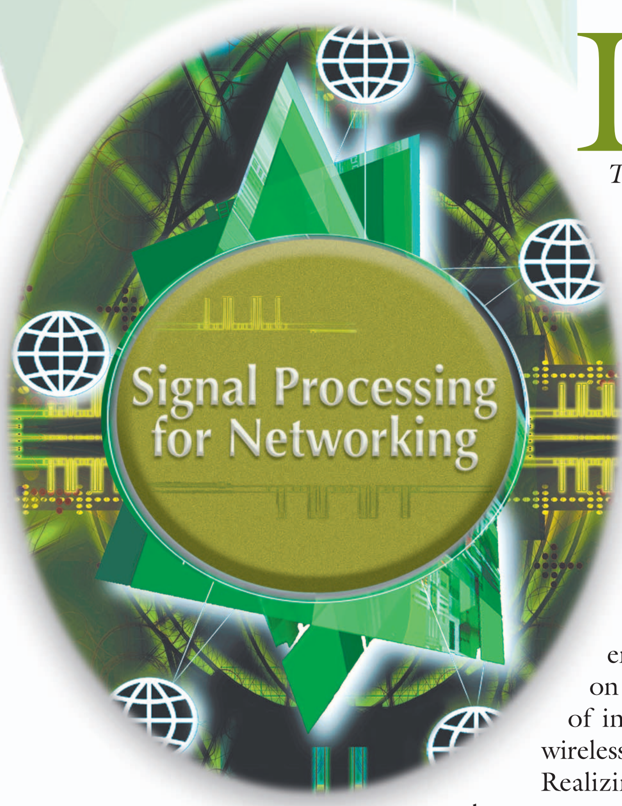
Compression via channel coding

Distributed Source Coding for Sensor Networks

Zixiang Xiong, Angelos D. Liveris, and Samuel Cheng



In recent years, sensor research has been undergoing a quiet revolution, promising to have a significant impact throughout society that could quite possibly dwarf previous milestones in the information revolution. *MIT Technology Review* ranked wireless sensor networks that consist of many tiny, low-power and cheap wireless sensors as the number one emerging technology. Unlike PCs or the Internet, which are designed to support all types of applications, sensor networks are usually mission driven and application specific (be it detection of biological agents and toxic chemicals; environmental measurement of temperature, pressure and vibration; or real-time area video surveillance). Thus they must operate under a set of unique constraints and requirements. For example, in contrast to many other wireless devices (e.g., cellular phones, PDAs, and laptops), in which energy can be recharged from time to time, the energy provisioned for a wireless sensor node is not expected to be renewed throughout its mission. The limited amount of energy available to wireless sensors has a significant impact on all aspects of a wireless sensor network, from the amount of information that the node can process, to the volume of wireless communication it can carry across large distances.

Realizing the great promise of sensor networks requires more than a mere advance in individual technologies; it relies on many components working together in an efficient, unattended, comprehensible, and trustworthy manner. One of the enabling technologies for sensor networks is *distributed source coding* (DSC), which refers to the compression of multiple correlated sensor outputs [1]–[4] that *do not* communicate with each other (hence distributed coding). These sensors send their compressed outputs to a central point [e.g., the base station (BS)] for *joint* decoding.

1053-5888/04/$20.00©2004IEEE

To motivate DSC, consider a wireless video sensor network consisting of clusters of low-cost video sensor nodes (VSNs), an aggregation node (AN) for each cluster, and a BS for surveillance applications. The lower tier VSNs are used for data acquisition and processing; the upper tier ANs are used for data fusion and transmitting information out of the network. This type of network is expected to operate unattended over an extended period of time. As such, VSN and AN power consumption cause severe system constraints; additionally, traditional, one-to-many, video processing that is routinely applied to sophisticated video encoders (e.g., MPEG compression) will not be suitable for use on a VSN. This is because under the traditional broadcast paradigm the video encoder is the computational workhorse of the video codec; consequently, computational complexity is dominated by the motion estimation operation. The decoder, on the other hand, is a relatively lightweight device operating in a "slave" mode to the encoder. The severe power constraints at VSNs thus bring about the following basic requirements: 1) an extremely low-power and low-complexity wireless video encoder, which is critical to prolonging the lifetime of a wireless video sensor node, and 2) a high ratio of compression efficiency, since bit rate directly impacts transmission power consumption at a node.

DSC allows a many-to-one video coding paradigm that effectively swaps encoder-decoder complexity with respect to conventional (one-to-many) video coding, thereby representing a fundamental conceptual shift in video processing. Under this paradigm, the encoder at each VSN is designed as simply and efficiently as possible, while the decoder at the BS is powerful enough to perform joint decoding. Furthermore, each VSN can operate independently of its neighbors; consequently, a receiver is *not* needed for video processing at a VSN, which enables the system to save a substantial amount of hardware cost and communication (i.e., receiver) power. In practice, depending also on the nature of the sensor network, the VSN might still need a receiver to take care of other operations, such as routing, control, and synchronization, but such a receiver will be significantly less sophisticated.

Under this new DSC paradigm, a challenging problem is to achieve the same efficiency (e.g., joint entropy of correlated sources) as traditional video coding, while not requiring sensors to communicate with each other. A moment of thought reveals that this might be possible because correlation exists among readings from closely placed neighboring sensors and the *decoder* can exploit such correlation with DSC; this is done at the *encoder* with traditional video coding. As an example, suppose we have two correlated 8-b gray-scale images X and Y whose same location pixel values $\mathbf{x}$ and $\mathbf{y}$ are related by $\mathbf{x} \in \{\mathbf{y}-3, \mathbf{y}-2, \mathbf{y}-1, \mathbf{y}, \mathbf{y}+1, \mathbf{y}+2, \mathbf{y}+3, \mathbf{y}+4\}$. In other words, the correlation of $\mathbf{x}$ and $\mathbf{y}$ is characterized by $-3 \leq \mathbf{x}-\mathbf{y} \leq 4$, or $\mathbf{x}$ assumes only eight different values around $\mathbf{y}$. Thus, joint coding of $\mathbf{x}$ would take 3 b. But in DSC, we simply take modulo of pixel value $\mathbf{x}$ with respect to eight, which also reduces the required bits to three. Specifically, let $\mathbf{x} = 121$ and $\mathbf{y} = 119$. Instead of transmitting both $\mathbf{x}$ and $\mathbf{y}$ at 8 b/p without loss, we transmit $\mathbf{y} = 119$ and $\mathbf{x}' = \mathbf{x} \pmod 8 = 1$ in distributed coding. Consequently, $\mathbf{x}'$ indexes the set that $\mathbf{x}$ belongs to, i.e., $\mathbf{x} \in \{1, 9, 17, \ldots, 249\}$, and the joint decoder picks the element $\mathbf{x} = 121$ closest to $\mathbf{y} = 119$.

The above is but one simple example showcasing the feasibility of DSC. Slepian and Wolf [1] theoretically showed that separate encoding (with increased complexity at the joint decoder) is as efficient as joint encoding for lossless compression. Similar results were obtained by Wyner and Ziv [2] with regard to lossy coding of joint Gaussian sources. Driven by applications like sensor networks, DSC has recently become a very active research area—more than 30 years after Slepian and Wolf laid its theoretical foundation [4].

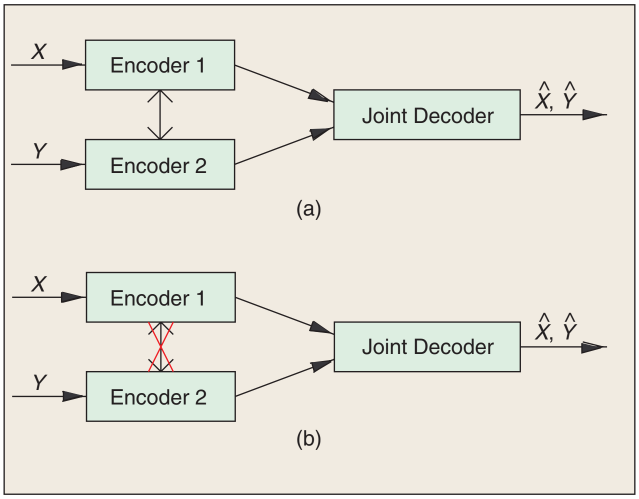
A tutorial article, "Distributed Compression in a Dense Microsensor Network" [5], appeared in *IEEE Signal Processing Magazine* in 2002. Central to [5] is a practical DSC scheme called DISCUS (distributed source coding using syndromes) [6]. This current article is intended as a sequel to [5] with the main aim of covering our work on DSC and other relevant research efforts ignited by DISCUS.

We note that DSC is only one of the communication layers in a network, and its interaction with the lower communication layers, such as the transport, the network, the medium access control (MAC), and the physical layers, is crucial for exploiting the promised gains of DSC. Issues related to cross-layer design (e.g., queuing management [7], resource allocation [8], call admission control [9], and MAC [10]) are addressed in other articles in this issue. DSC cannot be used without proper synchronization between the nodes of a sensor network, i.e., several assumptions are made for the routing and scheduling algorithms and their connection to the utilized DSC scheme. In addition, compared to a separate design, further gains can be obtained by *jointly* designing the distributed source codes with the underlying protocols, channel codes, and modulation schemes.

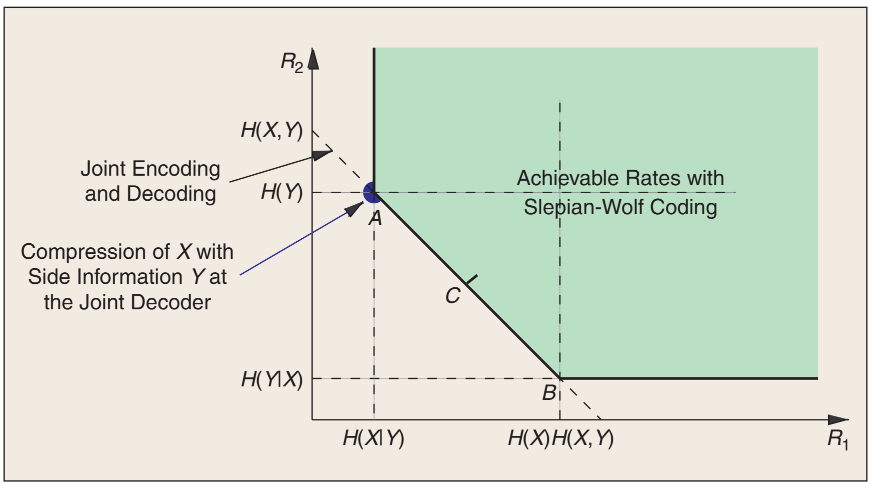
Many relevant results and references on DSC could not be included in this article. We refer readers to a longer version with more figures and tables and a comprehensive bibliography at http://lena.tamu.edu.

Slepian-Wolf Coding

Let $\{(X_i, Y_i)\}_{i=1}^{\infty}$ be a sequence of independent and identically distributed (i.i.d.) drawings of a pair of correlated discrete random variables X and Y. For lossless compression with $\hat{X} = X$ and $\hat{Y} = Y$ after decompression, we know from Shannon's source coding theory [3] that a rate given by the joint entropy $H(X, Y)$ of X and Y is sufficient if we are encoding them together [see Figure 1(a)]. For example, we can first compress



▲ 1. (a) Joint encoding of X and Y. The encoders collaborate and a rate $H(X, Y)$ is sufficient. (b) Distributed/separate encoding of X and Y. The encoders do not collaborate. The Slepian-Wolf theorem says that a rate $H(X, Y)$ is also sufficient provided that decoding of X and Y is done jointly.



▲ 2. The Slepian-Wolf rate region for two sources.

Y into $H(Y)$ bits per sample and based on the complete knowledge of Y at both the encoder and the decoder, we then compress X into $H(X|Y)$ bits per sample. But what if X and Y must be separately encoded for some user to reconstruct both of them?

One simple way is to do separate coding with rate $R = H(X) + H(Y)$, which is greater than $H(X, Y)$ when X and Y are correlated. In a landmark paper [1], Slepian and Wolf showed that $R = H(X, Y)$ is sufficient even for separate encoding of correlated sources [see Figure 1(b)]! The Slepian-Wolf theorem says that the achievable region of DSC for discrete sources X and Y is given by $R_1 \geq H(X|Y)$, $R_2 \geq H(Y|X)$ and $R_1 + R_2 \geq H(X, Y)$, which is shown in Figure 2. The proof of the Slepian-Wolf theorem is based on random *binning*. Binning is a key concept in DSC and refers to partitioning the space of all possible outcomes of a random source into disjoint subsets or *bins*. Examples explaining the binning process are given in "Slepian-Wolf Coding Examples." The achievability of Slepian-Wolf coding was generalized by Cover [3] to arbitrary ergodic processes, countably infinite alphabets, and arbitrary number of correlated sources.

Slepian-Wolf Coding of Two Binary Sources

Just like in Shannon's channel coding theorem [3], the random binning argument used in the proof of the Slepian-Wolf theorem is asymptotic and nonconstructive. For practical Slepian-Wolf coding, we can first try to design codes to approach the blue corner point A with $R_1 + R_2 = H(X|Y) + H(Y) = H(X, Y)$ in the Slepian-Wolf rate region of Figure 2. This is a problem of source coding of X with side information Y at the decoder as depicted in Figure 3 in "Slepian-Wolf Coding Examples." If this can be done, then the other corner point B of the Slepian-Wolf rate region can be approached by swapping the roles of X and Y and all points between these two corner points can be realized by time sharing—for example, using the two codes designed for the corner points 50% of the time each will result in the mid-point C.

Although constructive approaches (e.g., [12], [13]) have been proposed to directly approach the midpoint C in Figure 2 and progress has recently been made in practical code designs (see [14] and references therein) that can approach any point between A and B due to the space limitation, we only consider code designs for the corner points (or source coding with side information at the decoder) in this article. The former approaches are referred to as *symmetric* and the latter as *asymmetric*.

In asymmetric coding, our aim is to code X at a rate that approaches $H(X|Y)$ based on the conditional statistics of (or the correlation model between) X and Y but not the specific $\mathbf{y}$ at the encoder. Wyner first realized the close connection of DSC to channel coding and suggested the use of linear channel codes as a constructive approach for Slepian-Wolf coding in his 1974 paper [11]. The basic idea was to partition the space of all possible source outcomes into disjoint bins (sets) that are the cosets of some "good" linear channel code for the specific correlation model (see "Slepian-Wolf Coding Examples" for examples and more detailed explanation). Consider the case of binary symmetric sources and Hamming distance measure, with a linear (n, k) binary block code; there are 2^{n-k} distinct syndromes, each indexing a bin (set) of 2^k binary words of length n. Each bin is a coset code of the linear binary block code, which means that the Hamming distance properties of the original linear code are preserved in each bin. In compressing, a sequence of n input bits is mapped into its corresponding $(n-k)$ syndrome bits, achieving a compression ratio of $n:(n-k)$. This approach, known as "Wyner's scheme" [11] for some time, was only recently used in [6] for practical Slepian-Wolf code designs based on conventional channel codes like block and trellis codes.

If the correlation between X and Y can be modeled

Slepian-Wolf Coding Examples

1) Assume X and Y are equiprobable binary triplets with $X, Y \in \{0,1\}^3$ and they differ at most in one position. (We first start with the same example given in [5] and [6] and extend it to more general cases later.) Then $H(X) = H(Y) = 3$ b. Because the Hamming distance between X and Y is $d_H(X,Y) \leq 1$, for a given Y, there are four equiprobable choices of X. For example, when $Y = 000$, $X \in \{000, 100, 010, 001\}$. Hence $H(X|Y) = 2$ b. For joint encoding of X and Y, three bits are needed to convey Y and two additional bits to index the four possible choices of X associated with Y, thus a total of $H(X,Y) = H(Y) + H(X|Y) = 5$ b suffice.

For source coding with side information at the decoder as depicted in Figure 3, the side information Y is perfectly known at the decoder but *not* at the encoder, by the Slepian-Wolf theorem, it is still possible to send $H(X|Y) = 2$ b instead of $H(X) = 3$ b for X and decode it without loss at the joint decoder. This can be done by first partitioning the set of all possible outcomes of X into four bins (sets) Z_{00}, Z_{01}, Z_{10}, and Z_{11} with $Z_{00} = \{000, 111\}$, $Z_{01} = \{001, 110\}$, $Z_{10} = \{010, 101\}$ and $Z_{11} = \{011, 100\}$ and then sending two bits for the index $\mathbf{s}$ of the bin (set) $Z_{\mathbf{s}}$ that X belongs to. In forming the bins $Z_{\mathbf{s}}$'s, we make sure that each of them has two elements with Hamming distance $d_H = 3$. For joint decoding with $\mathbf{s}$ (hence $Z_{\mathbf{s}}$) and side information Y, we pick in bin $Z_{\mathbf{s}}$ the X with $d_H(X,Y) \leq 1$. Unique decoding is guaranteed because the two elements in each bin $Z_{\mathbf{s}}$ have Hamming distance $d_H = 3$. Thus we indeed achieve the Slepian-Wolf limit of $H(X,Y) = H(Y) + H(X|Y) = 3 + 2 = 5$ bits in this example with lossless decoding.

To cast the above example in the framework of coset codes and syndromes [11], we form the parity-check matrix $\mathbf{H}$ of rate 1/3 repetition channel code as

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}. \tag{1}$$

Then if we think of $\mathbf{x}$ as a length-3 binary word, the index $\mathbf{s}$ of the bin $Z_{\mathbf{s}}$ is just the syndrome $\mathbf{s} = \mathbf{x}\mathbf{H}^T$ associated with all $\mathbf{x} \in Z_{\mathbf{s}}$. Sending the 2-b syndrome $\mathbf{s}$ instead of the original 3-b $\mathbf{x}$ achieves a compression ratio of 3:2. In partitioning the eight $\mathbf{x}$ according to their syndromes into four disjoint bins $Z_{\mathbf{s}}$, we preserve the Hamming distance properties of the repetition code in each bin $Z_{\mathbf{s}}$. This ensures the same decoding performance for different syndromes.

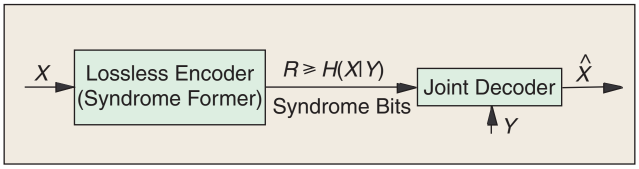
In channel coding, the set of the length three vectors $\mathbf{x}$ satisfying $\mathbf{s} = \mathbf{x}\mathbf{H}^T$ is called a coset code $\mathcal{C}_{\mathbf{s}}$ of the linear channel code $\mathcal{C}_{00}$ defined by $\mathbf{H}$ ($\mathbf{s} = \mathbf{x}\mathbf{H}^T = 00$ for the linear code). It is easy to see in this example that each coset code corresponds to a bin $Z_{\mathbf{s}}$, i.e., all the members of the bin $Z_{\mathbf{s}}$ are code words of the coset code $\mathcal{C}_{\mathbf{s}}$ and vice versa, i.e., all the code words of $\mathcal{C}_{\mathbf{s}}$ also belong to $Z_{\mathbf{s}}$. The Hamming distance properties between the code words of the linear rate 1/3 repetition channel code $\mathcal{C}_{00}$, are the same as those between the code words of each coset code $\mathcal{C}_{\mathbf{s}}$. Given the index $\mathbf{s}$ of the bin $Z_{\mathbf{s}}$, i.e., a specific coset code $\mathcal{C}_{\mathbf{s}}$, the side information Y can indicate its closest code word in $\mathcal{C}_{\mathbf{s}}$, and thus, recover X, or equivalently $\mathbf{x}$, without an error.

2) We now generalize the above example to the case when X and Y are equiprobable $(2^m - 1)$-bit binary sources. Here $m \geq 3$ is a positive integer. The correlation model between X and Y is again characterized by $d_H(X,Y) \leq 1$. Let $n = 2^m - 1$ and $k = n - m$, then $H(X) = H(Y) = n$ bits, $H(X|Y) = m$ bits and $H(X,Y) = n + m$ bits. Assuming that $H(Y) = n$ bits are spent in coding Y so that it is perfectly known at the decoder, for Slepian-Wolf coding of X, we enlist the help of the $m \times n$ parity-check matrix $\mathbf{H}$ of the (n,k) binary Hamming channel code. When $m = 3$,

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}. \tag{2}$$

For each n bit input $\mathbf{x}$, its corresponding syndrome $\mathbf{s} = \mathbf{x}\mathbf{H}^T$ is coded using $H(X|Y) = m$ bits, achieving the Slepian-Wolf limit with a compression ratio of $n:m$ for X. In doing so, we again partition the total of 2^n binary words according to their syndromes into 2^m disjoint bins (or $Z_{\mathbf{s}}$'s), indexed by the syndrome $\mathbf{s}$ with each set containing 2^k elements. With this partition, all 2^k code words of the linear (n,k) binary Hamming channel code $\mathcal{C}_{\mathbf{0}}$ are included in the bin $Z_{\mathbf{0}}$ and the code distance properties (minimum Hamming distance of three) is preserved in each coset code $\mathcal{C}_{\mathbf{s}}$. This way X, or equivalently $\mathbf{x}$, is recovered correctly.

Again, a linear channel code with its coset codes were used to do the binning and hence to construct a Slepian-Wolf limit achieving source code. The reason behind the use of channel codes is that the correlation between the source X and the side information Y can be modeled with a virtual "correlation channel." The input of this "channel" is X and its output is Y. For a received syndrome $\mathbf{s}$, the decoder uses Y together with the correlation statistics to determine which code word of the coset code $\mathcal{C}_{\mathbf{s}}$ was input to the "channel." So if the linear channel code $\mathcal{C}_{\mathbf{0}}$ is a good channel code for the "correlation channel," then the Slepian-Wolf source code defined by the coset codes $\mathcal{C}_{\mathbf{s}}$ is also a good source code for this type of correlation.



▲ *3. Lossless source coding with side information at the decoder as one case of Slepian-Wolf coding.*

by a binary channel, Wyner's syndrome concept can be extended to all binary linear codes, and state-of-the-art near-capacity channel codes such as turbo and LDPC codes [15] can be employed to approach the Slepian-Wolf limit. A short summary on turbo and LDPC codes is provided in "Turbo and LDPC Codes." There is, however, a small probability of loss in general at the Slepian-Wolf decoder due to channel coding (*zero-error* coding of correlated sources is outside the scope of this article). In practice, the linear channel code rate and code design in Wyner's scheme depend on the correlation model. Toy examples like those in "Slepian-Wolf Coding Examples" are tailor-designed with correlation models satisfying $H(X) = n$ bits and $H(X|Y) = n - k$ bits so that binary (n, k) Hamming codes can be used to exactly achieve the Slepian-Wolf limits.

A more practical correlation model than those of "Slepian-Wolf Coding Examples" is the binary symmetric model, where $\{(X_i, Y_i)\}_{i=1}^{\infty}$ is a sequence of i.i.d. drawings of a pair of correlated binary Bernoulli(0.5) random variables X and Y and the correlation between X and Y is modeled by a "virtual" binary symmetric case (BSC) with crossover probability p. In this channel $H(X|Y) = H(p) = -p \log_2 p - (1-p) \log_2 (1-p)$. Although this correlation model looks simple, the Slepian-Wolf coding problem is not trivial. As the BSC is a well-studied channel model in channel coding with a number of capacity-approaching code designs available [15], through the close connection between Slepian-Wolf coding and channel coding, this progress could be exploited to design Slepian-Wolf limit approaching codes. The first such practical designs [13], [16], [17] borrowed concepts from channel coding, especially turbo codes, but did not establish the more direct link with channel coding through the syndromes and the coset codes, as in "Slepian-Wolf Coding Examples." A

Turbo and LDPC Codes

Although Gallager discovered low-density parity-check (LDPC) codes 40 years ago, much of the recent developments in near-capacity codes on graphs and iterative decoding [15], ranging from turbo codes to the rediscovery of LDPC codes only happened in the last decade.

Turbo codes and more generally concatenated codes are a class of near-capacity channel codes that involve concatenation of encoders via random interleaving in the encoder and iterative maximum-a-posteriori (MAP) decoding, also called Bahl-Cocke-Jelinek-Raviv (BCJR) decoding, for the component codes in the decoder. A carefully designed concatenation of two or more codes performs better than each of the component codes alone. The role of the pseudo-random interleaver is to make the code appear random while maintaining enough code structure to permit decoding. The BCJR decoding algorithm allows exchange of information between decoders and thus an iterative interaction between the decoders (hence turbo), leading to near-capacity performance.

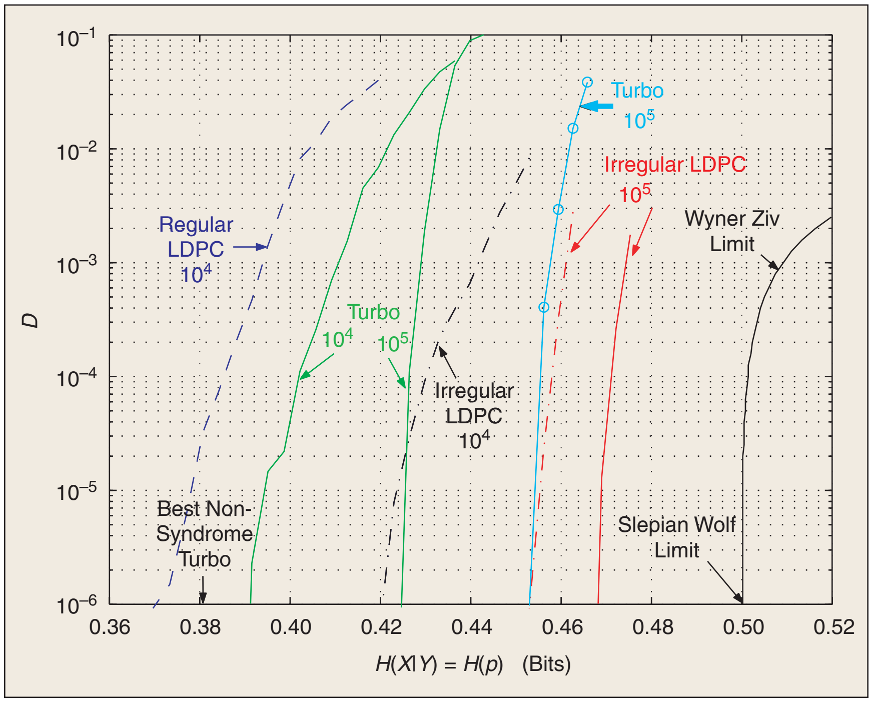
LDPC codes [15] are block codes, best described by their parity-check matrix **H** and the associated bipartite graph. The parity-check matrix **H** of a binary LDPC code has a small number of ones (hence low density). The way these ones are spread in **H** is described by the degree distribution polynomials $\tilde{\lambda}(x)$ and $\tilde{\rho}(x)$, which indicate the percentage of columns and rows of **H** respectively, with different Hamming weights (number of ones). When both $\tilde{\lambda}(x)$ and $\tilde{\rho}(x)$ have only a single term, the LDPC code is *regular*, otherwise it is *irregular*. In general, an optimized irregular LDPC code is expected to be more powerful than a regular one of the same code word length and code rate. Given both $\tilde{\lambda}(x)$ and $\tilde{\rho}(x)$, the code rate is exactly determined, but there are several channel codes and thus, several **H**'s that can be formed. Usually one is constructed randomly.

The bipartite graph of an LDPC code is an equivalent representation of the parity-check matrix **H**. Each column is represented with a variable or left node and each row with a check or right node. All variable (left) nodes are put in one column, all check (right) nodes in a parallel column and then wherever there is an one in **H**, there is an edge connecting the corresponding variable and check node. The bipartite graph is used in the decoding procedure, allowing the application of the message-passing algorithm [15]. (Iterative decoding algorithms including the sum-product algorithm, the forward/backward algorithm in speech recognition and the belief propagation algorithm in artificial intelligence are collectively called message-passing algorithms.) LDPC codes are the most powerful channel codes nowadays [15] and they can be designed via density evolution [15].

Both turbo and LDPC codes exhibit near-approaching performance with reasonable complexity over most conventional channels; and algorithms have been devised for the design of very good turbo and LDPC codes. However, the design procedure for LDPC codes is more flexible and less complex and therefore allows faster, easier and more precise design. This has made LDPC codes the most powerful channel codes over both conventional and unconventional channels. But turbo and more generally concatenated codes are still employed in many applications when short block lengths are needed or when one of the component codes is required to satisfy additional properties (e.g., for joint source-channel coding).

turbo scheme with structured component codes was used in [16] and parity bits were sent in [13] and [17] instead of syndrome bits as advocated in Wyner's scheme. Code design that did follow Wyner's scheme for this problem was done by Liveris et al. [18] with turbo/LDPC codes, achieving performance better than in [13], [16], and [17] and very close to the Slepian-Wolf limit $H(p)$.

Some simulation results from [18] are shown in Figure 4. The horizontal axis in Figure 4 shows the amount of correlation, i.e., lower $H(p)$ means higher correlation, and the vertical axis shows the probability of error for the decoded X. All coding schemes in Figure 4 achieve 2:1 compression, and therefore the Slepian-Wolf limit is shown to be 0.5 b at almost zero error probability. The performance limit of the more



▲ *4. Slepian-Wolf coding of binary X with side information Y at the decoder based on turbo/LDPC codes. The rate for all codes is fixed at 1/2.*

Wyner-Ziv Coding: The Binary Symmetric Case and the Quadratic Gaussian Case

Wyner-Ziv coding generalizes the setup of Slepian-Wolf coding [1] in that coding of X in Figure 3 is with respect to a fidelity criterion rather than lossless.

The Binary Symmetric Case

X and Y are binary symmetric sources, the correlation between them is modeled as a BSC with crossover probability p and the distortion measure is the Hamming distance. We can write $X = Y \oplus E$, where E is a Bernouli(p) source. Then the rate-distortion function $R_E(D)$ for E serves as the performance limit $R_{X|Y}(D)$ of lossy coding of X given Y at both the encoder and the decoder. From [3] we have

$$R_{X|Y}(D) = R_E(D) = \begin{cases} H(p) - H(D), & 0 \le D \le \min\{p, 1-p\}, \\ 0, & D > \min\{p, 1-p\}. \end{cases} \tag{3}$$

On the other hand, the Wyner-Ziv rate-distortion function in this case is [2], [21]

$$R^*_{WZ}(D) = l.c.e\{H(p * D) - H(D), (p, 0)\}, 0 \le D \le p, \tag{4}$$

the lower convex envelope of $H(p*D) - H(D)$ and the point $(D = p, R = 0)$, where $p * D = (1-p)D + (1-D)p$. For $p \le 0.5$, $R^*_{WZ}(D) \ge R_{X|Y}(D)$ with equality only at two distortion-rate points: the zero-rate point $(p, 0)$ and the zero-distortion (or Slepian-Wolf) point $(0, H(p))$. See as depicted later in Figure 6 for $p = 0.27$. Thus Wyner-Ziv coding suffers rate loss in this binary symmetric case. When $D = 0$, the Wyner-Ziv problem degenerates to the Slepian-Wolf problem with $R^*_{WZ}(0) = R_{X|Y}(0) = H(X|Y) = H(p)$.

The Quadratic Gaussian Case

X and Y are zero mean and stationary Gaussian memoryless sources and the distortion metric is mean-squared error (MSE). Let the covariance matrix of X and Y be

$$\Lambda = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

with $|\rho| < 1$, then [2]

$$R^*_{WZ}(D) = R_{X|Y}(D) = \frac{1}{2}\log^+\left[\frac{\sigma_X^2(1-\rho^2)}{D}\right], \tag{5}$$

where $\log^+ x = \max\{\log_2 x, 0\}$. There is no rate loss with Wyner-Ziv coding in this quadratic Gaussian case. If Y can be written as $Y = X + Z$, with independent $X \sim N(0, \sigma_X^2)$ and $Z \sim N(0, \sigma_Z^2)$, then

$$R^*_{WZ}(D) = R_{X|Y}(D) = \frac{1}{2}\log^+\left[\frac{\sigma_Z^2}{(1+\sigma_Z^2/\sigma_X^2)D}\right]. \tag{6}$$

On the other hand, if $X = Y + Z$, with independent $Y \sim N(0, \sigma_Y^2)$ and $Z \sim N(0, \sigma_Z^2)$, then

$$R^*_{WZ}(D) = R_{X|Y}(D) = \frac{1}{2}\log^+\left(\frac{\sigma_Z^2}{D}\right). \tag{7}$$

The seemingly source coding problem of Slepian-Wolf coding is actually a channel coding problem.

general Wyner-Ziv coding described in "Wyner-Ziv Coding: The Binary Symmetric Case and the Quadratic Gaussian Case" and the section "The Binary Symmetric Case" is also included. The code-word length of each Slepian-Wolf code is also given in Figure 4. It can be seen that the higher the correlation, the lower the probability of error for a certain Slepian-Wolf code. The more powerful the Slepian-Wolf code, the lower the correlation needed for the code to achieve very low probability of error. Clearly, stronger channel codes, e.g., same family of codes with longer code word length in Figure 4, for the BSC result in better Slepian-Wolf codes in the binary symmetric correlation setup.

This last statement can be generalized to any correlation model: if the correlation between the source output X and the side information Y can be modeled with a "virtual" correlation channel, then a good channel code over this channel can provide us with a good Slepian-Wolf code through the syndromes and the associated coset codes. Thus the seemingly source coding problem of Slepian-Wolf coding is actually a channel coding one and near-capacity channel codes such as turbo and LDPC codes can be used to approach the Slepian-Wolf limits.

Slepian-Wolf Coding of Multiple Sources with Arbitrary Correlation

Practical Slepian-Wolf code designs for other correlation models and more than two sources have appeared recently in the literature (e.g., [19]) using powerful turbo and LDPC codes. However, there is no systematic approach to general practical Slepian-Wolf code design yet, in the sense of being able to account for an arbitrary number of sources with nonbinary alphabets and possibly with memory in the marginal source and/or correlation statistics. Most designs do not consider sources with memory and/or the memory in the correlation between the sources. The only approach taking into account correlation with memory is [20]. The lack of such results is due to the fact that the channel coding analog to such a general Slepian-Wolf coding problem is a channel code design problem over channels that resemble more involved communication channels, and thus, has not been adequately studied yet.

Nevertheless, there has been a significant amount of work, and for a number of different scenarios the available Slepian-Wolf code designs perform well. These designs are very important, as asymmetric or symmetric Slepian-Wolf coding plays the role of conditional or joint entropy coding, respectively. Not only can Slepian-Wolf coding be considered the analog to entropy coding in classic lossless source coding but also the extension of entropy coding to problems with side information and/or distributed sources. In these more general problems, near-lossless compression of a source down to its entropy is a special case of Slepian-Wolf coding when there is no correlation either between the source and the side information (asymmetric setup) or between the different sources (symmetric setup).

So, apart from its importance as a separate problem, when combined with quantization, Slepian-Wolf can provide a practical approach to lossy DSC problems, such as the Wyner-Ziv problem considered next, similarly to the way quantization and entropy coding are combined in classic lossy source coding.

Wyner-Ziv Coding

In the previous section, we focused on lossless source coding of discrete sources with side information at the decoder as one case of Slepian-Wolf coding. In sensor network applications, we are often dealing with continuous sources; then the problem of rate distortion with side information at the decoder arises. The question to ask is how many bits are needed to encode X under the constraint that the average distortion between X and the coded version $\hat{X}$ is $E\{d(X, \hat{X})\} \leq D$, assuming the side information Y is available at the decoder but not at the encoder. This problem, first considered by Wyner and Ziv in [2], is one instance of DSC with Y available uncoded as side information at the decoder. It generalizes the setup of [1] in that coding of discrete X is with respect to a fidelity criterion rather than lossless. For both discrete and continuous alphabet cases and general distortion metrics $d(\cdot)$, Wyner and Ziv [2] gave the rate-distortion function $R^*_{WZ}(D)$ for this problem. We include in "Wyner-Ziv Coding: The Binary Symmetric Case and the Quadratic Gaussian Case" the Wyner-Ziv rate distortion functions for the binary symmetric case and the quadratic Gaussian case.

The important thing about Wyner-Ziv coding is that it usually suffers rate loss when compared to lossy coding of X when the side information Y is available at both the encoder and the decoder (see the binary symmetric case in "Wyner-Ziv Coding: The Binary Symmetric Case and the Quadratic Gaussian Case"). One exception is when X and Y are jointly Gaussian with MSE measure (the quadratic Gaussian case in "Wyner-Ziv Coding: The Binary Symmetric Case and the Quadratic Gaussian Case"). There is no rate loss with Wyner-Ziv coding in this case, which is of special interest in practice (e.g., sensor networks) because many image and video sources can be modeled as jointly Gaussian (after mean subtraction). Pradhan et al. [22] recently extended the no rate loss condition for Wyner-Ziv coding to $X = Y + Z$, where Z is independently Gaussian but X and Y could follow more general distributions.

Main Results in Source Coding (Quantization Theory)

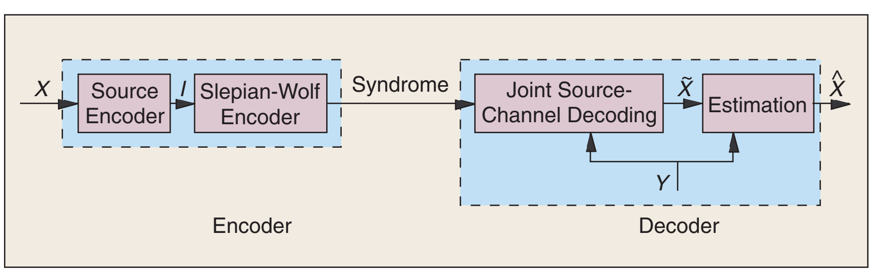
For Gaussian source $X \sim N(0, \sigma_X^2)$ with MSE distortion metric, the rate-distortion function [3] is $R_X(D) = (1/2)\log^+(\sigma_X^2/D)$ and the distortion-rate function is $D_X(R) = \sigma_X^2 2^{-2R}$. At high rate, the minimum MSE (or Lloyd-Max) scalar quantizer, which is nonuniform, performs 4.35 dB away from $D_X(R)$ [24]; and the entropy-coded scalar quantizer (or uniform scalar quantization followed by ideal entropy coding) for X suffers only 1.53 dB loss with respect to $D_X(R)$ [24]. High dimensional lattice vector quantizers [24] can be found in dimensions 2, 4, 8, 16, and 24 that perform 1.36, 1.16, 0.88, 0.67 and 0.50 dB away, respectively, from $D_X(R)$. In addition, trellis-coded quantization (TCQ) can be employed to implement equivalent vector quantizers in even higher dimensions, achieving better results. For example, TCQ can perform 0.46 dB away from $D_X(R)$ at 1 b/s for Gaussian sources [24] and entropy-coded TCQ can get as close as 0.2 dB to $D_X(R)$ for any source with smooth PDF [24].

From an information-theoretical perspective, according to [23], there are granular gain and boundary gain in source coding and packing gain and shaping gain in channel coding. Wyner-Ziv coding is foremost a source coding (i.e., a rate-distortion) problem; one should consider the granular gain and the boundary gain. In addition, the side information necessitates channel coding for compression (e.g., via Wyner's syndrome-based binning scheme [11]), which utilizes a linear channel code together with its coset codes. Thus, channel coding in Wyner-Ziv coding is not conventional in the sense that there is only packing gain but no shaping gain. One needs to establish the equivalence between the boundary gain in source coding and the packing gain in channel coding for Wyner-Ziv coding; this is feasible because channel coding for compression in Wyner-Ziv coding can perform conditional entropy coding to achieve the boundary gain—the same way as entropy coding achieves the boundary gain in classic source coding [23], [24 p. 123]. Then in Wyner-Ziv coding, he/she can shoot for the granular gain via source coding and the boundary gain via channel coding.

From a practical viewpoint, because we are introducing loss/distortion to the source with Wyner-Ziv coding, source coding is needed to quantize X. "Main Results in Source Coding (Quantization Theory)" reviews main results in source coding (quantization theory) regarding Gaussian sources.

Usually there is still correlation remaining in the quantized version of X and the side information Y, and Slepian-Wolf coding should be employed to exploit this correlation to reduce the rate. Since Slepian-Wolf coding is based on channel coding, Wyner-Ziv coding is, in a nutshell, a source-channel coding problem. There are quantization loss due to source coding and binning loss due to channel coding. To reach the Wyner-Ziv limit, one needs to employ both source codes [e.g., trellis coded quantization (TCQ)] that can achieve the granular gain and channel codes (e.g., turbo and LDPC codes) that can approach the Slepian-Wolf limit. In addition, the side information Y can be used in jointly decoding and estimating $\hat{X}$ at the decoder to help reduce the distortion $d(X, \hat{X})$ for nonbinary sources, especially at low bit rates. The intuition is that in decoding X, the joint decoder should rely more on Y when the rate is too low to make the coded version of X to be useful in terms of lowering the distortion. On the other hand, when the rate is high, the coded version of X becomes more reliable than Y so the decoder should put more weight on the former in estimating $\hat{X}$. Figure 5 depicts the block diagram of a generic Wyner-Ziv coder.

To illustrate the basic concepts of Wyner-Ziv coding we consider nested lattice codes, which were introduced by Zamir et al. [21] as codes that can achieve the Wyner-Ziv limit asymptotically, for large dimensions. Practical nested lattice code implementation was first done in [25]. The coarse lattice is nested in the fine lattice in the sense that each point of the coarse lattice is also a point of the fine lattice but not vice versa. The fine code in the nested pair plays the role of source coding while each coset coarse code does channel coding. To encode, **x** is first quantized with respect to the fine source code, resulting in quantization loss. However, only the index identifying the bin (coset channel code) that contains the quantized **x** is coded to save rate. Using this coded index, the decoder finds in the bin (coset code) the code word closest to the side information **y** as the best estimate of **x**. Due to the binning process employed in nested coding, the Wyner-Ziv decoder suffers a small probability of error. To



▲ *5. Block diagram of a generic Wyner-Ziv coder.*

Wyner-Ziv coding is, in a nutshell, a source-channel coding problem.

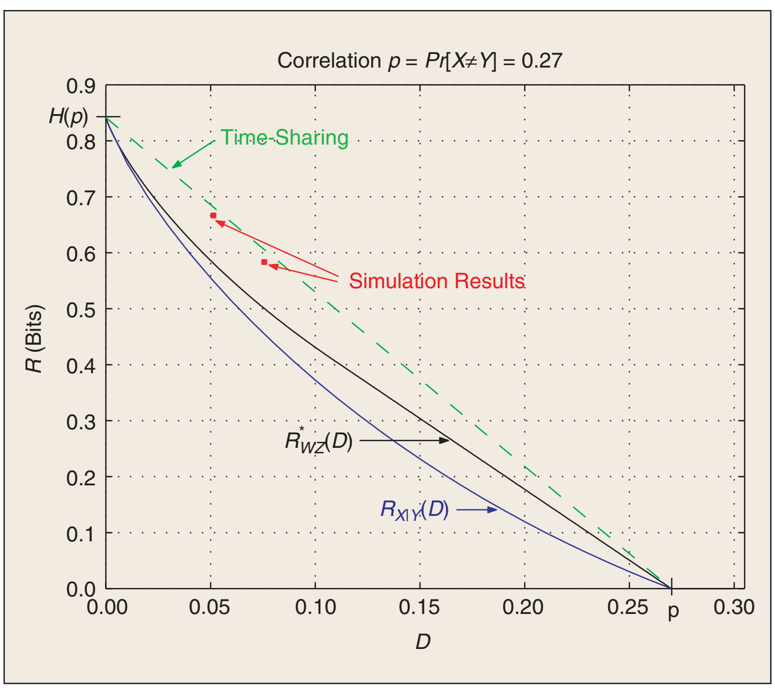
reduce this decoding error probability, the coarse channel code has to be strong with large minimum distance between its code words. This means that the dimensionality of the coarse linear/lattice code needs to be high. It is proven in [21] that infinite dimensional source and channel codes are needed to reach the Wyner-Ziv limit. This result is only asymptotical and not practical. Implemented code designs for the binary symmetric and the quadratic Gaussian cases are discussed next.

The Binary Symmetric Case

Recall that in Wyner's scheme [11] for lossless Slepian-Wolf coding, a linear (n, k) binary block code is used. There are 2^{n-k} distinct syndromes, each indexing a set (bin) of 2^k binary words of length n that preserve the Hamming distance properties of the original code. In compressing, a sequence of n input bits is mapped into its corresponding $(n-k)$ syndrome bits, achieving a compression ratio of $n : (n-k)$.

For lossy Wyner-Ziv coding, Shamai, Verdu, and Zamir generalized Wyner's scheme using nested linear binary block codes [21], [26]. According to this nested scheme, a linear (n, k_2) binary block code is again used to partition the space of all binary words of length n into 2^{n-k_2} bins of 2^{k_2} elements, each indexed by a unique syndrome value. Out of these 2^{n-k_2} bins only $2^{k_1-k_2}$ $(k_1 \geq k_2)$ are used, and the elements of the remaining $2^{n-k_2} - 2^{k_1-k_2}$ sets are "quantized" to the closest, in Hamming distance sense, binary word of the allowable $2^{k_1-k_2} \times 2^{k_2} = 2^{k_1}$ ones. This "quantization" can be viewed as a (n, k_1) binary block source code. Then the linear (n, k_2) binary block code can be considered to be a coarse channel code nested inside the (n, k_1) fine source code.

To come close to the Wyner-Ziv limit, both codes in the above nested scheme should be good, i.e., a good fine source code is needed with a good coarse channel subcode [21], [26]. Knowing how to employ good channel codes based on Wyner's scheme $(k_1 = n)$ [18], Liveris et al. proposed a scheme in [27] based on concatenated codes, where from the constructions in [18] the use of good channel codes is guaranteed. As for the source code, its operation resembles that of TCQ and hence, it is expected to be a good source code. The scheme in [27] can come within 0.09 b from the theoretical limit (see Figure 6). This is the only result reported so far for the binary Wyner-Ziv problem.



▲ *6. Simulated performance of the nested scheme in [27] for binary Wyner-Ziv coding for correlation* $p = 0.27$*. The time-sharing line between the zero-rate point* $(p, 0)$ *and the Slepian-Wolf point* $(0, H(p))$ *is also shown.*

The binary symmetric Wyner-Ziv problem does not seem to be practical, but due to its simplicity, it provides useful insight into the interaction between source and channel coding. The rate loss from the Wyner-Ziv limit in the case of binary Wyner-Ziv coding can be clearly separated into source coding loss and channel coding loss. For example, in Figure 6 the rate gap in bits between the simulated points and the Wyner-Ziv limit can be separated into source coding rate loss and channel coding rate loss. This helps us understand how to quantify and combine rate losses in Wyner-Ziv coding.

In the binary setup, there is no estimation involved as in the general scheme of Figure 5. The design process consists of two steps. First, a good classic binary quantizer is selected, i.e., a quantizer that can minimize distortion D close to the distortion-rate function of a single Bernoulli(0.5) source at a given rate. The second step is to design a Slepian-Wolf encoder matched to the quantizer codebook. The better the matching of the Slepian-Wolf code constraints (parity check equations) to the quantizer codebook, the better the performance of the decoder. Joint source-channel decoding in Figure 5 refers to the fact that the decoder combines the Slepian-Wolf code constraints with the quantizer codebook to reconstruct X. This binary design approach is very helpful in understanding the Gaussian Wyner-Ziv code design.

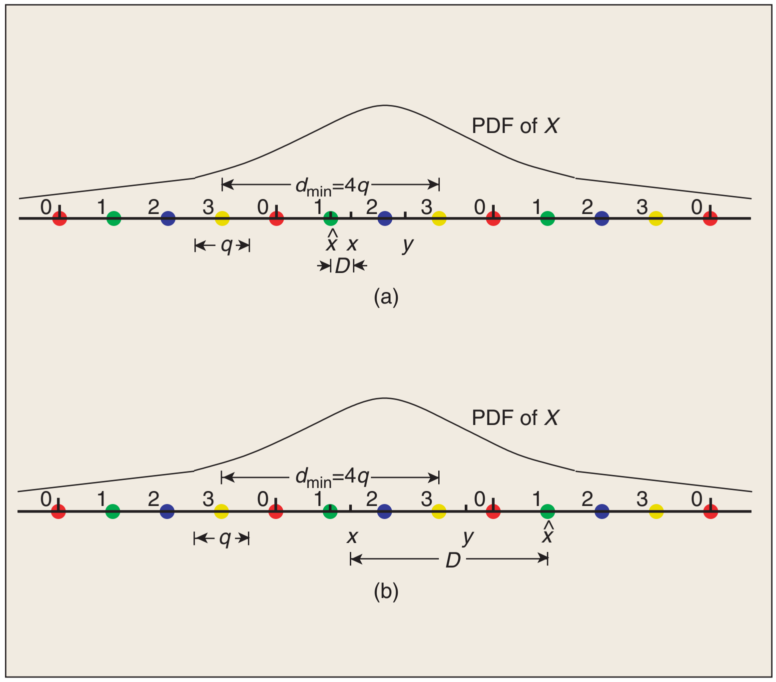
The Quadratic Gaussian Case

For practical code design in this case, one can first consider lattice codes [28] and trellis-based codes [23] that have been used for both source and channel coding in the past and focus on finding good nesting codes among them. Following Zamir et al.'s theoretical nested lattice coding scheme, Servetto [25] proposed explicit nested lattice constructions based on similar sublattices for the high correlation case. Figure 7 shows the simplest one-dimensional (1-D) nested lattice/scalar quantizer with $N = 4$ bins, where the fine source code employs a uniform scalar quantizer with stepsize q and the coarse channel code uses a 1-D lattice code with minimum distance $d_{\min} = Nq$. The distortion consists of two parts: the "good" distortion introduced from quantization by the source code and the "bad" distortion from decoding error of the channel code. To reduce the "good" distortion, it is desirable to choose a small quantization stepsize q; on the other hand, to limit the "bad" distortion, $d_{\min} = Nq$ should be maximized to minimize the channel decoding error probability P_e. Thus for a fixed N, there exists an optimal q that minimizes the total distortion.

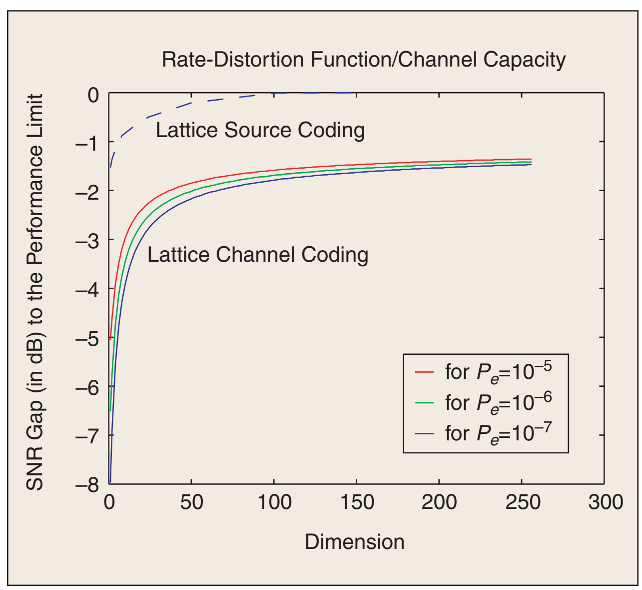
Research on trellis-based nested codes as a way of realizing high-dimensional nested lattice codes has just started recently. For example, in DISCUS [6], two source codes (scalar quantization and TCQ) and two channel codes (scalar coset code and trellis-based coset code [23]) are used in source-channel coding for the Wyner-Ziv problem, resulting in four combinations. One of them (scalar quantization with scalar coset code) is nested scalar quantization and another one (TCQ with trellis-based coset code) can effectively be considered as nested TCQ.

Nested lattice or TCQ constructions in [6] might be the first approach one would attempt because source and channel codes of about the same dimension are utilized. However, in this setup, the coarse channel code is not strong enough. This can be seen clearly from Figure 8, where performance bounds [28], [29] of lattice source and channel codes are plotted together. With nested scalar quantization, the fine source code (scalar quantization) leaves unexploited the maximum granular gain of only 1.53 dB [24] but the coarse channel code (scalar coset code) suffers more than 6.5 dB loss with respect to the capacity (with $P_e = 10^{-6}$). On the other hand, Figure 8 indicates that lattice channel code at dimension 250 still performs more than 1 dB away from the capacity. Following the 6-dB rule, i.e., that every 6 dB correspond to 1 b, which is approximately true for both source and channel coding, the decibel gaps can be converted into rate losses (bits) and then combined into a single rate loss from the Wyner-Ziv rate-distortion function.

Nested TCQ employed in DISCUS [6] can be viewed as a nested lattice vector quantizer, where the lattice source code corresponding to TCQ has a smaller gap from the performance limit than the lattice channel code (trellis-based coset code) of about the same dimension does. As the dimensionality increases, lattice source codes reach the ceiling of 1.53 dB in granular



▲ *7. A 1-D nested lattice/uniform quantizer with four bins for the quadratic Gaussian Wyner-Ziv problem, where Y is the side information only available at the decoder. (a) "Good" distortion only. (b) "Good" and "bad" distortion.*

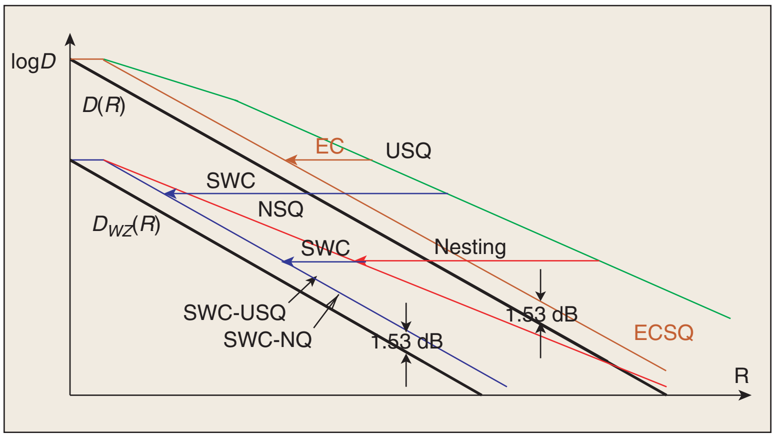


▲ *8. Lower bound in terms of the performance gap (in dB) of lattice source code from the rate-distortion function of Gaussian sources and lattice channel codes from the capacity of Gaussian channels, as a function of dimensionality.*

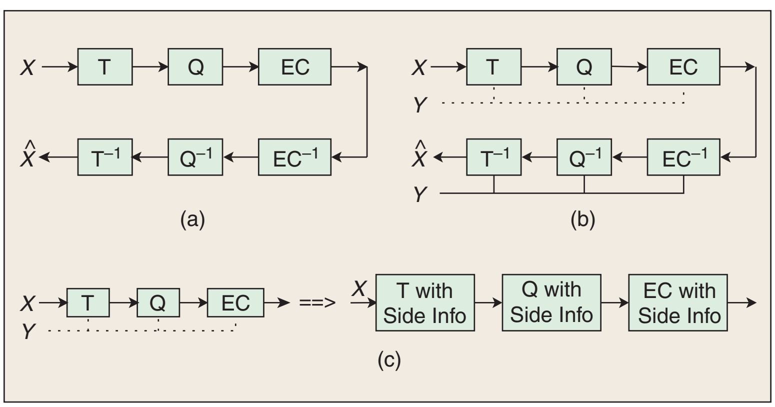
Wyner first realized the close connection of DSC to channel coding.

gain much faster than lattice channel codes approach the capacity. Consequently one needs channel codes of much higher dimension than source codes to achieve the same loss, and the Wyner-Ziv limit should be approached with nesting codes of different dimensionalities in practice. The need of strong channel codes for Wyner-Ziv coding was also emphasized in [21].

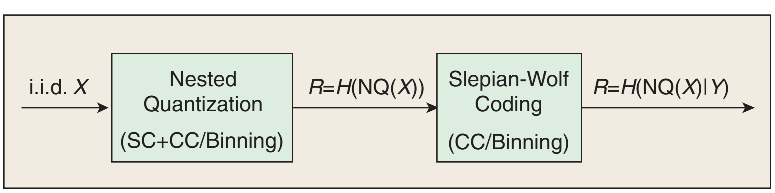
This leads to the second approach [30] based on Slepian-Wolf coded nested quantization (SWC-NQ), i.e., nested scheme followed by a second layer of binning. At high rate, asymptotic performance bounds of SWC-NQ similar to those in classic source coding were established in [30], showing that ideal Slepian-Wolf coded 1-D/two-dimensional (2-D) nested lattice quantization performs 1.53/1.36 dB worse than the Wyner-Ziv distortion-rate function $D^*_{WZ}(R)$ with probability almost one. Performances close to the corresponding theoretical limits were obtained by using 1-D and 2-D nested lattice quantization, together with irregular LDPC codes for Slepian-Wolf coding.



▲ 9. Performance of Gaussian Wyner-Ziv coding with scalar quantization and Slepian-Wolf coding.



▲ 10. (a) Classic source coding. (b) DSC (source coding with side information at the decoder). (c) Equivalent encoder structure of DSC for correlated sources.



▲ 11. SWC-NQ for Wyner-Ziv coding of i.i.d. sources.

The third practical nested approach to Wyner-Ziv coding involves combined source and channel coding. The main scheme in this approach has been the combination of a classic scalar quantizer (no binning in the quantization) and a powerful Slepian-Wolf code. The intuition that all the binning should be left to the Slepian-Wolf code, allows the best possible binning (a high dimensional channel code). This limits the performance loss of such a Wyner-Ziv code to that from source coding alone. Some first interesting results were given in [17], where assuming ideal Slepian-Wolf coding and high rate the use of classic quantization seemed to be sufficient, leading to a similar 1.53 dB gap for classic scalar quantization with ideal SWC. In a more general context, this approach could be viewed as a form of nesting with fixed finite source code dimension and larger channel code dimension. This generalized context can include the turbo-trellis Wyner-Ziv codes introduced in [31], where the source code is a TCQ nested with a turbo channel code. However, the scheme in [31] can also be classified as a nested one in the second approach. Wyner-Ziv coding based on TCQ and LDPC codes was presented in [32], which shows that at high rate, TCQ with ideal Slepian-Wolf coding performs 0.2 dB away from the theoretical limit $D^*_{WZ}(R)$ with probability almost one. Practical designs with TCQ, irregular LDPC code based Slepian-Wolf coding and optimal estimation at the decoder can perform 0.82 dB away from $D^*_{WZ}(R)$ at medium bit rates (e.g., $\geq$ 1.5 b/s). With 2-D trellis-coded vector quantization (TCVQ), the performance gap to $D^*_{WZ}(R)$ is only 0.66 dB at 1.0 b/s and 0.47 dB at 3.3 b/s [32]. Thus we are approaching the theoretical performance limit of Wyner-Ziv coding.

From Classic Source Coding to Wyner-Ziv Coding

We shall start with the general case that in addition to the correlation between X and Y, $\{X_i\}_{i=1}^{\infty}$ are also correlated (we thus implicitly assume that $\{Y_i\}_{i=1}^{\infty}$ are correlated as well). The classic transform, quantization, and entropy coding (T-Q-EC) paradigm for X is show in Figure 10(a). When the side information Y is available at the decoder, we immediately obtain the DSC paradigm in Figure 10(b). Note that Y is not available at the encoder—the dotted lines in Figure 10 (b) only serve as a reminder that the design of each of the T,Q, and EC components should reflect the fact that Y is available at the decoder.

The equivalent *encoder* structure of DSC in Figure 10 (b) is redrawn in Figure 10(c). Each component in the classic T-Q-EC paradigm is now replaced with one that takes into account the side information at the *decoder*. For example, the first component "T with side info" could be the conditional Karhunen-Loeve transform [33], which is beyond the scope of this article.

Assuming that "T with side information" is doing a perfect job in the sense that it completely decorrelates X conditional on Y, from this point on we will assume i.i.d. X and Y and focus on "Q with side information" and "EC with side information" in Figure 10(c). We rename the former as *nested* quantization (the latter is exactly Slepian-Wolf coding) and end up with the encoder structure of SWC-NQ for Wyner-Ziv coding of i.i.d. sources in Figure 11.

Nested quantization in SWC-NQ plays the role of source-channel coding, in which the source coding component relies on the fine code and the channel coding component on the coarse code of the nested pair of codes. The channel coding component is introduced to the nested quantizer precisely because the side information Y is available at the decoder. It effectively implements a binning scheme to take advantage of this fact in the quantizer. Nested quantization in Figure 11 thus corresponds to quantization in classic source coding.

For practical lossless source coding, conventional techniques (e.g., Huffman coding, arithmetic coding, Lempel-Ziv coding, and others referred in [34]) have dominated so far. However, if one regards lossless source coding as a special case of Slepian-Wolf coding without side information at the decoder, then channel coding techniques can also be used for source coding based on syndromes (see references in [34]). In this light, the Slepian-Wolf coding component in Figure 11 can be viewed as the counterpart of entropy coding in classic source coding. Although the idea of using channel codes for source coding dates back to the Shannon-MacMillan theorem [3] and theoretical results appeared later [34], practical turbo/LDPC code based noiseless data compression schemes did not appear until very recently [34], [35].

Starting from syndrome based approaches for entropy coding, one can easily make the schematic connection between entropy-coded quantization for classic source coding and SWC-NQ for Wyner-Ziv coding, as syndrome based approaches can also be employed for Slepian-Wolf coding (or source coding with side information at the decoder) in the latter case. Performance-wise, the work in [30], [32] reveals that the performance gap of high-rate Wyner-Ziv coding (with ideal Slepian-Wolf coding) to $D^*_{WZ}(R)$ is exactly the same as that of high-rate classic source coding (with ideal entropy coding) to the distortion-rate function $D_X(R)$. This interesting and important finding is highlighted in Table 1.

Comparison of Different Approaches to the Quadratic Gaussian Case

Based on scalar quantization, Figure 9 illustrates the performance difference between classic uniform scalar quantization (USQ), the first approach with nested scalar quantization (NSQ), the second with SWC-NQ, and the third with ideal Slepian-Wolf coded uniform scalar quantization (SWC-USQ).

Although the last two approaches perform roughly the same, using nesting/binning in the quantization step in the second approach of SWC-NQ has the advantage that even without the additional Slepian-Wolf coding step, nested quantization (e.g., NSQ) alone performs better than the third approach of SWC-USQ without Slepian-Wolf coding, which degenerates to just classic quantization (e.g., USQ). At high rate, the nested quantizer asymptotically becomes almost a nonnested regular one so that strong channel coding is guaranteed and there is a constant gain in rate at the same distortion level due to nesting when compared with classic quantization. The role of Slepian-Wolf coding in both SWC-NQ and SWC-USQ is to exploit the correlation between the quantized source and the side information for further compression and in SWC-NQ, as a complementary binning layer to make the overall channel code stronger.

SWC-NQ generalizes the classic source coding approach of quantization (Q) and entropy coding (EC) in the sense that the quantizer performs quite well alone and can exhibit further rate savings by

Table 1. High-rate classic source coding versus high-rate Wyner-Ziv coding.

Classic Source Coding		Wyner-Ziv Coding	
Coding Scheme	Gap to $D_X(R)$	Coding Scheme	Gap to $D^*_{WZ}(R)$
ECSQ [24]	1.53 dB	SWC-NSQ [30]	1.53 dB
ECLQ (2-D) [28]	1.36 dB	SWC-NQ (2-D) [30]	1.36 dB
ECTCQ [24]	0.2 dB	SWC-TCQ [32]	0.2 dB

Slepian and Wolf theoretically showed that separate encoding is as efficient as joint encoding for lossless compression.

employing a powerful Slepian-Wolf code. This connection between entropy-coded quantization for classic source coding and SWC-NQ for Wyner-Ziv coding is developed in "From Classic Source Coding to Wyner-Ziv Coding."

Successive Wyner-Ziv Coding

Successive or scalable image coding made popular by SPIHT [36] is attractive in practical applications such as networked multimedia. For Wyner-Ziv coding, scalability is also a desirable feature in applications that go beyond sensor networks. Steinberg and Merhav [37] recently extended Equitz and Cover's work [38] on successive refinement of information to Wyner-Ziv coding, showing that both the doubly symmetric binary source and the jointly Gaussian source are successively refinable. Cheng et al. [39] further pointed out that the broader class of sources that satisfy the general condition of no rate loss [22] for Wyner-Ziv coding is also successively refinable. Practical layered Wyner-Ziv code design for Gaussian sources based on nested scalar quantization and multilevel LDPC code for Slepian-Wolf coding was also presented in [39]. Layered Wyner-Ziv coding of real video sources was introduced in [40].

Applications of DSC in Sensor Networks

In the above discussions we mainly considered lossless (Slepian-Wolf) and lossy (Wyner-Ziv) source coding with side information only at the decoder, as most of the work so far in DSC has been focusing on these two problems. For any network and especially a sensor network, this means that the nodes transmitting correlated information need to cooperate in groups of two or three so that one node provides the side information and another one can compress its information down to the Slepian-Wolf or the Wyner-Ziv limit. This approach has been followed in [41].

As pointed out in [41], such cooperation in groups of two and three sensor nodes means that less complex decoding algorithms should be employed to save some decoding processing power as the decoder is also a sensor node or a data processing node of limited power [41]. This is not a big issue as several low complexity channel decoding algorithms have been studied in the past years and so the decoding loss for using lower complexity algorithms can be minimized.

A way to change this assumption of cooperation in small groups and the associated low complexity decoding is to employ DSC schemes for multiple sources. Several Slepian-Wolf coding approaches for multiple sources (lossless DSC) have been discussed earlier. However, theoretical performance limits in the more general setting of lossy DSC for multiple sources still remain elusive. Even Wyner-Ziv coding assumes perfect side information at the decoder, i.e., it cannot really be considered a two sources problem. There are scant code designs (except [12], [14], and [42]) in the literature for the case of lossy DSC with two sources, where the side information is also coded.

The main issue for practical deployment of DSC is the correlation model. Although there has been significant effort in DSC designs for different correlation models, in some cases even application specific [5], [12], [41], [45], in practice it is usually hard to come up with a joint probability mass or density function in sensor networks, especially if there is little room for training or little information about the current network topology. In some applications, e.g., video surveillance networks, the correlation statistics can be mainly a function of the location of the sensor nodes. In case the sensor network has a training mode option and/or can track the varying network topology, adaptive or universal DSC that could achieve gains for time-varying correlation could be used to follow the time-varying correlation. Such universal DSC that can work well for a variety of correlation statistics seems to be the most appropriate approach for sensor networks, but it is still an open and very challenging DSC problem.

Measurement noise, which is another important issue in sensor networks, can be addressed through DSC. Following the Wyner-Ziv coding approach presented in the previous section, the existence of measurement noise, if not taken into account, causes some mismatch between the actual correlation between the sources and the noiseless correlation statistics used to do the decoding, which means worse performance. One way to resolve this issue is the robust code design discussed before. But if the noise statistics are known, even approximately, they can be incorporated into the correlation model and thus considered in the code design for DSC. There is actually one specific DSC

Table 2. Dual problems in source and channel coding, where the encoder for the source coding problem is the functional dual of the decoder for the corresponding channel coding problem and vice versa

Source Coding	Channel Coding
DSC [1], [2] (source coding with side information at the decoder)	Data hiding [46], [47] (channel coding with side information at the encoder)
Multiterminal source coding [4]	Coding for MIMO/broadcast channels [21], [48]
Multiple description coding [49]	Coding for multiple access channels [3]

problem, the chief executive officer (CEO) problem [44], which considers such a scenario. In this problem the CEO of a company employs a number of agents to observe an event, and each of the agents provides the CEO with his/her own (noisy) version of the event. The agents are not allowed to convene, and the goal of the CEO is to recover as much information as possible about the actual event from the noisy observations received from the agents, while minimizing the total information rate from the agents (sum rate). The CEO problem, hence, can account for the measurement noise at the sensor nodes. Preliminary practical code constructions for the CEO problem appeared in [12] and [42], based on the Wyner-Ziv coding approaches, but they are only limited to special cases.

Another issue is the cross-layer design aspect of DSC. DSC can be considered to be at the top of the sensor networks protocol stack, the application layer.Therefore, DSC sets several requirements for the underlying layers, especially the next lower layer, the transport layer, regarding synchronization between packets from correlated nodes.

DSC can also be designed to work together with the transport layer to make retransmissions smarter. In that sense, scalable DSC [39] seems to be the way to implement such smart retransmissions.

One last aspect of the cross-layer design is the joint design with the lowest layer in the protocol stack, the physical layer. In a packet transmitted through the network, the correlated compressed data can have weaker protection than the associated header, thus saving some overhead, because the available side information at the decoding node can make up for this weaker protection.

Related Topics

So far we have motivated DSC with applications to sensor networks. Research on this application area [5], [12], [41], [43] has just begun. Significant work remains to be done to fully harness the potential of these next-generation sensor networks, as discussed in the last section. However, DSC is also related to a number of different source and channel coding problems in networks.

First of all, due to the duality [22], [45] between DSC (source coding with side information) and channel coding with side information [46], [47] (data hiding/digital watermarking), there are many application areas related to DSC (e.g., coding for the multiple access channel [3] and the MIMO/broadcast channels [21], [48]). Table 2 summarizes different dual problems in source and channel coding, where the encoder for the source coding problem is the functional dual of the decoder for the corresponding channel coding problem and vice versa. We can see that DSC only represents the starting point of a class of related research problems.

Furthermore, a recent theoretical result established multiple description coding [49] as a special case of DSC with *colocated* sources, with multiple descriptions easily generated by embedded coding [36] plus unequal error protection. Under this context, iterative decoding approaches in DSC immediately suggest that the same iterative (turbo) technique can be employed for decoding multiple descriptions. Yet another new application example of DSC is layered coding for video streaming, where the error drifting problem in standard MPEG-4 FGS coding can be potentially eliminated with distributed video coding [17], [40], [43], [50].

Last but not least, DSC principles can be applied in reliable communications with uncoded side information and systematic source-channel coding [26]. The latter includes embedding digital signals into analog (e.g., TV) channels and communication over channels with unknown SNRs. Applying DSC algorithms to these real-world problems will prove to be extremely exciting and yield the most fruitful results.

Acknowledgments

This work was supported by NSF CAREER Grant MIP-00-96070, NSF Grant CCR-01-04834, ONR YIP Grant N00014-01-1-0531, and ARO YIP Grant DAAD19-00-1-0509.

Zixiang Xiong received the Ph.D. degree in electrical engineering in 1996 from the University of Illinois, Urbana-Champaign. From 1997 to 1999, he was with the University of Hawaii. Since 1999, he has been with the Department of Electrical Engineering at Texas A&M University as an associate professor. His current research interests are coding for multiterminal communication networks, joint source-channel coding and genomic signal processing. He received an NSF Career Award in 1999, an ARO Young Investigator Award in 2000, and an ONR Young Investigator Award in 2001. He also received faculty fellow awards in 2001, 2002 and 2003 from Texas A&M University. He is currently an associate editor for *IEEE Transactions on Circuits and Systems for Video Technology*, *IEEE Transactions on Signal Processing*, and *IEEE Transactions on Image Processing*.

Angelos D. Liveris received the Diploma in electrical and computer engineering from the National Technical University of Athens, Greece, in 1998. Since 1999, he has been working toward the Ph.D. degree in electrical engineering at Texas A&M University, College Station. His research interests include network information theory, distributed source and channel coding, detection, and wireless communications.

Samuel Cheng received the B.Eng. degree in electrical and electronic engineering from the University of Hong Kong and the M.Phil. degree in physics and the M.S. degree in electrical engineering from the Hong Kong University of Science and Technology and the University of Hawaii, Honolulu, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at Texas A&M University, College Station.

References

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

[2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[3] T. Cover and J. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[4] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1977.

[5] S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Mag.*, vol. 19, pp. 51–60, Mar. 2002.

[6] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inform. Theory*, vol. 49, pp. 626–643, Mar. 2003.

[7] S. Bohacek, K. Shah, G. Arce, and M. Davis, "Signal processing challenges in active queue management," *IEEE Signal Proc. Mag.*, vol. 21, Sept. 2004.

[8] R. Berry and E. Yeh, "Cross-layer wireless resource allocation: Fundamental performance limits," *IEEE Signal Proc. Mag.*, vol. 21, Sept. 2004.

[9] R. Rao, C. Comaniciu, T. Lakshman, and H.V. Poor, "Integrated call admission control in DS-CDMA wireless multimedia networks," *IEEE Signal Proc. Mag.*, vol. 21, Sept. 2004.

[10] L. Tong, V. Naware, and P. Venkitasubramaniam, "Signal processing in random access: A cross-layer perspective, *IEEE Signal Proc. Mag.*, vol. 21, Sept. 2004.

[11] A. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, vol. 20, no. 1, pp. 2–10, 1974.

[12] S. Pradhan and K. Ramchandran, "Distributed source coding: Symmetric rates and applications to sensor networks," in *Proc. DCC'00*, Snowbird, UT, 2000, pp. 363–372.

[13] Y. Zhao and J. Garcia-Frias, "Data compression of correlated non-binary sources using punctured turbo codes," in *Proc. DCC'02*, Snowbird, UT, 2002, pp. 242–251.

[14] V. Stanković, A. Liveris, Z. Xiong, and C. Georghiades, "Design of Slepian-Wolf codes by channel code partitioning," in *Proc. DCC'04*, Snowbird, UT, 2004, pp. 302–311.

[15] "Special Issue on Codes on Graphs and Iterative Algorithms," *IEEE Trans. Inform. Theory*, vol. 47, Feb. 2001, pp. 493–849.

[16] P. Mitran and J. Bajcsy, "Coding for the Wyner-Ziv problem with turbo-like codes," in *Proc. ISIT'02*, Lausanne, Switzerland, 2002, p. 91.

[17] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, 2004 (to appear).

[18] A. Liveris, Z. Xiong, and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 440–442, 2002.

[19] C. Lan, A. Liveris, K. Narayanan, Z. Xiong, and C. Georghiades, "Slepian-Wolf coding of multiple *M*-ary sources using LDPC codes," *Proc. DCC'04*, Snowbird, UT, 2004, p. 549.

[20] J. Garcia-Frias and W. Zhong, "LDPC codes for compression of multiterminal sources with hidden Markov correlation," *IEEE Commun. Lett.*, vol. 7, no. 3, pp. 115–117, 2003.

[21] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.

[22] S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1181–1203, 2003.

[23] M.V. Eyuboglu and G.D. Forney, "Lattice and trellis quantization with lattice- and trellis-bounded codebooks—High-rate theory for memoryless sources," *IEEE Trans. Inform. Theory*, vol. 39, no. 1, pp. 46–59, 1993.

[24] D. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards, and Practice.* Norwell, MA: Kluwer, 2001.

[25] S. Servetto, "Lattice quantization with side information," in *Proc. DCC'00*, Snowbird, UT, 2000, pp. 510–519.

[26] S. Shamai, S. Verdu, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 564–579, 1998.

[27] A. Liveris, Z. Xiong, and C. Georghiades, "Nested turbo codes for the binary Wyner-Ziv problem," in *Proc. ICIP'03*, Barcelona, Spain, 2003, pp. 601–604.

[28] J. Conway and N. Sloane, *Sphere Packings, Lattices and Groups.* New York: Springer Verlag, 1998.

[29] V. Tarokh, A. Vardy, and K. Zeger, "Universal bounds on the performance of lattice codes," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 670–681, 1999.

[30] Z. Liu, S. Cheng, A. Liveris, and Z. Xiong, "Slepian-Wolf coded nested quantization (SWC-NQ) for Wyner-Ziv coding: Performance analysis and code design," in *Proc. DCC'04*, Snowbird, UT, 2004, pp. 322–331.

[31] J. Chou, S. Pradhan, and K. Ramchandran, "Turbo and trellis-based constructions for source coding with side information," in *Proc. DCC'03*, Snowbird, UT, 2003, pp. 33–42.

[32] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, 2003, pp. 825–829.

[33] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed, partial, and conditional Karhunen-Loeve transforms" in *Proc. DCC'03*, Snowbird, UT, 2003, pp. 283–292.

[34] G. Caire, S. Shamai, and S. Verdu, "Lossless data compression with low-density parity-check codes," in *Multiantenna Channels: Capacity, Coding and Signal Processing*, G. Foschini and S. Verdu, Eds. Providence, RI: American Math. Soc., 2003.

[35] J. Garcia-Frias and Y. Zhao: "Compression of binary memoryless sources using punctured turbo codes," *IEEE Commun. Letters*, vol. 6, pp. 394–396, Sept. 2002.

[36] J.M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3463, Dec. 1993.

[37] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Inform. Theory*, to be published.

[38] W. Equitz and T. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–274, Mar. 1991.

[39] S. Cheng and Z. Xiong, "Successive refinement for the Wyner-Ziv problem and layered code design," in *Proc. DCC'04*, Snowbird, UT, 2004, p. 531.

[40] Q. Xu and Z. Xiong, "Layered Wyner-Ziv video coding," in *Proc. VCIP'04*, San Jose, CA, 2004, pp. 83–91.

[41] J. Chou, D. Petrović, and K. Ramchandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," in *Proc. INFOCOM 2003*, San Francisco, CA, 2003, pp. 1054–1062.

[42] Y. Yang, V. Stanković, Z. Xiong, and W. Zhao, "Asymmetric code design for remote multiterminal source coding," in *Proc. DCC'04*, Snowbird, UT, 2004, p. 572.

[43] R. Puri and K. Ramchandran, "PRISM: A new 'reversed' multimedia coding paradigm," in *Proc. ICIP'03*, Barcelona, Spain, 2003, pp. 617–620.

[44] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," *IEEE Trans. Inform. Theory*, vol. 42, pp. 887–902, May 1996.

[45] R. Barron, B. Chen, and G. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1159–1180, May 2003.

[46] S.I. Gelfand and M.S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[47] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[48] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[49] V. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Mag.*, vol. 18, no. 5, pp. 74–93, 2001.

[50] A. Sehgal, A. Jagmohan, and N. Ahuja, "Wyner-Ziv coding of video: An error-resilient compression framework," *IEEE Trans. Multimedia*, vol. 6, pp. 249–258, Apr. 2004