Data and Error Rate Bounds for Binary Data Gathering Wireless Sensor Networks

Xin He*†, Xiaobo Zhou†, Markku Juntti† and Tad Matsumoto*†
*Japan Advanced Institute of Science and Technology (JAIST)
1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan
Email:{hexin, matumoto}@jaist.ac.jp
†Centre for Wireless Communications, FI-90014 University of Oulu, Finland
Email: {firstname.lastname}@ee.oulu.fi

***Abstract*—The goal of this paper is to provide analytical assessment that justifies the performance tendencies of practical encoding/decoding techniques for a binary data gathering wireless sensor network (WSN). The theoretical rate region of the WSN is approximately analyzed based on the Slepian-Wolf theorem. We also derive the bit error rate (BER) floor given the observation error probability by using the Poisson-binomial distribution approximation. The simulation results show that the BER performance of our proposed technique is close to theoretical limits supported by the Slepian-Wolf theorem. The extrinsic information transfer chart analysis is also used to verify the BER performance. Moreover, the simulation results show that the bit error probability floor is matched with the Poisson-binomial approximation.**

I. Introduction

Wireless sensor networks (WSNs) have attracted significant attention by the sensing and wireless communication research communities, and their practical deployment is also gaining momentum. The applications of WSNs are numerous, and they can be deployed in unpredictable environments to perform various distributed sensing tasks, for which it is extraordinarily important to design good transmission and scheduling techniques in order to make the sensor network highly energy efficient. The information gathered by different sensors is often correlated. Therefore, distributed source coding (DSC) techniques inspired, e.g., by the Slepian-Wolf theorem [1], can provide an effective framework to optimize the network. This can be used to decrease the transmission rate or to reduce the transmit power. A tutorial paper by Xiong *et al.* [2] discussed the DSC schemes for WSNs in the framework of the Slepian-Wolf theorem.

We consider a binary data gathering WSN, where multiple sensors observe a single binary source to produce erroneous observations. A fusion center (FC) reconstructs the binary source using multiple versions of observations. In the network information theory, the problem of estimating a single source by deploying multiple unreliable sensors (agents) over rate limited noiseless links is called the chief executive officer (CEO) problem [3], the goal of which is to derive the rate-distortion function for a temporally memoryless source.

Numerous encoding/decoding algorithms have been proposed for the binary data gathering WSN. In [4], authors proposed an encoding scheme using parallel concatenated convolutional codes and a joint decoding method using weighted extrinsic log-likelihood ratio (LLR) to utilize the correlation knowledge. An adaptive bi-modal decoder for the binary data gathering WSN having two sensors was proposed in [5]. In [6], the convergence properties of the iterative decoding process were evaluated. We studied the binary data gathering WSN in [7] and [8]. A very simple encoding and joint decoding algorithm using LLR updating function [9] was proposed, the bit error rate (BER) performance of which is significantly improved [7]. We further proposed an algorithm for estimating the correlation parameters in [8].

In additive white Gaussian noise (AWGN) channels, the BER curve of the technique presented in [7] and [8] exhibits a sharp turbo cliff at a certain threshold signal-to-noise ratio (SNR) like the standard turbo codes do. However, the BER curve exhibits an error floor. It is clearly understood that the floor is due to the fact that the observations suffer from errors before being channel-encoded for transmission. Then, there arises a question: how the values of the threshold SNR and the error floor should/could be justified, and what the reference values are to be compared with?

The major contribution of this work is threefold:

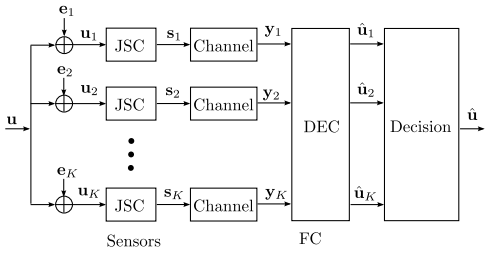

Fig. 1. An abstract model of a binary CEO problem.

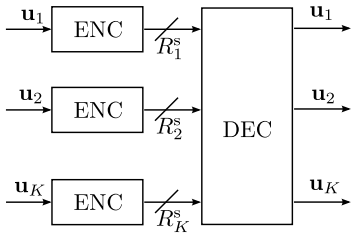

Fig. 2. The source coding model of the binary data gathering WSN.

(1) the theoretical threshold SNR of the data gathering WSN is derived based on the Slepian-Wolf theorem; (2) the error floor of the BER curve is analytically calculated from the probabilities of several events in a Poisson binomial process; (3) the threshold SNR in the BER curve is verified through the extrinsic information transfer (EXIT) chart analysis.

II. Theoretical Rate Region Analysis

In this section, the theoretical rate region of the binary data gathering sensor network is approximated by the Slepian-Wolf source coding, where account is taken of the observation error probability.

A. Problem Definition

The abstract model of the binary data gathering WSN is depicted in Fig. 1. An independently, identically distributed random data sequence $\mathbf{u}$ taking values from a set $\{0, 1\}$ with equal probability is observed by K sensors. Let $\mathbf{u}_k$, $k = 1, \cdots, K$, be the output random data sequence of a binary symmetric channel (BSC) with associated crossover probability p_k and $\mathbf{u}$ being the input. Each sensor independently encodes its observation $\mathbf{u}_k$ by a joint source channel (JSC) encoder and send the encoded version $\mathbf{s}_k$ over AWGN channel to a common decoder. It jointly reconstructs $\mathbf{u}_k$ with an arbitrarily small error probability. Finally, the estimate $\hat{\mathbf{u}}$ of $\mathbf{u}$ is obtained by using the majority logic decision.

In network information theory, the binary data gathering WSN is modeled by a binary CEO problem, the rate-distortion region of which is still not known. Therefore, we use the Slepian-Wolf theorem to evaluate the core part of the main problem, which is illustrated in Fig. 2.

B. Threshold for K Users

Slepian and Wolf [1] characterized the rate region for the lossless source coding of two correlated sources. It can be extended to the case having an arbitrary number of correlated sources [10]. The Slepian-Wolf theorem of this case is summarized in *Theorem 1*. Let $\Lambda = \{1, \cdots, K\}$ and $\mathcal{S} \subseteq \Lambda$.

Theorem 1 (Slepian-Wolf [10]): In order to achieve lossless transmission of multiple correlated sources, the source coding rate R_k^{s} should satisfy the following conditions

$$\sum_{k \in \mathcal{S}} R_k^{\mathrm{s}} \geq H(\mathbf{U}_{\mathcal{S}} | \mathbf{U}_{\mathcal{S}^{\mathrm{C}}}), \tag{1}$$

where $\mathcal{S}^{\mathrm{C}} = \Lambda \setminus \mathcal{S}$ represents the complementary set of $\mathcal{S}$ and $\mathbf{U}_{\mathcal{S}} = [\mathbf{u}_i; \mathbf{u}_j; \mathbf{u}_k]$ when $i, j, k \in \mathcal{S}$.

Since the channels between the sensors and the FC are assumed to be orthogonal with each other, source-channel separation is hold in this case. Therefore, according to the source-channel separation theorem [9], the threshold SNR is then calculated by

$$SNR_{\mathrm{lim}} = 10 \log_{10} \left(2^{\frac{\sum_k R_k^{\mathrm{s}}}{\sum_k 1/R_k^{\mathrm{c}}}} - 1 \right), \tag{2}$$

where R_k^{c} represents the rate of the channel code used at the k-th sensor node. It should be emphasized here that SNR_{lim} corresponds to the threshold SNR when the minimal distortion is achieved. The minimal distortion is equivalent to the error floor of the BER performance, which is analyzed in the next section. Now we need to calculate $R_{\mathrm{sum}}^{\mathrm{s}} = \sum_{k \in \Lambda} R_k^{\mathrm{s}}$.

Given the fact that u_k, $k = 1, \cdots, K$, is the result of passing u through a BSC with crossover probability p_k, where u_k and u represent the realizations of $\mathbf{u}_k$ and $\mathbf{u}$, respectively, the joint probability $\Pr(u_1, u_2, \cdots, u_K)$ is formulated as

$$\begin{aligned} &\Pr(u_1, u_2, \cdots, u_K) \\ &= \frac{1}{2} \prod_{i \in \mathcal{A}} (1 - p_i) \prod_{j \in \mathcal{A}^{\mathrm{C}}} p_j + \frac{1}{2} \prod_{i \in \mathcal{A}} p_i \prod_{j \in \mathcal{A}^{\mathrm{C}}} (1 - p_j), \end{aligned} \tag{3}$$

where $\mathcal{A} = \{k \in \Lambda | u_k = 0\}$ and $\mathcal{A}^{\mathrm{C}} = \Lambda \setminus \mathcal{A}$ is the complement set of $\mathcal{A}$. For example, setting $K = 3$ with $u_1 = 0$, $u_2 = 1$ and $u_3 = 0$, the set $\mathcal{A}$ is equal to $\{1, 3\}$ and $\mathcal{A}^{\mathrm{C}} = \{2\}$.

Therefore, the joint entropy $H(\mathbf{U})$ with $\mathbf{U} = [\mathbf{u}_1; \cdots; \mathbf{u}_K]$, which is equivalent to $R_{\text{sum}}^{\text{s}}$, is calculated as

$$H(\mathbf{U}) = - \sum_{u_k \in \{0,1\}} \Pr(u_1, u_2, \cdots, u_K) \log_2(\Pr(u_1, u_2, \cdots, u_K)). \tag{4}$$

III. BEP Floor Analysis

To analyze the error floor that appears in the BER performance curve of the encoding/decoding technique proposed in [7] and [8], we assume the channels between the sensors and the FC are noiseless since the BEP floor appears in the high SNR regime and it is only determined by the observation error probabilities.

A. Poisson-Binomial Approximation

Without loss of generality, we assume the source $\mathbf{u}$ is an all zero sequence with n bits. The majority decision rule to generate $\hat{u}(i)$ is

$$\hat{u}(i) = \begin{cases} 1, & \text{if } \mathbf{1}(\mathbf{U}^i) > \mathbf{0}(\mathbf{U}^i), \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where $\mathbf{1}(\mathbf{U}^i)$ and $\mathbf{0}(\mathbf{U}^i)$ represent the number of 1's and 0's in the i-th column of $\mathbf{U}$, respectively.

The decision error occurs when $\hat{u}(i)$ is decided to be 1. Hence, the error floor is analyzed by determining the probability of the number of 1's from K independent Bernoulli sequences having different probabilities of "1". In the statistics, the Poisson-binomial distribution is the probability distribution of the sum of independent Bernoulli trials that are not necessarily identically distributed. The probability of J-times occurrence of the error in K-times repeated binary trials with different crossover probabilities is [11]

$$\Pr(J = j) = \begin{cases} \prod_{k=1}^{K} (1 - p_k), & j = 0, \\ \frac{1}{j} \sum_{k=1}^{j} (-1)^{(k-1)} \Pr(J = j - k) L(k), & j > 0, \end{cases} \tag{6}$$

where $L(k) = \sum_{l=1}^{K} \left(\frac{p_l}{1 - p_l} \right)^k$ and $0 \leq j \leq K$.

Hence, the error floor with different observation noise probabilities is calculated by

$$\Pr(\hat{u}(i) \neq u(i)) = \begin{cases} \sum_{j=\frac{K+1}{2}}^{K} \Pr(J = j), & \text{if } K \text{ is odd}, \\ \frac{1}{2} \Pr(J = \frac{K}{2}) + \sum_{j=\frac{K}{2}+1}^{K} \Pr(J = j), & \text{if } K \text{ is even}. \end{cases} \tag{7}$$

B. Theoretical Lower Bound

The BEP floor analysis provided in the previous subsection is based only on the generalized majority logic (Poisson-binomial) analysis. However, it does not take into account the impact of soft-combining of the LLRs. The exact BEP floor analysis taking soft-combining into account is left as a future study. Instead, we further provide a theoretical lower bound of the BEP floor p_{lb}.

According to the rate-distortion theory for binary source [12], the theoretical lower bound of the BEP floor is given by [13]

$$p_{lb} = H_b^{-1}[1 + \sum_{k=1}^{K} H_b(p_k) - H(\mathbf{U})], \tag{8}$$

where $H_b(x) : [0, 0.5] \rightarrow [0, 1]$ and $H_b^{-1}(x) : [0, 1] \rightarrow [0, 0.5]$ represents the binary entropy function and its inverse function, respectively.

IV. Simulations for Verification

A series of simulations have been conducted to verify the results presented in this work. This section provides the results of the simulations.

A. BER Evaluation

The encoding/decoding algorithm [8] is used to obtain the BER performance. The decoding process is divided into two parts: 1)the *local iteration* (LI) decodes the error correcting coding chains applied at the sensor nodes; 2)the *global iteration* (GI) combines the extrinsic LLRs output from each LI and feedbacks the updated LLRs to the LI. The common parameters assumed in the simulations were set as follows

- Block length $n = 10000$ bits.
- Interleavers: random.
- Encoder: rate $R_k^{\text{c}} = \frac{1}{2}$ nonrecursive systematic convolutional code with generator polynomial $G = [03, 02]_8$, where $[\cdot]_8$ represents the argument is an octal number.
- Doping ratio of the ACC: 1.

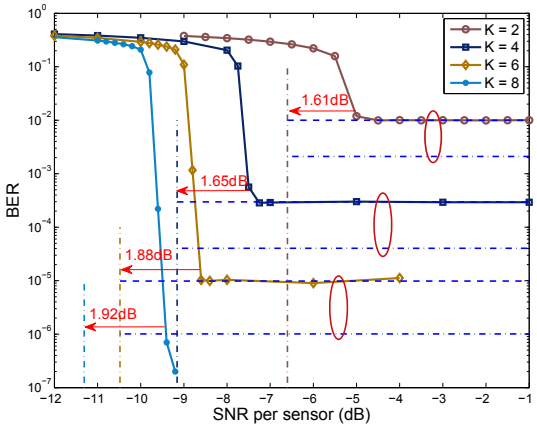

Fig. 3. BER performance of our proposed technique with the number of sensors $K = \{2, 4, 6, 8\}$. p_k are set at 0.01. The corresponding SNR_{lim} are plotted in vertical dash-dot lines. The approximated error floors and the lower bounds p_{lb} of the error floors are presented in horizontal dashed and dash-dot lines, respectively.

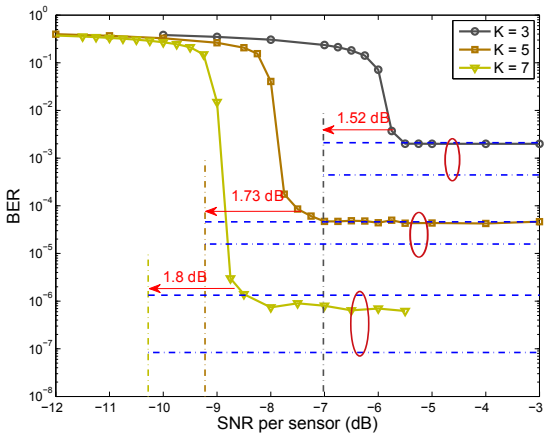

Fig. 4. BER performance comparison with the number of sensors $K = \{3, 5, 7\}$. The observation error probabilities are random distributed as $p_k = \{0.025, 0.075, 0.002\}$, $p_k = \{0.0145, 0.005, 0.025, 0.015, 0.03\}$ and $p_k = \{0.01, 0.015, 0.02, 0.05, 0.005, 0.0003, 0.02\}$.

- Decoding algorithm: log-maximum *a posteriori*.
- The number of iterations: 50 times[1].
- The SNR of each sensor node is assumed to be the same.

Fig. 3 illustrates the BER performance of the technique presented in [7] and [8] with the number of sensors $K = \{2, 4, 6, 8\}$. The results of the theoretical analysis presented in the previous sections are also shown. In the simulations, the error probabilities p_k were all set to 0.01. It can be clearly seen that in all the cases, the BER curves exhibit a sharp drop in the values at their corresponding threshold SNR values, like turbo-cliff of the turbo-codes. The difference between the threshold SNR obtained by the simulations and the theoretical analysis presented in Section II-B are $\{1.61, 1.65, 1.88, 1.92\}$ dB for $K = \{2, 4, 6, 8\}$. Furthermore, such sharp a decrease in BER suddenly plateaus at a value, however, the appearance of the error floor is different from that with the turbo codes; the floor is flat with the techniques shown in [7] and [8], while that with the turbo codes it still has a decay which is due to the property of the consistent codes.

Fig. 4 shows the BER performance with $K = \{3, 5, 7\}$, where the observation error probability p_k of each sensor was randomly set as those values shown in the caption of Fig. 4. Even in these cases, the BER performance gap to the theoretical limits are about $\{1.52, 1.73, 1.8\}$ dB.

It is also found that the BEP floors shown in Figs. 3 and 4 are placed between the results of the Poisson-binomial approximation and the rate-distortion lower bound. From the simulation results, it is clearly found that, the BEP floors are well matched with the values obtained in Poisson-binomial approximation.

B. Verification by EXIT Analysis

The convergence behavior of the k-th LI is evaluated by the EXIT chart analysis [14], [15]. The three dimensional (3D) EXIT chart analysis is used to verify the convergence of LI, taking into the account of the helper information of GI.

For obtaining the 3D EXIT chart, the *a priori* input LLR fed back to ACC^{-1} was first artificially generated for different values of $I^{\text{a}}_{\text{ACC}^{-1}}$, $0 \le I^{\text{a}}_{\text{ACC}^{-1}} \le 1$. The mutual information $I^{\text{e}}_{\text{ACC}^{-1}}$ was then calculated by evaluating the histogram of the output extrinsic LLR output from ACC^{-1}. Furthermore, the EXIT function for D_k is determined by evaluating the histogram of LLR $\mathbf{L}^{\text{e}}_{\mathbf{c}_k}$ in the form of

$$I^{\text{e}}_{\mathbf{c}_k} = T_k(\mathbf{L}^{\text{a}}_{\mathbf{c}_k}, \mathbf{L}^{\text{a}}_{\mathbf{u}_k}), \tag{9}$$

where $I^{\text{e}}_{\mathbf{c}_k}$ is the mutual information between the extrinsic LLR $\mathbf{L}^{\text{e}}_{\mathbf{c}_k}$ and the coded bits $\mathbf{c}_k$, indicating that $I^{\text{e}}_{\mathbf{c}_k} = I(\mathbf{L}^{\text{e}}_{\mathbf{c}_k}; \mathbf{c}_k)$. Since $I^{\text{e}}_{\mathbf{c}_k}$ is the output of a function with two inputs, $\mathbf{L}^{\text{a}}_{\mathbf{c}_k}$ and $\mathbf{L}^{\text{a}}_{\mathbf{u}_k}$, they were artificially generated for different values of $I^{\text{a}}_{\mathbf{c}_k}$, $0 \le I^{\text{a}}_{\mathbf{c}_k} \le 1$, and $I^{\text{a}}_{\mathbf{u}_k} = I(\mathbf{L}^{\text{a}}_{\mathbf{u}_k}; \mathbf{u}_k)$, $0 \le I^{\text{a}}_{\mathbf{u}_k} \le I(\mathbf{U}_{\Lambda \setminus k}; \mathbf{u}_k)$, respectively.

[1]In [7], [8], we just set the number of iterations empirically, however, in this paper, we set this number according to the EXIT analysis, which results in a better convergence performance.

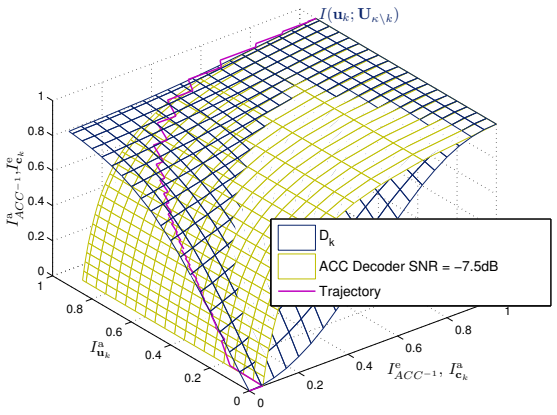

Fig. 5. 3D EXIT curves of ACC^{-1} and D_k at $SNR = -7.5$ dB and $K = 4$.

The mutual information $I^{\mathrm{e}}_{\mathbf{c}_k}$ was then calculated by evaluating the histogram of $\mathbf{L}^{\mathrm{e}}_{\mathbf{c}_k}$ output from D_k. Finally, the 3D EXIT chart is plotted based on the values of the mutual information.

The 3D EXIT chart shown in Fig. 5 was obtained by setting $p_k = 0.01$, $k = 1, \cdots, 4$, $SNR = -7.5$ dB. From Fig. 5, it is found that the effectiveness of GI is significant. The two EXIT planes are stuck at the initial stage, however, the tunnel between two EXIT planes opens as $I^{\mathrm{a}}_{\mathbf{u}_k}$ becomes large. The 3D trajectory obtained by evaluating $I^{\mathrm{a}}_{\mathbf{u}_k}$, $I^{\mathrm{e}}_{\mathrm{ACC}^{-1}}$ and $I^{\mathrm{e}}_{\mathbf{c}_k}$ in the real simulations is also presented in Fig. 5. As we can see, the trajectory asymptotically reaches a point very close to $(1.0, I(\mathbf{u}_k; \mathbf{U}_{\Lambda\setminus k}), 1.0)$ mutual information point, indicating that the information $\mathbf{u}_k$ transmitted from the k-th sensor node can be recovered with an arbitrary low error probability.

V. Conclusion

We investigated the transmission techniques for the binary data gathering sensor network, where multiple sensors observe a common binary source and transmit their erroneous observations to the fusion center. The theoretical limit was analyzed based on the frame work of the Slepian-Wolf theorem in lossless distributed source coding problem. Numerical results shown that the SNR values where turbo cliffs happened in the BER curves are only around $1.5 \sim 2$ dB to the corresponding threshold SNRs. Furthermore, the EXIT chart analysis clearly verified the threshold SNR values in the BER performance. The BEP floor, which is common to the CEO problem, was also approximated by evaluating the Poisson-binomial distribution. The simulation results were very close to the approximated BEP floor. In addition, the lower bound of the BEP floor was derived by the rate-distortion function of binary sources.

Acknowledgements

This work has been in part performed in the framework of the FP7 project ICT-619555 RESCUE (Links-on-the-fly Technology for Robust, Efficient and Smart Communication in Unpredictable Environments) which is partly funded by the European Union, and in part by Academy of Finland. The authors wish to acknowledge the helpful discussions with Dr. Reza Asvadi.

References

[1] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[2] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Mag.*, vol. 21, no. 5, pp. 80–94, Sept. 2004.

[3] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[4] A. Razi, K. Yasami, and A. Abedi, "On minimum number of wireless sensors required for reliable binary source estimation," in *Proc. IEEE Wireless Commun. and Networking Conf.*, Mexico, Mar. 2011, pp. 1852–1857.

[5] A. Razi and A. Abedi, "Adaptive bi-modal decoder for binary source estimation with two observers," in *Proc. Conf. Inform. Sciences Syst. (CISS)*, Princeton, NJ, USA, Mar. 2012, pp. 1–5.

[6] ——, "Convergence analysis of iterative decoding for binary ceo problem," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2944–2954, May 2014.

[7] X. Zhou, X. He, K. Anwar, and T. Matsumoto, "GREAT-CEO: larGe scale distRibuted dEcision mAking Technique for wireless Chief Executive Officer problems," *IEICE Trans. Commun.*, vol. E95-B, no. 12, pp. 3654–3662, Dec. 2012.

[8] X. He, X. Zhou, K. Anwar, and T. Matsumoto, "Estimation of Observation Error Probability in Wireless Sensor Networks," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1073–1076, Jun. 2013.

[9] J. Garcia-Frias and Y. Zhao, "Near-Shannon/Slepian-Wolf performance for unknown correlated sources over AWGN channels," *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 555–559, Apr. 2005.

[10] A. Gamal and Y. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[11] Y. H. Wang, "On the number of successes in independent trials," *Statistica Sinica*, vol. 3, no. 2, pp. 295–312, 1993.

[12] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. USA: John Wiley & Sons, Inc., 2006.

[13] X. He, X. Zhou, P. Komulainen, M. Juntti, and T. Matsumoto, "A lower bound on bit error probability for a binary CEO problem," *IEEE Trans. Commun.*, Mar. 2015, submitted for publication.

[14] J. Hagenauer, "The EXIT Chart - Introduction to Extrinsic Information Transfer," in *In Proc. 12th Europ. Signal Proc. Conf in Iterative Processing (EUSIPCO)*, 2004, pp. 1541–1548.

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