Estimating the Information Rate of Noisy Two-Dimensional Constrained Channels

Mehdi Molkaraie and Hans-Andrea Loeliger
Dept. of Information Technology and Electrical Engineering
ETH Zurich, Switzerland
Email: {molkaraie,loeliger}@isi.ee.ethz.ch

***Abstract*—The problem of computing the information rate of noisy two-dimensional constrained source / channel models has been an unsolved problem. In this paper, we propose two Monte Carlo methods for this problem. The first method, which is exact in expectation, combines tree-based Gibbs sampling with importance sampling. The second method uses generalized belief propagation and is shown to yield a good approximation of the information rate.**

I. Introduction

Simulation-based techniques as in [1] and [2] can be used to compute the information rate of noisy one-dimensional (1-D) constrained source / channel models. For two-dimensional (2-D) noisy constrained source / channel models, however, computing the information rate has remained an unsolved problem.

In a previous paper [3], we used tree-based Gibbs sampling to compute the noiseless capacity of 2-D constrained channels. In this paper, we extend this approach (using ideas from [4] that were not used in [3]) to compute the information rate of noisy constrained 2-D source / channel models. In addition, we propose also a second Monte Carlo method which uses generalized belief propagation (GBP) to compute an approximation of the information rate. The second method works also at high SNR where the first method suffers from slow convergence.

II. Problem Set-up

We consider the problem of computing the information rate of noisy 2-D constrained source / channel models. For a 2-D grid of size $N = M \times M$, let $\mathbf{X} = \{X_1, X_2, \ldots, X_N\}$ be the input and $\mathbf{Y} = \{Y_1, Y_2, \ldots, Y_N\}$ be the output of the channel. Let x_i denote a realization of X_i and let $\mathbf{x}$ denote $\{x_1, x_2, \ldots, x_N\}$. We assume that each input X_i takes values in a finite set $\mathcal{X}$.

In constrained channels, not all sequences of symbols from the channel input alphabet may be transmitted. Let $\mathcal{S}_\mathbf{X} \subset \mathcal{X}^N$ be the set of admissible input sequences. We define the indicator function

$$f(\mathbf{x}) \triangleq \begin{cases} 1, & \mathbf{x} \in \mathcal{S}_\mathbf{X} \\ 0, & \mathbf{x} \notin \mathcal{S}_\mathbf{X}. \end{cases} \tag{1}$$

A basic assumption in this paper is that $f(\mathbf{x})$ can be factored into "local" constraints. For example, let $f(\mathbf{x})$ be the product of kernels of the form

$$f_a(x_k, x_\ell) = \begin{cases} 0, & \text{if } x_k = x_\ell = 1 \\ 1, & \text{else,} \end{cases} \tag{2}$$

with one such kernel for each adjacent pair (x_k, x_ℓ). This example is known as the 2-D $(1, \infty)$ constrained channel and will be used for all numerical results of this paper. The corresponding Forney-style factor graph of f is shown in Fig. 1, where the boxes labeled "=" are equality constraints [5]. (Fig. 1 may also be viewed as a factor graph as in [6] where the boxes labeled "=" are the variable nodes.)

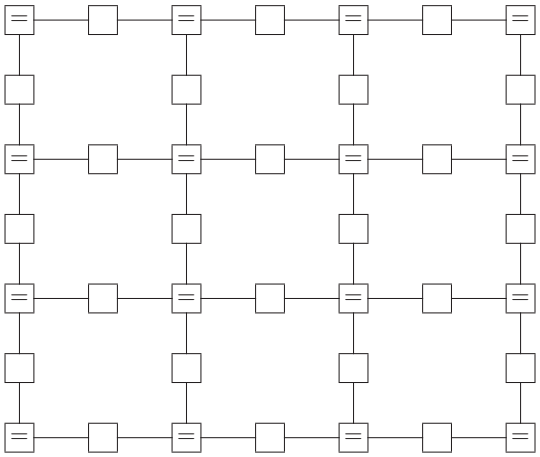

Fig. 1. Forney-style factor graph of the indicator function of a $(1, \infty)$ constraint. The unlabeled boxes represent factors as in (2).

We define

$$Z \triangleq \sum_{\mathbf{x} \in \mathcal{X}^N} f(\mathbf{x}) \tag{3}$$

$$= |\mathcal{S}_\mathbf{X}|, \tag{4}$$

which is also known as the *partition function* in statistical physics.

The mutual information rate is

$$\frac{1}{N} I(\mathbf{X}, \mathbf{Y}) = \frac{1}{N} \big(H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) \big). \tag{5}$$

For simplicity, we suppose that $H(\mathbf{Y}|\mathbf{X})$ is analytically available and $p(\mathbf{y}|\mathbf{x})$ factors as

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{N} p(y_i|x_i). \tag{6}$$

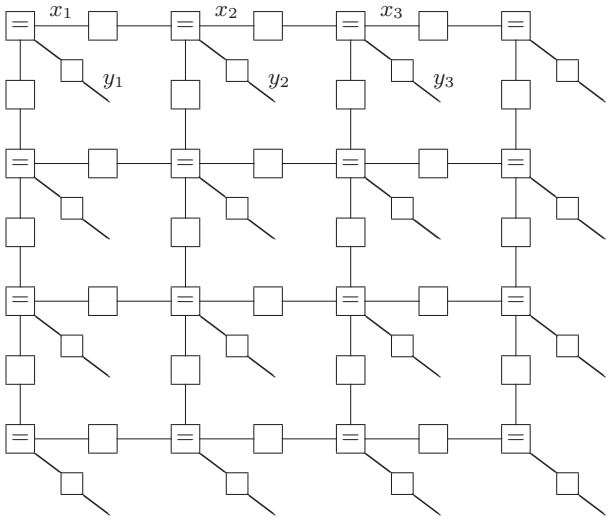

Fig. 2. Extension of Fig. 1 to a factor graph of $p(\mathbf{x}, \mathbf{y})$ with $p(\mathbf{x})$ uniform over the admissible configurations $\mathcal{S}_\mathbf{X}$ and with $p(\mathbf{y}|\mathbf{x})$ as in (6).

The corresponding extension of Fig. 1 is shown in Fig. 2.

In this case, the problem of estimating the mutual information rate reduces to estimating the entropy of the channel output, which is

$$H(\mathbf{Y}) = -\,\mathrm{E}\left[\log(p(\mathbf{Y}))\right]. \tag{7}$$

As in [1], we can approximate the expectation in (7) by the empirical average as

$$H(\mathbf{Y}) \approx -\frac{1}{L}\sum_{\ell=1}^{L} \log(p(\mathbf{y}^{(\ell)})), \tag{8}$$

where samples $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(L)}$ are drawn according to $p(\mathbf{y})$.

The following difficulties remain:

1) Drawing input samples $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(L)}$ from $\mathcal{S}_\mathbf{X}$ according to $p(\mathbf{x})$ and therefrom creating $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(L)}$ according to (6).
2) Computing $p(\mathbf{y}^{(\ell)})$ for each $\ell \in \{1, 2, \ldots, L\}$.

We will compute $p(\mathbf{y}^{(\ell)})$ based on

$$p(\mathbf{y}^{(\ell)}) = \sum_{\mathbf{x}} p(\mathbf{y}^{(\ell)}|\mathbf{x})p(\mathbf{x}) \tag{9}$$

$$= \mathrm{E}\left[p(\mathbf{y}^{(\ell)}|\mathbf{X})\right] \tag{10}$$

where $p(\mathbf{x})$ is a probability mass function on $\mathcal{S}_\mathbf{X}$. (More about this below.)

If $f(\mathbf{x})$ defined in (1) has a cycle-free factor graph representation with not too many states, then generating samples from $p(\mathbf{x})$ can be done efficiently, cf. Appendix A. Moreover, Z defined in (3) can then also be computed efficiently by sum-product message passing [1], [6].

In this paper, however, a different approach is needed. We will propose two methods: an exact sampling method in Section III and a GBP-based method in Section IV.

III. Exact Sampling Method

A. Tree-Based Gibbs Sampling

Tree-based Gibbs sampling (as described in Appendix A) can be applied to draw input samples $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(L)}$ from $\mathcal{S}_\mathbf{X}$ according to $p(\mathbf{x})$. The input samples can then be used to create output samples $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(L)}$.

Tree-based Gibbs sampling can also be applied to compute a Monte carlo estimate of $1/Z$, see [3] and Appendix B.

Tree-based Gibbs sampling mixes much faster than naive Gibbs sampling [3], [7].

B. Importance Sampling

Adapting an idea from [4], an estimate of $p(\mathbf{y}^{(\ell)})$ in (9) can be obtained using importance sampling [8] as follows.

1) Draw samples $\mathbf{x}^{(1,\ell)}, \mathbf{x}^{(2,\ell)}, \ldots, \mathbf{x}^{(K,\ell)}$ from $\mathcal{S}_X$ according to some auxiliary distribution $q^{(\ell)}(\mathbf{x})$ using tree-based Gibbs sampling.
2) Compute

$$\hat{p}(\mathbf{y}^{(\ell)}) = \frac{1}{K}\sum_{k=1}^{K} \frac{p(\mathbf{x}^{(k,\ell)})}{q^{(\ell)}(\mathbf{x}^{(k,\ell)})} p(\mathbf{y}^{(\ell)}|\mathbf{x}^{(k,\ell)}). \tag{11}$$

It is easily verified that $\mathrm{E}\left[\hat{p}(\mathbf{y}^{(\ell)})\right] = p(\mathbf{y}^{(\ell)})$. The special case $q^{(\ell)}(\mathbf{x}) = p(\mathbf{x})$ amounts to uniform sampling.

In the numerical experiments of Section V, we will use

$$q^{(\ell)}(\mathbf{x}) = \frac{p(\mathbf{x})}{Z_q^{(\ell)}} e^{-\alpha d_\mathrm{H}(\mathbf{x}, \mathbf{x}^{(\ell)})}, \tag{12}$$

where d_H denotes the Hamming distance, α is a nonnegative constant, and $\mathbf{x}^{(\ell)}$ is the input that gave rise to the output $\mathbf{y}^{(\ell)}$. Note that $q^{(\ell)}(\mathbf{x})$ has essentially the same factor graph as $p(\mathbf{x})$, which implies that sampling from $q^{(\ell)}(\mathbf{x})$ can also be done by tree-based Gibbs sampling.

The quantity $Z_q^{(\ell)}$ can be computed with tree-based Gibbs sampling using estimators in (21) and (23) with a factor graph partitioning as in Fig. 7 as a by-product of tree-based sampling, see Appendix B.

IV. GBP-Based Method

A. Region Graph

GBP has been shown to provide good approximations for 2-D channels with memory [9]. We start by constructing a region graph which provides the graphical framework for GBP. GBP operates by sending messages between the regions while performing exact computations inside each region.

To construct the region graph, a set of basic regions is required. In our numerical experiments in Section V with a factorization as in (2), we used basic regions of size 2×2 as shown in Fig. 3. By sliding the basic regions along the rows and the columns of the channel, we used the cluster variation method to construct a region graph with counting numbers as illustrated in Fig. 4, see [10], [11] for details. The set of variables involved in each region R is denoted by $\mathbf{x}_R$.

At each region R, the region beliefs $b_R(\mathbf{x}_R)$ were computed after convergence. The region beliefs $b_R(\mathbf{x}_R)$ are GBP approximations to the corresponding marginals of $p(\mathbf{x})$.

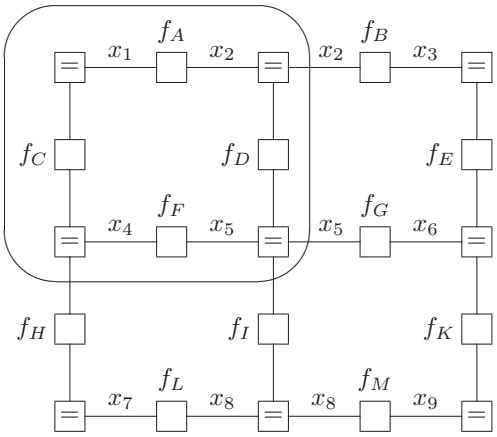

Fig. 3. The basic region to construct the region graph in Fig. 4.

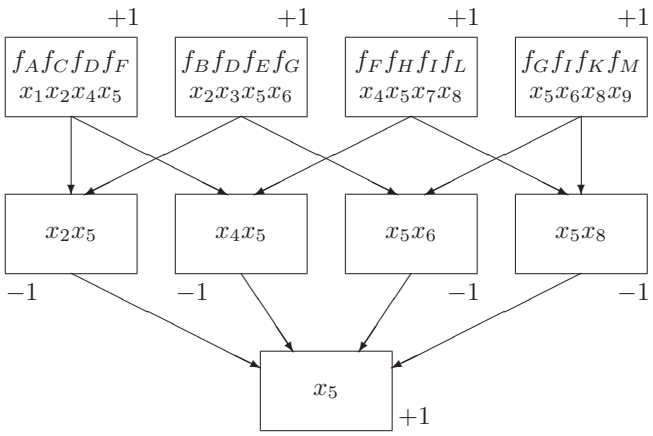

Fig. 4. Region graph with counting numbers constructed from Fig. 3.

B. Sampling and Estimation

Samples $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(L)}$ are generated as follows. Each sample $\mathbf{x}^{(\ell)}$ is generated piecewise sequentially according to the belief $b_R(\mathbf{x}_R)$ in each basic region. For example, in the region graph of Fig. 4, after computing $b_R(x_1, x_2, x_4, x_5)$, sample x_1 is drawn according to $b_R(x_1)$, sample x_2 is drawn according to $b_R(x_2|x_1)$, etc.

In the basic regions, the beliefs are directly proportional to the factor nodes involved in each region, which guarantees that the samples are drawn from $\mathcal{S}_{\mathbf{X}}$. Since beliefs are good approximations to the marginal probabilites, one expects that the samples are drawn from a distribution close to $p(\mathbf{x})$, see [11].

An estimate of $p(\mathbf{y}^{\ell})$ can then be obtained as follows:

1) Draw samples $\mathbf{x}^{(1,\ell)}, \mathbf{x}^{(2,\ell)}, \ldots, \mathbf{x}^{(K,\ell)}$ from $\mathcal{S}_X$ according to $b_R(\mathbf{x}_R)$ of the basic regions.
2) Compute the expectation in (10) by empirical average.

V. Numerical Experiments

In our numerical experiments, we consider a 12×12 grid with input alphabet $\mathcal{X} = \{-1, +1\}$ and with a factorization of f as in (2).

We assume an additive white Gaussian noise channel with noise variance σ^2. We thus have

$$H(\mathbf{Y}|\mathbf{X}) = \frac{N}{2} \log(2\pi e \sigma^2), \tag{13}$$

and $p(\mathbf{y}|\mathbf{x})$ in (6) factors into a product of kernels

$$p(y_i|x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{1}{2\sigma^2} \Big(y_i - x_i\Big)^2 \Big). \tag{14}$$

We define

$$\text{SNR} \triangleq 10 \log_{10} \frac{1}{\sigma^2}. \tag{15}$$

Using the methods of Sections III and IV, we obtain the information rates shown in Fig. 5. The horizontal dotted line in Fig. 5 is the noiseless capacity of this channel [3].

The solid curve in Fig. 5 was obtained by the exact sampling method of Section III. For the SNR values in the interval $[-10, -4]$ dB, the parameter α in (12) was set to zero, which means that importance sampling in (11) reduces to uniform sampling. Higher values of α were used for higher SNR. The convergence of the method is illustrated in Fig. 6. For SNR values above 4 dB, this method suffers from slow convergence.

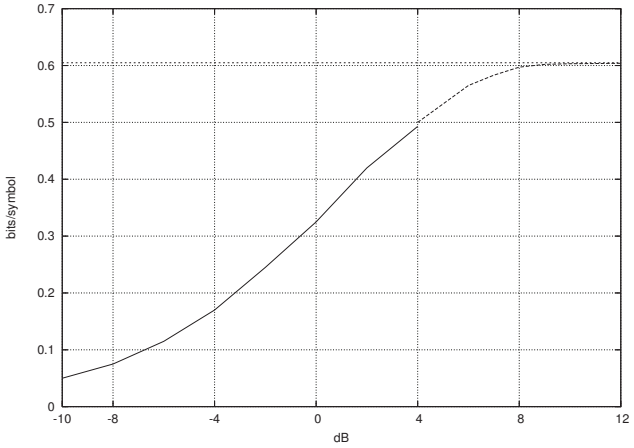

Fig. 5. Estimated information rate (in bits per symbol) vs. SNR (in dB) for a 12×12 channel with a $(1, \infty)$ constraint and additive white Gaussian noise. Solid line: exact sampling method; dotted line: GBP-based method.

The dashed curve in Fig. 5 was obtained by the GBP-based method of Section IV.

VI. Conclusion

We proposed both an exact sampling method and a GBP-based method to compute a Monte Carlo estimate of the information rate of noisy 2-D constrained source / channel models. The exact method has a convergence problem at high SNR. The GBP method yields only approximate results, but the approximation appears to be good and it seems to converge at every SNR.

The methods were demonstrated for a 2-D $(1, \infty)$ constrained channel with additive white Gaussian noise and with a uniform distribution over the admissible channel input configurations. However, the methods are applicable to any joint

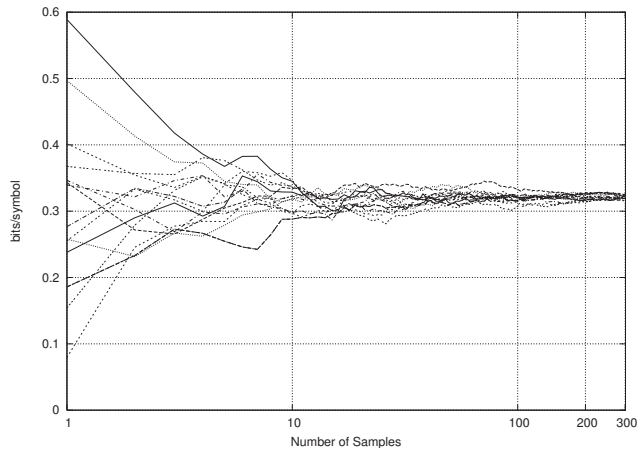

Fig. 6. Estimated information rate (in bits per symbol) vs. the number of samples L for a noisy 12×12 $(1, \infty)$ constraint at zero dB. The plot shows 15 independent sample paths, each with $K = 10^7$ and $\alpha = 0.5$.

source / channel model (including, in particular, nonuniform input) with a suitable factor graph.

ACKNOWLEDGEMENT

The authors wish to thank Giovanni Sabato for implementing the GBP-based method of Section IV.

APPENDIX A
TREE-BASED GIBBS SAMPLING [3], [7]

Let (A, B) be a partition of the index set $\{1, 2, \ldots, N\}$ such that,

- For fixed $\mathbf{x}_A$, the factor graph of $f(\mathbf{x}) = f(\mathbf{x}_A, \mathbf{x}_B)$ is a tree.
- For fixed $\mathbf{x}_B$, the factor graph of $f(\mathbf{x}) = f(\mathbf{x}_A, \mathbf{x}_B)$ is also a tree.

An example of such a partition is shown in Fig. 7.

Starting from some initial configuration $\mathbf{x}^{(0)} = (\mathbf{x}_A^{(0)}, \mathbf{x}_B^{(0)})$, the samples $\mathbf{x}^{(k)} = (\mathbf{x}_A^{(k)}, \mathbf{x}_B^{(k)})$, $k = 1, 2, \ldots$, are created as follows.

First, $\mathbf{x}_A^{(k)}$ is sampled according to

$$p(\mathbf{x}_A | \mathbf{x}_B = \mathbf{x}_B^{(k-1)}) \propto f(\mathbf{x}_A, \mathbf{x}_B^{(k-1)}); \tag{16}$$

then $\mathbf{x}_B^{(k)}$ is sampled according to

$$p(\mathbf{x}_B | \mathbf{x}_A = \mathbf{x}_A^{(k)}) \propto f(\mathbf{x}_A^{(k)}, \mathbf{x}_B). \tag{17}$$

Both sampling steps can be implemented efficiently using the assumed tree structure [3], [7]. Tree-based Gibbs sampling mixes much faster than naive Gibbs sampling.

APPENDIX B
ESTIMATING $1/Z$ USING TREE-BASED GIBBS SAMPLING [3]

Tree-based Gibbs sampling can be used to estimate $1/Z$ (and Z itself) using the following algorithm.

Let

$$f_A(\mathbf{x}_A) \triangleq \sum_{\mathbf{x}_B} f(\mathbf{x}_A, \mathbf{x}_B), \tag{18}$$

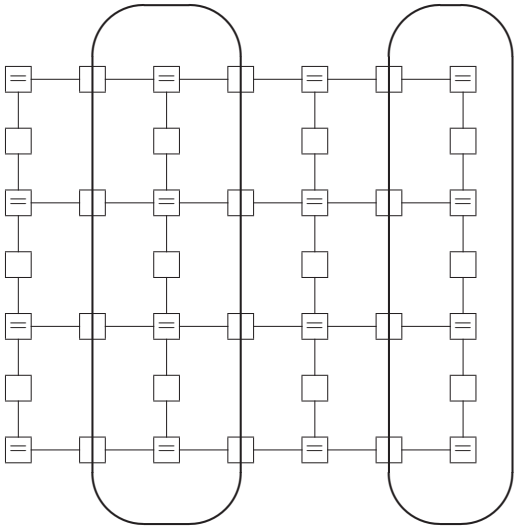

Fig. 7. Partition of Fig. 1 into two cycle-free parts (one part inside the two ovals, the other part outside the ovals).

and

$$f_B(\mathbf{x}_B) \triangleq \sum_{\mathbf{x}_A} f(\mathbf{x}_A, \mathbf{x}_B). \tag{19}$$

Note that

$$\sum_{\mathbf{x}_A} f_A(\mathbf{x}_A) = \sum_{\mathbf{x}_B} f_B(\mathbf{x}_B) = \sum_{\mathbf{x}} f(\mathbf{x}) = Z. \tag{20}$$

With the above assumptions, an estimate Γ_A of $1/Z$ is formed as follows.

1) Draw samples $\mathbf{x}_A^{(1)}, \mathbf{x}_A^{(2)}, \ldots, \mathbf{x}_A^{(K)}$ from $\mathcal{S}_{\mathbf{X}_A}$ according to $p(\mathbf{x}_A) \triangleq \sum_{\mathbf{x}_B} p(\mathbf{x}_A, \mathbf{x}_B) = f_A(\mathbf{x}_A)/Z$.
2) Compute

$$\Gamma_A \triangleq \frac{1}{K|\mathcal{S}_{\mathbf{X}_A}|} \sum_{k=1}^{K} \frac{1}{f_A(\mathbf{x}_A^{(k)})}, \tag{21}$$

where

$$\mathcal{S}_{\mathbf{X}_A} \triangleq \{\mathbf{x}_A : f_A(\mathbf{x}_A) > 0\}. \tag{22}$$

By symmetry, we also have an analogous estimate Γ_B

$$\Gamma_B \triangleq \frac{1}{K|\mathcal{S}_{X_B}|} \sum_{k=1}^{K} \frac{1}{f_B(\mathbf{x}_B^{(k)})} \tag{23}$$

The computation of

$$f_A(\mathbf{x}_A^{(k)}) = \sum_{\mathbf{x}_B} f(\mathbf{x}_A^{(k)}, \mathbf{x}_B), \tag{24}$$

which is required in (21), is easy since the corresponding factor graph is a tree.

The quantity $\mathcal{S}_{\mathbf{X}_A}$ in (22) may be easy to determine, in particular, in our numerical experiments in Section V

$$\mathcal{S}_{\mathbf{X}_A} = \{\mathbf{x}_A : f(\mathbf{x}_A, -\mathbf{1}) > 0\}. \tag{25}$$

In this case, $|\mathcal{S}_{\mathbf{X}_A}| = \sum_{\mathbf{x}_A} f(\mathbf{x}_A, -\mathbf{1})$ is easily computed by sum-product message passing in the (cycle-free) factor graph of $f(\mathbf{x}_A, -\mathbf{1})$.

REFERENCES

[1] D. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3498–3508, August 2006.

[2] H. D. Pfister, J.-B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite-state ISI channels," in *Proc. 2001 IEEE Globecom*, San Antonio, TX, Nov. 2001, pp. 2992–2996.

[3] H.-A. Loeliger and M. Molkaraie, "Estimating the partition function of 2-D fields and the capacity of constrained noiseless 2-D channels using tree-based Gibbs sampling," *Proc. 2009 IEEE Information Theory Workshop,* Taormina, Italy, October 11–16, pp. 228–232.

[4] H.-A. Loeliger and M. Molkaraie, "Simulation-based estimation of the partition function and the information rate of two-dimensional models," *Proc. 2008 IEEE Int. Symp. on Information Theory,* Toronto, Canada, July 6–11, 2008, pp. 1113–1117.

[5] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Proc. Mag.,* Jan. 2004, pp. 28–41.

[6] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory,* vol. 47, pp. 498–519, Feb. 2001.

[7] F. Hamze and N. de Freitas, "From fields to trees," *Proc. Conf. on Uncertainty in Artificial Intelligence,* Banff, July 2004.

[8] D. J. C. MacKay, "Introduction to Monte Carlo methods," in *Learning in Graphical Models,* M. I. Jordan, ed., Kluwer Academic Press, 1998, pp. 175–204.

[9] O. Shental, N. Shental, S. Shamai (Shitz), I. Kanter, A. J. Weiss, and Y. Weiss, "Discrete-input two-dimensional Gaussian channels with memory: estimation and information rates via graphical models and statistical mechanics," *IEEE Trans. Inform. Theory,* vol. 54, pp. 1500–1513, April 2008.

[10] G. Sabato, *Simulation-Based Techniques to Study Two-Dimensional ISI Channels and Constrained Systems.* Master thesis, Dept. Inform. Techn. & Electr. Eng, ETH Zurich, Switzerland, 2009.

[11] J. S. Yedidia, W. T. Freeman, and Y. Weiss. "Constructing free energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inform. Theory,* vol. 51, pp. 2282–2312, July 2005.