Extremal Problems of Information Combining

Yibo Jiang, Alexei Ashikhmin, Ralf Koetter, Andrew C. Singer

***Abstract*— In this paper we study moments of soft-bits of binary-input symmetric-output channels and solve some extremal problems of the moments. We use these results to solve the extremal information combining problem. Further, we extend the information combining problem by adding a constraint on the second moment of soft-bits, and find the extremal distributions for this new problem. The results for this extension problem are used to improve the prediction of convergence of the belief propagation decoding of low-density parity-check codes, provided that another extremal problem related to the variable nodes be solved.**

***Index Terms*— EXIT functions, Information combining, LDPC codes.**

I. Introduction

Low-density parity-check (LDPC) codes were invented by Gallager [1] in 1960. In recent years, there has been surging interest in LDPC codes. A lot of work has been done in areas such as asymptotic performance analysis, capacity-approaching code design, and low complexity decoding algorithms for LDPC codes (see [2]). Among various low complexity decoding algorithms for LDPC codes, the belief propagation decoding algorithm is a popular and powerful one [3]. It is well known that an LDPC code can be represented by a Tanner graph [4], which is a bipartite graph consisting of variable nodes and check nodes. A variable node corresponds to a codeword bit, while a check node corresponds to a row of the check matrix. In belief propagation decoding, each node on the Tanner graph works as a local decoder. More specifically, at a degree d variable node, the local variable node decoder (VND) is essentially an extrinsic *a posteriori* probability (APP) decoder for a $[d+1, 1]$ repetition code (remember there is an incoming channel message for each variable node). At a degree d check node, the local check node decoder (CND) is essentially an extrinsic APP decoder for a $[d, d-1]$ single parity-check code. Overall, the belief propagation decoder works iteratively. In each iteration, first, every VND uses the channel message and incoming check-to-variable messages (from the previous iteration) to compute the outgoing variable-to-check messages, then every CND uses the incoming variable-to-check messages to compute the outgoing check-to-variable messages.

Y. Jiang was with Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. He is with Qualcomm, 5775 Morehouse Drive, San Diego, CA 92121, USA (email: yjiang@qualcomm.com)

R. Koetter was with Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. He is with Institute for Communications Engineering, Technische Universitaet Muenchen, D-80290 Muenchen. (email: ralf.koetter@tum.de)

A. C. Singer is with Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA (email: acsinger@uiuc.edu)

A. Ashikhmin is with Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, NJ 07974, USA (email: aea@research.bell-labs.com)

Density evolution was proposed in [3], [5], which rigorously analyzes the convergence behavior of belief propagation decoding. A one dimensional approximation method of density evolution, the extrinsic information transfer (EXIT) chart, originally suggested by ten Brink [6] for analyzing and designing turbo codes, was a simpler method for convergence prediction. In [7], [8], the EXIT chart method was extended for the analysis and design of LDPC codes, and a rigorous analysis of the method was conducted when the *a priori* channel is a binary erasure channel (BEC). VND and CND EXIT functions are defined to characterize how the mutual information contained in messages changes during the variable and check nodes processing at each iteration of the belief propagation decoding. In the computation of VND and CND EXIT functions, every incoming message to a VND or a CND is usually modelled as the output log-likelihood-ratio (LLR) of an additive white Gaussian noise (AWGN) channel whose input is the binary phase shift keying (BPSK) modulated codeword bit corresponding to the message [9]. The EXIT chart method has proven useful in practice, see, e.g., [9], and simple since it tracks only one parameter, namely the mutual information. However, since the distributions of messages are often non-Gaussian, it only provides an approximation to the real decoding trajectory. Thus it is worthwhile to conduct a more rigorous analysis. Bounds on the extrinsic mutual information at either a VND or a CND will be useful and lead to upper and lower bounds on the decoding trajectory of mutual information. These bounds are still simple in the sense that mutual information is the only parameter tracked, and useful since they can lead to rigorous convergence predictions. This motivates the study of information combining problems.

The notion of information combining was introduced by Huettinger et al. in [10], [11] for the study of concatenated coding systems. In the belief propagation decoding of LDPC codes, the processing at a VND or a CND can be interpreted as an operation of information combining, i.e., combining the mutual information contained in those incoming messages. Let $(X_1, \cdots, X_d)$ be a codeword of either a $[d, 1]$ repetition code or a $[d, d-1]$ single parity-check code, which models a degree $d-1$ variable node or a degree d check node, respectively. Assume $d \geq 3$ to avoid trivial cases. Assume X_i is BPSK modulated under the mapping $0 \rightarrow +1$ and $1 \rightarrow -1$, and transmitted through a binary-input symmetric-output channel with output Y_i, $1 \leq i \leq d$, i.e.,

$$p(Y_i = y | X_i = 1) = p(Y_i = -y | X_i = -1).$$

These d channels are assumed to be independent. The notation $X_i \rightarrow Y_i$ is used to indicate the i-th channel. The soft-bit T_i for the i-th channel is defined as

$$T_i = p(X_i = 1 \mid Y_i) - p(X_i = -1 \mid Y_i)$$

and is a sufficient statistic. Although the transition probability distribution of each channel is unknown, the mutual information of each channel is assumed to be known. Without loss of generality, one can focus on X_d and ask the following question: can we find tight lower and upper bounds on the combined information $I(X_d; Y_1, Y_2, \cdots, Y_{d-1})$, i.e., the extrinsic mutual information? Such an extremal problem was first considered in [12] for a $[3, 1]$ repetition code. It was shown that $I(X_3; Y_1, Y_2)$ is maximized (or minimized) when both $X_1 \to Y_1$ and $X_2 \to Y_2$ are BECs (or binary symmetric channels (BSCs)) with prescribed mutual information values. In [13], the case of $[3, 2]$ single parity-check codes was studied, and it was shown that BSCs achieve the upper bound, and BECs achieve the lower bound. In [14], [15], [16], [17], the above results were extended to arbitrary codeword length d for both repetition codes and single parity-check codes. In [18], instead of mutual information, conditional expectations of messages were used to bound on the performance of belief propagation decoding of LDPC codes.

In [12], [13], [14], [15], the extrinsic mutual information $I(X_d; Y_1, Y_2, \cdots, Y_{d-1})$ is optimized over all $d-1$ channels subject to their individual mutual information constraints. In [16], [17], the problems were generalized by optimizing the extrinsic mutual information with respect to a single channel. Fix the binary-input symmetric-output channels $X_i \to Y_i$, $1 \leq i \leq d-2$, and fix the mutual information of $X_{d-1} \to Y_{d-1}$. It is shown by Sutskover et al. in [17] that for a repetition code, when the channel $X_{d-1} \to Y_{d-1}$ is a BEC (BSC), the extrinsic mutual information $I(X_d; Y_1, Y_2, \cdots, Y_{d-1})$ is maximized (minimized). For a single parity-check code, the roles of BEC and BSC are reversed.

In the above information combining problems, all the results were obtained by proving new mutual information inequalities and using various existing inequalities and identities of mutual information. The work by Sharon et al. in [19], [20] tells us that when the input is equiprobable, the mutual information of a binary-input symmetric-output channel can be expressed as

$$I(X;T) = \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} m_{2i},$$

where m_{2i} is the $2i$-th conditional moment of the channel soft-bit T. For a $[d, d-1]$ single parity-check code, the extrinsic mutual information at the output of its APP decoder is

$$\begin{aligned} & I(X_d; Y_1, Y_2, \cdots, Y_{d-1}) \\ =& \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} \prod_{k=1}^{d-1} \mathrm{E}[T_k^{2i} \mid X_k = 1], \end{aligned} \tag{1}$$

where $\mathrm{E}[T_k^{2i} \mid X_k = 1]$ is the $2i$-th conditional moment of the channel soft-bit T_k. These results lay a foundation and motivate us to solve the extremal information combining problem from a moments approach. In other words, we first solve an extremal problem of soft-bit moments, determining the channel distributions maximizing or minimizing the second conditional moment subject to a mutual information constraint. We find that the extremal channel distributions are BSC and BEC, respectively. Then, we determine the ordering between moment sequences of the BSC and an arbitrary channel that satisfies the mutual information constraint, and the ordering between moment sequences of the BEC and that (arbitrary) channel. We also find the connection between the extremal problem of the second moment and a related minimum mean squared error (MMSE) estimation problem. The theory of Tchebycheff systems [21], [22] is used to solve the extremal problem of the second moment.

Next, we use the ordering properties of moment sequences to prove the results for a $[d, d-1]$ single parity-check code obtained in [17]; i.e., BSC and BEC are the most and least informative channels, respectively.

Bounds obtained from the information combining problems are used to bound the VND and CND EXIT functions [13], [17]. In [17], the lower bounds on the extrinsic mutual information for the variable nodes and check nodes were combined to obtain the overall worst performance bound, which was used to find sufficient conditions on communication channels for successful decoding. Similarly, the upper bounds were combined to obtain the overall best performance bound, which was used to find necessary conditions on communication channels for successful decoding. The independence assumption on channels in the information combining problems are satisfied due to the property of asymptotic tree-like decoding neighborhoods for very long ensembles of LDPC codes [5]. However, as computed in [17], the gap between the predicted threshold and the exact threshold is significant. For binary input AWGN communication channels and LDPC codes of various rates, the upper bounds on the thresholds are about 0.5 dB above the exact thresholds, and the lower bounds on the thresholds are about 1.2 dB below the exact thresholds. This motivates us to consider improving the bounds to reduce threshold gaps.

To this end, we extend the information combining problem for a $[d, d-1]$ single parity-check code by adding one more constraint on the channel distribution [23], namely, a constraint on the second conditional moment of the channel soft-bit. The solution to this extension problem can be used to improve the prediction of convergence of the belief propagation decoding of LDPC codes, provided that another extremal problem related to the variable nodes be solved (details can be found in Section VIII). Again, we use the moments approach to solve the extension problem, first solving an extremal problem of the 4-th moment, then determining the ordering among moment sequences. We also use the results obtained in the extension problem to bound the CND EXIT function with Gaussian priors. The gap of our bounds is very small and we find that an approximation function for the CND EXIT functions in [9, (9)] is very accurate.

Recently new stronger bounds on information combining were reported in [24]. In [24], for obtaining stronger bounds, the authors added additional constraint on the probability of error, which is different from the approach used in [23] and the current paper.

The paper is organized as follows. In Section II, we introduce some basic concepts and results from [3] and [20]. In Section III, we solve an extremal problem of second moments of channel soft-bits. In Section IV, we apply the obtained

results for analysis of MMSE estimators. In Section V, we study the properties of sequences of soft-bit moments. In Section VI, we use the properties of soft-bit moment sequences to solve the original information combining problem at the check nodes. In Section VII, an extension of the information combining problem is proposed and solved. In Section VIII, the results from the extension problem are used to compute potentially improved best and worst performance bounds on the decoding trajectory of mutual information. Some key concepts and results in the Tchebycheff system theory are outlined in Appendix I.

II. T-consistency and Mutual Information

In [3] and [5] the concept of consistency, later called symmetry, of channel probability distributions were proposed. In [19], [20] the concept of consistency, called T-consistency, was formulated for soft-bits and several new mutual information results were obtained. The concept of consistency and results obtained in [3], [5], [19], and [20] lay a foundation for solving the extremal problems of information combining with the help of soft-bit moments. In this section, we give a short summary of the results for completeness. For convenience of notation and exposition we mostly follow presentation in [19], and [20].

Let X and Y be random variables at the input and output of a binary-input symmetric-output channel, respectively. Since the channel is symmetric, its transition probability density function satisfies

$$f(Y = y \mid X = 1) = f(Y = -y \mid X = -1).$$

We assume the input is equiprobable, i.e., $P(X = 1) = P(X = -1) = 0.5$. Define the channel soft-bit T as

$$T = \Pr(X = 1 \mid Y) - \Pr(X = -1 \mid Y). \tag{2}$$

It follows that

$$T = (e^L - 1)/(e^L + 1) = \tanh(L/2), \tag{3}$$

$$L = \ln \frac{1+T}{1-T} = \frac{\Pr(X = 1 \mid Y)}{\Pr(X = -1 \mid Y)}. \tag{4}$$

Thus, soft-bit T is a change of variable of the log-likelihood ratio (LLR) L. Many properties for LLRs, such as symmetry, are well established in [3], [5], [25], and similar properties can be obtained for soft-bits as follows.

Let $p(T = t \mid X = x)$ denote the conditional probability density function of T. Let $p(t) \triangleq p(T = t \mid X = 1)$ for notational convenience. Since the channel is symmetric, it follows that

$$p(t) = p(-t)\frac{1+t}{1-t}, \tag{5}$$

and

$$p(T = t \mid X = 1) = p(T = -t \mid X = -1), \tag{6}$$

which is also true if $p(T = t \mid X = x)$ contains Dirac delta functions. If a random variable satisfies (5) and (6), it is called T-consistent (like symmetry for LLR). For a BSC with capacity I, its T variable takes two values:

$$\{1 - 2h^{-1}[1 - I], 2h^{-1}[1 - I] - 1\},$$

where $h[x] = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function and $h^{-1}[x] \in [0, 0.5]$. For a BEC, its T variable always takes three values $\{-1, 0, 1\}$.

From (5) and (6), it follows that the mutual information between X and T is

$$\begin{aligned} I(X;T) &= \int_{-1}^{+1} \log_2(1+t)p(t)dt \\ &= \frac{1}{\ln 2}\sum_{i=1}^{\infty}\frac{1}{2i(2i-1)}m_{2i}, \end{aligned} \tag{7}$$

where the $2i$-th conditional moment m_{2i} of the channel soft-bit is defined by

$$m_{2i} = \int_{-1}^{+1} t^{2i}p(t)dt. \tag{8}$$

Note that $0 \le m_{2i} \le 1$ and

$$\frac{1}{\ln 2}\sum_{i=1}^{\infty}\frac{1}{2i(2i-1)} = \log_2(1+t)|_{t=1} = 1;$$

therefore, the right-hand side of (7) is convergent. We also point out that the singularity of $\log_2(1+t)$ at $t = -1$ does not affect the integral (7) since from (5) it follows that $p(T = -1 \mid X = 1) = 0$. Furthermore, it is not difficult to show that

$$\begin{aligned} m_{2i} &= \int_{-1}^{+1} t^{2i}p(t)dt \\ &= \int_{-1}^{+1} t^{2i-1}p(t)dt = m_{2i-1}. \end{aligned} \tag{9}$$

Consider an $[n, k]$ binary linear code. Its codeword is BPSK modulated and transmitted through n binary-input symmetric-output independent channels. Let $\underline{X} = (X_1, \cdots, X_n)$ and $\underline{Y} = (Y_1, \cdots, Y_n)$ indicate the input vector and output vector of those n channels, respectively. Assume that $\underline{X}$ is equally probable to be any BPSK modulated codeword. The extrinsic soft-bit for X_j of an APP decoder is defined as

$$T_{E,j} = \Pr(X_j = 1 \mid \underline{Y}_{[j]}) - \Pr(X_j = -1 \mid \underline{Y}_{[j]}),$$

where $\underline{Y}_{[j]}$ is used to denote the vector obtained by deleting the j-th element of a vector $\underline{Y}$. Let $\underline{c}_0$ be the all-zero codeword and $C_j^{(0)}$ be the set of codewords whose j-th position is 0. The random variable $T_{E,j}$ is T-consistent and

$$\begin{aligned} &p(T_{E,j} = t \mid X_j = 1) \\ =&p(T_{E,j} = t \mid \underline{c}_0 \text{ transmitted}) \\ =&p(T_{E,j} = t \mid \underline{c} \text{ transmitted}), \quad \forall \underline{c} \in C_j^{(0)}. \end{aligned} \tag{10}$$

Furthermore, for an $[n, n-1]$ single parity-check code, the extrinsic soft-bit satisfies

$$T_{E,j} = \prod_{k=1, k\neq j}^{n} T_k, \tag{11}$$

where $T_k = \Pr(X_k = 1 \mid Y_k) - \Pr(X_k = -1 \mid Y_k)$. In general, the extrinsic mutual information at the output of the

APP decoder of an $[n, n-1]$ single parity-check code is

$$\begin{aligned} I_{E,j} = I(X_j; T_{E,j}) &= \frac{1}{\ln 2}\sum_{i=1}^{\infty}\frac{1}{2i(2i-1)}\mathrm{E}[T_{E,j}^{2i} \mid X_j = 1] \\ &= \frac{1}{\ln 2}\sum_{i=1}^{\infty}\frac{1}{2i(2i-1)}\mathrm{E}[T_{E,j}^{2i} \mid \underline{c}_0 \text{ transmitted}] \\ &= \frac{1}{\ln 2}\sum_{i=1}^{\infty}\frac{1}{2i(2i-1)}\mathrm{E}[\prod_{k=1,k\neq j}^{n} T_k^{2i} \mid \underline{c}_0 \text{ transmitted}] \\ &= \frac{1}{\ln 2}\sum_{i=1}^{\infty}\frac{1}{2i(2i-1)}\prod_{k=1,k\neq j}^{n}\mathrm{E}[T_k^{2i} \mid X_k = 1]. \end{aligned} \tag{12}$$

As a special case, if all n channels have the same distribution, the average extrinsic mutual information is

$$I_E = \frac{1}{n}\sum_{j=1}^{n} I(X_j; T_{E,j}) = \frac{1}{\ln 2}\sum_{i=1}^{\infty}\frac{1}{2i(2i-1)} m_{2i}^{n-1} \tag{13}$$

where the $2i$-th moment m_{2i} is defined in (8).

III. An Extremal Problem of the Second Moment

In this section, we focus on the following optimization problem:

Problem 1: Among all binary-input symmetric-output channels with a fixed mutual information $I(X;Y) = I(X;T) = I$, determine the maximum and minimum of the conditional second moment m_2 of the channel soft-bit.

We will use T-system theory to solve this problem.

T-system theory (see Appendix I) requires that all integrand functions must be continuous on a closed interval. According to (7)

$$I(X;T) = \int_{-1}^{1}\log_2(1+t)p(t)dt,$$

and one can notice that the integrand $\log(1+t)$ is not continuous on $[-1,1]$. To avoid this problem we use the map $\mathcal{F}(p(t))$ defined by

$$\hat{p}(t) = \mathcal{F}(p(t)) = \begin{cases} p(t) + p(-t) = 2p(t)/(1+t), & t \in (0,1] \\ p(0), & t = 0. \end{cases} \tag{14}$$

The new density $\hat{p}(t)$ is defined on $[0,1]$. From this we have

$$\begin{aligned} I = I(X;T) &= \int_{-1}^{+1}\log_2(1+t)p(t)dt \\ &= \int_{0}^{+1}(1-h[(1-t)/2])\hat{p}(t)dt \\ &= \frac{1}{\ln 2}\sum_{i=1}^{\infty}\frac{1}{2i(2i-1)}m_{2i}, \end{aligned} \tag{15}$$

where

$$m_{2i} = \int_{-1}^{+1} t^{2i}p(t)dt = \int_{0}^{+1} t^{2i}\hat{p}(t)dt. \tag{16}$$

We point out that the odd moments with respect to $p(t)$ are at most as large as those moments with respect to $\hat{p}(t)$, i.e., $\int_{-1}^{+1} t^{2i-1}p(t)dt \le \int_{0}^{+1} t^{2i-1}\hat{p}(t)dt$.

The derivation from (14) to (16) establishes the following two facts:

1) The mapping $\mathcal{F}(\cdot)$ in (14) is a bijection from

$$\begin{aligned} \mathbb{S}_1 = \{&p \mid \text{p is a T-consistent probability density;} \\ &\int_{-1}^{+1}\log_2(1+t)p(t)dt = I\} \end{aligned} \tag{17}$$

to

$$\begin{aligned} \mathbb{S}_2 = \{&\hat{p} \mid \hat{p} \text{ is a probability density on [0,1];} \\ &\int_{0}^{+1}(1-h[(1-t)/2])\hat{p}(t)dt = I\}. \end{aligned} \tag{18}$$

2) Let $p \in \mathbb{S}_1$, then

$$\int_{-1}^{+1} t^2 p(t)dt = \int_{0}^{+1} t^2\hat{p}(t)dt.$$

Therefore, we have

$$\max_{p:p\in\mathbb{S}_1}\int_{-1}^{+1} t^2p(t)dt = \max_{\hat{p}:\hat{p}\in\mathbb{S}_2}\int_{0}^{+1} t^2\hat{p}(t)dt, \tag{19}$$

$$\min_{p:p\in\mathbb{S}_1}\int_{-1}^{+1} t^2p(t)dt = \min_{\hat{p}:\hat{p}\in\mathbb{S}_2}\int_{0}^{+1} t^2\hat{p}(t)dt. \tag{20}$$

Note that $1-h[(1-t)/2]$ is a continuous function on $[0,1]$. Thus we can formulate the following equivalent optimization problem, which is solvable by T-system theory.

Problem 2: Among all probability distributions on $[0,1]$ which satisfy the mutual information constraint, determine the probability distribution σ_b (σ_w) which maximizes (minimizes) the second moment. Mathematically,

$$\begin{aligned} \sigma_b &= \arg\max_{\sigma\in\mathbb{S}}\int_0^1 t^2 d\sigma(t), \\ \sigma_w &= \arg\min_{\sigma\in\mathbb{S}}\int_0^1 t^2 d\sigma(t), \end{aligned}$$

and

$$\mathbb{S} = \left\{\sigma \mid \int_0^1 1 d\sigma(t) = 1; \int_0^1 (1-h[(1-t)/2])d\sigma(t) = I\right\}, \tag{21}$$

where σ is a distribution on $[0,1]$.

Remark 1: To avoid trivial cases, throughout the remainder of this section, we assume $0 < I < 1$.

Let us define:

$$u_0(t) \triangleq 1, \quad u_1(t) \triangleq 1 - h[(1-t)/2], \tag{22}$$

$$\Omega(t) \triangleq -t^2, \quad \underline{c} \triangleq (1, I). \tag{23}$$

Lemma 1: The system $\{u_0, u_1\}$ is a T-system on $[0,1]$.

Proof: Let $0 \le t_0 < t_1 \le 1$. Since $u_1(t) = 1 - h[(1-t)/2]$ is a strictly increasing function, one has

$$\det\begin{pmatrix} u_0(t_0) & u_0(t_1) \\ u_1(t_0) & u_1(t_1) \end{pmatrix} = u_1(t_1) - u_1(t_0) > 0.$$

Hence, according to Definition 2 of Appendix I, $\{u_0, u_1\}$ form a T-system. ∎

Before dealing with more complicated T-systems, we need to establish the following lemmas.

Lemma 2: The systems $\{1, t^2, t^{2n}\}$ with $n \geq 2$ and $\{1, t^2, t^4, t^{2n}\}$ with $n \geq 3$ are T-systems on $[0, 1]$.

Proof: See Appendix II. ∎

As an application, we can prove the following Lemma.

Lemma 3: The augmented system $\{u_0, u_1, \Omega\}$ is a T-system on $[0, 1]$.

Proof: Since $\log_2(1+t)$ has the following series expansion:

$$\log_2(1+t) = \frac{1}{\ln 2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} t^k, \quad -1 < t \leq 1,$$

one can see that

$$\begin{aligned} u_1(t) &= 1 - h[(1-t)/2] \\ &= \frac{1+t}{2} \log_2(1+t) + \frac{1-t}{2} \log_2(1-t) \\ &= \frac{1}{\ln 2} \left(\frac{1+t}{2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} t^k - \frac{1-t}{2} \sum_{k=1}^{\infty} \frac{1}{k} t^k \right) \\ &= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \frac{1}{2k(2k-1)} t^{2k}, \quad t \in [0, 1]. \end{aligned}$$

At $t = 1$, $\frac{1-t}{2} \log_2(1-t) = 0$ by convention. Let $0 \leq t_0 < t_1 < t_2 \leq 1$. We have

$$\det \begin{pmatrix} 1 & 1 & 1 \\ u_1(t_0) & u_1(t_1) & u_1(t_2) \\ \Omega(t_0) & \Omega(t_1) & \Omega(t_2) \end{pmatrix}$$

$$= \frac{1}{\ln 2} \cdot$$

$$\det \begin{pmatrix} 1 & 1 & 1 \\ \sum_{k=1}^{\infty} \frac{t_0^{2k}}{2k(2k-1)} & \sum_{k=1}^{\infty} \frac{t_1^{2k}}{2k(2k-1)} & \sum_{k=1}^{\infty} \frac{t_2^{2k}}{2k(2k-1)} \\ -t_0^2 & -t_1^2 & -t_2^2 \end{pmatrix}$$

$$= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \frac{1}{2k(2k-1)} \det \begin{pmatrix} 1 & 1 & 1 \\ t_0^{2k} & t_1^{2k} & t_2^{2k} \\ -t_0^2 & -t_1^2 & -t_2^2 \end{pmatrix}$$

$$= \frac{1}{\ln 2} \sum_{k=2}^{\infty} \frac{1}{2k(2k-1)} \det \begin{pmatrix} 1 & 1 & 1 \\ t_0^2 & t_1^2 & t_2^2 \\ t_0^{2k} & t_1^{2k} & t_2^{2k} \end{pmatrix} > 0.$$

The second equality is due to the interchangeability of finite sum and infinite sum, and the last inequality is due to Lemma 2. ∎

Now we are ready to solve Problems 2 and 1.

Theorem 1: Among binary-input symmetric-output channels with a fixed mutual information, BSC (BEC) strictly maximizes (minimizes) m_2.

Proof: We first use T-system theory to solve Problem 2. Next, using the map $\mathcal{F}^{-1}$, we find the solutions of Problem 1.

We use $u_0(t), u_1(t), \Omega(t)$, and $\underline{c}$ defined in (22) and (23). Let

$$V(\underline{c}) = \left\{ \sigma \mid \int_0^1 u_i(t) d\sigma(t) = c_i, \ i = 0, 1 \right\}$$

be the set of distributions which are probability distributions and satisfy the mutual information constraint (15). Notice that $\underline{c} \in \mathrm{Int}\mathcal{M}_2$ since $0 < I < 1$. From Lemma 1, Lemma 3, and Theorem 8 of Appendix I, we conclude that the distribution σ^* associated with the upper principal representation of $\underline{c}$ strictly maximizes $-\int_0^1 t^2 d\sigma(t)$ and the distribution σ_* associated with the lower principal representation of $\underline{c}$ strictly minimizes $-\int_0^1 t^2 d\sigma(t)$. Thus, $\sigma_w = \sigma^*$ and $\sigma_b = \sigma_*$.

According to Theorem 7, since $n = 1$ is odd, the entire probability mass of σ_b is concentrated at an interior point t_1, and the entire probability mass of σ_w is concentrated on the two endpoints 0 and 1. By solving (15) for σ_b, i.e., $u_1(t_1) = I$, we obtain $t_1 = 1 - 2h^{-1}[1 - I]$. According to (14), the T-consistent distribution $\mathcal{F}^{-1}(\sigma_b)$ has probability mass of $(1 + t_1)/2$ at t_1 and probability mass of $(1 - t_1)/2$ at $-t_1$. This exactly corresponds to a BSC with capacity I. Similarly, for σ_w, one can conclude that the probability mass at 1 is I. The T-consistent distribution $\mathcal{F}^{-1}(\sigma_w)$ has probability mass of I at 1, and probability mass of $1 - I$ at 0. This exactly corresponds to a BEC with capacity I. ∎

For a BEC with mutual information I, its m_2 is equal to I. For a BSC with mutual information I, we have

$$m_2 = t_1^2 = (1 - 2h^{-1}[1 - I])^2.$$

Define a function

$$\Phi(x) = (1 - 2h^{-1}[1 - x])^2, \quad x \in [0, 1]. \tag{24}$$

The following property of $\Phi(x)$ will be useful for us in the sequel.

Lemma 4: $\Phi(x)$ is a strictly increasing and concave function on $[0, 1]$; its inverse $\Phi^{-1}(x)$ is strictly increasing and convex on $[0, 1]$.

Proof: $\Phi(x)$ is obviously strictly increasing on $[0, 1]$. Thus, its inverse function exists:

$$\Phi^{-1}(x) = 1 - h\left(\frac{1 - \sqrt{x}}{2} \right).$$

It is easy to see that $\Phi^{-1}(x)$ is strictly increasing and convex on $[0, 1]$. Thus it follows that $\Phi(x)$ is concave. ∎

IV. MMSE and Mutual Information

In this section we show how to apply results from the previous section for analysis of the MMSE estimator of a binary-input symmetric-output channel.

It is easy to see that the channel soft-bit T defined in (2) is in fact the MMSE estimator for the channel input X. It is well known that the minimum mean square error is equal to

$$\mathrm{MMSE} = \mathrm{E}[(X - T)^2] = \mathrm{E}[X^2] - \mathrm{E}[T^2] = 1 - \mathrm{E}[T^2].$$

Further

$$\begin{aligned} \mathrm{E}[T^2] &= (\mathrm{E}[T^2|X = 1] + \mathrm{E}[T^2|X = -1])/2 \\ &= \mathrm{E}[T^2|X = 1] = m_2, \end{aligned} \tag{25}$$

where the second equality is due to (6). Hence

$$\mathrm{MMSE} = 1 - m_2. \tag{26}$$

Theorem 2: Among binary-input symmetric-output channels with a fixed mutual information, minimum mean squared error for estimating the channel input in a BSC is strictly minimal.

Proof: Combine (26) and Theorem 1. ∎

For a BSC with mutual information I, we have derived that

$$m_2 = \Phi(I) = (1 - 2h^{-1}[1 - I])^2.$$

Thus for BSC, we have

$$\text{MMSE} = 1 - \Phi(I).$$

It is easy to see that as I increases, its MMSE decreases, as expected.

Similarly, we have

Lemma 5: Among binary-input symmetric-output channels with a fixed mutual information, minimum mean squared error for estimating the channel input in a BEC is strictly maximal.

Proof: Combine (26) and Theorem 1. ∎

For a BEC with mutual information I, we have

$$\text{MMSE} = 1 - I.$$

V. Properties of Soft-Bit Moments

In this section we will characterize the ordering between moment sequences of channel soft-bits of a BSC (or BEC) and any other binary-input symmetric-output channels with the same mutual information value. These ordering properties will be used to prove the information combining results. We first prove the following auxiliary lemma.

Lemma 6: The moment sequence $\{m_{2i}\}_{i=1}^{\infty}$ of a channel soft-bit defined by (8) has the following properties:

1) $\{m_{2i}\}_{i=1}^{\infty}$ is nonnegative and nonincreasing.
2) The ratio sequence $\{\frac{1}{m_2}, \frac{m_2}{m_4}, \frac{m_4}{m_6}, \cdots\}$ is nonincreasing.

Proof: Nonnegativity is obvious. For $i \geq 1$,

$$m_{2i} - m_{2i+2} = \int_{-1}^{+1} (t^{2i} - t^{2i+2}) p(t) dt \geq 0;$$

thus, $\{m_{2i}\}_{i=1}^{\infty}$ is nonincreasing.

By Holder's inequality, for $i \geq 1$,

$$\begin{aligned} m_{2i+2} &= \mathrm{E}[T^{2i+2}] = \mathrm{E}[|T^i T^{i+2}|] \leq \mathrm{E}[|T^i|^2]^{\frac{1}{2}} \cdot \mathrm{E}[|T^{i+2}|^2]^{\frac{1}{2}} \\ &= \mathrm{E}[T^{2i}]^{\frac{1}{2}} \cdot \mathrm{E}[T^{2i+4}]^{\frac{1}{2}} = \sqrt{m_{2i} m_{2i+4}}, \end{aligned}$$

where the expectation is with respect to $p(t)$. Due to the nonnegativity of m_{2i}, we have for $i \geq 1$

$$\frac{m_{2i}}{m_{2i+2}} \geq \frac{m_{2i+2}}{m_{2i+4}}.$$

It remains to show $\frac{1}{m_2} \geq \frac{m_2}{m_4}$. By Holder's inequality,

$$\mathrm{E}[T^2] = \mathrm{E}[|1 \cdot T^2|] \leq (\mathrm{E}[1]\mathrm{E}[T^4])^{1/2} = \mathrm{E}[T^4]^{1/2},$$

thus $m_2^2 = \mathrm{E}[T^2]^2 \leq \mathrm{E}[T^4] = m_4$. This concludes the proof. ∎

Remark 2: Another conclusion [26, p. 110] related to ratios of moments is: let F be an arbitrary probability distribution function and $v_i = \int |z|^i dF(z)$; then

$$\frac{v_{2i}}{v_{2i+1}} \geq \frac{v_{2i+1}}{v_{2i+2}}.$$

The following lemma can be used to prove that BSC is the best channel (or most informative channel) for information combining at a check node of an LDPC code. Recall that $\mathbb{S}_1$ defined in (17) is the set of T-consistent distributions on $[-1, 1]$ which satisfy the mutual information constraint.

Lemma 7: There are two possible types of ordering for the moment sequence $\{m_{b,2i}\}_{i=1}^{\infty}$ of the BSC and the moment sequence $\{m_{p,2i}\}_{i=1}^{\infty}$ of an arbitrary channel with $p \in \mathbb{S}_1$:

1) $m_{b,2i} > m_{p,2i}$ for $1 \leq i < i_0$, and $m_{b,2i} < m_{p,2i}$ for $i \geq i_0$;
2) $m_{b,2i} > m_{p,2i}$ for $1 \leq i < i_0$, $m_{b,2i_0} = m_{p,2i_0}$ and $m_{b,2i} < m_{p,2i}$ for $i > i_0$,

where i_0 is an integer depending on p.

Proof: For the BSC, its moment sequence $\{m_{b,2i} = t_1^{2i}\}_{i=1}^{\infty}$ is a geometric sequence, therefore the moment ratio

$$r_{b,2i} = \frac{m_{b,2i}}{m_{b,2i+2}} = \frac{1}{t_1^2}, \quad i \geq 1.$$

By Theorem 1, we have $m_{b,2} > m_{p,2}$. Since $p \in \mathbb{S}_1$, we have

$$\frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} m_{p,2i} = \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} m_{b,2i} = I$$

and therefore there exists at least one index i such that $m_{p,2i} \geq m_{b,2i}$. Let i_0 be the smallest index with this property. Then there are two possible cases:

1) $m_{b,2i_0} < m_{p,2i_0}$.
   By the definition of i_0, we have $m_{b,2i} > m_{p,2i}$ for $1 \leq i < i_0$. It is easy to see that
   $$r_{b,2i_0-2} = \frac{m_{b,2i_0-2}}{m_{b,2i_0}} > \frac{m_{p,2i_0-2}}{m_{p,2i_0}} = r_{p,2i_0-2}.$$
   By Lemma 6, we have $r_{p,2i_0} \leq r_{p,2i_0-2} < r_{b,2i_0-2} = r_{b,2i_0}$, i.e.,
   $$r_{p,2i_0} = \frac{m_{p,2i_0}}{m_{p,2i_0+2}} < r_{b,2i_0} = \frac{m_{b,2i_0}}{m_{b,2i_0+2}}.$$
   Thus
   $$m_{b,2i_0+2} < \frac{m_{b,2i_0}}{m_{p,2i_0}} m_{p,2i_0+2} < m_{p,2i_0+2}. \tag{27}$$
   Using the above argument repeatedly and sequentially for $i \geq i_0 + 2$, one can conclude that $m_{b,2i} < m_{p,2i}$ for $i \geq i_0$. This concludes the proof of the first case.
2) $m_{b,2i_0} = m_{p,2i_0}$.
   The remaining proof is exactly the same as in the first case except that the second inequality ($<$) in (27) becomes an equality.

∎

Similarly, the following lemma will be used to prove that the BEC is the worst channel (or least informative channel) for information combining at a check node of an LDPC code.

Lemma 8: There are two possible types of ordering for the moment sequence $\{m_{w,2i}\}_{i=1}^{\infty}$ of the BEC and the moment sequence $\{m_{p,2i}\}_{i=1}^{\infty}$ of an arbitrary channel with $p \in \mathbb{S}_1$:

1) $m_{p,2i} > m_{w,2i}$ for $1 \leq i < i_1$, and $m_{p,2i} < m_{w,2i}$ for $i \geq i_1$;
2) $m_{p,2i} > m_{w,2i}$ for $1 \leq i < i_1$, $m_{p,2i_1} = m_{w,2i_1}$ and $m_{p,2i} < m_{w,2i}$ for $i > i_1$,

where i_1 is an integer depending on p.

Proof: It is easy to see that $\{m_{w,2i} = I\}_{i=1}^{\infty}$. By Theorem 1, one has $m_{w,2} < m_{p,2}$. Since $p \in \mathbb{S}_1$ we have

$$\frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} m_{p,2i} = I = \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} m_{w,2i}, \tag{28}$$

and therefore there exists at least one index i such that $m_{p,2i} \leq m_{w,2i}$. Let i_1 be the smallest index with this property. Then

$$m_{w,2i} < m_{p,2i}, \quad 1 \leq i < i_1. \tag{29}$$

For the $2i_1$-th moment, there are two possible cases:

1) $m_{w,2i_1} > m_{p,2i_1}$.
   For $i > i_1$, one has

$$m_{w,2i} = m_{w,2i_1} > m_{p,2i_1} \geq m_{p,2i},$$

   where the last inequality is due to Lemma 6. This concludes the proof of the first case.
2) $m_{w,2i_1} = m_{p,2i_1}$.
   We claim that $m_{p,2i_1+2} < m_{p,2i_1}$ and prove it by contradiction. If not so, it must be $m_{p,2i_1+2} = m_{p,2i_1}$ since moments are nonincreasing. Therefore, $r_{p,2i_1} = m_{p,2i_1}/m_{p,2i_1+2} = 1$. Together with the fact that a ratio of two consecutive moments is at least 1 and Lemma 6, we obtain $r_{p,2i} = 1, \quad i > i_1$. Thus,

$$m_{p,2i} = m_{p,2i_1} = m_{w,2i_1} = m_{w,2i} = I, \quad i > i_1.$$

   Together with (29), it is obvious that (28) is violated. Thus we have proved that $m_{p,2i_1+2} < m_{p,2i_1}$, and it follows $m_{p,2i_1+2} < m_{w,2i_1+2}$. The rest of the proof is the same as in the first case.

■

As an example, in Fig. 1, we plot the initial parts of the moment sequences of a BSC, a binary-input AWGN channel, and a BEC, all having mutual information 0.5.

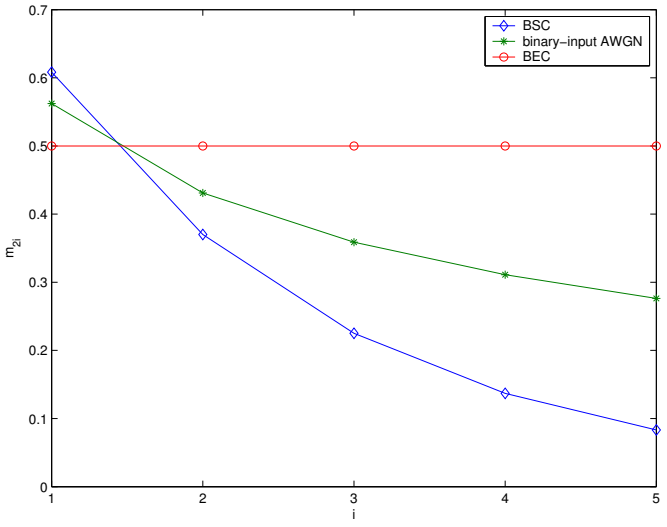


Fig. 1. The moment sequences of a BSC, a binary-input AWGN channel, and a BEC. All three channels have mutual information 0.5.

VI. Information Combining

Let $(X_1, \cdots, X_d)$ be a codeword of either a $[d, 1]$ repetition code, or a $[d, d-1]$ single parity-check code. The repetition code is a model for a variable node of degree $d-1$ with an incoming message from a communication channel. The single parity-check code is a model for a check node of degree d. Assume $d \geq 3$ to avoid trivial cases. Assume X_i is BPSK modulated under the mapping $0 \to +1$ and $1 \to -1$, and transmitted through a binary-input symmetric-output channel with output Y_i, $1 \leq i \leq d$. These d channels are assumed independent. In this section, we focus on the type of information combining problem proposed in [16], [17].

Problem 3: Fix the binary-input symmetric-output channels $X_i \to Y_i$, $1 \leq i \leq d-2$ and fix the mutual information of $X_{d-1} \to Y_{d-1}$. Find symmetric channel distributions for $X_{d-1} \to Y_{d-1}$ such that the extrinsic mutual information $I(X_d; Y_1, Y_2, \cdots, Y_{d-1})$ is maximized or minimized.

Due to [17, Lemma 3], we only need to solve either the variable node problem or the check node problem. In this section, we will solve Problem 3 for a degree d check node, i.e., a $[d, d-1]$ single parity-check code, using the properties of the moments m_{2i}.

Since we will work with moments of soft-bits, we will reformulate Problem 3 in terms of soft-bits. Without loss of generality, the extremal problem of information combining at a check node can be stated as follows.

Problem 4: Let $(X_1, \ldots, X_d)$ be a codeword of a $[d, d-1]$ single parity-check code. Fix the first $d-2$ binary-input symmetric-output channels $X_i \to Y_i$, $1 \leq i \leq d-2$, and let each channel have a positive mutual information. Find symmetric channel distributions for $X_{d-1} \to Y_{d-1}$ that maximize and minimize, respectively, the extrinsic mutual information $I(X_d; T_{E,d})$ subject to the constraint $I(X_{d-1}; T_{d-1}) = I_{d-1}$, $0 < I_{d-1} < 1$.

Remark 3: If any of the first $d-2$ channels has zero mutual information, $I(X_d; T_{E,d})$ is also zero. Thus we assume they all have positive mutual information. The assumption $0 < I_{d-1} < 1$ is also used to avoid trivial cases.

Before solving Problem 4, we need to prove the following lemma.

Lemma 9: Assume $a_i > 0$ for $i \geq 1$, $1 \geq b_1 \geq b_2 \geq \cdots \geq b_i \geq \cdots \geq 0$. Let sequences $\{x_i\}_{i=1}^{\infty}$, $\{y_i\}_{i=1}^{\infty}$ satisfy

1) $x_i \geq 0$, $y_i \geq 0$, for $i \geq 1$,
2) $\sum_{i=1}^{\infty} a_i x_i = \sum_{i=1}^{\infty} a_i y_i$,
3) $x_i > y_i$ for $1 \leq i \leq k$, and $x_i < y_i$ for $i > k$.

Then

$$\sum_{i=1}^{\infty} a_i b_i x_i \geq \sum_{i=1}^{\infty} a_i b_i y_i.$$

and a necessary and sufficient condition to achieve the equality is $b_i = const, \ i = 1, \ldots, \infty$.

Proof: See Appendix III. ■

Now we are ready to show the following theorem, which solves Problem 4.

Theorem 3: ([17, Theorem 3]) Under the constraints of Problem 4 the BSC and BEC maximize and minimize the extrinsic mutual information $I(X_d; T_{E,d})$, respectively.

Proof: Let $\{m_{2i}\}_{i=1}^{\infty}$ indicate the moment sequence for T_{d-1}. Define

$$\begin{aligned} a_i &= \frac{1}{\ln 2} \frac{1}{2i(2i-1)} > 0, \\ b_i &= \prod_{k=1}^{d-2} \mathrm{E}[T_k^{2i} | X_k = 1]. \end{aligned}$$

Obviously, $1 \geq b_1 \geq b_2 \geq \cdots \geq b_i \geq \cdots \geq 0$. By (7) and (12), we have

$$\begin{aligned} I_{d-1} &= I(X_{d-1};T_{d-1}) = \sum_{i=1}^{\infty} a_i m_{2i}, \\ I(X_d;T_{E,d}) &= \sum_{i=1}^{\infty} a_i b_i m_{2i}. \end{aligned}$$

By Lemma 7 and Lemma 9, we conclude a BSC with the prescribed mutual information I_{d-1} maximizes $I(X_d;T_{E,d})$. Similarly, by Lemma 8 and Lemma 9, we conclude a BEC with the prescribed mutual information I_{d-1} minimizes $I(X_d;T_{E,d})$. ■

We would like to note that the condition $b_1 = b_2 = \cdots = b_n = b_{n+1} = \cdots$ is equivalent to requiring each channel $X_i \to Y_i$, $1 \leq i \leq d-2$, has a constant moment sequence, i.e., all of them are BEC channels. Thus, if $X_i \to Y_i$, $1 \leq i \leq d-2$, are all BECs with mutual information I_i, the extrinsic mutual information

$$I(X_d;T_{E,d}) = \prod_{k=1}^{d-1} I_k$$

does not depend on the channel distribution of $X_{d-1} \to Y_{d-1}$. It means for $X_{d-1} \to Y_{d-1}$, a BSC, BEC, or any other channel distribution with mutual information I_{d-1} leads to a constant extrinsic mutual information. Otherwise, if any channel among $X_i \to Y_i$, $1 \leq i \leq d-2$, is not a BEC, by Lemma 9, the BSC for $X_{d-1} \to Y_{d-1}$ strictly maximizes the extrinsic mutual information $I(X_d;T_{E,d})$, and the BEC strictly minimizes it.

For a $[d,1]$ repetition code, we know that $X_1 = X_2 = \cdots = X_d$. Assume at least one of the first $d-2$ channels has positive mutual information. If $X_i \to Y_i$, $1 \leq i \leq d-2$, are all BECs with mutual information I_i, for $X_{d-1} \to Y_{d-1}$, a BSC, or a BEC, or any other symmetric channel distribution (with mutual information I_{d-1}) leads to a constant extrinsic mutual information. Otherwise, the BEC is a strict maximizer, and the BSC is a strict minimizer.

Now consider the case [13] when the distribution of each channel is allowed to vary subject to a mutual information equality constraint in $(0,1)$. At a check node, by applying Theorem 3 sequentially, one can conclude that when $X_i \to Y_i$, $1 \leq i \leq d-1$, are all BSCs, the extrinsic mutual information $I(X_d;Y_1,Y_2,\cdots,Y_{d-1})$ is strictly maximized. However, the minimizing distribution is not unique. As long as there is at most one non-BEC distribution for $d-1$ channels, $I(X_d;Y_1,Y_2,\cdots,Y_{d-1})$ is minimized. Similarly, for a $[d,1]$ repetition code, when $X_i \to Y_i$, $1 \leq i \leq d-1$, are all BSCs, the extrinsic mutual information $I(X_d;Y_1,Y_2,\cdots,Y_{d-1})$ is strictly minimized. However, the maximizing distribution is not unique. As long as there is at most one non-BEC distribution for $d-1$ channels, $I(X_d;Y_1,Y_2,\cdots,Y_{d-1})$ is maximized.

In [13] and [17], the bounds obtained in the information combining problem were used to bound the EXIT chart of LDPC codes. For a degree d variable node (corresponding to bit $b \in \{-1,+1\}$), let $r(x)$ be the probability density function of the channel LLR message conditioned on $b = +1$, and I_c be the channel mutual information. Let I_A be the mutual information for each incoming message. The upper bound on the extrinsic mutual information is [17]

$$I_{UB,VND} = 1 - (1-I_A)^{d-1}(1-I_c), \tag{30}$$

and the lower bound is [17]

$$\begin{aligned} &I_{LB,VND} \\ &= 1 - \sum_{k=0}^{d-1} \binom{d-1}{k} p^k (1-p)^{d-1-k} \\ &\cdot \int_{-\infty}^{\infty} \log_2 \left(1 + e^{-x} \left(\frac{p}{1-p}\right)^{d-1-2k}\right) r(x)dx, \end{aligned} \tag{31}$$

where $p = h^{-1}(1-I_A)$. These two bounds will be used in Section VIII.

Furthermore, in [17], best and worst performance bounds were derived and used to predict the convergence of the belief propagation decoding of LDPC codes. However, the gap between the predicted threshold and the exact threshold is significant. For instance, in [17], it was computed that, for binary input AWGN communication channels and LDPC codes of various rates, the upper bounds on the thresholds are about 0.5 dB above the exact thresholds, and the lower bounds on the thresholds are about 1.2 dB below the exact thresholds. This observation motivates us to improve the bounds. To this end, in the next section, we study an extended information combining problem.

VII. An Extended Information Combining Problem

A. More on Properties of Soft-Bits

In this section, we consider an extension of the original extremal information combining problem by adding a constraint on the second conditional moment m_2. The motivation is to narrow the gap between the maximum and minimum of extrinsic mutual information at a check node. In turn, this will potentially help improve the prediction of convergence of the belief propagation decoding of LDPC codes, which will be analyzed in Section VIII. Throughout this section, we keep assuming all channels $X_i \to Y_i$, $1 \leq i \leq d$, are independent.

Problem 5: Let $(X_1,\cdots,X_d)$ be a codeword of a $[d,d-1]$ single parity-check code. Fix the first $d-2$ binary-input symmetric-output channels $X_i \to Y_i$, $1 \leq i \leq d-2$, and each channel has a positive mutual information. Find symmetric channel distributions for $X_{d-1} \to Y_{d-1}$ that maximize and minimize the extrinsic mutual information $I(X_d;T_{E,d})$, respectively, subject to the constraints

- $I(X_{d-1};T_{d-1}) = I_{d-1}$,
- $m_2 = \mathrm{E}[T_{d-1}^2 | X_{d-1} = 1] = \theta$,

where $0 < I_{d-1} < 1$ and $I_{d-1} < \theta < \Phi(I_{d-1})$.

Recall that $\Phi(x)$ is defined in (24). Since we will work with soft-bits, a channel distribution for $X_{d-1} \to Y_{d-1}$ means a probability distribution of T_{d-1} conditioned on $X_{d-1} = 1$. We use P_b and P_w to indicate the maximizing and minimizing distributions, respectively. Since $0 < I_{d-1} < 1$, by Theorem 1, we have

1) If $I_{d-1} = \theta$, there is only one channel distribution, namely BEC, satisfying both constraints.

2) If $\theta = \Phi(I_{d-1})$, there is only one channel distribution, namely BSC, satisfying both constraints.
3) If $\theta \in (I_{d-1}, \Phi(I_{d-1}))$, there are infinitely many channel distributions satisfying both constraints.

Thus, Problem 5 is nontrivial if and only if $0 < I_{d-1} < 1$ and $I_{d-1} < \theta < \Phi(I_{d-1})$. Notice that these two conditions imply $0 < \theta < 1$.

Recall that the extremal distributions for Problem 4 are exactly the extremal distributions for Problem 1. Following that approach, first, we consider the following optimization problem. Later we will show that the extremal distributions of this problem are exactly the same as those of Problem 5.

Problem 6: Among all binary-input symmetric-output channels with a fixed mutual information $I(X;Y) = I(X;T) = I$ and a fixed conditional second moment m_2, determine the maximum and minimum of the conditional fourth moment m_4 of the channel soft-bit.

Again, since we will work with the channel soft-bit, we focus on its conditional probability distribution and use $P_{4,b}$ and $P_{4,w}$ to indicate the maximizing and minimizing distributions, respectively.

We will use T-system theory to solve Problem 6. As in the approach to Problem 1, we transform Problem 6 to an equivalent problem on the domain $[0,1]$. For convenience, we first make the following definitions:

$$\begin{aligned} u_0(t) &\triangleq 1, \\ u_1(t) &\triangleq t^2, \\ u_2(t) &\triangleq 1 - h[(1-t)/2], \\ \Omega(t) &\triangleq -t^4, \\ \underline{c} &= (c_0, c_1, c_2) \triangleq (1, \theta, I_{d-1}). \end{aligned}$$

The equivalent problem is:

Problem 7: Among all probability distributions on $[0,1]$ which satisfy an equality constraint on mutual information and an equality constraint on conditional second moment, determine the probability distribution $\tilde{\sigma}_b$ ($\tilde{\sigma}_w$) which maximizes (minimizes) the 4-th moment m_4. Mathematically,

$$\begin{aligned} \tilde{\sigma}_b &= \arg\max_{\sigma \in \mathbb{S}} \int_0^1 t^4 d\sigma(t), \\ \tilde{\sigma}_w &= \arg\min_{\sigma \in \mathbb{S}} \int_0^1 t^4 d\sigma(t), \\ \tilde{\mathbb{S}} &= \{\sigma \mid \int_0^1 u_i(t) d\sigma(t) = c_i, 0 \le i \le 2\}, \end{aligned}$$

where σ is a distribution on $[0,1]$.

After this problem is solved, by taking an inverse mapping of (14), we obtain the extremal distributions to Problem 6.

Next, we prove a couple of lemmas related to T-systems.

Lemma 10: $\{u_0, u_1, u_2\}$ is a T-system on $[0,1]$.

Proof: We have already shown in Lemma 3 that $\{1, 1 - h[(1-t)/2], -t^2\}$ is a T-system on $[0,1]$. By Definition 2, $\{1, t^2, 1 - h[(1-t)/2]\}$ is a T-system on $[0,1]$ as well. ∎

Lemma 11: The augmented system $\{u_0, u_1, u_2, \Omega\}$ is a T-system on $[0,1]$.

Proof: The proof is similar to the proof of Lemma 3, with the help of Lemma 2. ∎

Since we have proved both $\{u_0, u_1, u_2\}$ and $\{u_0, u_1, u_2, \Omega\}$ are T-systems on $[0,1]$, we will use the notation $V(\underline{c})$ from T-system theory, i.e., $V(\underline{c}) = \tilde{\mathbb{S}}$, where $\tilde{\mathbb{S}}$ is defined in Problem 7. Now we are ready to solve Problem 7 (equivalently, Problem 6).

Obviously,

$$V(\underline{c}) = \left\{ \sigma \middle| \int_0^1 u_i(t) d\sigma(t) = c_i,\ i = 0,1,2 \right\}$$

is a set of distributions which are probability distributions and satisfy constraints on the mutual information and the conditional second moment. Notice that $\underline{c} \in \mathrm{Int}\mathcal{M}_3$ since $0 < I_{d-1} < 1$ and $I_{d-1} < \theta < \Phi(I_{d-1})$. By Lemma 10, Lemma 11, and Theorem 8 of Appendix I, we concludes that the distribution σ^* associated with the upper principal representation of $\underline{c}$ strictly maximizes $-\int_0^1 t^4 d\sigma(t)$, thus $\tilde{\sigma}_w = \sigma^*$. For the same reason the distribution σ_* associated with the lower principal representation of $\underline{c}$ strictly minimizes $-\int_0^1 t^4 d\sigma(t)$, thus $\tilde{\sigma}_b = \sigma_*$.

Next, we will determine $\tilde{\sigma}_b$, $\tilde{\sigma}_w$ and map them by $\mathcal{F}^{-1}$ ($\mathcal{F}$ defined in (14)) back to T-consistent distributions on $[-1,1]$, which will be $P_{4,b}$ and $P_{4,w}$, respectively. According to Theorem 7, since $n = 2$ is even, $q = 1$, all probability mass of $\tilde{\sigma}_b$ concentrates at two points $\{0, t_1\}$, where t_1 is an interior point, and all probability mass of $\tilde{\sigma}_w$ concentrates at two points $\{s_1, 1\}$, where $0 \le s_1 \le t_1$. Correspondingly, $P_{4,b} = \mathcal{F}^{-1}(\tilde{\sigma}_b)$ has probability mass of p_0 at 0, probability mass of $p_{b,1}$ at t_1, and probability mass of $1 - p_0 - p_{b,1}$ at $-t_1$. From the constraints and the distribution structure, we can determine that

$$t_1 = f^{-1}\left(\frac{I_{d-1}}{\theta}\right), \tag{32}$$

$$p_{b,1} = \frac{\theta \cdot (1 + t_1)}{2{t_1}^2}, \tag{33}$$

$$p_0 = 1 - \frac{2p_{b,1}}{1 + t_1}, \tag{34}$$

$$f(x) = \frac{1 - h[(1-x)/2]}{x^2}. \tag{35}$$

Since $f(x)$ is a strictly increasing function on $[0,1]$, f^{-1} is well defined.

In Fig. 2, we give a not-to-scale illustration for $P_{4,b}$.

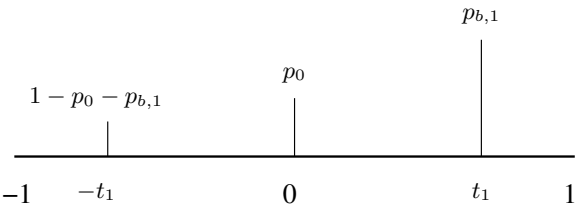


Fig. 2. A not-to-scale illustration for the distribution $P_{4,b}$.

Similarly, $P_{4,w} = \mathcal{F}^{-1}(\tilde{\sigma}_w)$ has probability mass of $p_{w,1}$ at s_1, probability mass of $p_{w,1}\frac{1-s_1}{1+s_1}$ at $-s_1$, and probability mass of p_2 at 1. From the constraints and the distribution structure,

we can determine that

$$s_1 = g^{-1}\left(\frac{1-I_{d-1}}{1-\theta}\right), \quad (36)$$

$$p_{w,1} = \frac{1-\theta}{2(1-s_1)}, \quad (37)$$

$$p_2 = \frac{\theta - s_1^2}{1-s_1^2}, \quad (38)$$

$$g(x) = \frac{h[(1-x)/2]}{1-x^2}. \quad (39)$$

Since $g(x)$ is a strictly increasing function on $[0,1]$, g^{-1} is well defined.

In Fig. 3, we give a not-to-scale illustration for $P_{4,w}$.

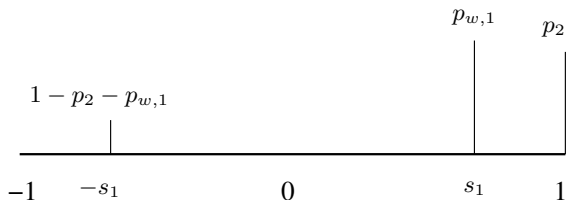


Fig. 3. A not-to-scale illustration for the distribution $P_{4,w}$.

In summary, we have

Theorem 4: Among binary-input symmetric-output channels with a fixed mutual information I_{d-1} and a fixed conditional second moment θ, where $0 < I_{d-1} < 1$ and $I_{d-1} < \theta < \Phi(I_{d-1})$, the channel corresponding to $P_{4,b}$ ($P_{4,w}$) strictly maximizes (minimizes) m_4.

For convenience, we define the following set

$$\tilde{\mathbb{S}}_1 = \{p \,|\text{p is a T-consistent probability density;} \int_{-1}^{+1} \log_2(1+t)p(t)dt = I, \int_{-1}^{+1} t^2 p(t)dt = \theta\}. \quad (40)$$

For $P_{4,b}$, the moment sequence $\{\tilde{m}_{b,2i} = \theta t_1^{2i-2}\}_{i=1}^{\infty}$ is a geometric sequence. Thus, similar to Lemma 7, we have

Lemma 12: There are two possible types of ordering for the moment sequence $\{\tilde{m}_{b,2i}\}_{i=1}^{\infty}$ of $P_{4,b}$ and the moment sequence $\{m_{p,2i}\}_{i=1}^{\infty}$ of any $p \in \tilde{\mathbb{S}}_1$ other than $P_{4,b}$:

1) $\tilde{m}_{b,2i} > m_{p,2i}$ for $2 \le i < i_0$, and $\tilde{m}_{b,2i} < m_{p,2i}$ for $i \ge i_0$;
2) $\tilde{m}_{b,2i} > m_{p,2i}$ for $2 \le i < i_0$, $\tilde{m}_{b,2i_0} = m_{p,2i_0}$ and $\tilde{m}_{b,2i} < m_{p,2i}$ for $i > i_0$,

where i_0 is an integer depending on p. Obviously $\tilde{m}_{b,2} = m_{p,2} = \theta$.

Proof: Similar to the proof of Lemma 7. ∎

Before proving ordering properties related to $P_{4,w}$, we need to establish the following lemma.

Lemma 13: Let $\{m_{2i}\}_{i=1}^{\infty}$ be the moment sequence of a channel soft-bit. Define $\Delta_{2i} = m_{2i} - m_{2i+2}$ for $i \ge 1$, and $\Delta_0 = 1 - m_2$. $\{\Delta_{2i}\}_{i=0}^{\infty}$ has the following properties:

1) $\{\Delta_{2i}\}_{i=0}^{\infty}$ is nonnegative and nonincreasing.
2) The ratio sequence $\{\frac{\Delta_{2i}}{\Delta_{2i+2}}\}_{i=0}^{\infty}$ is nonincreasing.

Proof: Nonnegativity is true since the moment sequence is nonincreasing, and $1 \ge m_2$. Due to the second property in Lemma 6, we have for $i \ge 1$, $m_{2i+2}^2 \le m_{2i}m_{2i+4}$. It follows that

$$\begin{aligned} m_{2i+2} \le & [m_{2i}m_{2i+4}]^{1/2} \le \frac{m_{2i}+m_{2i+4}}{2} \\ \Rightarrow & m_{2i} - m_{2i+2} \ge m_{2i+2} - m_{2i+4} \\ \Rightarrow & \Delta_{2i} \ge \Delta_{2i+2}. \end{aligned}$$

Due to Lemma 6, we have $1/m_2 \ge m_2/m_4$. Thus

$$\begin{aligned} m_2 \le & [1 \cdot m_4]^{1/2} \le \frac{1+m_4}{2} \\ \Rightarrow & 1 - m_2 \ge m_2 - m_4 \\ \Rightarrow & \Delta_0 \ge \Delta_2. \end{aligned}$$

By Holder's inequality, for $i \ge 1$,

$$\begin{aligned} \Delta_{2i+2} &= \mathrm{E}[T^{2i+2}(1-T^2)] = \mathrm{E}[|T^i\sqrt{1-T^2} \cdot T^{i+2}\sqrt{1-T^2}|] \\ &\le \mathrm{E}[|T^i\sqrt{1-T^2}|^2]^{\frac{1}{2}} \cdot \mathrm{E}[|T^{i+2}\sqrt{1-T^2}|^2]^{\frac{1}{2}} \\ &= \mathrm{E}[T^{2i}(1-T^2)]^{\frac{1}{2}} \cdot \mathrm{E}[T^{2i+4}(1-T^2)]^{\frac{1}{2}} \\ &= \sqrt{\Delta_{2i}\Delta_{2i+4}}, \end{aligned}$$

where the expectation is with respect to $p(T|x=1)$. Due to the nonnegativity,

$$\frac{\Delta_{2i}}{\Delta_{2i+2}} \ge \frac{\Delta_{2i+2}}{\Delta_{2i+4}}.$$

Also by Holder's inequality, we have

$$\begin{aligned} \Delta_2 &= \mathrm{E}[T^2(1-T^2)] = \mathrm{E}[|\sqrt{1-T^2} \cdot T^2\sqrt{1-T^2}|] \\ &\le \mathrm{E}[|\sqrt{1-T^2}|^2]^{\frac{1}{2}} \cdot \mathrm{E}[|T^2\sqrt{1-T^2}|^2]^{\frac{1}{2}} \\ &= \mathrm{E}[(1-T^2)]^{\frac{1}{2}} \cdot \mathrm{E}[T^4(1-T^2)]^{\frac{1}{2}} \\ &= \sqrt{\Delta_0\Delta_4}, \end{aligned}$$

thus

$$\frac{\Delta_0}{\Delta_2} \ge \frac{\Delta_2}{\Delta_4}.$$

∎

Now we are ready to prove the following result for $P_{4,w}$.

Lemma 14: There are two possible types of ordering for the moment sequence $\{\tilde{m}_{w,2i}\}_{i=1}^{\infty}$ of $P_{4,w}$ and the moment sequence $\{m_{p,2i}\}_{i=1}^{\infty}$ of any $p \in \tilde{\mathbb{S}}_1$ other than $P_{4,w}$:

1) $m_{p,2i} > \tilde{m}_{w,2i}$ for $2 \le i < i_1$, and $m_{p,2i} < \tilde{m}_{w,2i}$ for $i \ge i_1$;
2) $m_{p,2i} > \tilde{m}_{w,2i}$ for $2 \le i < i_1$, $m_{p,2i_1} = \tilde{m}_{w,2i_1}$ and $m_{p,2i} < \tilde{m}_{w,2i}$ for $i > i_1$,

where i_1 is an integer depending on p. Obviously $\tilde{m}_{w,2} = m_{p,2} = \theta$.

Proof: See Appendix IV. ∎

For example, when $I_{d-1} = 0.5$ and $\theta = 0.562223$, we plot the initial parts of the moment sequences of $P_{4,b}$, a binary-input AWGN channel and $P_{4,w}$ in Fig. 4.

Now we are ready to show the following theorem, which solves Problem 5.

Theorem 5: Under the constraints of Problem 5 the distributions $P_{4,b}$ and $P_{4,w}$ maximize and minimize the extrinsic mutual information $I(X_d; T_{E,d})$, respectively.

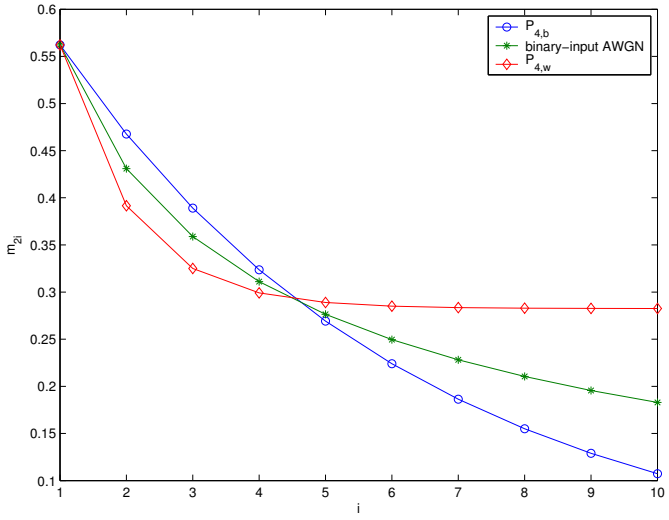


Fig. 4. The moment sequences of $P_{4,b}$, a binary-input AWGN channel and $P_{4,w}$ at $I_{d-1} = 0.5$ and $\theta = 0.562223$.

Proof: Let $\{m_{2i}\}_{i=1}^{\infty}$ be the moment sequence for T_{d-1}. Define

$$\begin{aligned} a_i &= \frac{1}{\ln 2}\frac{1}{2i(2i-1)} > 0, \\ b_i &= \prod_{k=1}^{d-2} \mathrm{E}[T_k^{2i}|X_k = 1]. \end{aligned}$$

Obviously, $1 \geq b_1 \geq b_2 \geq \cdots \geq b_i \geq \cdots \geq 0$. By (7) and (12), we have

$$\begin{aligned} I_{d-1} &= I(X_{d-1}; T_{d-1}) = \sum_{i=1}^{\infty} a_i m_{2i}, \\ I(X_d; T_{E,d}) &= \sum_{i=1}^{\infty} a_i b_i m_{2i}. \end{aligned}$$

By Lemma 12 and Lemma 9, we conclude $P_{4,b}$ maximizes $I(X_d; T_{E,d})$. Similarly, by Lemma 14 and Lemma 9, we conclude $P_{4,w}$ minimizes $I(X_d; T_{E,d})$. ■

Remark 4: According to our notation, we have $P_b = P_{4,b}$ and $P_w = P_{4,w}$.

As pointed out after the proof of Theorem 3, if $X_i \to Y_i$, $1 \leq i \leq d-2$, are all BECs with mutual information I_i, the extrinsic mutual information

$$I(X_d; T_{E,d}) = \prod_{k=1}^{d-1} I_k$$

does not depend on the channel distribution of $X_{d-1} \to Y_{d-1}$. It means the gap between the maximum and minimum of $I(X_d; T_{E,d})$ is zero. On the other hand, if any channel among $X_i \to Y_i$, $1 \leq i \leq d-2$ is not a BEC, by Lemma 9, $P_{4,b}$ for $X_{d-1} \to Y_{d-1}$ strictly maximizes the extrinsic mutual information $I(X_d; T_{E,d})$, and $P_{4,w}$ strictly minimizes it.

Let us now consider a case where the distribution of each of the $d-1$ channels $X_k \to Y_k$, $1 \leq k \leq d-1$, is allowed to vary subject to a mutual information equality constraint I_k and a conditional second moment equality constraint θ_k. We assume that $0 < I_k < 1$ and $I_k < \theta_k < \Phi(I_k)$. By sequentially applying Theorem 5, we can conclude that when all of the $d-1$ channels are of the type P_b (with appropriate parameters), the extrinsic mutual information $I(X_d; T_{E,d})$ is strictly maximized. Similarly, when all of the $d-1$ channels are of the type P_w (with appropriate parameters), the extrinsic mutual information $I(X_d; T_{E,d})$ is strictly minimized.

B. Computation of bounds

In this section, we derive the formulae to compute the maximum and minimum of the extrinsic mutual information $I(X_d; T_{E,d})$ at a degree d check node for the following scenario: the distribution of each channel $X_k \to Y_k$, $1 \leq k \leq d-1$, is allowed to vary subject to an equality constraint on mutual information and an equality constraint on conditional second moment. For simplicity, we assume that for all $d-1$ channels, the mutual information constraint value is I_A and the m_2 constraint value is θ. In fact, this kind of constraint is typical for EXIT chart analysis of LDPC codes. We assume $0 < I_A < 1$. It implies $0 < \theta < 1$.

We first treat the nontrivial case where $I_A < \theta < \Phi(I_A)$.

We know that when all $d-1$ channels have the same channel distribution P_b (i.e., $P_{4,b}$), $I(X_d; T_{E,d})$ is maximized. Recall that for P_b, T_k for the k-th channel has three values:

$$\{-t_1, 0, t_1\},$$

where t_1 is computed by (32) as

$$t_1 = f^{-1}\left(\frac{I_A}{\theta}\right).$$

By (11), we have

$$T_{E,d} = \prod_{k=1}^{d-1} T_k \in \{-t_1^{d-1}, 0, t_1^{d-1}\}.$$

By (10), we can determine that

$$\begin{aligned} \Pr[T_{E,d} = 0 | X_d = 1] &= 1 - (1-p_0)^{d-1}, \quad (41) \\ \Pr[T_{E,d} = t_1^{d-1} | X_d = 1] &= (1-p_0)^{d-1}\frac{1+t_1^{d-1}}{2}, \quad (42) \end{aligned}$$

where p_0 is computed by (34). Then we obtain the upper bound as

$$\begin{aligned} I_{E,U} &= I(X_d; T_{E,d}) \\ &= \int_{-1}^{+1} \log_2(1+t) p(T_{E,d} = t \mid X_d = 1) dt \\ &= \left(\frac{\theta}{t_1^2}\right)^{d-1} \left[1 - h\left(\frac{1-t_1^{d-1}}{2}\right)\right]. \end{aligned} \quad (43)$$

To obtain the lower bound, all $d-1$ channels should have the same channel distribution P_w (i.e., $P_{4,w}$). Recall that for P_w, T_k for the k-th channel has four values:

$$\{-1, -s_1, s_1, 1\},$$

where s_1 is computed as

$$s_1 = g^{-1}\left(\frac{1-I_A}{1-\theta}\right).$$

Since $0 < I_A < 1$ and $I_A < \theta < \Phi(I_A)$, we have $0 < s_1 < \sqrt{\theta} < 1$. By (11), we have

$$T_{E,d} = \prod_{k=1}^{d-1} T_k \in \{-1, -s_1^{d-1}, \cdots, -s_1, s_1, \cdots, s_1^{d-1}, 1\}.$$

Due to $\Pr[T_k = -1 | X_k = 1] = 0$ and (10), when we compute the probability mass function of $T_{E,d}$ conditioned on $X_d = 1$, we can neglect any combination of T_k, $1 \le k \le d-1$, if any T_k is equal to -1. Through some computation, we have

$$\Pr(T_{E,d} = 1 | X_d = 1) = p_2^{d-1}, \tag{44}$$

and

$$\begin{aligned} &\Pr(T_{E,d} \in \{s_1^i, -s_1^i\} | X_d = 1) \\ &= \binom{d-1}{i} \frac{(1-\theta)^i (\theta - s_1^2)^{d-1-i}}{(1-s_1^2)^{d-1}}. \end{aligned} \tag{45}$$

Thus, the lower bound is

$$\begin{aligned} &I_{E,L} \\ &= I(X_d; T_{E,d}) = \int_{-1}^{+1} \log_2(1+t) p(T_{E,d} = t \mid X_d = 1) dt \\ &= \sum_{i=0}^{d-1} \binom{d-1}{i} \frac{(1-\theta)^i (\theta - s_1^2)^{d-1-i}}{(1-s_1^2)^{d-1}} \left[1 - h\left(\frac{1-s_1^i}{2}\right)\right]. \end{aligned} \tag{46}$$

For the trivial case where $I_A = \theta$, i.e., all $d-1$ channels are BECs, the lower and upper bounds coincide as I_A^{d-1}. For another trivial case where $\theta = \Phi(I_A)$, i.e., all $d-1$ channels are BSCs, the lower and upper bounds coincide as [1]

$$1 - h\left(\frac{1 - [1 - 2h^{-1}(1-I_A)]^{d-1}}{2}\right).$$

Combining the nontrivial and trivial upper bounds together, we get the complete upper bound for $I_A \le \theta \le \Phi(I_A)$ as

$$I_{E,U} = \left(\frac{\theta}{t_1^2}\right)^{d-1} \left[1 - h\left(\frac{1-t_1^{d-1}}{2}\right)\right], \tag{47}$$

where $t_1 = f^{-1}(\frac{I_A}{\theta})$. We can establish the following properties related to the complete upper bound (47). Recall that by Lemma 4

$$\Phi^{-1}(x) = 1 - h\left(\frac{1-\sqrt{x}}{2}\right),$$

and $\Phi^{-1}(x)$ a strictly increasing function on $[0,1]$.

Lemma 15: 1) Assume θ is fixed. The upper bound $I_{E,U}$ (47) is a strictly increasing function of I_A, provided that $\Phi^{-1}(\theta) \le I_A \le \theta$.

2) Assume I_A is fixed. The upper bound $I_{E,U}$ (47) is a strictly increasing function of θ, provided that $I_A \le \theta \le \Phi(I_A)$.

Proof: See Appendix V. ∎

Combining the nontrivial and trivial lower bounds together, we get the complete lower bound for $I_A \le \theta \le \Phi(I_A)$ as

$$\begin{aligned} &I_{E,L} \\ &= \left(\frac{\theta - s_1^2}{1-s_1^2}\right)^{d-1} + \left(\frac{1-\theta}{1-s_1^2}\right)^{d-1} \left[1 - h\left(\frac{1-s_1^{d-1}}{2}\right)\right] \\ &+ \sum_{i=1}^{d-2} \binom{d-1}{i} \frac{(1-\theta)^i (\theta - s_1^2)^{d-1-i}}{(1-s_1^2)^{d-1}} \left[1 - h\left(\frac{1-s_1^i}{2}\right)\right], \end{aligned} \tag{48}$$

where $s_1 = g^{-1}(\frac{1-I_A}{1-\theta})$. We can establish the following properties related to the complete lower bound (48).

Lemma 16: 1) Assume θ is fixed. The lower bound $I_{E,L}$ (48) is a strictly increasing function of I_A, provided that $\Phi^{-1}(\theta) \le I_A \le \theta$.

2) Assume I_A is fixed. The lower bound $I_{E,L}$ (48) is a strictly increasing function of θ, provided that $I_A \le \theta \le \Phi(I_A)$.

Proof: See Appendix VI. ∎

C. Bounding the CND EXIT function with Gaussian priors

In the design of LDPC codes, for instance [9], it is often assumed that the *a priori* information is conditionally Gaussian distributed. Then, the CND EXIT function is approximated by [9, (9)]

$$I_{E,CND}(I_A, d) \approx 1 - J\left(\sqrt{d-1} \cdot J^{-1}(1-I_A)\right), \tag{49}$$

which is based on the duality theorem from [8]. The $J(\cdot)$ function is the mutual information between the input and output of a binary-input AWGN channel. To facilitate the computation, both $J(\cdot)$ and $J^{-1}(\cdot)$ are approximated piecewisely by some elementary functions. In this section, we use our lower bound in (46) and upper bound in (43) to bound the CND EXIT function.

For a given $I_A \in (0,1)$, we first compute the conditional second moment of the channel soft-bit. Let us indicate it by θ. Then we use (46) and (43) to compute the bounds.

In Figure 5, for a $d = 4$ check node, we plot the upper bound (43), lower bound (46), and the approximation function (49) together. We can see that the gap between the lower and upper bounds is very small, and the approximation function is very accurate. We did the computation for many other check degrees, and made similar observations. Thus, our bounds are very tight for the CND EXIT functions with Gaussian priors, and (49) is very accurate.

VIII. Prediction of Convergence of LDPC Decoding

In this section, we will describe how the results from the extended information combining problem can be used to predict the convergence of the belief propagation decoding of LDPC codes. Later we will argue why this method has the potential to improve the prediction.

Let $\{\lambda_k\}$ and $\{\rho_k\}$ be the left edge (variable) degree distribution and right edge (check) degree distribution, respectively. The codeword of the LDPC code is sent through a memoryless

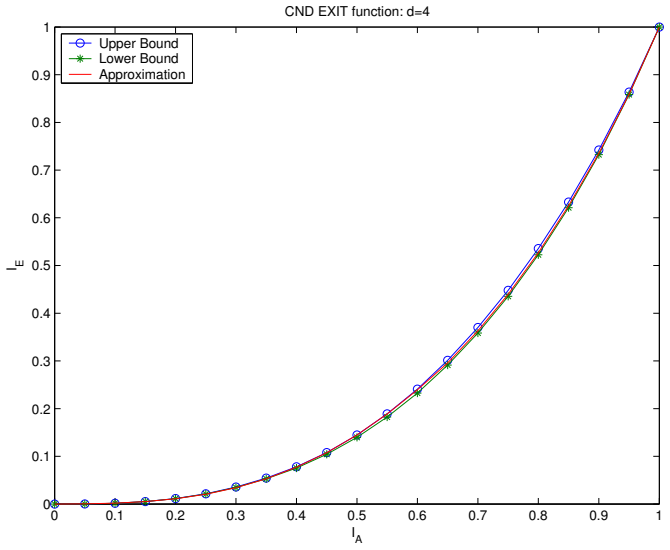


Fig. 5. For a $d = 4$ check node, compare the upper bound (43), lower bound (46), and an approximation function (49) on its EXIT function.

binary-input symmetric-output communication channel, which ensures that in the belief propagation decoding messages are symmetric [3]. Let I_c indicate the mutual information of the channel, and θ_c indicate the conditional second moment of the channel soft-bit. Recall that the relation between an LLR L and a soft-bit T is $T = \tanh(L/2)$.

First, we will describe how to obtain a best performance bound on the decoding trajectory of mutual information. At the first iteration, a variable node decoder will directly send the incoming channel LLR to each variable-to-check edge as the outgoing message. Thus the average mutual information of variable-to-check messages is

$$I_{vc}^{(1)} = \sum_{k=1}^{\infty} \lambda_k I_c = I_c,$$

and the average conditional second moment of variable-to-check messages (in the soft-bit domain) is

$$\theta_{vc}^{(1)} = \sum_{k=1}^{\infty} \lambda_k \theta_c = \theta_c.$$

Based on these two quantities, we use (47) to compute the upper bound on the mutual information of an outgoing check-to-variable message for a degree k check node. Let $I_{cv,k}^{(1)}$ be the upper bound. The upper bound on the average mutual information of check-to-variable messages is

$$I_{cv}^{(1)} = \sum_{k=1}^{\infty} \rho_k I_{cv,k}^{(1)}.$$

This concludes the first iteration.

Now we consider the second iteration of decoding. For a degree k variable node, the upper bound $I_{vc,k}^{(2)}$ on the mutual information of an outgoing variable-to-check message is obtained when all incoming check-to-variable messages are BECs; thus we have

$$I_{vc,k}^{(2)} = 1 - (1 - I_{cv}^{(1)})^{k-1} \cdot (1 - I_c).$$

Next, we need to upper bound (i.e., maximize) the conditional second moment of an outgoing variable-to-check message (in the soft-bit domain), given the mutual information $I_{cv}^{(1)}$ of incoming messages. This is an unsolved problem, and its connection to MMSE will be discussed in Section VIII-A. First, we can show that the maximum and minimum exist. It follows from sequential compactness of the space of symmetric distributions [25] (under point-wise convergence at points of continuity of the cumulative distributions) and continuity of mutual information and the moments. Assume the upper bound $\theta_{vc,k}^{(2)}$ is

$$\theta_{vc,k}^{(2)} = \varphi(I_{cv}^{(1)}, I_c, k), \tag{50}$$

where $\varphi(\cdot,\cdot,\cdot)$ needs to be determined. The upper bound on the average conditional second moment of variable-to-check messages (in the soft-bit domain) is

$$\theta_{vc}^{(2)} = \sum_{k=1}^{\infty} \lambda_k \theta_{vc,k}^{(2)}.$$

The upper bound on the average mutual information of variable-to-check messages is

$$I_{vc}^{(2)} = \sum_{k=1}^{\infty} \lambda_k I_{vc,k}^{(2)}.$$

Based on these two quantities, for a degree k check node, we use (47) to compute the upper bound $I_{cv,k}^{(2)}$ on the mutual information of an outgoing check-to-variable message. The upper bound on the average mutual information of check-to-variable messages is

$$I_{cv}^{(2)} = \sum_{k=1}^{\infty} \rho_k I_{cv,k}^{(2)}.$$

This concludes the second iteration.

The computation for the remaining iterations is the same as that for the second iteration. The following property is useful to ensure that this procedure leads to a best performance bound. The proof of the property was told to us by T. Richardson.

Lemma 17: (T. Richardson) For given I_c and k, $\varphi(I_{cv}, I_c, k)$ in (50) is a nondecreasing function of I_{cv}.

Proof: Let $I_{cv,2} > I_{cv,1}$ and f, g be the maximizing LLR distributions for $I_{cv,1}$ and $I_{cv,2}$ respectively. Let r be the distribution for channel message. For $I_{cv,1}$, the output LLR has the distribution $\mathrm{r} \otimes \mathrm{f}^{\otimes(k-1)}$, where $\otimes$ stands for convolution of distributions. Let θ_1 be the associated second moment. Similarly, let θ_2 be the associated second moment to $\mathrm{r}\otimes\mathrm{g}^{\otimes(k-1)}$. One can upgrade f to g (this concept is established in [25]). It follows that $\mathrm{r}\otimes\mathrm{g}^{\otimes(k-1)}$ is upgraded with respect to $\mathrm{r}\otimes\mathrm{f}^{\otimes(k-1)}$. Since the second moment of soft-bit is monotonic under degradation (Chapter 4, [25]), it follows that the second moment θ_2 of $\mathrm{r}\otimes\mathrm{g}^{\otimes(k-1)}$ is greater than or equal to the second moment θ_1 of $\mathrm{r} \otimes \mathrm{f}^{\otimes(k-1)}$. ■

Remark 5: Similarly, one can show that the minimum second moment is a nondecreasing function of I_{cv}.

Another property needs to be shown is as follows.

Lemma 18: If (I_2, θ_2) and (I_1, θ_1) satisfy

$$\begin{aligned} I_2 &> I_1, \quad \theta_2 > \theta_1, \\ I_2 &\leq \theta_2 \leq \Phi(I_2), \quad I_1 \leq \theta_1 \leq \Phi(I_1), \end{aligned}$$

the upper bound value in (47) for (I_2, θ_2) is strictly larger than the value for (I_1, θ_1).

Proof: In the first step, we define a new pair

$$I^{(1)} = I_1, \quad \theta^{(1)} = \min\{\Phi(I_1), \theta_2\}.$$

Since $\theta^{(1)} \geq \theta_1$, by Lemma 15, the upper bound value for $(I^{(1)}, \theta^{(1)})$ is not smaller than the value for (I_1, θ_1). In the second step, define

$$I^{(2)} = \min\{I_2, \theta^{(1)}\}, \quad \theta^{(2)} = \theta^{(1)}.$$

Since $\theta^{(1)} > I_1$ and $I_2 > I_1$, we have $I^{(2)} > I_1 = I^{(1)}$. By Lemma 15, the upper bound value for $(I^{(2)}, \theta^{(2)})$ is strictly larger than the value for $(I^{(1)}, \theta^{(1)})$. In general, for $i \geq 1$, define

$$\begin{aligned} (I^{(2i+1)}, \theta^{(2i+1)}) &= (I^{(2i)}, \min\{\Phi(I^{(2i)}), \theta_2\}), \\ (I^{(2i+2)}, \theta^{(2i+2)}) &= (\min\{I_2, \theta^{(2i+1)}\}, \theta^{(2i+1)}). \end{aligned}$$

Keep increasing i until the newest pair becomes (I_2, θ_2). Starting from the second step, the upper bound value keeps strictly increasing. Since it requires at least two steps to reach (I_2, θ_2), this lemma is proved. ∎

Remark 6: The trajectory of the pairs on the (θ, I) plane is a staircase.

Thus, if the upper bound function $\varphi(\cdot, \cdot, \cdot)$ in (50) can be determined, one gets a best performance bound, which can be used to derive a necessary condition for successful decoding. Next, we will show that this bound is tighter than the best performance bound in [17]. At any iteration, assume the upper bound on the average mutual information of variable-to-check messages is

$$I_{vc} = \sum_{k=1}^{\infty} \lambda_k I_{vc,k}.$$

Consider a degree d check node. In [17], the upper bound on the mutual information of check-to-variable messages is obtained when all incoming messages are BSCs with mutual information I_{vc}. In our method, we take into account another quantity θ_{vc}, which is an upper bound on the average conditional second moment of variable-to-check messages (in the soft-bit domain):

$$\theta_{vc} = \sum_{k=1}^{\infty} \lambda_k \theta_{vc,k}.$$

If we can show that $\theta_{vc} < \Phi(I_{vc})$, then by Lemma 15, the upper bound value in (47) for (I_{vc}, θ_{vc}) is smaller than the upper bound value for $(I_{vc}, \Phi(I_{vc}))$, which corresponds to the case of all BSCs. In other words, our method gives a better bound. Thus it is enough to show $\theta_{vc} < \Phi(I_{vc})$. By Theorem 1 and Lemma 4, it is easy to see that

$$I_{vc,k} \leq \theta_{vc,k} \leq \Phi(I_{vc,k}).$$

To achieve $I_{vc,k}$, there must be at most one non-BEC channel among the incoming message channels and the communication channel. It means that the outgoing message cannot be a BSC. Thus $\theta_{vc,k} < \Phi(I_{vc,k})$. Now we have

$$\begin{aligned} \Phi(I_{vc}) &= \Phi(\sum_{k=1}^{\infty} \lambda_k I_{vc,k}) \\ &\geq \sum_{k=1}^{\infty} \lambda_k \Phi(I_{vc,k}) \\ &> \sum_{k=1}^{\infty} \lambda_k \theta_{vc,k} = \theta_{vc}, \end{aligned}$$

where the second step is due to the concavity of $\Phi(x)$ in Lemma 4. Thus, we have shown our method potentially gives a better best performance bound.

To obtain a worst performance bound, the procedure is similar. At a variable node, the lower bound on the mutual information of an outgoing message is obtained when all incoming message channels are BSCs, and computed according to (31). Similarly, we need to find a lower bound (i.e., minimum) $\tilde{\theta}_{vc,k}$ on the conditional second moment of an outgoing variable-to-check message (in the soft-bit domain), given the mutual information $\tilde{I}_{cv}$ of incoming messages. At a check node, the lower bound on the mutual information of an outgoing message is computed by (48). We can also show a similar version of Lemma 18 for the lower bound (48) and that our method potentially gives a better worst performance bound, compared to [17]. The key is to show $\tilde{\theta}_{vc} > \tilde{I}_{vc}$.

Finally, the independence assumption on channels in the original and extended information combining problems is satisfied since it is well known that the decoding neighborhoods (for any fixed depth) of a very long ensemble of LDPC codes are asymptotically tree-like [5].

A. Bounding the conditional second moment at a variable node

As pointed out before, we need to solve the following problem:

Problem 8: Assume the communication channel is given. For a degree d variable node, the first $d-1$ incoming message channels have the same mutual information I_A, but their distributions are unknown. Let $T_{E,d}$ indicate the soft-bit of the d-th outgoing variable-to-check message. Find the maximum and minimum of the conditional second moment m_2 of $T_{E,d}$.

Let X indicate the BPSK modulated code bit represented by that variable node and Y_i indicate the incoming messages. As shown in Section IV,

$$\text{MMSE} = 1 - m_2.$$

Thus the optimization of m_2 is equivalent to the optimization of MMSE for the estimation of X from $Y_1, \cdots, Y_{d-1}$. Intuitively, as I_A increases, both the maximum and minimum of MMSE will decrease. It means both the maximum and minimum of m_2 will increase, as I_A increases.

IX. Conclusions

In this paper, we have studied an extremal problem of moments. Among all binary-input symmetric-output channels

with a fixed mutual information value, the BSC and BEC have been shown to maximize and minimize the conditional second moment of the channel soft-bit, respectively. We have also determined the ordering between the moment sequences of the BSC (or BEC) and an arbitrary binary-input symmetric-output channel (with the same mutual information). The connection between the extremal problem of moments and a related MMSE estimation problem has been found.

We have used the results on moments to solve the information combining problem (of the type in [17]) at the check nodes of LDPC codes. Namely, for the information combining at a check node of an LDPC code, the BSC and BEC are the most and least informative channels, respectively. Aiming at obtaining a better prediction of the convergence behavior of the belief propagation decoding of LDPC codes, we have extended the information combining problem at the check nodes by adding a constraint on the conditional second moment of the channel soft-bit. To solve the extension problem, first, an extremal problem with respect to the conditional fourth moment was solved, then the ordering between the moment sequences of an extremal distribution and an arbitrary distribution (of course, satisfying the constraints) was determined. We have also derived the formulae for computing the maximum and minimum of the extension problem. The results have been used to bound the CND EXIT function with Gaussian priors. Along with the results of information combining at the variable nodes (the BEC and BSC are the most and least informative channels, respectively), the results from the extension problem have been used to derive the best and worst performance bounds of the belief propagation decoding. The performance bounds are expected to be better than the performance bounds derived in [17], provided that Problem 8 be solved. We will leave this as future work.

X. Acknowledgment

The authors are grateful to the associate editor Tom Richardson for his proof of the existence of maximum and minimum of second moment at the variable node and their nondecreasing property, and useful comments from the reviewers.

Appendix I
Tchebycheff Systems

Tchebycheff systems have broad applications in a wide variety of subjects, for example, the theory of approximations, boundary value problems, and the theory of inequalities. For a good introduction into the theory of Tchebysheff systems see for example [21], [22]. In our research on extremal problems of moments and information combining, Tchebycheff system theory plays an important role. For the sake of completeness, in this section, we will summarize some of the key concepts and results of the theory.

Definition 1: A set of real continuous functions $\{u_i(t)\}_{i=0}^n$ defined on a real closed finite interval $[a,b]$ is called a Tchebycheff system (T-system) [21], [22] if every nontrivial real linear combination $\sum_{i=0}^n a_i u_i(t)$ has at most n distinct zeros in $[a,b]$.

It is easy to show that $\{u_i(t)\}_{i=0}^n$ is a T-system if and only if the determinant

$$\det \begin{pmatrix} u_0(t_0) & u_0(t_1) & \cdots & u_0(t_n) \\ u_1(t_0) & u_1(t_1) & \cdots & u_1(t_n) \\ \vdots & \vdots & \ddots & \vdots \\ u_n(t_0) & u_n(t_1) & \cdots & u_n(t_n) \end{pmatrix} \tag{51}$$

does not vanish whenever $a \le t_0 < t_1 < \cdots < t_n \le b$. Since the determinant (51) is a continuous function of t_i, it is equivalent to requiring that the determinant (51) maintains a fixed strict sign over variations of t_i. Without loss of generality we will assume that the sign is positive. Thus we have the following equivalent definition of T-systems.

Definition 2: $\{u_i(t)\}_{i=0}^n$ is called a T-system if the determinant (51) is strictly positive whenever $a \le t_0 < t_1 < \cdots < t_n \le b$.

The following definition introduces an important concept of distribution.

Definition 3: [21, p.15] A distribution $\sigma(t)$ on $[a,b]$ is a nondecreasing, right-continuous (except at the left endpoint a) function. For a distribution $\sigma(t)$, the mass at a point $\xi \in (a,b]$ is $\sigma(\xi) - \sigma(\xi - 0)$, and the mass at the left endpoint a is $\sigma(a+0) - \sigma(a)$.

The notations $\sigma(\xi - 0)$ and $\sigma(a+0)$ indicate the left and right limits, respectively. Note that a distribution is not necessarily a probability distribution (total mass on $[a,b]$ is 1, i.e., $\int_a^b d\sigma = 1$).

The moment space $\mathcal{M}_{n+1}$ induced by the T-system $\{u_i(t)\}_{i=0}^n$ is

$$\mathcal{M}_{n+1} = \left\{ (c_0, \cdots, c_n) \mid c_i = \int_a^b u_i(t) d\sigma(t),\ \sigma \in \mathbb{S}_d,\ 0 \le i \le n \right\}, \tag{52}$$

where $\mathbb{S}_d$ is the set of all valid distributions (the subscript "d" indicates "distribution"). One can show that geometrically the moment space $\mathcal{M}_{n+1}$ is a closed convex cone.

Assume $\underline{c} = (c_0, c_1, \ldots, c_n) \in \mathcal{M}_{n+1}$ (such a $\underline{c}$ is called a "positive sequence" in [21]). Define a set $V(\underline{c})$ to be

$$V(\underline{c}) = \{\sigma | \int_a^b u_i(t) d\sigma(t) = c_i, 0 \le i \le n\}.$$

In general, $V(\underline{c})$ contains either one distribution or infinitely many distributions. $V(\underline{c})$ consists of a single distribution if and only if $\underline{c}$ is a boundary point of $\mathcal{M}_{n+1}$ (such $\underline{c}$ is called a "singularly positive sequence" [21]). $V(\underline{c})$ consists of infinitely many distributions if and only if $\underline{c}$ is an interior point of $\mathcal{M}_{n+1}$ (denoted by $\underline{c} \in \mathrm{Int}\mathcal{M}_{n+1}$) (such $\underline{c}$ is called a "strictly positive sequence" [21]).

Assume $\{u_i(t)\}_{i=0}^n$ is a T-system. Let σ be a distribution in $V(\underline{c})$. If $\sigma(t)$ is a distribution with finitely many points of increase $a \le t_1 < t_2 < \cdots < t_m \le b$, we have

$$c_i = \int_a^b u_i(t) d\sigma(t) = \sum_{k=1}^m \rho_k u_i(t_k), \quad 0 \le i \le n, \tag{53}$$

where ρ_k is the mass at the point t_k. We will say that (53) is a representation of $\underline{c}$. The points $\{t_k\}_1^m$ are called roots of

the representation and $\sigma(t)$ is called the distribution associated with the representation. As in [21], an index function $\epsilon(t)$ is defined as

$$\epsilon(t) = \begin{cases} 2, & a < t < b, \\ 1, & t = a,\ b. \end{cases}$$

The index of the representation (53) is defined as the sum $\sum_{k=1}^{m} \epsilon(t_k)$. Obviously, it is equal to either $2m$, $2m-1$, or $2m-2$.

Lemma 19: ([22, Theorem 2.1] or [21, Theorem 4.1, Sec. III.4]) $\underline{c}$ is a boundary point of $\mathcal{M}_{n+1}$ if and only if it admits a representation of index not greater than n.

If $\underline{c}$ is a boundary point of $\mathcal{M}_{n+1}$ then, since $V(\underline{c})$ contains only one distribution, $\underline{c}$ admits a unique representation. If $\underline{c} \in \mathrm{Int}\mathcal{M}_{n+1}$, then $\underline{c}$ admits infinitely many representations. Among them, there are two important types of representations, namely, a canonical representation and a principal representation. If the index of a representation is not greater than $n+2$, the representation is said to be canonical. If the index is equal to $n+1$, the representation is said to be principal. By Lemma 19, it is obvious that $n+1$ is the smallest possible index for an interior point of $\mathcal{M}_{n+1}$. Furthermore, if a canonical or principal representation has a root at the right endpoint b, it is further called an upper canonical representation or an upper principal representation. On the other hand, if b is not a root, it is further called a lower canonical representation or a lower principal representation.

Theorem 6: ([22, Coro. 3.1, Sec. II.3] or [21, Theorem 5.1, Sec. III.5]) For each $\underline{c} \in \mathrm{Int}\mathcal{M}_{n+1}$, there exist exactly one lower principal representation and exactly one upper principal representation. The roots of these two representations strictly interlace.

The following theorem characterizes the roots of these two principal representations.

Theorem 7: (p.77 of [21], p.45 of [22]): The roots of lower and upper principal representations have the following properties:

(1). For n odd ($n = 2q-1$):

- **lower principal representation**: all mass is concentrated at q interior points of $[a, b]$, i.e., $a < t_1 < t_2 < \cdots < t_q < b$;
- **upper principal representation**: all mass is concentrated at $q-1$ interior points of $[a, b]$, and at both endpoints a, b, i.e., $a = s_1 < s_2 < \cdots < s_q < s_{q+1} = b$.

(2). For n even ($n = 2q$):

- **lower principal representation**: all mass is concentrated at q interior points of $[a, b]$, and at the endpoint a, i.e., $a = t_1 < t_2 < \cdots < t_{q+1} < b$;
- **upper principal representation**: all mass is concentrated at q interior points of $[a, b]$, and at the endpoint b, i.e., $a < s_1 < s_2 < \cdots < s_q < s_{q+1} = b$.

It is easy to verify that all the above forms of representations have an index equal to $n+1$.

Let $\Omega(t)$ be a continuous function and $\{u_i(t)\}_{i=0}^{n}$ be a T-system and let $\underline{c} \in \mathcal{M}_{n+1}$. The following theorem determines the maximum or minimum of the integral

$$\int_a^b \Omega(t) d\sigma(t),$$

where $\sigma(t)$ belongs to the set $V(\underline{c})$. For convenience, we define $u_{n+1}(t) = \Omega(t)$.

Theorem 8: ([21, Theorem 1.1, Sec. IV.1], [22, Theorem 1.1, Sec. III.1]) Let $\underline{c} = \{c_0, c_1, \cdots, c_n\} \in \mathrm{Int}\mathcal{M}_{n+1}$. If both $\{u_i(t)\}_{i=0}^{n}$ and the augmented system $\{u_i(t)\}_{i=0}^{n+1}$ are T-systems, $\max\limits_{\sigma \in V(\underline{c})} \int_a^b \Omega(t) d\sigma(t)$ is attained uniquely for the distribution σ^* associated with the upper principal representation of $\underline{c}$, and $\min\limits_{\sigma \in V(\underline{c})} \int_a^b \Omega(t) d\sigma(t)$ is attained uniquely for the distribution σ_* associated with the lower principal representation of $\underline{c}$.

It is remarkable that as long as the augmented system $\{u_i(t)\}_{i=0}^{n+1}$ is a T-system, the maximizing and the minimizing distributions (σ^* and σ_*) are independent of the function $\Omega(t)$.

Appendix II
Proof of Lemma 2

Let $0 \le t_0 < t_1 < t_2 < t_3 \le 1$ and $a_k = t_k^2$ for $0 \le k \le 3$. For the system $\{u_0(t) = 1, u_1(t) = t^2, u_2(t) = t^{2n}\}$ with $n \ge 2$, we have

$$\begin{aligned} \Delta &= \det \begin{pmatrix} u_0(t_0) & u_0(t_1) & u_0(t_2) \\ u_1(t_0) & u_1(t_1) & u_1(t_2) \\ u_2(t_0) & u_2(t_1) & u_2(t_2) \end{pmatrix} \\ &= \det \begin{pmatrix} 1 & 1 & 1 \\ t_0^2 & t_1^2 & t_2^2 \\ t_0^{2n} & t_1^{2n} & t_2^{2n} \end{pmatrix} \\ &= \prod_{0 \le i < j \le 2} (a_j - a_i) \cdot \sum_{i_0 + i_1 + i_2 = n-2} a_0^{i_0} a_1^{i_1} a_2^{i_2}, \end{aligned}$$

where i_0, i_1 and i_2 are nonnegative integers. Thus $\Delta > 0$ and $\{1, t^2, t^{2n}\}$ is a T-system.

For the system $\{u_0(t) = 1, u_1(t) = t^2, u_2(t) = t^4, u_3(t) = t^{2n}\}$ with $n \ge 3$, it can be shown that

$$\begin{aligned} \Delta &= \det \begin{pmatrix} 1 & 1 & 1 & 1 \\ t_0^2 & t_1^2 & t_2^2 & t_3^2 \\ t_0^4 & t_1^4 & t_2^4 & t_3^4 \\ t_0^{2n} & t_1^{2n} & t_2^{2n} & t_3^{2n} \end{pmatrix} \\ &= \prod_{0 \le i < j \le 3} (a_j - a_i) \cdot \sum_{i_0 + i_1 + i_2 + i_3 = n-3} a_0^{i_0} a_1^{i_1} a_2^{i_2} a_3^{i_3}, \end{aligned}$$

where i_k, $0 \le k \le 3$ are nonnegative integers. Thus $\Delta > 0$ and $\{1, t^2, t^4, t^{2n}\}$ is a T-system on $[0, 1]$.

A more general proof is as following. Let $A = [a_{i,j}]$ be a $n \times n$ matrix with entries $a_{i,j} = t_{j-1}^{r_{i-1}}$, where $0 \le t_0 < t_1 < \cdots < t_{n-1} \le 1$ and integers r_i satisfy $0 \le r_0 < r_1 < \cdots < r_{n-1}$. Let $V = [v_{i,j}]$ be the Vandermonde matrix with $v_{i,j} = t_{j-1}^{i-1}$. By the theory of generalized Vandermonde determinants [27], we have

$$\det(A) = \det(V) s(t_0, t_1, \cdots, t_{n-1}),$$

where $s(t_0, t_1, \cdots, t_{n-1})$ is a homogeneous symmetric function [28] with positive coefficients (called the Schur function). Since $t_0 < t_1 < \cdots < t_{n-1}$, we have $\det(V) > 0$. Furthermore, since all t_i belong to $[0, 1]$, we conclude that $s(t_0, t_1, \cdots, t_{n-1}) > 0$. Thus $\det(A) > 0$.

APPENDIX III
PROOF OF LEMMA 9

For convenience, we call $\sum_{i=1}^{\infty} a_i b_i z_i$ the target value for a sequence $\{z_i\}$. First we prove the following lemma.

Lemma 20: If $a_1 > 0$, $a_2 > 0$, $b_1 \geq b_2 \geq 0$, $x_1 > y_1 \geq 0$, $0 \leq x_2 < y_2$, $a_1 x_1 + a_2 x_2 = a_1 y_1 + a_2 y_2$, then

$$a_1 b_1 x_1 + a_2 b_2 x_2 \geq a_1 b_1 y_1 + a_2 b_2 y_2.$$

A necessary and sufficient condition to achieve equality is $b_1 = b_2$.

Proof:

$$\begin{aligned} &a_1 b_1 x_1 + a_2 b_2 x_2 - (a_1 b_1 y_1 + a_2 b_2 y_2) \\ =&b_1(a_1 x_1 + a_2 x_2) + a_2 x_2 (b_2 - b_1) \\ &- b_1(a_1 y_1 + a_2 y_2) - a_2 y_2 (b_2 - b_1) \\ =&a_2(x_2 - y_2)(b_2 - b_1) \geq 0. \end{aligned}$$

Since $a_2 > 0$ and $x_2 - y_2 < 0$, the necessary and sufficient condition to achieve equality is $b_2 - b_1 = 0$. ∎

Pick the pairs (x_1, x_{k+1}) and (y_1, y_{k+1}). If $a_1 x_1 + a_{k+1} x_{k+1} \geq a_1 y_1 + a_{k+1} y_{k+1}$, then define $\tilde{y}_{k+1} = x_{k+1}$ and solve

$$a_1 \tilde{y}_1 + a_{k+1} \tilde{y}_{k+1} = a_1 y_1 + a_{k+1} y_{k+1} \tag{54}$$

to obtain that

$$\tilde{y}_1 = y_1 + \frac{a_{k+1}}{a_1}(y_{k+1} - x_{k+1}) \leq x_1.$$

It is easy to see that $\tilde{y}_{k+1} < y_{k+1}$ and $\tilde{y}_1 > y_1$. Along with (54) and Lemma 20, we conclude that

$$a_1 b_1 \tilde{y}_1 + a_{k+1} b_{k+1} \tilde{y}_{k+1} \geq a_1 b_1 y_1 + a_{k+1} b_{k+1} y_{k+1}. \tag{55}$$

Overall, the new sequence $\{y_j^{(1)}\}_{j=1}^{\infty} = (\tilde{y}_1, y_2, \cdots, y_k, \tilde{y}_{k+1}, y_{k+2}, \cdots)$ satisfies

$$\sum_{i=1}^{\infty} a_i y_i^{(1)} = \sum_{i=1}^{\infty} a_i y_i, \tag{56}$$

$$\sum_{i=1}^{\infty} a_i b_i y_i^{(1)} \geq \sum_{i=1}^{\infty} a_i b_i y_i. \tag{57}$$

If $a_1 x_1 + a_{k+1} x_{k+1} < a_1 y_1 + a_{k+1} y_{k+1}$, then define $\tilde{y}_1 = x_1$ and solve (54) to obtain that

$$\tilde{y}_{k+1} = y_{k+1} + \frac{a_1}{a_{k+1}}(y_1 - x_1) > x_{k+1}.$$

It is easy to see that $\tilde{y}_{k+1} < y_{k+1}$ and $\tilde{y}_1 > y_1$. So similarly, (55), (56), and (57) hold. Therefore, the operation on the pair (y_1, y_{k+1}) leads to a new pair and sequence which has a larger target value as shown in (57).

Next, if $y_1^{(1)} < x_1$, we pick the pairs (x_1, x_{k+2}) and $(y_1^{(1)}, y_{k+2}^{(1)})$; if $y_{k+1}^{(1)} > x_{k+1}$, we pick the pairs (x_2, x_{k+1}) and $(y_2^{(1)}, y_{k+1}^{(1)})$; if $y_1^{(1)} = x_1$ and $y_{k+1}^{(1)} = x_{k+1}$, we pick the pairs (x_2, x_{k+2}) and $(y_2^{(1)}, y_{k+2}^{(1)})$. Then we repeat the same process as described above to obtain a new sequence $\{y_j^{(2)}\}_{j=1}^{\infty}$ satisfying

$$\sum_{i=1}^{\infty} a_i y_i^{(2)} = \sum_{i=1}^{\infty} a_i y_i^{(1)}, \tag{58}$$

$$\sum_{i=1}^{\infty} a_i b_i y_i^{(2)} \geq \sum_{i=1}^{\infty} a_i b_i y_i^{(1)}. \tag{59}$$

As the operations continue, $\{y_j^{(m)}\}_{j=1}^{\infty}$ converges to $\{x_j\}_{j=1}^{\infty}$ as m goes to infinity, and the target value $\sum_{i=1}^{\infty} a_i b_i y_i^{(m)}$ is nondecreasing in m. It follows that

$$S_y = \sum_{i=1}^{\infty} a_i b_i y_i \leq S_x = \sum_{i=1}^{\infty} a_i b_i x_i.$$

In order to achieve $S_x = S_y$, in each step of operation, the target value of the newly obtained sequence should remain the same as the old sequence. By Lemma 20, it requires $b_1 = b_2 = \cdots = b_n = b_{n+1} = \cdots$. On the other hand, if $b_1 = b_2 = \cdots = b_n = b_{n+1} = \cdots$, it is trivially true that $S_x = S_y$.

This concludes the proof.

APPENDIX IV
PROOF OF LEMMA 14

Define

$$\begin{aligned} \Delta_{w,2i} &= \tilde{m}_{w,2i} - \tilde{m}_{w,2i+2} \\ &= (s_1^{2i} p_{w,1} \frac{2}{1+s_1} + p_2) - (s_1^{2i+2} p_{w,1} \frac{2}{1+s_1} + p_2) \\ &= s_1^{2i} \cdot 2 p_{w,1}(1 - s_1), \end{aligned}$$

thus $\{\Delta_{w,2i}\}$ is a geometric sequence. Define $\Delta_{p,2i} = m_{p,2i} - m_{p,2i+2}$.

Lemma 21: There are two possible types of ordering for $\{\Delta_{w,2i}\}_{i=1}^{\infty}$ and $\{\Delta_{p,2i}\}_{i=1}^{\infty}$:

1) $\Delta_{w,2i} > \Delta_{p,2i}$ for $1 \leq i < k_0$, and $\Delta_{w,2i} < \Delta_{p,2i}$ for $i \geq k_0$;
2) $\Delta_{w,2i} > \Delta_{p,2i}$ for $1 \leq i < k_0$, $\Delta_{w,2k_0} = \Delta_{p,2k_0}$ and $\Delta_{w,2i} < \Delta_{p,2i}$ for $i > k_0$,

where k_0 is an integer depending on p.

Proof: Obviously, we have

$$\sum_{i=1}^{\infty} \Delta_{w,2i} = \tilde{m}_{w,2} = \theta = m_{p,2} = \sum_{i=1}^{\infty} \Delta_{p,2i}. \tag{60}$$

By Theorem 4, we obtain

$$\Delta_{w,2} = \theta - \tilde{m}_{w,4} > \theta - m_{p,4} = \Delta_{p_2}.$$

Based on these two facts, the remaining proof is similar to the proof of Lemma 7. ∎

Since $p \in \tilde{\mathbb{S}}_1$, we have $\tilde{m}_{w,2} = m_{p,2} = \theta$ and

$$\begin{aligned} \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} m_{p,2i} &= I_{d-1} \\ &= \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} \tilde{m}_{w,2i}. \end{aligned} \tag{61}$$

By Theorem 4, one has $\tilde{m}_{w,4} < m_{p,4}$. Thus for (61) to hold, after the fourth moment, there should exist at least one moment for p to be not greater than its counterpart for $P_{4,w}$. Letting the smallest index for such a moment be $2i_1$, then there are two possible cases:

1) $\tilde{m}_{w,2i_1} > m_{p,2i_1}$.
   By the definition of i_1, one has $\tilde{m}_{w,2i} < m_{p,2i}$ for $2 \leq i < i_1$. It is easy to see that

$$\begin{aligned} \Delta_{w,2i_1-2} &= \tilde{m}_{w,2i_1-2} - \tilde{m}_{w,2i_1} \\ &< m_{p,2i_1-2} - m_{p,2i_1} = \Delta_{p,2i_1-2}. \end{aligned} \tag{62}$$

By Lemma 21, one has

$$\Delta_{w,2i_1} < \Delta_{p,2i_1}.$$

Then one obtains

$$\begin{aligned}\tilde{m}_{w,2i_1+2} &= \tilde{m}_{w,2i_1} - \Delta_{w,2i_1} \\ &> m_{p,2i_1} - \Delta_{p,2i_1} = m_{p,2i_1+2}.\end{aligned} \tag{63}$$

Using the above argument repeatedly and sequentially for $i \geq i_1 + 2$, one can conclude that $\tilde{m}_{w,2i} > m_{p,2i}$ for $i \geq i_1$. This concludes the proof of the first case of Lemma 14.

2) $\tilde{m}_{w,2i_1} = m_{p,2i_1}$.
The remaining proof is exactly the same as in the first case.

Thus this lemma is proved.

Appendix V
Proof of Lemma 15

We know that I_A and θ should satisfy $I_A \leq \theta \leq \Phi(I_A)$. Thus, for a fixed θ, I_A should satisfy $\Phi^{-1}(\theta) \leq I_A \leq \theta$. By definition of $f(x)$ in (35), we have

$$f(t_1^{d-1}) = \frac{1 - h[(1 - t_1^{d-1})/2]}{(t_1^{d-1})^2}.$$

Thus,

$$1 - h[(1 - t_1^{d-1})/2] = f(t_1^{d-1})(t_1^{d-1})^2.$$

The upper bound (47) can be rewritten as

$$\begin{aligned}I_{E,U} &= \left(\frac{\theta}{t_1^2}\right)^{d-1} f(t_1^{d-1})(t_1^{d-1})^2 \\ &= \theta^{d-1} f(t_1^{d-1}).\end{aligned}$$

As I_A increases and θ fixed, $t_1 = f^{-1}(\frac{I_A}{\theta})$ strictly increases and so does $f(t_1^{d-1})$. Thus $I_{E,U}$ is a strictly increasing function of I_A.

Now fix I_A and allow θ to change provided that $I_A \leq \theta \leq \Phi(I_A)$. For convenience, we write the complete upper bound (47) as a function of θ, i.e., $I_{E,U}(\theta)$. The right endpoint of the feasible interval corresponds to all BSCs, which strictly maximizes the extrinsic mutual information among all channel combinations subject to mutual information constraints (see Section VI). Thus,

$$I_{E,U}(\theta = \Phi(I_A)) > I_{E,U}(\theta = x), \quad x \in [I_A, \Phi(I_A)).$$

The left endpoint corresponds to all BECs, which minimizes the extrinsic mutual information among all channel combinations subject to mutual information constraints (see Section VI). Since for $\theta > I_A$, no channel could be a BEC; therefore, we have

$$I_{E,U}(\theta = I_A) < I_{E,U}(\theta = x), \quad x \in (I_A, \Phi(I_A)].$$

It remains to show that $I_{E,U}(\theta)$ is a strictly increasing function when $\theta \in (I_A, \Phi(I_A))$. Recall that P_b is the type of maximizing distribution to Problem 5 and determined by (32)-(35). We first prove the following lemma:

Lemma 22: Assume $(X_1, \cdots, X_d)$ be a codeword of a $[d, d-1]$ single parity-check code. The first $d-2$ channels $X_i \to Y_i$, $1 \leq i \leq d-2$ are fixed and each channel has a positive mutual information. Let $\theta_2 > \theta_1$, P_2 and P_1 indicate the P_b distributions corresponding to (I_A, θ_2) and (I_A, θ_1), respectively. Let $I_{E,2}$ and $I_{E,1}$ indicate the extrinsic mutual information $I(X_d; T_{E,d})$ when $X_{d-1} \to Y_{d-1}$ has the channel distribution P_2 and P_1, respectively. It is true that $I_{E,2} > I_{E,1}$.

Proof: Since $\theta_2 > \theta_1$, by (32), P_2 has a smaller t_1 value than P_1. Since the moment sequences for P_2 and P_1 have the form $\{\theta t_1^{2i-2}\}_{i=1}^{\infty}$, i.e., a geometric sequence, it means P_2's sequence has a larger initial value but decays faster than P_1's sequence. We can easily conclude that the moment sequence of P_2 and the moment sequence of P_1 have the same two possible types of ordering as in Lemma 7. Since they have the same mutual information I_A, by Lemma 9, we have $I_{E,2} > I_{E,1}$. ∎

Recall that when all of the first $d-1$ channels have the same P_b distribution, $I(X_d; T_{E,d})$ becomes $I_{E,U}$. For θ_1, it means they all have the channel distribution P_1. For θ_2, they all have the channel distribution P_2. Starting from all P_1 distributions, we can replace a P_1 channel by a P_2 channel. By the above lemma, we get a larger extrinsic mutual information $I(X_d; T_{E,d})$. Repeat the procedure $d-1$ times and $I(X_d; T_{E,d})$ keeps strictly increasing. As a result, all $d-1$ channels have the same channel distribution P_2 and $I(X_d; T_{E,d})$ becomes the upper bound for θ_2. Therefore, $I_{E,U}$ for θ_2 has a larger value than $I_{E,U}$ for θ_1. This concludes the proof.

Appendix VI
Proof of Lemma 16

Assume θ is fixed. For convenience, we define the following function:

$$m_{2i}(I_A) = \begin{cases}\theta, & I_A = \theta \\ 2p_{w,1}s_1^{2i}/(1+s_1) + p_2, & \Phi^{-1}(\theta) < I_A < \theta \\ \theta^i, & I_A = \Phi^{-1}(\theta).\end{cases}$$

Namely, when $I_A = \theta$, $m_{2i}(I_A)$ is the $2i$-th moment for a BEC. When $I_A = \Phi^{-1}(\theta)$, $m_{2i}(I_A)$ is the $2i$-th moment for a BSC. When $\Phi^{-1}(\theta) < I_A < \theta$, $m_{2i}(I_A)$ is the $2i$-th moment for a P_w channel corresponding to (I_A, θ), i.e.,

$$\begin{aligned}m_{2i}(I_A) &= \frac{2p_{w,1}}{1+s_1}s_1^{2i} + p_2 \\ &= \frac{1-\theta}{1-s_1^2}s_1^{2i} + \frac{\theta - s_1^2}{1-s_1^2} \\ &= \theta + \sum_{k=1}^{i-1} s_1^{2k}(\theta - 1).\end{aligned}$$

For a fixed θ, it is obvious that $s_1 = g^{-1}((1-I_A)/(1-\theta))$ is a strictly decreasing function of I_A. Due to $\theta < 1$, for $i \geq 2$, $m_{2i}(I_A)$ is a strictly increasing function on $(\Phi^{-1}(\theta), \theta)$. We also have for $\Phi^{-1}(\theta) < I_A < \theta$,

$$0 < s_1 < \sqrt{\theta}$$

and

$$\theta + \sum_{k=1}^{i-1}(\sqrt{\theta})^{2k}(\theta - 1) = \theta^i < m_{2i}(I_A) < \theta.$$

Thus, we have shown that $m_{2i}(I_A)$ is a strictly increasing function on $[\Phi^{-1}(\theta), \theta]$.

Recall that the lower bound $I_{E,L}$ is equal to the extrinsic mutual information $I(X_d; T_{E,d})$ when all of the first $d-1$ channels are BECs, or of the same P_w distribution, or BSCs. Thus, by (12), $I_{E,L}$ can also be computed as

$$I_{E,L}(I_A) = \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} m_{2i}(I_A)^{d-1}.$$

As shown above, $m_{2i}(I_A)$ is a strictly increasing function of I_A. Thus we can conclude $I_{E,L}(I_A)$ is a strictly increasing function of I_A, too.

Now assume I_A is fixed. Based on the same argument as in Section V, it suffices to show that $I_{E,U}(\theta)$ is a strictly increasing function on $(I_A, \Phi(I_A))$. Let $\theta_2 > \theta_1$. P_2 and P_1 indicate the P_w distributions corresponding to (I_A, θ_2) and (I_A, θ_1), respectively (determined by (36)-(39)). We first show this lemma.

Lemma 23: Assume $(X_1, \cdots, X_d)$ be a codeword of a $[d, d-1]$ single parity-check code. The first $d-2$ channels $X_i \to Y_i$, $1 \le i \le d-2$ are fixed and each channel has a positive mutual information. Let $I_{E,2}$ and $I_{E,1}$ indicate the extrinsic mutual information $I(X_d; T_{E,d})$ when $X_{d-1} \to Y_{d-1}$ has the channel distribution P_2 and P_1, respectively. It is true that $I_{E,2} > I_{E,1}$.

Proof: Let $\{m_{2i}^{(2)}\}_{i=1}^{\infty}$ and $\{m_{2i}^{(1)}\}_{i=1}^{\infty}$ be the moment sequences for P_2 and P_1, respectively. Let $\{\Delta_{2i}^{(2)}\}_{i=0}^{\infty}$ and $\{\Delta_{2i}^{(1)}\}_{i=0}^{\infty}$ be the corresponding Δ sequence defined in Lemma 13. Since $m_2^{(2)} = \theta_2 > \theta_1 = m_2^{(1)}$, we have

$$\Delta_0^{(2)} < \Delta_0^{(1)}.$$

Both Δ sequences are geometric sequences. Since P_2 has a larger s_1 value than P_1, $\{\Delta_{2i}^{(1)}\}_{i=0}^{\infty}$ decays faster than $\{\Delta_{2i}^{(2)}\}_{i=0}^{\infty}$. Since

$$\sum_{i=0}^{\infty} \Delta_{2i}^{(1)} = \sum_{i=0}^{\infty} \Delta_{2i}^{(2)} = 1,$$

it is easy to see that these two Δ sequences have the following two possible orderings:

1) $\Delta_{2i}^{(1)} > \Delta_{2i}^{(2)}$ for $0 \le i < k_0$, and $\Delta_{2i}^{(1)} < \Delta_{2i}^{(2)}$ for $i \ge k_0$;
2) $\Delta_{2i}^{(1)} > \Delta_{2i}^{(2)}$ for $1 \le i < k_0$, $\Delta_{2k_0}^{(1)} = \Delta_{2k_0}^{(2)}$ and $\Delta_{2i}^{(1)} < \Delta_{2i}^{(2)}$ for $i > k_0$,

where k_0 is an integer depending on θ_1 and θ_2. Since $m_2^{(2)} > m_2^{(1)}$, and

$$\frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} m_{2i}^{(1)} = I_A = \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} m_{2i}^{(2)},$$

based on the same technique as in the proof of Lemma 14, it is easy to conclude that the two moment sequences have the following two possible orderings:

1) $m_{2i}^{(2)} > m_{2i}^{(1)}$ for $1 \le i < k_1$, and $m_{2i}^{(2)} < m_{2i}^{(1)}$ for $i \ge k_1$;
2) $m_{2i}^{(2)} > m_{2i}^{(1)}$ for $1 \le i < k_1$, $m_{2k_1}^{(2)} = m_{2k_1}^{(1)}$ and $m_{2i}^{(2)} < m_{2i}^{(1)}$ for $i > k_1$,

where k_1 is an integer depending on θ_1 and θ_2. Thus by Lemma 9, we have $I_{E,2} > I_{E,1}$. ■

Recall that when all of the first $d-1$ channels have the same P_w distribution, $I(X_d; T_{E,d})$ becomes $I_{E,L}$. For θ_1, it means they all have the channel distribution P_1. For θ_2, they all have the channel distribution P_2. Starting from all P_1 distributions, we can replace a P_1 channel by a P_2 channel. By the above lemma, we get a larger extrinsic mutual information $I(X_d; T_{E,d})$. Repeat the procedure $d-1$ times and $I(X_d; T_{E,d})$ keeps strictly increasing. As a result, all $d-1$ channels have the same channel distribution P_2 and $I(X_d; T_{E,d})$ becomes the lower bound for θ_2. Therefore, $I_{E,L}$ for θ_2 has a larger value than $I_{E,L}$ for θ_1. This concludes the proof.

REFERENCES

[1] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1963.

[2] B. J. Frey, R. Koetter, G. D. Forney, F. R. Kschischang, R. J. McEliece, and D. A. Spielman, Eds., *IEEE Transactions on Information Theory, Special Issue on Codes on Graphs and Iterative Algorithms*, vol. 47, no. 2, Feb. 2001.

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

[4] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, no. 5, pp. 533–547, Sept. 1981.

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

[6] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[7] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: A model and two properties," in *Proceedings of Conference on Information Sciences and Systems*, Princeton, NJ, Mar. 2002, pp. 742–747.

[8] ——, "Extrinsic information transfer functions: Model and erasure channel properties," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2657–2673, Nov. 2004.

[9] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.

[10] S. Huettinger, J. Huber, R. Johannesson, and R. Fischer, "Information processing in soft-output decoding," in *Proc. Allerton Conf. on Communications, Control, and Computing*, Monticello, IL, Oct. 2001.

[11] S. Huettinger, J. Huber, R. Fischer, and R. Johannesson, "Soft-output-decoding: Some aspects from information theory," in *Proc. Int. ITG Conf. on Source and Channel Coding*, Berlin, Germany, Jan. 2002, pp. 81–90.

[12] I. Land, S. Huettinger, P. A. Hoeher, and J. Huber, "Bounds on information combining," in *Proc. Int. Symp. on Turbo Codes & Rel. Topics*, Brest, France, Sept. 2003, pp. 39–42.

[13] ——, "Bounds on information combining," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 612–619, Feb. 2005.

[14] I. Land, P. A. Hoeher, and J. Huber, "Bounds on information combining for parity-check equations," in *Proc. Int. Zurich Seminar on Communications (IZS)*, Zurich, Switzerland, Feb. 2004, pp. 68–71.

[15] I. Land, S. Huettinger, P. A. Hoeher, and J. Huber, "Bounds on mutual information for simple codes using information combining," *Annals of Telecomm.*, vol. 60, no. 1-2, pp. 184–214, Jan. 2005.

[16] I. Sutskover, S. Shamai, and J. Ziv, "Extremes of information combining," in *Proceedings of the 41st Annual Allerton Conference on Communications, Conrol and Computing*, Monticello, IL, Oct. 2003, pp. 1446–1455.

[17] ——, "Extremes of information combining," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1313–1326, Apr. 2005.

[18] D. Burshtein and G. Miller, "Bounds on the performance of belief propagation decoding," *IEEE Trans. Inform. Theory*, vol. 48, no. 1, pp. 112–122, Jan. 2002.

[19] E. Sharon, A. Ashikhmin, and S. Litsyn, "EXIT functions for the Gaussian channel," in *Proceedings of 41st Annual Allerton Conference on Communication, Control and Computing*, Monticello, Illinois, Oct. 2003, pp. 972–981.

[20] ——, "EXIT functions for binary input memoryless symmetric channels," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1207–1214, July 2006.
[21] M. G. Krein and A. A. Nudel'man, *The Markov Moment Problem and Extremal Problems*. Providence, Rhode Island: American Mathematical Society, 1977.
[22] S. Karlin and W. J. Studden, *Tchebycheff Systems: with Applications in Analysis and Statistics*. New York: Interscience Publishers, 1966.
[23] Y. Jiang, A. Ashikhmin, R. Koetter, and A. C. Singer, "Extremal problems of information combining," in *IEEE International Symposium on Information Theory*, Adelaide, Australia, Sept. 2005, pp. 1236–1240.
[24] I. Sutskover, S. Shamai, and J. Ziv, "Constrained information combining: Theory and applications for LDPC coded systems," *IEEE Trans. Inform. Theory*, vol. 53, no. 5, pp. 1617–1643, May 2007.
[25] T. Richardson and R. Urbanke, *Modern Coding Theory*. preliminary online version, 2007.
[26] R. Durrett, *Probability: Theory and Examples*, 2nd ed. Belmont, California: Duxbury Press, 1996.
[27] R. P. Flowe and G. A. Harris, "A note on generalized Vandermonde determinants," *SIAM J. Matrix Anal. Appl.*, vol. 14, no. 4, pp. 1146–1151, Oct. 1993.
[28] I. G. Macdonald, *Symmetric functions and orthogonal polynomials*, ser. University Lecture Series. American Mathematical Society, 1998, vol. 12.

Yibo Jiang was born in Ningbo, China, in 1976. He received the B. Eng. degree in information engineering, M. Eng. degree in communications and information systems, from Zhejiang University, Hangzhou, China, in 1997 and 2000, respectively. In 2005, he received the Ph.D. degree in Electrical Engineering from University of Illinois at Urbana-Champaign. Since July 2005, he has been working on physical layer of wireless communications in Qualcomm.

Alexei Ashikhmin is a member of technical staff in the Communications and Statistical Sciences Department, Bell Laboratories, Murray Hill, New Jersey. He received his Ph.D. degree in Electrical Engineering from the Institute of Information Transmission Problems, Russian Academy of Science, Moscow, Russia, in 1994. From September 1995 to September 1996, he was with the Electrical Engineering, Mathematics and Computer Science Department, Delft University of Technology, The Netherlands. From January 1997 to July 1999, he was a Postdoctoral Fellow at Modeling, Algorithms, and Informatics Group of Los Alamos National Laboratory , Los Alamos, NM. Since 1999, he has been with Bell Laboratories. His research interests include classical and quantum information theory and communication theory From 2003 to 2006 Dr. Ashikhmin served as an Associate Editor for Coding Theory for *IEEE Transactions on Information Theory*. In 2002, Dr. Ashikhmin received *Bell Laboratories President's Gold Award* for breakthrough research resulting in the ability to deliver unprecedented wireless bit-rates. In 2005 Dr. Ashikhmin was honored by the *IEEE Communications S.O. Rice Prize Paper Award* for work on LDPC codes for information transmission with multiple antennas.

Ralf Koetter (M 91, SM 06) received a Diploma in Electrical Engineering from the Technical University Darmstadt, Germany in 1990 and a Ph.D. degree from the Department of Electrical Engineering at Linkoeping University, Sweden. From 1996 to 1997, he was a visiting scientist at the IBM Almaden Research Laboratory in San Jose, California. He was a visiting assistant professor at the University of Illinois at Urbana/Champaign and a visiting scientist at CNRS in Sophia Antipolis, France from 1997 to 1998. In the years 1999-2006 he was member of the faculty of the University of Illinois at Urbana/Champaign. In 2006 he joined the faculty of the TU Muenchen as the head of the Institute for Communications Engineering.

Dr. Koetter's research interest include coding and information theory and their application to communication systems.

In the years 1999-2001, he served as associate editor for coding theory & techniques for the IEEE Transactions on Communications. In 2003, he concluded a term as associate editor for coding theory of the IEEE Transactions on Information Theory. He received an IBM Invention Achievement Award in 1997, an NSF CAREER Award in 2000, an IBM Partnership Award in 2001 and a 2006 XEROX award for faculty research. He is co-recipient of the 2004 Paper Award of the Information Theory Society. Since 2003 he has been a member of the Board of Governors of the IEEE Information Theory Society.

Andrew C. Singer received the S.B., S.M., and Ph.D. degrees, all in electrical engineering and computer science, from the Massachusetts Institute of Technology. Since 1998, he has been on the faculty of the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign, where he is currently an Associate Professor in the ECE department, a Research Associate Professor in the Coordinated Science Laboratory and a Donald Biggar Willett Faculty Scholar. In 2005 he was appointed Director the Technology Entrepreneur Center in the College of Engineering. During the academic year 1996, he was a Postdoctoral Research Affiliate in the Research Laboratory of Electronics at MIT. From 1996 to 1998, he was a Research Scientist at Sanders, A Lockheed Martin Company in Manchester, New Hampshire. His research interests include statistical signal processing, communications, and machine learning. He was a Hughes Aircraft Masters Fellow, and was the recipient of the Harold L. Hazen Memorial Award for excellence in teaching. In 2000, he received the National Science Foundation CAREER Award, in 2001 he recieved the Xerox Faculty Research Award, and in 2002 was named a Willett Faculty Scholar. Dr. Singer serves as an Associate Editor for the IEEE Transactions on Signal Processing and is a member of the MIT Educational Council, and of Eta Kappa Nu and Tau Beta Pi. In 2000, he co-founded Intersymbol Communications, a manufacturer of high-speed integrated circuits for the optical communications industry, where he served as Chief Executive Officer. In 2007, Intersymbol Communications was acquired by Finisar Corporation (NASD:FNSR), where he now serves as a Senior Scientist.