Multiterminal Communication Systems

© DIGITAL VISION

[Martin J. Wainwright]

Sparse Graph Codes for Side Information and Binning

[Sparse graph codes and message-passing algorithms for binning and coding with side information]

The pioneering work of Shannon provides fundamental bounds on the rate limitations of communicating information reliably over noisy channels (the channel coding problem), as well as the compressibility of data subject to distortion constraints (the lossy source coding problem). However, Shannon's theory is nonconstructive in that it only establishes the existence of coding schemes that can achieve the fundamental bounds but provides neither concrete codes nor computationally efficient algorithms. In the case of channel coding, the past two decades have witnessed dramatic advances in practical constructions and algorithms, including the invention of turbo codes [3] and the surge of interest in low-density parity check (LDPC) codes [12], [17], [34]. Both these classes of codes are based on sparse graphs and yield excellent error-correction performance when decoded using computationally efficient methods such as the message-passing sum-product algorithm [17]. Moreover, their performance limits are well characterized, at least in the asymptotic limit of large block lengths, via the density evolution method [21], [34].

Despite these impressive advances in the pure channel coding problem, for many other communication and signal processing problems, the current gap between theory and practice remains substantial. Various applications under

IEEE SIGNAL PROCESSING MAGAZINE [**47**] SEPTEMBER 2007

development—among them, distributed video coding [16], [33], MIMO communication systems [46], as well as digital watermarking and data hiding [11], [29]—require efficient methods for various types of coding with side information. Classical information-theoretic solutions to coding with side information are based on binning, a general procedure that also underlies various other communication problems. Given the wealth of applications, there is great deal of contemporary interest in developing practical codes and computationally efficient algorithms. Although a large number of practical approaches have been proposed, current theoretical understanding of their performance remains incomplete. The common thread tying together these approaches is the framework of sparse graphical codes and an associated set of message-passing algorithms for decoding and encoding [17], [20]. These iterative algorithms, in which neighboring nodes in the graph defining the code exchange information, may either be of exact nature, such as the Viterbi algorithm for trellis codes, or an approximate nature, such as the sum-product algorithm for LDPC codes.

> **THERE IS NOW GREAT INTEREST AND CONSIDERABLE LITERATURE ON PRACTICAL CODES AND ALGORITHMS FOR PERFORMING BINNING.**

The purpose of this article is to provide an overview and introduction to this rapidly developing area of research. We begin in the following section with a brief introduction to the basics of codes defined by factor graphs as well as iterative message-passing algorithms for solving decoding and encoding problems. Then we describe the problems of source coding with side information (SCSI) and channel coding with side information (CCSI) along with their applications to distributed compression, MIMO communication, and information embedding. In addition, we discuss the more general notion of binning, which underlies the classical information-theoretic solutions to these and related communication problems. Although conceptually appealing, naive random binning is practically infeasible, which motivates the study of structured schemes for constructing nested source-channel codes. Accordingly, the core material in the section following is devoted to an overview of structured methods for lossy compression and binning, with a particular emphasis on the sparse graphical codes studied in our recent work. Next, we discuss some of the algorithmic challenges associated with sparse graphical codes. In particular, many problems involving lossy compression and binning appear to require novel message-passing algorithms, related to but distinct from the usual sum-product updates [17], [20]. Given the space constraints of this article, we limit our discussion to simple but intuitively revealing examples, mainly involving binary sources and channels. However, it should be stressed that the basic ideas are more broadly applicable to general sources and channels.

BACKGROUND
This section is devoted to background material that underlies our discussion. We begin with a brief description of binary linear codes and associated channel and source coding problems. We then discuss how coding problems can be represented in terms of sparse graphs and the use of message-passing algorithms for decoding and encoding in such sparse graphical codes.

BASIC CHANNEL AND SOURCE CODING
Any binary linear code $\mathbb{C}$ of block length n and rate $R = 1 - (m/n)$ can be specified as the null space (in modulo two arithmetic) of a given parity check matrix $\mathbf{H} \in \{0, 1\}^{m \times n}$. If the parity check matrix is of full rank (m), then the code $\mathbb{C}$ consists of $2^{n-m} = 2^{nR}$ codewords. In the channel coding problem, the transmitter chooses some codeword $\mathbf{x} \in \mathbb{C}$ and transmits it over a noisy channel, so the receiver observes a realization of a noisy random variable $\mathbf{Y} = \mathbf{y}$. Frequently, the channel is modeled as memoryless, meaning that it can be specified as a factorized distribution $\mathbb{P}(\mathbf{y} \mid \mathbf{x}) = \prod_{i=1}^{n} f_i(x_i ; y_i)$. As a simple example, in the binary symmetric channel (BSC), the channel flips each transmitted bit x_i independently with some probability p. The goal of the receiver is to solve the channel decoding problem: estimate the most likely transmitted codeword, given by $\hat{\mathbf{x}} = \max_{\mathbf{x} \in \mathbb{C}} \mathbb{P}(\mathbf{y}|\mathbf{x})$. This decoding task is typically a difficult problem in combinatorial optimization since, in general, it entails searching over an exponentially large set of possible codewords. The Shannon capacity of a channel specifies an upper bound on the rate R of any code for which transmission can be asymptotically error-free. For example, for the BSC with flip probability p, the capacity is given by $C = 1 - h(p)$, where $h(p) = -p \log_2 p - (1 - p) \log_2(1 - p)$ is the binary entropy function.

The goal of source coding, on the other hand, is to compress a data source subject to some bound on the amount of distortion. To be concrete, one might be interested in compressing a symmetric Bernoulli source, meaning a bit string $\mathbf{Y} \in \{0, 1\}^n$, with each bit Y_i equally likely to be 0 or 1. Here a natural distortion metric would be the Hamming distance $d(\hat{\mathbf{y}}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|$ between a fixed source sequence $\mathbf{y}$ and the compressed version $\hat{\mathbf{y}}$. Classical rate-distortion theory specifies the optimal tradeoff between the compression rate R, which indexes the number 2^{nR} of possible compressed sequences $\hat{\mathbf{y}}$, and the average distortion $D = \mathbb{E}[d(\hat{\mathbf{Y}}, \mathbf{Y})]$, where the expectation is taken over the random source sequence $\mathbf{Y}$ and its compressed version $\hat{\mathbf{Y}}$. For the symmetric Bernoulli source considered here, this rate-distortion tradeoff is specified by the function $R(D) = 1 - h(D)$, where h is the previously defined binary entropy function. Given a code $\mathbb{C}$, optimal source encoding corresponds to finding the codeword $\hat{\mathbf{y}} \in \mathbb{C}$ that is closest to a given source sequence $\mathbf{y}$—namely, solving the combinatorial optimization problem $\hat{\mathbf{y}} = \arg \min_{\mathbf{x} \in \mathbb{C}} d(\mathbf{x}, \mathbf{y})$.

SPARSE GRAPHICAL CODES
Both the channel decoding and source encoding problems, if viewed naively, require searching over an exponentially large

codebook, since $|\mathbb{C}| = 2^{nR}$ for a code of rate R. Therefore, any practically useful code must have special structure that facilitates decoding and encoding operations. The success of a large subclass of modern codes in use today—including trellis codes, turbo codes, and LDPC codes—is based on the framework of factor graphs and message-passing algorithms [17], [20]. Given a binary linear code $\mathbb{C}$, specified by parity check matrix $\mathbf{H}$, the code structure can be captured by a bipartite graph, in which circular nodes ($\bigcirc$) represent the binary values x_i (or columns of $\mathbf{H}$), and square nodes ($\blacksquare$) represent the parity checks (or rows of $\mathbf{H}$). For instance, Figure 1(a) shows the factor graph for a rate $R = 1/2$ code in parity check form, with $m = 6$ checks acting on $n = 12$ bits. The edges in this graph correspond to 1s in the parity check matrix, and reveal the subset of bits on which each parity check acts. The code illustrated is regular with bit degree three and check degree six. Such low-density constructions, meaning that both the bit degrees and check degrees remain bounded independently of the block length n, are of most practical use since they can be efficiently represented and stored and yield excellent performance under message-passing decoding. In the context of a channel coding problem, the shaded circular nodes at the top of Figure 1(a) represent the observed variables y_i received from the noisy channel.

Trellis codes are another widely used class of codes. In contrast to the classical trellis representation shown at the bottom of Figure 1(b), the top part shows how a trellis code can be represented in terms of a linearly ordered factor graph (i.e., a chain). As we discuss in more detail in the following, the cycle-free structure of this graph is desirable since message-passing algorithms for decoding/encoding are guaranteed to be exact for such graphs [17], [31]. Note that, in contrast to the cycle-free structure of the trellis code in Figure 1(b), the factor graph representing the LDPC code in Figure 1(a) has many cycles. For such graphs with cycles, message-passing algorithms are no longer exact but nonetheless perform well for many graphs, as we discuss in the following.
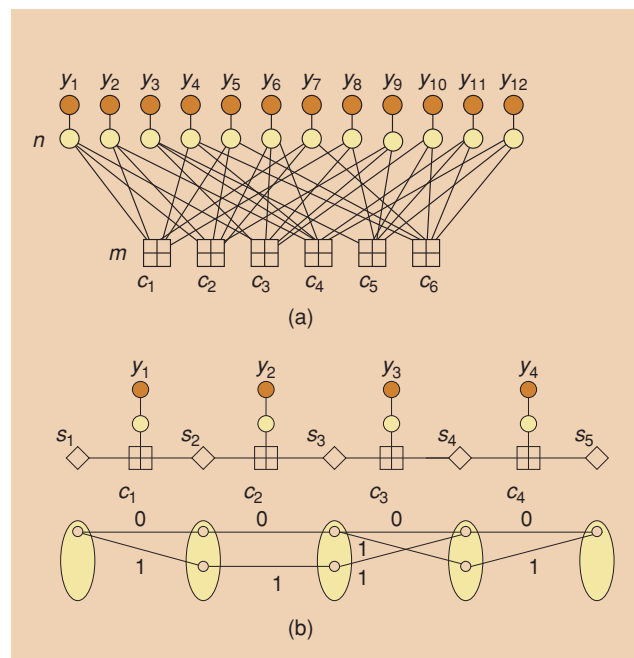
MESSAGE-PASSING ALGORITHMS
The significance of codes based on sparse factor graphs is in facilitating the use of message-passing algorithms for encoding and decoding operations. Here we briefly discuss of the sum-product and max-product (min-sum) algorithms for factor graphs; given space constraints, our treatment is necessarily superficial, but see the tutorial papers [17], [20] for more detail. In the previous section, we described the use of factor graphs for representing the structure of a binary linear code. More broadly, one can think of factor graphs as specifying the factorization of some probability distribution in the form $\mathbb{P}_X(\mathbf{x}) = \frac{1}{Z} \prod_{a \in C} f_a(x_{N(a)})$, where C represents the set of factors a in the decomposition, and the factor f_a is a function of the variables in its factor graph neighborhood $N(a)$. As a concrete example, in the case of binary linear codes discussed previously, each factor f_a cor-

responds to an indicator function for the parity check over a particular subset of bits, so $f_a(x_{N(a)}) = 1$ if $\bigoplus_{i \in N(a)} x_i = 0$, and $f_a(x_{N(a)}) = 0$ otherwise. The constant Z, known as the partition function, serves to normalize the probability distribution to unity. Observations $\mathbf{y}$, which might either be the output of a noisy channel or a realization of some data source, are represented by a conditional distribution $\mathbb{P}_{Y|X}(\mathbf{y} | \mathbf{x}) = \prod_{i=1}^{n} f_i(x_i, y_i)$ and lead via Bayes' rule to the posterior distribution

$$\mathbb{P}_{X|Y}(\mathbf{x} | \mathbf{y}) = \frac{1}{Z'} \prod_{a \in C} f_a(x_{N(a)}) \prod_{i=1}^{n} f_i(x_i\, y_i). \qquad (1)$$

Given this posterior distribution, the following two computational tasks are of interest: 1) computing marginal distributions $\mathbb{P}(x_i | \mathbf{y})$ at each node $i = 1, \ldots, n$ and 2) computing the maximum a posteriori (MAP) assignment $\hat{\mathbf{x}}_{MAP} := \arg \max_{\mathbf{x}} \times \mathbb{P}_{X|Y}(\mathbf{x} | \mathbf{y})$. Note that when the prior distribution $\mathbb{P}_X$ is uniform, as in the standard coding set-up, then the MAP codeword is equivalent to the maximum likelihood (ML) codeword previously defined.

The sum-product and max-product algorithms are efficient approaches, based on a divide-and-conquer strategy, for solving the marginalization and MAP problems, respectively. The updates can be cast as message-passing operations, in which nodes adjacent in the factor graph convey the results of intermediate computations by passing messages (vectors in the case of discrete random variables). In a factor graph, there are two types of messages: variable to

[FIG1] (a) Factor graph representation of a rate $R = 0.5$ LDPC code with bit degree $d_v = 3$ and check degree $d_c = 6$. (b) Top: Factor graph representation of a trellis code as a chain-structured graph. Bottom part: more standard trellis-diagram, illustrating both the chain structure and the pattern of zeros in the transition diagram.

check messages (M_{ia}) and check to variable messages (M_{ai}). In the sum-product algorithm, these messages are updated according to the recursive equations

$$M_{ai}(x_i) \leftarrow \sum_{x_j} f_a(x_{N(a)}) \prod_{j \in N(a)/i} M_{ja}(x_{N(a)}), \quad \text{and}$$

$$M_{ia}(x_{N(a)}) \leftarrow f(x_i; y_i) \prod_{b \in N(i)/a} M_{bi}(x_i). \tag{2}$$

The max-product (or min-sum) algorithm has a very similar form, with the summation in (2) replaced by maximization. For graphs without cycles, also known as trees, the updates converge after a finite number of updates and provide exact solutions [17], [31]. For instance, when applied to the chain-structured factor graph in Figure 1(b) associated with a trellis code, the min-sum algorithm is equivalent to the Viterbi algorithm, and will compute the most likely configuration (or shortest trellis path) in a finite number of steps. However, these algorithms are also frequently applied to graphs with cycles, for which they are no longer exact but rather approximate methods. Nonetheless, for certain classes of graphs with cycles (e.g., those corresponding to LDPC codes with large block lengths), the performance of these approximate message-passing schemes for decoding is excellent, as substantiated by both empirical and theoretical results [34].

> **NOVEL ALGORITHMS ARE REQUIRED FOR SOLVING CODING PROBLEMS OTHER THAN CHANNEL DECODING.**

BEYOND THE BASICS: CODING WITH SIDE INFORMATION AND BINNING

Many problems in communication and signal processing require coding approaches that go beyond channel or source coding alone. Important examples include the problems of source coding with side information (SCSI) at the receiver, and the problem of CCSI at the transmitter. In this section, we begin with a brief description of these problems, and their significance in various applications. We then discuss the more general notion of binning, which is the key ingredient in the information-theoretic solutions to these problems.

LOSSY SOURCE CODING WITH SIDE INFORMATION

It is often desirable to perform joint compression of a collection of dependent data sources. Without any constraints on communication between the sources, one could simply pool all of the data into a single source and apply standard compression methods to this pooled source. However, many application domains—among them sensor networks [43] and distributed video coding [6], [16]—require that compression be performed in a distributed manner, without any communication between the sources. In the information theory literature, one version of lossless distributed compression is the Slepian-Wolf problem [37]. Here we focus on the Wyner-Ziv extension to lossy SCSI[42], in which the goal is to perform lossy compression of a source X up to some distortion constraint, using side information Y available only at the receiver. For instance, in the context of distributed compression in a sensor network, the side information represents the compressed information provided by adjacent sensors.

As a simple but pedagogically useful example, let us suppose that we wish to compress a symmetric Bernoulli source vector $X \in \{0, 1\}^n$. As discussed previously, classical rate distortion theory asserts that the minimum rate required to achieve average distortion D is given by $R(D) = 1 - h(D)$. In the Wyner-Ziv extension [42] of this problem, there is an additional source of side information about the source sequence X—say in the form $Y = X \oplus W$, where $W \sim \text{Ber}(p)$ is observation noise—that is available only at the decoder. In this setting, the minimum achievable rate is essentially specified by the curve $R_{wz} = -h(D * P) - h(D)$, where $D * p = D(1 - p) + (1 - D)p$ is the binary convolution. [Strictly speaking, the Shannon limit is given as the lower convex envelope of this function with the point $(p, 0)$.] In comparison with the standard rate-distortion function $R(D) = 1 - h(D)$, we see that the value of the side information is measured by the gap between $h(D * p)$ and 1, with equality for $p = 1/2$ corresponding to useless side information.

CHANNEL CODING WITH SIDE INFORMATION

The problem of CCSI takes a symmetric form with respect the encoder and decoder roles: here the side information is available only at the encoder, and the goal is to perform (almost) as well as if the side information were available at both the encoder and decoder. One variant is known as the Gelfand-Pinsker problem [15]. Given an input vector U, the channel adds to it some fixed host signal S and possibly adds a noise term as well. Even though the encoder knows the host signal S, it cannot simply cancel this known interference (i.e., tidy up the dirty paper [9]) due to some form of channel constraint on the average transmitted power. This class of models provides an appropriate abstraction for various applications, including the problem of coding on defective memory cells, problems in digital watermarking and steganography, and MIMO communication. The papers [2], [5], [29], and [46] and references therein contain further background on the problem and these applications.

As an illustrative example, let us formalize the special case of coding on binary dirty paper, also known as binary information embedding. More specifically, suppose that for a given binary input $U \in \{0, 1\}^n$, the channel output takes the form $Y = U \oplus S \oplus Z$, where $S \in \{0, 1\}^n$ is a host signal (not under control of the encoder), and $Z \sim \text{Ber}(p)$ is a channel noise vector. Typically, the host signal S is modeled as a realization from some stochastic model (e.g., a symmetric Bernoulli source). The encoder is free to choose the input vector U, subject to the average channel constraint $\mathbb{E}[\|U\|_1] \leq \delta n$, so as to

maximize the rate of information transfer. Conceptually, it is useful to write $\mathbf{U} \equiv \mathbf{U}_\mathbf{M}$, where $\mathbf{M}$ is the underlying message that is embedded, and note that the decoder's goal is to recover this embedded message $\mathbf{M}$ from the corrupted observation $\mathbf{Y}$. The capacity in this set-up is essentially determined by the curve $R_{\mathrm{IE}}(\delta, p) = h(\delta) - h(p)$. [To be precise [2], the Shannon limit is specified by time-sharing between the curve R_{IE} and the point $(0, 0)$.] As one special case, note that when $\delta = 1/2$, the channel input constraint $\mathbb{E}[\|\mathbf{U}\|_1] \leq n/2$ effectively plays no role, and the information embedding capacity reduces to the classical capacity $C = 1 - h(p)$.

BINNING SOLUTIONS
The term *binning* refers to a standard information-theoretic approach to proving capacity results for coding with side information, as well as related multiterminal problems [9], [15], [33], [42], [46]. To illustrate the basic idea, let us consider the binary information embedding problem described previously but with no channel noise ($p = 0$). In this special case, the problem is one of constrained channel coding—how to transmit information over a channel that maps a binary input input $\mathbf{U} \in \{0, 1\}^n$ to a binary output $\mathbf{Y} = \mathbf{U} \oplus \mathbf{S}$ subject to the average input constraint $\mathbb{E}[\|\mathbf{U}\|_1] \leq \delta n$. Note that the curve R_{IE} reduces to $h(\delta)$ when $p = 0$. If the host signal were also known at the decoder, , then it would be straightforward to achieve the rate $h(\delta)$ by the following procedure. First, the encoder chooses a binary sequence $\mathbf{M} \in \{0, 1\}^n$ with δn ones. The number of such sequences is $\binom{n}{\delta n}$, which by Stirling's approximation to factorial coefficients scales as $2^{nh(\delta)}$. The encoder then simply transmits $\mathbf{U} = \mathbf{M}$, which satisfies the channel input constraint. In the second step, the decoder receives $\mathbf{Y} = \mathbf{M} \oplus \mathbf{S}$, and given knowledge of the host signal $\mathbf{S}$, it recovers the message $\mathbf{M}$ by a simple XOR operation. The overall transmission rate for this scheme is $(1/n) \log \binom{n}{\delta n} \approx h\delta$, and it is the best that can be achieved.

What approach should be taken when the host signal $\mathbf{S}$ is not known at the decoder? Simply attempting to cancel out the host signal at the encoder (by transmitting $\mathbf{U} = \mathbf{M} \oplus \mathbf{S}$) is not an option, since the sequence $\mathbf{M} \oplus \mathbf{S}$ fails to satisfy the channel input constraint. Although this naive approach fails, the remarkable phenomenon is that even when the decoder does not know the host signal, the optimal rate $h(\delta)$ can be achieved by a binning strategy. For this problem, the binning strategy is to partition the set of all 2^n binary sequences into a total of $B := \binom{n}{\delta n} \approx 2^{nh(\delta)}$ bins. Figure 2(a) provides a toy illustration of this binning procedure, where the set of all circular nodes (representing all 2^n possible binary sequences) are partitioned into $B = 3$ bins.
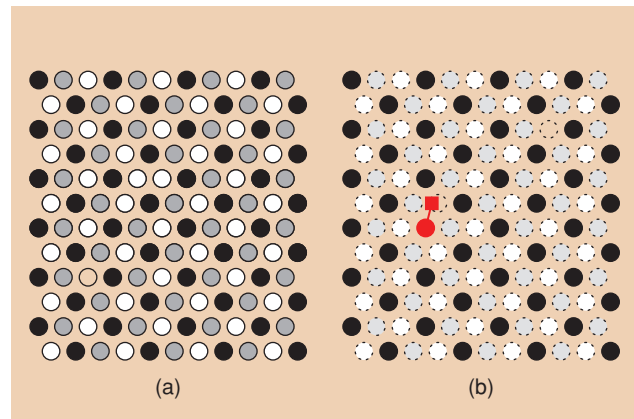
Both the encoder and decoder are given knowledge of the partition, and the message vectors $\mathbf{M}$ are used to index the bins. The encoder proceeds as follows: in order to transmit a particular fixed message $\mathbf{M}$, it restricts itself to the bin indexed by $\mathbf{M}$ [see Figure 2(b)], and searches for a codeword $\hat{\mathbf{S}}_\mathbf{M}$ that differs from the host signal $\mathbf{S}$ in, at most, δn positions. Each randomly constructed bin contains approximately

$2^n / B \approx 2^{n[1 - h\delta]}$ elements and so can be viewed as a source code with rate $R \approx 1 - h\delta$. In fact, the scheme requires that each bin on its own acts as a good source code for quantizing up to average distortion D. Under this assumption, standard rate-distortion theory guarantees that it will be possible to quantize the host $\mathbf{S}$ such that the average distortion satisfies $\mathbb{E}[d(\hat{\mathbf{S}}_\mathbf{M}, \mathbf{S})] \leq \delta n$. The encoder then transmits the quantization error $\mathbf{E}_\mathbf{M} := \mathbf{S} \oplus \hat{\mathbf{S}}_\mathbf{M}$, which satisfies the channel input constraint by construction. Finally, the encoder receives $\mathbf{Y} = \mathbf{E}_\mathbf{M} \oplus \mathbf{S} = \hat{\mathbf{S}}_\mathbf{M}$, on which basis it can identify the message $\mathbf{M}$ uniquely specified by the codeword $\hat{\mathbf{S}}_\mathbf{M}$.

Note this binning scheme can be viewed as a collection of $2^{nh(\delta)}$ source codes for quantizing the host signal, all nested within the full space of 2^n binary strings. In order to solve the more general information embedding problem with channel noise $p > 0$, it is necessary to partition a good channel code (with $\asymp 2^{n[1 - H(p)]}$ codewords) into a collection of good source codes. Conversely, for the dual problem of source coding with side information, one requires a good source code that can be partitioned into good channel codes [42], [46]. However, the preceding discussion has been rather idealized, in that it is based on completely unstructured codes and bins, and provides no concrete construction of such nested codes. Accordingly, the following section is devoted to structured and practical approaches based on sparse graphical codes for constructing such nested source-channel codes.

PRACTICAL CONSTRUCTIONS AND ALGORITHMS
We now turn to an overview of practical constructions and algorithms, beginning with the simpler problem of lossy compression, before moving onto practical codes for binning and coding with side information.

[FIG2] Illustration of binning for binary information embedding. (a) The original set of binary strings is partitioned into a set of B bins ($B = 3$ in this diagram, with white, grey and black nodes respectively). Each bin on its own is required to be a good source code, meaning that it can be used to quantize the host signal $\mathbf{S}$ up to average Hamming distortion δn. (b) The message $\mathbf{M}$ specifies a particular bin (in this case, the bin with black nodes); the message is transmitted implicitly, by using the chosen bin as a source code to quantize the host signal $\mathbf{S}$ (square) by the nearest codeword $\hat{\mathbf{S}}_\mathbf{M}$, and transmitting the quantization error $\mathbf{E}_\mathbf{M} := \mathbf{S} \oplus \hat{\mathbf{S}}_\mathbf{M}$.

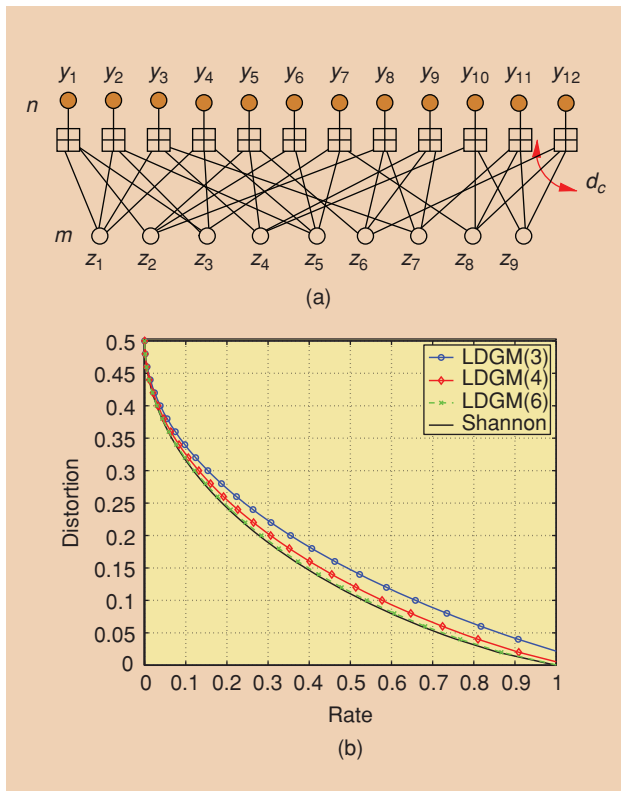## PRACTICAL CODES FOR LOSSY COMPRESSION

For problems of lossless compression, the use of practical codes and algorithms is quite well-understood, both for standard lossless compression [4], [14], [33] and for Slepian-Wolf type extensions [13], [33], [35]. Many lossless problems can be converted into equivalent channel coding problems (e.g., via syndrome forming procedures), which allows known constructions and methods from channel coding to be leveraged. For lossy compression, in contrast, the current gap between theory and practice is more substantial.

The class of trellis codes provides one widely-used method for lossy compression [23], [39]. As illustrated in Figure 1(b), a trellis code can be represented as a chain-structured factor graph, which allows encoding/decoding problems to be solved exactly via the sum-product or max-product algorithms. The complexity of these message-passing algorithms is linear in the trellis length but exponential in the trellis constraint length. When used for lossy compression, the shaded nodes at the top of Figure 1(b) represent the incoming stream of source symbols ($y_1, y_2 \ldots$, etc.). Note that these source symbols align with the edges in the trellis diagram shown in Figure 1(b). The distortion metric is used to assign weights to the edges of trellis diagram, and the minimum-distortion

[FIG3] (a) Factor graph representation of a low-density generator matrix (LDGM) code with $n = 12$, $m = 9$ and rate $R = 0.75$. An LDGM code has bit and check degrees bounded independently of blocklength; for this example, $d_c = 3$ and $d_v = 4$. (b) Plots of the effective rate-distortion function of several finite degree LDGM ensembles in comparison to the Shannon bound, obtained using the second moment bound (3).

encoding, or equivalently the shortest path through the trellis, with edge lengths corresponding to distortion, can then be computed exactly by running the min-sum algorithm (see [23] for details). Various researchers have demonstrated the practical utility of trellis codes for single-source and distributed compression [6], [41] as well as for information embedding problems [5], [10]. In terms of theoretical guarantees, trellis-based codes can saturate the rate-distortion bound if the trellis constraint length is taken to infinity [39]. This constraint length controls the cardinality of the hidden random variables, represented by diamonds in Figure 1(b); in particular, the cardinality of these random variables, and hence the computational complexity of the Viterbi algorithm, grows exponentially in the constraint length. Consequently, provably achieving the rate-distortion bound using trellis codes requires increasing computational complexity [39], even with optimal encoding and decoding.

Given the limitations of trellis-based approaches, a parallel line of recent work has explored the use of low-density generator matrix (LDGM) codes for lossy compression [7], [27], [30], [40]. The class of LDGM codes is closely related to but distinct from the class of LDPC codes. As opposed to a parity check representation, an LDGM code of rate $R = m/n$ is described by a sparse generator matrix $G \in \{0, 1\}^{n \times m}$. For instance, Figure 3(a) shows a LDGM code with $n = 12$ code bits arranged across the top row, and $m = 9$ information bits along the bottom row, corresponding to an overall rate $R = 0.75$. When used for lossy compression, say of a binary sequence $y \in \{0, 1\}^n$, source decoding using an LDGM code is very simple. For a given information sequence $\hat{z} \in \{0, 1\}^m$, the associated source reconstruction $\hat{y} \in \{0, 1\}^n$ is obtained by the matrix-vector product $\hat{y} = G\hat{z}$, performed in modulo two arithmetic. On the other hand, the source encoding step is, in general, nontrivial, since it requires solving the optimization problem $\hat{z} = \arg\min_{z \in \{0, 1\}^m} d(Gz, y)$ for the given distortion metric d.

Focusing on an idealized compression problem known as binary erasure quantization, Martinian and Yedidia [27] showed that LDGM codes combined with modified message-passing can saturate the associated rate-distortion bound. This approach was subsequently extended to a capacity-achieving scheme for the deterministic broadcast channel [8]. Another line of research, using techniques from both statistical physics [7], [30] and probabilistic combinatorics [25], [26], has focused on the fundamental performance limits of sparse LDGM codes for compression of binary symmetric Hamming sources. To provide a flavor for such results, Figure 3(b) plots some rigorous bounds on the effective rate-distortion performance of LDGM code ensembles formed by having each of the n checks choose at random a set of d_c information bits. [The code in Figure 3(a) corresponds to the case $d_c = 3$.] For a given $R \in [0, 1]$, the curve is an upper bound on the average Hamming distortion D incurred by the given code famil, when encoding and decoding are performed optimally. Figure 3(b) shows curves for degrees $d_c \in \{3, 4, 6\}$; note how these effective rate-distortion function approaches the Shannon limit as the degree d_c increases.

The curves in Figure 3(b) are derived using the first- and second-moment methods [1], as applied to the random variable $N \equiv N(\mathbf{Y})$ that simply counts the number of codewords that achieve distortion D for a given source sequence $\mathbf{Y}$. Note that the code achieves distortion D for source sequence $\mathbf{Y}$ if and only if $\{N(\mathbf{Y}) > 0\}$. To study the probability of this event, we make note of the upper and lower bounds

$$\frac{(\mathbb{E}[N])^2}{\mathbb{E}[N^2]} \overset{(a)}{\leq} \mathbb{P}[N > 0] \overset{(b)}{\leq} \mathbb{E}[N]. \qquad (3)$$

The first-moment upper bound (a) is simply a form of Markov's inequality, whereas the second-moment lower bound (b) follows by applying the Cauchy-Schwarz inequality. (In particular, we have $\mathbb{E}[N] = \mathbb{E}[N \mathbb{I}(N > 0)] \leq \sqrt{\mathbb{E}[N^2] \mathbb{P}[N > 0]}$.) For a symmetric Bernoulli source, it is straightforward to show that for any code of rate R, regardless of its structure, the first moment $\mathbb{E}[N]$ scales as $2^{n(R-1+h(D))}$. Consequently, by the first moment upper bound (3)(a), we recover Shannon's rate-distortion bound for a symmetric Bernoulli source—namely, that achieving distortion D is impossible unless $R > 1 - h(D)$.

Unfortunately however, the first moment $\mathbb{E}[N]$ need not be representative of typical behavior of the random variable N, nd hence overall distortion performance of the code. As a simple illustration, consider an imaginary code consisting of 2^{nR} copies of the all-zeroes codeword. Even for this nonsensical code, as long as $R > 1 - h(D)$, the expected number of distortion D optimal codewords grows exponentially. Indeed, if a fair coin is tossed n times, the number of heads will almost always be larger than Dn, in which case $N = 0$. However, our imaginary code also exhibits a jackpot effect: once in a while (more precisely, with probability $\approx 2^{-n[1-h(D)]}$), the coin flip will yield less than Dn heads, in which case the random variable N takes on the enormous value 2^{nR}. Overall, the first moment $\mathbb{E}[N]$ grows exponentially as long as $R > 1 - h(D)$; however, this exponential growth is entirely misleading because the average distortion incurred by using this code will be ≈ 0.5 for any rate.

With this cautionary example in mind, the second moment lower bound (3)(b) allows us to assess the representativeness of the first moment by comparing its squared value to the second moment $\mathbb{E}[N^2]$. Intuitively, there are two factors that control the behavior of this second moment, and hence the rate-distortion performance of the code. The first factor is the overlap probability: for each $\omega \in (0, 1]$, how likely is it that two codewords at Hamming distance ωn both lie within Hamming distance Dn of the same source sequence? This overlap probability typically decays exponentially with the distance. The second factor is the weight enumerator, or the number of codewords at a given distance ωn, which typically grows exponentially in the distance. The behavior of the second moment is governed by this balancing act between the overlap probability and the weight enumerator, and the second moment bound (3)(b) yields the curves shown in Figure 3(b).

BINNING AND COMPOUND CODES WITH NESTED STRUCTURE

As discussed previously, information-theoretic arguments show that the fundamental limits for both source coding and channel coding with side-information (the Wyner-Ziv and Gelfand-Pinsker problems, respectively) are achieved by binning schemes. However, the typical methods proposed in such information-theoretic arguments are impractical, involving unstructured codes. Accordingly, there is now great interest and considerable literature on practical codes and algorithms for performing binning. Several researchers have studied binning schemes for SCSI based on trellis codes [33], [41], combinations of trellis and turbo codes [6], [19], and trellis codes with LDPC codes [44]. Many of these approaches yield good practical performance, close to the information-theoretic bounds in certain regimes. As discussed earlier, however, provably achieving information-theoretic limits with trellis constructions requires taking the constraint length to infinity [39], which incurs exponential decoding complexity. Other work has studied the use of lattice codes for SCSI problems [46] or lattice codes in conjunction with LDPC codes [18]. However, the complexity of lattice decoding also grows exponentially in dimension, in the absence of efficient approximate algorithms. A parallel line of research has focused on practical constructions for CCSI, also known as the Gelfand-Pinsker or dirty paper problem, including schemes based on combinations of trellis coding in conjunction with channel coding methods, such as LDPC or turbo-like codes [5], [10], [38]. Using various types of iterative message-passing, it has been shown empirically that such constructions can lead to excellent practical performance.

Here we describe some of our recent theoretical work [24]–[26] on low-density codes that have a nested structure naturally suited to binning procedures. The basic construction, shown in Figure 4(a), is quite simple, consisting of an LDGM code (top section of figure) compounded with an LDPC code (bottom part of figure). Compound constructions of this nature have been studied in past work, notably on raptor codes [36] and related punctured codes [32], but with exclusive focus on channel coding problems, as opposed to lossy compression and binning that is of interest here. The code shown in Figure 4(a) has block length n, and a rate $R = (m - k_1)/n$. Any codeword $\mathbf{x} \in \{0, 1\}^n$ at the top layer can be generated as $\mathbf{x} = \mathbf{Gz}$, where $\mathbf{z} \in \{0, 1\}^m$ is a string of information bits in the middle layer. The information bits, in turn, are required to satisfy a set of k_1 parity checks (bottom layer), via the relation $\mathbf{H}_1 \mathbf{z} = 0$. Figure 4(b) shows a closely related code $\mathbb{C}'$ with rate $R' = (m - k_1 - k_2)/n < R$, i.e., in fact, a subcode of the original code. This subcode is formed by requiring that the information bits $\mathbf{z}$ that generate some codeword $\mathbf{x} \in \mathbb{C}$ satisfy an additional set of k_2 parity check constraints ($\mathbf{H}_2 \mathbf{z} = \mathbf{u}$)—concretely, for the subcode shown, $k_2 = 2$ and $\mathbf{u} = [1 \ 0]$. Varying the choice of the parity check vector $\mathbf{u} \in \{0, 1\}^{k_2}$ yields a set of 2^{k_2} such subcodes or bins, and the union of all these subcodes is equivalent to the original code $\mathbb{C}$. In this way, the compound

LDGM/LDPC construction leads to sparse graphical codes that are naturally nested: i.e., the original code $\mathbb{C}$ is partitioned into a set of 2^{k_2} subcodes, as is required in binning procedures [compare to Figure 2(a) and (b)]. Moreover, as we discuss previously, there always exist finite choices of the check and node degrees for which codes from the associated compound family are Shannon optimal for both source and channel coding.

PRACTICAL BINNING FOR DIRTY PAPER CODING

To cement this connection, let us now illustrate how the compound construction in Figure 4 can be used to implement the binning steps needed for dirty paper coding, again focusing on the binary case and rates on the curve $R_{IE}(\delta, p) = h(\delta) - h(p)$ for simplicity (see "Channel Coding with Side Information"). We begin by specifying a good channel code and then showing how it can be

> COMMUNICATION PROBLEMS INVOLVING QUANTIZATION AND BINNING OFTEN PRESENT CHALLENGES NOT PRESENT IN ORDINARY CHANNEL CODING OR LOSSLESS SOURCE CODING PROBLEMS.

partitioned into a family of good source codes. Starting with the compound code $\mathbb{C}$ shown in Figure 4(a), assume that the parameters n, m and k_1 are chosen such that its rate satisfies $R_{CHA} = (m - k_1)/n = 1 - h(p) - (\epsilon/2)$ for some $\epsilon > 0$. Given the Shannon optimality of the code family, this rate choice will guarantee that $\mathbb{C}$ is a channel code that can yield asymptotically reliable communication on any BSC with parameter p.

ENCODING

In the encoding step, we use the subcode of $\mathbb{C}$ shown in Figure 4(b) to embed a message $\mathbf{M} \in \{0, 1\}^{k_2}$ through our choice of the coset (i.e., we enforce the add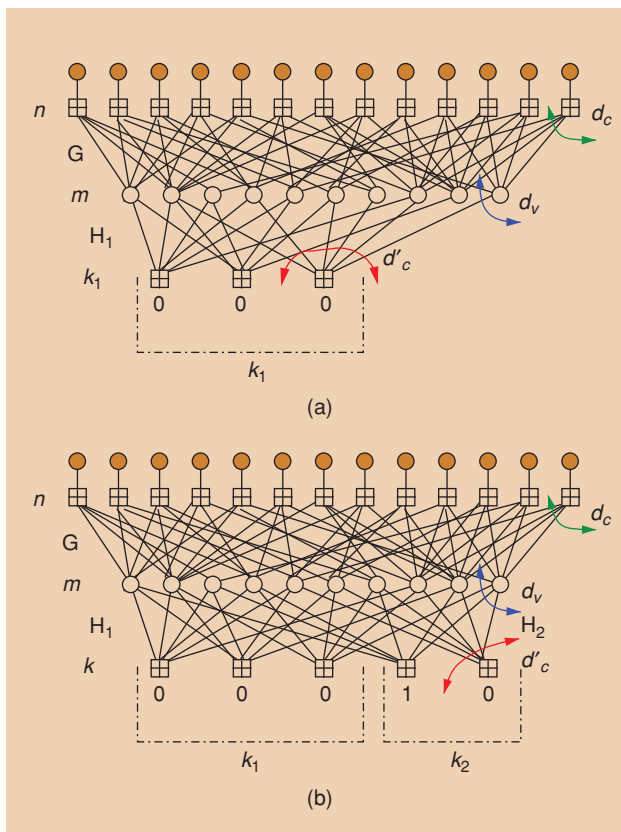itional parity check constraints $\mathbf{H}_2\mathbf{z} = \mathbf{M}$). Moreover, let us assume that k_2 is chosen such that the rate $R_{SOU} = (m - k_1 - k_2)/n = 1 - h(\delta) + (\epsilon/2)$. Given this subcode $\mathbb{C}(\mathbf{M})$, we now use it to quantize the host signal $\mathbf{S} \in \{0, 1\}^n$, thereby yielding a quantized version $\hat{\mathbf{S}}_{\mathbf{M}} \in \{0, 1\}^n$. By construction, the quantized sequence is a codeword of $\mathbb{C}(\mathbf{M})$, meaning that there exists an information sequence $\mathbf{z}_{\mathbf{M}} \in \{0, 1\}^m$ with $\hat{\mathbf{S}}_{\mathbf{M}} = \mathbf{G}\,\mathbf{z}_{\mathbf{M}}$ and moreover $\mathbf{H}_1\,\mathbf{z}_{\mathbf{M}} = \mathbf{0}$ and $\mathbf{H}_2\,\mathbf{z}_{\mathbf{M}} = \mathbf{M}$. This latter parity check constraint is the key, in that it corresponds to the implicit embedding of the message $\mathbf{M}$ into the quantized sequence $\hat{\mathbf{S}}_{\mathbf{M}}$. Moreover, since $\mathbb{C}(\mathbf{M})$ is a good source code and its rate R_{SOU} is sufficiently large, the quantization error $\mathbf{E} = \mathbf{S} \oplus \hat{\mathbf{S}}_{\mathbf{M}}$ has average Hamming weight less than δp, so that it can be transmitted over the channel without violating the input constraint.

DECODING

The decoder receives the sequence $\mathbf{y} = \mathbf{E} \oplus \mathbf{S} \oplus \mathbf{W} = \hat{\mathbf{S}}_{\mathbf{M}} \oplus \mathbf{W}$, where $\mathbf{W}$ is an independent identically distributed (i.i.d.) Bernoulli (p) noise sequence, and its goal is to first recover $\hat{\mathbf{S}}_{\mathbf{M}}$, from which it can then recover the unknown message $\mathbf{M} \in \{0, 1\}^{k_2}$. Since it does not know this message, the decoder operates over the original code $\mathbf{C}$, consisting of the union of all possible cosets, and its task of recovering $\hat{\mathbf{S}}_{\mathbf{M}}$ is equivalent to performing channel decoding over a BSC with flip probability p. Therefore, our earlier choice of rate R_{CHA} guarantees that the decoder will be able to recover $\hat{\mathbf{S}}_{\mathbf{M}}$ with high probability. In summary, the overall scheme succeeds in transmitting k_2 bits of information in each n uses of the channel. Hence, the overall rate $R = (k_2)/n$ satisfies $R = R_{CHA} - R_{SOU} = h(\delta) - h(p) - \epsilon$, showing that we can come arbitrarily close to the curve R_{IE}.

FINITE DEGREES ARE SUFFICIENT

In the preceding discussion, we assumed that the compound construction could be used to perform source and channel coding arbitrarily close to the Shannon limits. Indeed, our recent work [25], [26] shows that it is possible to achieve rates arbitrarily close to the Shannon limits, for both source and channel coding, with all node degrees (check and variable)

[FIG4] (a) A compound LDGM/LDPC construction, consisting of a (n, m) LDGM code specified by generator matrix $\mathbf{G}$ (top part of figure) compounded with a (m, k_1) LDPC code specified by parity check matrix $\mathbf{H}_1$ (bottom part of figure). The n variable nodes (denoted by x) at the top level are the codeword bits, whereas the m variable nodes in the middle layer (denoted by z) are information or state bits. (b) A coset subcode of the code in panel (a), obtained by enforcing an additional set of parity checks . This nesting property of the construction is key for coding with side information.

remaining strictly bounded independently of blocklength. Retaining bounded degrees is crucial if the graphs are to remain sparse and amenable to efficient message-passing algorithms. In order to understand why the compound construction yields an optimal source code, let us return to the second-moment analysis discussed previously. Recall the two factors that control the rate-distortion performance of a code: the overlap probability of two codewords separated by a distance ωn both being distortion D-optimal for the same source sequence, and the weight enumerator that specifies the number of codewords at distance ωn. Note that for any fixed distance ω, the overlap probability is purely a function of the source, whereas the weight enumerator is a code-specific quantity. Accordingly, the purpose of the LDPC code at the bottom of Figure 4(a) is to control this weight enumerator. It is known from early work of Gallager [12] that suitable LDPC codes have linear minimum distance, so that the LDPC code component of the compound construction reshapes the weight enumerator so that the proportion of low-weight sequences is asymptotically vanishing. Application of the second moment method then yields that the overall compound construction saturates the rate-distortion bound; we refer the interested reader to the papers [24]–[26] for further details.
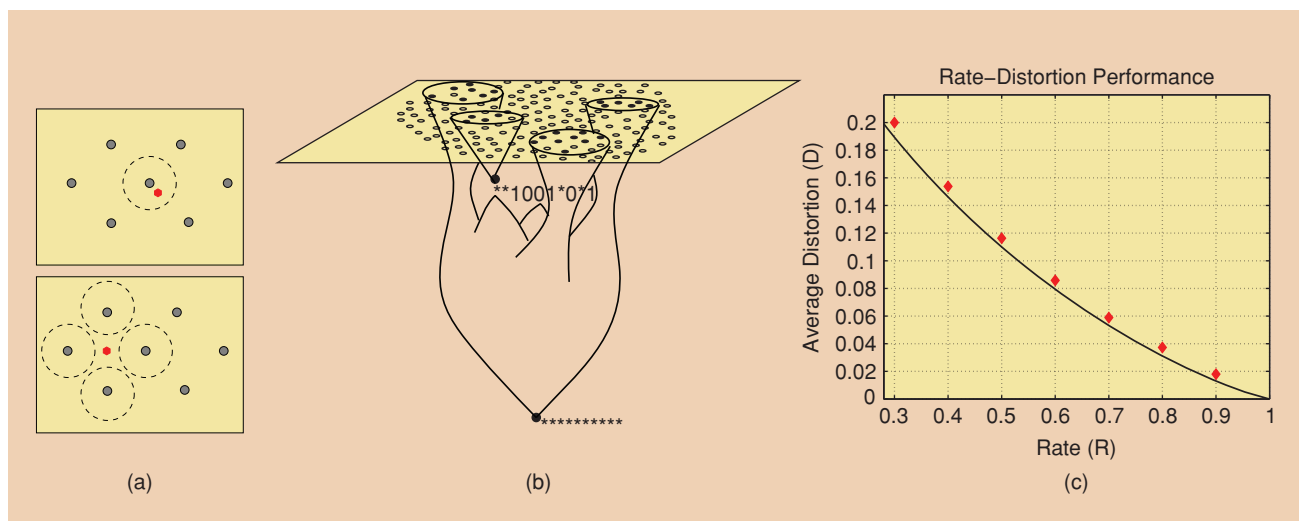
ALGORITHMIC CHALLENGES

The preceding sections focused primarily on the structure of sparse graph codes, and their use in performing lossy compression, coding with side information, and binning. In this section, we turn to some algorithmic issues associated these codes when applied to lossy compression and binning. The significance of bounded degree graph codes is their suitability to efficient algorithms, such as the sum-product and min-sum algorithms, that operate by passing messages along the edges of the factor graph. For pure channel decoding problems, the behavior of message-passing decoding is relatively well-understood, especially in the large block length limit [21], [34]. Many problems in lossless source coding share key features with channel coding problems, which allow similar algorithms to be leveraged. In contrast, for lossy source coding and multistage binning, there remain various open challenges associated with the design and analysis of message-passing algorithms.

GRAPH AND SOLUTION SPACE STRUCTURE

We begin by providing some intuition for the features of channel coding problems (and many lossless source coding problems) that allow iterative message-passing, despite its low complexity, to be such a successful decoding algorithm. As discussed previously, message-passing algorithms such as sum-product and min-sum are exact methods for trees (graphs without cycles). Herein lies the first reason for the success of iterative message-passing: for bounded degree factor graphs such as those associated with LDPC codes [Figure 1(a)], random constructions ensure that the length of the typical cycle is of the order $\log n$. Consequently, from the local perspective of a message-passing algorithm, these graphs are tree-like in that it is possible to take a large number of steps before wrapping around a cycle. It is this locally tree-like nature that underlies the density evolution analysis of message-passing for LDPC codes [21], [34]. A second reason for the success of iterative message-passing involves the structure of the posterior distribution $\mathbb{P}_{X \mid Y}$, defined in equation (1), obtained after observing a particular noisy channel realization $\mathbf{Y} = \mathbf{y}$. Channel coding always involves transmitting a particular codeword $\mathbf{x}$; after transmission over a noisy channel, the received version $\mathbf{y}$ is typically a short distance away, with the distance proscribed by the noise level [see top part of Figure 5(a)]. Consequently, at least for suitable noise levels, the posterior

[FIG5] (a) Top: in the channel coding problem, received channel sequence (pentagon) is a noise-perturbed version of the transmitted codeword. Bottom: in the lossy source coding problem, source sequence (pentagon) is typically at similar distance from many codewords. (b) Cluster structure of solutions in satisfiability problems, and the alternative representation used by the survey propagation algorithm. (c) Empirical results [40] for lossy compression of the symmetric Bernoulli source for blocklengths $n = 10,000$.

distribution typically has a unimodal shape, heavily peaked around the transmitted codeword, so that local message-passing algorithms like the sum-product algorithm are able to recover the transmitted codeword.

NOVEL ALGORITHMIC APPROACHES

By way of contrast, let us consider the problem of lossy source coding, say using an LDGM code illustrated in Figure 3(a). Like an LDPC code, such an LDGM code with bounded degrees also exhibits locally tree-like structure. However, for a lossy compression problem, the structure of the posterior distribution is radically different. Due to the geometry of spheres in high dimensions, it is very unlikely that a typical source sequence is extremely close to any particular codeword; rather, it is more likely to be roughly equidistant from a large number of codewords [see bottom part of Figure 5(a)]. Consequently, unlike the unique mode associated with a channel coding posterior, the source coding posterior will typically have many local modes, corresponding to many codewords whose distortion is close to minimal. Indeed, as shown by our moment analysis (see previous), the typical number of such D-good codewords scales exponentially in the block length for suitably large rates. This multimodal structure in the posterior presents substantial challenges for naive message-passing algorithms. Indeed, if the unmodified sum-product algorithm is applied to an LDGM code for lossy source coding, it either fails to converge or provides uninformative marginal information.

These difficulties suggest that novel algorithms are required for solving coding problems other than channel decoding. Some recent inspiration comes from the statistical physics community, and their work [28] on the survey propagation algorithm for k-SAT problems. The k-SAT problem is a classical problem in computer science, with important connections to circuit design, artificial intelligence, and complexity theory. It is closely related to lossy compression problems: indeed, the encoding problem for an LDGM code can be recast as an instance of a MAX-XORSAT problem, a related class of satisfiability problems. The k-SAT problem has similar structure, in that for appropriate clause densities (the analog of code rate), there are exponentially many satisfying assignments. Moreover, statistical physicists [28] have conjectured that the space of satisfying assignments has a clustered structure, in which solutions are grouped into an exponential number of well-separated clusters. Working from this perspective, Mézard et al. [28] introduced the survey propagation algorithm, a novel type of message-passing explicitly designed to deal with clustering in the space of feasible solutions, that turns out to be extraordinarily successful for solving satisfiability problems. Maneva et al. [22] showed that survey propagation can actually be derived as an instance of the ordinary sum-product algorithm, but as applied to an ingenious factor graph representation of satisfiability problems, one which encodes the cluster structure [see Figure 5(b) for an illustration].

The k-SAT problem shares common features with source coding, which suggests that similarly modified message-passing algorithms should also be useful for lossy data compression and related binning problems. Based on this intuition, a number of researchers have explored modifications of standard message-passing [30] or survey propagation algorithms [7], [40] for quantizing binary sources, as well as adaptation of survey propagation to coding for the Blackwell channel [45]. As with survey propagation applied to k-SAT problems, typical use of these algorithms entails multiple rounds. In each round, the algorithm is run until convergence, and the results are used to determine a subset of strongly biased variables. In the decimation phase, the biased variables are either set to their preferred values, known as hard decimation [7], [40], [45], or reinforced towards their preferred values, known as soft decimation [30]. Wainwright and Maneva [40] showed that LDGM codes constructed using suitable degree distributions, with encoding performed using message-passing and decimation steps, yield performance extremely close to the symmetric Bernoulli rate-distortion bound. Figure 5(c) shows some representative results, using LDGM codes with block-length $n = 10,000$ and suitably irregular degree distributions [34]. In other recent work, Fridrich and Filler [11] extended this approach to achieve state-of-the-art performance for embedding in digital steganography.

DISCUSSION

In this article, we have provided an overview of codes based on factor graphs for lossy compression, and coding with side information. Effective constructions for these problems have a wide range of applications, including MIMO communication, sensor networks, distributed video coding, and digital watermarking, as well as for related communication problems where binning plays an important role. We have emphasized the importance of low-density constructions—meaning that the variable and check degrees in the associated factor graphs remain bounded independently of blocklength—in permitting the application of efficient message-passing algorithms. We also pointed out that communication problems involving quantization and binning often present challenges not present in ordinary channel coding or lossless source coding problems, including nonuniqueness of optimal solutions and possible clustering in the solution space. Recent work has started addressing these challenges, and shown that novel variants on standard message-passing algorithms can yield excellent practical performance for various quantization and binning problems. Nonetheless, the theoretical understanding of such combined message-passing and decimation procedures is still incomplete.

ACKNOWLEDGMENTS

Garcia-Frias and the anonymous reviewers for their constructive criticism and helpful suggestions.

AUTHOR

Martin J. Wainwright (wainwrig@eecs.berkeley.edu) is currently an assistant professor at University of California at Berkeley, with a joint appointment between the Department of Statistics and the Department of Electrical Engineering and Computer Sciences. His research interests include coding and information theory, statistical signal processing, and statistical machine learning. He received the Ph.D. degree in electrical engineering and computer science from Massachusetts Institute of Technology for which he was awarded the George M. Sprowls Award for outstanding doctoral dissertation. He has been awarded an NSF CAREER award (2006), an Alfred P. Sloan Foundation Fellowship (2005), and an Okawa Foundation Research Fellowship (2005).

REFERENCES

[1] N. Alon and J. Spencer, *The Probabilistic Method*. New York: Wiley Interscience, 2000.

[2] R.J. Barron, B. Chen, and G.W. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1159–1180, 2003.

[3] C. Berroux, A. Glavieux, and P. Thitmajshima, "Near Shannon limit error-correcting coding and decoding," in *Proc. ICC*, 1993, pp. 1064–1070.

[4] G. Caire, S. Shamai, and S. Verdu, "Noiseless data compression with low-density parity check codes," in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 66, Oct. 2004, pp. 263–284.

[5] J. Chou, S.S. Pradhan, and K. Ramchandran, "Turbo coded trellis-based constructions for data embedding: Channel coding with side information," in *Proc. Asilomar Conf.*, Nov. 2001, pp. 305–309.

[6] J. Chou, S.S. Pradhan, and K. Ramchandran, "Turbo and trellis-based constructions for source coding with side information," in *Proc. Data Compression Conf.*, 2003.

[7] S. Ciliberti, M. Mézard, and R. Zecchina, "Message-passing algorithms for nonlinear nodes and data compression," Tech. Rep., Nov. 2005, #0508723. [Online]. Available: http://arXiv:cond-mat/0508723.

[8] T.P. Coleman, E. Martinian, M. Effros, and M. Medard, "Interface management via capacity-achieving codes for the deterministic broadcast channel," in *Proc. IEEE Inform. Theory Workshop*, Rotorua, New Zealand, Sept. 2005, pp. 29–33.

[9] M.H.M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, 1983.

[10] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3417–3432, 2005.

[11] J. Fridrich and T. Filler, "Practical methods for minimizing embedding impact in steganography," in *Proc. SPIE Electron. Imaging*, vol. 66, pp. 1–15, Jan. 2007.

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

[13] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Commun. Lett.*, vol. 5, no. 10, pp. 417–419, Oct. 2001.

14] J. Garcia-Frias and Y. Zhao, "Compression of binary memoryless sources using punctured turbo codes," *IEEE Commun. Lett.*, vol. 6, no. 9, pp. 394–396, Sept. 2002.

[15] S.I. Gelfand and M.S. Pinsker, "Coding for channel with random parameters," *Probl. Pered. Inform. (Probl. Inf. Transmission)*, vol. 9, no. 1, pp. 19–31, 1983.

[16] B. Girod, A.M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, 2005.

[17] F.R. Kschischang, B.J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[18] Z. Liu, Z. Xiong, A. Liveris, and S. Cheng, "Slepian-Wolf coded nested lattice quantization for Wyner-Ziv coding: High-rate performance analysis and code design," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4358–4379, 2006.

[19] A. Liveris, Z. Xiong, and C. Georghiades, "Nested convolutional/turbo codes for the binary Wyner-Ziv problem," in *Proc. Int. Conf. Image Processing*, Sept. 2003, vol. 1, pp. 601–604.

[20] H.A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Mag.*, vol. 21, no. 28–41, 2004.

[21] M. Luby, M. Mitzenmacher, M.A. Shokrollahi, and D. Spielman, "Improved low-density parity check codes using irregular graphs," *IEEE Trans. Inform. Theory*, vol. 47, pp. 585–598, Feb. 2001.

[22] E. Maneva, E. Mossel, and M.J. Wainwright, "A new look at survey propagation and its generalizations," *J. Assoc. Comput. Mach.*, to be published.

[23] M.W. Marcellin and T.R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 82–93, 1990.

[24] E. Martinian and M.J. Wainwright, "Analysis of LDGM and compound codes for lossy compression and binning," in *Proc. Workshop Inform. Theory and Applicat.*, Feb. 2006, pp. 229–233. [Online]. Available: http://arxiv:cs.IT/0602046.

[25] E. Martinian and M.J. Wainwright, "Low density codes achieve the rate-distortion bound," in *Proc. Data Compression Conf.*, vol. 1, Mar. 2006, pp. 153–162. [Online]. Available: http:// arxiv:cs.IT/061123.

[26] E. Martinian and M.J. Wainwright, "Low density codes can achieve the Wyner-Ziv and Gelfand-Pinsker bounds," in *Proc. Int. Symp. Inform. Theory*, July 2006, pp. 484–488. [Online]. Available: http://arxiv:cs.IT/0605091.

[27] E. Martinian and J.S. Yedidia, "Iterative quantization using codes on graphs," in *Proc. Allerton Conf. Control, Comput. Commun.*, Oct. 2003.

[28] M. Mézard and R. Zecchina, "Random k-satisfiability: from an analytic solution to an efficient algorithm," *Phys. Rev. E*, vol. 66, no. 1, 2002, pp. 2001–3007.

[29] P. Moulin and R. Koetter, "Data-hiding codes," *Proc. IEEE*, vol. 93, no. 12, pp. 2083–2127, Dec. 2005.

[30] T. Murayama, "Thouless-Anderson-Palmer approach for lossy compression," *Phys. Rev. E*, vol. 69, no. 1, pp. 035105(1)–035105(4), 2004. [31] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufman, 1988.

[31] J. Pearl, *Probalistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufman, 1988.

[32] H. Pfister, I. Sason, and R. Urbanke, "Capacity-achieving ensembles for the binary erasure channel with bounded complexity," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2352–2379, 2005.

[33] S.S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 626–643, 2003.

[34] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

[35] D. Schonberg, S.S. Pradhan, and K. Ramchandran, "LDPC codes can approach the Slepian-Wolf bound for general binary sources," in *Proc. Allerton Conf. Control, Commun. Comput.*, pp. 576–585, Oct. 2002.

[36] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2551–2567, 2006.

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

[38] Y. Sun, A. Liveris, V. Stankovic, and Z. Xiong, "Near-capacity dirty-paper code designs based on TCQ and IRA codes," in *Proc. Int. Symp. Inform. Theory*, Adelaide, Australia. Sept. 2005, pp. 184–188.

[39] A.J. Viterbi and J.K. Omura, "Trellis encoding of memoryless discrete-time sources with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 3, pp. 325–332, 1974.

[40] M.J. Wainwright and E. Maneva, "Lossy source coding by message-passing and decimation over generalized codewords of LDGM codes," in *Proc. Int. Symp. Inform. Theory*, Adelaide, Australia, Sept. 2005. [Online]. Available: http:// arxiv:cs.IT/0508068.

[41] X. Wang and M.T. Orchard, "Design of trellis codes for source coding with side-information," in *Proc. Data Compression Conf.*, 2001, pp. 361–370.

[42] A.D. Wyner and J. Ziv, "The rate-distortion function for source encoding with side information at the encoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1–10, Jan. 1976.

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

[44] Y. Yang, V. Stankovic, Z. Xiong, and W. Zhao, "On multiterminal source code design," in *Proc. Data Compression Conf.*, 2005, pp. 43–52.

[45] W. Yu and M. Aleksic, "Coding for the Blackwell channel: A survey propagation approach," in *Proc. Int. Symp. Inform. Theory*, Adelaide, Australia, Sept. 2005, pp. 1583–1587.

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

SP