

Variable-Rate Universal Slepian-Wolf Coding with Feedback

Shriram Sarvotham, Dror Baron, and Richard G. Baraniuk
 Dept. of Electrical and Computer Engineering
 Rice University, Houston, TX 77005

Abstract— **Traditional Slepian-Wolf coding assumes known statistics and relies on asymptotically long sequences. However, in practice the statistics are unknown, and the input sequences are of finite length. In this finite regime, we must allow a non-zero probability of codeword error ϵ and also pay a penalty by adding redundant bits in the encoding process. In this paper, we develop a universal scheme for Slepian-Wolf coding that allows encoding at variable rates close to the Slepian-Wolf limit. We illustrate our scheme in a setup where we encode a uniform Bernoulli source sequence and the second sequence, which is correlated to the first via a binary symmetric correlation channel, is available as side information at the decoder. This specific setup is easily extended to more general settings. For length n source sequences and a fixed ϵ , we show that the redundancy of our scheme is $O(\sqrt{n}\Phi^{-1}(\epsilon))$ bits over the Slepian-Wolf limit. The prior art for Slepian-Wolf coding with known statistics shows that the redundancy is $\Omega(\sqrt{n}\Phi^{-1}(\epsilon))$. Therefore, we infer that for Slepian-Wolf coding, the penalty needed to accommodate universality is $\Theta(\sqrt{n}\Phi^{-1}(\epsilon))$.**

I. INTRODUCTION

The classical Slepian-Wolf theorem assumes known source statistics and applies to asymptotically long sequences [1, 2]. In practice, however, the input sequences are of a finite length n . Hence no matter what fixed rate we choose to encode, the coded sequences may be jointly atypical. Since the decoding process relies on joint typicality, we may be incapable of reconstructing the source sequences at the decoder. Therefore, in this *finite-length regime* we must allow a non-zero probability of codeword error ϵ . Another problem that arises in practice is that the source statistics are unknown, and we must use coding methods whose performance is *universal* with respect to a broad range of possible statistics. Of course these practical concerns impose a penalty in the coding rate in terms of redundant bits to describe the source.

A. Overview of the paper

In this work, we propose a variable-rate scheme for universal Slepian-Wolf coding. Our goal is not only to

This research was supported by NSF, AFOSR, ONR, and the Texas Instruments Leadership University Program.
 Email: {shri, drorb, richb}@rice.edu; Web: dsp.rice.edu.

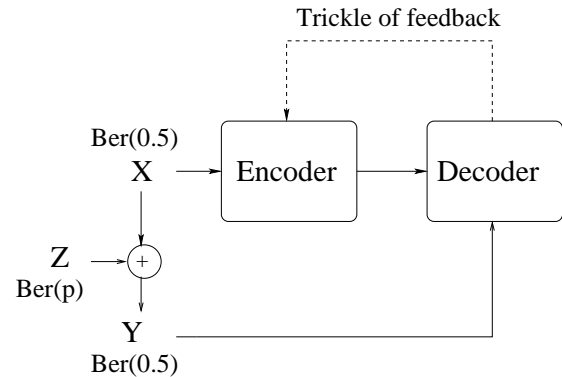


Fig. 1. Setup for universal Slepian-Wolf encoding with feedback.

achieve a small probability of codeword error, but also to encode at rates close to the Slepian-Wolf limit. Our scheme uses feedback from the decoder (Fig. 1) that enables the encoder to estimate the source statistics. The feedback allows the encoder to vary the rate of encoding based on the current estimate of the source statistics. In our scheme, the penalty needed is shown to be $O(\sqrt{n}\Phi^{-1}(\epsilon))$,¹ where n is the number of source bits, ϵ is the probability of codeword error, and $\Phi^{-1}(\cdot)$ is the inverse Gaussian error function. Since the penalty in the finite-length known statistics case is $\Omega(\sqrt{n}\Phi^{-1}(\epsilon))$ bits, we infer that the penalty for universality with feedback is $\Theta(\sqrt{n}\Phi^{-1}(\epsilon))$ bits.

B. Related work

We survey several related results that consider universality or non-asymptotic sources in Slepian-Wolf coding.

Fixed rate universal Slepian-Wolf setup: Csiszar and Korner [3] and Oohama and Han [4] consider a Slepian-Wolf setup where the coding rates are fixed, and the goal is to achieve universal error exponents similar to the known-

¹For two functions $f(n)$ and $g(n)$, $f(n) = O(g(n))$ if $\exists c, n_0 \in \mathbb{R}^+$, $0 \leq f(n) \leq cg(n)$ for all $n > n_0$. Similarly, $f(n) = o(g(n))$ if for all positive $c > 0$, $\exists n_0 \in \mathbb{R}^+$, $0 \leq f(n) < cg(n)$ for all $n > n_0$, and $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$ for all $n > n_0$. Finally, $f(n) = \Theta(g(n))$ if $\exists c_1, c_2, n_0 \in \mathbb{R}^+$, $0 \leq c_1g(n) \leq f(n) \leq c_2g(n)$ for all $n > n_0$.

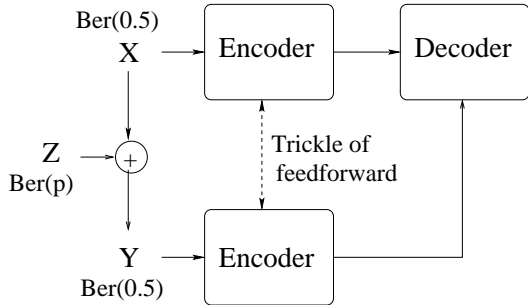


Fig. 2. Linked encoders for Slepian-Wolf encoding as considered by Kimura and Uyematsu [5].

statistics exponents. The limitation of this approach is that the fixed rates may greatly exceed the Slepian-Wolf limit.

Variable-rates via linked encoders: Kimura and Uyematsu proposed variable-rate Slepian-Wolf coding with *linked-encoders*, where a trickle of feed-forward between encoders is allowed (Fig. 2) [5]. In this setup, the penalty also includes communication between the encoders. Their result can be summarized in the following theorem:

Theorem 1: [5] For unknown statistics, a zero-rate penalty (redundancy over Slepian-Wolf limit) suffices for Slepian-Wolf coding using the linked encoder setup.

Lower bound via fixed rate coding (finite regime): Theorem 1 implies that for a fixed ϵ the penalty in the linked encoder setup is $o(n)$ bits [5]. The limitation here is that the $o(n)$ term may be quite large. Can we do better? One way to begin is with a lower bound by Wolfowitz on the penalty for using finite length sequences when the statistics are *known* [6].

Theorem 2: [6, 7] For known statistics and a fixed ϵ , the penalty for using finite length sequences (redundancy over Slepian-Wolf limit) is $\Theta(\sqrt{n}\Phi^{-1}(\epsilon))$ bits.

Wolfowitz proves Theorem 2 for Slepian-Wolf coding, channel coding and also discusses additional communication systems [6]. The theorem lower bounds the penalty needed to accommodate universality. This is because we cannot perform better for universality than the known statistics case. *However, Theorem 2 does not upper bound the penalty for universality.*

Upper bound via linked encoders: In previous work, we have shown that a penalty of $O(n^{2/3}\Phi^{-1}(\epsilon))$ bits suffice in the linked encoder setup in Slepian-Wolf coding [7]. Considering the lower bound of Theorem 2, the challenge is then to resolve the gap between $O(\sqrt{n}\Phi^{-1}(\epsilon))$ and $O(n^{2/3}\Phi^{-1}(\epsilon))$ bits.

C. Contributions

We propose a scheme that handles universality in a Slepian-Wolf setup. In our scheme we perform variable-

rate block sequential source coding on the input sequence. The input sequence is partitioned into blocks, and each block is encoded separately. The decoder transmits a trickle of feedback at the end of every block, which enables the encoder to estimate the statistics. The redundancy needed in our scheme is shown to be $O(\sqrt{n}\Phi^{-1}(\epsilon))$.

In our companion paper, we explained how our scheme can be applied to additional communication systems in which the decoder relies on joint typicality [8]. Specifically, we analyze our scheme for the channel coding setup, where we show that the penalty (in terms of gap to capacity) is $O(\sqrt{n}\Phi^{-1}(\epsilon))$, where n is the number of channel-coded bits. We also investigate the additional penalty incurred when we restrict the number of rounds of feedback, and show that for numbers of practical interest, 2–3 rounds of feedback suffice.

II. PROBLEM STATEMENT

Consider Fig. 1. Two binary sources X and Y each generate length- n Bernoulli sequences x and y with parameter 0.5. The correlation between the sequences x and y is captured by a binary symmetric channel (BSC) with parameter p . In other words, y can be obtained from x by adding a noise sequence z , which is Bernoulli with parameter p . In our setup, p is unknown to the encoder and the decoder. The decoder must reconstruct the sequences x and y with codeword error probability upper bounded by ϵ . The decoder is allowed to provide a small amount of (assumed error-free) feedback to the encoder to help the encoder estimate p . The goal is to design a strategy to reconstruct the sequences x and y at the decoder using minimal communication.

III. PROPOSED SOLUTION

Main idea: The encoder performs variable-rate block-sequential Slepian-Wolf coding on the sequence x . The decoder transmits a trickle of feedback (assumed error-free) to the encoder at the end of every block (Fig. 1). The feedback helps the encoder to refine the estimate of the correlation parameter p . The feedback encodes the number of bit positions in the last block where the bit values in the reconstructed signals differ. We then encode the next block using the improved estimate of p .

Notation: Let the number of rounds of feedback be k . This implies that the source sequences x and y are divided into $k + 1$ blocks for block-transmission. Let b_i be the length of the i^{th} block. Thus we have $n = \sum_{i=1}^{k+1} b_i$. Let $B_i = \sum_{j=1}^i b_j$ be the sum of the block lengths of the first i blocks. Let t be the total number of encoded bits (the sum of the block lengths of the *encoded* blocks). The probability of codeword error ϵ is allocated between the different

blocks. The block error probabilities, denoted by ϵ_i , satisfy the following constraint: $\sum_{i=1}^{k+1} \epsilon_i = \epsilon$.² Let s be the feedback costs involved. The number of bits required for the feedback in round i is $\log_2(b_i)$; this is because the feedback encodes the number of bit positions in block i in which the values differ in the source sequences. Hence $s = \sum_{i=1}^k \log_2(b_i)$. Finally, we define the penalty r as the number of redundant bits in the encoding process beyond the Slepian-Wolf limit:

$$r \triangleq t - nH(p) + s.$$

In our companion paper, we analyze the penalty due to universality in channel coding [8]. In that setting, if p is close to 0.5, then the channel capacity C of the BSC becomes very small. We avoid that by assuming $C > \eta$ for some $\eta > 0$. We do not require an analogous assumption for our Slepian-Wolf setup, because in the worst case when $p = 0.5$ we can send the source X uncompressed.

Partitioning into blocks: We propose that the block lengths increase as powers of 2 except for the last block; that is, $b_i = 2^{i-1}$ for $1 \leq i \leq k$. The last block consists of the bits that remain to be transmitted. We assign the block error probabilities such that $\epsilon_{i+1} = 2\epsilon_i$ and $\sum_{i=1}^{k+1} \epsilon_i = \epsilon$. Therefore, $\epsilon_i = \epsilon b_i / (2^{k+1} - 1)$.

IV. ANALYSIS OF THE VARIABLE-RATE SCHEME WITH FEEDBACK

Theorem 3: For unknown statistics and a fixed ϵ , the penalty (redundant bits needed) for Slepian-Wolf coding in our variable-rate scheme with feedback is $O(\sqrt{n}\Phi^{-1}(\epsilon))$ bits.

Parts of the proof that are tedious but less insightful are not included; See [9] for the complete proof.

Proof of Theorem 3: We compute the penalty for block i as

$$r_i = c_1 \frac{b_i}{\sqrt{B_{i-1}}} \Phi^{-1}(\epsilon_i) + c_2 \sqrt{b_i} \Phi^{-1}(\epsilon_i) + o(\sqrt{b_i}). \quad (1)$$

The first term captures the penalty due to universality. Note the $\sqrt{B_{i-1}}$ term in the denominator; this term captures the fact that for later blocks we have a better estimate of p and hence the penalty rate (i.e., penalty per transmitted bit) due to universality is smaller. The penalty due to universality is also proportional to b_i because block i is encoded at a constant rate that is computed using the current estimate of p . The remaining terms in (1) account for the

²Using the union bound, summing the individual block error probabilities to ϵ is sufficient to bound the overall codeword error probability to ϵ .

penalty in non-asymptotic Slepian-Wolf coding [6]. These terms represent the $\Theta(\sqrt{n}\Phi^{-1}(\epsilon))$ penalty in Theorem 2. The $\Phi^{-1}(\epsilon_i)$ comes from the Central Limit Theorem [6, 10]; providing a redundancy of δ standard deviations over the Slepian-Wolf limit buys us $\Phi(\delta)$ probability of error.

For blocks 1 through k , $b_i = 2^{i-1}$ and so

$$r_i = c_1 \frac{2^{i-1}}{\sqrt{2^{i-1} - 1}} \Phi^{-1}(\epsilon_i) + c_2 \sqrt{2^{i-1}} \Phi^{-1}(\epsilon_i) + o(2^{\frac{i}{2}}).$$

The first and second terms (corresponding to penalties associated with universality and finite length coding) are of the same order and can be combined. Therefore,

$$r_i = c_3 2^{\frac{i}{2}} \Phi^{-1}(\epsilon_i) + o(2^{\frac{i}{2}}),$$

for $1 \leq i \leq k$. In block $k+1$, we can transmit a maximum of 2^k bits and so

$$r_{k+1} \leq c_3 2^{\frac{k+1}{2}} \Phi^{-1}(\epsilon_{k+1}) + o(2^{\frac{k+1}{2}}).$$

The total penalty r is the sum of the penalties incurred for each block, yielding

$$r \leq \left[\sum_{i=1}^{k+1} c_3 2^{\frac{i}{2}} \Phi^{-1}(\epsilon_i) \right] + \left[\sum_{i=1}^{k+1} o(2^{\frac{i}{2}}) \right] + s,$$

where the inequality appears because of block $k+1$. It can be shown that $\sum_{i=1}^{k+1} o(2^{\frac{i}{2}}) = o(2^{\frac{k}{2}})$ [9], and so

$$r \leq \left[\sum_{i=1}^{k+1} c_3 2^{\frac{i}{2}} \Phi^{-1}(\epsilon_i) \right] + o(2^{\frac{k}{2}}) + s. \quad (2)$$

It can be easily shown that the feedback cost satisfies $s = O(\log^2(n))$. Therefore, s can be absorbed by the $o(2^{\frac{k}{2}})$ term. To simplify the summation term in (2), recall that $\epsilon_{i+1} = 2\epsilon_i$. The following upper-bound on $\frac{\Phi^{-1}(\frac{\xi}{2})}{\Phi^{-1}(\xi)}$ can be applied.

Lemma 1: For any $\delta > 0$, we can find a $\xi' > 0$ such that for all $\xi < \xi'$

$$\frac{\Phi^{-1}(\frac{\xi}{2})}{\Phi^{-1}(\xi)} < 1 + \delta.$$

We give an outline of the proof in the Appendix; the rigorous proof appears in [9].

For our proposed scheme, we pick a δ such that $1 + \delta \in (1, \sqrt{2})$ and determine the corresponding ξ' for which $\frac{\Phi^{-1}(\frac{\xi}{2})}{\Phi^{-1}(\xi)} < (1 + \delta)$.³ Because Lemma 1 is applicable only when $\epsilon_i < \xi'$, we split the summation in (2) into two sums,

³For $\delta = 0.4$, the value of ξ' is about 0.3093. This indicates the typical values of ξ' that we can expect in practical settings.

such that Lemma 1 can be applied to the $\Phi^{-1}(\epsilon_i)$ terms in the first sum:

$$\begin{aligned} r &\leq \sum_{i=1}^m c_3 2^{\frac{i}{2}} \Phi^{-1}(\epsilon_i) \\ &+ \sum_{i=m+1}^{k+1} c_3 2^{\frac{i}{2}} \Phi^{-1}(\epsilon_i) + o(2^{\frac{k}{2}}). \end{aligned} \quad (3)$$

If $\epsilon_{k+1} \geq \xi'$, then the index m is chosen such that $\epsilon_m < \xi'$ and $\epsilon_{m+1} > \xi'$. Else if $\epsilon_{k+1} < \xi'$, then we choose $m = k+1$. Consequently, the second summation is empty. Because we choose $\epsilon_{i-1} = \epsilon_i/2$, we have $\epsilon_i < \xi'$ for $i \leq m$, and hence the result from Lemma 1 can be applied to all the terms in the first summation. Denote the first summation in (3) by S_1 . By change of variables $j \leftarrow m-i$, we have

$$\begin{aligned} S_1 &= \sum_{i=1}^m c_3 2^{\frac{i}{2}} \Phi^{-1}(\epsilon_i) \\ &= \sum_{j=0}^{m-1} c_3 2^{\frac{m-j}{2}} \Phi^{-1}(\epsilon_{m-j}). \end{aligned}$$

From Lemma 1, $\Phi^{-1}(\epsilon_{m-j}) = \Phi^{-1}(\frac{\epsilon_m}{2^j}) < \Phi^{-1}(\epsilon_m)(1 + \delta)^j$, hence S_1 can be bounded by

$$\begin{aligned} S_1 &< c_3 2^{\frac{m}{2}} \Phi^{-1}(\epsilon_m) \sum_{j=0}^{m-1} \left(\frac{1 + \delta}{\sqrt{2}} \right)^j \\ &< c_3 2^{\frac{m}{2}} \Phi^{-1}(\epsilon_m) O(1) \\ &\quad \text{(summing the geometric series)} \\ &= O(2^{\frac{k}{2}} \Phi^{-1}(\epsilon_m)) \\ &\quad \text{(since } m \leq k+1 \text{)}. \end{aligned}$$

Note the dependence on $\Phi^{-1}(\epsilon_m)$. If ϵ is large enough, then $\epsilon_m \in (\xi'/2, \xi')$ and $\Phi^{-1}(\epsilon_m)$ is $O(1)$. However, if ϵ is small, then $m = k+1$, in which case $\Phi^{-1}(\epsilon_m) = \Phi^{-1}(\epsilon_{k+1}) = O(\Phi^{-1}(\epsilon))$. The second summation in (3), which we denote by S_2 , can be simplified as

$$\begin{aligned} S_2 &= \sum_{i=m+1}^{k+1} c_3 2^{\frac{i}{2}} \Phi^{-1}(\epsilon_i) \\ &< c_3 \Phi^{-1}(\xi') \sum_{i=m+1}^{k+1} 2^{\frac{i}{2}} \\ &\quad \text{(since } \Phi^{-1}(\epsilon_i) < \Phi^{-1}(\xi') \text{)} \\ &= O(2^{\frac{k}{2}}) \\ &\quad \text{(since } \sum_{i=m+1}^{k+1} 2^{\frac{i}{2}} = O(2^{\frac{k}{2}}) \text{)}. \end{aligned}$$

Note that S_2 is often empty in practice because ξ' is large (see Footnote 3). Using the results for S_1 and S_2 in (3), we infer that $r = O(2^{\frac{k}{2}} \Phi^{-1}(\epsilon))$. Because $n \in [2^k, 2^{k+1})$, we conclude that $r = O(\sqrt{n} \Phi^{-1}(\epsilon))$. \square

Because $\Phi^{-1}(\epsilon)$ also appears in refined versions of Theorem 3 [6, 7], we have a tight order bound.

V. DISCUSSION AND CONCLUSIONS

We have shown that the penalty due to universality in Slepian-Wolf coding is $O(\sqrt{n} \Phi^{-1}(\epsilon))$ bits. The prior art for known statistics lower bounds the penalty by $\Omega(\sqrt{n} \Phi^{-1}(\epsilon))$; hence it is $\Theta(\sqrt{n} \Phi^{-1}(\epsilon))$ bits. We have shown how to achieve this limit by proposing a variable-rate coding scheme with feedback. Our scheme can be applied to compression of encrypted data [11–13] and distributed compression of sensor network data [14].

Our vision is to develop the science of non-asymptotic information theory. Towards this goal, the current work characterizes the order terms of the penalty for universality in non-asymptotic Slepian-Wolf coding. As future work, we can refine these results to determine the constants involved in the order terms. Consider channel coding as an example. It has been shown that powerful channel coding schemes such as LDPC codes can operate near capacity and have the same $\Theta(\sqrt{n})$ penalty term as the optimal coding scheme; only the constants in the order terms differ [15]. In a similar vein, the constants involved in the order terms for universality will allow us to benchmark the effectiveness of practical schemes. Further work can also be done to extend the results for more complex communication systems such as sources and channels with memory, and systems in which the feedback channel is noisy.

Acknowledgments: We thank K. Ramchandran and T. Weissman, who inspired our study of and approach to this problem. We also thank M. A. Khojastepour for informative and inspiring conversations.

APPENDIX

Proof sketch for Lemma 1: The inverse complementary error function $\Phi^{-1}(\cdot)$ satisfies the following inequality [9]:

$$\sqrt{\frac{1}{4} + \log \frac{1}{\xi}} - \frac{1}{2} < \Phi^{-1}(\xi) < \sqrt{\log \frac{1}{\xi}},$$

for all $\xi < \xi_0$, where ξ_0 is some positive constant. Therefore,

$$\frac{\Phi^{-1}(\frac{\xi}{2})}{\Phi^{-1}(\xi)} < \frac{\sqrt{\log \frac{2}{\xi}}}{\sqrt{\frac{1}{4} + \log \frac{1}{\xi}} - \frac{1}{2}} \quad (4)$$

for $\xi < \xi_0$. Let

$$p(\xi) = \frac{\sqrt{\log \frac{2}{\xi}}}{\sqrt{\frac{1}{4} + \log \frac{1}{\xi} - \frac{1}{2}}}.$$

The following properties of $p(\xi)$ can be easily shown [9]:

- The limit $\lim_{\xi \rightarrow 0} p(\xi) = 1$.
- The derivative $\frac{d}{d\xi} p(\xi) > 0$ for any $0 < \xi < 1$.

These observations reveal that the function $p(\xi)$ is a monotonic increasing function of ξ for $0 < \xi < 1$ with $p(0) = 1$. For any $\delta > 0$, we can assign $\xi' = p^{-1}(1 + \delta)$ and hence $p(\xi) < 1 + \delta$ for all $\xi < \xi'$. Using this in (4) proves Lemma 1. \square

REFERENCES

- [1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, New York, 1991.
- [3] I. Csiszar and J. Korner, "Towards a general theory of source networks," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 2, pp. 155–165, Mar. 1980.
- [4] Y. Oohama and T. S. Han, "Universal coding for the Slepian-Wolf data compression system and the strong converse theorem," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1908–1919, Nov. 1994.
- [5] A. Kimura and T. Uyematsu, "Weak variable-length Slepian-Wolf coding with linked encoders for mixed source," *IEEE Trans. Inform. Theory*, vol. 50, no. 1, pp. 183–193, Jan. 2004.
- [6] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. Springer-Verlag, Berlin; New York, 1978.
- [7] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, "Redundancy rates of Slepian-Wolf coding," *Proc. 42nd Annual Allerton Conf. on Comm., Control, and Computing*, Oct. 2004.
- [8] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Variable-rate coding with feedback for universal communication systems," *Proc. 43rd Annual Allerton Conf. on Comm., Control, and Computing*, Monticello, IL, Sept. 2005.
- [9] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Penalty for universality in communication systems that rely on joint typicality," http://www.ece.rice.edu/~shri/univ_techrep.pdf, *Technical report*, Nov. 2005.
- [10] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, "How quickly can we approach channel capacity?," *Proc. 38th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2004.
- [11] M. Johnson, P. Ishwar, V. M. Prabhakaran, D. Schonberg, and K. Ramchandran "On compressing encrypted data," *IEEE Trans. Signal Processing*, vol. 52, no. 10, pp. 2992–3006, Oct. 2004.
- [12] D. Schonberg, S. Draper, and K. Ramchandran "On blind compression of encrypted data approaching the source entropy rate," *Proc. European Signal Processing Conf.*, Antalya, Turkey, Sept. 2005.
- [13] D. Schonberg, S. C. Draper, and K. Ramchandran, "On blind compression of encrypted correlated data approaching the source entropy rate," *Proc. 43rd Annual Allerton Conf. on Comm., Control, and Computing*, Monticello, IL, Sept. 2005.
- [14] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Mag.*, vol. 21, pp. 80–94, Sept. 2004.
- [15] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, "Finite-length scaling for iteratively decoded LDPC ensembles," *Proc. 41st Annual Allerton Conf. on Comm., Control, and Computing*, Oct. 2003.