On Universal Properties of Capacity-Approaching LDPC Ensembles

Igal Sason, *Member*

CCIT Report # 638 September 2007

Abstract

This paper provides some universal information-theoretic bounds related to capacity-approaching ensembles of low-density parity-check (LDPC) codes. These bounds refer to the behavior of the degree distributions of such ensembles, and also to the graphical complexity and the fundamental system of cycles associated with the Tanner graphs of LDPC ensembles. The transmission of these ensembles is assumed to take place over an arbitrary memoryless binary-input output-symmetric (MBIOS) channel. The universality of the bounds derived in this paper stems from the fact that they do not depend on the full characterization of the LDPC ensembles but rather depend on the achievable gap between the channel capacity and the design rate of the ensemble, and also on the required bit error (or erasure) probability at the end of the decoding process. Some of these bounds hold under maximum-likelihood decoding (and hence, they also hold under any sub-optimal decoding algorithm) whereas the others hold particularly under the sum-product iterative decoding algorithm. The tightness of some of these bounds is exemplified numerically for capacity-approaching LDPC ensembles under sum-product decoding; the bounds are reasonably tight for general MBIOS channels, and are tightened for the binary erasure channel.

Index Terms

Cycles, low-density parity-check (LDPC) codes, maximum-likelihood decoding, memoryless binary-input output-symmetric (MBIOS) channels, stability condition, sum-product decoding algorithm, Tanner graphs.

I. Introduction

Low-density parity-check (LDPC) codes were introduced by Gallager [7] in the early 1960s. These linear block codes are characterized by sparse parity-check matrices which facilitate their efficient decoding with sub-optimal iterative message-passing algorithms. In spite of the seminal work of Gallager [7], LDPC codes were ignored for a long time. Following the breakthrough in coding theory made by the introduction of turbo codes [2] and the rediscovery of LDPC codes [13] in the mid 1990s, it was realized that these codes and lots of other variants of capacity-approaching error-correcting codes can all be understood as codes defined on graphs. Graphs not only describe the codes, but more importantly, they structure the operation of efficient iterative decoding algorithms which are used to decode these codes. Various iterative algorithms, used to decode codes defined on graphs, enable to closely approach the channel capacity while maintaining reasonable decoding complexity. This breakthrough attracted coding theorists, and lots of research activity has been conducted during the last decade on these modern coding techniques and their practical decoding algorithms; the reader is referred to the special issue of the *IEEE Transactions on Information Theory* on codes on graphs and iterative algorithms [26].

This paper derives some universal information-theoretic bounds related to capacity-approaching LDPC ensembles whose transmission takes place over a memoryless binary-input output-symmetric (MBIOS) channel. The universality of the bounds derived in this paper stems from the fact that they do not depend on the full characterization of the LDPC ensembles but rather depend on the achievable gap between the channel capacity and the design rate of the ensemble, and also on the required bit error (or erasure) probability at the end of the decoding process. Most of these bounds apply to the asymptotic case where we let the block length tend to infinity, and the bounds are compared with some capacity-approaching LDPC ensembles. The design of such ensembles under iterative decoding lies on a solid background due to the *density evolution* technique which was developed by Richardson and Urbanke (see [10], [18], [19]). This technique is commonly used for a numerical search of the degree distributions of capacity-approaching LDPC ensembles where the target is to minimize the gap to capacity for infinite block length while

The paper was submitted in September 2007 as a regular paper to the *IEEE Trans. on Information Theory*. This research work was supported by the Israel Science Foundation (grant no. 1070/07).

Igal Sason is with the Department of Electrical Engineering at the Technion – Israel Institute of Technology, Haifa 32000, Israel (e-mail: sason@ee.technion.ac.il).

limiting the maximal degree and specifying the communication channel model. Some approximate techniques which optimize the degree distributions of LDPC ensembles under further practical constraints (e.g., an optimization of the degree distributions of LDPC ensembles for obtaining a good tradeoff between the asymptotic gap to capacity and the decoding complexity [1]) are of interest. For the binary erasure channel (BEC), the density evolution technique is tractable for analysis (as it becomes a one-dimensional analysis), and explicit expressions for capacity-achieving sequences of LDPC ensembles for the BEC are introduced in [12], [17], [24]. For general MBIOS channels, as of yet there are no closed-form expressions for capacity-achieving sequences of LDPC ensembles under iterative decoding, and the density evolution technique is used as a numerical tool for devising the degree distributions of capacity-approaching LDPC ensembles in the limit where their block length tends to infinity.

It is well known that linear block codes which are represented by cycle-free Tanner graphs have poor performance even under ML decoding [6]. The Tanner graphs of capacity-approaching LDPC codes should have cycles, and it is of interest to provide some quantitative bounds related to the average cardinality of these cycles for LDPC ensembles. Following the approach of Khandekar and McEliece in [11] where the decoding complexity is measured in terms of the achievable gap (in rate) to capacity, we also study in this paper the behavior of the degree distributions and the average cardinality of the fundamental system of cycles of LDPC ensembles in terms of their achievable gap to capacity. Several of the results derived in this paper hold under ML decoding (and, hence, they also hold under any sub-optimal decoding algorithm), while some other results are specialized to the iterative sum-product decoding algorithm.

This paper is structured as follows: Section II presents preliminary background and notation, Section III introduces the main results of this paper, Section IV provides their proofs followed by some discussions, and Section V exemplifies the numerical tightness of some of the bounds derived in this paper. Finally, Section VI summarizes this paper, and provides some open problems which are related to this research.

II. Preliminaries

We introduce here some background from [8], [21], [25] and [28] and notation which serve for the rest of this paper.

A. LDPC Ensembles

LDPC codes are linear block codes which are characterized by sparse parity-check matrices. Alternatively, a parity-check matrix can be represented by a bipartite (Tanner) graph where the variable and parity-check nodes which specify the binary linear block code are on the left and the right of this graph, respectively, and an edge connects between a variable node and a parity-check node if the corresponding code symbol is involved in the specific parity-check equation. This is illustrated in Fig. 1.

$$H := \begin{array}{c} \\ 1 \\ \mathbf{2} \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & \mathbf{8} & 9 & 10 \end{array} \\ \left(\begin{array}{cccccccccc} 1 & 1 & 1 & 1 & 0 & 1 & 1 & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & \mathbf{1} & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & \mathbf{1} & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & \mathbf{0} & 1 & 1 \end{array} \right) \end{array}$$

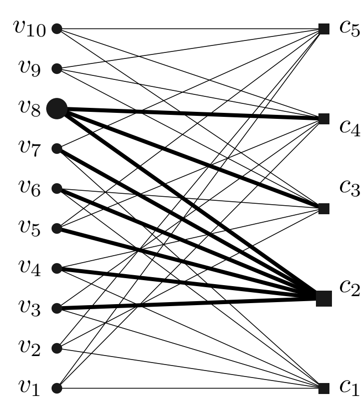

Fig. 1. A parity-check matrix H and the corresponding Tanner graph. For illustrating this relationship, column 8 and row 2 of H are emphasized; the corresponding variable and parity-check nodes, and the attached edges are also emphasized (this figure appears in [20], and it is used later in this section as a reference).

We now move to consider ensembles of LDPC codes. The requirement of the sparseness of the parity-check matrices representing LDPC codes is transformed to an equivalent requirement where the number of edges in the

Tanner graph scales linearly with the block length (note that by picking a parity-check matrix uniformly at random as a representative of a binary linear block codes, the number of non-zero elements in this matrix is likely to scale quadratically with the block length). Following standard notation in [21], let λ_i and ρ_i denote the fraction of *edges* attached to variable and parity-check nodes of degree i, respectively. In a similar manner, let Λ_i and Γ_i denote the fraction of variable and parity-check nodes of degree i, respectively. The LDPC ensemble is characterized by a triplet (n, λ, ρ) where n designates the block length of the codes, and the power series

$$\lambda(x) \triangleq \sum_{i=1}^{\infty} \lambda_i x^{i-1}, \qquad \rho(x) \triangleq \sum_{i=1}^{\infty} \rho_i x^{i-1}$$

represent, respectively, the left and right degree distributions from the *edge* perspective. Equivalently, this ensemble is also characterized by the triplet (n, Λ, Γ) where the power series

$$\Lambda(x) \triangleq \sum_{i=1}^{\infty} \Lambda_i x^{i}, \qquad \Gamma(x) \triangleq \sum_{i=1}^{\infty} \Gamma_i x^{i}$$

represent, respectively, the left and right degree distributions from the *node* perspective. We denote by LDPC(n, λ, ρ) (or LDPC(n, Λ, Γ)) the ensemble of codes whose bipartite graphs are constructed according to the corresponding pairs of degree distributions. The connections between the edges emanating from the variable nodes to the parity-check nodes are constructed by numbering the connectors on the left and the right of the graph (whose number is the same on both sides, and is equal to $s = n \sum_{i=1}^{\infty} i \Lambda_i$), and by using a random and uniform permutation $\pi : \{1, \ldots, s\} \rightarrow \{1, \ldots, s\}$ which associates connector number i (where $1 \leq i \leq s$) on the left side of this graph with the connector whose number is $\pi(i)$ on the right. One can switch between degree distributions w.r.t. to the nodes and edges of a bipartite graph, using the following equations:

$$\Lambda(x) = \frac{\int_0^x \lambda(u) du}{\int_0^1 \lambda(u) du}, \qquad \Gamma(x) = \frac{\int_0^x \rho(u) du}{\int_0^1 \rho(u) du}$$

$$\lambda(x) = \frac{\Lambda'(x)}{\Lambda'(1)}, \qquad \rho(x) = \frac{\Gamma'(x)}{\Gamma'(1)}.$$

An important characteristic of an ensemble of LPDC codes is its *design rate*. For an LDPC ensemble whose codes are represented by parity-check matrices of dimension $c \times n$, the design rate is defined to be $R_{\text{d}} \triangleq 1 - \frac{c}{n}$. This serves as a lower bound on the actual rate of any code from this ensemble, and the rate of such a code is equal to the design rate if the particular parity-check matrix representing this code is full rank. For an ensemble of LDPC codes, the design rate is given in terms of the degree distributions (either w.r.t. the edges or nodes of a Tanner graph), and it can be expressed in two equivalent forms:

$$R_{\text{d}} = 1 - \frac{\int_0^1 \rho(x) dx}{\int_0^1 \lambda(x) dx} = 1 - \frac{\Lambda'(1)}{\Gamma'(1)}. \tag{1}$$

Note that

$$a_{\text{L}} = \Lambda'(1) = \frac{1}{\int_0^1 \lambda(x) dx}, \qquad a_{\text{R}} = \Gamma'(1) = \frac{1}{\int_0^1 \rho(x) dx}$$

designate the average left and right degrees, respectively (i.e., the average degrees of the variable nodes and parity-check nodes, respectively).

In this paper, we rely on the *stability condition* which forms a necessary condition for a successful iterative message-passing decoding for LDPC ensembles. The reader is referred to [21, Chapter 4] for more background on the analytical tools used for the asymptotic analysis of LDPC ensembles over MBIOS channels, and the stability condition in particular.

B. Elements from Graph Theory

Definition 1: **[Tree]** A *tree* is a connected graph that has no cycles.
From Definition 1, trees are the smallest connected graphs; remove any edge from a tree and it becomes disconnected. A fundamental property of trees is that any two vertices are connected by a *unique path*.

Every graph $\mathcal{G}$ has subgraphs that are trees. This motivates the following definition:

Definition 2: **[Spanning tree]** A *spanning tree* of a graph $\mathcal{G}$ is a tree which spans all the vertices of $\mathcal{G}$.

Lemma 1: Every connected graph has a spanning tree.

Proof: Let $\mathcal{G}$ be a connected graph. If $\mathcal{G}$ has no cycles, then it is a spanning tree. Otherwise, choose a cycle $\mathcal{S}$ in the graph, and remove an arbitrary edge e which belongs to this cycle. The remaining graph is still connected since any path which uses the removed edge e can now be replaced by a path using $\mathcal{S} - \{e\}$, so that every two vertices of the graph $\mathcal{G}$ are still connected. One can repeat this process as many times as required until the resulting graph has no cycles. By construction, it is a spanning tree of $\mathcal{G}$. ■

Definition 3: **[Fundamental cycle of a connected graph]** Let $\mathcal{F}$ be a spanning tree of a graph $\mathcal{G}$, and let e be any edge in the relative complement of $\mathcal{F}$. The cycle of the subgraph $\mathcal{F} \cup \{e\}$ (whose existence and uniqueness is guaranteed by [8, Theorem 3.1.11]) is called a fundamental cycle of $\mathcal{G}$ which is associated with the spanning tree $\mathcal{F}$.

We turn now to discuss graphs with a finite number of components.

Definition 4: **[Number of components of a graph]** Let $\mathcal{G}$ be a graph (possibly disconnected). The *number of components* of $\mathcal{G}$ is the minimal number of its connected subgraphs whose union forms the graph $\mathcal{G}$ (clearly, a connected graph has a unique component).

Definition 5: **[Cycle rank]** Let $\mathcal{G}$ be an arbitrary graph with $|V_\mathcal{G}|$ vertices, $|E_\mathcal{G}|$ edges and $C(\mathcal{G})$ components. The *cycle rank* of $\mathcal{G}$, denoted by $\beta(\mathcal{G})$, equals the maximal number of edges which can be removed from the graph without increasing its number of components.

From Definition 5, the cycle rank of a graph is a measure of the edge redundancy with respect to the connectedness of this graph. The cycle rank satisfies the following equality (see [8, p. 154]):

$$\beta(\mathcal{G}) = |E_\mathcal{G}| - |V_\mathcal{G}| + C(\mathcal{G}). \tag{2}$$

Definition 6: **[Full spanning forest]** Let $\mathcal{G}$ be an arbitrary graph. The *full spanning forest* $\mathcal{F}$ of the graph $\mathcal{G}$ is the subgraph of $\mathcal{G}$ after removing the $\beta(\mathcal{G})$ edges from Definition 5. Clearly, the number of components of $\mathcal{F}$ and $\mathcal{G}$ is the same.

Definition 7: **[Fundamental cycle]** Let $\mathcal{F}$ be a full spanning forest of a graph $\mathcal{G}$, and let e be any edge in the relative complement of $\mathcal{F}$. The cycle of the subgraph $\mathcal{F} \cup \{e\}$ (whose existence and uniqueness is guaranteed by [8, Theorem 3.1.11]) is called a fundamental cycle of $\mathcal{G}$ which is associated with $\mathcal{F}$.

Remark 1: Each of the edges in the relative complement of a full spanning forest $\mathcal{F}$ gives rise to a *different* fundamental cycle of the graph $\mathcal{G}$.

Definition 8: **[Fundamental system of cycles]** The *fundamental system of cycles* of a graph $\mathcal{G}$ which is associated with a full spanning forest $\mathcal{F}$ is the set of all fundamental cycles of $\mathcal{G}$ associated with $\mathcal{F}$.

Remark 2: By Remark 1, the cardinality of the fundamental system of cycles of $\mathcal{G}$ associated with a full spanning forest of this graph is equal to the cycle rank $\beta(\mathcal{G})$.

Example 1: **[Illustration with the Tanner graph in Fig. 1]** We refer now to the Tanner graph in Fig. 1. This graph is connected, but it is clearly not a tree. As an example of a cycle in this graph, we refer to the path which connects v_9 to c_4, c_4 to v_{10}, v_{10} to c_5, and finally c_5 back to v_9; this cycle includes 4 edges, so it is a cycle of the shortest length. Since the number of vertices in this graph is 15 and the number of edges is 30, then from (2), the cycle rank of this connected Tanner graph is equal to $30 - 15 + 1 = 16$. In order to get a spanning tree of this graph, we follow the concept of the proof of Lemma 1, and remove sequentially 16 suitable edges from this graph so that it still remains connected. In the following, we show the parity-check matrix $\widetilde{H}$ which corresponds to the new Tanner graph after the removal of the above 16 edges from the original graph in Fig. 1; the new zero entries of $\widetilde{H}$ which correspond to these removed edges are bolded.

The matrix $\widetilde{H}$ with its bolded zeros defines a set of 16 fundamental cycles of the Tanner graph in Fig. 1. For example, by returning the edge which connects v_6 with c_1 (i.e., returning the value in the first row and sixth column of $\widetilde{H}$ to be 1), we get a fundamental cycle; this is the path which connects v_3 to c_2, c_2 to v_6, v_6 to c_1, and c_1 back

$$\widetilde{H} = \begin{array}{cc} & \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \left(\begin{array}{cccccccccc} 1 & 1 & 1 & \mathbf{0} & 0 & \mathbf{0} & \mathbf{0} & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & \mathbf{0} & 1 & 1 & 1 & 0 & 0 \\ 0 & \mathbf{0} & 0 & \mathbf{0} & 0 & \mathbf{0} & 0 & \mathbf{0} & 1 & \mathbf{0} \\ 1 & 0 & \mathbf{0} & 0 & 1 & 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 1 & \mathbf{0} & 0 & 0 & \mathbf{0} & 0 & \mathbf{0} & 0 & 1 & 1 \end{array} \right) \end{array}$$

Fig. 2. A parity-check matrix which corresponds to a spanning tree of the Tanner graph in Fig. 1. As compared to the parity-check matrix H in Fig. 1, the new parity-check matrix $\widetilde{H}$ is obtained by changing the values of the emphasized 16 entries from 1 to 0.

to v_3). This is clearly a fundamental cycle in this graph since it is a cycle of the shortest possible length. The new Tanner graph which corresponds to $\widetilde{H}$ is connected. For example, the variable nodes v_5 and v_6 are connected by the path of length 6 which connects v_6 to c_2, c_2 to v_3, v_3 to c_1, c_1 to v_1, v_1 to c_4 and c_4 to v_5. This path can be observed directly from the parity-check matrix $\widetilde{H} = [\widetilde{h}_{i,j}]$ by alternate horizontal and vertical moves through the ones of $\widetilde{H}$; explicitly, this path is determined by the horizontal move from $\widetilde{h}_{2,6}$ to $\widetilde{h}_{2,3}$, then the vertical move to $\widetilde{h}_{1,3}$, the horizontal move to $\widetilde{h}_{1,1}$, the vertical move to $\widetilde{h}_{4,1}$ and the horizontal move to $\widetilde{h}_{4,5}$. In a similar way one can show that each two vertices of the Tanner graph which corresponds to $\widetilde{H}$ are connected, and hence this graph is indeed a spanning tree of the Tanner graph in Fig. 1. Every single edge which is added to the new graph creates a fundamental cycle (like the one mentioned above).

C. Lower Bound on the Conditional Entropy for Binary Linear Block Codes

In this section, we outline the derivation of a lower bound on the conditional entropy of a transmitted codeword given the received sequence at the output of the channel; the full derivation is given in [25, Section 4] and its appendices, and it is outlined here in order to highlight the main steps used for this derivation. The lower bound on the conditional entropy which is presented in this section is crucial for the proof of Theorem 1 of this paper.

In the sequel to this discussion, we assume that the transmission of a binary linear block code takes place over an MBIOS channel. In the following, the block length and the code rate are designated by n and R, respectively. Let C designate the capacity of this communication channel in units of bits per channel use.

- Define an equivalent channel whose output is the log-likelihood ratio (LLR) of the original communication channel.
- The LLR is represented by a pair which includes its sign and absolute value.
- For the characterization of the equivalent channel, let the function a designate the conditional *pdf* of the LLR given that the channel input is the zero symbol.
- We randomly generate an i.i.d. sequence $\{L_i\}_{i=1}^n$ w.r.t. the conditional *pdf* a, and define

$$\Omega_i \triangleq |L_i|, \quad \Theta_i \triangleq \begin{cases} 0 & \text{if} \quad L_i > 0 \\ 1 & \text{if} \quad L_i < 0 \\ 0 \text{ or } 1 \text{ w.p. } \frac{1}{2} & \text{if} \quad L_i = 0 \end{cases}.$$

- The output of the equivalent channel is $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_n)$ where

  $$\widetilde{Y}_i = (\Phi_i, \Omega_i), \quad i = 1, \ldots, n$$

  and $\Phi_i = \Theta_i + X_i$ (this addition is modulo-2).
- The output of this equivalent channel at time i is therefore the pair (Φ_i, Ω_i) where $\Phi_i \in \{0, 1\}$ and $\Omega_i \in \mathbb{R}^+$. This defines the memoryless mapping

  $$X \to \widetilde{Y} \triangleq (\Phi, \Omega)$$

  where Φ is a binary random variable which is affected by X, and Ω is a non-negative random variable which is not affected by X.

- Due to the symmetry of the communication channel, the *pdf* of the absolute value of the LLR satisfies

$$f_\Omega(\omega) = \begin{cases} a(\omega) + a(-\omega) = (1 + e^{-\omega})\, a(\omega) & \text{if } \omega > 0, \\ a(0) & \text{if } \omega = 0. \end{cases}$$

Let $\mathcal{C}$ be a binary linear block code of length n and rate R. In addition, let $\mathbf{X}$ and $\mathbf{Y}$ be the transmitted codeword and received sequence, respectively. The conditional entropy of the transmitted codeword given the received sequence at the output of the MBIOS channel satisfies

$$\begin{aligned} H(\mathbf{X}|\mathbf{Y}) &= H(\mathbf{X}|\widetilde{\mathbf{Y}}) \\ &= H(\mathbf{X}) + H(\widetilde{\mathbf{Y}}|\mathbf{X}) - H(\widetilde{\mathbf{Y}}) \\ &= nR + nH(\widetilde{Y}_1|X_1) - H(\widetilde{\mathbf{Y}}) \\ &= nR + n[H(\widetilde{Y}_1) - I(X_1;\widetilde{Y}_1)] - H(\mathbf{Y}) \end{aligned} \tag{3}$$

and

$$I(X_1;\widetilde{Y}_1) = I(X_1;Y_1) \le C \tag{4}$$

$$\begin{aligned} H(\widetilde{Y}) &= H(\Phi, \Omega) \\ &= H(\Omega) + H(\Phi|\Omega) \\ &= H(\Omega) + 1. \end{aligned} \tag{5}$$

The last transition in (5) is due to the fact that given the absolute value of the LLR, its sign is equally likely to be positive or negative. The entropy $H(\Omega)$ is not expressed explicitly as it will cancel out later.

The entropy of the vector $\widetilde{\mathbf{Y}}$ satisfies

$$\begin{aligned} H(\widetilde{\mathbf{Y}}) &= H\big(\Phi_1, \Omega_1, \ldots, \Phi_n, \Omega_n\big) \\ &= H(\Omega_1, \ldots, \Omega_n) + H\big(\Phi_1, \ldots, \Phi_n \mid \Omega_1, \ldots, \Omega_n\big) \\ &= nH(\Omega) + H\big(\Phi_1, \ldots, \Phi_n \mid \Omega_1, \ldots, \Omega_n\big). \end{aligned} \tag{6}$$

- Define the syndrome vector $\mathbf{S} = (\Phi_1, \ldots, \Phi_n)H^T$ where H is an arbitrary full-rank parity-check matrix of $\mathcal{C}$.
- Let M be the index of the vector $(\Phi_1, \ldots, \Phi_n)$ in the coset.
- $H(M) = nR$ since all the codewords are transmitted with equal probability, and we get

$$\begin{aligned} & H\big((\Phi_1, \ldots, \Phi_n) \,|\, (\Omega_1, \ldots, \Omega_n)\big) \\ & = H\big(\mathbf{S}, M \,|\, (\Omega_1, \ldots, \Omega_n)\big) \\ & \le H(M) + H\big(\mathbf{S} \,|\, (\Omega_1, \ldots, \Omega_n)\big) \\ & \le nR + \sum_{j=1}^{n(1-R)} H\big(S_j \,|\, (\Omega_1, \ldots, \Omega_n)\big)\,. \end{aligned} \tag{7}$$

- Since $\mathbf{X}H^T = \mathbf{0}$ for any codeword $\mathbf{X}$, and $\Phi_i = X_i + \Theta_i$ for all i, then $\mathbf{S} = (\Theta_1, \ldots, \Theta_n)H^T$ which is independent of the transmitted codeword.

By combining (3)–(7), one obtains that

$$H(\mathbf{X}|\mathbf{Y}) \ge n(1-C) - \sum_{j=1}^{n(1-R)} H(S_j \big| \Omega_1, \ldots, \Omega_n)$$

where

- $S_j = 1$ if and only if $\Theta_i = 1$ for an odd number of i's in the j'th parity-check equation.
- Due to the symmetry of the channel

$$P(\alpha_i) \triangleq P(\Theta_i = 1 \big| \Omega_i = \alpha_i) = \frac{1}{1 + e^{\alpha_i}}.$$

In order to calculate the conditional entropy of a single component of the syndrome, the following lemma is used:

Lemma 2: If the j'th component of the syndrome $\mathbf{S}$ involves k active variables whose indices are $\{i_1, \ldots, i_k\}$ then

$$P(S_j = 1 \big| \Omega_{i_1} = \alpha_1, \ldots, \Omega_{i_k} = \alpha_k) = \frac{1}{2}\Big[1 - \prod_{m=1}^{k} \big(1 - 2P(\alpha_m)\big)\Big]$$

where

$$1 - 2P(\alpha) = \tanh\left(\frac{\alpha}{2}\right).$$

- For a parity-check equation of degree k, the conditional entropy $H(S_j \big| \Omega_1, \ldots, \Omega_n)$ is given by the following k-dimensional integral:

$$\int_0^\infty \ldots \int_0^\infty h_2\left(\frac{1}{2}\Big[1 - \prod_{m=1}^{k} \tanh(\frac{\alpha_m}{2})\Big]\right) \prod_{m=1}^{k} f_\Omega(\alpha_m)\, d\alpha_1 \ldots d\alpha_k$$

  where f_Ω designates the *pdf* of the absolute value of the LLR, and h_2 designates the binary entropy function to the base 2.
- Using the Taylor series expansion of h_2 (around one-half) transforms the above multi-dimensional integral to a one-dimensional integral raised to the k'th power.

For an arbitrary full-rank parity-check matrix of a binary linear block code $\mathcal{C}$, let Γ_i designate the fraction of the parity-checks involving i variables, and let $\Gamma(x) \triangleq \sum_i \Gamma_i x^i$. This leads in [25, Eq. (56)] to the following lower bound on the conditional entropy of the transmitted codeword given the received sequence at the channel output:

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq R - C + \frac{1-R}{2\ln 2} \sum_{k=1}^{\infty} \frac{\Gamma(g_k)}{k(2k-1)} \tag{8}$$

where

$$g_k \triangleq \int_0^\infty a(l)(1+e^{-l}) \tanh^{2k}\left(\frac{l}{2}\right) dl, \quad k \in \mathbb{N}. \tag{9}$$

The above lower bound on the conditional entropy holds for any representation of the code by a full-rank parity-check matrix. Note also that the symmetry condition states that $a(l) = e^l a(-l)$ for all $l \in \mathbb{R}$, and therefore (9) gives that

$$g_k = \mathbb{E}\left[\tanh^{2k}\left(\frac{L}{2}\right)\right], \quad k \in \mathbb{N} \tag{10}$$

where $\mathbb{E}$ designates the statistical expectation, and L is a random variable which stands for the LLR at the output of the channel given that the input bit is zero. This also implies that the sequence $\{g_k\}$ depends only on the communication channel (and not on the code).

We note that for a general MBIOS channel, the lower bound on the conditional entropy which is given in (8) is tighter than the bound in [3, Eq. (15)]. These two bounds coincide however for the case of a binary symmetric channel (BSC) since the derivation of the bound in (8) relies on the soft output of the channel whereas the derivation of the bound in [3, Eq. (15)] relies on a two-level quantization of this output (which turns the side information about the LLR at the output of the MBIOS channel to be equivalent to the side information which is obtained from a degraded BSC, unless this channel was originally a BSC where the two bounds then coincide).

D. Notation

We rely on the following standard notation (see [28]):

- $f(n) = O\big(g(n)\big)$ means that there are positive constants c and k, such that $0 \leq f(n) \leq c\, g(n)$ for all $n \geq k$. The values of c and k must be fixed for the function f and should not depend on n.
- $f(n) = \Omega\big(g(n)\big)$ means that there are positive constants c and k, such that $0 \leq c\, g(n) \leq f(n)$ for all $n \geq k$. The values of c and k must be fixed for the function f and should not depend on n.

In this paper, we refer to the gap (in rate) to capacity, denoted by ε, and discuss in particular the case where $0 \leq \varepsilon \ll 1$ (i.e., capacity-approaching ensembles). Accordingly, we have

- $f(\varepsilon) = O\big(g(\varepsilon)\big)$ means that there are positive constants c and δ, such that $0 \leq f(\varepsilon) \leq c\, g(\varepsilon)$ for all $0 \leq \varepsilon \leq \delta$. The values of c and δ must be fixed for the function f and should not depend on ε.
- $f(\varepsilon) = \Omega\big(g(\varepsilon)\big)$ means that there are positive constants c and δ, such that $0 \leq c\, g(\varepsilon) \leq f(\varepsilon)$ for all $0 \leq \varepsilon \leq \delta$. The values of c and δ must be fixed for the function f and should not depend on ε.

Throughout the paper

$$h_2(x) \triangleq -x \log_2(x) - (1-x)\log_2(1-x), \quad 0 \leq x \leq 1$$

designates the binary entropy function to the base 2, and $h_2^{-1} : [0,1] \rightarrow \big[0, \frac{1}{2}\big]$ stands for its inverse function.

III. Main Results

The following theorem provides an information-theoretic lower bound on the average degree of the parity-check nodes of an arbitrary Tanner graph representing a binary linear block code; it is assumed that the graph corresponds to a full-rank parity-check matrix of this code. The new bound forms a tightened version of a previously reported lower bound (see [25, Eq. (77)]). We later generalize this theorem for LDPC ensembles.

Theorem 1: **[On the average degree of the parity-check nodes]** Let $\mathcal{C}$ be a binary linear block code whose transmission takes place over an MBIOS channel. Let $\mathcal{G}$ be a standard bipartite (Tanner) graph which represents the code while referring to a full-rank parity-check matrix of the code, let C designate the channel capacity in units of bits per channel use, and a be the conditional *pdf* of the log-likelihood ratio (LLR) at the output of the channel given that the input is zero. Let the code rate be (at least) a fraction $1-\varepsilon$ of the channel capacity (where $\varepsilon \in (0,1)$ is arbitrary), and assume that this code achieves a bit error probability P_b under some decoding algorithm. Then the average right degree of the Tanner graph of the code (i.e., the average degree of the parity-check nodes in $\mathcal{G}$) satisfies

$$a_\mathrm{R} \geq \frac{2\ln\left(\frac{1}{1-2h_2^{-1}\left(\frac{1-C-h_2(P_\mathrm{b})}{1-(1-\varepsilon)C}\right)}\right)}{\ln\left(\frac{1}{g_1}\right)} \tag{11}$$

where g_1 only depends on the channel and is given by

$$g_1 \triangleq \int_0^\infty a(l)(1+e^{-l})\tanh^2\left(\frac{l}{2}\right)\, \mathrm{d}l. \tag{12}$$

For the BEC, this bound is tightened to

$$a_\mathrm{R} \geq \frac{\ln\left(1 + \frac{p-P_\mathrm{b}}{(1-p)\varepsilon+P_\mathrm{b}}\right)}{\ln\left(\frac{1}{1-p}\right)} \tag{13}$$

where p is the erasure probability of the BEC and P_b is the erasure probability of the code. Furthermore, among all the MBIOS channels with a fixed capacity C, and for fixed values of the gap (in rate) to capacity (ε) and bit error/ erasure probability (P_b), the lower bound on the average degree of the parity-check nodes given in (11) attains its maximal and minimal values for a BSC and BEC, respectively.

Remark 3: In the particular case where P_b vanishes, the lower bound on the average right degree introduced in (11) forms a tightened version of the bound given in [25, Eq. (77)]; note also that the bit error probability is taken into account in Theorem 1 while it is assumed to vanish in [25, Eq. (77)]. This point and some of its implications are clarified in Discussion 1, which proceeds the proof of Theorem 1. In the limit where the gap (in rate) to capacity vanishes, the lower bounds on the average right degree in (11) and [25, Eq. (77)] both grow like the logarithm of the inverse of the gap to capacity and they possess the same behavior. In this case

$$a_\mathrm{R} = \Omega\left(\ln\frac{1}{\varepsilon}\right). \tag{14}$$

Besides of tightening the bound in [25, Eq. (77)] and generalizing it to the case where the bit error (or erasure) probability is strictly positive, Theorem 1 also states the two extreme cases among all MBIOS channels with fixed capacity.

Remark 4: As is clarified later in Discussion 2, Theorem 1 can be adapted to hold for an arbitrary ensemble of (n, λ, ρ) LDPC codes. In this case, the requirement of a full-rank parity-check matrix of a particular code $\mathcal{C}$ from this ensemble is relaxed by requiring that the design rate of the LDPC ensemble forms a fraction $1-\varepsilon$ of the channel capacity. In this case, P_{b} stands for the average bit error (or erasure) probability of the ensemble under some decoding algorithm.

Corollary 1: Under the assumptions in Theorem 1, the cardinality of the fundamental system of cycles of a Tanner graph $\mathcal{G}$, associated with a full spanning forest of this graph, is larger than

$$n\big[(1-R)(a_{\mathrm{R}}-1)-1\big] \tag{15}$$

where from Theorem 1, a_{R} can be replaced by the lower bounds in (11) and (13) for a general MBIOS channel and a BEC, respectively. From this corollary and Remark 3, the cardinality of the fundamental system of cycles of the Tanner graph $\mathcal{G}$ which is associated with a full spanning forest of this graph is $\Omega\left(\ln \frac{1}{\varepsilon}\right)$.

Based on Remark 4 and Corollary 1, the following result is derived:

Corollary 2: **[On the asymptotic average cardinality of the fundamental system of cycles of LDPC ensembles]** Consider a sequence of LDPC ensembles, specified by an arbitrary pair of degree distributions (λ, ρ), whose transmission takes place over an MBIOS channel. Let the design rate of these ensembles be a fraction $1-\varepsilon$ of the channel capacity C (in units of bits per channel use), and assume that the average bit error/ erasure probability of such an LDPC ensemble vanishes under some decoding algorithm as we let the block length tend to infinity. Then, the average cardinality of the fundamental system of cycles of a Tanner graph $\mathcal{G}$ (denoted by $\beta(\mathcal{G})$) satisfies the following asymptotic property when the average is taken over all the Tanner graphs which represent codes from an LDPC ensemble of this sequence and as we let the block length tend to infinity:

$$\begin{aligned} &\liminf_{n\to\infty} \frac{\mathbb{E}_{\mathrm{LDPC}(n,\lambda,\rho)}\big[\beta(\mathcal{G})\big]}{n} \\ &\geq \frac{(1-C)\ \ln\left(g_1\ \left[1-2h_2^{-1}\left(\frac{1-C}{1-(1-\varepsilon)C}\right)\right]^{-2}\right)}{\ln\left(\frac{1}{g_1}\right)} - 1 \end{aligned} \tag{16}$$

where g_1 in introduced in (12). For a BEC whose erasure probability is p, a tightened version of this result gets the form

$$\liminf_{n\to\infty} \frac{\mathbb{E}_{\mathrm{LDPC}(n,\lambda,\rho)}\big[\beta(\mathcal{G})\big]}{n} \geq \frac{p\ \ln\left(1-p+\frac{p}{\varepsilon}\right)}{\ln\left(\frac{1}{1-p}\right)} - 1. \tag{17}$$

Furthermore, among all the MBIOS channels with a fixed capacity C and for a fixed value of the achievable gap in rate (ε) to the channel capacity, the lower bound (16) on the asymptotic average cardinality of the fundamental system of cycles attains its maximal and minimal values for a BSC and BEC, respectively.

Remark 5: Corollary 2 provides two results which are of the type $\Omega\left(\ln \frac{1}{\varepsilon}\right)$.

Theorem 2: **[On the degree distributions of capacity-approaching LDPC ensembles]** Let $\big(n, \lambda, \rho\big)$ be an ensemble of LDPC codes whose transmission takes place over an MBIOS channel. Assume that the design rate of the ensemble is equal to a fraction $1-\varepsilon$ of the channel capacity C, and let P_{b} designate the average bit error (or erasure) probability of the ensemble under ML decoding or any sub-optimal decoding algorithm. Then, the following properties hold for an arbitrary finite degree i from the node perspective in the case where $\varepsilon C + h_2(P_{\mathrm{b}}) \ll 1$

$$\Lambda_i = O(1) \tag{18}$$

$$\Gamma_i = O\big(\varepsilon C + h_2(P_{\mathrm{b}})\big) \tag{19}$$

and the following properties hold for the degree distributions from the edge perspective:

$$\lambda_i = O\left(\frac{1}{\ln \frac{1}{\varepsilon C + h_2(P_{\mathrm{b}})}}\right) \tag{20}$$

$$\rho_i = O\left(\frac{\varepsilon C + h_2(P_{\mathrm{b}})}{\ln \frac{1}{\varepsilon C + h_2(P_{\mathrm{b}})}}\right). \tag{21}$$

For the case where the transmission takes place over the BEC, the bounds above are tightened by replacing $h_2(P_\text{b})$ with P_b.

Remark 6: **[On the connection between Theorems 1 and 2]** Theorem 2 implies that for any capacity-approaching LDPC ensemble whose bit error probability vanishes and also for any finite degree i in their Tanner graphs, the fraction of variable nodes and parity-check nodes of degree i tends to zero as the gap to capacity (ε) vanishes. This conclusion is consistent with Theorem 1 which states that the average left and right degrees of the Tanner graphs scale at least like $\ln \frac{1}{\varepsilon}$; hence, these average degrees necessarily become unbounded as the gap to capacity vanishes.

Corollary 3: Under the assumptions of Theorem 2, in the limit where the bit error (or erasure) probability of a sequence of LDPC ensembles vanishes asymptotically (as we let the block length tend to infinity) and the design rate is a fraction $1-\varepsilon$ of the channel capacity, the following properties hold for an arbitrary finite degree i

$$\begin{aligned} \Lambda_i &= O(1)\,, & \Gamma_i &= O(\varepsilon)\,, \\ \lambda_i &= O\left(\frac{1}{\ln \frac{1}{\varepsilon}}\right), & \rho_i &= O\left(\frac{\varepsilon}{\ln \frac{1}{\varepsilon}}\right). \end{aligned}$$

We turn now our attention to ensembles of LDPC codes which achieve vanishing bit error (or erasure) probability under the sum-product decoding algorithm. The following theorem relies on information-theoretic arguments and the stability condition, and it provides an upper bound on the fraction of degree-2 variable nodes for a sequence of LDPC ensembles whose transmission takes place over an arbitrary MBIOS channel. As we will see, the following theorem provides a tight upper bound for capacity-achieving sequences of LDPC ensembles over the BEC.

Theorem 3: **[On the fraction of degree-2 variable nodes of LDPC ensembles]** Let $\bigl\{\bigl(n_m, \lambda(x), \rho(x)\bigr)\bigr\}_{m\geq 1}$ be a sequence of ensembles of LDPC codes whose transmission takes place over an MBIOS channel. Assume that this sequence asymptotically achieves a fraction $1-\varepsilon$ of the channel capacity under iterative sum-product decoding with vanishing bit error probability. Based on the notation in Theorem 1, the fraction of degree-2 variable nodes satisfies

$$\Lambda_2 < \frac{e^r(1-C)}{2}\left(1+\frac{\varepsilon C}{1-C}\right)\cdot\left[1+\frac{\ln\left(\frac{1}{g_1}\right)}{\ln\left(\frac{g_1}{\left[1-2h_2^{-1}\left(\frac{1-C}{1-(1-\varepsilon)C}\right)\right]^2}\right)}\right] \tag{22}$$

where

$$r \triangleq -\ln\left(\int_{-\infty}^{\infty} a(l)e^{-\frac{l}{2}}\,\mathrm{d}l\right) \tag{23}$$

is the Bhattacharyya distance which only depends on the channel. For sequences of LDPC ensembles whose transmission takes place over a BEC with an erasure probability p, if the bit erasure probability vanishes under iterative message-passing decoding, then the bound on the fraction of degree-2 variable nodes is tightened to

$$\Lambda_2 < \frac{1}{2}\left(1+\frac{\varepsilon(1-p)}{p}\right)\left[1+\frac{\ln\left(\frac{1}{1-p}\right)}{\ln\left(1-p+\frac{p}{\varepsilon}\right)}\right]. \tag{24}$$

Corollary 4: Under the assumptions of Theorem 3, in the limit where the gap to capacity vanishes under the sum-product decoding algorithm (i.e., $\varepsilon \to 0$), the fraction of degree-2 variable nodes satisfies

$$\Lambda_2 \leq \frac{e^r(1-C)}{2}\,. \tag{25}$$

The tightness of the latter upper bound on the fraction of degree-2 variable nodes for capacity-achieving sequences of LDPC ensembles under iterative sum-product decoding is considered in Discussion 5 which follows the proof of Theorem 3. We state there two conditions which ensure the tightness of the bound in (25).

Remark 7: Note that for capacity-achieving sequences of LDPC ensembles whose transmission takes place over the BEC, the bound in (25) is particularized to $\frac{1}{2}$ *regardless of the erasure probability of this channel.* This is indeed the case for some capacity-achieving LDPC ensembles over the BEC (see, e.g., [12], [17], [24]).

Remark 8: Let us consider sequences of LDPC ensembles which achieve the capacity of a BEC under iterative message-passing decoding. As mentioned above, for all such known sequences, the fraction of variable nodes of degree 2 tends to $\frac{1}{2}$ as the gap to capacity vanishes. This is in contrast to the behavior of the right degree-distribution where, for any fixed degree i, Corollary 3 implies that the fraction of the parity-check nodes of degree i tends to zero as the gap to capacity vanishes.
In order to complement the picture for degree-2 variable nodes, we provide in the following theorem an upper bound on the fraction of edges connected to degree-2 variable nodes, and show that this fraction of edges vanishes as the gap to capacity under iterative decoding tends to zero. Like the previous theorem, the derivation of the following theorem also relies on information-theoretic arguments and the stability condition.

Theorem 4: **[On the fraction of edges connected to degree-2 variable nodes]** Under the assumptions of Theorem 3, the fraction of edges connected to variable nodes of degree 2 satisfies

$$\lambda_2 < \frac{e^r \ln\left(\frac{1}{g_1}\right)}{\ln\left(\frac{g_1}{\left[1-2h_2^{-1}\left(\frac{1-C}{1-(1-\varepsilon)C}\right)\right]^2}\right)} \tag{26}$$

where g_1 and r are introduced in (12) and (23), respectively. For a BEC with erasure probability p, the bound in (26) is tightened to

$$\lambda_2 < \frac{\ln\left(\frac{1}{1-p}\right)}{p \ln\left(1-p+\frac{p}{\varepsilon}\right)}. \tag{27}$$

Corollary 5: **[A looser and simple version of the bounds in Theorem 4]** The upper bound on the fraction of edges connected to degree-2 variable nodes can be loosened to

$$\lambda_2 < \frac{1}{\left[c_1 + c_2 \ln\left(\frac{1}{\varepsilon}\right)\right]^+} \tag{28}$$

for some constants c_1 and c_2 which only depend on the MBIOS channel, and where $[x]^+ \triangleq \max(x, 0)$; the coefficient c_2 of the logarithm in (28) is given by

$$c_2 = \frac{e^{-r}}{\ln\left(\frac{1}{g_1}\right)} \tag{29}$$

and it is strictly positive.

Theorem 4 and Corollary 5 show that the fraction of edges connected to degree-2 variable nodes of capacity-achieving LDPC ensembles tends to zero, though the decay of the upper bound on λ_2 to zero is rather slow as the gap to capacity vanishes. In the following proposition, we show that for transmission over the BEC, the bounds in (27) and (28) are indeed tight.

Proposition 1: **[On the tightness of the upper bound on λ_2 for capacity-achieving sequences of LDPC ensembles over the BEC]** The upper bounds on the fraction of edges connected to degree-2 variable nodes in (27) and (28) are tight for the sequence of capacity-achieving right-regular LDPC ensembles over the BEC in [24]. For this capacity-achieving sequence, $\lambda_2 \triangleq \lambda_2(\varepsilon)$ decays to zero as $\varepsilon \to 0$ similarly to the upper bound in (28) with the same coefficient c_2 of $\ln\left(\frac{1}{\varepsilon}\right)$ as given in (29).

IV. Proofs and Discussions

A. Proof of Theorem 1

Let $\mathbf{X}$ be a random codeword from the binary linear block code $\mathcal{C}$. Let $\mathbf{Y}$ designate the output of the communication channel when $\mathbf{X}$ is transmitted. Based on the assumption that the code $\mathcal{C}$ is represented by a full-rank parity-check matrix and that $\mathcal{G}$ is the corresponding bipartite (Tanner) graph representing this code, the conditional entropy of $\mathbf{X}$ given $\mathbf{Y}$ satisfies the inequality in (8) (the derivation of this lower bound on the conditional entropy is outlined in Section II-C; for the full derivation, which includes all the mathematical details, the reader is referred to [25]). Using the convexity of $f(t) = x^t$ for all $x \geq 0$ and applying Jensen's inequality gives

$$\Gamma(x) = \sum_i \Gamma_i x^i \geq x^{\sum_i i \Gamma_i} = x^{a_{\mathrm{R}}}, \quad x \geq 0\,.$$

Substituting the inequality above in (8) implies that

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq R - C + \frac{1-R}{2\ln 2} \sum_{k=1}^{\infty} \frac{g_k^{a_{\mathrm{R}}}}{k(2k-1)}\,. \tag{30}$$

To continue, we apply the following lemma, which relates g_1 and g_k in (9) for all $k \in \mathbb{N}$.

Lemma 3:

$$g_k \geq (g_1)^k\,, \quad \forall k \in \mathbb{N} \tag{31}$$

where g_k is defined in (9).

Proof: For $k = 1$, (31) is trivial. For $k > 1$, the equality in (10) and Hölder's inequality give

$$\begin{aligned} g_k &= \mathbb{E}\left[\tanh^{2k}\left(\frac{L}{2}\right)\right] \\ &= \left(\mathbb{E}^{\frac{1}{k}}\left[\left(\tanh^2\left(\frac{L}{2}\right)\right)^k\right]\right)^k \\ &\geq \left(\mathbb{E}\left[\left(\tanh^2\left(\frac{L}{2}\right)\right)\right]\right)^k \\ &= (g_1)^k\,. \end{aligned}$$

■

The substitution of (31) in (30) gives

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq R - C + \frac{1-R}{2\ln 2} \sum_{k=1}^{\infty} \frac{(g_1^{a_{\mathrm{R}}})^k}{k(2k-1)}\,. \tag{32}$$

Expanding the binary entropy function into a power series around $\frac{1}{2}$ (see [25, Appendix II.A]) gives

$$h_2(x) = 1 - \frac{1}{2\ln 2} \sum_{k=1}^{\infty} \frac{(1-2x)^{2k}}{k(2k-1)}\,, \quad 0 \leq x \leq 1 \tag{33}$$

and assigning $x = \frac{1-\sqrt{u}}{2}$ gives

$$\frac{1}{2\ln 2} \sum_{k=1}^{\infty} \frac{u^k}{k(2k-1)} = 1 - h_2\left(\frac{1-\sqrt{u}}{2}\right), \quad 0 \leq u \leq 1. \tag{34}$$

Since $0 \leq \tanh^2(x) < 1$ for all $x \in \mathbb{R}$, we get from (9) that $0 \leq g_1 \leq 1$ (this property holds for all the sequence $\{g_k\}_{k=1}^{\infty}$). Substituting (34) into (32) gives

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq R - C + (1-R)\left[1 - h_2\left(\frac{1-g_1^{a_{\mathrm{R}}/2}}{2}\right)\right]. \tag{35}$$

Fano's inequality provides the following upper bound on the conditional entropy of $\mathbf{X}$ given $\mathbf{Y}$:

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \leq R\, h_2(P_{\mathrm{b}}) \tag{36}$$

where P_{b} designates the bit error probability of the code under some arbitrary decoding algorithm (without any loss of generality, one can assume that the first nR bits of the code are its information bits, and their knowledge is sufficient for determining the codeword). In order to make later the statement also valid for code ensembles (as is clarified later in Discussion 2), we loosen the bound in (36), and get

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \leq h_2(P_\mathrm{b}). \tag{37}$$

Combining the upper bound on the conditional entropy in (37) with the lower bound in (35) gives

$$h_2(P_\mathrm{b}) \geq R - C + (1-R)\left[1 - h_2\left(\frac{1-g_1^{a_\mathrm{R}/2}}{2}\right)\right]. \tag{38}$$

Since the RHS of (38) is monotonically increasing in R, then following our assumption, one can possibly loosen the bound by replacing R with $(1-\varepsilon)C$; this yields after some algebra that

$$h_2\left(\frac{1-g_1^{a_\mathrm{R}/2}}{2}\right) \geq \frac{1-C-h_2(P_\mathrm{b})}{1-(1-\varepsilon)C}\,.$$

Since the binary entropy function h_2 is monotonically increasing between 0 and $\frac{1}{2}$ then

$$g_1^{\frac{a_\mathrm{R}}{2}} \leq 1 - 2h_2^{-1}\left(\frac{1-C-h_2(P_\mathrm{b})}{1-(1-\varepsilon)C}\right)$$

which gives the lower bound on a_R in (11) (note that $g_1 \leq 1$ and this inequality is strict unless the channel is noiseless).

Let us now consider the particular case where the transmission is over the BEC. Note that for a BEC with erasure probability p, $g_k = 1-p$ for all $k \in \mathbb{N}$ and therefore (32) is particularized to

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq R - C + \frac{(1-R)(1-p)^{a_\mathrm{R}}}{2\ln 2} \sum_{k=1}^{\infty} \frac{1}{k(2k-1)}\,.$$

Substituting $u=1$ in (34) gives the equality

$$\frac{1}{2\ln 2}\sum_{k=1}^{\infty}\frac{1}{k(2k-1)} = 1$$

and hence, the last inequality gets the form

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq R - C + (1-R)(1-p)^{a_\mathrm{R}}\,. \tag{39}$$

Note that the RHS of (39) is monotonic increasing with the code rate R; following our assumption that $R \geq (1-\varepsilon)C$ and since $C = 1-p$ is the channel capacity of the BEC, we get

$$\frac{H(\mathbf{X}|\mathbf{Y})}{n} \geq -\varepsilon(1-p) + \big(1-(1-\varepsilon)(1-p)\big)(1-p)^{a_\mathrm{R}}\,. \tag{40}$$

The normalized conditional entropy $\frac{H(\mathbf{X}|\mathbf{Y})}{n}$ satisfies

$$\begin{aligned}
\frac{H(\mathbf{X}|\mathbf{Y})}{n} &= \frac{1}{n}\sum_{i=1}^{n} H\Big(X_i|\mathbf{Y}, \big\{X_j\big\}_{j<i}\Big) \\
&\stackrel{(a)}{=} \frac{1}{n}\sum_{i=1}^{nR} H\Big(X_i|\mathbf{Y}, \big\{X_j\big\}_{j<i}\Big) \\
&\leq \frac{1}{n}\sum_{i=1}^{nR} H(X_i|\mathbf{Y}) \\
&\stackrel{(b)}{=} R\,P_\mathrm{b} \\
&\leq P_\mathrm{b}
\end{aligned} \tag{41}$$

where (a) holds since the code dimension is at most nR and assuming without any loss of generality that the first nR bits form the information bits of the code, and (b) holds since given $\mathbf{Y}$, the decoder finds X_i with probability $1 - P_{\rm b}$ and otherwise the bit X_i is erased and due to the linearity of the code it takes the values 0 and 1 with equal probability so its entropy is equal to 1 bit. Combining (40) with (41) gives

$$P_{\rm b} \geq -\varepsilon(1-p) + \big(1 - (1-\varepsilon)(1-p)\big)(1-p)^{a_{\rm R}}\,. \tag{42}$$

Finally, the lower bound on the average right degree in (13) follows from (42) by simple algebra. Note that in the case where $P_{\rm b} = 0$, the resulting lower bound coincides with the result obtained in [22, p. 1619] (though it was obtained there in a different way), and it gets the form

$$a_{\rm R} \geq \frac{\ln\left(1 + \frac{p}{(1-p)\varepsilon}\right)}{\ln\left(\frac{1}{1-p}\right)}. \tag{43}$$

We wish now to show that among all the MBIOS channels with a fixed capacity C and for fixed values of the gap to capacity (ε) and the bit error/ erasure probability $(P_{\rm b})$, the lower bound on the average degree of the parity-check nodes as given in (11) attains its maximal and minimal values for a BSC and BEC, respectively. To this end, note that for fixed values of C, ε and $P_{\rm b}$, the numerator of the lower bound in (11) is fixed; hence, the value of this lower bound is maximized or minimized by maximizing or minimizing, respectively, the value of g_1 as given in (12) (note that g_1 lies in general between zero and one).

Lemma 4: **[Extreme values of g_1 among all MBIOS channels with a fixed capacity]** Among all the MBIOS channels with a fixed capacity C, the value of g_1 satisfies

$$C \leq g_1 \leq \big(1 - 2h_2^{-1}(1-C)\big)^2 \tag{44}$$

and these upper and lower bounds on g_1 are attained for a BSC and BEC, respectively.

Proof: The proof of the lower bound on g_1 follows from the calculations in [25, p. 565], though the proof there is not explicit (since something else was needed to prove there). We prove here this lower bound explicitly to improve readability. From (33), we get that for $0 \leq x \leq 1$

$$\begin{aligned} h_2(x) &= 1 - \frac{1}{2\ln 2} \sum_{k=1}^{\infty} \frac{(1-2x)^{2k}}{k(2k-1)} \\ &\geq 1 - \frac{(1-2x)^2}{2\ln 2} \sum_{k=1}^{\infty} \frac{1}{k(2k-1)} \\ &= 1 - (1-2x)^2 \end{aligned} \tag{45}$$

where the last inequality holds since $(1-2x)^{2k} \leq (1-2x)^2$ for $k \in \mathbb{N}$ and $0 \leq x \leq 1$, and the last transition follows from the equality $\sum_{k=1}^{\infty} \frac{1}{2k(2k-1)} = \ln 2$. By substituting $x = \frac{1}{1+e^l}$ for $l \geq 0$ in (45), it follows that

$$h_2\left(\frac{1}{1+e^l}\right) \geq 1 - \tanh^2\left(\frac{l}{2}\right), \quad \forall\, l \in [0, \infty). \tag{46}$$

The substitution of (46) in (12) gives

$$\begin{aligned} g_1 &= \int_0^{\infty} a(l)(1+e^{-l}) \tanh^2\left(\frac{l}{2}\right)\, {\rm d}l \\ &\geq \int_0^{\infty} a(l)(1+e^{-l}) \left[1 - h_2\left(\frac{1}{1+e^l}\right)\right]\, {\rm d}l \\ &= C \end{aligned} \tag{47}$$

where the last equality relies on a possible representation of the channel capacity for an arbitrary MBIOS channel (see [21, Section 4.1.8]). This completes the proof of the lower bound on g_1, as given on the LHS of (44). Note that this lower bound on g_1 is attained for a BEC (since for a BEC with an arbitrary erasure probability p, we get from (12) that $g_1 = 1 - p = C$).

In order to prove the upper bound on g_1, as given on the RHS of (44), observe that the channel capacity of an arbitrary MBIOS channel satisfies

$$\begin{aligned} C &= \int_0^\infty a(l)(1+e^{-l}) \left[1 - h_2\left(\frac{1}{1+e^{\,l}}\right)\right] \, \mathrm{d}l \\ &\stackrel{\text{(a)}}{=} \int_0^\infty a(l)(1+e^{-l}) \, \frac{1}{2\ln 2} \sum_{k=1}^{\infty} \frac{\tanh^{2k}\left(\frac{l}{2}\right)}{k(2k-1)} \, \mathrm{d}l \\ &= \frac{1}{2\ln 2} \sum_{k=1}^{\infty} \left\{ \frac{\int_0^\infty a(l)(1+e^{-l}) \tanh^{2k}\left(\frac{l}{2}\right) \, \mathrm{d}l}{k(2k-1)} \right\} \\ &\stackrel{\text{(b)}}{=} \frac{1}{2\ln 2} \sum_{k=1}^{\infty} \frac{g_k}{k(2k-1)} \end{aligned} \tag{48}$$

where equality (a) follows by substituting in (33) $x = \frac{1}{1+e^{\,l}}$ for $l \geq 0$, and equality (b) follows from (9); this provides an expression for the channel capacity in terms of the non-negative sequence $\{g_k\}_{k=0}^{\infty}$ defined in (9). Since we look for the maximal value of g_1 among all MBIOS channels with a fixed capacity, then we need to solve the maximization problem

$$\text{maximize} \left\{ g_1 \;\middle|\; \frac{1}{2\ln 2} \, \sum_{k=1}^{\infty} \frac{g_k}{k(2k-1)} = C \right\}. \tag{49}$$

Based on Lemma 3, for every MBIOS channel, $g_k \geq (g_1)^k$ for all $k \in \mathbb{N}$; therefore, one can write $g_k \triangleq (g_1)^k + \epsilon_k$ where $\epsilon_k \geq 0$ for all $k \in \mathbb{N}$. By substituting this in the infinite series (48), we get

$$\begin{aligned} &\frac{1}{2\ln 2} \sum_{k=1}^{\infty} \frac{g_k}{k(2k-1)} \\ &= \frac{1}{2\ln 2} \sum_{k=1}^{\infty} \frac{(g_1)^k}{k(2k-1)} + \frac{1}{2\ln 2} \sum_{k=1}^{\infty} \frac{\epsilon_k}{k(2k-1)} \\ &= 1 - h_2\left(\frac{1-\sqrt{g_1}}{2}\right) + \frac{1}{2\ln 2} \sum_{k=1}^{\infty} \frac{\epsilon_k}{k(2k-1)} \end{aligned} \tag{50}$$

where the last equality is based on (34). Since $\epsilon_k \geq 0$ for all $k \in \mathbb{N}$, the equality constraint in (49) and the equality in (50) yield that

$$1 - h_2\left(\frac{1-\sqrt{g_1}}{2}\right) \leq C$$

from which the RHS of (44) follows. Note that this upper bound on g_1 is achieved when the second term in (50) vanishes. Since $\{\epsilon_k\}_{k=1}^{\infty}$ is a non-negative sequence, this happens if and only if $\epsilon_k = 0$ for all $k \in \mathbb{N}$. For a BSC with an arbitrary crossover probability p, the LLR at the channel output is bimodal and it gets the values $\pm \ln\left(\frac{1-p}{p}\right)$, which implies from (10) that

$$\begin{aligned} g_k &\triangleq \mathbb{E}\left[\tanh^{2k}\left(\frac{L}{2}\right)\right] \\ &= (1-p) \tanh^{2k}\left(\frac{l}{2}\right)\bigg|_{l=\ln\left(\frac{1-p}{p}\right)} + p \tanh^{2k}\left(\frac{l}{2}\right)\bigg|_{l=-\ln\left(\frac{1-p}{p}\right)} \\ &= \tanh^{2k}\left(\frac{l}{2}\right)\bigg|_{l=\ln\left(\frac{1-p}{p}\right)} \\ &= \left(\frac{e^l-1}{e^l+1}\right)^{2k}\bigg|_{l=\ln\left(\frac{1-p}{p}\right)} \\ &= (1-2p)^{2k}\,, \quad \forall\, k \in \mathbb{N}. \end{aligned} \tag{51}$$

Hence for the BSC, $g_k = (g_1)^k$ and $\epsilon_k = 0$ for all $k \in \mathbb{N}$. The upper bound on g_1 on the RHS of (44) is therefore achieved for a BSC whose crossover probability is $p = h_2^{-1}(1-C)$. ■

Based on the paragraph which proceeds Lemma 4 and the result of this lemma, the last claim in Theorem 1 is proved.

Discussion 1: **[A discussion on the bounds in Theorem 1 and their comparison to [25, Eq. (77)]]** In the particular case where P_{b} vanishes, the lower bound on the average right degree, as introduced in (11), forms a tightened version of the bound given in [25, Eq. (77)]. We note that both bounds are derived from the lower bound (8) on the conditional entropy of the transmitted codeword given the channel output. However, the bound in (11) (even in the particular case where we set P_{b} to zero) is tighter than the bound given in [25, Eq. (77)]; this follows from the improved tightness of the lower bound on the RHS of (8), as obtained in the proof of Theorem 1. Particularly, the authors in [25] rely on the fact that all the terms of the infinite sum in the RHS of (8) are non-negative, and derive a simple lower bound on the conditional entropy by truncating this infinite sum after its first term. In the proof of Theorem 1, on the other hand, this lower bound on the conditional entropy is improved by applying Jensen's inequality and the Taylor series expansion in (34). Note that in the case where the gap to capacity vanishes (i.e., $\varepsilon \to 0$), the average right degree tends to infinity. Since $g_1 \leq 1$ where this inequality is strict unless the channel is noiseless, then when ε becomes small, $g_1^{a_{\mathrm{R}}} << 1$. As compared to the derivation of Theorem 1, the loss in the tightness of the bound in [25] due to the truncation of the infinite sum after its first term becomes marginal as the gap (in rate) to capacity vanishes; though, the difference between the two lower bounds on the average right degree given in (8) and [25, Eq. (77)]] becomes more significant as the gap to capacity is increased. Note also that the bound on the average right degree in Theorem 1 takes into account the bit error probability at the end of the decoding process, while the bound in [25, Eq. (77)] applies only to the case where P_{b} vanishes. The additional dependence of the bound (11) on P_{b} makes the bounds in Theorem 1 valid for LDPC ensembles of finite block length while the bound in [25, Eq. (77)] can be only applied to the asymptotic case of vanishing bit error (or erasure) probability by letting the block length tend to infinity. In addition to the tightening and generalization of the bound in [25, Eq. (77)], Theorem 1 also states the two extreme cases of (11) among all the MBIOS channels whose capacity is fixed.

Discussion 2: **[An adaptation of Theorem 1 for LDPC ensembles]** As mentioned in Remark 4, the statement in Theorem 1 can be easily adapted to hold for an LDPC ensemble (n, λ, ρ) whose transmission takes place over an arbitrary MBIOS channel. This modification is done by relaxing the requirement of a full-rank parity-check matrix for a particular code with the requirement that the design rate of the ensemble is (at least) a fraction $1 - \varepsilon$ of the channel capacity. In this case, by taking the statistical expectation over the codes from the considered LDPC ensemble (in addition to the original statistical expectation over the codewords of the code), the inequality in (35) holds where we also need here to average over the rate R of the codes from the ensemble. Instead of averaging over the rate R of codes from the ensemble, since the RHS of (35) is monotonically increasing with R, one can replace the rate of any code from this ensemble with the design rate of the ensemble (which forms a lower bound on R for every code from this ensemble). By assumption, the design rate is not less than $(1-\varepsilon)C$; hence, (35) and the requirement on the design rate yield that

$$\mathbb{E}\left[\frac{H(\mathbf{X}|\mathbf{Y})}{n}\right] \geq 1 - C - \big(1 - (1-\varepsilon)C\big)\, \mathbb{E}\left[h_2\left(\frac{1 - g_1^{a_{\mathrm{R}}/2}}{2}\right)\right] \tag{52}$$

where $\mathbb{E}$ designates the statistical expectation over all the codes from the LDPC ensemble (n, λ, ρ).

Since the binary entropy function is a concave function, Jensen's inequality gives

$$\mathbb{E}\left[h_2\left(\frac{1 - g_1^{a_{\mathrm{R}}/2}}{2}\right)\right] \leq h_2\left(\frac{1 - \mathbb{E}\left[g_1^{a_{\mathrm{R}}/2}\right]}{2}\right) \tag{53}$$

and another application of Jensen's inequality to the exponential function (which is convex) gives

$$\mathbb{E}\left[g_1^{a_{\mathrm{R}}/2}\right] \geq g_1^{\mathbb{E}[a_{\mathrm{R}}]/\,2}. \tag{54}$$

Note that $0 \leq g_1 \leq 1$ and for every code from the ensemble $a_{\mathrm{R}} \geq 2$ (since $\rho_i = 0$ for $i = 0, 1$, i.e., the degrees of all the parity-check nodes are at least 2, then also the average right degree of any code from the ensemble cannot be smaller than 2); this implies that the arguments inside the binary entropy functions of (53) lie between zero and one-half.

The loosening of the bound in the transition from (36) to (37) is due to the fact that we need an upper bound on the rate R of a code from this ensemble; since we consider binary codes, a trivial upper bound on the rate is 1 bit per channel use (note that the rate of an arbitrarily chosen code from this ensemble may exceed the channel capacity). Note also that due to the concavity of the binary entropy function and Jensen's inequality, one can replace P_{b} on the RHS of (37) which applies to a particular code with the average bit error probability of the ensemble, and the upper bound on the conditional entropy still holds. This gives

$$h_2(\overline{P_{\mathrm{b}}}) \geq \mathbb{E}\left[\frac{H(\mathbf{X}|\mathbf{Y})}{n}\right] \tag{55}$$

where $\overline{P_{\mathrm{b}}} \triangleq \mathbb{E}\big[P_{\mathrm{b}}\big]$ designates the average bit error probability of the LDPC ensemble.

The combination of the chain of inequalities in (52)–(55) leads to the adaptation of the statement in Theorem 1 for arbitrary LDPC ensembles with the proper modification of the requirement on the design rate of the ensemble (instead of the requirement of a full-rank parity-check matrix which is quite heavy for individual codes which are selected at random from an LDPC ensemble), and the reference to the average bit error (or erasure) probability of the ensemble.

Note that the adaptation of the statement in Theorem 1 for LDPC ensembles whose transmission takes place over the BEC is more direct. In this case, we refer to (42); since in the latter case, $h_2(P_{\mathrm{b}})$ is replaced by P_{b}, then two uses (out of three) of Jensen's inequality become irrelevant for the BEC.

Discussion 3: **[An adaptation of Theorem 1 for LDPC ensembles with random or intentional puncturing patterns]** In continuation to Discussion 2, it is also possible to adapt Theorem 1 to hold for ensembles of punctured LDPC codes whose transmission takes place over an MBIOS channel. To this end, we refer the reader to [23, Section 5] which is focused on the derivation of lower bounds on the average right degree and the graphical complexity of such ensembles. The derivation of these bounds relies on a lower bound on the conditional entropy of the transmitted codeword given the received sequence at the output of a set of parallel MBIOS channels (see [23, Eqs. (2) and (3)]); the latter bound was particularized in [23, Sections 2–4] to the two settings of randomly and intentionally punctured LDPC ensembles which are communicated over a single MBIOS channel. We note that the lower bound on the conditional entropy, as given in (8), forms a particular case of the bound in [23, Eqs. (2) and (3)] where the set of parallel MBIOS channels degenerates to a single MBIOS channel. The concept of the proof of Theorem 1 enables one to tighten the lower bounds on the average right degree and the graphical complexity, as presented in [23, Section 5], for both randomly and intentionally punctured LDPC ensembles. More explicitly, by comparing the proof of (11) with the derivation of [25, Eq. (77)] under the assumption of vanishing bit error probability, one notices that the tightening of the bound in the former case follows by combining Lemma 3 with the equality in (34) (instead of the truncation of a non-negative infinite series after its first term, as was done for the derivation of the looser bound in [25]); this difference can be exploited exactly in the same way in [23, Section 5] for improving the tightness of the lower bounds on the average right degree and the graphical complexity for punctured LDPC ensembles.

Discussion 4: **[Lower bounds on the conditional entropy which are based on statistical physics]** Theorem 1 is proved via the lower bound on the conditional entropy (8) which holds for an arbitrary binary linear block code that is represented by a full-rank parity-check matrix and whose transmission takes place over an MBIOS channel. This bound is adapted above to hold for LDPC ensembles under ML decoding or any sub-optimal decoding algorithm (see Discussion 2). A different approach for obtaining a lower bound on the same conditional entropy for ensembles of LDPC codes and ensembles of low-density generator-matrix (LDGM) codes was introduced by Montanari in [15]. This bounding technique is based on tools borrowed from statistical physics, and it provides lower bounds on the entropy of the transmitted message conditioned on the received sequence at the output of the MBIOS channel. The computational complexity of the bound in [15] grows exponentially with the maximal right and left degrees (see [15, Eqs. (6.2) and (6.3)]) which therefore imposes a difficulty on the calculation of this bound (especially, for continuous-output channels). Since the bounds in [15] are derived for ensembles of codes, they are probabilistic in their nature; based on concentration arguments, they hold asymptotically in probability 1 as the block length tends to infinity. Based on heuristic statistical mechanics calculations, it was conjectured that the bounds in [15], which hold for general LDPC and LDGM ensembles over MBIOS channels, are tight. As opposed to the lower bound on the conditional entropy which is given in (8), the bounding techniques in [14] and [15], where both rely on statistical physics, do not provide a bound which is valid for every binary linear block code. It would be interesting to get some theory that unifies the information-theoretic and statistical physics approaches and provides bounds that are tight on the average and valid code by code.

Proof of Corollary 1

From Remark 2, the cardinality of the fundamental system of cycles of the Tanner graph $\mathcal{G}$, which is associated with a full spanning forest of $\mathcal{G}$, is equal to the cycle rank $\beta(\mathcal{G})$. From Eq. (2), we get that $\beta(\mathcal{G}) > |E_{\mathcal{G}}| - |V_{\mathcal{G}}|$ where $|E_{\mathcal{G}}|$ and $|V_{\mathcal{G}}|$ designate the number of edges and vertices. Specializing this for a Tanner graph $\mathcal{G}$ which represents a full-rank parity-check matrix of a binary linear block code, the number of vertices satisfies $|V_{\mathcal{G}}| = n(2-R)$ (since there are n variable nodes and $n(1-R)$ parity-check nodes in the graph) and the number of edges satisfies $|E_{\mathcal{G}}| = n(1-R)a_{\mathrm{R}}$. Combining these equalities gives the lower bound on the cardinality of the fundamental system of cycles in (4).

Proof of Corollary 2

The proof of (16) and (17) is based on Remark 4 and Corollary 1. From the substitution of $P_{\mathrm{b}} = 0$ in (11), which is justified in Discussion 2 for LDPC ensembles, one obtains the following lower bound on the average right degree of the LDPC ensemble as the average bit error probability of this ensemble vanishes:

$$a_{\mathrm{R}} \geq \frac{2\ln\left(\frac{1}{1-2h_2^{-1}\left(\frac{1-C}{1-(1-\varepsilon)C}\right)}\right)}{\ln\left(\frac{1}{g_1}\right)}. \tag{56}$$

Since we assume here that the bit error probability of the ensemble vanishes as the block length tends to infinity, then asymptotically with probability 1, the code rate of an arbitrary code from the considered ensemble does not exceed the channel capacity. By substituting the lower bound on a_{R} from (56) and an upper bound on R (i.e., $R \leq C$) into (15), the asymptotic result in (16) readily follows. The proof of (17) follows similarly based on (13) (with $P_{\mathrm{b}} = 0$). From the last statement in Theorem 1 regarding the maximal and minimal values of the lower bound in (11) among all the MBIOS channels with a fixed capacity, it follows that the same conclusion also holds w.r.t. the lower bound in (16). Hence, the RHS of (16) also attains, respectively, its maximal and minimal values for the BSC and BEC whose capacities are equal to C.

B. *Proof of Theorem 2*

Since the fraction of variable nodes of degree i is not greater than 1 for any degree i, (18) clearly holds (we demonstrate later that this result is asymptotically tight as the gap to capacity vanishes, at least for degree-2 variable nodes).

We turn now to consider the degrees of the parity-check nodes. Similarly to the proof of Theorem 1, we denote by $\mathbf{X}$ a random codeword from the ensemble (n, λ, ρ) where the randomness is over the selected code from the ensemble and the codeword which is selected from the code. Let $\mathbf{Y}$ designate the output of the communication channel when $\mathbf{X}$ is transmitted. The lower bound on the conditional entropy of $\mathbf{X}$ given $\mathbf{Y}$ in [25, Eq. (56)] gives

$$\begin{aligned}
&\frac{H(\mathbf{X}|\mathbf{Y})}{n} \\
&\geq R - C + \frac{1-R}{2\ln 2} \sum_{k=1}^{\infty} \frac{\Gamma(g_k)}{k(2k-1)} \\
&= -\varepsilon C + \frac{1-(1-\varepsilon)C}{2\ln 2} \sum_{i=2}^{\infty} \left[\Gamma_i \sum_{k=1}^{\infty} \frac{g_k^i}{k(2k-1)}\right]
\end{aligned} \tag{57}$$

where the equality follows from the definition of Γ and since the design rate of the ensemble satisfies $R = (1-\varepsilon)C$.

Applying Lemma 3 to the RHS of (57), we get

$$\begin{aligned}
&\frac{H(\mathbf{X}|\mathbf{Y})}{n}\\
&\geq -\varepsilon\, C + \frac{1-(1-\varepsilon)C}{2\ln 2}\,\sum_{i=2}^{\infty}\left\{\Gamma_i\sum_{k=1}^{\infty}\frac{\left(g_1^i\right)^k}{k(2k-1)}\right\}\\
&= -\varepsilon\, C + \big(1-(1-\varepsilon)C\big)\sum_{i=2}^{\infty}\left\{\left[1-h_2\left(\frac{1-g_1^{i/2}}{2}\right)\right]\,\Gamma_i\right\}
\end{aligned}$$

where the last equality follows from (34). Combining the upper bound on the conditional entropy in (36) with the last result gives

$$h_2(P_\text{b}) \geq -\varepsilon\, C + \big(1-(1-\varepsilon)C\big)\sum_{i=2}^{\infty}\left[1-h_2\left(\frac{1-g_1^{i/2}}{2}\right)\right]\,\Gamma_i$$

and therefore

$$\sum_{i=2}^{\infty}\left\{\left[1-h_2\left(\frac{1-g_1^{i/2}}{2}\right)\right]\,\Gamma_i\right\} \leq \frac{\varepsilon\, C + h_2(P_\text{b})}{1-(1-\varepsilon)C} \tag{58}$$

where P_b designates the average bit error probability of the ensemble under the considered decoding algorithm. Since all the terms in the sum on the LHS of (58) are non-negative, this sum is lower bounded by its i'th term, for any degree i. This provides the following lower bound on the fraction of parity-check nodes of degree i:

$$\begin{aligned}
\Gamma_i \quad &\leq \quad \frac{\varepsilon C + h_2(P_\text{b})}{1-(1-\varepsilon)C}\,\frac{1}{1-h_2\left(\frac{1-g_1^{i/2}}{2}\right)}\\
&\leq \quad \big(\varepsilon C + h_2(P_\text{b})\big)\left[\frac{1}{1-C}\,\frac{1}{1-h_2\left(\frac{1-g_1^{i/2}}{2}\right)}\right].
\end{aligned} \tag{59}$$

This completes the proof of the statement in (19) when the transmission takes place over an arbitrary MBIOS channel. Let us now consider the particular case where the transmission is over a BEC with erasure probability p. In this case, $g_k = 1-p$ for all $k \in \mathbb{N}$, and the channel capacity is given by $C = 1-p$. Therefore, (57) is particularized to

$$\begin{aligned}
\frac{H(\mathbf{X}|\mathbf{Y})}{n} &\geq -\varepsilon(1-p) + \frac{1-(1-\varepsilon)(1-p)}{2\ln 2}\cdot\sum_{i=2}^{\infty}\left[\Gamma_i\,(1-p)^i\,\sum_{k=1}^{\infty}\frac{1}{k(2k-1)}\right]\\
&= -\varepsilon(1-p) + \big(1-(1-\varepsilon)(1-p)\big)\sum_{i=2}^{\infty}\Gamma_i\,(1-p)^i
\end{aligned} \tag{60}$$

where the equality holds since $\sum_{k=1}^{\infty}\frac{1}{k(2k-1)} = 2\ln 2$. Applying the upper bound on the conditional entropy (41) to the LHS of (60), we get

$$P_\text{b} \geq -\varepsilon(1-p) + \big(1-(1-\varepsilon)(1-p)\big)\sum_{i=2}^{\infty}\Gamma_i\,(1-p)^i$$

and therefore

$$\sum_{i=2}^{\infty}\left\{(1-p)^i\,\Gamma_i\right\} \leq \frac{\varepsilon\,(1-p)+P_\text{b}}{1-(1-\varepsilon)(1-p)} \tag{61}$$

where this time P_b denotes the average bit erasure probability of the ensemble. Following the same steps as above, we get that for the BEC

$$\begin{aligned}
\Gamma_i \quad &\leq \quad \frac{\varepsilon\,(1-p)+P_\text{b}}{\big(1-(1-\varepsilon)(1-p)\big)(1-p)^i}\\
&\leq \quad \big(\varepsilon\,(1-p)+P_\text{b}\big)\left(\frac{1}{p\,(1-p)^i}\right)
\end{aligned} \tag{62}$$

so indeed $h_2(P_b)$ is replaced with P_b in (19).

We turn now to consider the pair of degree distributions from the edge perspective. The average left degree (a_L) of an LDPC ensemble satisfies

$$\begin{aligned}\frac{1}{a_L} &= \int_0^1 \lambda(x)\,dx \\ &= \int_0^1 \sum_{i=2}^{\infty} \lambda_i x^{i-1}\,dx \\ &= \sum_{i=2}^{\infty} \frac{\lambda_i}{i}\end{aligned} \tag{63}$$

which implies that the fraction of edges connected to variable nodes of an arbitrary degree i satisfies

$$\lambda_i \le \frac{i}{a_L}\,. \tag{64}$$

Since the design rate of the LDPC ensemble is assumed to be a fraction $1-\varepsilon$ of the channel capacity, then the average right and left degrees are related via the equality

$$a_L = \big(1-(1-\varepsilon)C\big)a_R\,. \tag{65}$$

Substituting (65) on the RHS of (64) and applying the lower bound on a_R in (11) gives

$$\lambda_i \le \frac{i\ln\big(\frac{1}{g_1}\big)}{2\big(1-(1-\varepsilon)C\big)\;\ln\left(\frac{1}{1-2h_2^{-1}\left(\frac{1-C-h_2(P_b)}{1-(1-\varepsilon)C}\right)}\right)}\,. \tag{66}$$

Using the power series for the binary entropy function in (33) and truncating the sum on the RHS after the first term gives

$$1-h_2(x) \ge \frac{(1-2x)^2}{2\,\ln 2}\,.$$

Substituting $u = h_2(x)$ yields

$$\big(1-2h_2^{-1}(u)\big)^2 \le 2\,\ln 2\cdot(1-u)\,. \tag{67}$$

Substituting (67) into the denominator on the RHS of (66), we get

$$\begin{aligned}\lambda_i &\le \frac{i\ln\big(\frac{1}{g_1}\big)}{\big(1-(1-\varepsilon)C\big)\;\ln\left(\frac{1}{2\ln 2}\,\frac{1}{1-\frac{1-C-h_2(P_b)}{1-(1-\varepsilon)C}}\right)} \\ &= \frac{i\ln\big(\frac{1}{g_1}\big)}{\big(1-(1-\varepsilon)C\big)\;\ln\left(\frac{1}{2\ln 2}\,\frac{1-(1-\varepsilon)C}{\varepsilon C+h_2(P_b)}\right)} \\ &\le \frac{i\ln\big(\frac{1}{g_1}\big)}{(1-C)\;\left[\ln\left(\frac{1}{\varepsilon C+h_2(P_b)}\right)+\ln\left(\frac{1-C}{2\ln 2}\right)\right]}\end{aligned}$$

which completes the proof of (20) for general MBIOS channels. For the BEC, we substitute (65) and the lower bound on the average right degree in (13) into the RHS of (64) to get

$$\begin{aligned}\lambda_i &\le \frac{i\ln\big(\frac{1}{1-p}\big)}{\big(1-(1-\varepsilon)(1-p)\big)\;\ln\left(1+\frac{p-P_b}{\varepsilon\,(1-p)+P_b}\right)} \\ &= \frac{i\ln\big(\frac{1}{1-p}\big)}{\big(1-(1-\varepsilon)(1-p)\big)\;\ln\left(\frac{\varepsilon\,(1-p)+p}{\varepsilon\,(1-p)+P_b}\right)} \\ &\le \frac{i\ln\big(\frac{1}{1-p}\big)}{p\;\left[\ln\left(\frac{1}{\varepsilon\,(1-p)+P_b}\right)+\ln(p)\right]}\,.\end{aligned} \tag{68}$$

Hence, $h_2(P_\text{b})$ is replaced by P_b in (20) when the communication channel is a BEC. Considering the right degree distribution of the ensemble, we have

$$\begin{aligned}\frac{1}{a_\text{R}} &= \int_0^1 \rho(x)\,\mathrm{d}x \\ &= \sum_{i=2}^{\infty} \frac{\rho_i}{i}\,.\end{aligned}$$

Replacing (63) with the above equality and following the same steps as above, one obtains an upper bound on the fraction of edges connected to parity-check nodes of a given degree i. The asymptotic behavior of the resulting upper bound on ρ_i is similar to the upper bound on λ_i as given in (68). However, a tighter upper bound on the fraction of edges connected to parity-check nodes of degree i is derived from the equality

$$\rho_i = \frac{i\,\Gamma_i}{a_\text{R}}\,. \tag{69}$$

Substituting (11) and (59) in the above equality, we get

$$\rho_i \leq \frac{\varepsilon\,C + h_2(P_\text{b})}{1-C}\;\frac{\ln\left(\frac{1}{g_1}\right)}{2\ln\left(\frac{1}{1-2h_2^{-1}\left(\frac{1-C-h_2(P_\text{b})}{1-(1-\varepsilon)C}\right)}\right)}\;\frac{i}{1-h_2\left(\frac{1-g_1^{i/2}}{2}\right)}\,. \tag{70}$$

Applying (67) to the denominator of the second term on the RHS of (70) gives

$$\begin{aligned}\rho_i &\leq \frac{\varepsilon\,C + h_2(P_\text{b})}{1-C}\;\frac{\ln\left(\frac{1}{g_1}\right)}{\ln\left(\frac{1}{2\ln 2}\;\frac{1}{1-\frac{1-C-h_2(P_\text{b})}{1-(1-\varepsilon)C}}\right)}\;\frac{i}{1-h_2\left(\frac{1-g_1^{i/2}}{2}\right)} \\ &= \frac{\varepsilon\,C + h_2(P_\text{b})}{1-C}\;\frac{\ln\left(\frac{1}{g_1}\right)}{\ln\left(\frac{1}{2\ln 2}\;\frac{1-(1-\varepsilon)C}{\varepsilon C+h_2(P_\text{b})}\right)}\;\frac{i}{1-h_2\left(\frac{1-g_1^{i/2}}{2}\right)} \\ &\leq \frac{\ln\left(\frac{1}{g_1}\right)}{1-C}\;\frac{\varepsilon\,C + h_2(P_\text{b})}{\ln\left(\frac{1}{\varepsilon\,C+h_2(P_\text{b})}\right)+\ln\left(\frac{1-C}{2\ln 2}\right)}\;\frac{i}{1-h_2\left(\frac{1-g_1^{i/2}}{2}\right)}\,.\end{aligned}$$

This proves the statement in (21) regarding the fraction of edges connected to parity-check nodes of an arbitrary finite degree i.

When the communication takes place over the BEC, we substitute (13) and (62) in (69) to get

$$\rho_i \leq \frac{i\big[\varepsilon\,(1-p)+P_\text{b}\big]}{\big(1-(1-\varepsilon)(1-p)\big)(1-p)^i}\;\frac{\ln\left(\frac{1}{1-p}\right)}{\ln\left(1+\frac{p-P_\text{b}}{\varepsilon\,(1-p)+P_\text{b}}\right)}.$$

Applying some simple algebra, similarly to the previous derivations, the statement in (21) is validated even when $h_2(P_\text{b})$ is replaced with P_b.

C. *Proof of Theorem 3*

The average degrees of the variable nodes and the parity-check nodes of a bipartite graph which represents a code from an LDPC ensemble are expressible in terms of the pair of the degree distributions (λ,ρ) of this ensemble, and these average degrees are given, respectively, by $a_\text{L} = \left(\int_0^1 \lambda(x)\mathrm{d}x\right)^{-1}$ and $a_\text{R} = \left(\int_0^1 \rho(x)\mathrm{d}x\right)^{-1}$. Hence, the fraction of degree-2 variable nodes is given by

$$\Lambda_2 = \frac{\lambda_2\,a_\text{L}}{2} = \frac{\lambda_2}{2\int_0^1 \lambda(x)\mathrm{d}x} \tag{71}$$

and the design rate of this ensemble is given by

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

Using the last equation, we rewrite the denominator of (71) as

$$\int_0^1 \lambda(x)\mathrm{d}x = \frac{1}{1-R_{\mathrm{d}}} \int_0^1 \rho(x)\mathrm{d}x\,. \tag{72}$$

According to our assumption, the considered sequence of ensembles achieves vanishing bit error probability under sum-product decoding and hence the stability condition implies that

$$\lambda_2 = \lambda'(0) < \frac{e^r}{\rho'(1)}$$

where r is introduced in (23). Substituting (72) in (71) and applying the inequality above leads to an upper bound on Λ_2 of the form

$$\Lambda_2 < \frac{e^r(1-R_{\mathrm{d}})}{2\rho'(1)\int_0^1 \rho(x)\mathrm{d}x}\,. \tag{73}$$

Relying on the convexity of the function $f(t) = x^t$ for all $x > 0$, Jensen's inequality gives

$$\begin{aligned}
&\int_0^1 \rho(x)\mathrm{d}x \\
&= \int_0^1 \sum_i \rho_i x^{i-1}\mathrm{d}x \\
&\geq \int_0^1 x^{\sum_i \rho_i\,(i-1)}\mathrm{d}x \\
&= \int_0^1 x^{\rho'(1)}\mathrm{d}x \\
&= \frac{1}{\rho'(1)+1}
\end{aligned}$$

which implies that

$$\rho'(1) \geq \frac{1}{\int_0^1 \rho(x)\mathrm{d}x} - 1 = a_{\mathrm{R}} - 1\,. \tag{74}$$

Substituting (74) in (73) and since the ensembles are assumed to achieve a fraction $1-\varepsilon$ of the channel capacity (i.e., $R_{\mathrm{d}} = (1-\varepsilon)C$) under sum-product decoding with vanishing bit error probability then

$$\begin{aligned}
\Lambda_2 &< \frac{e^r(1-R_{\mathrm{d}})}{2}\left(1+\frac{1}{\rho'(1)}\right) \\
&\leq \frac{e^r(1-R_{\mathrm{d}})}{2}\left(1+\frac{1}{a_{\mathrm{R}}-1}\right) \\
&= \frac{e^r(1-C)}{2}\left(1+\frac{\varepsilon C}{1-C}\right)\left(1+\frac{1}{a_{\mathrm{R}}-1}\right).
\end{aligned} \tag{75}$$

Since the RHS of (75) is monotonically decreasing with average right degree, the bound still holds when a_{R} is replaced by a lower bound. For all $m \in \mathbb{N}$, let $P_{\mathrm{b},m}$ designate the average bit error probability of the ensemble $\big(n_m, \lambda(x), \rho(x)\big)$ under the considered decoding algorithm. Applying Theorem 1 and letting $P_{\mathrm{b},m}$ tend to zero gives

$$a_{\mathrm{R}} \geq \frac{2\ln\left(\frac{1}{1-2h_2^{-1}\left(\frac{1-C}{1-(1-\varepsilon)C}\right)}\right)}{\ln\left(\frac{1}{g_1}\right)}\,. \tag{76}$$

The upper bound in (22) follows by substituting (76) in (75).

Turning to consider transmission over the BEC, the improved upper bound on the degree-2 variable nodes follows by substituting the lower bound in (13) (when we let the bit erasure probability to vanish) into (75). Note that for a BEC with erasure probability p, $C = 1 - p$ and $e^r(1 - C) = 1$.

Discussion 5: **[On the tightness of the upper bound** (25) **on the fraction of degree-2 variable nodes for capacity-achieving LDPC ensembles over MBIOS channels]** As a direct consequence of Theorem 4, then in the limit were the gap to capacity vanishes under the sum-product decoding algorithm, the fraction of degree-2 variable nodes of the LDPC ensembles satisfies the upper bound in (25). In the following, we consider the tightness of this upper bound. To this end, we first present the following lemma:

Lemma 5: **[On the asymptotic fraction of degree 2 variable nodes for capacity-achieving sequences of LDPC ensembles]** Let

$$\Big\{\big(n_m, \lambda_m(x) \triangleq \sum_i \lambda_i^{(m)} x^{i-1}, \rho_m(x) \triangleq \sum_i \rho_i^{(m)} x^{i-1}\big)\Big\}_{m\in\mathbb{N}}$$

be a sequence of LDPC ensembles whose transmission takes place over an MBIOS channel, and let C designate the channel capacity (in units of bits per channel use). Assume that this sequence approaches the channel capacity under the sum-product iterative decoding algorithm, and also assume that the *flatness condition is asymptotically satisfied for this sequence* (i.e., let the stability condition be asymptotically satisfied with equality for this capacity-achieving sequence). Let $\Lambda_2^{(m)}$ be the fraction of the variable nodes of degree 2 of the m'th ensemble in this sequence. Also, let $d_{\text{R}}^{(m)}$ be a random variable distributed according to the degree distribution of the parity-check nodes of the m'th ensemble, and let $a_{\text{R}}^{(m)} = \mathbb{E}\,[d_{\text{R}}^{(m)}]$ designate the average right degree of this ensemble. Finally, let us assume that the limit of the ratio between the standard deviation and the expectation of the right degree distribution is finite, i.e.,

$$\lim_{m\to\infty} \frac{\text{std}\big(d_{\text{R}}^{(m)}\big)}{a_{\text{R}}^{(m)}} = K < \infty \tag{77}$$

where $\text{std}\big(d_{\text{R}}^{(m)}\big)$ denotes the standard deviation of the random variable $d_{\text{R}}^{(m)}$. Then, in the limit where m tends to infinity, the fraction of degree-2 variable nodes satisfies

$$\lim_{m\to\infty} \Lambda_2^{(m)} = \frac{e^r(1-C)}{2(1+K^2)} \tag{78}$$

where r is the Bhattacharyya distance, introduced in (23), which only depends on the communication channel.

Proof: By the assumption that the sequence of LDPC ensembles satisfies asymptotically the flatness condition, we get

$$\lim_{m\to\infty} \lambda_2^{(m)}\, \rho'_m(1) = e^r\,. \tag{79}$$

Based on the equality in (71)

$$\Lambda_2^{(m)} = \frac{\lambda_2^{(m)}}{2\displaystyle\int_0^1 \lambda_m(x)\mathrm{d}x}\,, \qquad \forall m \in \mathbb{N} \tag{80}$$

and therefore

$$\begin{aligned}
&\lim_{m\to\infty} \Lambda_2^{(m)} \\
&\stackrel{\text{(a)}}{=} \lim_{m\to\infty} \frac{e^r}{2\rho'_m(1)\displaystyle\int_0^1 \lambda_m(x)\mathrm{d}x} \\
&\stackrel{\text{(b)}}{=} \lim_{m\to\infty} \frac{e^r\,(1-R_m)}{2\rho'_m(1)\displaystyle\int_0^1 \rho_m(x)\mathrm{d}x} \\
&\stackrel{\text{(c)}}{=} \frac{e^r(1-C)}{2} \lim_{m\to\infty} \frac{1}{\rho'_m(1)\displaystyle\int_0^1 \rho_m(x)\mathrm{d}x}
\end{aligned} \tag{81}$$

where the equality in (a) relies on (80) and the requirement in (79), the equality in (b) follows from (1) where R_m designates the design rate of the ensemble (n_m, λ_m, ρ_m), and (c) follows by the assumption that the sequence of LDPC ensembles is capacity-achieving. Based on the convexity of the function $f(t) = x^t$ for all $x \geq 0$, then Jensen's inequality implies that for all $m \in \mathbb{N}$ and non-negative values of x

$$\rho_m(x) = \sum_i \rho_i^{(m)} x^{i-1} \geq x^{\sum_i (i-1)\rho_i^{(m)}} = x^{\rho'_m(1)}\,.$$

Integrating both sides of the inequality over the interval $[0, 1]$ gives

$$\int_0^1 \rho_m(x)\mathrm{d}x \geq \int_0^1 x^{\rho'_m(1)}\mathrm{d}x = \frac{1}{\rho'_m(1)+1}$$

which therefore implies that

$$\rho'_m(1) \geq \frac{1}{\int_0^1 \rho_m(x)\mathrm{d}x} - 1 = a_{\mathrm{R}}^{(m)} - 1\,. \tag{82}$$

From Theorem 1, the asymptotic average right degree of a capacity-achieving sequence of LDPC ensembles tends to infinity as the gap to capacity vanishes. Therefore, (82) implies that $\lim_{m\to\infty} \rho'_m(1) = \infty$. Substituting this into the RHS of (81) gives the equality

$$\lim_{m\to\infty} \Lambda_2^{(m)} = \frac{e^r(1-C)}{2} \lim_{m\to\infty} \frac{1}{\left(\rho'_m(1)+1\right)\int_0^1 \rho_m(x)\mathrm{d}x}\,. \tag{83}$$

Let $\Gamma_i^{(m)}$ be the fraction of parity-check nodes of degree i in the m'th LDPC ensemble, then the following equality holds:

$$\rho_i^{(m)} = \frac{i\,\Gamma_i^{(m)}}{\sum_j j\,\Gamma_j^{(m)}}$$

and

$$\begin{aligned}
&\rho'_m(1)+1\\
&= \sum_i \left\{(i-1)\rho_i^{(m)}\right\} + 1\\
&= \sum_i i\rho_i^{(m)}\\
&= \frac{\sum_i i^2\,\Gamma_i^{(m)}}{\sum_j j\,\Gamma_j^{(m)}}\,.
\end{aligned} \tag{84}$$

The denominator in the RHS of (83) therefore satisfies the following chain of equalities:

$$\begin{aligned}
&\left(\rho'_m(1)+1\right)\int_0^1 \rho_m(x)\mathrm{d}x\\
&= \frac{\sum_i i^2\,\Gamma_i^{(m)}}{\sum_j j\,\Gamma_j^{(m)}} \int_0^1 \rho_m(x)\mathrm{d}x\\
&= \frac{\sum_i i^2\,\Gamma_i^{(m)}}{\sum_j j\,\Gamma_j^{(m)}}\,\frac{1}{a_{\mathrm{R}}^{(m)}}\\
&= \frac{\sum_i i^2\,\Gamma_i^{(m)}}{\left(\sum_j j\,\Gamma_j^{(m)}\right)^2}\\
&= \frac{\mathrm{var}\left(d_{\mathrm{R}}^{(m)}\right)}{\left(a_{\mathrm{R}}^{(m)}\right)^2} + 1
\end{aligned} \tag{85}$$

where $\text{var}\left(d_{\text{R}}^{(m)}\right)$ designates the variance of the random variable $d_{\text{R}}^{(m)}$. Taking the limit where m tends to infinity on both sides of (85), the definition of K in (77) yields that

$$\lim_{m\to\infty} \left(\rho_m'(1)+1\right) \int_0^1 \rho_m(x)\mathrm{d}x = 1+K^2.$$

The proof of this lemma is completed by substituting the last equality in the RHS of (83). ■

Lemma 5 shows that for capacity-achieving sequences of LDPC ensembles for which the following assumptions hold:

- **Condition 1**: The stability condition is asymptotically satisfied with equality as the gap to capacity vanishes (i.e., the flatness condition is asymptotically satisfied)
- **Condition 2**: The limit of the ratio between the standard deviation and the expectation of the right degree distribution is finite (see (77))

then the asymptotic fraction of degree 2 variable nodes is bounded away from zero as the gap to capacity vanishes (under the sum-product decoding algorithm). Moreover, if the limit in Condition 2 is equal to zero (i.e., $K=0$), then the asymptotic fraction of degree-2 variable nodes coincides with the upper bound in (25); this makes the upper bound in (25) tight under the above two conditions for capacity-achieving sequences of LDPC ensembles under the sum-product decoding algorithm.

As an example of a capacity-achieving sequence of LDPC ensembles which satisfies the upper bound in (25) with equality, consider the sequence of right-regular LDPC ensembles as introduced by Shokrollahi (see [17], [24]); this capacity-achieving sequence satisfies the flatness condition on the BEC; also, by definition, $K=0$ for this sequence since the degree of the parity-check nodes is fixed for a right-regular LDPC ensemble. Note that for the BEC, the upper bound on the fraction of degree-2 variable nodes as given in (25) is particularized to one-half regardless of the erasure probability of the BEC.

Remark 9: We note that the property proved in Lemma 5 for the non-vanishing asymptotic fraction of degree-2 *variable nodes* of capacity-achieving sequences of LDPC ensembles is reminiscent of another information-theoretic property which was proved by Shokrollahi with respect to the non-vanishing fraction of degree-2 *output nodes* for capacity-achieving sequences of Raptor codes whose transmission takes place over an MBIOS channel (see [5, Theorem 11 and Proposition 12]). The concepts of the proofs of these two results are completely different, but there is a duality the two results.

D. Proof of Theorem 4 and its Corollary

Since the considered sequence of ensembles achieves vanishing bit error probability under sum-product decoding, the stability condition implies that

$$\lambda_2 = \lambda'(0) < \frac{e^r}{\rho'(1)} \tag{86}$$

where r is given by (23). Substituting (74) in (86) yields

$$\lambda_2 < \frac{e^r}{a_{\text{R}}-1} \tag{87}$$

where a_{R} designates the common average right degree of the sequence of ensembles. The upper bounds on λ_2 in (26) and (27) are obtained by substituting (76) and (43), respectively, in (87).

Discussion 6: **[Comparison between the two upper bounds on the fraction of edges connected to degree-2 variable nodes: ML versus iterative decoding]** In the proof of Theorem 2, we derive an upper bound on the fraction of edges connected to variable nodes of degree i for ensembles of LDPC codes which achieve a bit error (or erasure) probability P_{b} under an arbitrary decoding algorithm (see (66) and the tightened version (68) of this bound for the BEC). Referring to degree-2 variable nodes and letting P_{b} vanish, this bound is particularized to

$$\begin{aligned}\lambda_2 &\leq \frac{\ln\left(\frac{1}{g_1}\right)}{\big(1-(1-\varepsilon)C\big)\ \ln\left(\frac{1}{1-2h_2^{-1}\left(\frac{1-C}{1-(1-\varepsilon)C}\right)}\right)} \\ &\leq \frac{\ln\left(\frac{1}{g_1}\right)}{(1-C)\ \ln\left(\frac{1}{1-2h_2^{-1}\left(\frac{1-C}{1-(1-\varepsilon)C}\right)}\right)} \end{aligned} \tag{88}$$

with the following tightened version for the BEC:

$$\begin{aligned}\lambda_2 &\leq \frac{2\ln\left(\frac{1}{1-p}\right)}{\big(1-(1-\varepsilon)(1-p)\big)\ln\left(1+\frac{p}{\varepsilon(1-p)}\right)} \\ &\leq \frac{2\ln\left(\frac{1}{1-p}\right)}{p\left[\ln\left(1-p+\frac{p}{\varepsilon}\right)+\ln\left(\frac{1}{1-p}\right)\right]}\,. \end{aligned} \tag{89}$$

It is interesting to note on some similarity between the upper bounds on λ_2 as given in (88) and (89) and the corresponding bounds given in (26) and (27). Note that the bounds in (88) and (89) are valid under ML decoding or any other decoding algorithm while the two bounds in (26) and (27) are more restrictive in the sense that they are valid under the sum-product decoding algorithm. These two pairs of bounds are numerically compared in Section V.

Proof of Corollary 5: A truncation of the power series on the LHS of (34) after the first term gives the inequality

$$1-h_2\left(\frac{1-\sqrt{u}}{2}\right) \geq \frac{u}{2\ln 2}\,, \quad 0 \leq u \leq 1.$$

Assigning $u = \big(1-2h_2^{-1}(x)\big)^2$ and rearranging terms gives

$$h_2^{-1}(x) \geq \frac{1}{2}\left(1-\sqrt{2\ln 2\ \ (1-x)}\right), \quad 0 \leq x \leq 1. \tag{90}$$

Assigning $0 \leq x \triangleq \frac{1-C}{1-(1-\varepsilon)C} \leq 1$ in (90) gives

$$\begin{aligned} & h_2^{-1}\left(\frac{1-C}{1-C(1-\varepsilon)}\right) \\ & \geq \frac{1}{2}\left(1-\sqrt{2\ln 2\ \left(\frac{\varepsilon C}{1-(1-\varepsilon)C}\right)}\right) \\ & \geq \frac{1}{2}\left(1-\sqrt{2\ln 2\ \left(\frac{\varepsilon C}{1-C}\right)}\right) \end{aligned}$$

and

$$1-2h_2^{-1}\left(\frac{1-C}{1-(1-\varepsilon)C}\right) \leq \sqrt{2\ln 2\ \left(\frac{\varepsilon C}{1-C}\right)}\,. \tag{91}$$

Substituting (91) in (76) provides the following lower bound on the average right degree of the ensembles:

$$a_{\mathrm{R}} \geq \frac{\ln\left(\frac{1}{2\ln 2}\frac{1-C}{\varepsilon C}\right)}{\ln\left(\frac{1}{g_1}\right)}. \tag{92}$$

As, clearly, the average right degree of an LDPC ensemble is always greater than 1, then it follows from (92) that

$$\begin{aligned} a_{\mathrm{R}}-1 &\geq \left[\frac{\ln\left(\frac{g_1}{2\ln 2}\frac{1-C}{\varepsilon C}\right)}{\ln\left(\frac{1}{g_1}\right)}\right]_+ \\ &= \left[\frac{\ln\left(\frac{g_1}{2\ln 2}\frac{1-C}{C}\right)+\ln\left(\frac{1}{\varepsilon}\right)}{\ln\left(\frac{1}{g_1}\right)}\right]_+\,. \end{aligned} \tag{93}$$

The proof is completed by substituting (93) in (87).

E. Proof of Proposition 1

When the transmission takes place over a BEC whose erasure probability is p, the constant c_2 given in (29) takes the form

$$c_2 = \frac{p}{\ln\left(\frac{1}{1-p}\right)}\ . \tag{94}$$

For $0 < \alpha < 1$, let

$$\begin{aligned}\hat{\lambda}_\alpha(x) &= 1-(1-x)^\alpha = \sum_{k=1}^{\infty}(-1)^{k+1}\binom{\alpha}{k}x^k\,, \quad 0 \leq x \leq 1 \\ \rho_\alpha(x) &= x^{\frac{1}{\alpha}}.\end{aligned} \tag{95}$$

Note that all the coefficients in the power series expansion of $\hat{\lambda}_\alpha$ are positive for all $0 < \alpha < 1$. Let us now define the polynomials $\hat{\lambda}_{\alpha,N}$ and $\lambda_{\alpha,N}$, where $\hat{\lambda}_{\alpha,N}$ is the truncated power series of $\hat{\lambda}_\alpha$, where only the first $N-1$ terms are taken (i.e., $\hat{\lambda}_{\alpha,N}$ is of degree $N-1$), and the polynomial

$$\lambda_{\alpha,N}(x) \triangleq \frac{\hat{\lambda}_{\alpha,N}(x)}{\hat{\lambda}_{\alpha,N}(1)} \tag{96}$$

is normalized so that $\lambda_{\alpha,N}(1) = 1$. The right-regular sequences of LDPC ensembles in [24] are of the form $\big\{\big(n_m, \lambda_{\alpha,N}(x), \rho_\alpha(x)\big)\big\}_{m\geq 1}$ where $0 < \alpha < 1$ and $N \in \mathbb{N}$ are arbitrary parameters which need to be selected properly. Based on the analysis in [22, Theorem 2.3], this sequence achieves a fraction $1-\varepsilon$ of the capacity of the BEC with vanishing bit erasure probability under message-passing decoding when α and N satisfy

$$\frac{1}{N^\alpha} = 1-p \tag{97}$$

and

$$N = \max\left(\left\lceil \frac{1-k_2(p)\,(1-p)\,(1-\varepsilon)}{\varepsilon}\right\rceil, \left\lceil (1-p)^{-\frac{1}{p}}\right\rceil\right) \tag{98}$$

where

$$k_2(p) = (1-p)^{\frac{\pi^2}{6}}\ e^{\left(\frac{\pi^2}{6}-\gamma\right)\,p} \tag{99}$$

and γ is Euler's constant. Combining (95) and (96), and substituting

$$\sum_{k=1}^{N-1}(-1)^{k+1}\binom{\alpha}{k} = 1-\frac{N}{\alpha}\binom{\alpha}{N}(-1)^{N+1}$$

gives

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

Therefore, the fraction of edges adjacent to variable nodes of degree two is given by

$$\lambda_2 = \frac{\alpha}{1-\frac{N}{\alpha}\,(-1)^{N+1}\binom{\alpha}{N}}\ . \tag{100}$$

We now obtain upper and lower bounds on λ_2. From [22, Eq. (67)] we have that

$$\frac{c(\alpha,N)}{N^\alpha} < \frac{N}{\alpha}(-1)^{N+1}\binom{\alpha}{N} \leq \frac{1}{N^\alpha} \tag{101}$$

where

$$c(\alpha,N) = (1-\alpha)^{\frac{\pi^2}{6}}\ e^{\alpha\left(\frac{\pi^2}{6}-\gamma+\frac{1}{2N}\right)}\ . \tag{102}$$

Substituting (101) in (100) and using (97), we get

$$\frac{\alpha}{1-c(\alpha,N)\,(1-p)} < \lambda_2 \leq \frac{\alpha}{1-(1-p)} = \frac{\alpha}{p}\,. \tag{103}$$

Under the parameter assignments in (97) and (98), the parameters N and α satisfy

$$\alpha = \frac{\ln\left(\frac{1}{1-p}\right)}{\ln N} \tag{104}$$

$$N \geq \frac{1-(1-p)\,k_2(p)}{\varepsilon}\,. \tag{105}$$

Substituting (104) and (105) into the right inequality of (103) gives an upper bound on λ_2 which takes the form

$$\begin{aligned}\lambda_2 &\leq \frac{\alpha}{p}\\ &\leq \frac{\ln\left(\frac{1}{1-p}\right)}{p\,\ln\left(\frac{1-(1-p)\,k_2(p)}{\varepsilon}\right)}\\ &= \frac{\ln\left(\frac{1}{1-p}\right)}{p\Big[\ln\frac{1}{\varepsilon}+\ln\big(1-(1-p)\,k_2(p)\big)\Big]}\\ &= \frac{1}{c_3+c_2\ln\frac{1}{\varepsilon}}\end{aligned} \tag{106}$$

where c_2 is the coefficient of the logarithmic growth rate in $\frac{1}{\varepsilon}$, which coincides here with (29), and

$$c_3 \triangleq \frac{p\,\ln\big(1-(1-p)\,k_2(p)\big)}{\ln\left(\frac{1}{1-p}\right)} \tag{107}$$

is a constant which only depends on the BEC. We turn now to derive a lower bound on λ_2 for the asymptotic case where the gap to capacity vanishes. From (98), we have that for small enough values of ε, the parameter N satisfies

$$\begin{aligned}N &= \left\lceil\frac{1-k_2(p)\,(1-p)\,(1-\varepsilon)}{\varepsilon}\right\rceil\\ &\leq \frac{1-k_2(p)\,(1-p)\,(1-\varepsilon)}{\varepsilon}+1\\ &= \frac{1-k_2(p)\,(1-p)\,\left(1-\varepsilon\Big(1-\frac{1}{k_2(p)\,(1-p)}\Big)\right)}{\varepsilon}\,.\end{aligned} \tag{108}$$

Substituting (104) and (108) into the left inequality of (103), we get

$$\begin{aligned}\lambda_2 &> \frac{\alpha}{p}\,\frac{p}{1-c(\alpha,N)\,(1-p)}\\ &\geq \frac{\ln\left(\frac{1}{1-p}\right)}{p\,\ln\left(\frac{1-k_2(p)\,(1-p)\,\left(1-\varepsilon\left(1-\frac{1}{k_2(p)\,(1-p)}\right)\right)}{\varepsilon}\right)}\\ &\quad\cdot\frac{p}{1-c(\alpha,N)\,(1-p)}\\ &= \frac{\ln\left(\frac{1}{1-p}\right)}{p\left[\ln\left(\frac{1}{\varepsilon}\right)+\ln\big(1-(1-p)\,k_2(p)\big)+\ln(1-\varepsilon)\right]}\\ &\quad\cdot\frac{p}{1-(1-p)\,c(\alpha,N)}\\ &= \frac{1}{c_3+c_2\ln\left(\frac{1}{\varepsilon}\right)+\widetilde{\varepsilon}(\varepsilon,p)}\;\;\frac{p}{1-(1-p)\,c(\alpha,N)}\end{aligned} \tag{109}$$

where c_2 is the coefficient of the logarithm in the denominator of (28) and it coincides with (94), c_3 is given in (107), and

$$\widetilde{\varepsilon}(\varepsilon, p) = \frac{p \ln(1-\varepsilon)}{\ln\left(\frac{1}{1-p}\right)} \tag{110}$$

which therefore implies that for $0 \leq p < 1$

$$\lim_{\varepsilon \to 0} \widetilde{\varepsilon}(\varepsilon, p) = 0. \tag{111}$$

Using the lower bound on the parameter N in (105), then in the limit where ε vanishes, N tends to infinity (this holds since $1-(1-p)k_2(p) > 0$ for all $0 < p < 1$ where k_2 in introduced in (99)). Also, from (98) and (104), we get

$$\lim_{\varepsilon \to 0} \alpha = 0$$

which, from (102), yields that

$$\lim_{\varepsilon \to 0} c(\alpha, N) = 1\,. \tag{112}$$

Substituting (111) and (112) in (109) yields that in the limit where the gap to capacity vanishes (i.e., $\varepsilon \to 0$), the upper and lower bounds on λ_2 in (106) and (109) coincide. Specifically, we have shown that

$$\lim_{\varepsilon \to 0} \lambda_2(\varepsilon) \cdot c_2 \, \ln\left(\frac{1}{\varepsilon}\right) = 1\,.$$

Therefore, as $\varepsilon \to 0$, the upper bound on $\lambda_2 = \lambda_2(\varepsilon)$ in Corollary 5 becomes tight for the sequence of right-regular LDPC ensembles in [24] with the parameters chosen in (97) and (98). We note that the setting of the parameters N and α in (97) and (98) is identical to [22, p. 1615].

V. Numerical Results

In this section, we consider sequences of LDPC ensembles which achieve vanishing bit error probability and closely approach the channel capacity limit under sum-product decoding. As representatives of MBIOS channels, the considered communication channels are the binary erasure channel (BEC), binary symmetric channel (BSC) and the binary-input AWGN channel (BIAWGNC).

Example 2: **[BEC]** Consider a sequence of LDPC ensembles (n, λ, ρ) where the block length (n) tends to infinity and the pair of degree distributions is given by

$$\begin{aligned}
&\lambda(x) = 0.409x + 0.202x^2 + 0.0768x^3 + 0.1971x^6 + 0.1151x^7 \\
&\rho(x) = x^5.
\end{aligned}$$

The design rate of this ensemble is $R = 0.5004$, and the threshold under iterative message-passing decoding is equal to

$$p^{\mathrm{IT}} = \inf_{x \in (0,1]} \frac{x}{\lambda\big(1-\rho(1-x)\big)} = 0.4810$$

so the corresponding channel capacity of the BEC is $C = 1 - p^{\mathrm{IT}} = 0.5190$ bits per channel use, and the multiplicative gap to capacity is $\varepsilon = 1 - \frac{R}{C} = 0.0358$. The lower bound on the average right degree in (13) with vanishing bit erasure probability (i.e., $P_{\mathrm{b}} = 0$) gives that the average right degree should be at least 5.0189, and practically, since we consider here LDPC ensembles with fixed right degree then the right degree cannot be below 6. Hence, the lower bound is attained in this case with equality. An upper bound on the fraction of edges which are connected to degree-2 variable nodes (λ_2) is readily calculated based on (87) with $e^r = (p^{\mathrm{IT}})^{-1} = 2.0790$ and the above lower bound on a_{R} (for LDPC ensembles of a fixed right degree) which is equal to 6; this gives $\lambda_2 \leq 0.4158$ as compared to the exact value which is equal to 0.409. The exact value of the fraction of degree-2 variable nodes is

$$\Lambda_2 = \frac{\lambda_2 \, a_{\mathrm{L}}}{2} = \frac{\lambda_2 \, (1-R) \, a_{\mathrm{R}}}{2} = 0.6130$$

as compared to the upper bound in (75), combined with the tight lower bound $a_{\mathrm{R}} \geq 6$, which gives $\Lambda_2 \leq 0.6232$.

Example 3: **[BIAWGNC]** Table I considers two sequences of LDPC ensembles of design rate $\frac{1}{2}$ which are taken from [4, Table II]. The pair of degree distributions of the ensembles in each sequence is fixed and the block length of these ensembles tends to infinity. The LDPC ensembles in each sequence are specified by the following pairs of degree distributions:

Ensemble 1:

$$\begin{aligned}\lambda(x) &= 0.170031x + 0.160460x^2 + 0.112837x^5 \\ &\quad +0.047489x^6 + 0.011481x^9 + 0.091537x^{10} \\ &\quad +0.152978x^{25} + 0.036131x^{26} + 0.217056x^{99} \\ \rho(x) &= \frac{1}{16}\,x^9 + \frac{15}{16}\,x^{10}\end{aligned}$$

Ensemble 2:

$$\begin{aligned}\lambda(x) &= 0.153425x + 0.147526x^2 + 0.041539x^5 + \\ &\quad 0.147551x^6 + 0.047938x^{17} + 0.119555x^{18} \\ &\quad +0.036379x^{54} + 0.126714x^{55} + 0.179373x^{199} \\ \rho(x) &= x^{11}\end{aligned}$$

The asymptotic thresholds of the considered LDPC ensembles under iterative sum-product decoding are calculated with the density evolution technique when the transmission is assumed to take place over the BIAWGNC; these calculations provide the indicated gaps to capacity as given in Table I. The value of λ_2 for each sequence of LDPC ensembles (where we let the block length tend to infinity) is compared with the upper bound given in Theorem 4 which holds for any sequence of LDPC ensembles whose bit error probability vanishes under sum-product decoding with a certain gap (in rate) to capacity. Note that for calculating the bound in Theorem 4, the parameter r introduced in (23) gets the form $r = \frac{RE_{\text{b}}}{N_0}$ for the BIAWGNC where $\frac{E_{\text{b}}}{N_0}$ designates the energy per information bit over the one-sided noise spectral density, and we substitute here the threshold value of $\frac{E_{\text{b}}}{N_0}$ under the sum-product decoding. The average right degree of each sequence is also compared with the lower bound in Theorem 1. These comparisons exemplify that for the examined LDPC ensembles, both of the theoretical bounds are informative. It is noted that

LDPC ensemble	Gap to capacity (ε)	a_{R}	Lower bound on a_{R} (Theorem 1)	λ_2	Upper bound on λ_2 (Theorem 4)
1	$3.72 \cdot 10^{-3}$	10.938	9.249	0.170	0.205
2	$2.22 \cdot 10^{-3}$	12.000	10.134	0.153	0.185

TABLE I

COMPARISON OF THEORETICAL BOUNDS AND ACTUAL VALUES OF λ_2 AND a_{R} FOR TWO SEQUENCES OF LDPC ENSEMBLES OF DESIGN RATE $\frac{1}{2}$ TRANSMITTED OVER THE BIAWGNC. THE SEQUENCES ARE TAKEN FROM [4, TABLE II] AND ACHIEVE VANISHING BIT ERROR PROBABILITY UNDER SUM-PRODUCT DECODING WITH THE INDICATED GAPS TO CAPACITY.

the degree distributions of the examined LDPC ensembles were obtained in [27] by numerical search, based on the density evolution technique, where the goal was to minimize the gap to capacity without taking into consideration the values of λ_2 or a_{R}. Therefore, it might be possible to construct LDPC ensembles which achieve the same gap to capacity with values of λ_2 and a_{R} which are even closer to the theoretical bounds.

Example 4: **[BSC]** Table II considers two sequences of LDPC ensembles, taken from [27], where the pair of degree distributions of the ensembles in each sequence is fixed and the block length of these ensembles tends to infinity. The LDPC ensembles in each sequence are specified by the following pairs of degree distributions and design rates:

Ensemble 1:

$$\begin{aligned}\lambda(x) &= 0.291157x + 0.189174x^2 + 0.0408389x^4 \\ &\quad +0.0873393x^5 + 0.00742718x^6 + 0.112581x^7 \\ &\quad +0.0925954x^{15} + 0.0186572x^{20} + 0.124064x^{32} \\ &\quad +0.016002x^{39} + 0.0201644x^{44} \\ \rho(x) &= 0.8x^4 + 0.2x^5 \\ R &= 0.250\end{aligned}$$

Ensemble 2:

$$\begin{aligned}\lambda(x) &= 0.160424x + 0.160541x^2 + 0.0610339x^5 \\ &\quad +0.153434x^6 + 0.0369041x^{12} + 0.020068x^{15} \\ &\quad +0.0054856x^{16} + 0.128127x^{19} + 0.0233812x^{24} \\ &\quad +0.05285542x^{34} + 0.0574104x^{67} + 0.0898442x^{68} \\ &\quad +0.0504923x^{85} \\ \rho(x) &= x^{10} \\ R &= 0.500\end{aligned}$$

The thresholds of the considered LDPC ensembles under iterative sum-product decoding are calculated with the density evolution technique when the transmission takes place over the BSC. These thresholds under sum-product decoding, which correspond to the asymptotic case where we let the block length tend to infinity, provide the indicated gaps to capacity in Table II. The value of λ_2 for each sequence is compared with the upper bound given in Theorem 4. Note that for calculating the bound in Theorem 4, the parameter r introduced in (23) satisfies $e^r = \frac{1}{\sqrt{4p(1-p)}}$ for the BSC whose crossover probability is equal to p, and we substitute here the threshold value of p under the sum-product decoding. Also, for the calculation of this bound for such a BSC, Eq. (51) gives that $g_1 = (1-2p)^2$. The average right degree of each sequence is also compared with the lower bound in Theorem 1. These comparisons show that for the considered sequences of LDPC ensembles, both of the theoretical bounds are fairly tight; the upper bound on λ_2 is within a factor of 1.3 from the actual value for the two sequences of LDPC ensembles while the lower bound on the average right degree is not lower than 83% of the corresponding actual values. As with the LDPC ensembles in Table I, the ensembles referred to in Table II were obtained by the density

LDPC ensemble	Gap to capacity (ε)	a_{R}	Lower bound on a_{R} (Theorem 1)	λ_2	Upper bound on λ_2 (Theorem 4)
1	$1.85 \cdot 10^{-2}$	5.172	4.301	0.291	0.371
2	$6.18 \cdot 10^{-3}$	11.000	9.670	0.160	0.185

TABLE II

COMPARISON OF THEORETICAL BOUNDS AND ACTUAL VALUES OF λ_2 AND a_{R} FOR TWO SEQUENCES OF LDPC ENSEMBLES TRANSMITTED OVER THE BSC. THE SEQUENCES ARE TAKEN FROM [27] AND ACHIEVE VANISHING BIT ERROR PROBABILITY UNDER ITERATIVE SUM-PRODUCT DECODING WITH THE INDICATED GAPS TO CAPACITY.

evolution technique with the goal of minimizing the gap to capacity under a constraint on the maximal degree.

Example 5: **[On the fundamental system of cycles for capacity-approaching sequences of LDPC ensembles]** Corollary 2 considers an arbitrary sequence of LDPC ensembles, specified by a pair of degree distributions, whose transmission takes place over an MBIOS channel. This corollary refers to the asymptotic case where we let the block length of the ensembles in this sequence tend to infinity and the bit error (or erasure) probability vanishes; the design rate of these ensembles is assumed to be a fraction $1-\varepsilon$ of the channel capacity (for an arbitrary $\varepsilon \in (0,1)$). In Corollary 2, Eq. (16) applies to a general MBIOS channel and a tightened version of this bound is given in (17) for the BEC. Based on these results, the asymptotic average cardinality of the fundamental system of cycles for Tanner graphs representing codes from LDPC ensembles as above, where this average cardinality is

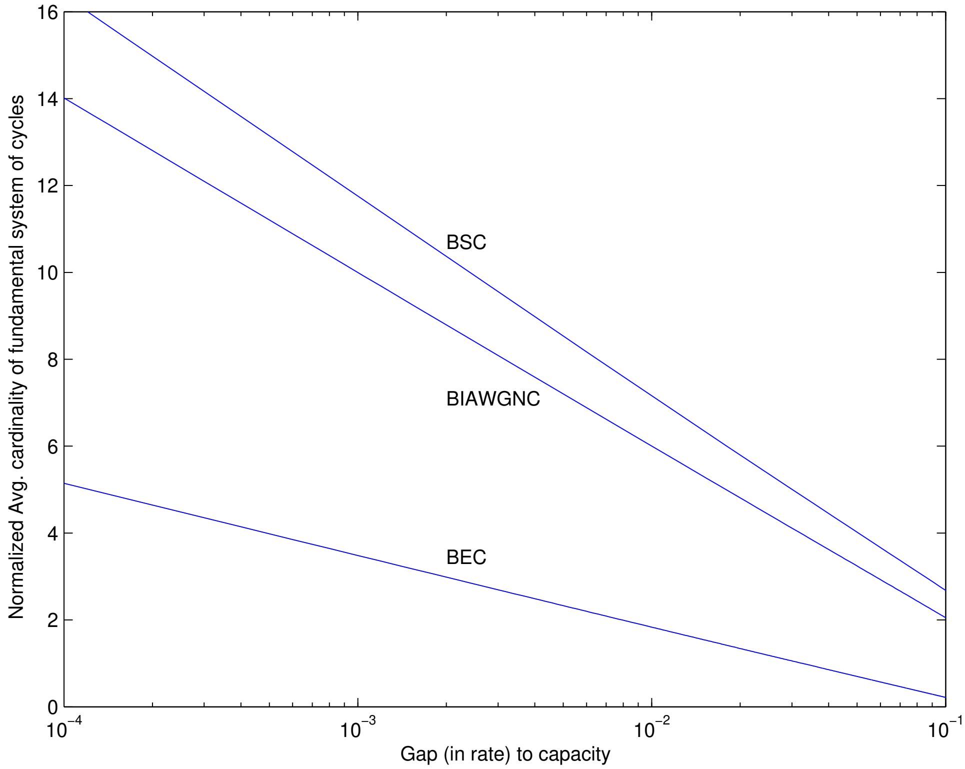

Fig. 3. Plot of the asymptotic lower bounds in Corollary 2 (see Eqs. (16) and (17)) for three memoryless binary-input output-symmetric (MBIOS) channels. These lower bounds correspond to the average cardinality of the fundamental system of cycles for Tanner graphs representing codes from an arbitrary LDPC ensemble; the above quantity is normalized w.r.t. the block length of the ensemble, and the asymptotic result shown in this figure refers to the case where we consider a sequence of LDPC ensembles whose block lengths tend to infinity (according to the statement in Corollary 2, the degree distributions of the sequence of LDPC ensembles are assumed to be fixed and the block lengths of these ensembles tend to infinity). The bounds are plotted versus the achievable gap (in rate) between the channel capacity and the design rate of the LDPC ensembles. This figure shows the bounds for the binary symmetric channel (BSC), binary-input AWGN channel (BIAWGNC) and the binary erasure channel (BEC) where it is assumed that the design rate of the LDPC ensembles is fixed in all cases and is equal to one-half bit per channel use.

normalized w.r.t. the block length, grows at least like $\log \frac{1}{\varepsilon}$. We consider here the BSC, BEC, and BIAWGNC as three representatives of the class of MBIOS channels, and assume that the design rate of the LDPC ensembles is fixed to one-half bit per channel use. It is shown in Fig. 3 that for a given gap (ε) to the channel capacity and for a fixed design rate, the extreme values of this lower bounds correspond to the BSC and BEC (which attain the maximal and minimal values, respectively). This observation is consistent with the last part of the statement in Corollary 2.

VI. Summary and Outlook

This paper considers properties related to the degree distributions of capacity-approaching LDPC ensembles, and to the graphical complexity and the average cardinality of the fundamental system of cycles of their Tanner graphs. Universal information-theoretic bounds which are related to these properties are derived when the transmission of these LDPC ensembles takes place over an arbitrary memoryless binary-input output-symmetric (MBIOS) channel. The universality of the bounds derived in this paper stems from the fact that they do not depend on the full characterization of the LDPC ensembles but rather depend on the achievable gap between the channel capacity and the design rate of the ensemble. Some of these bounds are also expressed in terms of the bit error (or erasure) probability and the block length of the ensembles, and several other bounds refer to the asymptotic case where we let the block length tend to infinity and the bit error probability vanishes. The first category of the results introduced in this paper provides the following bounds which hold under maximum-likelihood (ML) decoding (and hence, they also hold under any sub-optimal decoding algorithm):

- Theorem 1 provides a lower bound on the average degree of the parity-check nodes of a binary linear block code where it is assumed that the code is represented by an arbitrary bipartite (Tanner) graph which corresponds to a full-rank parity-check matrix. A lower bound on the average right degree of the Tanner graph is related to the graphical complexity of the representation of a binary linear block code by such a bipartite graph, and also to the decoding complexity per iteration which is associated with a message-passing iterative decoder operating on this graph. The bound in Theorem 1 applies to finite-length binary linear block codes, and it is expressed in terms of the required bit error probability of the code under ML decoding or any other decoding algorithm. This theorem is adapted to hold for LDPC ensembles while relaxing the rather heavy requirement of full-rank parity-check matrices for codes from an LDPC ensemble, and addressing instead the design rate of this ensemble with its corresponding gap to capacity (see Discussion 2 in Section IV). The adaptation of Theorem 1 is also considered for randomly and intentionally punctured LDPC ensembles (see Discussion 3 in Section IV). One of the implications of this theorem is introduced in Corollary 2 which provides asymptotic lower bounds on the average cardinality of the fundamental system of cycles of LDPC ensembles in terms of the capacity of the communication channel and the achievable gap to capacity. All of these lower bounds grow like the logarithm of the inverse of this gap (in rate) to capacity (see also [9], [22], [23], [25]), and they provide quantitative measures of the graphical complexity and the average number of fundamental cycles of capacity-approaching LDPC ensembles in terms of the gap between the channel capacity and their design rate. These bounds demonstrate that the above quantities become unbounded as the gap to capacity vanishes, and they also provide a quantitative tradeoff between the performance of these ensembles, and the graphical complexity and average cardinality of the fundamental system of cycles of their Tanner graphs (which are important parameters for studying the performance and decoding complexity of iterative message-passing decoders).
- Theorem 2 provides upper bounds on the degree distributions of capacity-approaching LDPC ensembles (both from the node and the edge perspectives), and it addresses the behavior of these degree distributions for any finite degree in terms of the achievable gap to capacity. We discuss the implication of this theorem later.

The second category of results is specialized to iterative message-passing decoding algorithms (addressing in particular the sum-product decoding algorithm). The derivation of these results is based on the proofs of the above bounds combined with the stability condition. Theorems 3 and 4 provide upper bounds on the fraction of degree-2 variable nodes and the fraction of the edges connected to these nodes for LDPC ensembles; these bounds are expressed in terms of the achievable gap to capacity of these ensembles under the sum-product decoding algorithm where it is assumed that the block length of these ensembles tends to infinity and the bit error (or erasure) probability vanishes. A byproduct of Theorems 3 and 4 and Lemma 5 is that while the fraction of degree-2 variable nodes stays positive (under some mild conditions) for capacity-approaching LDPC ensembles, the fraction of edges connected to degree-2 variable nodes vanishes, and it is upper bounded by the inverse of the logarithm of the achievable gap to capacity. The tightness of the bounds in Theorems 1–4 is exemplified for some capacity-approaching LDPC ensembles under sum-product decoding (see Section V). These bounds are shown to be reasonably tight for general MBIOS channels, and are particularly tight for the binary erasure channel (BEC).

It is interesting to note that there is a fundamental property which is in common and another property which distinguishes the behavior of the degree distribution of the variable nodes from the degree distribution of the parity-check nodes for capacity-approaching LDPC ensembles under iterative message-passing decoding. The property which is in common relies on Theorem 1 and its adaptation to LDPC ensembles in Discussion 2. First note that the behavior of the average left and right degrees is similar in the sense that their ratio depends on the design rate of the ensemble (so it depends on the channel capacity for sequences of capacity-achieving LDPC ensembles, and is positive as long as the channel capacity is less than 1 bit per channel use). According to Discussion 2, the average left and right degrees of an LDPC ensemble scale at least like the logarithm of the inverse of the gap (in rate) to capacity, and therefore become unbounded for capacity-achieving sequences of LDPC ensembles (even under ML decoding). The other property which distinguishes between the degree distributions of these two kinds of nodes is related to the fact that under some mild conditions, the fraction of degree-2 variable nodes stays strictly positive as the gap to capacity vanishes under the sum-product decoding algorithm (see Lemma 5 and Remark 9) whereas, according to Theorem 2, the fraction of parity-check nodes of an arbitrary finite degree vanishes for capacity-achieving LDPC ensembles. This observation conforms with the behavior of the optimized pairs of the degree distributions of capacity-approaching LDPC ensembles for various MBIOS channels (see, e.g.,

[27]). More explicitly, the degrees of the variable nodes, which are obtained numerically in [27] via the density evolution technique, span over a large range (i.e., the degrees of the variable nodes are distributed between 2 and the maximal allowed degree) whereas the parity-check nodes are almost right concentrated (i.e., the degrees of the parity-check nodes are almost fixed). This observation also conforms with the behavior of the right-regular capacity-achieving sequence of LDPC ensembles over the BEC [24].

In the following, we gather what we consider to be the most interesting open problems which are related to this research:

- The asymptotic bounds in Corollary 2 address the average cardinality of the fundamental system of cycles for Tanner graphs representing LDPC ensembles where the results are directly linked to the average right degree of these ensembles. Further study of the possible link between the statistical properties of the degree distributions of capacity-approaching LDPC ensembles and some other graphical properties related to the cycles in the Tanner graphs of these ensembles is of interest.
- The derivation of universal bounds on the number of iterations and the decoding complexity of code ensembles defined on graphs, measured in terms of the achievable gap (in rate) to capacity, is of theoretical and practical interest. In a currently ongoing work, this issue is addressed for the particular case of the BEC where we let the block length of these ensembles tend to infinity.
- The lower bound on the conditional entropy which is given in (8) plays a key role in the derivation of Theorem 1 and accordingly, it is also crucial for the derivation of the other results in this paper. As opposed to this information-theoretic bound, the bounding techniques presented in [14] and [15] rely on statistical physics, and therefore do not provide a bound on the conditional entropy which is valid for every binary linear block code from the considered ensembles (for a survey paper which provides an introduction to codes defined on graphs and highlights connections with statistical physics, the reader is referred to [16]). It would be interesting to get some theory that unifies the information-theoretic and statistical physics approaches and provides bounds that are tight on the average and valid code by code.
- Extension of the results in this paper to channels with memory (e.g., finite-state channels) is of interest. In this respect, the reader is referred to [9] which considers information-theoretic bounds on the achievable rates of LDPC ensembles for a class of finite-state channels.

Acknowledgment

A discussion with Gil Wiechman is acknowledged.

REFERENCES

[1] M. Ardakani, B. Smith, W. Yu and F. R. Kschischang, “Complexity-optimized low-density parity-check codes,” *Proceedings of the Forty-Third Annual Allerton Conference on Communication, Control and Computing*, pp. 45–54, Urbana-Champaign, Illinois, USA, September 28–30, 2005.

[2] C. Berrou, A. Glavieux and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding,” *Proceedings 1993 IEEE International Conference on Communications (ICC’93)*, pp. 1064–1070, Geneva, Switzerland, May 1993.

[3] 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.

[4] S. Y. Chung, G. D. Forney, T. J. Richardson, and R. Urbanke, “On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit,” *IEEE Communications Letters*, vol. 5, no. 2, pp. 58–60, February 2001.

[5] O. Etesami and A. Shokrollahi, “Raptor codes on binary memoryless symmetric channels,” *IEEE Trans. on Inofrmation Theory*, vol. 52, no. 5, pp. 2033–2051, May 2006.

[6] 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.

[7] R. G. Gallager, “Low-density parity-check codes,” *IEEE Trans. on Information Theory*, vol. 8, no. 1, pp. 21–28, January 1962.

[8] J. Gross and J. Yellen, *Graph Theory and its Applications*, CRC Press Series on Discrete Mathematics and its Applications, 1999.

[9] P. Grover and A. K. Chaturvedi, "Upper bounds on the rate of LDPC codes for a class of finite-state Markov channels," *IEEE Trans. on Information Theory*, vol. 53, no. 2, pp. 794–804, February 2007.

[10] H. Jin and T. Richardson, "A new fast density evolution," *Proceedings 2006 IEEE Information Theory Workshop*, Punta del Este, Uruguay, pp. 183–187, March 13–17, 2006.

[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*, p. 1, Washington, D.C., USA, June 2001.

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

[13] D. J. C. MacKay and R. M. Neal, "Near Shannon limit perofrmance of low-density parity-check codes," *IEE Electronic Letters*, vol. 32, no. 18, pp. 1545–1646, August 1996.

[14] N. Macris, "Giffith-Kelly-Sherman inequalities: a useful tool in the theory of error-correcting codes," *IEEE Trans. on Information Theory*, vol. 53, no. 2, pp. 664–683, February 2007.

[15] A. Montanari, "Tight bounds for LDPC and LDGM codes under MAP decoding," *IEEE Trans. on Information Theory*, vol. 51, no. 9, pp. 3221–3246, September 2005.

[16] A. Montanari and R. Urbanke, "Modern coding theory: The statistical mechanics and computer science point of view," Summer School on Complex Systems, Les Houches, France, July 2006. [Online]. Available: http://www.arxiv.org/abs/0704.2857.

[17] 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.

[18] 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.

[19] 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.

[20] T. Richardson and R. Urbanke, "The renaissance of Gallager's low-density parity-check codes," *IEEE Communications Magazine*, vol. 41, no. 8, pp. 126–131, August 2003.

[21] T. Richardson and R. Urbanke, *Modern Coding Theory*. [Online]. Available: http://lthcwww.epfl.ch/mct/index.php.

[22] I. Sason and R. Urbanke, "Parity-check density versus performance of binary linear block codes over memoryless symmetric channels," *IEEE Trans. on Information Theory*, vol. 49, no. 7, pp. 1611–1635, July 2003.

[23] I. Sason and G. Wiechman, "On achievable rates and complexity of LDPC codes over parallel channels: Bounds and applications," *IEEE Trans. on Information Theory*, vol. 53, no. 2, pp. 580–598, July 2007.

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

[25] G. Wiechman and I. Sason, "Parity-check density versus performance of binary linear block codes: New bounds and applications," *IEEE Trans. on Information Theory*, vol. 53, no. 2, pp. 550–579, February 2007.

[26] Special issue on *Codes on Graphs and Iterative Algorithms*, *IEEE Trans. on Information Theory*, vol. 47, no. 2, February 2001.

[27] Optimization of the degree distrbutions of LDPC ensembles. [Online]. Available: http://lthcwww.epfl.ch/research/ldpcopt/index.php.

[28] *Big-O notation*. [Online]. Available: http://www.nist.gov/dads/HTML/bigOnotation.html.