Cutset Width and Spacing for Reduced Cutset Coding of Markov Random Fields

Matthew G. Reyes
self-employed
mgreyes@umich.edu

David L. Neuhoff
University of Michigan
neuhoff@umich.edu

***Abstract*—This paper analyzes the rate, redundancy and complexity of a generalized version of Reduced Cutset Coding (RCC) [4] in which cutset *lines* can consist of more than one image row. We show that increasing the thickness of the lines reduces coding rate for the cutset, increasing the spacing of the lines increases the coding rate of the non-cutset pixels, though the coding rate of the latter is always strictly less than that of the former. We show that the redundancy of RCC can be decomposed into two terms, a correlation redundancy due to coding the components of the cutset independently, and a distribution redundancy due to coding the cutset as a reduced Ising model [3]. We show that the correlation redundancy decreases with increasing spacing of the lines, and that the parameter of the reduced Ising model approaches the true parameter monotonically in the thickness of the lines. We present a learning phase in which the true parameter of the model is learned as well as the parameter of the reduced model. Numerical simulations show that only small improvements in rate-complexity performance are possible by increasing line widths, though rates within 3 to 4 percent of the entropy rate can be achieved with only 20 to 50 operations per pixel through combinations of line thickness and spacing.**

I. INTRODUCTION

The general topic of this paper is lossless compression of a Markov random field. The notion of a multivariate probability distribution p being *Markov with respect to* an undirected graph $G = (V, E)$ is a rather general one, meaning simply that the conditional independence relations of p are depicted by G, in that if two nodes are not connected by an edge in G, then the corresponding random variables are conditionally independent of each other conditioned on the values of all other nodes. Thus by saying that we want to losslessly compress a Markov random field, what we are really saying is that we want to compress a source of information taking advantage of the graph G that depicts the source's conditional independence structure. While there has been relatively little development of algorithms or theory for lossless compression of MRFs [1–6], we feel that this is an important problem to consider.

The specific information source that we consider in the present paper is a uniform Ising model on a square grid graph G, whose nodes are the sites of an $M \times N$ rectangular lattice and whose edges are pairs of horizontally and vertically adjacent nodes. The random variable X_i associated with each node i assumes values in the alphabet $\mathcal{X} = \{-1, 1\}$ and a configuration $\mathbf{x} = (x_i : i \in V)$ has probability given by

$$p(\mathbf{x}; \theta) \quad = \quad \exp\{\, \theta \sum_{\{i,j\} \in E} x_i x_j - \Phi(\theta)\}, \qquad (1)$$

where $\Phi(\theta)$ is the log-partition function and θ is the exponential parameter of the model. The Ising model was introduced as a model to describe the spontaneous magnetization of iron [7] and has proved significant in many problems in statistical physics. Moreover, generalization of it led to the concept of a Markov random field. The Ising model has also been proposed as a model for bilevel images [8] called *scenic* which are complex bilevel images, such as landscapes and portraits, having numerous black and white regions with smooth or piecewise smooth boundaries between them. For exponential parameter θ, we define

$$\mu \quad \triangleq \quad \mathbb{E}_\theta\Big[\sum_{\{i,j\} \in E} x_i x_j \Big] \qquad (2)$$

to be the *moment* of the Ising model. It is well-known that θ and μ are *dual parameters* for the family of Ising models [9].

We consider encoding a configuration $\mathbf{x}$ using Reduced Cutset Coding (RCC) [4, 5], a two-stage algorithm for lossless compression of an MRF defined on an intractable graph, where tractability is with respect to Belief Propagation (BP) [9]. First, a cutset $U \subset V$ is encoded using the induced subgraph G_U, and then each component of the remainder $W = V \setminus U$ is conditionally encoded (using the full graph G) given the values on the cutset. It is called reduced cutset coding because the cutset is, in effect, encoded as a reduced MRF, i.e., an MRF on the induced subgraph.

We use a cutset U consisting of $k+1$ evenly spaced $n_L \times N$ rectangular regions $L_1, \ldots, L_{k+1}$, referred to as *lines*. This generalizes [4, 5] in which $n_L = 1$, and permits exploration of the benefits of $n_L > 1$. Because the induced subgraph G_U is the union of disconnected induced subgraphs G_{L_i}, encoding the cutset U using G_U amounts to encoding each line L_i (independently of all strips and all other lines) with an encoder, for example Arithmetic Encoding (AC), with a coding distribution generated by a uniform Ising model on the induced subgraph G_{L_i} with a parameter $\tilde{\theta}_{n_L}$ that we choose. We will refer to such an Ising model as a *reduced uniform Ising model*. For the sake of brevity, we will subsequently use the phrase *code with an Ising model* to mean "code with a coding distribution generated by a uniform Ising model". Note that the coding distribution used to encode the line is not the same as the true (marginal) distribution of the line.

For a subset of sites $L \subset V$, we define μ_L to be the moment for subset L, defined as in (2) with the summation over edges in L and expectation with respect to the marginal distribution for X_L. For the coding parameter $\tilde{\theta}_{n_L}$ used to encode the lines, the moment for the reduced Ising model on the lines is denoted $\tilde{\mu}_{n_L}$ and is defined analogously to (2) on G_L. Following [4], the redundancy of encoding the lines with reduced Ising models is minimized when the parameter $\tilde{\theta}_{n_L}$ is chosen so that the moment for the reduced Ising model $\tilde{\mu}_{n_L}$ equals the true moment μ_L for a line. This is referred to as the moment-matching condition, and the moment-matching parameter will be denoted $\theta^*_{n_L}$. We let $R^L_{n_L}$ denote the rate obtained in encoding the lines with moment-matched reduced (uniform) Ising models with line width n_L. Note both that the moment-matching parameter $\theta^*_{n_L}$ will vary with line width, and that even with the moment-matching parameter, the encoding will be suboptimal in that its rate $R^L_{n_L}$ will be strictly greater than the (normalized) entropy of the true distribution of the line.

The k components of the induced subgraph G_W are themselves $n_S \times N$ rectangular regions $S_1, \ldots, S_k$, referred to as *strips*. Each strip S_i is encoded with AC conditioned on the cutset (union of lines) with coding distribution equal to the conditional distribution generated by the uniform Ising model. By the Markov property, this reduces to the conditional distribution of the strip given its boundary, namely, the last row of line L_i, which we denote L_{i,n_L}, and the first row of line L_{i+1}, denoted $L_{i+1,1}$. Therefore the rate $R^S_{n_S}$ of encoding a strip is equal to the (normalized) conditional entropy of the strip given its boundary ∂S_i.

The results of this paper are as follows. We show that the coding rate $R^S_{n_S}$ of a strip increases with n_S, the coding rate $R^L_{n_L}$ of a line decreases with n_L when the moment-matching parameter $\theta^*_{n_L}$ is used to encode the lines, and that $R^S_{n_S} < R^L_{n_L}$ for all choices of n_S and n_L. We show that the redundancy of the coding method can be decomposed into a correlation redundancy due to encoding the lines independently and a distribution redundancy due to approximating the lines with reduced Ising models. We show the first of these decreases with n_S and that $\theta^*_{n_L}$ monotonically decreases to θ as n_L increases. From this, we conjecture that the second redundancy term decreases monotonically with n_L. We discuss a learning phase for RCC in which both an estimate $\hat{\theta}$ of the Ising model parameter is learned using the Minimum Conditional Description Length method introduced in [10], and an estimate $\hat{\theta}^*_{n_L}$ of the moment-matching parameter for lines is learned using an analogous Minimum Description Length approach. We present expressions for rate and complexity of the method as a function of n_S and n_L, thus allowing us to explore rate-complexity tradeoffs. Numerical experiments find that only small improvements in rate-complexity performance are possible by increasing line widths, but that it is possible to achieve within 3 to 4 percent of the entropy rate with between 20 and 50 operations per pixel. Proofs of propositions and a more general treatment of the problem can be found in [11].

Section II provides background on AC and BP, Section III analyzes rate and complexity of RCC, Section IV presents methods for estimating θ and $\theta^*_{n_L}$, and Section V presents numerical simulations and explores coding rate and complexity.

II. Background

A. Graphs, Markov Properties, and Subset Distributions

For any $U \subset V$, its *boundary* ∂U is the set of nodes not in U connected by an edge to a member of U. The subgraph G_U *induced by* U is the graph consisting of nodes and edges contained in U. A subset U is called a *cutset* if $G_{V \setminus U}$ has more than one component. A subset U is said to be *tractable* if the induced subgraph G_U is acyclic or if G_U can be clustered into an acyclic graph where the size of the largest clustered node is of moderate size, say at most 10.

We find it convenient to include an additional argument in the notation for the probability distribution of an Ising model. Specifically, $p(G; \mathbf{x}; \theta)$ indicates not only the configuration $\mathbf{x}$ in question and the exponential parameter θ of the distribution, but also the graph G on which the model is defined. For subset $L \subset V$, the marginal probability distribution on X_L is denoted $p(G; \mathbf{x}_L; \theta)$. The conditional probability of a configuration $\mathbf{x}_S$ on subset $S \subset V$ given the values $\mathbf{x}_U$ on another subset $U \subset V$ is denoted $p(G; \mathbf{x}_S | \mathbf{x}_U; \theta)$. It is straightforward to check that $p(G; \mathbf{x}_S | \mathbf{x}_{\partial S}; \theta) = p(G; \mathbf{x}_S | \mathbf{x}_{V \setminus S}; \theta)$ for all S, $\mathbf{x}_S$, and $\mathbf{x}_{\partial S}$. Indeed, this is the *Markov Property*. A *reduced Ising model* for X_L on G_L with exponential parameter $\tilde{\theta}_{n_L}$ is denoted $p(G_L; \mathbf{x}_L; \tilde{\theta}_{n_L})$ and has the same form as in (1).

B. Arithmetic Encoding and Belief Propagation

To encode a configuration $\mathbf{x}_L$ or $\mathbf{x}_S$ on a line or strip with Arithmetic Encoding (AC) involves passing to the encoder individual symbols of the configuration together with symbol probability distributions, the product of which is the *coding distribution* for the configuration. Details on the use of AC in encoding an MRF are given in [2, 4, 5].

For a line L of width n_L, configuration $\mathbf{x}_L$ is losslessly compressed with a reduced Ising coding distribution $p(G_L; \mathbf{x}_L; \tilde{\theta}_{n_L})$, and the average number of bits produced by AC is the *cross entropy* $H(G; X_L; \theta || G_L; X_L; \tilde{\theta}_{n_L})$ between the marginal distribution $p(G; \mathbf{x}_L; \theta)$ and the reduced MRF coding distribution $p(G_L; \mathbf{x}_L; \tilde{\theta}_{n_L})$ for X_L, defined as

$$H(G; X_L; \theta || G_L; X_L; \tilde{\theta}_{n_L}) \triangleq H(G; X_L; \theta) + D(p(G; \mathbf{x}_L; \theta) || p(G_L; \mathbf{x}_L; \tilde{\theta}_{n_L}))$$

where $H(G; X_L; \theta)$ is the entropy of X_L, $D(p(G; \mathbf{x}_L; \theta) || p(G_L; \mathbf{x}_L; \tilde{\theta}_{n_L}))$ is the *divergence* from $p(G; \mathbf{x}_L; \theta)$ to $p(G_L; \mathbf{x}_L; \tilde{\theta}_{n_L})$ and is the *redundancy* in the code.

We showed in [4] that the above divergence is minimized at $\theta^*_{n_L}$, the exponential parameter on G_L such that the corresponding moment $\mu^*_{n_L}$ is equal to the moment μ_L under the true marginal $p(G; \mathbf{x}_L; \theta)$. The reduced Ising model $p(G_L; \mathbf{x}_L; \theta^*_{n_L})$ is called the *moment-matching* reduced Ising model for X_L, and as shorthand, is sometimes denoted $\tilde{X}_L$. When the moment-matching reduced Ising model $p(G_L; \mathbf{x}_L; \theta^*_{n_L})$ is used as the coding distribution to encode X_L, the cross entropy is in fact the entropy $H(G_L; X_L; \theta^*_{n_L})$

of the moment-matching reduced Ising model [4]. Normalizing by the number of pixels, the coding rate for a line is then

$$\begin{aligned} R^L_{n_L} &= \frac{1}{n_L N} H(G; X_L; \theta || G_L; X_L; \theta^*_{n_L}) \\ &= \frac{1}{n_L N} H(G_L; X_L; \theta^*_{n_L}). \end{aligned}$$

For a strip S of width n_S, configuration $\mathbf{x}_S$ is losslessly compressed using coding distribution equal to the conditional distribution $p(G; \mathbf{x}_S | \mathbf{x}_{\partial S}; \theta)$ of $\mathbf{x}_S$ given its boundary $\mathbf{x}_{\partial S}$. The average number of bits produced by AC is the conditional entropy$H(G; X_S | X_{\partial S}; \theta)$. The coding rate for a strip is then

$$R^S_{n_S} = \frac{1}{n_S N} H(G; X_S \mid X_{\partial S}; \theta).$$

If n_L is moderate, for example no more than 10, one can cluster a line into a chain graph in which the columns of the line are clustered into supernodes with state space $\{-1, 1\}^{n_L}$. Then, when coding the line, Belief Propagation can be used to compute the coding distribution $p(G_L; \mathbf{x}_L; \tilde{\theta}_{n_L})$ generated by the reduced Ising model on G_L, for configuration $\mathbf{x}_L$. This involves, for example, an initial phase of leftward messages in which the message from column i to column $i-1$, assuming we have enumerated the columns from left to right, consists of the element-wise product of two 2^{n_L} length vectors and the subsequent product of the resulting 2^{n_L} length vector and a $2^{n_L} \times 2^{n_L}$ matrix. A subsequent rightward phase of messages consists of 2^{n_L} multiplications per message. Both phases are done for $N-1$ edges, so the total number of arithmetic operations is $c^L_{n_L} \approx (2^{2n_L+1} + 2^{n_L})/n_L$.

In encoding a configuration $\mathbf{x}_S$ on strip S of width n_S conditioned on the configuration $\mathbf{x}_{\partial S}$ on its boundary ∂S, the coding distribution $p(G; \mathbf{x}_S | \mathbf{x}_{\partial S}; \theta)$ can likewise be computed by clustering the columns into supernodes and running BP. The same computations that are performed in computing the coding distributions for a line are required in the computation of the coding distribution for a strip. Furthermore, an additional $N 2^{n_S+1}$ operations are needed to account for conditioning on the boundary. Thus the number of arithmetic operations is $c^S_{n_S} \approx (2^{2n_S+1} + 2^{n_S} + 2^{n_S+1})/n_S$.

III. Rate and Complexity

We let R_{n_S,n_L} denote the total rate of encoding the $k+1$ lines and k strips with cutset parameters n_S and n_L. If k is large, rate is well-approximated by

$$R_{n_S,n_L} \approx \frac{n_S}{n_L + n_S} R^S_{n_S} + \frac{n_L}{n_L + n_S} R^L_{n_L}. \quad (3)$$

Analogously, the complexity of RCC is

$$C_{n_S,n_L} \approx \frac{n_S}{n_S + n_L} C^S_{n_S} + \frac{n_L}{n_S + n_L} C^L_{n_L}. \quad (4)$$

We see that the rate-complexity performance of RCC with cutset parameters n_S and n_L is characterized by the rates $R^L_{n_L}$ and $R^S_{n_S}$, the complexities $C^L_{n_L}$ and $C^S_{n_S}$, and the fractions $\frac{n_L}{n_L+n_S}$ and $\frac{n_S}{n_L+n_S}$. The question then is how to choose n_S and n_L to optimize R_{n_S,n_L} and C_{n_S,n_L}. The following proposition provides a starting point for discussion of these tradeoffs.

Proposition 3.1: For all n_S and n_L,

1) $R^L_{n_L+1} < R^L_{n_L}$
2) $R^S_{n_S+1} > R^S_{n_S}$
3) $R^L_{n_L} > R^S_{n_S}$

The first statement asserts that the line rate $R^L_{n_L}$ decreases by making n_L larger. Recalling that the lines are encoded with the moment-matching Ising model, it is clear that if we could set n_L equal to the image height M, then the moment-matching parameter θ^*_M would equal the true parameter θ and the encoding would be optimal, that is R_{n_S,n_L} would equal the entropy rate of X. Additionally, the second statement of the proposition states that as the strip width n_S becomes larger, the effect of the conditioning diminishes and $R^S_{n_S}$ increases. Indeed, $R^S_{n_S}$ also approaches the entropy rate with increasing n_S. The third statement of the proposition summarizes that $R^L_{n_L}$ approaches the entropy rate from above, while $R^S_{n_S}$ from below. It is straightforward to see, then, that making either n_L or n_S very large will achieve an optimal or very nearly optimal rate. However, given that the complexities C^L_M and C^S_M are each infeasible, how do we choose n_S and n_L to achieve the best rate possible given a constraint on the maximum of the two. Thus bounding both n_S and n_L by some maximum width, we now see that there is a tradeoff in the choice of strip and line width. Specifically, by increasing n_L the line rate $R^L_{n_L}$ decreases, though the fraction $\frac{n_L}{n_S+n_L}$ of pixels encoded at the higher rate increases, while increasing n_S increases the fraction $\frac{n_S}{n_L+n_S}$ of pixels encoded at the lower rate, though the strip rate $R^S_{n_S}$ increases.

The appropriate balance between n_S and n_L ultimately depends on how $R^L_{n_L}$ decreases and $R^S_{n_S}$ increases. However, we can get a more detailed perspective on the tradeoffs between n_S and n_L by examining the per-site redundancy $\Delta_{n_S,n_L} \triangleq \frac{1}{|V|} D(X_U || \tilde{X}_U)$ in the coding rate, where $D(X_L || \tilde{X}_L)$ is shorthand for $D(p(G; \mathbf{x}_L; \theta) || p(G_L; \mathbf{x}_L; \theta^*_{n_L}))$.

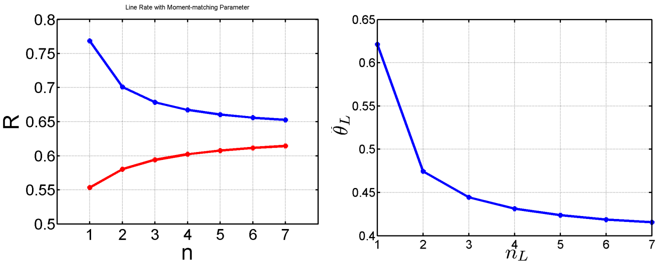
Proposition 3.2:

$$\Delta_{n_S,n_L} = \frac{I(X_{r_1}; X_{r_{-n_S}})}{N(n_S + n_L)} + \frac{D(X_L || \tilde{X}_L)}{N(n_S + n_L)}$$

where $I(\cdot;\cdot)$ denotes information, r_1 is the 1st row of a line, and r_{-n_S} is the last row of the previous line.

This proposition shows specifically how the redundancy of RCC has two components: a correlation redundancy $I(X_{r_1}; X_{r_{-n_S}})$ due to encoding the lines independently of one another, and a distribution redundancy $D(X_L || \tilde{X}_L)$ due to approximating the lines as moment matching reduced Ising models. Moreover, while the redundancy of RCC is entirely due to the encoding of the lines, it is still a function of strip width n_S through the correlation redundancy. Intuitively one expects that as n_S increases, the information between successive lines decreases. This intuition is born out by the following proposition.

Proposition 3.3: $I(X_{r_1}; X_{r_{-n_S}})$ is decreasing in n_S.



(a) (b)

Fig. 1. For $\theta = .4$, (a) rates R_n^L for lines (blue) and R_n^S for strips (red); (b) estimate of moment-matching parameter $\hat{\theta}^*_{n_L}$.

Regarding the distribution redundancy between X_L and $\tilde{X}_L$, we showed in [3] that $\theta^*_{n_L} > \theta$, which intuitively makes sense since a stronger $\tilde{\theta}_{n_L}$ would be required to enforce the correlation on X_L. In light of this, the following proposition is also intuitive.

Proposition 3.4 (Monotonicity of $\theta^*_{n_L}$):

$$\theta^*_{n_L+1} \quad < \quad \theta^*_{n_L}.$$

Again, when n_L equals the height of the image, $\theta^*_{n_L} = \theta$, and $D(X_L||\tilde{X}_L) = 0$. Having no reason to suppose that $D(X_L||\tilde{X}_L)$ is non-monotonic, we extrapolate the following.

Conjecture 3.5: $D(X_L||\tilde{X}_L)$ is decreasing in n_L.

We now consider the effects of changing n_S and n_L on Δ_{n_S,n_L}, as expressed in Proposition 3.2. Increasing n_S decreases distribution redundancy through the factor $\frac{1}{n_S+n_L}$. It likewise decreases the correlation redundancy, not only through the fraction $\frac{1}{n_S+n_L}$, but also because the information $I(X_{r_1};X_{r_{-n_S}})$ decreases with n_S. Similarly, increasing n_L decreases the correlation redundancy through the factor $\frac{1}{n_S+n_L}$. Assuming the above conjecture, it also decreases the distribution redundancy, both through the fraction $\frac{1}{n_S+n_L}$, and by decreasing the divergence $D(X_L||\tilde{X}_L)$. From the point of view of minimizing R_{n_S,n_L}, we conclude that one should make n_S and n_L each as large as possible. However, whether a given combination of n_S and n_L is feasible depends on the computational resources and demands of the particular application. Moreover, how $D(X_L||\tilde{X}_L)$ decreases with n_L and how $I(X_{r_1};X_{r_{-n_S}})$ decreases with n_S will play a role in the choice of n_S and n_L, as one will likely decrease faster than the other and there may be a point of diminishing returns on one or the other. In Section V we consider how R_{n_S,n_L} and C_{n_S,n_L} change by increasing either n_S or n_L while keeping the other fixed at some value; by increasing both n_S and n_L simultaneously; and by keeping the sum n_S+n_L constant for a particular value of θ.

IV. Learning Phase: Estimating θ and $\theta^*_{n_L}$

To encode the lines we need an estimate of the moment-matching parameter $\theta^*_{n_L}$, and to encode the strips we need to estimate the true parameter θ of the Ising model. We can learn these estimates from observations $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ on G.

We form the estimate $\hat{\theta}^*_{n_L}$ by minimizing the empirical cross-entropy

$$H(\tilde{\theta}_{n_L}) \quad = \quad \frac{1}{nK}\sum_{j=1}^{n}\sum_{L_i} -\log p(G_{L_i};\mathbf{x}^{(j)}_{L_i};\tilde{\theta}_{n_L}), \quad (5)$$

where K is the number of lines in each image. This is Minimum Description Length estimation as $H(\tilde{\theta}_{n_L})$ is the approximate number of bits need to encode the lines $\{x^{(j)}_{L_i}\}$ as reduced Ising models with parameter $\tilde{\theta}_{n_L}$.

Proposition 4.1: $\hat{\theta}^*_{n_L}$ is consistent.

We form the estimate $\hat{\theta}$ by minimizing the empirical cross-entropy

$$H(\tilde{\theta}) \quad = \quad \frac{1}{nK}\sum_{j=1}^{n}\sum_{S_i} -\log p(G;\mathbf{x}^{(j)}_{S_i}|\mathbf{x}^{(j)}_{\partial S_i};\tilde{\theta}). \quad (6)$$

where K is the number of strips in each image. This is Minimum Conditional Description Length estimation [10], which can be seen as a re-formulation and generalization of Pseudo-Likelihood estimation. The strips S_i can be chosen to be any width, though it is not yet understood how accuracy of the estimate $\hat{\theta}$ depends on the width.

Proposition 4.2: $\hat{\theta}$ is consistent.

V. Numerical Simulations

Using Gibbs sampling, we generated configurations $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(17)}$ with $\theta = 0.4$ on a 200×200 lattice where each interior site is connected to its four nearest neighbors. We obtained the estimate $\hat{\theta} = 0.4025$ for the parameter of the Ising model, and estimates $\hat{\theta}^*_{n_L}$ for the reduced Ising model are shown in Figure 1 (b) for different values of n_L. Note that for $n_L \geq 4$, the estimate $\hat{\theta}^*_{n_L}$ is within 10% of θ. The line rate $R^L_{n_L}$ and the strip rate $R^S_{n_S}$ are approximated by (5) and (6) using the estimates $\hat{\theta}^*_{n_L}$ and $\hat{\theta}$, respectively, for $n_L, n_S = 1, \ldots, 7$. As shown in Figure 1 (a), and predicted by Proposition 3.1, $\hat{R}^S_{n_S}$ is increasing in n_S, $\hat{R}^L_{n_L}$ is decreasing in n_L, and $\hat{R}^S_{n_S} < \hat{R}^L_{n_L}$ for all n_S, n_L.

We computed $\hat{R}_{n_S,n_L}$ from $\hat{R}^L_{n_L}$ and $\hat{R}^S_{n_S}$ using (3), and Figure 2 shows results from holding n_L fixed and increasing n_S; holding n_S fixed and increasing n_L; increasing both n_S and n_L; and holding the sum n_S+n_L constant. In comparison to RCC, the rate of simple *line scan coding*, in which each pixel is coded conditioned on the previous pixel in the same row, has rate $R^L_1 = 0.77$, which lies considerably above the rates of all RCC points. As anticipated, we found that $\hat{R}_{n_S,n_L}$ decreases while increasing n_L and holding n_S constant, increasing n_S and holding n_L constant, and increasing both n_S and n_L. Noting the $n_S = 1$ and $n_L = 1$ curves, we observe an interesting tradeoff within RCC. In particular, holding n_S fixed at 1 and increasing n_L, the strip rate $\hat{R}^S_1 = 0.55$ is constant and the line rate $\hat{R}^L_{n_L}$ is decreasing, while the weighting on

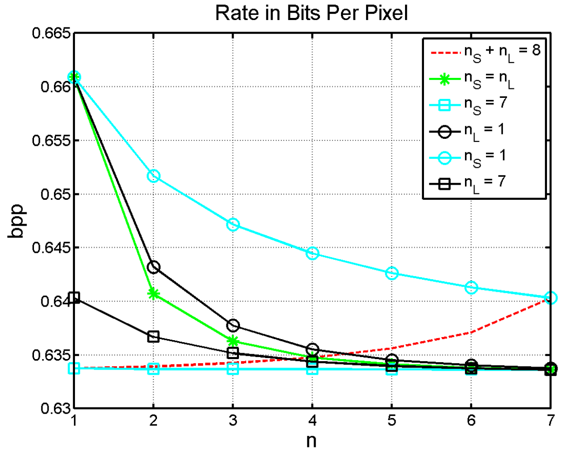


Fig. 2. Encoding rates for different choices of varying n_S of n_L. The cyan curves are holding n_S fixed and increasing n_L; the black holding n_L fixed and increasing n_S; the green increasing both n_S and n_L; and the red is holding the sum $n_S + n_L$ constant and is shown as a function of n_L.

$\hat{R}^L_{n_L}$ is increasing, resulting in an expected decrease in $\hat{R}_{1,n_L}$. On the other hand, holding n_L fixed at 1 and increasing n_S, the line rate $\hat{R}^L_1 = 0.77$ is constant, while both the strip rate $\hat{R}^S_{n_S}$ and the weighting on $\hat{R}^S_{n_S}$ are increasing. Nevertheless, not only does $\hat{R}_{n_S,1}$ decrease, it decreases more rapidly than when holding n_S fixed at 1 and increasing n_L. This is because although $\hat{R}^S_{n_S}$ is increasing, there is less redundancy in the line rate, as increasing n_S decreases the correlation between lines. This is made more explicit by noting the curve for constant $n_L + n_S$. In this case, the fact that $\hat{R}_{n_S,n_L}$ increases in n_L indicates that the correlation redundancy decreases in n_S faster than the distribution redundancy decreases in n_L. This same tradeoff can be observed in the $n_S = 7$ and $n_L = 7$ curves.

Figure 3 (a) shows the number of operations per pixel C_{n_S,n_L} as a function of line or strip width, computed from (4). Figure 3 (b) shows coding rate $\hat{R}_{n_S,n_L}$ vs. complexity C_{n_S,n_L}. The $n_S = n_L$ and $n_L = 1$ curves give the best rate-complexity performance, and the two provide tradeoffs between rate and complexity for comparable data points. Interestingly, the $n_L = 1$ case is what we initially explored in [4]. The $n_S = n_L$ curve offers a potential benefit in simplicity in that having equal line and strip widths means that there is essentially a single algorithm used for computing coding distributions, with a pre-processing step to account for the boundary in the case of strips. Looking at the indicated (n_S, n_L) points on this curve, and using $\hat{R}^S_7$ as a lower bound for the entropy rate, we see that with $n_S = n_L = 2$, RCC achieves within 4.25% of the entropy rate with only 20 operations per pixel, or with $n_S = n_L = 3$, within 3.5% of entropy rate with 48 operations per pixel.

In conclusion, we have analyzed rate, redundancy and complexity of a generalized version of RCC in which lines can consist of more than one image row, described methods for estimating the key MRF parameter and the optimum coding parameter for RCC. Numerical experiments then explored the rate-complexity tradeoffs. It is found that only small improvements in rate-complexity performance are possible by increasing line widths.

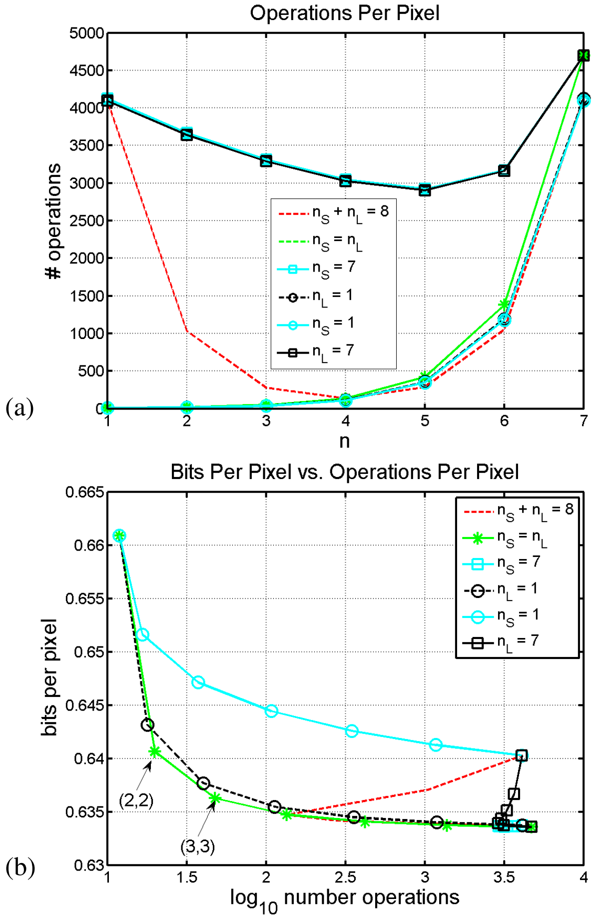


Fig. 3. (a) Complexity of RCC for different variations of n_S and n_L; (b) R_{n_S,n_L} vs. C_{n_S,n_L} for different variations.

References

[1] D. Anastassiou and D.J. Sakrison, "Some Results Regarding the Entropy Rate of Random Fields," *IEEE Trans. Inform. Thy.*, vol. IT-28, pp. 340–343, Mar. 1982.

[2] M.G. Reyes and D.L. Neuhoff, *Arithmetic Compression of Markov Random Fields*, Seoul, Korea, ISIT 2009.

[3] M. G. Reyes and D. L. Neuhoff, "Entropy Bounds for a Markov Random Subfield," *Proc. ISIT*, Seoul, Korea, pp. 309–313, July 2009.

[4] M.G. Reyes and D.L. Neuhoff, *Lossless Reduced Cutset Coding of Markov Random Fields*, Snowbird, UT, DCC 2010.

[5] M.G. Reyes, *Cutset Based Processing and Compression of Markov Random Fields*, Ph.D. thesis, University of Michigan, April 2011.

[6] S. Wu, *Lossless Compression of Markov Random Fields with Higher Order Cliques*, Master's thesis, Queen's University, August 2013.

[7] R.J. Baxter, *Exactly Solved Models in Statistical Mechanics*, New York: Academic, 1982.

[8] M. G. Reyes, D. L. Neuhoff, T. N. Pappas, "Lossy Cutset Coding of Bilevel Images Based on Markov Random Fields," *IEEE Trans. Img. Proc.*, vol. 23, pp. 1652-1665, April 2014.

[9] M. J. Wainwright and M. I. Jordan, *Graphical models, exponential families and variational inference*, Berkeley Tech. Report 649, Sept. 2003.

[10] M. G. Reyes and D. L. Neuhoff, "Minimum Description Length Estimation for Markov Random Fields", ITA, San Diego, CA, February 2016.

[11] M. G. Reyes and D. L. Neuhoff, "Cutset Width and Spacing for Reduced Cutset Coding of Markov Random Fields," available online at http://arxiv.org/abs/1602.04835.