 - R_m) \cdot \sum_k \left\{ d_{k,m} \cdot h\left(\frac{1-(1-2w)^k}{2} \right) \right\}. \tag{29}$$

Since the function $f(x) = h\left(\frac{1-(1-2w)^x}{2}\right)$ (for $x \geq 0$) is concave for every $0 \leq w \leq \frac{1}{2}$,[13] we obtain by Jensen's inequality that

$$\frac{H(\mathbf{u}_m|\mathbf{v}_m)}{n_m} \geq 1 - C - (1 - R_m) \cdot h\left(\frac{1-(1-2w)^{a_{\mathrm{R}}(m)}}{2} \right), \tag{30}$$

where $a_{\mathrm{R}}(m) \triangleq \sum_k k d_{k,m}$ denotes the average right degree.
Since the sequence $\{\mathcal{C}_m\}$ achieves a fraction $1 - \varepsilon$ of the channel capacity with vanishing bit error probability, then according to Definition 2.1, there exists a decoding algorithm (e.g., ML decoding) so that the average bit error probability of the code $\mathcal{C}_m$ tends to zero as m gets large, and $\lim_{m\to\infty} R_m = (1-\varepsilon)C$. Let $P_{\mathrm{b}}^{(i)}(m)$ designate the bit error probability of the digit u_i in the code $\mathcal{C}_m$ (where $1 \leq i \leq n_m$), and $P_{\mathrm{b}}(m) = \frac{\sum_{i=1}^{n_m} P_{\mathrm{b}}^{(i)}(m)}{n_m}$ be the average bit error probability, then

$$\begin{aligned} \tfrac{H(\mathbf{u}_m|\mathbf{v}_m)}{n_m} &\overset{\text{(a)}}{=} \tfrac{\sum_{i=1}^{n_m} H(u_i|\mathbf{v}_m, u_1, \ldots, u_{i-1})}{n_m} \\ &\overset{\text{(b)}}{\leq} \tfrac{\sum_{i=1}^{n_m} H(u_i|\mathbf{v}_m)}{n_m} \\ &\overset{\text{(c)}}{\leq} \tfrac{\sum_{i=1}^{n_m} h\left(P_{\mathrm{b}}^{(i)}(m)\right)}{n_m} \\ &\overset{\text{(d)}}{\leq} h(P_{\mathrm{b}}(m)) \,, \end{aligned} \tag{31}$$

where equality (a) is based on the chain rule for the entropy, inequality (b) is since conditioning reduces the entropy, inequality (c) follows from Fano's inequality and since the code $\mathcal{C}_m$ is binary, and inequality (d) is based on Jensen's inequality. This implies that if the bit error probability of a sequence of binary linear codes $\{\mathcal{C}_m\}$ tends to zero (as $m \to \infty$), then $\lim_{m\to\infty} \frac{H(\mathbf{u}_m|\mathbf{v}_m)}{n_m} = 0$. By letting m tend to infinity, we obtain from Eq. (30) that

$$1 - C - \Big(1 - (1-\varepsilon)C\Big) \cdot h\left(\frac{1-(1-2w)^{a_{\mathrm{R}}(\infty)}}{2} \right) \leq 0, \tag{32}$$

[13] If $0 \leq w < \frac{1}{2}$, the function $f(\cdot)$ is concave since $f'(x) = \frac{(1-2w)^x}{2\ln 2} \cdot \ln\left(\frac{1}{1-2w}\right) \cdot \ln\left(\frac{1+(1-2w)^x}{1-(1-2w)^x}\right)$, is a monotonically decreasing function on the interval $[0, \infty)$, and therefore $f''(x) \leq 0$ for $x \geq 0$. If $w = \frac{1}{2}$, $f(\cdot)$ is constant.

where $a_{\mathrm{R}}(\infty) \triangleq \liminf_{m\to\infty} a_{\mathrm{R}}(m)$. For the continuation of the proof, we prove the following lemma.

Lemma 3.1. $h(x) \leq 1 - \frac{2}{\ln 2} \cdot (\frac{1}{2} - x)^2$ for $0 \leq x \leq \frac{1}{2}$.

Proof. Define the function $m(x) = h(x) - \left[1 - \frac{2}{\ln 2} \cdot \left(\frac{1}{2} - x\right)^2\right]$ for $0 \leq x \leq \frac{1}{2}$. The first three derivatives of $m(\cdot)$ are

$$m'(x) = \frac{\ln\left(\frac{1-x}{x}\right) - 2(1-2x)}{\ln 2}, \quad m''(x) = \frac{4 - \left(\frac{1}{x} + \frac{1}{1-x}\right)}{\ln 2}, \quad m^{(3)}(x) = \frac{\frac{1}{x^2} - \frac{1}{(1-x)^2}}{\ln 2}\,.$$

The third derivative of $m(\cdot)$ is positive on the interval $(0, \frac{1}{2})$ and vanishes at $x = \frac{1}{2}$, and therefore the second derivative of $m(\cdot)$ is monotonically increasing on the interval $(0, \frac{1}{2}]$. Since $m''(\frac{1}{2}) = 0$, then $m''(x) \leq 0$ for $0 < x \leq \frac{1}{2}$ (with equality if and only if $x = \frac{1}{2}$), which yields that the first derivative of $m(\cdot)$ is a monotonically decreasing function on the interval $(0, \frac{1}{2}]$. Since the derivative of $m(\cdot)$ also vanishes at $x = \frac{1}{2}$, then it yields that $m'(x) > 0$ for $0 < x < \frac{1}{2}$, and therefore $m(\cdot)$ is a monotonically increasing function on the interval $(0, \frac{1}{2}]$. Finally, since $m(\frac{1}{2}) = 0$, then it yields that $m(x) \leq 0$ for $0 \leq x \leq \frac{1}{2}$. □

Based on Lemma 3.1, it follows that $h\left(\frac{1-(1-2w)^{a_{\mathrm{R}}(\infty)}}{2}\right) \leq 1 - \frac{1}{2\ln 2} \cdot (1-2w)^{2a_{\mathrm{R}}(\infty)}$ for $0 \leq w \leq \frac{1}{2}$. The substitution of this upper bound on $h(\cdot)$ in the left-hand side of Eq. (32) yields that

$$a_{\mathrm{R}}(\infty) \geq \frac{\ln\left(\frac{1}{2\ln 2} \cdot \left(1 + \frac{1-C}{\varepsilon C}\right)\right)}{2\ln\left(\frac{1}{1-2w}\right)}.$$

Since $a_{\mathrm{R}}(m)$ and Δ_m designate the normalized number of ones in a parity-check matrix which represents the binary linear code $\mathcal{C}_m$ (where the normalization is per *parity* bit or per *information* bit, respectively), then $\Delta_m = \left(\frac{1-R_m}{R_m}\right) a_{\mathrm{R}}(m)$, and

$$\begin{aligned}
\liminf_{m\to\infty} \Delta_m &= \left(\frac{1-(1-\varepsilon)C}{(1-\varepsilon)C}\right) a_{\mathrm{R}}(\infty) \\
&\geq \frac{(1-C)\cdot a_{\mathrm{R}}(\infty)}{(1-\varepsilon)C} \\
&> \frac{1-C}{(1-\varepsilon)C} \cdot \frac{\ln\left(\frac{1}{2\ln 2} \cdot \frac{1-C}{\varepsilon C}\right)}{2\ln\left(\frac{1}{1-2w}\right)}\,,
\end{aligned}$$

which yields the lower bound (1) with the coefficients K_1, K_2 in (2).

For the BEC, we will derive a lower bound on $\frac{H(\mathbf{u}_m|\mathbf{v}_m)}{n_m}$ in a different way. For the sake of notational simplicity, we will replace $\mathbf{u}_m$ and $\mathbf{v}_m$ by $\mathbf{U}$ and $\mathbf{V}$, respectively. In the following derivation, let $\mathbf{K}$ and $\mathbf{E}$ designate the random vectors which indicate the positions of the known and erased digits in the received vector ($\mathbf{V}$), respectively (note that knowing one of these two random vectors implies the knowledge of the other vector). The random vector $\mathbf{V_K}$ denotes the sub-vector of $\mathbf{V}$ with the known digits of the received vector (i.e., those digits which are not erased by the BEC). Note that there is a *one-to-one* correspondence between the received vector $\mathbf{V}$ and the pair of vectors $(\mathbf{V_K}, \mathbf{E})$. We designate by $\mathbf{U_E}$ and $\mathbf{U_K}$ the sub-vectors of the transmitted codeword $\mathbf{U}$, such that they correspond to digits of $\mathbf{U}$ in the erased and known positions of the received vector, respectively (so that $\mathbf{U_K} = \mathbf{V_K}$). Finally, let $H_{\mathbf{E}}$ denote the matrix of those columns of H (a

parity-check matrix representing the code $\mathcal{C}_m$) whose variables are indexed by $\mathbf{E}$, and $|\mathbf{e}|$ denotes the number of elements of a vector $\mathbf{e}$. Then

$$\begin{aligned} H(\mathbf{U}|\mathbf{V}) &= H(\mathbf{U}|\mathbf{V_K}, \mathbf{E}) \\ &= H(\mathbf{U_E}, \mathbf{U_K}|\mathbf{V_K}, \mathbf{E}) \\ &= H(\mathbf{U_E}|\mathbf{V_K}, \mathbf{E}) \\ &= \textstyle\sum_{\mathbf{v_k}, \mathbf{e}} p(\mathbf{v_k}, \mathbf{e}) \cdot H(\mathbf{U_E}|\mathbf{V_K} = \mathbf{v_k}, \mathbf{E} = \mathbf{e}) \\ &= \textstyle\sum_{\mathbf{v_k}, \mathbf{e}} p(\mathbf{v_k}, \mathbf{e}) \cdot (|\mathbf{e}| - \mathrm{rank}(H_{\mathbf{e}})) \\ &= \textstyle\sum_{\mathbf{e}} p(\mathbf{e}) \cdot (|\mathbf{e}| - \mathrm{rank}(H_{\mathbf{e}})) \\ &= n_m p - \textstyle\sum_{\mathbf{e}} p(\mathbf{e}) \cdot \mathrm{rank}(H_{\mathbf{e}}) \ . \end{aligned} \tag{33}$$

Note that the rank of $H_{\mathbf{e}}$ is upper bounded by the number of non-zero rows of $H_{\mathbf{e}}$ which is equal to the number of parity-check nodes which involve erased bits (the summation $\sum_{\mathbf{e}} p(\mathbf{e}) \cdot \mathrm{rank}(H_{\mathbf{e}})$ is therefore upper bounded by the average number of parity-check sets which involve erased bits). If a parity-check node is of degree k, then the probability that it involves at least one erased bit is equal to $1-(1-p)^k$. The average number of the parity-check nodes which therefore involve at least one erased bit is equal to $n_m(1-R_m)\sum_k d_{k,m}\left(1-(1-p)^k\right)$ (where as before, $d_{k,m}$ designates the fraction of parity-check nodes of degree k), and therefore

$$\sum_{\mathbf{e}} p(\mathbf{e}) \cdot \mathrm{rank}(H_{\mathbf{e}}) \leq n_m(1-R_m)\left(1 - \sum_k d_{k,m}(1-p)^k\right). \tag{34}$$

From Jensen's inequality, it follows that

$$\sum_k d_{k,m}(1-p)^k \geq (1-p)^{a_{\mathrm{R}}(m)} \tag{35}$$

where $a_{\mathrm{R}}(m) \triangleq \sum_k k d_{k,m}$ is the average right degree. Eqs. (33), (34) and (35) yield that

$$\frac{H(\mathbf{u}_m|\mathbf{v}_m)}{n_m} \geq p - (1-R_m) \cdot \left(1 - (1-p)^{a_{\mathrm{R}}(m)}\right). \tag{36}$$

If $\{\mathcal{C}_m\}$ is a sequence of codes which achieves a fraction $1-\varepsilon$ of the channel capacity with vanishing bit error (erasure) probability (i.e., $\lim_{m\to\infty} R_m = (1-\varepsilon)(1-p)$ and the bit error probability of the sequence of codes $\{\mathcal{C}_m\}$ tends asymptotically to zero), then from Eq. (31) $\lim_{m\to\infty} \frac{H(\mathbf{u}_m|\mathbf{v}_m)}{n_m} = 0$. From Eq. (36), this implies that $a_{\mathrm{R}}(\infty) \triangleq \liminf_{m\to\infty} a_{\mathrm{R}}(m) \geq \frac{\ln\left(1+\frac{p}{\varepsilon(1-p)}\right)}{\ln\left(\frac{1}{1-p}\right)} > \frac{\ln\left(\frac{p}{\varepsilon(1-p)}\right)}{\ln\left(\frac{1}{1-p}\right)}$. Since $\Delta_m = \left(\frac{1-R_m}{R_m}\right) \cdot a_{\mathrm{R}}(m)$, it follows that $\liminf_{m\to\infty} \Delta_m \geq \left(\frac{p}{(1-p)(1-\varepsilon)}\right) a_{\mathrm{R}}(\infty)$, which yields Eq. (1) with the improved coefficients K_1 and K_2 in Eq. (3).

- **A consequence of the proof of Theorem 2.1**

  Based on the proof of Theorem 2.1, we prove and discuss an upper bound on the asymptotic rate of every sequence of binary linear codes for which reliable communication is achievable. The bound refers to optimal ML decoding, and is therefore valid for any sub-optimal decoding algorithm. Hence, the following result also provides an upper bound on the achievable rate of ensembles of LDPC codes under iterative decoding, where the transmission takes place over an MBIOS channel.

  Corollary 3.1. Let $\{\mathcal{C}_m\}$ be a sequence of binary linear codes whose codewords are transmitted with equal probability over an MBIOS channel. Let $d_{k,m}$ be the fraction of the parity-check nodes of degree k in a representation of the code $\mathcal{C}_m$ by a bipartite graph, and define w as in Theorem 2.1.[14]

[14] For a sequence of ensembles of binary linear codes $\{\mathcal{C}_m\}$, we denote by $d_{k,m}$ the probability of picking (with a uniform distribution) a parity-check node of degree k from a bipartite graph which represents a code in the ensemble $\mathcal{C}_m$.

Then, a necessary condition on the achievable rate (R) for reliable communication is

$$R \leq 1 - \max\left\{ \frac{1-C}{\sum_k d_{k,m}\, h\left(\frac{1-(1-2w)^k}{2}\right)} \,,\; \frac{2w}{1-\sum_k d_{k,m}\,(1-2w)^k} \right\} \tag{37}$$

in the limit where $m \to \infty$. The necessary condition in Eq. (37) can be loosened to

$$R \leq 1 - \max\left\{ \frac{1-C}{h\left(\frac{1-(1-2w)^{a_{\mathrm{R}}(m)}}{2}\right)} \,,\; \frac{2w}{1-(1-2w)^{a_{\mathrm{R}}(m)}} \right\} \,, \tag{38}$$

where $a_{\mathrm{R}}(m) \triangleq \sum_k k d_{k,m}$ denotes the average right degree of the bipartite graph of the code $\mathcal{C}_m$. Otherwise, under any decoding algorithm, the average *bit error probability* of the codes in the sequence $\{\mathcal{C}_m\}$ is bounded away from zero by a constant which is independent of m.
For a BSC with crossover probability p (where $p < \frac{1}{2}$) or a BEC with erasure probability p, Eq. (38) is equivalent to

$$R \leq 1 - \frac{h(p)}{h\left(\frac{1-(1-2p)^{a_{\mathrm{R}}(m)}}{2}\right)} \quad \text{or} \quad R \leq 1 - \frac{p}{1-(1-p)^{a_{\mathrm{R}}(m)}} \,, \tag{39}$$

respectively.

Proof. From Eqs. (29) and (31), one obtains that

$$h(P_{\mathrm{b}}(m)) \geq 1 - C - (1-R_m) \cdot \sum_k d_{k,m}\, h\left(\frac{1-(1-2w)^k}{2}\right). \tag{40}$$

Under the assumption of vanishing bit error probability (i.e., $\lim_{m\to\infty} P_{\mathrm{b}}(m) = 0$), it follows from Eq. (40) that

$$R \leq 1 - \frac{1-C}{\sum_k d_{k,m}\, h\left(\frac{1-(1-2w)^k}{2}\right)} \tag{41}$$

where $R \triangleq \lim_{m\to\infty} R_m$ is the asymptotic rate of the sequence.
Based on the erasure decomposition lemma [20, Appendix B], an arbitrary MBIOS channel is physically degraded with respect to a BEC whose erasure probability is $p = 2w$ (according to the notation in [20], the equality $w = P_{\mathrm{e}}(f)$ holds with the definitions of $f(\cdot)$ and w in Theorem 2.1 here). Let $\mathbf{u}_m$ designate a transmitted codeword in the code $\mathcal{C}_m$, and let $\mathbf{v}_m$ and $\mathbf{z}_m$ designate the received sequence at the output of the BEC above and at the output of the considered MBIOS channel, respectively. Based on this notation, one obtains that $I(\mathbf{u}_m;\mathbf{v}_m) \geq I(\mathbf{u}_m;\mathbf{z}_m)$ where we rely here on the erasure decomposition lemma and the data processing theorem. The latter inequality implies that

$$H(\mathbf{u}_m|\mathbf{v}_m) \leq H(\mathbf{u}_m|\mathbf{z}_m). \tag{42}$$

Based on Eq. (31)

$$\frac{H(\mathbf{u}_m|\mathbf{z}_m)}{n_m} \leq h(P_{\mathrm{b}}(m)) \tag{43}$$

where $P_{\mathrm{b}}(m)$ is the bit error probability of the code $\mathcal{C}_m$ at the output of the (original) MBIOS channel, and n_m is the length of the code $\mathcal{C}_m$. Moreover, based on Eqs. (33) and (34)

$$\frac{H(\mathbf{u}_m|\mathbf{v}_m)}{n_m} \geq 2w - (1-R_m) \cdot \left(1 - \sum_k d_{k,m}\,(1-2w)^k\right) \tag{44}$$

where $p = 2w$ is the erasure probability of the BEC in the erasure decomposition lemma. From Eqs. (42), (43) and (44), it follows that

$$h(P_{\mathrm{b}}(m)) \geq 2w - (1 - R_m) \cdot \left(1 - \sum_k d_{k,m} \, (1 - 2w)^k \right) . \tag{45}$$

The necessary condition for reliable communication with vanishing bit error probability which then follows from Eq. (45) is

$$R \leq 1 - \frac{2w}{1 - \sum_k d_{k,m} \, (1 - 2w)^k}. \tag{46}$$

Finally, Eq. (37) follows immediately from the necessary conditions which are imposed in Eqs. (41) and (46). Jensen's inequality and the explanation in footnote 13 justify that the necessary condition in Eq. (38) is loosened as compared to that one in Eq. (37).
The transition from Eq. (38) to Eq. (39) is based on the equalities $w = \min(p, 1-p)$ and $C = 1-h(p)$ for a BSC with crossover probability p. For a BEC with erasure probability p, the equalities $w = \frac{p}{2}$ and $C = 1 - p$ hold, and we also rely on the inequality $h\left(\frac{1-x}{2}\right) > 1 - x$ for $0 < x < 1$.

□

Corollary 3.1 provides a generalization of the statement in [8, pp. 37–38] which was proved by Gallager for right-regular LDPC codes used on a BSC, and was extended by Richardson et al. [20] for the case where $a_{\mathrm{R}}(m)$ designates the *maximal right degree*. Corollary 3.1 asserts that this conclusion is also valid with respect to the *average right degree*, and not only that the block error probability is bounded away from zero by a constant which does not depend on the block length n (as was stated in [8] and [20]), but also the *bit error probability* has the same property. Eq. (37) in Corollary 3.1 suggests an improved upper bound on the rates for which reliable communication is achieved, where the improvement is with respect to Eq. (16) in [4] (this improvement is pronounced for a BEC and is not useful for a BSC, see Eq. (39)). We note that under *iterative message-passing decoding*, refined bounds on the achievable rates of LDPC codes used over a BEC were derived in [1] (for a BEC, the bound in Corollary 3.1 coincides with the bound of Shokrollahi [25] and with the zero-order bound in [1]). However, the bound in Corollary 3.1 differs from the bound of Shokrollahi and the refined bounds in [1] in the sense that the latter bounds apply to the sub-optimal iterative message-passing decoding algorithm, and they are high probability results which rely on the density evolution over a BEC. We also note that in the limit where the sequence of codes achieves the capacity of a BEC (which then yields that their average right degrees tend to infinity), the refined bounds on the achievable rates in [1] and the bound presented in Corollary 3.1 asymptotically coincide, but the latter bound is stronger in the sense that it applies to ML decoding (and not only to sub-optimal iterative decoding), and since it also applies to every sequence of binary linear codes.

3.2 Proof of Theorem 2.2

The weight distribution of linear block codes plays a crucial role in their performance analysis under ML decoding (see [24] and references therein). Gallager has derived an upper bound on the average weight distribution of an ensemble of regular LDPC codes; the bound provides the correct behavior of the exponential growth rate of their average weight distribution [8, pp. 14–16]. For a given value of the code rate and an increasing right degree, the average weight distribution of Gallager's ensemble of regular LDPC codes approaches the binomial distribution (see [14], where the latter distribution characterizes the average weight distribution of fully random block codes). We provide here a quantitative measure of this observation (see Proposition 3.1), and apply it to the

performance analysis of Gallager's ensemble under ML decoding. For an arbitrary MBIOS channel, an upper bound on the decoding error probability which combines the Shulman and Feder bound [27] with the union bound provides for two ensembles of LDPC codes the same asymptotic behavior as the lower bound (see [17]). Following Miller and Burshtein [17], and based on tight bounds on the exponential growth rate of the average weight distribution of Gallager's ensemble (which are derived in subsection 3.2.1), we determine the parameters of a sequence of Gallager's ensembles so that it achieves a fraction $1-\varepsilon$ of the channel capacity of an arbitrary MBIOS channel. Finally, we verify that the asymptotic density of the parity-check matrices which represent this sequence satisfies inequality (4), and therefore their asymptotic density behaves similarly to the information theoretic lower bound (1). We note that for fully random block codes, the Shulman and Feder bound [27] coincides with the random coding bound of Gallager, and therefore the former bound achieves the channel capacity of an arbitrary MBIOS channel.[15]

3.2.1 Derivation of simple and tight bounds on the exponential growth rate of the weight distribution

Consider Gallager's ensemble of (n, j, k) LDPC codes (where j and k designate the number of ones in every column or row, respectively, of a parity-check matrix which represents a code from this ensemble) [8]. The rate R of this ensemble is at least $1-\frac{j}{k}$, and it follows from the analysis in [8] that the asymptotic exponential growth rate of the average weight distribution of this ensemble is

$$r(\delta) = j\left(\frac{\mu_k(s)}{k} - s\delta + \left(1-\frac{1}{k}\right)\ln 2\right) - (j-1)\, h_e(\delta), \tag{47}$$

where δ ($0 \leq \delta \leq 1$) is the normalized Hamming weight of the codewords with respect to their block length n, $r(\cdot)$ is the exponential growth rate of the average weight distribution (i.e., the average number of codewords of Hamming weight $l = n\delta$ is $\overline{N(l)} \doteq \exp(n\ r(\delta))$), $h_e(\cdot)$ designates the binary entropy function to the natural base, and

$$\mu_k(s) = \ln\left[\left(\frac{1+e^s}{2}\right)^k + \left(\frac{1-e^s}{2}\right)^k\right]. \tag{48}$$

The parameter s ($-\infty < s < \infty$) in (47) is related to δ so that

$$\mu_k^{'}(s) = k\delta \tag{49}$$

where $\mu_k^{'}(\cdot)$ designates the derivative of $\mu_k(\cdot)$ [8]. In order to proceed in our analysis, we will derive in this subsection simple bounds on the exponential growth rate of the average weight distribution $r(\cdot)$ which become very tight for large values of k. By the substitution $u = \frac{1-e^s}{1+e^s}$ (or equivalently $s = \ln\left(\frac{1-u}{1+u}\right)$ where $-1 < u < 1$), Eq. (49) is converted to the polynomial equation

$$(1-2\delta)u^k - u^{k-1} - u + (1-2\delta) = 0. \tag{50}$$

We provide here three lemmas which we will rely on later in this subsection.

[15]The interested reader is referred to Appendix A in [24] which provides a generalization of the Shulman and Feder bound [27], and considers its error exponent. We note that the latter bound is a particular case of the DS2 bound (a generalization of the second version of Duman and Salehi bounds) [24]. However, as will be clarified in the continuation of this proof, the utilization of a combination of the Shulman and Feder bound and the union bound provides a sufficiently tight bound for our purpose.

Lemma 3.2. Let $k \geq 2$ be an integer and $\delta \in (0, \frac{1}{2})$. Then there exists a unique root $u^* = u^*(\delta)$ of the polynomial equation (50) which is in the interval $(0, 1)$, and it satisfies the inequality

$$(1-2\delta)\,\big[1-2(1-2\delta)^{k-2}\big] < u^* < 1-2\delta. \tag{51}$$

The transformation $s^* = \ln\left(\frac{1-u^*}{1+u^*}\right)$ yields that

$$\ln\left(\frac{\delta}{1-\delta}\right) < s^* < \ln\left(\frac{\delta}{1-\delta}\right) + \frac{2(1-2\delta)^{k-1}}{\delta}\,. \tag{52}$$

Proof. Define the function $f(u) = (1-2\delta)u^k - u^{k-1} - u + (1-2\delta)$ for $u \in [0,1]$.
If $0 < \delta < \frac{1}{2}$ then $f(0) = 1-2\delta > 0$ and $f(1) = -4\delta < 0$, so there exists a root of the polynomial equation (50) inside the interval $(0,1)$. The uniqueness of this root is proved by showing that the function $f(\cdot)$ is monotonically decreasing in the interval $[0,1]$: the derivative of $f(\cdot)$ is $f'(u) = ku^{k-2}\,\big[(1-2\delta)u - 1\big] + u^{k-2} - 1$. For $u \in [0,1]$ and $k \geq 2$, $u^{k-2} - 1 \leq 0$, and also for $0 < \delta < \frac{1}{2}$ and $u \in [0,1]$, $ku^{k-2}\,\big[(1-2\delta)u - 1\big] \leq 0$. This yields that $f'(u) < 0$ for $u \in (0,1)$, which therefore implies the uniqueness of a root $u^* = u^*(\delta)$ of equation (50) inside the interval $(0,1)$.
The polynomial equation (50) is equivalent to the equation $[(1-2\delta)u-1]u^{k-1} = u-(1-2\delta)$. Since $\big[(1-2\delta)u-1\big]\,u^{k-1} < 0$ for $\delta \in (0, \frac{1}{2})$ and $u \in (0,1)$ (and in particular for $u = u^*$), then $u^* < 1-2\delta$ (thus proving the upper bound on u^* in (51)). To derive the lower bound on u^*, lets assume that $u^* = 1-2\delta-\varepsilon_0$ is the unique solution of (50) in the interval $(0, 1-2\delta)$ (so that $0 < \varepsilon_0 < 1-2\delta$). The substitution of u^* in (50) gives

$$\big[4\delta(1-\delta) + \varepsilon_0(1-2\delta)\big]\,(1-2\delta)^{k-1}\left(1 - \frac{\varepsilon_0}{1-2\delta}\right)^{k-1} = \varepsilon_0\,. \tag{53}$$

Since $0 < \left(1 - \frac{\varepsilon_0}{1-2\delta}\right)^{k-1} < 1$ and $0 < 4\delta(1-\delta) + \varepsilon_0(1-2\delta) < 2$, then Eq. (53) implies that $\varepsilon_0 < 2(1-2\delta)^{k-1}$, which yields the lower bound on u^* in (51). The function $s = \ln\left(\frac{1-u}{1+u}\right)$ is monotonically decreasing in the interval $(-1,1)$, so based on the upper bound on u^* in (51), $s^* > \ln\left(\frac{1-(1-2\delta)}{1+(1-2\delta)}\right) = \ln\left(\frac{\delta}{1-\delta}\right)$. Since $k \geq 2$ and $0 < \delta < \frac{1}{2}$, then from the lower bound on u^* in Eq. (51)

$$\begin{aligned}
s^* &= \ln\left(\tfrac{1-u^*}{1+u^*}\right) \\
&< \ln\left(\tfrac{1-(1-2\delta)(1-2(1-2\delta)^{k-2})}{1+(1-2\delta)(1-2(1-2\delta)^{k-2})}\right) \\
&= \ln\left(\tfrac{\delta+(1-2\delta)^{k-1}}{1-\delta-(1-2\delta)^{k-1}}\right) \\
&= \ln\left(\tfrac{\delta}{1-\delta}\right) + \ln\left(1+\tfrac{(1-2\delta)^{k-1}}{\delta}\right) + \ln\left(1+\tfrac{\frac{(1-2\delta)^{k-1}}{1-\delta}}{1-\frac{(1-2\delta)^{k-1}}{1-\delta}}\right) \\
&< \ln\left(\tfrac{\delta}{1-\delta}\right) + \tfrac{(1-2\delta)^{k-1}}{\delta} + \tfrac{\frac{(1-2\delta)^{k-1}}{1-\delta}}{1-\frac{(1-2\delta)^{k-1}}{1-\delta}} \\
&\leq \ln\left(\tfrac{\delta}{1-\delta}\right) + \tfrac{(1-2\delta)^{k-1}}{\delta} + \tfrac{\frac{(1-2\delta)^{k-1}}{1-\delta}}{1-\frac{1-2\delta}{1-\delta}} \\
&= \ln\left(\tfrac{\delta}{1-\delta}\right) + \tfrac{2(1-2\delta)^{k-1}}{\delta}\,,
\end{aligned}$$

which prove the bounds on the parameter s^* in (52). □

Lemma 3.3. Let k be an even number. Then $r(\cdot)$ is symmetric around $\delta = \frac{1}{2}$, i.e., $r(\delta) = r(1-\delta)$.

Proof. Since k is even, the all-one word is a codeword, and therefore, by linearity, $N(l) = N(n-l)$. This implies that $r(\delta) = r(1-\delta)$ where $\delta \triangleq \frac{l}{n}$ $(0 \leq \delta \leq 1)$. $\square$

Lemma 3.4. Let the function $g_{k,v}(s) = \frac{\mu_k(s)}{k} - vs$ be a function of a real parameter s where $k \geq 1$ is an integer, v is a real number, and the function $\mu_k(\cdot)$ is defined in (48). Then the *maximal value* of $g_{k,v}(\cdot)$ in any closed interval is achieved at one of its *endpoints.*

Proof. By differentiating twice the function $g_{k,v}(\cdot)$, one obtains that

$$g''_{k,v}(s) = \tfrac{k-1}{4} \cdot \exp\left(\mu_{k-2}(s) - \mu_k(s) + 2s\right),$$

which is non-negative for $-\infty < s < \infty$. This yields that the function $g_{k,v}(\cdot)$ is convex, and therefore the maximal value of $g_{k,v}(\cdot)$ in any closed interval is achieved at one of its endpoints. $\square$

At this stage, we are ready to derive simple and tight bounds on the exponential growth rate $r(\cdot)$ of the average weight distribution of Gallager's ensemble of LDPC codes.

Proposition 3.1. Consider Gallager's ensemble of (n, j, k) regular LDPC codes, and let $R = 1 - \frac{j}{k}$ be the asymptotic rate of this ensemble. Let $\delta_0 \in (0, \frac{1}{2})$ be chosen arbitrarily, and suppose $k \geq 2$ is an even number satisfying $k > z(\delta_0)$ (where $z(u) \triangleq \frac{\ln\left(\frac{2}{u(1-2u)\ln\left(\frac{1-u}{u}\right)}\right)}{\ln\left(\frac{1}{1-2u}\right)}$ for $0 < u < \frac{1}{2}$).
If $\delta_0 \leq \delta \leq \frac{1}{2}$, then the asymptotic exponential growth rate $r(\cdot)$ of the average weight distribution of this ensemble satisfies

$$h_e(\delta) - (1-R)\ln 2 \leq r(\delta) \leq h_e(\delta) + (1-R)\left[\ln\left(\tfrac{1+(1-2\delta)^k}{2}\right) + k\delta\left(\exp\left(\tfrac{2(1-2\delta)^{k-1}}{\delta}\right) - 1\right)\right], \quad (54)$$

where $h_e(\cdot)$ denotes the binary entropy function to the natural base. If $\frac{1}{2} < \delta \leq 1 - \delta_0$, then $r(\delta) = r(1-\delta)$.

Proof. The lower bound on $r(\cdot)$ is provided for completeness, as we only need the upper bound in the next subsection. It is well known that the exponential growth rate of the average weight distribution of Gallager's ensemble of LDPC codes is not below the one which corresponds to fully random codes (they coincide at $\delta = \frac{1}{2}$, and otherwise (for other values of δ in the interval $[0, 1]$), the former exponent is strictly bigger than the latter exponent).[16] Since the weight distribution of fully random codes is binomial, then this immediately implies the lower bound in (54).
We will now prove the upper bound in (54): Let s_{l} and s_{u} be the lower and upper bounds on s^* which are provided in (52), respectively. Based on (47), the bounds on s^* in (52), and Lemma 3.4, it follows that if $\delta_0 \leq \delta \leq \frac{1}{2}$ then

$$r(\delta) \leq j\left[\max\left(\frac{\mu_k(s_{\mathrm{l}})}{k} - s_{\mathrm{l}}\delta,\ \frac{\mu_k(s_{\mathrm{u}})}{k} - s_{\mathrm{u}}\delta\right) + \left(1 - \frac{1}{k}\right)\ln 2\right] - (j-1)\, h_e(\delta), \quad (55)$$

where from (52), $s_{\mathrm{l}} = \ln\left(\frac{\delta}{1-\delta}\right)$ and $s_{\mathrm{u}} = \ln\left(\frac{\delta}{1-\delta}\right) + \frac{2(1-2\delta)^{k-1}}{\delta}$.
Since $R = 1 - \frac{j}{k}$ is the asymptotic rate of Gallager's ensemble, then a short calculation reveals that

$$j\left(\frac{\mu_k(s_{\mathrm{l}})}{k} - s_{\mathrm{l}}\delta + \left(1 - \frac{1}{k}\right)\ln 2\right) - (j-1)\, h_e(\delta) = h_e(\delta) + (1-R)\cdot\ln\left(\frac{1+(1-2\delta)^k}{2}\right). \quad (56)$$

[16] This phenomenon is illustrated in Fig. 1 of [23], and it is also illustrated there that as the values of j and k are increased so that $\frac{j}{k} = 1 - R$, then the average weight distribution of Gallager's ensemble of (n, j, k) regular LDPC codes approaches more and more to the binomial distribution of fully random codes. The bounds in (54) provide a quantitative meaning to the latter phenomenon.

It can be verified that $z(\cdot)$ is a monotonically decreasing function in the interval $(0, \frac{1}{2})$. Therefore, if $k > z(\delta_0)$ as assumed in this proposition, then it implies that $k > z(\delta)$ for $\delta \in [\delta_0, \frac{1}{2})$, which is equivalent to fulfilling the condition $\frac{\delta}{1-\delta} \cdot \exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right) < 1$ for these values of δ (with an equality if $\delta = \frac{1}{2}$). Calculations (which are presented in Appendix A) show that if $\delta \in [\delta_0, \frac{1}{2}]$

$$\begin{aligned} & j\left(\tfrac{\mu_k(s_{\mathrm{u}})}{k} - s_{\mathrm{u}}\delta + \left(1 - \tfrac{1}{k}\right)\ln 2\right) - (j-1)\, h_e(\delta) \\ & \leq h_e(\delta) + (1-R)\, \ln\left(\tfrac{1+(1-2\delta)^k}{2}\right) + k(1-R)\delta \cdot \left(\exp\left(\tfrac{2(1-2\delta)^{k-1}}{\delta}\right) - 1\right). \end{aligned} \tag{57}$$

The combination of the results in (55), (56) and (57) yields the upper bound on $r(\cdot)$ in (54). The symmetry of $r(\cdot)$ around one-half is a direct consequence of the fact that k is even (see Lemma 3.3). □

3.2.2 Performance analysis under ML decoding

Following Miller and Burshtein [17, Theorem 1], our performance analysis under ML decoding combines the union bound with the Shulman and Feder bound [27]:
Let $U \subseteq \{1, 2, \cdots, n\}$ and denote the complementary set by U^c. The following upper bound on the average block error probability under ML decoding was derived in [17] for a binary linear code (or an ensemble of such codes) of block length n and a number of codewords $M = 2^{nR}$ which is used over an MBIOS channel:

$$P_{\mathrm{B}} \leq \sum_{l \in U} \left\{\overline{N(l)} D^l\right\} + 2^{-nE_{\mathrm{r}}(R + \frac{\ln \alpha}{n \ln 2})}\,, \tag{58}$$

where

$$\alpha = \max_{l \in U^c} \frac{\overline{N(l)}}{M-1} \cdot \frac{2^n}{\binom{n}{l}}\,, \tag{59}$$

$E_{\mathrm{r}}(\cdot)$ is the random coding exponent, and $D \triangleq \sum_y \sqrt{p(y|0)p(y|1)}$.
For fully random codes, if we set U to be the empty set then the upper bound (58) coincides with the random coding bound (and it therefore achieves the channel capacity).

Let $\delta_0 \in (0, \frac{1}{2})$ be an arbitrary number (δ_0 will be determined later in this proof), and define $U \triangleq \left\{l :\ 0 < \frac{l}{n} < \delta_0 \ \text{ or } \ 1 - \delta_0 < \frac{l}{n} \leq 1\right\}$. Since $\overline{N(l)} \doteq e^{nr(\delta)}$ and $\binom{n}{l} \doteq e^{nh_e(\delta)}$ where $\delta \triangleq \frac{l}{n}$, then

$$\lim_{n \to \infty} \frac{\ln \alpha}{n} = r(\delta_0) - \Big[h_e(\delta_0) - (1-R)\, \ln 2\Big], \tag{60}$$

since the weight distribution of Gallager's ensemble of LDPC codes deviates further from the binomial distribution as δ moves away from $\frac{1}{2}$, and since from Lemma 3.3, the function $r(\cdot)$ (and $h_e(\cdot)$) are symmetric around one-half if $k \geq 2$ is an even integer. Under the assumptions in Proposition 3.1, by combining Eqs. (55), (56), (57) and (60), one obtains that

$$\lim_{n \to \infty} \frac{\ln \alpha}{n} \leq (1-R)\ \left[\ln\left(1 + (1-2\delta_0)^k\right) + k\delta_0 \cdot \left(\exp\left(\frac{2(1-2\delta_0)^{k-1}}{\delta_0}\right) - 1\right)\right]\,. \tag{61}$$

As we are interested in constructing a sequence of ensembles so that it achieves a fraction $1 - \varepsilon$ of the channel capacity, then based on the error exponent of the second term in the upper bound (58) and inequality (61), it suffices that

$$(1-R)\left[\ln\left(1 + (1-2\delta_0)^k\right) + k\delta_0 \cdot \left(\exp\left(\frac{2(1-2\delta_0)^{k-1}}{\delta_0}\right) - 1\right)\right] < \varepsilon C \ln 2\,, \tag{62}$$

where Eq. (62) follows from the fact that the random coding exponent $E_{\mathrm{r}}(R)$ is positive for $R < C$. Then we need to determine the parameter δ_0 so that the first term (the union bound) in the upper bound (58) will tend asymptotically to zero. It is shown in Appendix B that if k satisfies the requirements in (B.3) and (B.6) (with the coefficients in (B.7)), then it also satisfies (62).

Let $\eta \in (0,1)$ be chosen arbitrarily, and define $\delta_0 \triangleq \eta \cdot h^{-1}(1-C)$. As δ_0 was defined independently of ε, and it is strictly smaller than the normalized Gilbert-Varshamov distance, the asymptotic normalized minimum distance of a typical code in the ensemble of Gallager's LDPC codes is above δ_0 for large values of k, and therefore $r(\delta) < 0$ for $\delta \in (0, \delta_0]$. Based on the upper bound on $r(\cdot)$ in (54), in order to ensure the latter condition, we impose the following stronger requirement on k

$$h_e(\delta_0) + (1-R)\left[\ln\left(\frac{1+(1-2\delta_0)^k}{2}\right) + k\delta_0\left(\exp\left(\frac{2(1-2\delta_0)^{k-1}}{\delta_0}\right) - 1\right)\right] < 0. \tag{63}$$

It is shown in Appendix B that if k satisfies (B.3) and (B.9), then it also satisfies (63).

So far we have derived conditions on the parameter k which ensure that the second term in the upper bound (58) tends asymptotically to zero for all rates R which do not exceed a fraction $1-\varepsilon$ of the channel capacity (and the convergence in this case is exponential in the block length n). We need also to verify that the same is true for the first term in the upper bound (58):
Based on [8, Theorem 2.4], the minimum distance distribution function of Gallager's ensemble of (n, j, k) LDPC codes has the property that there exists a positive constant $\delta_{j,k}$ so that the probability of having codewords whose normalized Hamming weight is below this constant converges asymptotically to zero, and the asymptotic behavior of this convergence is upper bounded by $\frac{k-1}{2n^{j-2}} + o\left(\frac{1}{n^{j-2}}\right)$ (if $j \geq 3$, then the above probability tends asymptotically to zero). Since

$$\begin{aligned}
&\sum_{l \in U} \overline{N(l)} D^l \\
&\leq 2 \sum_{l:\, \frac{l}{n} < \delta_0} \overline{N(l)} D^l \\
&\leq 2 \left(\sum_{l:\, \frac{l}{n} < \delta_{j,k}} \overline{N(l)} D^l + \sum_{l:\, \delta_{j,k} \leq \frac{l}{n} < \delta_0} \overline{N(l)} D^l \right) \\
&\leq \frac{k-1}{n^{j-2}} + o\left(\frac{1}{n^{j-2}}\right) + 2 \sum_{l:\, \delta_{j,k} \leq \frac{l}{n} < \delta_0} \overline{N(l)} D^l \\
&\leq \frac{k-1}{n^{j-2}} + o\left(\frac{1}{n^{j-2}}\right) + 2n\delta_0 \cdot \exp\left(n \cdot \max_{\delta_{j,k} \leq \delta \leq \delta_0} r(\delta)\right)
\end{aligned}$$

where we used in the last transition the fact that $D \leq 1$, and upper bounded $\sum_{l:\, \delta_{j,k} \leq \frac{l}{n} < \delta_0} \overline{N(l)} D^l$ by the expression $n\delta_0 \cdot \max_{\delta_{j,k} \leq \delta \leq \delta_0} \overline{N(n\delta)}$.
Since $\max_{\delta \in [\delta_{j,k}, \delta_0]} r(\delta) < 0$ (from the construction of δ_0), then it follows that also the first term in the upper bound (58) indeed converges to zero as $n \to \infty$, and therefore the same is true for the overall upper bound on the block error probability in (58). However, we note the convergence of the first and the second terms of the upper bound (58) are different: The convergence of the first term is polynomial in the block length, and the convergence of the second term is exponential in the block length, so the overall bounds tends to zero polynomially in the block length.

From the discussion so far, it follows that if k satisfies the requirements in Proposition 3.1, (B.3), (B.6) and (B.9), and $j \geq 3$ satisfies the requirement in Proposition 3.1 (i.e., $R = 1 - \frac{j}{k}$ or equivalently $j = (1-R)k$), then the sequence of Gallager's ensemble achieves the required fraction of the channel

capacity, and the asymptotic degree of the parity-check nodes of this sequence of ensembles is k. By choosing the minimal value of k which satisfies all these requirements, it can be verified that

$$k \leq A \ln\left(\frac{1}{\varepsilon}\right) + C \cdot \max(\xi_1, \xi_2, \xi_3, \xi_4) + 2,$$

where $\xi_1, \xi_2, \xi_3, \xi_4$ are defined in Theorem 2.2 and A is defined in (B.7), and the $+2$ in the last term above results from the requirement in Proposition 3.1 that k is an even positive integer. Then, the resulting asymptotic density of the parity-check matrices which represent the considered sequence of ensembles is

$$\begin{aligned}
\lim_{n\to\infty} \Delta_n &= \tfrac{(1-R)k}{R} \\
&= \tfrac{(1-(1-\varepsilon)C)k}{(1-\varepsilon)C} \\
&= \tfrac{A(1-C)}{(1-\varepsilon)C} \cdot \ln\left(\tfrac{1}{\varepsilon}\right) + \tfrac{A}{1-\varepsilon} \cdot \varepsilon \, \ln\left(\tfrac{1}{\varepsilon}\right) + \Big[C \cdot \max(\xi_1, \xi_2, \xi_3, \xi_4) + 2\Big] \cdot \left(\tfrac{1-(1-\varepsilon)C}{(1-\varepsilon)C}\right) \\
&\overset{\text{(a)}}{\leq} \tfrac{A(1-C)}{(1-\varepsilon)C} \cdot \ln\left(\tfrac{1}{\varepsilon}\right) + \tfrac{A}{(1-\varepsilon)\cdot e} + \tfrac{1}{(1-\varepsilon)C} \Big[C \cdot \max(\xi_1, \xi_2, \xi_3, \xi_4) + 2\Big] \\
&= \tfrac{1}{1-\varepsilon} \cdot \left(\tfrac{A(1-C)}{C} \cdot \ln\left(\tfrac{1}{\varepsilon}\right) + \max(\xi_1, \xi_2, \xi_3, \xi_4) + \tfrac{2}{C} + \tfrac{A}{e}\right) \\
&\overset{\text{(b)}}{=} \tfrac{1}{1-\varepsilon} \cdot \left(K_4 \ln\tfrac{1}{\varepsilon} + K_3\right),
\end{aligned}$$

where inequality (a) follows from the inequality $\varepsilon \ln\left(\frac{1}{\varepsilon}\right) \leq \frac{1}{e}$ for $0 < \varepsilon < 1$, and equality (b) follows from Eqs. (5) and (B.7). This completes the proof of Theorem 2.2.

Proof of Corollary 2.1
Since K_2 and K_4 are the coefficients of $\ln\frac{1}{\varepsilon}$ in the lower bound (1) and the upper bound (4) on the asymptotic density, respectively, then we are interested to obtain bounds on the ratio $\frac{K_4}{K_2}$ for an arbitrary MBIOS channel. From Eqs. (2) and (5)

$$\frac{K_4}{K_2} = \frac{2 \ln\left(\frac{1}{1-2w}\right)}{(1-a)\ln\left(\frac{1}{1-2\delta}\right)}. \tag{64}$$

The minimal value of $\frac{K_4}{K_2}$ is achieved in the limit where $a \to 0$ and $\eta \to 1$ (see Eq. (10)). In order to proceed in our proof, we will first prove the following lemma.

Lemma 3.5. For an arbitrary MBIOS channel, the channel capacity satisfies the inequality

$$1 - h(w) \leq C \leq 1 - 2w$$

where w is introduced in Theorem 2.1. Moreover, the upper and lower bounds on the channel capacity are achieved for a BEC and for a BSC, respectively.

Proof. From the erasure decomposition lemma [20, Appendix B], an arbitrary MBIOS channel can be decomposed to a BEC with erasure probability $2w$ which is followed by another MBIOS channel. From the data processing theorem, it follows that $C \leq 1 - 2w$, and clearly equality is achieved if the MBIOS channel is a BEC. On the other hand, assume that an arbitrary MBIOS channel (whose binary input is x, its output is y, its capacity is C, and its conditional probability distribution is $p(y|x)$) is followed by a channel whose output is $+1$ or -1 if $p(y|x=1) > p(y|x=0)$ or $p(y|x=1) < p(y|x=0)$, respectively, and whose output is equally likely $+1$ or -1 if the equality $p(y|x=1) = p(y|x=0)$ holds. Then, from the symmetry property of the former channel (i.e., since the equality $p(y|x=1) = p(-y|x=0)$ holds for all y), it follows that the equivalent channel is a BSC whose crossover probability is equal to w (and whose channel capacity is equal to $1 - h(w)$ bits per channel use). From the data processing theorem, one obtains that $C \geq 1 - h(w)$, and equality is clearly achieved if the MBIOS channel is a BSC. □

From Lemma 3.5, since $C \geq 1-h(w)$, then Eq. (10) implies that $w \geq \delta$ (which becomes an equality for a BSC). Eq. (64) then yields that $\frac{K_4}{K_2} \geq 2$ (with an equality for a BSC where also $a \to 0$). On the other hand, from Lemma 3.5, the inequality $C \leq 1-2w$ holds for an arbitrary MBIOS channel (with equality for a BEC). The maximal value of the right-hand side of Eq. (64) is therefore achieved for a BEC in the limit where $a \to 0$ and $\eta \to 1$. In the latter case, $C = 1-2w$, and it follows from Eq. (10) that $\delta = h^{-1}(1-C)$. The right-hand side of Eq. (64) then achieves the upper bound on $\frac{K_4}{K_2}$. It is clear from this proof that the upper and lower bounds on $\frac{K_4}{K_2}$ are achieved for a BEC and a BSC, respectively. As we will see in Theorem 2.3, for a BEC, the maximal value of $\frac{K_4}{K_2}$ can be reduced to *unity* by considering another sequence of ensembles of LDPC codes; the bounds in Eq. (11) refer to the sequence of ensembles of regular LDPC codes which is considered in Theorem 2.2.

3.3 Proof of Theorem 2.3

3.3.1 Proof of the statement on the bit error probability (Eqs. (13)–(17))

Based on the asymptotic analysis of iterative message-passing decoding for the BEC, sequences of capacity-approaching ensembles of LDPC codes are constructed so that one first chooses functions $\hat{\lambda}_\alpha(\cdot)$ and $\rho_\alpha(\cdot)$ which satisfy the equality $\hat{\lambda}_\alpha(1-\rho_\alpha(1-x)) \equiv x$, and so that all the coefficients in the power series expansions of $\hat{\lambda}_\alpha(\cdot)$ and $\rho_\alpha(\cdot)$ around zero are *non-negative*, and $\rho_\alpha(1) = 1$. The choice of the sequence of ensembles of LDPC in (12) was initiated by the choice $\hat{\lambda}_\alpha(x) = 1-(1-x)^\alpha$ and $\rho_\alpha(x) = x^{\frac{1}{\alpha}}$ where $0 < \alpha < 1$. In this case

$$\hat{\lambda}_\alpha(x) = \sum_{k=1}^{\infty} (-1)^{k+1} \binom{\alpha}{k} x^k \ , \quad |x| < 1 \ ,$$

so that all the coefficients in the power series expansion of $\hat{\lambda}_\alpha(x)$ are positive for $0 < \alpha < 1$. The polynomial $\hat{\lambda}_{\alpha,N}(x)$ is defined to be the truncated power series of $\hat{\lambda}_\alpha(x)$ (by taking the first $N-1$ non-zero terms (up to x^{N-1}) in this expansion), and $\lambda_{\alpha,N}(x)$ is defined so that $\lambda_{\alpha,N}(x) \triangleq \frac{\hat{\lambda}_{\alpha,N}(x)}{\hat{\lambda}_{\alpha,N}(1)}$. This approach yields the construction of the sequence of ensembles of LDPC codes in (12) (see [25]). Further, for a finite value of N, the sequence of ensembles of $(n, \lambda_{\alpha,N}, \rho_\alpha)$ LDPC codes achieves asymptotically (as $n \to \infty$) a fraction $1-\varepsilon_{\alpha,N}$ of the capacity of the BEC where

$$\varepsilon_{\alpha,N} \leq \frac{1-\hat{\lambda}_{\alpha,N}(1) - r(\alpha,N)}{1-\hat{\lambda}_{\alpha,N}(1)} \ , \tag{65}$$

and $r(\alpha,N)$ is the rate of this sequence of ensembles. In order to construct sequences of capacity-achieving ensembles of LDPC codes with vanishing bit error probability on the BEC, it is sufficient to choose the functions $\hat{\lambda}_\alpha(\cdot)$ and $\rho_\alpha(\cdot)$ so that in addition to the requirements above, also $\lim_{N\to\infty} \frac{1-\hat{\lambda}_{\alpha,N}(1)-r(\alpha,N)}{1-\hat{\lambda}_{\alpha,N}(1)} = 0$.

For the sequence of ensembles of LDPC codes in (12)

$$\begin{aligned} r(\alpha,N) &= 1 - \frac{\int_0^1 \rho_\alpha(x)\,\mathrm{d}x}{\int_0^1 \lambda_{\alpha,N}(x)\,\mathrm{d}x} \\ &= 1 - \frac{\alpha - N\binom{\alpha}{N}(-1)^{N+1}}{\alpha - \binom{\alpha}{N}(-1)^{N+1}} \ , \end{aligned} \tag{66}$$

where the latter equality relies on the the following equalities (which can be easily proved by

mathematical induction)

$$\sum_{k=1}^{N-1} \frac{(-1)^{k+1}}{k+1}\binom{\alpha}{k} = \frac{\alpha - \binom{\alpha}{N}(-1)^{N+1}}{\alpha+1}\,, \quad \sum_{k=1}^{N-1}(-1)^{k+1}\binom{\alpha}{k} = 1 - \frac{N}{\alpha}\binom{\alpha}{N}(-1)^{N+1}\,, \quad N \geq 2.$$

A slight refinement of the derivation of Proposition 1 in [25] yields that for $0 < \alpha < 1$ and for an integer $N \geq 1$

$$\frac{\alpha \cdot c(\alpha, N)}{N^{\alpha+1}} < (-1)^{N+1}\binom{\alpha}{N} \leq \frac{\alpha}{N^{\alpha+1}}\,, \tag{67}$$

where $c(\alpha, N) = (1-\alpha)^{\frac{\pi^2}{6}} \cdot e^{\alpha(\frac{\pi^2}{6} - \gamma + \frac{1}{2N})}$, and γ is Euler's constant (see Appendix C for a proof).[17] Based on (66) and (67), it can be shown that

$$\frac{\frac{c(\alpha,N)}{N^\alpha} - \frac{1}{N^{\alpha+1}}}{1 - \frac{1}{N^{\alpha+1}}} < r(\alpha, N) < \frac{\frac{1}{N^\alpha} - \frac{c(\alpha,N)}{N^{\alpha+1}}}{1 - \frac{c(\alpha,N)}{N^{\alpha+1}}}\,,$$

and therefore a choice of α and N so that $\frac{1}{N^\alpha} = 1 - p$ yields the inequality

$$\frac{(1-p)\left(c(\alpha,N) - \frac{1}{N}\right)}{1 - \frac{1-p}{N}} < r(\alpha, N) < \frac{(1-p)\left(1 - \frac{c(\alpha,N)}{N}\right)}{1 - \frac{(1-p)\,c(\alpha,N)}{N}}\,.$$

These bounds on the rate yield that $\lim_{N\to\infty} r(\alpha, N) = 1 - p$ (if $N \to \infty$ then $\alpha = \frac{\ln\left(\frac{1}{1-p}\right)}{\ln N} \to 0$ for $0 < p < 1$, and $\lim_{\alpha \to 0} c(\alpha, N) = 1$), which implies that in the limit where $N \to \infty$ (in addition to the assumption of the asymptotic analysis where the block length tends to infinity, i.e. $n \to \infty$), the sequence of ensembles of LDPC codes in (12) achieves asymptotically the capacity of a BEC with an erasure probability p. Therefore, α and N are chosen to be related according to (14). In order to choose α and N so that this sequence of ensembles achieves asymptotically a fraction $1-\varepsilon$ of the capacity of the BEC with vanishing bit error probability, then based on (65), it is sufficient to find an integer N so that $\frac{1-\hat{\lambda}_{\alpha,N}(1) - r(\alpha,N)}{1-\hat{\lambda}_{\alpha,N}(1)} \leq \varepsilon$, where $\hat{\lambda}_{\alpha,N}(1) = \sum_{k=1}^{N-1}(-1)^{k+1}\binom{\alpha}{k} = 1 - \frac{N}{\alpha}\binom{\alpha}{N}(-1)^{N+1}$ and $r(\alpha, N)$ is introduced in (66). A short calculation yields that the last inequality is equivalent to

$$\frac{1 - \frac{1}{N}}{1 - \frac{1}{\alpha}\binom{\alpha}{N}(-1)^{N+1}} \geq 1 - \varepsilon. \tag{68}$$

The main difference in the choice of the parameters in (14) as compared to (13) is based on the following step: Let $f(\cdot)$ be a function which satisfies the condition $0 < f(p) < 1$ for $0 < p < 1$. For $0 < \alpha \leq f(p)$ and $N \geq 1$, it can be verified that $c(\alpha, N) \geq c_2(f(p))$, where $c_2(\cdot)$ is introduced in (14) (we note that $c_2(f(p))$ is the value of $c(\alpha, N)$ for $\alpha = f(p)$ and $N \to \infty$). Based on (67), inequality (68) can be replaced by the stronger condition $\frac{1-\frac{1}{N}}{1-\frac{c_2(f(p))}{N^{\alpha+1}}} \geq 1 - \varepsilon$. Since α, N were chosen so that $\frac{1}{N^\alpha} = 1 - p$, then the solution of the last inequality is $N \geq \frac{1 - c_2(f(p)) \cdot (1-p)(1-\varepsilon)}{\varepsilon}$. The assumption of the analysis where $0 < \alpha \leq f(p)$ and the relation between α and N in (14) also requires that $N \geq (1-p)^{-\frac{1}{f(p)}}$. These two requirements on the integer N imply that

$$N \geq \max\left(\left\lceil \frac{1 - c_2(f(p)) \cdot (1-p)(1-\varepsilon)}{\varepsilon} \right\rceil, \left\lceil (1-p)^{-\frac{1}{f(p)}} \right\rceil\right) \triangleq N_3(\varepsilon, p, f(\cdot))\,. \tag{69}$$

[17] Proposition 1 in [25] states the existence of a constant c (instead of the function $c(\cdot,\cdot)$ in the lower bound in (67)) which is *independent* of α and N for $0 < \alpha \leq \frac{1}{2}$. The stronger version of the lower bound in (67) is required here, and in particular we make use of the equality $\lim_{\alpha\to 0} c(\alpha, N) = 1$.

As will be clarified shortly, in order to obtain a function $g(\cdot,\cdot)$ which does not add too much to the upper bound on asymptotic density (15) (as compared to (1)), we need to choose N to be as small as possible, and therefore N will be determined such that (69) is satisfied with equality. The choice of the function $f(p) = p$ $(0 < p < 1)$ and the connection above between N and α yields the setting of the parameters in (14) (i.e., in this case $N_3(\varepsilon, p, f(\cdot)) \equiv N_2(\varepsilon, p)$). We note that for small values of ε (i.e., for the most appealing case where the gap to capacity is small enough), the alternative choice of $f(\cdot) \approx 0.5295$ yields (13) (since in the latter case $N_3(\varepsilon, p, f(\cdot)) \equiv N_1(\varepsilon, p)$), and Eq. (13) could be derived as a consequence of the proof of Proposition 1 in [25].[18]

For the sequence of ensembles in (12) and (14), the asymptotic right degree of the LDPC codes is

$$\begin{aligned}
\lim_{n\to\infty} a_{\mathrm{R}}(n) &= \frac{1}{\alpha} + 1 \\
&= \frac{\ln N_2(\varepsilon, p)}{\ln\left(\frac{1}{1-p}\right)} + 1 \\
&= \frac{\ln\left(\frac{1}{\varepsilon}\right)}{\ln\left(\frac{1}{1-p}\right)} + \frac{\ln\left(\frac{p}{1-p}\right)}{\ln\left(\frac{1}{1-p}\right)} + \frac{\ln\left(\frac{\varepsilon N_2(\varepsilon,p)}{p}\right)}{\ln\left(\frac{1}{1-p}\right)} ,
\end{aligned}$$

which implies the following for the asymptotic density per information bit of the parity-check matrices which represent this sequence of ensembles (since the sequence achieves asymptotically a fraction $1 - \varepsilon$ of the capacity of a BEC (which is $1 - p$) with vanishing bit error probability)

$$\begin{aligned}
\lim_{n\to\infty} \Delta_n &= \frac{1-r(\alpha,N)}{r(\alpha,N)} \cdot \lim_{n\to\infty} a_{\mathrm{R}}(n) \\
&\leq \frac{1-(1-p)(1-\varepsilon)}{(1-p)(1-\varepsilon)} \cdot \lim_{n\to\infty} a_{\mathrm{R}}(n) \\
&= \frac{1+\frac{\varepsilon(1-p)}{p}}{1-\varepsilon} \cdot \frac{p}{1-p} \cdot \left(\frac{\ln\left(\frac{1}{\varepsilon}\right)}{\ln\left(\frac{1}{1-p}\right)} + \frac{\ln\left(\frac{p}{1-p}\right)}{\ln\left(\frac{1}{1-p}\right)} + \frac{\ln\left(\frac{\varepsilon N_2(\varepsilon,p)}{p}\right)}{\ln\left(\frac{1}{1-p}\right)} \right) \qquad (70) \\
&= \frac{K_1 + K_2 \ln\frac{1}{\varepsilon} + g(\varepsilon,p)}{1-\varepsilon} ,
\end{aligned}$$

where K_1 and K_2 are given in (3), and $g(\cdot,\cdot)$ is introduced in (16) (which completes the proof of (15) and (16)). Eq. (17) is derived from (16) as a result of the equality $\lim_{\varepsilon\to 0} \varepsilon \ln\left(\frac{1}{\varepsilon}\right) = 0$, and because the coefficients K_1 and K_2 in (3) only depend on p, and $x \leq \lceil x \rceil < x + 1$. The last inequality and the inequality $0 < \varepsilon \ln\left(\frac{1}{\varepsilon}\right) \leq \frac{1}{e}$ for $0 < \varepsilon < 1$, enables to prove that $g(\cdot, p)$ is bounded between two functions which only depend on p. The function $g(\cdot,\cdot)$ is clearly positive, as is reflected from a comparison between the lower bound (1) (which applies to all binary linear codes) and the upper bound (15) (see also Fig. 1). It can be verified numerically that the maximal value of the right-hand side of Eq. (17) is achieved at $p^* \approx 0.5009$. Hence, for $0 < p < 1$

$$\lim_{\varepsilon\to 0} g(\varepsilon, p) \leq \frac{p^* \cdot \ln\left(\frac{1-c_2(p^*)\cdot(1-p^*)}{p^*}\right)}{(1-p^*) \cdot \ln\left(\frac{1}{1-p^*}\right)} \approx 0.5407$$

where $c_2(\cdot)$ is introduced in (14), and $\gamma \approx 0.5772$ designates Euler's constant in (17). The latter result is also reflected in curve 5 in the right-hand side plot of Fig. 1.

[18] Proposition 1 in [25] restricts the observation to the interval $\alpha \in [0, \frac{1}{2}]$, since this is sufficient for the proof that the sequence of ensembles of LDPC codes achieves asymptotically the capacity of the BEC, while the degree of freedom which was introduced here for the function $f(\cdot)$ serves to reduce the value of the function $g(\cdot,\cdot)$ in (16) (see the improvement in the right-hand side plot as compared to the left-hand side plot of Fig. 1), and enhances the tightness of the lower bound (1) on the asymptotic density.

3.3.2 Proof of the statement on the block error probability (Eq. (18))

For constructing a sequence of ensembles achieving a fraction $1-\varepsilon$ of the capacity of a BEC with vanishing *block error probability*, we rely on subsection 3.3.1 which yields the existence of such a sequence with vanishing *bit error probability*, and we follow Luby et al. [13, p. 577].
Let η be an arbitrary positive number, and construct a sequence of ensembles of irregular LDPC codes, based on (12) and (14), where ε in the definition of $N_2(\varepsilon, p)$ in (14) is replaced by $\varepsilon' \triangleq \frac{\varepsilon}{1+\eta}$. Based on the proof in subsection 3.3.1, this sequence of ensembles achieves a fraction $1-\varepsilon'$ of the capacity of a BEC with vanishing bit error probability. In order to obtain vanishing block error probability, we will add a second set of parity-check nodes (whose role is similar to the concept in Lemma 3 of [13]), which typically adds to the code a small number of parity-check nodes, and therefore yields a slight reduction in the asymptotic rate of this sequence (i.e., a proper design yields that the reduction in the code rate is by a factor which is not below $\frac{1-\varepsilon}{1-\varepsilon'}$. This scaling factor becomes very close to unity (but is still below unity) for small positive values of η). The second set of parity-check nodes is characterized by the property that together with the variable nodes of this sequence of ensembles, we construct an ensemble of *regular* LDPC codes, with left and right degrees of 3 and d_{r}, respectively.[19] To summarize, we construct a sequence of ensembles of LDPC codes which are characterized by two sets of parity-check nodes: the first set of parity-check nodes are connected to the variable nodes by edges according to (12) and (14) with ε' replacing ε (characterizing sequence of ensembles of irregular LDPC codes), and the second set of parity-check nodes are connected to the variable nodes so that it specify a sequence of ensembles of regular LDPC codes. The asymptotic (total) rate of this sequence of ensembles is not below a fraction $1-\varepsilon$ of the channel capacity (i.e., it is at least $(1-\varepsilon)(1-p)$).

On a BEC, an iterative message-passing decoder fails to reveal part of the bits which are erased by the channel if a subset of these bits contains a non-empty *stopping set*, and the set of variable nodes which are not revealed at the end of this decoding process coincides with the *maximal* stopping set (the reader is referred to Section 1 in [6] for more details). For the ensemble of regular LDPC codes of block length n, left degree d_{l} and right degree d_{r}, it follows from the analysis in [22] that there exists a *positive* number $\underline{\omega}(d_{\mathrm{l}}, d_{\mathrm{r}})$ such that *at most a fraction* $O\left(\frac{1}{n}\right)$ *of the codes from this ensemble contain stopping sets of size* $\underline{\omega}(d_{\mathrm{l}}, d_{\mathrm{r}})n$ *or less* [22, Lemma 4.1].

The decoding of this sequence of ensembles for the BEC will be performed as follows: iterative message-passing decoding is first used for the sequence of codes which are induced by the variable nodes and the *first* set of the parity-check nodes. According to the proof in the previous section, vanishing bit erasure probability will be asymptotically achieved (as the block length tends to infinity). That implies that at the end of this message-passing decoding, the fraction of the variable nodes which remain unknown tends asymptotically to zero. It yields that in the second stage, an iterative message-passing decoder which relies on the output of the decoder from the first stage, and also relies on the connections between the variable nodes and the second set of parity-check nodes (representing codes from the ensemble of regular LDPC codes with left degree $d_{\mathrm{l}} = 3$ and right degree d_{r}) will finally succeed to decode successfully all the block. Otherwise, at the end of the second stage, the decoder would end with a stopping set whose size is more than $\underline{\omega}(d_{\mathrm{l}}, d_{\mathrm{r}})n$, but we already obtain after the first decoding stage a fraction of unknown variable nodes which tends asymptotically to zero, and this forms a sufficiently good starting point for the second decoding

[19] The rate of a code from the sequence of ensembles of regular LDPC codes with left and right degrees of 3 and d_{r}, respectively, is not below $1-\frac{3}{d_{\mathrm{r}}}$. Therefore, one can choose a sufficiently large value of d_{r}, so that the rate of the overall code will be reduced by a factor which is not below $\frac{1-\varepsilon}{1-\varepsilon'}$ (as a consequence of adding the small set of parity-check nodes, as described above). It can be verified that for this purpose, a right degree of $d_{\mathrm{r}} = \left\lceil \frac{3(1+\eta)}{\varepsilon\eta C} \right\rceil$ is sufficiently large.

stage to ensure that with probability 1, the decoder successfully decodes all the block. This approach suggests therefore a sequence of ensembles of codes which achieves a fraction $1-\varepsilon$ of the capacity of the BEC with vanishing *block error probability*. Based on Eq. (70), we obtain the following inequality for the asymptotic density of the parity-check matrices which represent this sequence

$$\lim_{n\to\infty} \Delta_n \leq \frac{1-\varepsilon'}{1-\varepsilon} \cdot \frac{K_1 + K_2 \ln \frac{1}{\varepsilon'} + g(\varepsilon', p)}{1-\varepsilon'} + \frac{3}{(1-p)(1-\varepsilon)} \,,$$

where the factor $\frac{1-\varepsilon'}{1-\varepsilon}$ in the first term is a consequence of the slight reduction in the rate of this sequence by adding the second set of parity-check nodes, and the second term is the contribution to the asymptotic density (per information bit) which is made by the regular LDPC codes whose left degree is three. Since $\eta > 0$ is a parameter in the construction of this sequence of ensembles which can be chosen arbitrarily small (yielding that ε' can be made as close as desired to ε, though $\varepsilon' < \varepsilon$), then the last inequality coincides with (18).

3.4 Proof of Theorem 2.4

We start in proving the first part of Theorem 2.4 which refers to the block error probability. Let $\mathbf{u}$ and $\mathbf{v}$ be the transmitted codeword in the code $\mathcal{C}$ and the received sequence, respectively (both are vectors of length n). From Fano's inequality

$$\frac{H(\mathbf{u}|\mathbf{v})}{n} \leq \frac{h(P_{\mathrm{B}})}{n} + R P_{\mathrm{B}} \,, \tag{71}$$

where P_{B} is the average block error probability of the code $\mathcal{C}$ under an arbitrary decoding algorithm (or an upper bound on the decoding error probability). If the transmission takes place over an MBIOS channel, then the combination of (30)[20] and (71) yields that

$$1 - C - (1-R) \cdot h\left(\frac{1-(1-2w)^{a_{\mathrm{R}}}}{2}\right) \leq \frac{h(P_{\mathrm{B}})}{n} + R P_{\mathrm{B}}.$$

Since $h(x) \leq 1 - \frac{2}{\ln 2} \cdot (\frac{1}{2} - x)^2$ for $0 \leq x \leq \frac{1}{2}$ (see Lemma 3.1), then

$$1 - C - (1-R)\left(1 - \frac{1}{2\ln 2} \cdot (1-2w)^{2a_{\mathrm{R}}}\right) \leq \frac{h(P_{\mathrm{B}})}{n} + R P_{\mathrm{B}} \,.$$

Since $R = (1-\varepsilon)C$, the last inequality yields that

$$\begin{aligned} a_{\mathrm{R}} &\geq \frac{\ln\left(\frac{1}{2\ln 2} \cdot \frac{1-R}{C-(1-P_{\mathrm{B}})R + \frac{h(P_{\mathrm{B}})}{n}}\right)}{2\ln\left(\frac{1}{1-2w}\right)} \\ &= \frac{\ln\left(\frac{1}{2\ln 2} \cdot \frac{1-C+\varepsilon C}{\delta_1 C + \delta_2}\right)}{2\ln\left(\frac{1}{1-2w}\right)} \end{aligned}$$

where δ_1, δ_2 are introduced in (21). Since Δ and a_{R} designate the normalized number of ones in a parity-check matrix which represent the binary linear block code $\mathcal{C}$, normalized per information bit and per parity bit, respectively, then clearly $\Delta = \left(\frac{1-R}{R}\right) a_{\mathrm{R}}$, which yields that

$$\Delta \geq \frac{1-(1-\varepsilon)C}{(1-\varepsilon)C} \cdot \frac{\ln\left(\frac{1}{2\ln 2} \cdot \frac{1-C+\varepsilon C}{\delta_1 C + \delta_2}\right)}{2\ln\left(\frac{1}{1-2w}\right)} \,,$$

[20] The subscript m in (30) is irrelevant here, and it only serves in the continuation of the proof of Theorem 2.1, where we let m tend to infinity. Here we consider a code and not a sequence of codes as in Theorem 2.1.

which coincides with (19). The proof of the lower bound for the BEC with respect to the block error probability is similar, except for the beginning of the proof which combines Fano's inequality with (36) (the subscript m in (36) is again irrelevant here).

The proof on the lower bounds on Δ with respect to the bit error probability (i.e., the derivation of (19) and (20) with the parameters δ_1, δ_2 in (22)) is based on the same way, except for replacing the upper bound (71) on $\frac{H(\mathbf{u}|\mathbf{v})}{n}$ (which is given in terms of the block error probability) by inequality (31) (i.e., $\frac{H(\mathbf{u}|\mathbf{v})}{n} \leq h(P_{\rm b})$ where $P_{\rm b}$ designates the bit error probability of the code $\mathcal{C}$).

3.5 Proof of Theorem 2.5

Before proving Theorem 2.5, we will first verify the statement in Lemma 2.1: Consider a code whose factor graph is a tree. If the code has block length n and rate R, then this factor graph has exactly n variable nodes and $(1-R)n$ parity-check nodes. Since the factor graph is a tree, there are exactly $(2-R)n-1$ edges in the graph. Consider now the parity-check matrix. Note that the number of ones in the matrix is equal to the number of edges in the graph. Therefore the density of the parity-check matrix of a cycle-free code is $\Delta = \frac{2-R}{R} - \frac{1}{nR}$, and according to Definition 2.4, its normalized density is therefore $t = 1 - \frac{1}{n(2-R)}$.

Inequality (23) follows easily from Theorem 2.4. More specifically, it follows from inequality (19), Eq. (22) (where in this case $\varepsilon = 1 - \frac{R}{C}$), and the equality $\Delta = \left(\frac{2-R}{R}\right) t$ (see Definition 2.4). Similarly, inequality (24) follows from inequality (20), Eq. (22) and Definition 2.4.

From inequality (23), the lower bound on the bit error probability ($P_{\rm b}$) is meaningful if the right-hand side of this inequality is positive, i.e.,

$$R - C + \frac{1}{2\ln 2} \cdot (1-R)(1-2w)^{\frac{2t(2-R)}{1-R}} > 0.$$

Let the code rate be a fraction $1-\varepsilon$ of the channel capacity, i.e., $R = (1-\varepsilon)C$. The substitution of R in the last inequality gives

$$-C\varepsilon + \frac{1}{2\ln 2} \cdot \big(1-(1-\varepsilon)C\big)(1-2w)^{\frac{2t(2-(1-\varepsilon)C)}{1-(1-\varepsilon)C}} > 0. \tag{72}$$

The solution of (72) is $\varepsilon < \varepsilon_0$, where $\varepsilon^* = \varepsilon_0$ is the value which sets the left-hand side of (72) to zero, i.e.,

$$\frac{1}{2\ln 2} \cdot \left(1 + \frac{1-C}{\varepsilon_0 C}\right) = \left(\frac{1}{1-2w}\right)^{2t\left(1+\frac{1}{1-(1-\varepsilon_0)C}\right)}. \tag{73}$$

By doing the substitutions

$$x = \frac{1}{1-(1-\varepsilon_0)C}\ ,\ u = 1-(1-C)x\ ,\ z = \frac{2ut\cdot\ln(1-2w)}{1-C}\ , \tag{74}$$

Eq. (73) transforms to the equation

$$ze^z = \frac{t\cdot\ln(1-2w)\cdot(1-2w)^{\frac{2t(2-C)}{1-C}}}{\ln 2\cdot(1-C)}\ . \tag{75}$$

The solution of Eq. (75) is

$$z = W\left(\frac{t\cdot\ln(1-2w)\cdot(1-2w)^{\frac{2t(2-C)}{1-C}}}{\ln 2\cdot(1-C)}\right)\ , \tag{76}$$

where the function $W(\cdot)$ is the Lambert W-function [28]. Finally, the reverse substitutions of those in (74) transform the value of z in (76) to the value of ε_0 in (25), where $1-\varepsilon_0$ is the achievable fraction of the channel capacity.

For the model of a BEC, the improved lower bound (24) on the bit erasure probability (as compared to (23) with $w = \frac{p}{2}$) yields to a wider range of values of ε for which the former lower bound is meaningful. The lower bound on the bit erasure probability (24) is useful if its right-hand side is positive. By setting $R = (1-\varepsilon)(1-p)$ in the right-hand side of (24), one obtains the condition

$$-(1-p)\varepsilon + \big(1-(1-p)(1-\varepsilon)\big)\cdot(1-p)^{\frac{(2-(1-p)(1-\varepsilon))t}{1-(1-p)(1-\varepsilon)}} > 0 \tag{77}$$

whose solution is $\varepsilon < \varepsilon_0$, where $\varepsilon^* = \varepsilon_0$ sets the left-hand side of (77) to zero, i.e.,

$$1 + \frac{p}{\varepsilon_0(1-p)} = \left(\frac{1}{1-p}\right)^{\left(1+\frac{1}{1-(1-\varepsilon_0)(1-p)}\right)t} . \tag{78}$$

By doing sequentially the substitutions

$$x = \frac{1}{1-(1-\varepsilon_0)(1-p)} \ , \ u = 1-px \ , \ z = \frac{ut\cdot\ln(1-p)}{p} \ , \tag{79}$$

Eq. (78) transforms to the equation

$$ze^z = \frac{t\cdot\ln(1-p)\cdot(1-p)^{\left(1+\frac{1}{p}\right)t}}{p} . \tag{80}$$

The solution of Eq. (80) is

$$z = W\left(\frac{t\cdot\ln(1-p)\cdot(1-p)^{\left(1+\frac{1}{p}\right)t}}{p}\right) \ , \tag{81}$$

and then, the reverse substitutions of those in Eq. (79) transform the value of z in Eq. (81) to the value of ε_0 in Eq. (26).

- **A consequence of Theorem 2.5**
  The lower bounds on the bit error probability in Eqs. (23) and (24) can be easily applied to ensembles of binary linear block codes of a given rate, where the bit error probability of a code (P_{b}) is replaced with the average bit error probability over the ensemble ($\overline{P_{\mathrm{b}}}$), and the normalized density (t) of a parity-check matrix which represents a code is replaced with the average normalized density ($\overline{t}$). This is readily justified by applying Jensen's inequality to both sides of Eqs. (23) and (24), i.e.,

$$E\left[h(P_{\mathrm{b}})\right] \le h\left(\overline{P_{\mathrm{b}}}\right) , \ \ E\Big[(1-2w)^{\beta t}\Big] \ge (1-2w)^{\beta\overline{t}} \ \ \forall\,\beta\in\mathbb{R}.$$

  Based on the proof of Theorem 2.5, it follows that the expression of ε_0 in Eqs. (25) and (26) is valid for an arbitrary ensemble of binary linear block codes of a given rate (R); to this end, one replaces t in Eqs. (25) and (26) with $\overline{t}$. Corollary 2.3 and the discussion above suggests the following interpretation to ε_0 in Eqs. (25) and (26); it is a lower bound on the gap to capacity for an arbitrary sequence of binary linear block codes (or a sequence of ensembles of these codes) under optimal decoding (and hence, under an arbitrary sub-optimal decoding algorithm).

  Consider a sequence of ensembles of (n,λ,ρ) LDPC codes, whose block length (n) tends to infinity. Since the asymptotic rate and the normalized density are calculable in terms of the polynomials

$\lambda(\cdot)$ and $\rho(\cdot)$, then we will show that it is possible to compute a lower bound on the asymptotic gap to capacity for this sequence of ensembles; the bound is valid under ML decoding (and also under an arbitrary iterative decoding algorithm). The asymptotic rate of the considered sequence of ensembles is

$$R = 1 - \frac{\int_0^1 \rho(x)\,\mathrm{d}x}{\int_0^1 \lambda(x)\,\mathrm{d}x}\,. \tag{82}$$

Since the average number of ones in a parity-check matrix of a code from an ensemble of (n, λ, ρ) LDPC codes is equal to

$$\frac{n}{\int_0^1 \lambda(x)\,\mathrm{d}x}$$

and (by Definitions 2.2 and 2.4), it is also equal to $n(2-R)\bar{t}$, then a short calculation shows that the average value of the asymptotic normalized density is equal to

$$\bar{t} = \frac{1}{\int_0^1 \lambda(x)\,\mathrm{d}x + \int_0^1 \rho(x)\,\mathrm{d}x}\,. \tag{83}$$

From Theorem 2.5 and the discussion above, a lower bound on the inherent gap to capacity (ε_0) of a sequence of ensembles of LDPC codes can be calculated by solving numerically the equation $1 - \frac{R}{C} = \varepsilon_0$, where ε_0 is given in Eq. (25) for a general MBIOS channel and is improved in Eq. (26) for a BEC, the asymptotic rate (R) is given in Eq. (82), and C designates the channel capacity (in bits per channel use). For an arbitrary sequence of ensembles of LDPC codes whose transmission takes place over an MBIOS channel, a lower bound on the gap to the channel capacity is therefore calculated numerically by solving the equation

$$\frac{\int_0^1 \lambda(x)\,\mathrm{d}x}{\int_0^1 \rho(x)\,\mathrm{d}x} = \frac{1}{1-C} - \left(\frac{1}{2\bar{t}\ln(1-2w)}\right) \cdot W\Big(\frac{\bar{t}\cdot\ln(1-2w)\cdot(1-2w)^{\frac{2\bar{t}(2-C)}{1-C}}}{\ln 2\cdot(1-C)}\Big) \tag{84}$$

where C and w depend on the channel, $\bar{t}$ is given in Eq. (83), the function $W(\cdot)$ designates the Lambert W-function [28], and Eq. (84) follows from Eqs. (25) and (82).

For a BEC with an erasure probability p, an improved lower bound on the gap to capacity can be calculated numerically by solving the equation

$$\frac{\int_0^1 \lambda(x)\,\mathrm{d}x}{\int_0^1 \rho(x)\,\mathrm{d}x} = \frac{1}{p} - \left(\frac{1}{\bar{t}\ln(1-p)}\right) \cdot W\Big(\frac{\bar{t}\cdot\ln(1-p)\cdot(1-p)^{\bar{t}\left(1+\frac{1}{p}\right)}}{p}\Big) \tag{85}$$

where Eq. (85) is equivalent to the original equation (i.e., $1 - \frac{R}{C} = \varepsilon_0$), and it is based on the improved value of ε_0 for a BEC in Eq. (26) (as compared to the value of ε_0 in Eq. (25) which refers to an arbitrary MBIOS channel). To conclude, this discussion leads to the following result

Corollary 3.2. Consider an arbitrary sequence of ensembles of (n, λ, ρ) binary LDPC codes whose transmission takes place over an MBIOS channel. A lower bound on the asymptotic gap to capacity under *optimal decoding* (i.e., ML decoding) can be calculated numerically by solving Eq. (84). This lower bound can be improved for a BEC by solving Eq. (85).

4 Numerical Results

Our goal here is to show numerical results for the information theoretic bounds on the limitations of capacity-approaching binary linear codes over MBIOS channels. These numerical results mainly refer to Theorem 2.5 and Corollary 3.1.

Tables 1–3 present numerical results for thresholds of ensembles of LDPC codes whose transmission takes place over a BEC, a BSC or a binary-input AWGN channel. Following Burshtein et al. [4], the ultimate Shannon capacity limit of the channel is compared with bounds on their thresholds under ML decoding and iterative message-passing decoding. The numerical calculation of the latter threshold (the RU threshold) is based on the density evolution analysis [19]. The difference between [4, Theorem 1, Eq. (16)] and Eq. (37) here provides an improved upper bound on the maximal achievable rate for reliable communications over a BEC, and also suggests a certain improvement on the maximal achievable rate under ML decoding for a general MBIOS channel (with the exception of a BSC, where [4, Theorem 1] and Eq. (37) coincide). We note that for a BEC and a BSC, Corollary 3.1 provides an upper bound on the threshold under ML decoding, and it also provides a lower bound on the $\frac{E_\mathrm{b}}{N_\mathrm{o}}$–threshold for a binary-input AWGN channel (since if the erasure (crossover) probability of a BEC (BSC) is increased, then the channel is degraded; similarly, the same happens if the value of $\frac{E_\mathrm{b}}{N_\mathrm{o}}$ for a binary-input AWGN channel is decreased). It also follows from Theorem 2.1 that the value of w in Corollary 3.1 is equal to $w = \frac{p}{2}$ for a BEC with erasure probability p, $w = \min(p, 1-p)$ for a BSC with crossover probability p, and $w = Q\left(\sqrt{\frac{2RE_\mathrm{b}}{N_\mathrm{o}}}\right)$ for a binary-input AWGN channel with antipodal signaling (where $Q(\cdot)$ stands for the complementary Gaussian cumulative distribution function).

From Table 3, it can be verified that even for ensembles of LDPC codes which achieve near-Shannon capacity limit performance under iterative message-passing decoding, there exists an inherent gap between the channel capacity and the calculated bounds on the thresholds under optimal ML decoding; the gap is attributed to the moderate values of the average normalized density ($\overline{t}$) of the parity-check matrices of these ensembles (according to Theorem 2.1 and Definition 2.4, the normalized density scales at least like $\ln \frac{1}{\varepsilon}$ where ε designates the gap to capacity, and it therefore tends to infinity as this gap vanishes). A comparison of bounds on the thresholds under ML decoding with the RU thresholds provides an upper bound on the inherent loss in performance of the sub-optimal message-passing decoder (as compared to optimal ML decoding). For example, from Table 2, it follows that for a binary-input AWGN channel and an ensemble of $(3, 6)$ regular LDPC codes (whose asymptotic rate is one-half), the gap between the threshold under ML decoding and the capacity limit is between 0.062 and 0.486 dB, and the loss in $\frac{E_\mathrm{b}}{N_\mathrm{o}}$ due to the sub-optimality of the iterative message-passing decoding (as compared to the ML decoding algorithm) is between 0.437 and 0.861 dB.

Code Ensemble	Channel	Rate	Capacity	Upper Bound	Lower Bound	RU Threshold
(3,6) LDPC	BEC	$\frac{1}{2}$	0.5000	0.4913	0.483	0.429
(4,6) LDPC	BEC	$\frac{1}{3}$	0.6667	0.6657	0.665	0.506
(3,4) LDPC	BEC	$\frac{1}{4}$	0.7500	0.7469	0.744	0.647
(3,6) LDPC	BSC	$\frac{1}{2}$	0.1100	0.1025	0.092	0.084
(4,6) LDPC	BSC	$\frac{1}{3}$	0.1740	0.1726	0.170	0.116
(3,4) LDPC	BSC	$\frac{1}{4}$	0.2145	0.2109	0.205	0.167

Table 1: Comparison of thresholds for a BEC/ BSC and ensembles of regular LDPC codes. The upper bound on the threshold for the erasure/ crossover probability (p) refers to ML decoding and is based on Eq. (37) in Corollary 3.1. The lower bound on the threshold under ML decoding is based on an upper bound on the error performance under 'typical pairs' decoding (see [12, Table 2.1]), and the RU threshold is under iterative message-passing decoding [19].

Code Ensemble	Rate	Capacity	Lower Bound	Upper Bound	RU Threshold
(3,6) LDPC	$\frac{1}{2}$	+0.187 dB	+0.249 dB	+0.673 dB	+1.110 dB
(4,6) LDPC	$\frac{1}{3}$	−0.495 dB	−0.488 dB	−0.423 dB	+1.674 dB
(3,4) LDPC	$\frac{1}{4}$	−0.794 dB	−0.761 dB	−0.510 dB	+1.003 dB

Table 2: Comparison of thresholds for a binary-input AWGN channel and ensembles of regular LDPC codes. The lower bound on the threshold of $\frac{E_b}{N_o}$ refers to ML decoding and is based on Eq. (37) in Corollary 3.1. The upper bound on the threshold of $\frac{E_b}{N_o}$ under ML decoding is based on an upper bound on the error performance under 'typical pairs' decoding [10], and the RU threshold is under iterative message-passing decoding [19].

The upper plots of Figs. 2 and 3 present lower bounds on the bit erasure/ error probability of binary linear codes which are transmitted over a BEC or a BSC, respectively. The bounds rely on Theorem 2.5, and are plotted as a function of the normalized density of an arbitrary parity-check matrix which represents such a binary linear code. We note that the values of p in Figs. 2 and 3 were chosen so that the capacity (C) of the BEC and the BSC is equal to $\frac{1}{2}$ bit per channel use (so the probability of erasure of the BEC is $p = 1 - C = \frac{1}{2}$, and the crossover probability of the BSC is $p = h^{-1}(1 - C) = 0.110$). The bounds in the upper plot of Fig. 2 are based on Eq. (24), and are depicted for binary linear codes whose rate is a fraction $1 - \varepsilon$ of the channel capacity. For example, assume that one wishes to design a binary LDPC code which achieves a bit erasure probability of 10^{-6} at a rate which is 99.9% of the capacity of a BEC whose erasure probability is $\frac{1}{2}$. Then curve 3 in the upper plot of Fig. 2 implies that the normalized density of every parity-check matrix of such an LDPC code should be at least $t_{\min} = 3.325$. Since the designed rate of the code is nearly one-half (i.e., $R = 0.999C = 0.4995$ bits per channel use), then it yields that the density of every parity-check matrix of such an LDPC code should be at least $\Delta_{\min} = \frac{(2-R)t_{\min}}{R} = 9.987$. Similarly, the upper plot in Fig. 3 refers to a BSC, and is based on Eq. (23). It is reflected from the upper plots of Figs. 2 and 3 that the lower bounds on the normalized density of an arbitrary parity-check matrix (t) which represents a binary linear code whose bit erasure/ error probability is low, grow significantly as the accepted gap to capacity (εC) tends to zero. This observation agrees with the statement in Theorem 2.1 which implies that the minimal value of the normalized density (t) behaves like $\ln\left(\frac{1}{\varepsilon}\right)$.[21] It also explains why curves 2–10 in the upper plots of Figs. 2 and 3 have infinite slope as P_b goes to zero (the reason is that if $\varepsilon > 0$ and $P_b \to 0$, the lower bounds on the density (Δ) in Eqs. (19) and (20) tend to finite numbers, and therefore from Definition 2.4, the corresponding values of t also tend to finite numbers (i.e., $\frac{C}{2-C}$ times the limit of the density (Δ)). Since the lower bounds on t are finite in the latter case, then curves 2–10 in the upper plots of Figs. 2 and 3 should have an infinite slope as $P_b \to 0$). If $\varepsilon \to 0$ (i.e., the code achieves the channel capacity), then the normalized density of an arbitrary parity-check matrix which represents this code tends to infinity (as is also reflected in curve 1 of Figs. 2 and 3).

The lower plots in Figs. 2 and 3 depict information theoretic lower bounds on the achievable gap to capacity with vanishing bit erasure/ error probability (where this gap is normalized w.r.t. the channel capacity). These bounds refer to the BEC and BSC, respectively, and are valid for any sequence of codes and for optimal ML decoding (and hence, are valid for any sub-optimal decoding algorithm). These lower bounds on the normalized gap to capacity are plotted for a BEC and a BSC as a function of the channel parameter p (which designates the erasure probability or the crossover probability, respectively). Every curve in the lower plots of Figs. 2 and 3 refers to a fixed

[21] From Definition 2.4, $\lim_{\varepsilon \to 0} \frac{t}{\Delta} = \frac{C}{2-C}$. Theorem 2.1 yields therefore that as $\varepsilon \to 0$, both t and Δ grow at least like $\ln\left(\frac{1}{\varepsilon}\right)$.

$\lambda(x)$	$\rho(x)$	$\bar{t}$	Lower Bound	RU Threshold
$0.38354x + 0.04237x^2 + 0.57409x^3$	$0.24123x^4 + 0.75877x^5$	1.908	+0.269 dB	+0.809 dB
$0.23802x + 0.20997x^2 + 0.03492x^3 + 0.12015x^4 + 0.01587x^6 + 0.00480x^{13} + 0.37627x^{14}$	$0.98013x^7 + 0.01987x^8$	2.673	+0.201 dB	+0.335 dB
$0.21991x + 0.23328x^2 + 0.02058x^3 + 0.08543x^5 + 0.06540x^6 + 0.04767x^7 + 0.01912x^8 + 0.08064x^{18} + 0.22798x^{19}$	$0.64854x^7 + 0.34747x^8 + 0.00399x^9$	2.776	+0.198 dB	+0.310 dB
$0.19606x + 0.24039x^2 + 0.00228x^5 + 0.05516x^6 + 0.16602x^7 + 0.04088x^8 + 0.01064x^9 + 0.00221x^{27} + 0.28636x^{29}$	$0.00749x^7 + 0.99101x^8 + 0.00150x^9$	2.998	+0.194 dB	+0.274 dB

Table 3: Comparison of thresholds for the binary-input AWGN channel and ensembles of irregular LDPC codes with good degree distribution pairs of rate one-half. The Shannon capacity limit corresponds to $\frac{E_{\mathrm{b}}}{N_{\mathrm{o}}} = +0.187$ dB. The lower bound on the threshold of $\frac{E_{\mathrm{b}}}{N_{\mathrm{o}}}$ refers to ML decoding and is based on Eq. (37) in Corollary 3.1, and the RU threshold is under iterative message-passing decoding [19]. The degree distributions of the ensembles and their RU thresholds are taken from [20, Tables 1 and 2]. The average normalized densities ($\bar{t}$) of the parity-check matrices of these ensembles are also provided based on (83).

value of the normalized density of the parity-check matrix (t), which is depicted for values between 1 and 4.5 in increments of 0.25 (the greater the value of t is, the smaller is the lower bound on the achievable gap to capacity for all values of p). These lower bounds on the achievable gap to capacity for a BEC or a BSC are based on Eqs. (26) and (25), respectively. We note that if $p = 0$ (i.e., the channel is noiseless), then the capacity of the channel (which is clearly 1 bit per channel use) is achieved without any channel coding, and therefore it is clear why the lower bound on ε in this case is zero. On the other hand, if $p = 1$ for the BEC or if $p = 0.5$ for the BSC, then the channel capacity is zero ($C = 0$), and again it explains why the lower bound on ε is also zero in the latter case. Therefore, the fact that the lower bound on the achievable gap to capacity is not a monotonic function of p is not surprising, and this is indeed reflected in curves 1–15 which depict the lower plots of Figs. 2 and 3. For example, it follows from curves 1–13 in the lower plot of Fig. 2 that if there exists a parity-check matrix which represents a binary LDPC code and whose normalized density does not exceed 4, then the normalized gap to capacity of this LDPC code cannot be below 0.1% of the channel capacity if the probability of erasure of the BEC lies in the range $0.08 \leq p \leq 0.26$ (even if this code is ML decoded). It is also reflected from a comparison of the lower bounds on the achievable gap to capacity for the BEC and BSC (see the lower plots of Figs. 2 and 3, respectively) that the lower bounds on the gap to capacity for the BSC are more pessimistic than those for the BEC. This may be a result of the provable tightness of the bounds for the BEC, while for the BSC these bounds reflect the correct behavior but are seemingly somewhat less tight than those for the BEC (see Corollary 2.1 and Theorem 2.3). We finally note that the reason that we present in Figs. 2 and 3 curves corresponding to normalized densities (t) which are at least equal to unity is because cycle-free codes have parity-check matrices whose density is $t = 1 - \frac{1}{(2-R)n} \geq 1 - \frac{1}{n}$ (this readily results in from Lemma 2.1 and Definition 2.4), and hence

since the block length n is typically much larger than 1, then for cycle-free codes $t \approx 1$ (and in the limit where the block length tends to infinity, the corresponding density for cycle-free codes goes to unity). Clearly, if a binary code is represented by a bipartite graph with cycles, then the normalized density (t) of its corresponding parity-check matrix should increase as compared to the cycle-free case (see Corollary 2.5), which justifies our interest on values of t above unity.
Figs. 2 and 3 (which are based on Theorem 2.5) and the result in Corollary 2.5 which connects the normalized density of a parity-check matrix with the cardinality of the set of fundamental cycles in the corresponding bipartite graph of a binary linear block code, illustrate that bipartite graphs which represent good error correction codes should have cycles (see also [7]).

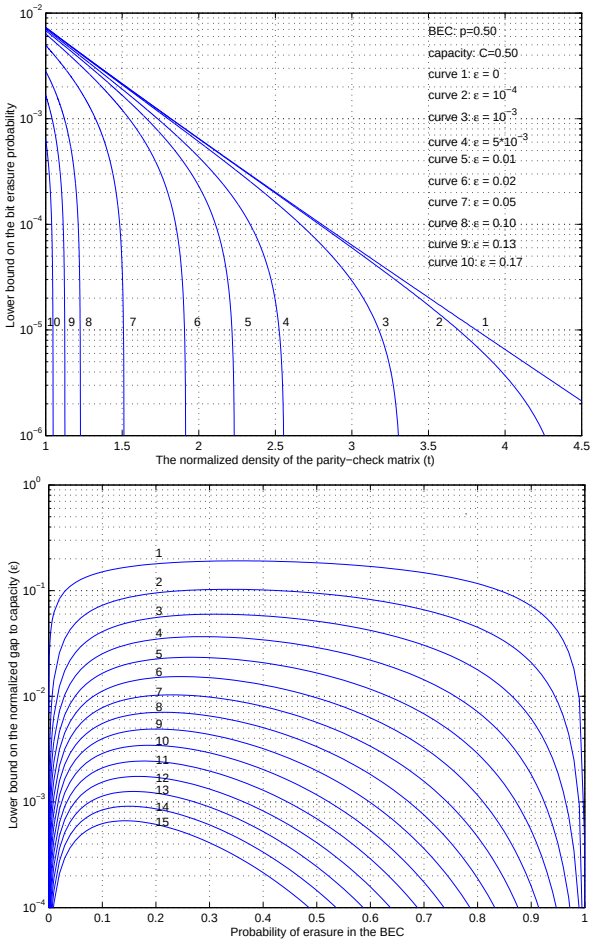

Figure 2: Upper plot: lower bounds on the bit erasure probability for any binary linear code which is transmitted over a BEC. The bounds are depicted in terms of the normalized density of an arbitrary parity-check matrix which represents the code, and the curves correspond to code rates which are a fraction $1-\varepsilon$ of the channel capacity (for different values of ε). The erasure probability of the BEC is $p = 0.500$ (which yields a capacity of one-half bits per channel use). Lower plot: lower bounds on the achievable gap to capacity with vanishing bit erasure probability (where this gap is normalized w.r.t. the channel capacity) for any sequence of codes operating over a BEC. The bounds are depicted as a function of the probability of erasure of the channel, and every single curve corresponds to a fixed value of the asymptotic normalized density of the parity-check matrices which represent this sequence of codes: curve no. i (where $i = 1, 2, \cdots, 15$) corresponds to a normalized density which is equal to $t_i = 1 + 0.25(i-1)$.

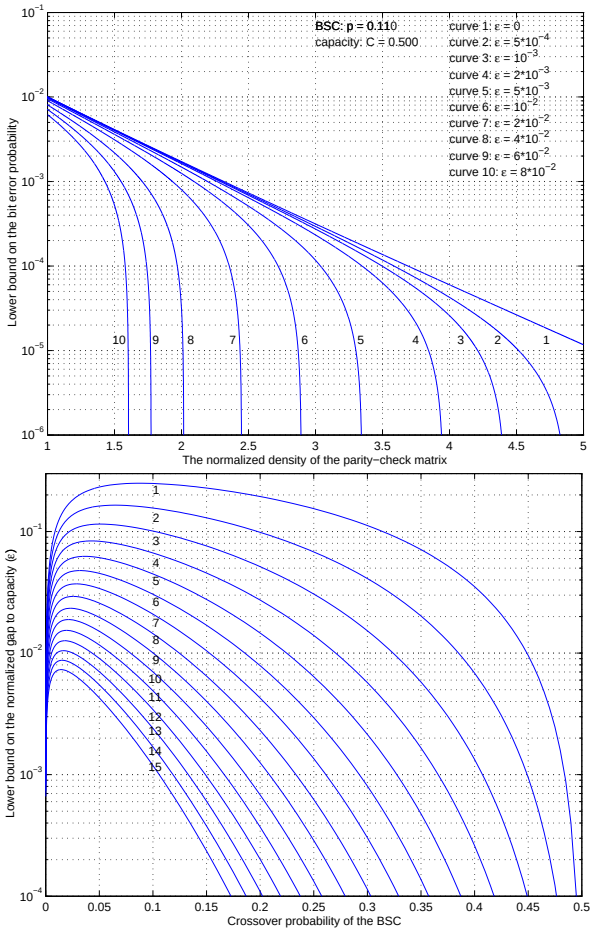

Figure 3: Upper plot: lower bounds on the bit error probability for any binary linear code which is transmitted over a BSC. The bounds are depicted in terms of the normalized density of an arbitrary parity-check matrix which represents the code, and the curves correspond to code rates which are a fraction $1-\varepsilon$ of the channel capacity (for different values of ε). The crossover probability of the BSC is $p = 0.110$ (which yields a capacity of one-half bits per channel use). Lower plot: lower bounds on the achievable gap to capacity with vanishing bit error probability (where this gap is normalized w.r.t. the channel capacity) for any sequence of codes operating over a BSC. The bounds are depicted as a function of the crossover probability of the channel, and every single curve corresponds to a fixed value of the asymptotic normalized density of the parity-check matrices which represent this sequence of codes: curve no. i ($i = 1, 2, \cdots, 15$) corresponds to a normalized density which is equal to $t_i = 1 + 0.25(i-1)$.

5 Outlook

We gather here what we consider to be the most interesting open problems in this research.

1. Theorems 2.1 and 2.3 show that for any iterative decoder which is based on the standard Tanner graph, there is a tradeoff between performance and complexity which cannot be surpassed. For the BEC, it can be achieved up to a small constant. This begs the question if better tradeoffs can be achieved by allowing more complicated graphical models (e.g., graphs which also involve state nodes, in addition to variable nodes and parity-check nodes used for representing codes by bipartite graphs). More generally, is it true that for any sequence of codes which under an arbitrary decoding algorithm (iterative or not) achieves a certain fraction of capacity with vanishing bit or block error probability, the encoding/decoding complexity can be linked to the density of the parity-check matrices which represent these codes (or to their gap to capacity) ?

2. Is it possible to improve the tightness of the lower bound (1) for general MBIOS channels ? From Theorem 2.2, it is clear that the logarithmic growth rate of the lower bound (1) reflects (up to a scaling factor) the real behavior of the best possible asymptotic density, but there may be a possibility to increase the coefficient K_2 in (2) (i.e., to increase the coefficient of the logarithm in the lower bound (1)) so that it will coincide with the logarithmic growth rate of the asymptotic density for a certain capacity-achieving sequence of ensembles (as was demonstrated for the BEC in Theorem 2.3). However, we note on the large gap between the current understanding of iterative message-passing decoding over a BEC and other types of channels (for a general MBIOS channel, it is not even known whether capacity can be achieved under iterative decoding).

3. Based on Theorem 2.2, it was noted in Section 2 that for the BSC, the coefficient of the logarithm in the upper bound (4) can be made as close as desired to twice the coefficient of the logarithm in the lower bound (1). For the BEC, the tightness of the lower bound (1) with the improved coefficients in (3) was demonstrated in Theorem 2.3 even under a sub-optimal decoding algorithm (the iterative message-passing decoding). It will be interesting to see if expander codes which attain the capacity of the BSC under iterative decoding [2], enable one to approach the information theoretic lower bound (1) with the coefficients in (2) for the BSC.

4. Extensions of the results in this paper to channels with memory (e.g., channels with ISI), and to non-binary linear block codes.

Appendix A
Derivation of inequality (57)

For $\delta \in (0, \frac{1}{2})$ and an integer $k \geq 2$, one obtains

$$
\begin{aligned}
& j\left(\tfrac{\mu_k(s_{\mathrm{u}})}{k} - s_{\mathrm{u}}\delta + \left(1-\tfrac{1}{k}\right)\ln 2\right) - (j-1)\, h_e(\delta) \\
& \overset{\text{(a)}}{=} \tfrac{j}{k}\cdot \ln\left[\left(\tfrac{1+\exp(s_{\mathrm{u}})}{2}\right)^k + \left(\tfrac{1-\exp(s_{\mathrm{u}})}{2}\right)^k\right] - j s_{\mathrm{u}}\delta + j\left(1-\tfrac{1}{k}\right)\ln 2 - (j-1)\, h_e(\delta) \\
& \overset{\text{(b)}}{=} \tfrac{j}{k}\cdot \ln\left[\left(\frac{1+\frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right)}{2}\right)^k + \left(\frac{1-\frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right)}{2}\right)^k\right] \\
& \quad -j\delta\left[\ln\left(\tfrac{\delta}{1-\delta}\right) + \tfrac{2(1-2\delta)^{k-1}}{\delta}\right] + j\left(1-\tfrac{1}{k}\right)\ln 2 - (j-1)\, h_e(\delta) \\
& = j\ \ln\left(\frac{1+\frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right)}{2}\right) + \tfrac{j}{k}\ \ln\left[1 + \left(\frac{1-\frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right)}{1+\frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right)}\right)^k\right] \\
& \quad +j\left(\delta\ln\left(\tfrac{1}{\delta}\right) + (1-\delta)\ln\left(\tfrac{1}{1-\delta}\right)\right) - j\ln\left(\tfrac{1}{1-\delta}\right) - 2j(1-2\delta)^{k-1} + j\left(1-\tfrac{1}{k}\right)\ln 2 \\
& \quad -(j-1)\, h_e(\delta) \\
& = j\ \ln\left(1 + \tfrac{\delta}{1-\delta}\cdot\exp\left(\tfrac{2(1-2\delta)^{k-1}}{\delta}\right)\right) + \tfrac{j}{k}\ \ln\left(\tfrac{1}{2}\cdot\left[1 + \left(\frac{1-\frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right)}{1+\frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right)}\right)^k\right]\right) \\
& \quad +h_e(\delta) - j\ln\left(\tfrac{1}{1-\delta}\right) - 2j(1-2\delta)^{k-1} \\
& = j\ \ln\left(1-\delta+\delta\cdot\exp\left(\tfrac{2(1-2\delta)^{k-1}}{\delta}\right)\right) + \tfrac{j}{k}\ \ln\left(\tfrac{1}{2}\cdot\left[1 + \left(\frac{1-\frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right)}{1+\frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right)}\right)^k\right]\right) \\
& \quad +h_e(\delta) - 2j(1-2\delta)^{k-1} \\
& \overset{\text{(c)}}{=} h_e(\delta) + (1-R)\ln\left(\tfrac{1}{2}\cdot\left[1 + \left(\frac{1-\frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right)}{1+\frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right)}\right)^k\right]\right) \\
& \quad +k(1-R)\ln\left(1+\delta\cdot\left(\exp\left(\tfrac{2(1-2\delta)^{k-1}}{\delta}\right)-1\right)\right) - 2k(1-R)(1-2\delta)^{k-1} \\
& \overset{\text{(d)}}{\leq} h_e(\delta) + (1-R)\ln\left(\tfrac{1}{2}\cdot\left[1 + \left(\frac{1-\frac{\delta}{1-\delta}}{1+\frac{\delta}{1-\delta}}\right)^k\right]\right) + k(1-R)\delta\cdot\left(\exp\left(\tfrac{2(1-2\delta)^{k-1}}{\delta}\right)-1\right) \\
& = h_e(\delta) + (1-R)\left[\ln\left(\tfrac{1+(1-2\delta)^k}{2}\right) + k\delta\cdot\left(\exp\left(\tfrac{2(1-2\delta)^{k-1}}{\delta}\right)-1\right)\right]
\end{aligned}
$$

where equalities (a), (b) and (c) are based on Eqs. (48) and (52) (s_{u} is the upper bound on s^* in Eq. (52)), and on the equality $R = 1 - \frac{j}{k}$ for the asymptotic rate of Gallager's ensemble, respectively. Inequality (d) is based on the inequality $\ln(1+x) \leq x$ for $x \geq 0$ and by neglecting the last term before transition (d) (which is non-positive). For transition (d) we also rely on the fact that $f(u) = \ln\left(\frac{1-u}{1+u}\right)$ is a monotonically decreasing function on the interval $(-1, 1)$, and that $\frac{\delta}{1-\delta} < \frac{\delta}{1-\delta}\cdot\exp\left(\frac{2(1-2\delta)^{k-1}}{\delta}\right) < 1$; the right-hand side of the latter inequality is fulfilled according to the assumption in Proposition 3.1 (see the explanation before Eq. (57)).

Appendix B
Derivation of sufficient conditions for the fulfillment of inequalities (62) and (63)

A sufficient condition for the fulfillment of inequality (62): Since the left side of (62) tends to zero as $k \to \infty$, then there is a hope to find a value of k in terms of ε which satisfies this inequality. Unfortunately, since inequality (62) does not lend itself to an analytical solution of k in terms of δ_0 and ε, then we will replace the left side of (62) by an upper bound, so that the new inequality which implies a stronger requirement on k can be solved analytically. First we have for $\delta_0 \in (0, \frac{1}{2})$

$$\begin{aligned}
&(1-R)\left[\ln\left(1+(1-2\delta_0)^k\right)+k\delta_0\cdot\left(\exp\left(\tfrac{2(1-2\delta_0)^{k-1}}{\delta_0}\right)-1\right)\right] \\
&<\ln\left(1+(1-2\delta_0)^k\right)+k\cdot\left(\exp\left(\tfrac{2(1-2\delta_0)^{k-1}}{\delta_0}\right)-1\right) \\
&<(1-2\delta_0)^k+k\cdot\left(\exp\left(\tfrac{2(1-2\delta_0)^{k-1}}{\delta_0}\right)-1\right) \\
&<\exp\left((1-2\delta_0)^k\right)-1+k\cdot\left(\exp\left(\tfrac{2(1-2\delta_0)^{k-1}}{\delta_0}\right)-1\right) \\
&<(k+1)\cdot\left(\exp\left(\tfrac{2(1-2\delta_0)^{k-1}}{\delta_0}\right)-1\right)\ .
\end{aligned} \tag{B.1}$$

Suppose that k is chosen so that $\frac{2(1-2\delta_0)^{k-1}}{\delta_0} \le \ln 2$, or equivalently

$$k \ge \frac{\ln\left(\frac{2}{\delta_0(1-2\delta_0)\ln 2}\right)}{\ln\left(\frac{1}{1-2\delta_0}\right)}. \tag{B.2}$$

Since $\exp(x) \le 1+\frac{x}{\ln 2}$ for $0 \le x \le \ln 2$, then $\exp\left(\frac{2(1-2\delta_0)^{k-1}}{\delta_0}\right)-1 \le \frac{2(1-2\delta_0)^{k-1}}{\delta_0 \ln 2}$ under the requirement in (B.2).
Let a be a parameter so that $0<a<1$, and suppose that k also satisfies the inequality

$$k+1 \le \left(\frac{2(1-2\delta_0)^{k-1}}{\delta_0 \ln 2}\right)^{-a}.$$

By taking the logarithms of both sides of this inequality, and utilizing the inequality $\ln(k+1) < \sqrt{k}$ for $k \ge 1$, one can impose the following stronger requirement on k

$$\sqrt{k} \le ak\ \ln\left(\frac{1}{1-2\delta_0}\right) - a\ \ln\left(\frac{2}{\delta_0(1-2\delta_0)\ln 2}\right)$$

which is satisfied if

$$k \ge \left(\sqrt{\frac{\ln\left(\frac{2}{\delta_0(1-2\delta_0)\ln 2}\right)}{\ln\left(\frac{1}{1-2\delta_0}\right)}+\left(\frac{1}{2a\ln\left(\frac{1}{1-2\delta_0}\right)}\right)^2}+\frac{1}{2a\ln\left(\frac{1}{1-2\delta_0}\right)}\right)^2. \tag{B.3}$$

It is easy to see that that the requirement on k in (B.3) yields the requirement in (B.2), and therefore one can ignore the condition on k in (B.2). The requirement (B.3) on k yields that

$$\begin{aligned}
&(k+1)\cdot\left(\exp\left(\tfrac{2(1-2\delta_0)^{k-1}}{\delta_0}\right)-1\right) \\
&\overset{(a)}{\le}\left(\tfrac{2(1-2\delta_0)^{k-1}}{\delta_0\ln 2}\right)^{-a}\cdot\left(\exp\left(\tfrac{2(1-2\delta_0)^{k-1}}{\delta_0}\right)-1\right) \\
&\overset{(b)}{\le}\left(\exp\left(\tfrac{2(1-2\delta_0)^{k-1}}{\delta_0}\right)-1\right)^{1-a}
\end{aligned} \tag{B.4}$$

where inequality (a) relies on the requirement on k which led to the derivation of (B.3), and inequality (b) is based on the inequality $x \geq \ln 2 \cdot (\exp(x) - 1)$ for $0 \leq x \leq \ln 2$, and by taking $x = \frac{2(1-2\delta_0)^{k-1}}{\delta_0}$ (which is not above $\ln 2$ under the requirement in (B.3)). We note that inequality (b) also relies on the fact that $a > 0$, which therefore yields from (B.3) that

$$k+1 \leq \left(\frac{2(1-2\delta_0)^{k-1}}{\delta_0 \ln 2}\right)^{-a} \leq \left(\exp\left(\frac{2(1-2\delta_0)^{k-1}}{\delta_0}\right) - 1\right)^{-a}.$$

Under the requirement on k in (B.3), then the combination of inequalities (B.1) and (B.4) yields that inequality (62) can be replaced by the stronger requirement

$$\left(\exp\left(\tfrac{2(1-2\delta_0)^{k-1}}{\delta_0}\right) - 1\right)^{1-a} \leq \varepsilon C \ln 2,$$

which is equivalent (since $0 < a < 1$) to the inequality

$$(1-2\delta_0)^k \leq \tfrac{\delta_0(1-2\delta_0)}{2} \cdot \ln\left(1 + (\varepsilon C \ln 2)^{\frac{1}{1-a}}\right). \tag{B.5}$$

Since the channel capacity (C) is not above one bit per channel use (as it is a binary-input channel) and $0 < \varepsilon < 1$, then based on the inequality $\ln(1+x) > \ln 2 \cdot x$ for $0 < x < 1$, one can replace inequality (B.5) by the following stronger requirement on k

$$(1-2\delta_0)^k \leq \left(\tfrac{\delta_0(1-2\delta_0)\ln 2}{2}\right) \cdot (\varepsilon C \ln 2)^{\frac{1}{1-a}}$$

which is satisfied if

$$k \geq A \ \ln\left(\frac{1}{\varepsilon}\right) + B \tag{B.6}$$

where

$$A = \tfrac{1}{(1-a)\ln\left(\frac{1}{1-2\delta_0}\right)}, \qquad B = \tfrac{\ln\left(\frac{2}{\delta_0(1-2\delta_0)\ln 2}\left(\frac{1}{C\ln 2}\right)^{\frac{1}{1-a}}\right)}{\ln\left(\frac{1}{1-2\delta_0}\right)}. \tag{B.7}$$

To conclude, for the fulfillment of inequality (62), it is sufficient to determine k as the minimal integer which satisfies the two conditions in (B.3) and (B.6). This suggests a closed form expression of k in terms of ε and δ_0 which satisfies inequality (62), so that k behaves like $\ln \frac{1}{\varepsilon}$ (since δ_0 was appropriately determined in subsection 3.2.2 to be a positive number which does not depend on ε).

A sufficient condition for the fulfillment of inequality (63): Under the requirement in (B.3), it follows immediately from (B.1) and (B.4) (with the slight difference that we do not upper bound $1-R$ in (B.1) by unity) that inequality (63) can be replaced by the stronger requirement

$$(1-R) \cdot \left(\exp\left(\frac{2(1-2\delta_0)^{k-1}}{\delta_0}\right) - 1\right)^{1-a} < (1-R)\ln 2 - h_e(\delta_0),$$

where $0 < a < 1$ is arbitrary, and the requirement on k in Eq. (B.3) depends on a.
Since $R < C$, the latter inequality could be replaced by the following inequality which imposes a stronger requirement on k

$$\left(\exp\left(\frac{2(1-2\delta_0)^{k-1}}{\delta_0}\right) - 1\right)^{1-a} < (1-C)\ln 2 - h_e(\delta_0) \tag{B.8}$$

whose solution is

$$k > \frac{\ln\left(\frac{2}{\delta_0(1-2\delta_0)} \cdot \frac{1}{\ln\left[1+\left((1-C)\ln 2 - h_e(\delta_0)\right)^{\frac{1}{1-a}}\right]}\right)}{\ln\left(\frac{1}{1-2\delta_0}\right)}. \tag{B.9}$$

Appendix C
Derivation of inequality (67)

For $0 < \alpha < 1$ and $N \geq 2$, Shokrollahi has derived the equality (see [25], Eq. (7))

$$\ln\left((-1)^{N+1}\binom{\alpha}{N}\right) = \ln\left(\frac{\alpha}{N}\right) - \alpha\sum_{k=1}^{N-1}\frac{1}{k} - \frac{\alpha^2}{2}\sum_{k=1}^{N-1}\frac{1}{k^2} - \frac{\alpha^3}{3}\sum_{k=1}^{N-1}\frac{1}{k^3} - \cdots, \quad 0 < \alpha < 1. \tag{C.1}$$

For the derivation of an improved lower bound on $(-1)^{N+1}\binom{\alpha}{N}$ (as compared to the one which was derived in [25]), we rely on the equality $\sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{\pi^2}{6}$, and thus for $p \geq 2$ and $N \geq 2$

$$\sum_{k=1}^{N-1}\frac{1}{k^p} < \frac{\pi^2}{6}. \tag{C.2}$$

From Eqs. (C.1) and (C.2), one obtains the inequality

$$\begin{aligned}\ln\left((-1)^{N+1}\tbinom{\alpha}{N}\right) &> \ln\left(\tfrac{\alpha}{N}\right) - \alpha\textstyle\sum_{k=1}^{N-1}\frac{1}{k} - \frac{\pi^2}{6}\cdot\sum_{k=2}^{\infty}\frac{\alpha^k}{k} \\ &= \ln\left(\tfrac{\alpha}{N}\right) - \alpha\textstyle\sum_{k=1}^{N-1}\frac{1}{k} + \frac{\pi^2}{6}\cdot\left[\ln(1-\alpha)+\alpha\right], \quad 0 < \alpha < 1,\ N \geq 2.\end{aligned} \tag{C.3}$$

By the exponentiation of both sides of inequality (C.3), and based on the inequality

$$\ln(N) + \gamma < \sum_{k=1}^{N}\frac{1}{k} < \ln(N) + \gamma + \frac{1}{2N} \tag{C.4}$$

where γ designates Euler's constant, one obtains that for $0 < \alpha < 1$ and $N \geq 2$

$$\begin{aligned}(-1)^{N+1}\tbinom{\alpha}{N} &> \tfrac{\alpha}{N}\cdot(1-\alpha)^{\frac{\pi^2}{6}}\cdot\exp\left(\tfrac{\alpha\pi^2}{6}\right)\cdot\exp\left\{-\alpha\left(\textstyle\sum_{k=1}^{N}\frac{1}{k} - \frac{1}{N}\right)\right\} \\ &> \tfrac{\alpha}{N}\cdot(1-\alpha)^{\frac{\pi^2}{6}}\cdot\exp\left(\tfrac{\alpha\pi^2}{6}\right)\cdot\exp\left\{-\alpha\left(\ln(N)+\gamma+\tfrac{1}{2N}-\tfrac{1}{N}\right)\right\} \\ &= \tfrac{\alpha}{N^{\alpha+1}}\cdot(1-\alpha)^{\frac{\pi^2}{6}}\cdot\exp\left(\alpha\left(\tfrac{\pi^2}{6}-\gamma+\tfrac{1}{2N}\right)\right) \\ &= \tfrac{\alpha\cdot c(\alpha,N)}{N^{\alpha+1}}\end{aligned}$$

where $c(\cdot,\cdot)$ is defined in Eq. (67).

For an upper bound on $(-1)^{N+1}\binom{\alpha}{N}$, Shokrollahi [25] has lower bounded the sums $\sum_{k=1}^{N-1}\frac{1}{k^p}$ (where $N \geq 2$ and $p \geq 2$ are integers) by unity (i.e., by the first term of these series). Then, from (C.1), (C.4), and the inequality $(1-\alpha)\cdot\exp(\alpha) < 1$ for $0 < \alpha < 1$, the right-hand side of inequality (67) follows directly.

Acknowledgment

The authors are indebted to the anonymous reviewers for their detailed reports and for their many valuable comments which improved the lucidity of the presentation. They are also grateful to Emre Telatar for his comments on an earlier version of the paper, and to Ralf Koetter for kindly handling their paper. The authors wish to acknowledge interesting discussions with Tom Richardson and Amin Shokrollahi.

References

[1] O. Barak, D. Burshtein and M. Feder, "Bounds on achievable rates of LDPC codes used over the binary erasure channel," submitted to *IEEE Trans. on Information Theory*.

[2] A. Barg and G. Zémor, "Error exponents of expander codes," *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1725–1729, June 2002.

[3] E. R. Berlekamp, R. J. McEliece and H. C. A. van Tilborg, "On the inherent intractability of certain coding problems," *IEEE Trans. on Information Theory*, vol. 24, no. 3, pp. 384–386, May 1978.

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

[5] J. T. Coffey and A. B. Kiely, "The capacity of coded systems," *IEEE Trans. on Information Theory*, vol. 43, no. 1, pp. 113–127, January 1997.

[6] C. Di, D. Proietti, T. Richardson, E. Telatar and R. Urbanke, "Finite length analysis of low-density parity-check codes on the binary erasure channel," *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1570–1579, June 2002.

[7] T. Etzion, A. Trachtenberg and A. Vardy, "Which codes have cycle-free Tanner graphs ?," *IEEE Trans. on Information Theory*, vol. 45, no. 6, pp. 2173–2181, September 1999.

[8] R. G. Gallager, *Low-density parity-check codes*, Cambridge, MA, USA, MIT Press, 1963. [Online]. Available: http://lthiwww.epfl.ch/∼eeigal/Gallager_monograph.ps.

[9] J. Gross and J. Yellen, *Graph Theory and Its Applications*, CRC Press, 1999.

[10] H. Jin and R. J. McEliece, "Typical pairs decoding on the AWGN channel", *Proceedings 2000 International Symposium on Information Theory and Its Applications*, pp. 180–183, Honolulu, Hawaii, U.S.A., November 5–8, 2000.

[11] A. Khandekar and R. J. McEliece, "On the complexity of reliable communication on the erasure channel," *Proceedings 2001 IEEE International Symposium on Information Theory (ISIT2001)*, p. 1, Washington, D.C., USA, June 2001.

[12] A. Khandekar, *Graph-based codes and iterative decoding*, Ph.D. dissertation, California Institute of Technology, Pasadena, California, June 2002. [Online]. Available: http://etd.caltech.edu/etd/available/etd-06202002-170522/.

[13] M. Luby, M. Mitzenmacher, A. Shokrollahi and D. Spielman, "Efficient erasure correcting codes," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 569–584, February 2001.

[14] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. on Information Theory*, vol. 45, no. 2, pp. 399–431, March 1999.

[15] S. J. MacMullan and O. M. Collins, "The capacity of binary channels that use linear codes and decoders," *IEEE Trans. on Information Theory*, vol. 44, no. 1, pp. 197–214, January 1998.

[16] R. J. McEliece, "Achieving the Shannon limit: A progress report," plenary talk given at the *38th Annual Allerton Conference on Communication, Control and Computing*, Allerton, Illinois, USA, October 5, 2000. [Online]. Available: http://www.systems.caltech.edu/EE/Faculty/rjm.

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

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

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

[20] T. Richardson, A. Shokrollahi and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 619–637, February 2001.

[21] T. Richardson and R. Urbanke, "Efficient encoding of low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 638–656, February 2001.

[22] T. Richardson, A. Shokrollahi and R. Urbanke, "Error floor analysis of various low-density parity-check ensembles for the binary erasure channel," *in preparation.*

[23] 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, no. 3, pp. 89–91, March 2000.

[24] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections and applications," *IEEE Trans. on Information Theory*, vol. 48, no. 12, pp. 3029–3051, December 2002.

[25] A. Shokrollahi, "New sequences of time erasure codes approaching channel capacity," *Proceedings of the 13th International Symposium on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, Lectures Notes in Computer Science 1719, Springer Verlag, pp. 65–76, 1999.

[26] A. Shokrollahi, "Capacity-achieving sequences," in *IMA Volumes in Mathematics and its Applications*, vol. 123, pp. 153–166, 2000.

[27] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes," *IEEE Trans. on Information Theory*, vol. 45, no. 6, pp. 2101–2104, September 1999.

[28] The Lambert W-function. [Online]. Available: http://mathworld.wolfram.com/LambertsW-Function.html.

Igal Sason (S'98–M'02) was born in Haifa, Israel, on May 1969. He received the B.Sc. and Ph.D. degrees in electrical engineering from the Technion–Israel Institute of Technology, Haifa, Israel, in 1992 and 2001, respectively.
During 1993–1998, he served in the army as an electrical engineer. During 2001–2003, he has been a Scientific Collaborator in the Faculty of Computer Science and Communications, EPFL– Swiss Federal Institute of Technology, Lausanne, Switzerland. On October 2003, he will join the Department of Electrical Engineering at the Technion as a Senior Lecturer.
Dr. Sason is a co-recipient of the 2003 IEEE Information Theory Society and Communications Society joint paper award. His research interests include information theory and coding theory. He is especially interested in codes on graphs and iterative decoding algorithms, performance bounds of linear codes, and the tradeoff between their performance and complexity.

Rüdiger Urbanke received the Diplomingenieur degree from the Vienna Institute of Technology, Vienna, Austria, in 1990 and the M.S. and Ph.D. degrees in electrical engineering from Washington University, St.-Louis, MO, in 1992 and 1995 respectively.
From 1995 to 1999, he held a position at the Mathematics of Communications Department at Bell Labs. Since November 1999, he has been a Professor at the faculty of Computer Science and Communications in the Swiss Federal Institute of Technology, Lausanne, Switzerland.
Dr. Urbanke is a recipient of a Fulbright Scholarship and since October 2000, has been an Associate Editor of the IEEE Transactions on Information Theory. He is a co-recipient of the 2002 IEEE Information Theory Best Paper Award.