

# Extremal Problems of Information Combining

Yibo Jiang\*, Alexei Ashikhmin<sup>†</sup>, Ralf Koetter\*, and Andrew C. Singer\*

\*Coordinated Science Laboratory

University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Email: {yjiang, koetter, acsinger}@uiuc.edu

<sup>†</sup>Bell Laboratories, Lucent Technologies

600 Mountain Avenue, Murray Hill, NJ 07974, USA

Email: aea@research.bell-labs.com

**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 extreme distributions for this new problem.

## I. INTRODUCTION

The extrinsic information transfer (EXIT) chart method [1] has proven useful in practice, for example, in the design of low-density parity-check (LDPC) codes [2], and simple since it tracks only one parameter, namely mutual information. It is usually assumed that *a priori* information is conditionally Gaussian distributed. In the belief propagation decoding of LDPC codes [3], however, the distributions of messages are often non-Gaussian. Therefore, the EXIT chart method 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 variable node decoder (VND) or a check node decoder (CND) will be useful and lead to upper and lower bounds on the decoding trajectory of mutual information. This motivates the study of information combining problems.

The notion of information combining was introduced by Huettinger *et al.* in [4], [5] for the study of concatenated coding systems. In particular, in the belief propagation decoding of LDPC codes, the processings at variable and check nodes can be interpreted as operations of information combining, i.e. combining the mutual information contained messages incoming to a node. Let  $(X_1, \dots, X_d)$  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 binary phase shift keying (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

$$T_i = p(X_i = 1 | Y_i) - p(X_i = -1 | Y_i),$$

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  $I(X_d; Y_1, Y_2, \dots, Y_{d-1})$ , i.e. the extrinsic mutual information? Such an extremal problem was first considered in [6] for a  $[3, 1]$  repetition code. It was shown that  $I(X_3; Y_1, Y_2)$  is maximized (or minimized) when both  $X_1 \rightarrow Y_1$  and  $X_2 \rightarrow Y_2$  are binary erasure channels (BECs) (or binary symmetric channels (BSCs)) with prescribed mutual information values. In [7], 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 [8], [9], [10], [11] the above results were extended to arbitrary codeword length  $d$  for both repetition codes and single parity-check codes.

In [6], [7], [8], [9], the extrinsic mutual information  $I(X_d; Y_1, Y_2, \dots, Y_{d-1})$  is optimized over all channels subject to their individual mutual information constraints. In [10], [11], the problems were generalized by optimizing with respect to a single channel. Fix the transition probability distributions of  $X_i \rightarrow Y_i$ ,  $1 \leq i \leq d-2$  and fix the mutual information of  $X_{d-1} \rightarrow Y_{d-1}$ . It was shown in [11] that for a repetition code, when  $X_{d-1} \rightarrow Y_{d-1}$  is a BEC (BSC), the extrinsic mutual information  $I(X_d; Y_1, Y_2, \dots, Y_{d-1})$  is maximized (minimized). For a single parity-check code, the roles of BEC and BSC are reversed.

All the results above were obtained by proving new mutual information inequalities and using various existing inequalities and identities of mutual information. The work by Sharon *et al.*, [12], [13], provides a framework of the moments of soft-bits and expresses both the mutual information of a binary-input symmetric-output channel and the extrinsic mutual information of a  $[d, d-1]$  single parity-check code as serieses of moments of channel soft-bits.

In this paper, we study various properties of the moments and moment sequences. We prove that among all binary-input symmetric-output channels with a fixed mutual information, the BSC (BEC) maximizes (minimizes) the second moment of channel soft-bits. We also determine the ordering between the moment sequences of a BSC (or BEC) and any other channel.

By using the properties of the moment sequences, we solve

the extremal information combining problem at the check nodes proposed in [11]. Later, we extend the problem by adding a constraint on the second moment of channel soft-bits. It is also solved by a moments approach, and the best and worst channel distributions in the sense of maximizing and minimizing the extrinsic mutual information are determined.

The following notations are used in this paper. Underscores denote vectors. The symbol  $\underline{y}_{[j]}$  is used to denote the vector obtained by deleting the  $j$ -th element of  $\underline{y}$ . Subscript  $b$  and  $w$  stand for “best” and “worst”, respectively. All proofs are omitted due to the limitation of space.

## II. T-CONSISTENCY AND MUTUAL INFORMATION

In [12], [13], the concept of T-consistency and some related mutual information results were proposed. We give a short summary for completeness.

Let  $X$  and  $Y$  be random variables at the input and output of a binary-input symmetric-output channel respectively, with transition probability density function  $f$  satisfying  $f_{Y|X}(Y|X=1) = f_{Y|X}(-Y|X=-1)$ . Define the channel soft-bit  $T$  as  $T = Pr(X=1|Y) - Pr(X=-1|Y)$ . It is easy to see that  $T = (e^L - 1)/(e^L + 1) = \tanh(L/2)$  where  $L$  is the log-likelihood-ratio (LLR). It was shown in [13] that the conditional probability density  $p(T|X)$  satisfies

$$p(T = t|X = 1) = p(T = -t|X = -1) \frac{1+t}{1-t}, \quad (1)$$

and  $p(T = t|X = 1) = p(T = -t|X = -1)$ . Such a random variable is called T-consistent. For a BSC channel 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 channel, its T variable always takes three values  $\{-1, 0, 1\}$ .

Assume  $X$  is an equiprobable binary random variable. The mutual information between  $X$  and  $T$  is [12], [13]

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

where

$$m_{2^i} = \int_{-1}^{+1} t^{2^i} p(T = t|X = 1) dt. \quad (3)$$

Note that  $0 \leq m_{2^i} \leq 1$  and  $\frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2^i(2^i - 1)} = \log_2(1+t)|_{t=1} = 1$ , therefore the right-hand-side of Eq. (2) is convergent. We point out that the singularity of  $\log_2(1+t)$  at  $t = -1$  does not affect the integral (2) since one knows  $p(T = -1|X = 1) = 0$  from (1).

Consider an  $[n, k]$  binary linear code. Its codeword is BPSK modulated and transmitted through  $n$  binary-input symmetric-output independent channels. Let  $\underline{x}$  and  $\underline{y}$  indicate the input and output vector. The extrinsic soft-bit for  $X_j$  of an *a posteriori* probability (APP) decoder is defined as

$$T_{E,j} = Pr(X_j = 1|\underline{Y}_{[j]}) - Pr(X_j = -1|\underline{Y}_{[j]}). \quad (4)$$

Let  $\underline{c}_0$  be the all-zero codeword. It is shown [13] that  $T_{E,j}$  is T-consistent and

$$p(T_{E,j} = t|X_j = 1) = p(T_{E,j} = t|\underline{c}_0 \text{ transmitted}). \quad (5)$$

Furthermore, for an  $[n, n-1]$  single parity-check code,  $T_{E,j} = \prod_{k=1, k \neq j}^n T_k$  where  $T_k = Pr(X_k = 1|Y_k) - Pr(X_k = -1|Y_k)$ . Thus the extrinsic mutual information is

$$\begin{aligned} I_{E,j} &= I(X_j; T_{E,j}) \\ &= \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2^i(2^i - 1)} \prod_{k=1, k \neq j}^n E[T_k^{2^i} | X_k = 1]. \end{aligned} \quad (6)$$

## III. TCHEBYCHEFF SYSTEM

Tchebycheff system (T-system) theory is used to solve extremal problems of moments. For completeness, we give an introduction to some of the key concepts and results of T-system theory.

For  $a < b$  finite and real, a set of real continuous functions  $\{u_i(t)\}_{i=0}^n$  defined on  $[a, b]$  is called a T-system [14], [15] 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

$$\det \begin{pmatrix} u_0(t_0) & u_0(t_1) & \cdots & u_0(t_n) \\ u_1(t_0) & u_1(t_1) & \cdots & u_1(t_n) \\ \vdots & \vdots & \ddots & \vdots \\ u_n(t_0) & u_n(t_1) & \cdots & u_n(t_n) \end{pmatrix} \quad (7)$$

does not vanish whenever  $a \leq t_0 < t_1 < \cdots < t_n \leq b$ . Since the determinant (7) is a continuous function of  $t_i$ , it is equivalent to require that the determinant (7) maintains a fixed strict sign. Without loss of generality, in [15], [14],  $\{u_i(t)\}_{i=0}^n$  is called a T-system if the determinant (7) is strictly positive whenever  $a \leq t_0 < t_1 < \cdots < t_n \leq b$ .

Let us introduce the concept of a distribution [14, p.15] on  $[a, b]$ . A distribution is a nondecreasing, right-continuous (except at the left endpoint) 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  $\sigma(a + 0) - \sigma(a)$ , where  $\sigma(\xi - 0)$  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

$$\mathcal{M}_{n+1} = \left\{ (c_0, \cdots, c_n) \mid c_i = \int_a^b u_i(t) d\sigma(t), 0 \leq i \leq n \right\} \quad (8)$$

where  $\sigma$  goes through the set of all valid distributions. Geometrically,  $\mathcal{M}_{n+1}$  is a closed convex cone.

Assume  $\underline{c} = \{c_k\}_0^n \in \mathcal{M}_{n+1}$ . Define a set  $V(\underline{c})$  to be  $V(\underline{c}) = \{\sigma \mid \int_a^b u_i(t) d\sigma(t) = c_i, 0 \leq i \leq n\}$ . In general,  $V(\underline{c})$  contains either one distribution (if and only if  $\underline{c}$  is a boundary point of  $\mathcal{M}_{n+1}$ ) or infinitely many distributions (if and only if  $\underline{c}$  is an interior point of  $\mathcal{M}_{n+1}$  (denoted by  $\underline{c} \in \text{Int} \mathcal{M}_{n+1}$ )).

Assume  $\{u_i(t)\}_{i=0}^n$  is a T-system. Let  $\sigma$  be a distribution in  $V(\underline{c})$ . If  $\sigma(t)$  has finitely many points of increase  $a \leq t_1 <$

$t_2 < \dots < t_m \leq b$ , the representation for  $\underline{c}$  becomes

$$c_i = \int_a^b u_i(t) d\sigma(t) = \sum_{k=1}^m \rho_k u_i(t_k), \quad 0 \leq i \leq n, \quad (9)$$

where  $\rho_k$  is the mass at the point  $t_k$ . The points  $\{t_k\}_1^m$  are called roots of the representation (9), and  $\sigma(t)$  is called the distribution associated with the representation (9). As in [14], an index function  $\epsilon(t)$  is defined as  $\epsilon(t) = 2$  for  $a < t < b$ , and  $\epsilon(a) = \epsilon(b) = 1$ . The index of the representation (9) is defined as the sum  $\sum_{k=1}^m \epsilon(t_k)$ . Now we introduce two important concepts, a canonical representation and a principle representation. If the index of a representation is  $\leq n+2$ , the representation is said to be canonical. If the index is  $n+1$ , the representation is said to be principle. Furthermore, if a canonical (principle) representation has a root at  $b$ , it is further called an upper canonical (principle) representation. On the other hand, if  $b$  is not a root, it is called a lower canonical (principle) representation. From [15, Coro. 3.1, Sec. II.3] or [14, Theorem 5.1, Sec. III.5], it is true that

*Theorem 1:* For each  $\underline{c} \in \text{Int}\mathcal{M}_{n+1}$ , there exists exactly one lower principle representation and exactly one upper principle representation. The roots of these two representations strictly interlace.

*Theorem 2:* (p.77 of [14], p.45 of [15]): The roots of lower and upper principle representations have the following properties (1). for  $n$  odd ( $n = 2q - 1$ ):

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

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

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

Let  $\Omega(t)$  be a continuous function. Define  $u_{n+1}(t) = \Omega(t)$ .

*Theorem 3:* ([14, Theorem 1.1, Sec. IV.1], [15, Theorem 1.1, Sec. III.1]): Let  $\underline{c} \in \text{Int}\mathcal{M}_{n+1}$ . If both  $\{u_i(t)\}_{i=0}^n$  and the augmented system  $\{u_i(t)\}_{i=0}^{n+1}$  are T-systems,  $\max_{\sigma \in V(\underline{c})} \int_a^b \Omega(t) d\sigma(t)$  is attained uniquely for the distribution  $\sigma^*$  associated with the upper principle representation of  $\underline{c}$ , and  $\min_{\sigma \in V(\underline{c})} \int_a^b \Omega(t) d\sigma(t)$  is attained uniquely for the distribution  $\sigma_*$  associated with the lower principle 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 ( $\sigma^*$  and  $\sigma_*$ ) are independent of the function  $\Omega(t)$ .

#### IV. THE EXTREMAL INFORMATION COMBINING PROBLEM

In this section, we are focusing on the type of extremal problems of information combining proposed in [11]. Let us

first consider a check node with degree  $d$ . One can treat such a node as a  $[d, d - 1]$  single parity-check code. Its codeword is BPSK modulated. Let  $\underline{x}$  and  $\underline{y}$  stand for a BPSK-modulated codeword and channel output vector respectively. All channels  $X_i \rightarrow Y_i$  are binary-input symmetric-output, i.e. T-consistent, but not necessarily have the same distribution. The channels are independent, i.e.  $p(\underline{y}|\underline{x}) = \prod_{i=1}^d p(y_i|x_i)$ . Without loss of generality, the extremal problem of information combining at a check node can be stated as follows.

*Problem 1:* Fix the channels  $X_i \rightarrow Y_i$ ,  $1 \leq i \leq d - 2$ . Find T-consistent probability densities  $P_b$  and  $P_w$  for  $X_{d-1} \rightarrow Y_{d-1}$  that maximize and minimize the extrinsic mutual information  $I(X_d; T_{E,d})$  respectively, subject to  $I(X_{d-1}; T_{d-1}) = I_{d-1}$ .

Let  $p(t) \triangleq p(T_{d-1} = t | X_{d-1} = 1)$  and  $p(t)$  should be T-consistent, i.e.  $p(t) = p(-t)(1+t)/(1-t)$ . In the constraint  $I_{d-1} = I(X_{d-1}; T_{d-1})$ , where the right-hand-side is related to  $p(t)$  by (2), one can see the integrand  $\log(1+t)$  is not continuous on  $[-1, 1]$ . One knows T-system theory requires all the integrand functions to be continuous on a closed interval. Thus we use the T-consistency of  $p(t)$  to transform the integrand into a continuous function. First,  $p(t)$  is mapped into a new density  $\hat{p}(t)$  on  $[0, 1]$

$$\hat{p}(t) = \begin{cases} p(t) + p(-t) = p(t)2/(1+t), & t \in (0, 1] \\ p(0), & t = 0. \end{cases} \quad (10)$$

Then one obtains

$$\begin{aligned} I_{d-1} &= I(X_{d-1}; T_{d-1}) = \int_0^{+1} (1 - h[(1-t)/2]) \hat{p}(t) dt \\ &= \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} m_{2i} \end{aligned} \quad (11)$$

where

$$m_{2i} = \int_0^{+1} t^{2i} \hat{p}(t) dt.$$

Note that  $1 - h[(1-t)/2]$  is a continuous function on  $[0, 1]$ . Another merit of (10) is that the T-consistency requirement on  $p(t)$  is automatically taken into account in (11).

Define  $\beta_{2i} = \prod_{k=1}^{d-2} \mathbb{E}[T_k^{2i} | X_k = 1]$ , from (6), one obtains

$$I_{E,d} = I(X_d; T_{E,d}) = \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} \beta_{2i} m_{2i}. \quad (12)$$

Fix  $\beta_{2i}$  for  $i \geq 1$ . The extremal information combining problem at the check node can be reformulated as: subject to (11), determine the best or worst densities of  $\hat{p}$  such that  $I_{E,d}$  (12) is maximized or minimized respectively.

In what follows we will show that the extreme densities for this optimization problem are exactly the densities corresponding to the extreme distributions of the following one.

*Problem 2:* Among all probability distributions on  $[0, 1]$  which satisfy the constraint (11), determine the probability distribution  $\sigma_b$  ( $\sigma_w$ ) which maximizes (minimizes) the 2nd moment  $m_2$ .

T-system theory is a perfect tool to solve it. Let

$$\begin{aligned} u_0(t) &\triangleq 1, & u_1(t) &\triangleq 1 - h[(1-t)/2], \\ \Omega(t) &\triangleq -t^2 \end{aligned}$$

and  $\underline{c} \triangleq (1, I_{d-1})$ . We prove that

*Lemma 1:* Both  $\{u_0, u_1\}$  and the augmented system  $\{u_0, u_1, \Omega\}$  are T-systems on  $[0, 1]$ .

$V(\underline{c}) = \{\sigma | \int_0^1 u_i(t) d\sigma(t) = c_i, i = 0, 1\}$  is the set of distributions which are probability distributions and satisfy (11). By Theorem 3 one concludes that the distribution  $\sigma^*$  associated with the upper principle representation of  $\underline{c}$  maximizes  $-m_2$ , thus  $\sigma_w = \sigma^*$ . The distribution  $\sigma_*$  associated with the lower principle representation of  $\underline{c}$  minimizes  $-m_2$ , thus  $\sigma_b = \sigma_*$ .

According to Theorem 2, since  $n = 1$  is odd,  $q = 1$ , all probability mass of  $\sigma_b$  concentrates at an interior point  $t_1$ , and all probability mass of  $\sigma_w$  concentrates at two endpoints  $\{0, 1\}$ . By solving (11) for  $\sigma_b$ , i.e.  $u_1(t_1) = I_{d-1}$ , one can obtain  $t_1 = 1 - 2h^{-1}[1 - I_{d-1}]$ . According to (10), the T-consistent distribution corresponding to  $\sigma_b$  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 channel with capacity  $I_{d-1}$ . Similarly, for  $\sigma_w$ , one can easily obtain that the probability mass at 1 is  $I_{d-1}$ . The T-consistent distribution corresponding to  $\sigma_w$  has probability mass of  $I_{d-1}$  at 1, and probability mass of  $1 - I_{d-1}$  at 0. This exactly corresponds to a BEC channel with capacity  $I_{d-1}$ . Thus we conclude that

*Theorem 4:* Among binary-input symmetric-output channels with a fixed mutual information, BSC (BEC) maximizes (minimizes)  $m_2$ .

*Lemma 2:* A moment sequence  $\{m_{2i}\}_{i=1}^\infty$  has the following properties:

- 1)  $\{m_{2i}\}_{i=1}^\infty$  is nonincreasing;
- 2) The ratio sequence  $\{r_{2i} = \frac{m_{2i}}{m_{2i+2}}\}_{i=1}^\infty$  is nonincreasing.

For BSC, its moment sequence  $\{m_{b,2i} = t_1^{2i}\}_{i=1}^\infty$  is a geometric sequence. For BEC, its moment sequence  $\{m_{w,2i} = I_{d-1}\}_{i=1}^\infty$  is a constant sequence. Using these facts, Theorem 4, Lemma 2 and the constraint (11), we show that

*Lemma 3:* There are two possible types of ordering for the moment sequence  $\{m_{b,2i}\}_{i=1}^\infty$  of BSC and the moment sequence  $\{m_{\sigma,2i}\}_{i=1}^\infty$  for any non-BSC  $\sigma \in V(\underline{c})$

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

where  $i_0$  is an integer depending on  $\sigma$ .

Similarly, we can show

*Lemma 4:* There are two possible types of ordering for the moment sequence  $\{m_{w,2i}\}_{i=1}^\infty$  of BEC and the moment sequence  $\{m_{\sigma,2i}\}_{i=1}^\infty$  for any non-BEC  $\sigma \in V(\underline{c})$

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

where  $i_1$  is an integer depending on  $\sigma$ .

The following lemma is useful. We prove that

*Lemma 5:* Assume  $a_i > 0$  for  $i \geq 1$ ,  $1 \geq b_1 \geq b_2 \geq \dots \geq b_i \geq \dots \geq 0$ . For sequences  $\{x_i\}_{i=1}^\infty, \{y_i\}_{i=1}^\infty$  which 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  $i > k$ ,

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

Using Lemma 3, Lemma 4, Lemma 5, (11) and (12), we conclude that subject to the mutual information constraint (11), BSC maximizes and BEC minimizes  $I_{E,d}$  (12), i.e.  $P_b = \text{BSC}$  and  $P_w = \text{BEC}$ . As pointed out before, they are also the extreme distributions with respect to optimizing  $m_2$ .

For the case [7] where the distribution of each channel is allowed to vary subject to a mutual information constraint, one can easily conclude that when all of them are BSCs (BECs), the extrinsic mutual information is maximized (minimized).

For the extremal information combining problem at the variable nodes, by Lemma 3 [11, Sec.IV], one can conclude that BEC and BSC are the maximizer and minimizer respectively.

## V. EXTENSION OF THE ORIGINAL EXTREMAL INFORMATION COMBINING PROBLEM

We consider an extension of the original extremal information combining problem by adding a constraint on the 2nd moment  $m_2$ .

*Problem 3:* Fix the channels  $X_i \rightarrow Y_i, 1 \leq i \leq d-2$ . Find T-consistent probability densities  $\tilde{P}_b$  and  $\tilde{P}_w$  for  $X_{d-1} \rightarrow Y_{d-1}$  that maximize and minimize the extrinsic mutual information  $I(X_d; T_{E,d})$  respectively, subject to constraints:

- $I(X_{d-1}; T_{d-1}) = I_{d-1}$ ;
- $m_2 = \theta$ .

The motivation for considering this problem is to obtain a smaller gap between maximized and minimized  $I_{E,d}$  compared to Problem 1.

First we again consider a different optimization problem. Later we will show that the solution to the following problem leads to the solution of Problem 3 by taking an inverse mapping of (10).

*Problem 4:* Among all probability distributions on  $[0, 1]$  which satisfy the constraint (11) and  $m_2 = \theta$ , determine the probability distribution  $\tilde{\sigma}_b$  ( $\tilde{\sigma}_w$ ) which maximizes (minimizes) the 4th moment  $m_4$ .

Again, we use T-system theory. Define

$$\begin{aligned} u_0(t) &\triangleq 1, & u_1(t) &\triangleq t^2, \\ u_2(t) &\triangleq 1 - h[(1-t)/2], & \Omega(t) &\triangleq -t^4 \end{aligned}$$

and  $\underline{c} \triangleq (1, \theta, I_{d-1})$ . We prove that

*Lemma 6:* Both  $\{u_0, u_1, u_2\}$  and the augmented system  $\{u_0, u_1, u_2, \Omega\}$  are T-systems on  $[0, 1]$ .

$V(\underline{c}) = \{\sigma | \int_0^1 u_i(t) d\sigma(t) = c_i, i = 0, 1, 2\}$  is a set of distributions which are probability distributions and satisfy both (11) and  $m_2 = \theta$ . By Theorem 3 one concludes that the distribution  $\sigma^*$  associated with the upper principle representation of  $\underline{c}$  maximizes  $-m_4$ , thus  $\tilde{\sigma}_w = \sigma^*$ . The distribution  $\sigma_*$  associated with the lower principle representation of  $\underline{c}$  minimizes  $-m_4$ , thus  $\tilde{\sigma}_b = \sigma_*$ .

According to Theorem 2, 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  $0 \leq s_1 \leq t_1$ . The T-consistent density  $P_{4,b}$  corresponding to  $\tilde{\sigma}_b$  has probability mass of  $p_0$  at 0, probability mass of  $p_1$  at  $t_1$  and probability mass of  $1 - p_0 - p_1$  at  $-t_1$ . From the constraints and the distribution structure, we can determine that

$$t_1 = f^{-1}\left(\frac{I_{d-1}}{\theta}\right), \quad \text{where } f(x) = \frac{1 - h[(1-x)/2]}{x^2},$$

$$p_1 = \frac{\theta \cdot (1 + t_1)}{2t_1^2}, \quad p_0 = 1 - \frac{2p_1}{1 + t_1}.$$

Similarly, the T-consistent density  $P_{4,w}$  corresponding to  $\tilde{\sigma}_w$  has probability mass of  $p_1$  at  $s_1$ , probability mass of  $p_1 \frac{1-s_1}{1+s_1}$  at  $-s_1$  and probability mass of  $p_2$  at 1. From the constraints and the distribution structure, we can determine that

$$s_1 = g^{-1}\left(\frac{1 - I_{d-1}}{1 - \theta}\right), \quad \text{where } g(x) = \frac{h[(1-x)/2]}{1 - x^2},$$

$$p_1 = \frac{1 - \theta}{2(1 - s_1)}, \quad p_2 = \frac{\theta - s_1^2}{1 - s_1^2}.$$

In summary, we have

*Lemma 7:* Among binary-input symmetric-output channels with a fixed mutual information and fixed  $m_2$ , the channel corresponding to  $P_{4,b}$  ( $P_{4,w}$ ) maximizes (minimizes)  $m_4$ .

Observe that the moment sequence  $\{\tilde{m}_{b,2i} = \theta t_1^{2i-2}\}_{i=1}^{\infty}$  of the best density  $P_{4,b}$  is a geometric sequence. Along with Lemma 2, Lemma 7 and (11), we show that

*Lemma 8:* 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  $\{m_{\sigma,2i}\}_{i=1}^{\infty}$  of any non- $P_{4,b}$   $\sigma \in V(\underline{c})$

- 1)  $\tilde{m}_{b,2i} > m_{\sigma,2i}$  for  $2 \leq i \leq i_0$ , and  $\tilde{m}_{b,2i} < m_{\sigma,2i}$  for  $i > i_0$ ;
- 2)  $\tilde{m}_{b,2i} > m_{\sigma,2i}$  for  $2 \leq i < i_0$ ,  $\tilde{m}_{b,2i_0} = m_{\sigma,2i_0}$  and  $\tilde{m}_{b,2i} < m_{\sigma,2i}$  for  $i > i_0$ ;

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

Although its proof is quite complex, we still can show

*Lemma 9:* 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  $\{m_{\sigma,2i}\}_{i=1}^{\infty}$  of any non- $P_{4,w}$   $\sigma \in V(\underline{c})$

- 1)  $m_{\sigma,2i} > \tilde{m}_{w,2i}$  for  $2 \leq i \leq i_1$ , and  $m_{\sigma,2i} < \tilde{m}_{w,2i}$  for  $i > i_1$ ;
- 2)  $m_{\sigma,2i} > \tilde{m}_{w,2i}$  for  $2 \leq i < i_1$ ,  $m_{\sigma,2i_1} = \tilde{m}_{w,2i_1}$  and  $m_{\sigma,2i} < \tilde{m}_{w,2i}$  for  $i > i_1$ ;

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

By Lemma 8, Lemma 9, Lemma 5, we can conclude that subject to the mutual information constraint (11) and  $m_2 = \theta$ , the distributions  $P_{4,b}$  and  $P_{4,w}$  maximize and minimize the extrinsic mutual information  $I_{E,d}$  respectively. Thus as pointed out before,  $\tilde{P}_b = P_{4,b}$  and  $\tilde{P}_w = P_{4,w}$ .

Extended to the case where the distribution of each channel is allowed to vary subject to (11) and  $m_2 = \theta$  constraints,

it is not difficult to see that when all of the channels are of the type of  $\tilde{P}_b$  (with appropriate parameters), the extrinsic mutual information is maximized. Similarly, when all of the channels are of the type of  $\tilde{P}_w$  (with appropriate parameters), the extrinsic mutual information is minimized.

In [16], we use the results of this section to bound the CND EXIT functions with Gaussian priors, and as part of a procedure which computes a (potentially) improved overall best and worst performance bounds on the mutual information trajectory of the belief propagation decoding of LDPC codes.

## VI. CONCLUSIONS

In this paper, we showed that BSC (BEC) maximizes (minimizes)  $m_2$  among all binary-input symmetric-output channels with a fixed mutual information. We also proved some ordering properties of moment sequences. We solved both the original and extended extremal problems of information combining. Ongoing and future work includes finding more applications for the extension problem, considering more complicated codes and channels.

## REFERENCES

- [1] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.
- [2] 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.
- [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] 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.
- [5] 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.
- [6] 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.
- [7] —, "Bounds on information combining," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 612–619, Feb. 2005.
- [8] 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.
- [9] I. Land, S. Huettinger, P. A. Hoeher, and J. Huber, "Bounds on mutual information for simple codes using information combining," *Annals of Telecommun.*, Jan. 2005, accepted for publication.
- [10] I. Sutskever, S. Shamai, and J. Ziv, "Extremes of information combining," in *Proceedings of the 41st Annual Allerton Conference on Communications, Control and Computing*, Monticello, IL, Oct. 2003, pp. 1446–1455.
- [11] —, "Extremes of information combining," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1313–1326, Apr. 2005.
- [12] 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.
- [13] —, "EXIT functions for continuous channels: Part I. Constituent codes," *IEEE Trans. Commun.*, submitted for publication, 2004.
- [14] M. G. Krein and A. A. Nudel'man, *The Markov Moment Problem and Extremal Problems*. Providence, Rhode Island: American Mathematical Society, 1977.
- [15] S. Karlin and W. J. Studden, *Tchebycheff Systems: with Applications in Analysis and Statistics*. New York: Interscience Publishers, 1966.
- [16] Y. Jiang, A. Ashikhmin, R. Koetter, and A. C. Singer, "Extremal problems of information combining," 2005, in preparation.