Discrete-Input Two-Dimensional Gaussian Channels With Memory: Estimation and Information Rates Via Graphical Models and Statistical Mechanics

Ori Shental, *Member, IEEE*, Noam Shental, Shlomo Shamai (Shitz), *Fellow, IEEE*, Ido Kanter, Anthony J. Weiss, *Fellow, IEEE*, and Yair Weiss

Abstract—Discrete-input two-dimensional (2-D) Gaussian channels with memory represent an important class of systems, which appears extensively in communications and storage. In spite of their widespread use, the workings of 2-D channels are still very much unknown. In this work, we try to explore their properties from the perspective of estimation theory and information theory.

At the heart of our approach is a mapping of a 2-D channel to an undirected graphical model, and inferring its *a posteriori* probabilities (APPs) using generalized belief propagation (GBP). The derived probabilities are shown to be practically accurate, thus enabling optimal maximum *a posteriori* (MAP) estimation of the transmitted symbols. Also, the Shannon-theoretic information rates are deduced either via the vector-wise Shannon–McMillan–Breiman (SMB) theorem, or via the recently derived symbol-wise Guo–Shamai–Verdú (GSV) theorem. Our approach is also described from the perspective of statistical mechanics, as the graphical model and inference algorithm have their analogues in physics.

Our experimental study, based on common channel settings taken from cellular networks and magnetic recording devices, demonstrates that under nontrivial memory conditions, the performance of this fully tractable GBP estimator is almost identical to the performance of the optimal MAP estimator. It also enables a practically accurate simulation-based estimate of the information rate. Rationalization of this excellent performance of GBP in the 2-D Gaussian channel setting is addressed.

Index Terms—Cluster variation method, generalized belief propagation (GBP), Guo–Shamai–Verdú (GSV) theorem, information rate, intersymbol interference (ISI), magnetic recording channels, maximum *a posteriori* (MAP) estimation, multiple-access (MA) channels, Shannon–McMillan–Breiman (SMB) theorem, two-dimensional (2-D) channels.

I. INTRODUCTION

IN discrete-input Gaussian channels with memory, also known as (a.k.a.) dispersive channels, a coded information symbol, chosen from a finite alphabet, is impaired by other transmitted symbols, and by an ambient additive white Gaussian noise (AWGN). The dependence between observations of the transmitted symbols, i.e., the effect of memory or interference, may be formed over time, space, frequency, or any other transmission aspect. The memory may be introduced either by design, as in the case of error-correcting codes, or by nature, as information is dispersed while propagating via the physical channel medium. The memory is primarily related to the dimensionality of the system, i.e., the particular extent and structure of interference.

In this paper, we consider two-dimensional (2-D) channels, which are characterized by symbols ordered, at least logically, on a 2-D grid, causing interference in a limited neighborhood. Such channels play a fundamental role in various applications in modern communications and storage, and several examples of such applications are concisely stated below.

A popular and important instance of this class of channels is 2-D intersymbol interference (ISI) channels, which appear, e.g., in magnetic and optical recording devices [1]. In the 2-D ISI channel, symbols, stored on a magnetic or optical planar medium, interfere with adjacent symbols during their retrieval process.

A second example concerns multiple-access (MA) channels in cellular networks. In a seminal work [2], Wyner introduced a simple, yet insightful, analytically solvable uplink model for cellular networks with continuous Gaussian signaling. This model yields a considerable insight into the ultimate information-theoretic limits of realistic cellular networks (e.g., the third-generation Universal Mobile Telecommunication System standard, UMTS [3]) and commercial wireless local area networks (W-LAN). In addition to a naive one-dimensional (1-D) cell array extension of a single-cell system, Wyner also analyzed the traditional 2-D hexagonal topology, where multiple-access interference (MAI) arises not only between

intra-cell users, but is also caused by users within geographically neighboring cellular tiers. Hence, Wyner's model can be viewed as an instance of dispersive channels. Another example of spatial interference stems from symbols transmitted (received) simultaneously from (in) nearby antennas in an array used in multiple-in multiple-out (MIMO) systems, for instance, in a single-cell base station with a planar antenna array and widely spaced antenna elements or in pixelated wireless optical channels [4].

A thorough analysis of 2-D channels involves both estimation-theoretic and information-theoretic aspects. In the following, the state-of-the-art in these two major facets is reviewed, starting with estimation, then moving to information-theoretic issues.

As in any other dispersive channel, the main task in communication through a 2-D channel is to overcome interference and noise, in order to correctly estimate the transmitted symbols. In the case of ISI this process is termed channel equalization [5], while in MA it is called multiuser detection (MUD, [6]). Optimal estimation, or detection, can be achieved by a maximum *a posteriori* (MAP) joint sequence decision based on the matched filter (MF) outputs of all corrupted symbols, and yields a significant capacity improvement over the conventional MF detector with symbol-wise hard decision. However, as we shall see, its complexity for 2-D channels is exponential in the grid's width, thus being impractical even for rather small channels. In fact, the NP completeness of the 2-D channel detection problem has recently been shown [7].

Hence, various practical suboptimal detection methods have been proposed (e.g., [8]–[15]). For example, Marrow and Wolf [14] evaluated the performance of several iterative estimation methods for the binary-input 2-D ISI channel. Their detection schemes operate iteratively on the rows and columns of the 2-D channel and approximate the optimal detector's bit-error rate (BER) to within 0.5 dB.

As for information-theoretic aspects, the capacity of discrete-input dispersive channels is given as the maximum mutual information rate over all discrete-input distributions. Computing this capacity for 1- and 2-D channels is a long-standing open problem. Calculating the mutual information rate in the case of a predefined stationary input distribution is, in principle, a simpler problem. For example, for input symbols which are independent and identically distributed (i.i.d.) and equiprobable, this is termed the symmetric (a.k.a. uniform-input) information rate (SIR), thus providing a limit on the achievable rate of reliable communication in this common case.

Various bounds, either rigorous [16]–[18], numerical [19], [20], or conjectured [21], on the capacity and SIR of certain discrete-input 1-D dispersive channels have been proposed. Several authors have recently introduced simulation-based methodologies for computing such information rates (Arnold *et al.* [22] and references therein). In this approach, the forward recursion of the sum-product (Bahl–Cocke–Jelinek–Raviv, BCJR) algorithm [23] is used for estimating the *a posteriori* probability (APP) and consequently deriving the 1-D information rates. As for 2-D channels, due to their inherent complexity, Chen and Siegel introduced upper and lower bounds on the information rate [24]–[27].

In spite of the efforts reported in the aforementioned contributions on 2-D channels, neither a near-optimal practical detector nor an accurate, yet tractable, computation of the information rate has been proposed for this important class of channels. In this paper, we apply simulation-based methodologies, originally developed in the interrelated fields of probabilistic graphical models [28]–[30] and statistical mechanics [31], [32], for the analysis of 2-D channels. This analysis yields practically accurate schemes for optimal detection [33] and information rate computation [34].

The basic observation is that 2-D channels can be viewed as an undirected graphical model, a.k.a. pairwise Markov random field (MRF, [35]), and that performing inference in the graphical model yields the required estimation-theoretic and information-theoretic values. As our inference engine we apply a fully tractable generalized belief propagation (GBP, [36], [37]) algorithm, which yields a practically accurate APP. The marginal probabilities are then used for detection, and also for estimating the information rate, via the Guo–Shamai-Verdú (GSV) theorem. As an alternative approach, the information rate is also estimated using the joint APP, via the Shannon–McMillan–Breiman (SMB) theorem, where the connection to statistical mechanics is established.

Graphical models provide powerful tools for exact (optimal) or approximate inference. For example, loopy belief propagation (LBP, sum–product algorithm, [38]) is an efficient way to solve inference problems in graphical models, which has been shown to serve remarkably well as a decoding engine [39] in low-density parity-check (LDPC, [40]) codes and turbo codes [41]–[43]. One-dimensional channels, depicted by a tree graph, can be also accurately analyzed, as was shown by Arnold *et al.* [22], using LBP which is equivalent in this case to the BCJR algorithm. Hence, our approach can be viewed as an extension of its 1-D Monte Carlo counterpart, where GBP replaces LBP, due to the latter's poor performance on loopy graphs.[1]

Our findings are of practical and theoretical interest. They may enable more reliable communications through the 2-D channel and denser data storage devices, along with a better understanding of the theoretical limits of such systems.

The paper is organized as follows. Section II introduces the dispersive 2-D channel model, while Section III presents its connection to graphical models. Section IV discusses the issue of exact and approximate inference in the graphical model. The outputs of our inference algorithm are both the marginal and the joint probability distributions. In Section V, we apply the marginal probabilities for performing detection and for estimating the information rate via the GSV theorem. In Section VI, we use the joint probability distribution for estimating the information rate via the SMB theorem, which leads to the connection to statistical mechanics. Section VII provides thorough simulation results for both detection and information rate of 2-D ISI and MA channels in various topologies. The remarkable performance of GBP in this problem is also discussed in Section VII. We conclude the paper in Section VIII.

---

[1]Recently, Pakzad and Anantharam [44] have applied GBP to the problem of joint decoding of a LDPC code and a 1-D ISI channel. The performance of GBP in this 1-D case was shown to outperform the best conventional iterative method.

We shall use the following notations. The operator $\{\cdot\}^T$ denotes a vector or matrix transpose, the matrix $\boldsymbol{I}_N$ is an $N \times N$ identity matrix, while the symbols $\{\cdot\}_i$ and $\{\cdot\}_{ij}$ denote entries of a vector and matrix, respectively. The operators $\mathbb{E}\{\cdot\}$ and $\|\cdot\|$ stand for expectation and Euclidean norm of a vector, respectively, while $\log$ is the natural logarithm.

II. CHANNEL MODEL

Consider an $N \times N$ 2-D discrete-input channel with memory in the form

$$\mathsf{y}_{k,l} = \mathsf{x}_{k,l} + \mathsf{v}_{k,l} + \sum_{(i,j) \in \langle k,l \rangle} \alpha_{i,j,k,l} \mathsf{x}_{i,j}, \quad \forall k,l = 1, \dots, N \tag{1}$$

where $\mathsf{y}_{k,l}$, the channel's output observation at symbol $(k,l) \in \mathbb{Z}^2$, is the sum of the finite-alphabet input symbol $\mathsf{x}_{k,l}$, assumed to be taken from a stationary process, and two additional terms. The first term $\mathsf{v}_{k,l}$ represents the common ambient AWGN, while the second term is the scaled interference caused by adjacent symbols to (k,l), denoted by $\langle k,l \rangle$. The parameter $\alpha_{i,j,k,l}$ $(|\alpha_{i,j,k,l}| \leq 1)$[2] controls the interference attenuation. The interference term is assumed to be spatially invariant[3] (excluding boundary symbols, as we do not assume periodic boundary conditions), which together with the assumptions regarding $\mathsf{x}_{k,l}$ and $\mathsf{v}_{k,l}$ guarantees that $\mathsf{y}_{k,l}$ $(k,l = 1, \dots, N)$ are stationary random variables.[4] We also assume that the channel (including the noise variance) is perfectly known on the receiver's side, which can jointly process all observations.

Stacking all the observations, data symbols, and noise samples into $N^2 \times 1$ vectors $\boldsymbol{y}$, $\boldsymbol{x}$, and $\boldsymbol{v}$, respectively, the channel (1) can be rewritten as

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{v} \tag{2}$$

where the $N^2 \times N^2$ square matrix $\boldsymbol{S}$ encapsulates the memory/interference structure, which uniquely defines the channel. Our basic assumption, which later allows for a graphical model interpretation, is that interference is caused by neighboring symbols, i.e., matrix $\boldsymbol{S}$ is a relatively sparse matrix.

The upper pane in Fig. 1 represents the interference structure of four 2-D topologies: two ISI examples (Fig 1(a) and (b)), a rectangular cellular network (Fig. 1(c)), and the typical Wyner hexagonal cellular network (Fig. 1(d)). In all four examples, nine symbols (full nodes denoting bits in this case) are ordered on $N \times N = 3 \times 3$ grids, which differ in the way the bits interfere with each other. In the ISI examples, where the memory is unidirectional, an information bit is interfered with preceding bits along the same row and column (Fig 1(a)) and also along the diagonal (Fig 1(b)). In the case of cellular networks the memory is bidirectional, hence, a bit is corrupted by both preceding and succeeding bits along its row and column (rectangular network,

Fig 1 (c)) and also along the diagonal (hexagonal network, Fig 1(d)). Note that ISI in 2-D storage systems can also exhibit bidirectional memory effects.

In the following derivations, real-space data signaling $\boldsymbol{x}$, interference $\boldsymbol{S}$, and noise $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N)$ are assumed. An extension to the complex domain is straightforward.

The joint posterior probability distribution of the channel is

$$\Pr(\boldsymbol{x}|\boldsymbol{y}) \propto \Pr(\boldsymbol{x}) \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right) \tag{3}$$

where $\Pr(\boldsymbol{x})$ is the prior input distribution. Hereinafter, for purposes of exposition only, we consider an equiprobable i.i.d. binary-input alphabet, i.e., $\boldsymbol{x} \in (-1, +1)^{N \times N}$. The binary alphabet is of interest, in both multiuser frameworks, as well as the magnetic recording setting. Hence, the conditional distribution (3) in this case takes the following form:

$$\Pr(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left(-\frac{1}{\sigma^2}\left(\sum_{(i>j)} R_{ij}x_i x_j - \sum_i h_i x_i\right)\right) \tag{4}$$

where the symmetric matrix $\boldsymbol{R} = \boldsymbol{S}^T \boldsymbol{S}$ is the interference cross-correlation matrix, and the vector $\boldsymbol{h} = \boldsymbol{S}^T \boldsymbol{y}$ is the output vector of a filter matched to the channel's interference structure. The notation $(i > j)$ stands for a summation over all nonzero entries in the upper triangular of the symmetric cross-correlation matrix $\boldsymbol{R}$.

In the remainder of this paper, the main goal is to efficiently calculate either the joint posterior probability $\Pr(\boldsymbol{x}|\boldsymbol{y})$ or its marginal probabilities $\Pr(x_i|\boldsymbol{y}) = \sum_{\boldsymbol{x} \backslash x_i} \Pr(\boldsymbol{x}|\boldsymbol{y})$, where summation is performed over all symbols except x_i. In order to do so, we adopt the methodology of graphical models.

III. THE CONNECTION TO UNDIRECTED GRAPHICAL MODELS

Undirected graphical models appear in many fields of research, e.g., spin systems in statistical physics [45], computer vision [46], digital communications, and error correcting codes [30], [47]. An undirected graphical model with pairwise potentials (a.k.a. pairwise MRF [35]), consists of a graph G and real positive potential functions $\psi_{ij}(x_i, x_j)$ and $\phi_i(x_i)$ such that the probability of an assignment $\boldsymbol{x}$ is given by

$$\Pr(\boldsymbol{x}) \propto \prod_{(i>j)} \psi_{ij}(x_i, x_j) \prod_i \phi_i(x_i). \tag{5}$$

Here, the previously defined notation $(i > j)$ represents the set of all edges connecting pairs of nodes in the graph. The implicit assumption is that $(i > j)$ contains only relatively small sets of variables, such that the dependencies between the variables are "local" (as opposed to "global" dependency as in, for example, a fully connected graph.) The model is undirected since the potentials do not contain causal dependencies among the variables as in the case of directed graphical models (i.e., Bayesian networks [28]).

Hence, the posterior probability distribution (4) defines the following undirected graphical model:

$$\Pr(\boldsymbol{x}|\boldsymbol{y}) \propto \prod_{(i>j)} \psi_{ij}(x_i, x_j) \prod_i \phi_i(x_i, h_i) \tag{6}$$

---

[2] This limit on α arises from the physical model, but is not crucial to our approach.

[3] Assuming $\alpha_{i,j,k,l} = \alpha_{k-i,l-j}$ being spatially invariant, the summation in the channel model (1) is simply a convolution.

[4] Note that stationarity is needed only in order to enable the proper definition of the mutual information rate. For estimation purposes, stationarity does not have to be assumed (e.g., see Section VII-A where $\alpha_{i,j,k,l}$ are taken randomly).
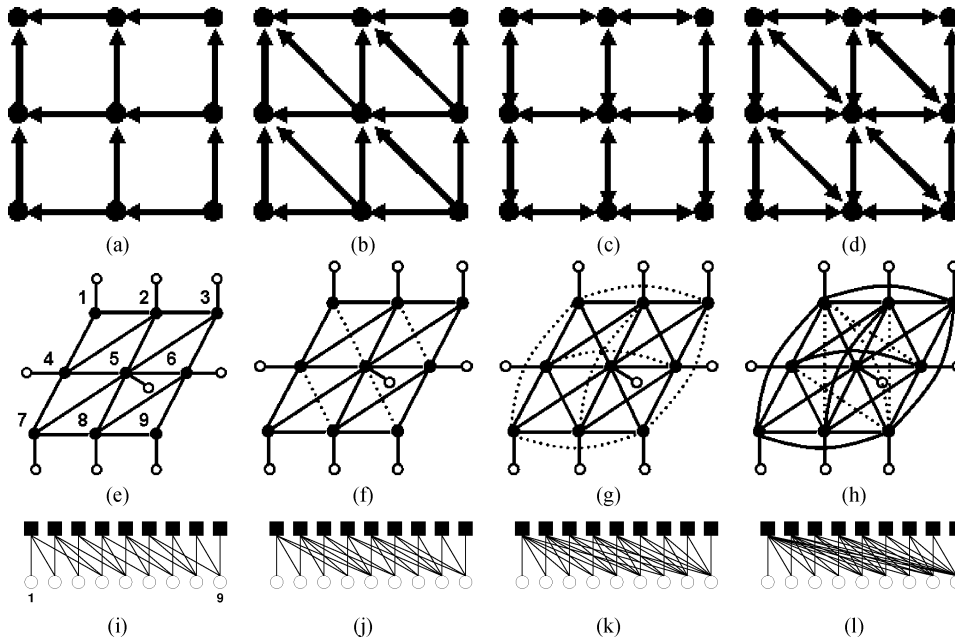
Fig. 1. **Upper pane**: Interference structures for four types of 3×3 2-D channels: (a) and (b) ISI grids; (c) rectangular cellular network; (d) hexagonal Wyner cellular network. The arrows mark the direction of interference. **Middle pane**: The corresponding undirected graphical model representation of the channels in the upper pane: (e) and (f) ISI grids; (g) and (h) rectangular and hexagonal cellular networks, respectively. Full nodes represent (hidden) transmitted bits, while empty nodes correspond to the observations. Interaction couplings (compatibility functions) ψ_{ij} are denoted by a solid line connecting two full nodes, while the external field potential (evidence) ϕ_i is depicted by a solid line connecting a full node and an empty node. For clarity we use dotted edges in (f), (g), and (h) to represent the extra edges added compared to the graphs (e), (f) and (g), respectively. **Lower pane**: The corresponding factor graph representation of the channels in the upper pane: (i) and (j) ISI grids; (k) and (l) rectangular and hexagonal cellular networks, respectively. Following common notation, the variables are denoted by empty circles and the local factor functions by full squares. The factor graph's variables and functions are enumerated from left to right, corresponding to a left-right/up-down numbering of the nodes in the 2-D grids.

where

$$\psi_{ij}(x_i, x_j) = \exp\left(-\frac{R_{ij}x_i x_j}{\sigma^2}\right) \qquad (7)$$

is a compatibility function (interaction coupling) i.e., "potential" representing the structure of the system, and the potential

$$\phi_i(x_i, y_i) = \exp\left(\frac{h_i x_i}{\sigma^2}\right) \qquad (8)$$

is the "evidence" (local likelihood or external field), which describes the statistical dependency between the hidden variable x_i and the observed variable h_i.[5]

The middle pane in Fig. 1 presents the corresponding graphical model of our four application examples: ISI (Fig. 1(e) and (f)), a rectangular cellular network (Fig. 1(g)), and a hexagonal cellular network (Fig. 1(h)). Full nodes represent transmitted bits, while empty nodes correspond to the observations. Compatibility functions ψ_{ij} are denoted by a solid line connecting

---

[5]In the case of a nonequiprobable i.i.d binary-input alphabet the MRF modeling is identical, except for an additional external field potential operating on each hidden variable node, which can be absorbed into ϕ_i. The same occurs for the nonbinary discrete-input alphabet case, where an external field potential, which arises from the autocorrelations R_{ii}, is added and should be absorbed into ϕ_i. In the case of correlated input (e.g., a Markov process), additional general potentials appear, and the graph may no longer be defined as a pairwise MRF, although the proposed approach still holds.

two full nodes, while the evidence ϕ_i is depicted by a solid line connecting a full node and an empty node. Fig. 1 also displays (lower pane) the corresponding factor graphs, c.f. [47].

Estimating the joint posterior probability (4) or its marginals is termed "inference," and the field of graphical models has developed tools for either calculating them exactly or approximately. In the next section, we present the intractability of exact inference over 2-D channels, and then discuss the issue of approximate inference using belief propagation and its generalizations.

IV. INFERENCE IN GRAPHICAL MODELS

A. Exact Inference

The canonical algorithm for exact inference is the junction-tree algorithm [48]. The algorithm converts a general graph into an equivalent tree whose nodes contain clusters (cliques) of nodes of the original graph. Inference can then be performed by passing messages forward and backward among neighboring cliques in the tree. The complexity of the algorithm is exponential in the size of the largest clique in the derived tree, which for $N \times N$ grid-like graphs, such as our 2-D problem, is N times the memory depth of the channel. Hence, exact inference becomes impractical for large grids, or even for moderate size graphs (e.g., a small network of tens of cells) with nonbinary modulation, or for nontrivial memory effects. Therefore, one must resort to approximate inference methods.
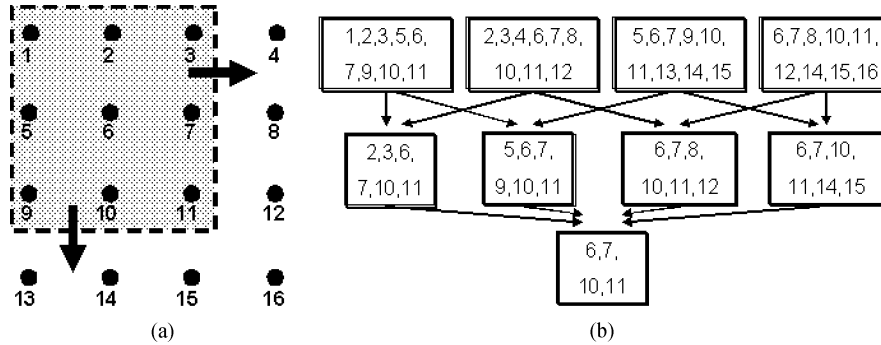
Fig. 2.   (a) Covering a 4×4 nodes graph using a 3×3 basic region in an overlapping left-and-right up-and-down sliding-window manner, and (b) the corresponding region graph, with two-region intersection levels.

B. Approximate Inference

1) Loopy Belief Propagation: Loopy belief propagation (LBP) is equivalent to applying Pearl's local message-passing algorithm [38], originally derived for trees, to a general graph even if it contains cycles (loops). As mentioned previously, LBP has been found to have outstanding empirical success in many applications, e.g., in decoding turbo codes and LDPC codes. The excellent performance of LBP in these applications may be attributed to the sparsity of the graphs, which ensures that cycles in the graph are long, and inference may be performed as if it were a tree.

Can LBP be used over dispersive 2-D channels? Our studies show that LBP frequently fails to converge (for both synchronous and asynchronous message update scheduling) and therefore its associated performance is poor. This result is not surprising since, in contrast to the sparse graphs of LDPC codes, the graph of 2-D channels contains many short cycles. As a result LBP's implicit tree-like assumption does not hold, and its approximation is poor. In order to recover the near-optimal characteristics of message-passing inference one must circumvent this problem.

2) Generalized Belief Propagation: The generalized belief propagation (GBP) algorithm [37] is an extension of LBP, which has been shown to provide better approximations than LBP in many applications. In GBP, messages are sent between clusters (regions) of nodes, in contrast to LBP, where messages are passed along the graph's edges. The GBP algorithm has the nice property of reducing to LBP when taking any pair of nodes connected by an edge on the MRF to be a region.

The "region graph" approach to GBP begins by defining "basic" regions, which completely cover the graph and include all linked nodes. Let $\mathcal{R}$ be this set of regions, their intersections, the intersections of the intersection, and so on. Based on the set $\mathcal{R}$, a "region graph" of the graphical model can be composed, along which messages are passed in an analogous way to LBP. An example of such basic regions and the resulting region graph appears in Fig. 2.

The initial issue in applying GBP concerns the selection of the basic regions. A smart choice of the basic regions will be such that it encompasses all nodes along the shortest loops. Since inference within a basic cluster is performed exactly, we may avoid the short loops which probably caused LBP not to converge.

As for complexity, it is striking to reveal that the GBP involves only minimally more computations than LBP, as the complexity of GBP grows exponentially only with the size of the chosen basic region. Enlarging the basic region often entails a more accurate inference at the cost of complexity. For an elaborate presentation of GBP, see Yedidia *et al.* [37], and more specifically [37, Appendix E] for the "two-way GBP algorithm" which we apply.

Can GBP be applied to 2-D channels? Since the graphical models of our 2-D channels contain interactions between nearest neighbors and next-nearest neighbors, as displayed in Fig. 1(e)–(h), a natural choice of overlapping regions is a sliding 3×3 square of nodes (e.g., see Fig. 2(h)), which were used in all of our simulations. This size of basic regions is small enough to perform exact inference within each region, yet still can well capture the interaction between neighboring nodes. Our empirical study shows that the GBP message-passing algorithm, which is fully tractable, yields remarkable results in 2-D channels, even for harsh interference conditions.

V. MARGINAL PROBABILITIES-BASED ANALYSIS

A. Detection

The marginal probabilities $\Pr(x_i|\boldsymbol{y})$ estimated by GBP are applied in two contexts. The first and more straightforward application is detection, as the estimated symbol is simply $\arg\max_{x_i} \Pr(x_i|\boldsymbol{y})$. As described in Section VII, our detected symbols are practically identical to optimal detection, which basically reflects the accuracy of estimated marginal probabilities.

B. Information Rate: Guo–Shamai–Verdú (GSV) Theorem

The second application of the marginal probabilities is to estimate the symmetric information rate (SIR) via the recently derived GSV theorem [49]. The GSV theorem reveals a simple, yet powerful relationship between information and estimation theories, as it bridges over the notions of Shannon's mutual information and minimum mean-square error (MMSE) for the common AWGN channel.

The GSV theorem for the vector Gaussian channel may be stated in the following way. Consider a real-valued channel with

input and output random variable vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, of the form

$$\boldsymbol{y} = \sqrt{\mathsf{snr}}\, \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \qquad (9)$$

where $\mathsf{snr} \geq 0$ is the channel's signal-to-noise ratio (SNR), $\boldsymbol{H}$ is a deterministic $L \times K$ matrix, and $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_L)$ is a standard Gaussian noise independent of $\boldsymbol{x}$. The input $\boldsymbol{x}$ and the output $\boldsymbol{y}$ are column vectors of appropriate dimensions. The input–output mutual information $I(\boldsymbol{x}; \boldsymbol{y})$ and the MMSE in estimating $\boldsymbol{H}\boldsymbol{x}$ denoted by $\mathsf{mmse}(\mathsf{snr})$ maintain the following relation.

GSV Theorem: [49, Theorem 2] For every input distribution $P(\boldsymbol{x})$ that satisfies $\mathbb{E}\|\boldsymbol{x}\|^2 < \infty$

$$\frac{d}{d\mathsf{snr}} I(\boldsymbol{x}; \boldsymbol{y}) = \frac{1}{2} \mathsf{mmse}(\mathsf{snr}). \qquad (10)$$

Hence, by adapting the GSV theorem to our problem, we may evaluate

$$\mathsf{mmse}(\mathsf{snr}) = \mathbb{E}\{\| \boldsymbol{S}\boldsymbol{x} - \boldsymbol{S}\hat{\boldsymbol{x}}(\boldsymbol{y}; \mathsf{snr}) \|^2\} \qquad (11)$$

via the marginal probabilities $\Pr(x_i|\boldsymbol{y})$ estimated by GBP, where the conditional mean estimate is defined by

$$\hat{\boldsymbol{x}}(\boldsymbol{y}; \mathsf{snr}) \triangleq \{\hat{x}_1, \dots, \hat{x}_{N^2}\}^T$$

with

$$\hat{x}_i \triangleq \Pr(x_i = 1|\boldsymbol{y}) - \Pr(x_i = -1|\boldsymbol{y}) \qquad (12)$$

and $\mathsf{snr} = 1/\sigma^2$. The expectation over all instances of the input $\boldsymbol{x}$ and output $\boldsymbol{y}$ in calculating the MMSE (11) is performed empirically by averaging over a sufficiently large 2-D channel grid. Integrating over snr one gets the desired mutual information $I(\boldsymbol{x}; \boldsymbol{y})$. Now the information rate is given by the mutual information per symbol over the sufficiently large symbol block.

VI. Joint Probability-Based Analysis

The second approach to estimating the information rate corresponds to estimating the joint probability distribution of the channel. This is a complementary approach to the former GSV approach, which also draws an interesting connection to statistical mechanics, as described below.

The information rate, i.e., mutual information per symbol, between the channel's input $\mathcal{X}$ and output $\mathcal{Y}$ is

$$I(\mathcal{X}; \mathcal{Y}) = h(\mathcal{Y}) - h(\mathcal{Y}|\mathcal{X}) \qquad (13)$$

where $h(\cdot)$ are (differential) entropy rates, where, by definition, the entropy rate $h(\mathcal{Q})$ of a stationary process $\boldsymbol{q} = \{q_1, \dots, q_L\}^T$ is given by $\lim_{L \to \infty} h(\boldsymbol{q})/L$. Let us deal separately with the two terms of the information rate (13).

The second term, $h(\mathcal{Y}|\mathcal{X})$, is given by $\lim_{N \to \infty} h(\boldsymbol{y}|\boldsymbol{x})/N^2$, but since $h(\boldsymbol{y}|\boldsymbol{x}) = h(\boldsymbol{v})$, it is straightforward to validate that $h(\mathcal{Y}|\mathcal{X}) = (\log 2\pi e\sigma^2)/2$. As for the term $h(\mathcal{Y})$ we apply the

SMB theorem,[6] which states that for a stationary and ergodic channel the entropy rate can be calculated by

$$-\frac{1}{N^2} \log p(\boldsymbol{y}) \xrightarrow{N \to \infty} h(\mathcal{Y}) \text{ with probability } 1 \qquad (14)$$

where $p(\boldsymbol{y})$ is the joint distribution of the channel's output $\boldsymbol{y}$. Hence, in order to calculate the information rate one needs to calculate $p(\boldsymbol{y})$ in the limit of large systems.

Applying Bayes' law and using the channel model (1), $p(\boldsymbol{y})$ can be rewritten as

$$p(\boldsymbol{y}) = \sum_{\boldsymbol{x}} p(\boldsymbol{y}|\boldsymbol{x}) \Pr(\boldsymbol{x}) = \sum_{\boldsymbol{x}} p_{\boldsymbol{v}}(\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}) \Pr(\boldsymbol{x}) \qquad (15)$$

where $\sum_{\boldsymbol{x}}$ corresponds to a sum over all the possible values of the transmitted symbols $\boldsymbol{x}$. As we consider the case of an equiprobable i.i.d. binary-input alphabet, using the AWGN's probability density function $p_{\boldsymbol{v}}(\cdot)$, the joint distribution of the channel's output (15) can be rewritten as

$$p(\boldsymbol{y}) = \mathcal{Z}(\boldsymbol{y}) \cdot (2\mathcal{C})^{-N^2} \qquad (16)$$

where $\mathcal{C} \triangleq (2\pi\sigma^2)^{1/2}$, and

$$\mathcal{Z}(\boldsymbol{y}) \triangleq \sum_{\boldsymbol{x}} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right) \qquad (17)$$

is the partition function of the channel's joint posterior distribution (3).

Inserting (16) into the entropy rate (14), $I(\mathcal{X}; \mathcal{Y})$, in nats, can be written as

$$\log 2 - 1/2 + \mathcal{F} \xrightarrow{N \to \infty} I(\mathcal{X}; \mathcal{Y}) \text{ with probability } 1 \qquad (18)$$

where

$$\mathcal{F} \triangleq -\frac{1}{N^2} \log(\mathcal{Z}(\boldsymbol{y})) \qquad (19)$$

is recognized as the free energy per symbol [31], [55]. Hence, the problem of calculating the symmetric information rate boils down to estimating the free energy of an infinite system. The next section provides the statistical mechanics view of 2-D channels and draws the connection to GBP.

A. A Statistical Mechanics View of 2-D Channels and the Connection to GBP

The connection between 2-D channels and basic models in statistical mechanics stems from the 2-D channel's model. The posterior probability (4) actually corresponds to a Boltzmann distribution of an Ising Hamiltonian, with pairwise interactions R_{ij} and external random fields h_i [31].

The difficulty in estimating the posterior probability (4) lies in estimating the partition function (17), or similarly, the free energy (19). The field of statistical mechanics has devoted considerable effort to the development of methods

---

[6]Proof of the SMB theorem for multidimensional channels, extending its 1-D version [50], can be found in Orenstein and Weiss [51] and the references therein (e.g., for processes indexed by $\mathbb{Z}^d$, $d \geq 2$[52] or even for all discrete amenable groups [53]). The SMB theorem also applies to continuous random variables [54].
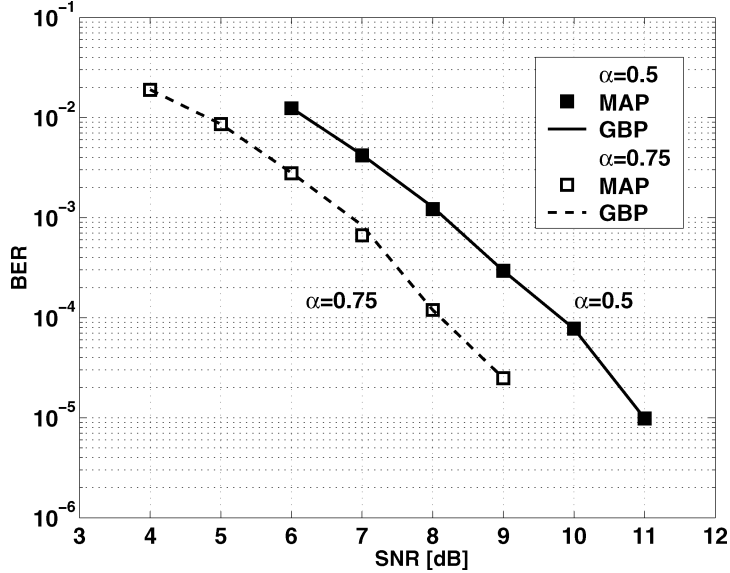
Fig. 3. 2-D ISI equalization in a 6×6 channel for $\alpha_{i,j,k,l} = \alpha = 0.5$ and $\alpha = 0.75$. Equalization performance of GBP (line) versus optimal (MAP, squares) detectors in terms of average BER per symbol, as a function of SNR. The full squares and solid line correspond to $\alpha = 0.5$, while empty squares and a dashed line correspond to $\alpha = 0.75$.

for calculating the free energy. However, evaluating the free energy of infinitely large 2-D channels (1) is infeasible due to the intractability of computing $p(\boldsymbol{y})$, and one must resort to approximate methods.

One of the classic approximation methods for free energies is the Kikuchi approximation, also known as the cluster variation method (CVM, [37]). The CVM follows a variational principle: It defines the free energy as a functional of $p(\boldsymbol{y})$, $\mathcal{F}(p(\boldsymbol{y}))$, replaces $p(\boldsymbol{y})$ by a tractable trial belief vector $b(\boldsymbol{y})$, and seeks to minimize $\mathcal{F}(b(\boldsymbol{y}))$. The trial belief vector in CVM is taken as $b(\boldsymbol{y}) \triangleq \prod_{\lambda \in \mathcal{M}} p(\boldsymbol{y}_\lambda)^{c_\lambda}$, where λ is a 'cluster' of "neighboring" observations $\boldsymbol{y}_\lambda$, taken from the set of "clusters" $\mathcal{M}$. The integers c_λ, a.k.a. counting numbers, are provided by the CVM in order to ascertain that each symbol is counted exactly once in the corresponding free energy. Since $b(\boldsymbol{y})$ depends only on local probabilities, $p(\boldsymbol{y}_\lambda)$, its computation is tractable.[7] Then, minimizing the functional $\mathcal{F}(b(\boldsymbol{y}))$ w.r.t $b(\boldsymbol{y})$, yields an approximation to the free energy.

Recently, Yedidia *et al.* [37] have shown a correspondence between the stationary points of the CVM-based free energy and the fixed points of GBP. Note, in passing, that Yedidia *et al.*, have also shown that the fixed points of LBP correspond to the stationary points of the Bethe free energy, which is a special case of the CVM, in the same way that belief propagation (BP) is a special case of GBP. For an elaborate discussion of CVM and its relation to GBP, see Yedidia *et al.* [37].

Hence, our idea for estimating the information rate is to use GBP to find the minimum (or stationary points) of the CVM of a large enough, yet finite, system, thus assuming that the computed free energy per symbol converges to its exact value for infinite systems.

VII. RESULTS AND DISCUSSION

A. Estimation

The performance of the proposed GBP detector was evaluated using Monte Carlo simulations of three examples of the dispersive 2-D channel: a 2-D ISI channel, and two topologies of a cellular network—rectangular and hexagonal (see Fig. 1(a), (c), and (d), respectively).

The performance of GBP was compared to the optimal detector and to several other standard detectors. GBP with a standard[8] parallel (synchronous) message-passing schedule was applied. Convergence is achieved when the total difference between beliefs along iterations is lower than 10^{-8}. In many of our experiments, the LBP detector did not converge in a substantial percentage of the simulations, for both synchronous (parallel) and asynchronous (serial) message scheduling,[9] in contrast to the consistent convergence of GBP. As a result, its performance was worse than an MF detector and in certain cases was close to a random guess. Hence, LBP's results are omitted from the performance evaluation figures.

1) 2-D ISI Channel: Following a recent contribution [14], we simulated a 6×6 binary ISI channel, under two nontrivial constant interference levels, $\alpha_{i,j,k,l} = \alpha = 0.5$ and $\alpha = 0.75$. Outside the grid the symbols were assumed to have value (-1).

Fig. 3 compares the equalization performance of the optimal detector and GBP in terms of average BER per symbol as a function of SNR.[10]

---

[7]However, $b(\boldsymbol{y})$ need not necessarily form a valid probability distribution function [37].

[8]GBP always converged without the need to use "damping," i.e., averaging old and newly computed messages.

[9]The stopping criteria used for LBP were much looser than the one for GBP.

[10]There are different reasonable definitions of SNR for channels with ISI. Besides the definition we use, i.e., $\mathsf{snr} = 1/\sigma^2$, it is also common to define $\mathsf{snr} = \| h \|^2 / \sigma^2$, where h is the Kronecker delta impulse response of the channel. Evidently, the comparison of the error and information rates, under different α values and SNR may change depending on the definition of the latter.
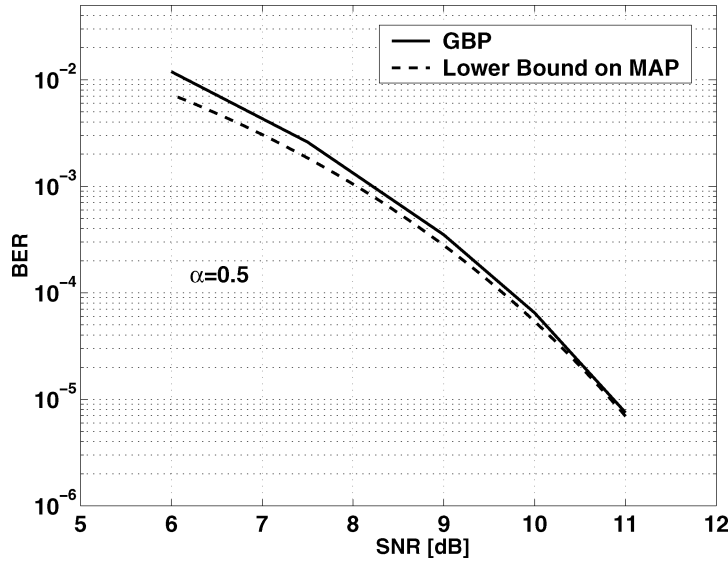
Fig. 4. 2-D ISI equalization in a 20×20 channel for $\alpha = 0.5$. Equalization performance of a GBP detector (solid line) versus lower bound (dashed line) on optimal (MAP) detector error probability in terms of average BER per symbol, as a function of SNR.
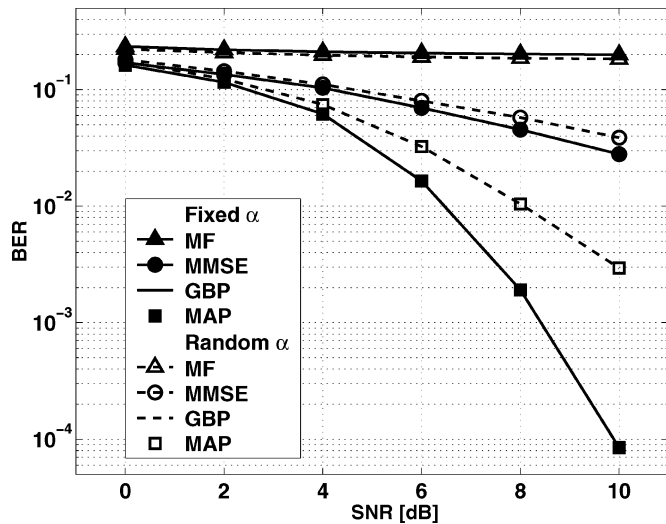

Fig. 5. MUD in a 9×9 rectangular topology. Average BER per cell as a function of SNR, for the optimal (MAP, squares) and GBP (line) detectors. Also shown are the BER for the linear MMSE (circles and line) and single-user MF (triangles and line) detectors. The BER was evaluated for two channel profiles of fixed ($\alpha_{i,j,k,l} = \alpha = 0.5$, full markers and solid lines) and random interference ($\alpha \sim 0.5 \cdot \mathcal{N}(0,1)$ and hard-limited at ± 1, empty markers and dashed lines).
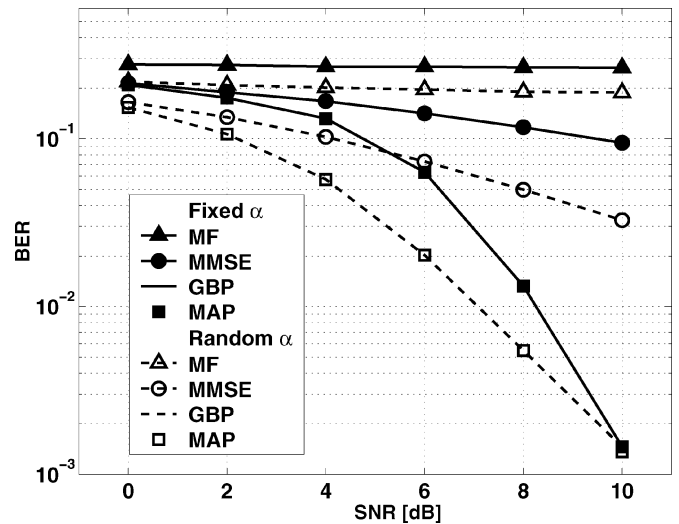

Fig. 6. MUD in a 9×9 hexagonal topology. Average BER per cell as a function of SNR, for the optimal (MAP, squares) and GBP (line) detectors. Also shown are the BER for the linear MMSE (circles and line) and single-user MF (triangles and line) detectors. The BER was evaluated for two channel profiles of fixed ($\alpha_{i,j,k,l} = \alpha = 0.5$, full markers and solid lines) and random interference ($\alpha \sim 0.5 \cdot \mathcal{N}(0,1)$ and hard-limited at ± 1, empty markers and dashed lines).

As can be seen, the GBP error decreases with SNR and its performance coincides with MAP error levels or is extremely close to it. The standard deviation of the results (including those in subsequent sections) was small, thus here, and hereinafter it is omitted from the figures (unless otherwise specifically stated). For comparison, note that the performance of the best iterative detector proposed by Marrow and Wolf [14] for this setting is approximately 2/3 dB above the optimal BER. To avoid confusion, note that in Figs. 3–7, the GBP curves have no markers, but the MAP performance points—represented by markers alone—fall exactly on top of the GBP lines.

The performance of the GBP detector in a similar setting, but for a larger 20×20 channel with $\alpha = 0.5$, was also evaluated. As MAP equalization is infeasible for such a system, GBP's performance was compared to an analytical lower bound on the optimal error probability. This bound was obtained by assuming that all interfering bits are known at the receiver's side, thus transforming the dispersive channel into a memoryless one, where optimal detection is achieved by using a simple MF. Fig. 4 displays the equalization performance of GBP and the lower bound on the optimal error probability. The results show that the GBP detector's BER is very close to the bound,
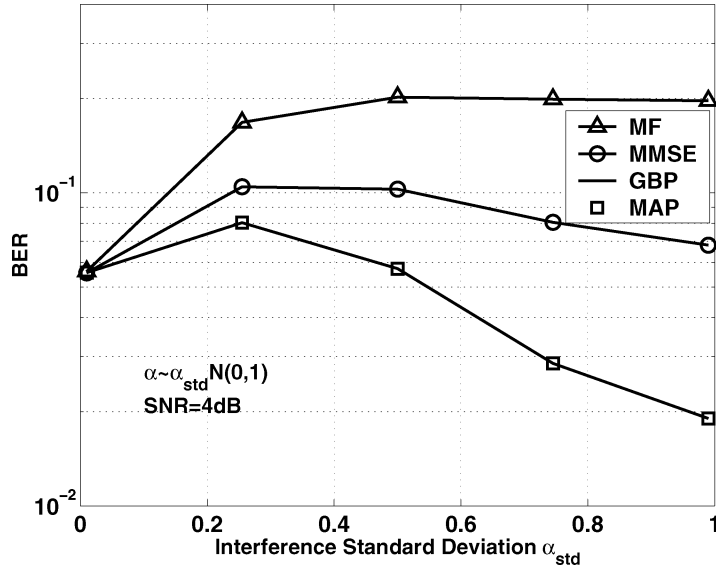
Fig. 7. MUD in 9×9 hexagonal topology over a range of random interferences. Average BER per cell as a function of random interference standard deviation α_{std}, for the optimal (MAP, squares), GBP (solid line), MMSE (circles and solid line), and MF (triangles and solid line) detectors. The BER was evaluated for a random interference scenario $\alpha \sim \alpha_{std} \cdot \mathcal{N}(0, 1)$ (hard-limited at ± 1) and a fixed SNR (= 4 dB).

which implies that its performance is also near-optimal for this system.

Moreover, apart from providing the correct hard decisions, GBP infers the marginal probabilities which may become useful in coded systems. We observed empirically in all our experiments that the marginal beliefs in GBP well-approximate the APP. Thus, GBP may also operate, e.g., as a detection stage in an iterative joint detection and decoding scheme.

2) Rectangular Topology Cellular Network: A cellular network of 9×9 cells[11] with rectangular topology (Fig. 1(c)) was also simulated. Intercell interference scaling was either constant $\alpha_{i,j,k,l} = \alpha = 0.5$, or taken from a zero-mean Gaussian distribution with standard deviation $\alpha_{std} = 0.5$, hard-limited at $\alpha = \pm 1$. The latter corresponds to a random MAI which may be interpreted as being caused by channel fading.

Fig. 5 presents the performance of the optimal MAP detector, the GBP-MUD, and also two other standard detectors—the linear MMSE and the naive single-user MF detectors [6]. (Note that the MF and MMSE BER in Figs. 5–7, in contrast to the GBP and MAP BER, are represented by curves composed of lines and markers.) It may be observed that GBP practically coincides with the optimal MUD for both constant and random interference scenarios, while MMSE and MF are substantially inferior.

Note that another attractive property of the GBP detector in the Wyner-like cellular MUD context is its potential decentralized implementation. GBP's messages between neighboring regions (or cells) may be implemented in the network itself, instead of being computed in a central processor, as imposed by the other detectors.

3) Hexagonal Topology Cellular Network: Another instance of the 2-D model, where a 9×9 cellular network is planned according to a hexagonal topology (Fig. 1(d)), was studied empirically. As in the rectangular case, the inter-cell interference

[11]This was the largest system for which exact inference was possible.

was either constant or random. Fig. 6 compares the GBP-MUD to the optimal MAP, and to MMSE and MF detectors, showing similar performance as in the rectangular case.

We also evaluated the performance of the GBP-MUD in an additional complementary setting, in which the SNR level was fixed (4 dB), and the interference scaling α was varied. We used $\alpha \sim \alpha_{std} \cdot \mathcal{N}(0, 1)$, hard-limited at $\alpha = \pm 1$, where $0 \leq \alpha_{std} \leq 1$. Fig. 7 displays the performance of GBP, MAP, MMSE, and MF detectors in this setting. GBP is extremely close to the optimum for all interference levels, even for harsh ($\alpha \to 1$) conditions.

As for computation, note the enormous advantage in using a GBP detector with complexity dominated by the order of the number of all discrete-input combinations within the basic region (e.g., $\mathcal{O}(2^9)$ for the systems examined in Figs. 3–7), in contrast to the $\mathcal{O}(|\aleph|^{N \times N})$ complexity of the optimal MAP detector, where $|\aleph|$ denotes the size of the input alphabet. As shown in the experimental study illustrated in Figs. 3–7, this significant reduction in complexity almost does not affect the BER. The suboptimal MF and MMSE detectors are of polynomial complexity, namely, $\mathcal{O}(N \times N)$ and $\mathcal{O}(\{N \times N\}^3)$, respectively, however resulting in considerably inferior error performance.

To conclude this section, we show how well the marginal beliefs (and not only the BER) inferred by GBP approximate the true (MAP) marginals. For a 9×9 hexagonal topology network with interference scaling $\alpha = 0.5$, the absolute value of the difference between the marginals computed using GBP and optimal MAP detectors is obtained for each of the 81 nodes. The maximal difference (out of the 81 results) is then recorded for 100 simulation rounds. Fig. 8 plots the median over this maximal difference record, i.e.,

$$\delta = \mathrm{med}(\max_{i=1,\ldots,81} |\Pr(x_i = 1|\boldsymbol{y})^{\mathrm{MAP}} - \Pr(x_i = 1|\boldsymbol{y})^{\mathrm{GBP}}|)$$
$$(20)$$

as a function of SNR. As expected, based on the above BER results, the difference is miniscule, on the order of 10^{-3} and
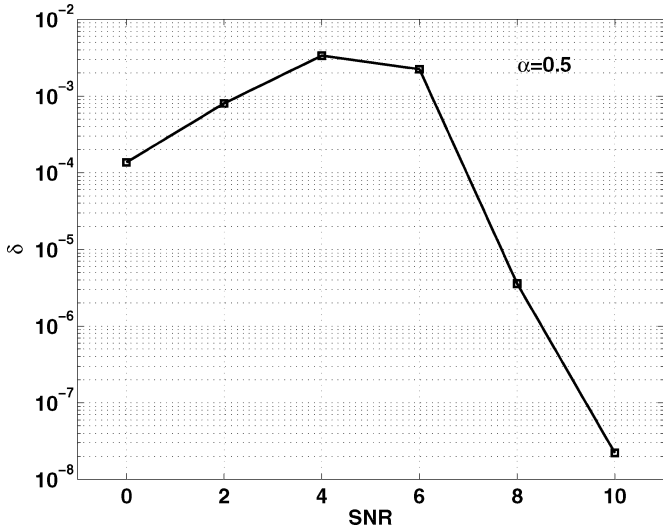
Fig. 8. The median (over 100 rounds), δ, of the maximal difference between the marginals computed using GBP and optimal MAP detectors versus SNR in 9×9 hexagonal topology with $\alpha = 0.5$.
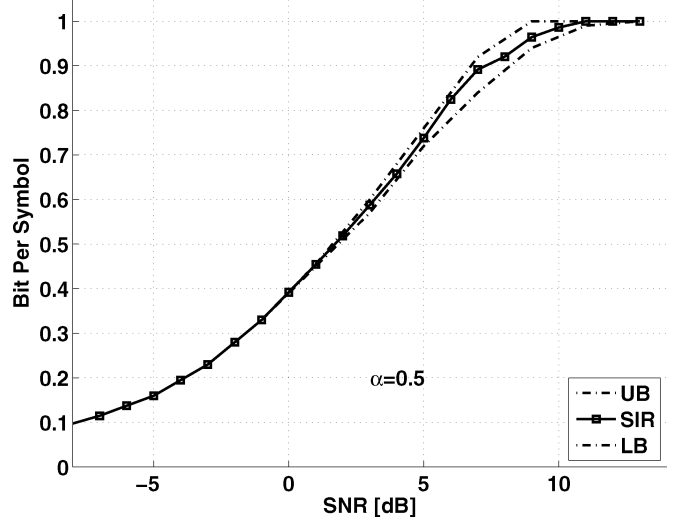

Fig. 9. 2-D ISI channel. Estimated SIR, in terms of bit per symbol, evaluated using GBP-based simulations (squares and a solid line), as a function of SNR for $\alpha = 0.5$. Also shown are upper (UB) and lower (LB) bounds (dashed–dotted) on the SIR [24], [25].

less. As LBP did not converge, in either synchronous or asynchronous modes, for a majority of the simulation rounds in the hexagonal topology case, it is omitted from the figure.

B. Information Rate

The proposed GBP-based algorithm is also used for estimating the SIR of two examples of dispersive 2-D channels: a 2-D ISI channel and a hexagonal Wyner cellular network. The two aforementioned schemes for computing the information rate, i.e., the symbol-wise GSV-based and the vector-wise SMB-based, were applied. Both approaches gave practically the same results, labeled as "SIR" in the following figures. Since GBP well-approximates the joint probability, and its subsequent marginals, both schemes may equally be applied (and were applied in our experimental study, producing the same information rates). The following results were obtained by averaging over 1000 realizations of 30×30 channels.[12]

1) 2-D ISI Channel: Fig. 9 presents the estimated SIR, in terms of bit per symbol, computed using the GBP-based algorithms, as a function of SNR, for a binary ISI channel with nontrivial ($\alpha = 0.5$) interference structure as depicted in Fig. 1(b). Also drawn are the lower and upper bounds on the SIR, recently suggested by Chen and Siegel [24], [25]. As can be seen, the evaluated SIR increases with SNR up to the trivial 1-bit Shannon bound. The estimated SIR curve falls within these tight bounds and agrees with them perfectly.

We would like to comment that recently another simulation-based method for estimating the mutual information of a certain instance of 2-D channels in optical data storage systems, known as the full-surface channel, was presented [56]. In this method, the joint APP is approximated using a heuristic greedy algorithm which has a stochastic element. However, our approach seems to be more promising in obtaining good estimates of the information rates.
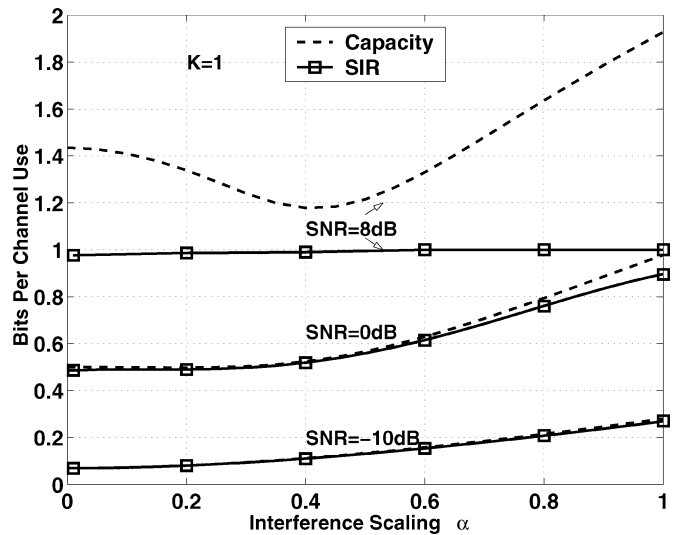

Fig. 10. 2-D Hexagonal Wyner cellular network. Estimated SIR (squares and solid line) and Gaussian signaling capacity (dashed line), in bits per channel use, as a function of the inter-cell interference scaling α for three SNR values: -10, 0, and 8 dB, and a single user within each cell ($K = 1$).

2) 2-D Hexagonal Wyner Cellular Network: In a similar way, the estimated SIR of a hexagonal topology Wyner model (as conceptually described in Fig. 1(d)), under binary signaling, with a single user within each cell (i.e., $K = 1$ in Wyner's original notation [2], describing the case of intra-cell time-division multiple access (TDMA) or orthogonal code-division multiple access (CDMA)) was computed. Fig. 10 displays the estimated SIR calculated for the possible range of inter-cell interference scaling α, for three SNR levels. For comparison, the Gaussian signaling capacity as derived by Wyner, is also presented. As may be expected for low SNR (-10 dB) the estimated SIR and Wyner's capacity almost coincide. For the intermediate SNR level (0 dB), Wyner's capacity provides a tight upper bound on the SIR for $\alpha < 0.5$. As for high SNR (8 dB), the SIR saturates the 1-bit bound for almost all values of α. The SIR, as opposed

---

[12]A grid width of 30 was chosen based on our free-energy approximation quality analysis, as discussed in Section VII-C.
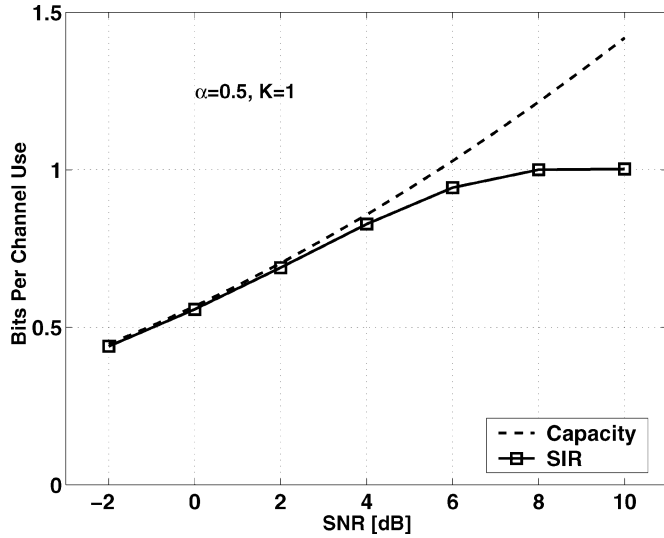
Fig. 11. 2-D Hexagonal Wyner cellular network. Estimated SIR (squares and solid line) and Gaussian signaling capacity (dashed line), in bits per channel use, as a function of SNR for $\alpha = 0.5$ and $K = 1$.
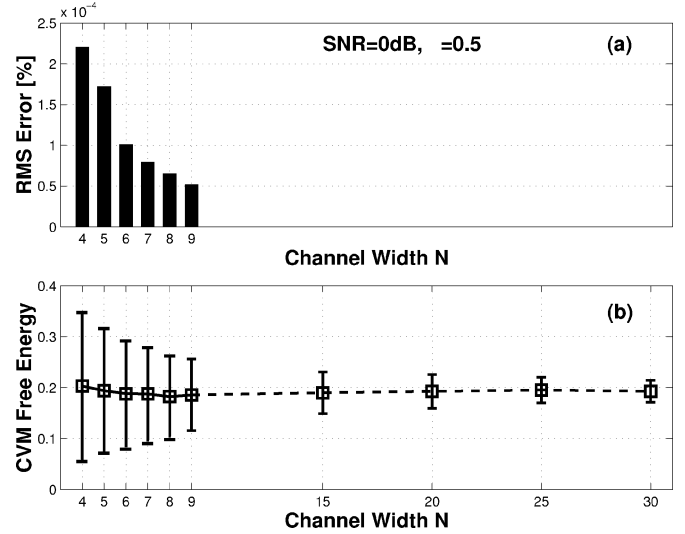

Fig. 12. Quality of the free-energy approximation. (a) Root mean square (rms) error (in %) in computing the free energy per symbol exactly and using the CVM, for $N \times N$ channels. The results were obtained using 500 realizations of Wyner's hexagonal cellular networks (assuming a single user per cell), with $\alpha = 0.5$ and SNR = 0 dB. (b) The corresponding CVM free energy per symbol as a function of N. For $N \leq 9$ we used the same realizations as in (a). For larger systems (dashed line) the exact free energy cannot be calculated. Vertical bars stand for standard deviation in simulation results.

to the capacity, always increases with the inter-cell interference scaling α.

Note, in passing, that since the capacity of a binary channel is bounded between the SIR and Wyner's Gaussian signaling capacity, one can also infer the ultimate (possibly nonuniform) information rate in these low and intermediate SNR regimes. Also note that a proper optimal joint processing of the channel observations yields an increase in the achievable error-free information throughput as the interference intensity grows. Thus, interference, conceived at first as having a detrimental effect, may assist in the process of reliable information transfer.

Fig. 11 evaluates the SIR for the complementary case of a fixed $\alpha = 0.5$ as a function of SNR. The estimated SIR coincides with Wyner's capacity for SNR $\lesssim 4$ dB.

It should be emphasized that the analysis could have been performed in a similar manner for the case of several intra-cell users, i.e., $K > 1$. In this case, the sum of all K intra-cell users' (scaled) transmissions is composed of $K + 1$ discrete amplitude levels. Thus, the scenario of K intra-cell users with equiprobable binary signaling can be seamlessly replaced by the setting of a single intra-cell user with $(K+1)$-ary signaling taken from a binomial distribution.

C. Quality of the Free-Energy Approximation

Our free-energy approximations were based on simulating channels of size 30×30. In this subsection, we aim to justify this choice. In order to evaluate the quality of our free-energy approximation we first performed Monte Carlo simulations of several channels with different sizes and compared our results to an exact calculation. The following results, averaged over 500 realizations, were obtained for a specific channel, i.e., Wyner's hexagonal cellular network, as depicted in Fig. 1(d), with $\alpha = 0.5$, SNR of 0 dB, and assuming a single user per cell.

Fig. 12(a) displays the root mean square (rms) error per symbol, in percentage, between the approximated and exact free energies as a function of the channel's size N^2 (width

$N = 4, \ldots, 9$, where a 9×9 channel is the largest case for which exact computation was feasible.) As can be observed, the difference between the approximated and exact free energies is miniscule (in the order of $10^{-4}\%$) and further decreases with the channel's width.

Fig. 12(b) presents the CVM approximation of the free energy per symbol as a function of the channel's width. It may be seen that the free energy per symbol converges with the size of the system, and that the differences among realizations become smaller.

In principle, larger systems can be simulated, for which these differences are expected to be even smaller. However, it seems that a 30×30 channel size suffices as an effective approximation of the exact free energy per symbol of infinite size systems, thus providing a proper estimate of the information rate. Similar behavior was observed for all other channels throughout the entire interference range and for a wide scope of SNR.

D. Remarks About GBP and 2-D Channels

In this subsection, we aim to justify the use of GBP for 2-D channels and try to understand the reasons for its excellent performance. We first compare its performance to two straightforward alternatives.

- **LBP Over a Galois Field (GF)**

  The system is coarsened using 3×3 (nonoverlapping) blocks. We apply LBP, using symbols with 2^9 states (i.e., LBP over GF($q = 2^9$), [57]), and estimate the free energy using the Bethe approximation. Detection is then performed by marginalizing a single block's length-2^9 belief vector. We refer to this method as Coarsened LBP (CLBP). The main difference between CLBP and GBP is the selection of the basic regions. While GBP uses a sliding 3×3 window, the regions of CLBP are nonoverlapping. Hence,
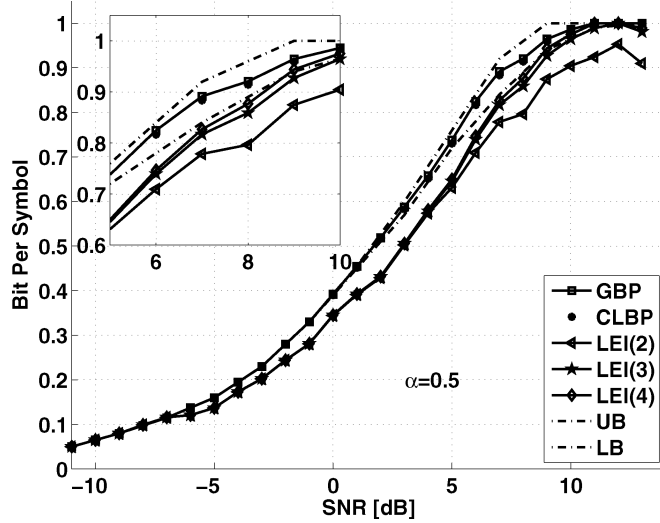
Fig. 13. 2-D ISI channel. Estimated SIR in terms of bit per symbol as a function of SNR for $\alpha = 0.5$. The SIR is evaluated using GBP (squares and a solid line), CLBP (circles), LEI of width ± 2 (triangles and a solid line), ± 3 (stars and a solid line), or ± 4 (diamonds and a solid line). Also shown are upper (UB) and lower (LB) bounds (dashed–dotted) on the SIR. The inset displays an enlargement of the curves.

better performance of GBP over CLBP may indicate the importance of the region selection in GBP. The complexity of CLBP is similar to GBP.

- **Local Exact Inference**

  In this local estimate method we perform exact inference over a certain square region around each symbol, and ignore the rest of the system. The width of a region was ± 2, ± 3, or ± 4, where the latter corresponds to a 9×9 system centered around the relevant symbol. Estimation is performed for each symbol separately, and the free energy corresponds to a mean-field approximation [45], as the joint probability distribution is completely factorized. We refer to this method as Local Exact Inference (LEI). LEI is highly nonefficient as the computation is performed independently for each symbol and we include it in order to check the necessity of passing messages.

Fig. 13 presents the information rate as a function of SNR, for an ISI channel (Fig. 1(b)) with interference level of $\alpha = 0.5$, together with the known bounds [24], [25]. The information rate is estimated using three methods: GBP, CLBP (with parallel message-passing schedule), and LEI. One observes that GBP and CLBP yield an identical estimate, which falls within the theoretical bounds, while LEI falls outside the bounds starting from SNR of -7 dB. Therefore, local estimates are adequate only for low SNR values.

Our empirical study has shown that the difference between GBP and CLBP appears only when the interference conditions are more harsh. Fig. 14 presents the percentage of 9×9 systems (Fig. 1(b), SNR = 5 dB) over which CLBP has converged as a function of α. As α becomes larger, CLBP tends not to converge, thus providing poor detection and information rate estimates, while GBP converges for all α. It appears that at increased values of α the system is "rescaled" such that CLBP effectively suffers the same problems as a regular LBP over lower values of α.
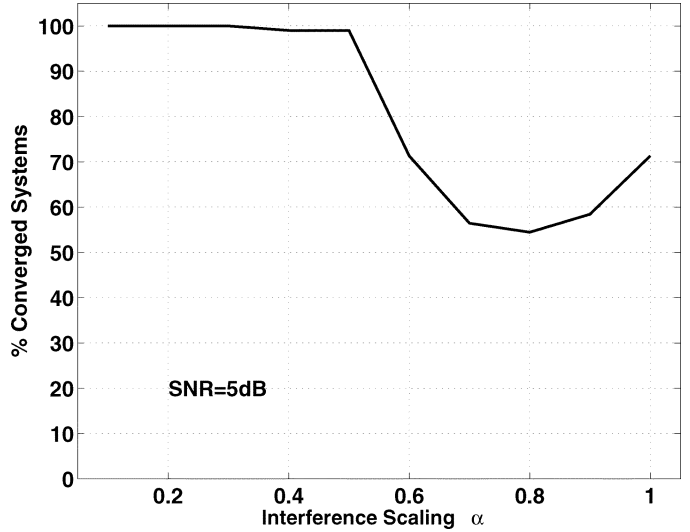

Fig. 14. Convergence of CLBP. Percentage of 2-D ISI systems (SNR = 5 dB) over which CLBP has converged as a function of the interference scaling α.

1) A Necessary Condition for GBP to Succeed: The physics and graphical models literature does not provide a rigorous explanation for the quality of approximations of GBP-based CVM for general graphs, or even for graphs representing 2-D channels.

We conjecture that the reason for the success of GBP stems from the local nature of the system. More specifically, we claim that the correlation length ξ, which measures the range over which symbols are correlated, is of the order of a GBP region over the relevant range of SNR. In order to show that we first examined the classical 2-D Ising ferromagnet spin-glass model without external fields. Fig. 15(a) presents the difference, Δ, between the exact free energy of an infinite system[13] and the CVM free energy of a 30×30 system.[14] Fig. 15(b) presents the 2-D Ising ferromagnet's correlation length as a function of SNR. The correlation length in this case is the "typical" distance over which two spin sites tend to be correlated and is calculated as

$$\xi = 1/\log(\lambda_0/\lambda_1) \qquad (21)$$

where λ_0 and λ_1 are the largest and second largest eigenvalues of the transfer matrix, respectively. The transfer matrix $\boldsymbol{M}$ in this case is a $2^4 \times 2^4$ matrix where M_{ij} is proportional to the probability that a 2×2 block is in a certain (four-symbol) state i and its adjacent block is in state j [59]. One observes that the CVM successfully approximates the free energy, as long as the correlation length is close to the block size, i.e., $\xi = 1$ which corresponds to a single block.

As for the case of 2-D channels with memory, there is no simple way of estimating the correlation length, due to the random external fields. Hence, in order to supply an estimate we calculate the correlation length for a 30×30 hexagonal topology (Fig. 1(d)) in the following way. We take each pair of adjacent 3×3 blocks and calculate the correlation length as if they were replicated indefinitely.

---

[13]A closed-form solution for the exact free energy of the 2-D Ising ferromagnet was obtained by Onsager in his renowned work of 1944 [58].

[14]In this case, the GBP regions were taken as a sliding 2×2 window, which is the natural choice for nearest neighbor interactions.
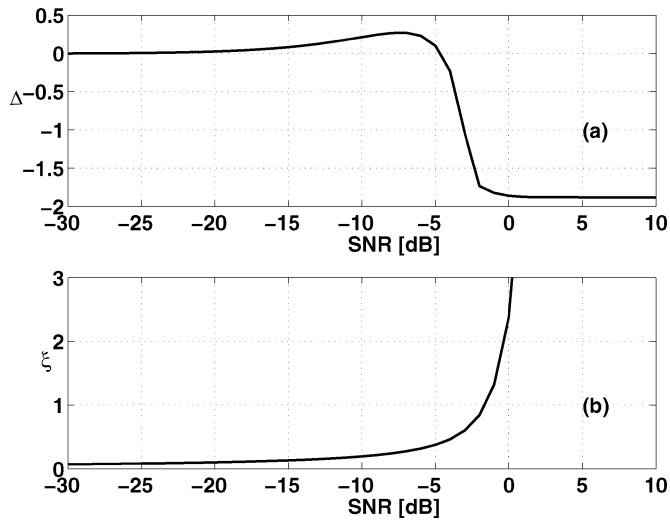
Fig. 15. 2-D Ising ferromagnet (without external fields). (a) The difference Δ between the exact free energy of an infinite system and the CVM free energy of a 30×30 system. (b) Correlation length ξ (in 2×2 block units) as a function of SNR.
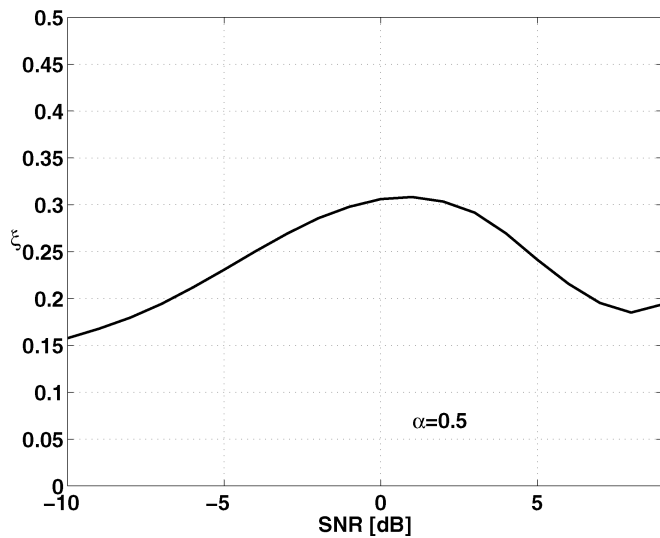

Fig. 16. Mean "correlation length" ξ (in 3×3 block units) over all pairs of adjacent blocks as a function of SNR ($\alpha = 0.5$).

Therefore, due to this replication, the 2-D channel can be now viewed as a classical 2-D Ising spin-glass model for which the correlation length is known (21), except here the eigenvalues are of a $2^9 \times 2^9$ transfer matrix. Now, the correlation length estimate can be computed by averaging over the correlation lengths of the different replicated 2-D channels, corresponding to all the pairs of adjacent blocks in the original 2-D channel's graph.

Fig. 16 presents the mean "correlation length" as a function of SNR ($\alpha = 0.5$). One observes that this value is lower than 1 for all SNR, which probably accounts for the excellent CVM approximation. The reason for the low correlation length, even at high SNR values, is the positive effect of external fields, originating from the observations. These effects can be related to the pinning and percolation theories in the physics literature [60].

VIII. CONCLUSION

A GBP estimator for 2-D channels with memory was introduced. This estimator arose from the frameworks of probabilistic graphical models and statistical mechanics. Extensive simulation results of four different 2-D memory types strongly indicate a practically optimal behavior of GBP in inferring the channel's both marginal and joint posterior probabilities. The proposed fully tractable message-passing scheme then shows a near-optimal error performance in detecting the channel input symbols. Based on this observation, two complementary methods for a simulation-based computation of the information rates of 2-D discrete-input channels with memory are developed. In the first method, the inferred marginals are used to evaluate the MMSE, then the GSV theorem is being applied. The second approach is based on the computed joint APP and the SMB theorem.

In order to validate our method, we compared the resulting information rate to previously calculated bounds. The quality of the free-energy approximation, which is identical (according to the previously derived relation (18)) to the quality of the information rate approximation was compared to the exact free energy using small grid-size channels and was found to exhibit practically accurate behavior. The error performance of GBP is found to be almost optimal throughout the whole SNR and interference ranges. As rigorous analysis of the workings of GBP is a challenging open problem, we have suggested an explanation for its remarkable performance. The attractive usage of GBP illustrated for 2-D channels naturally calls for its real-time implementation as an inference engine in communication and storage systems.

REFERENCES

[1] B. Vasic and E. M. Kurtas, Eds., *Coding and Signal Processing for Magnetic Recording Systems*. Boca Raton, FL: CRC, 2005.
[2] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1713–1727, Nov. 1994.
[3] The 3rd Generation Partnership Project Website 2003 [Online]. Available: http://www.3gpp.org
[4] S. Hranilovic and F. R. Kschischang, "An indoor wireless optical MIMO channel using pixelated arrays of transmitters and receivers," in *Proc. 22nd Bienn. Symp. Communications*, Kingston, ON, Canada, May 2004, pp. 392–394.
[5] J. G. Proakis, *Digital Communications*, 4th ed. New York, USA: McGraw-Hill, 2000.
[6] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[7] E. Ordentlich and R. M. Roth, "On the computational complexity of 2D maximum-likelihood sequence detection," Hewlett-Packard Labs, Palo Alto, CA, 2006, Tech. Rep. HPL-2006-69 .
[8] P. S. Kumar and S. Roy, "Two-dimensional equalization: Theory and application to high density magnetic recording," *IEEE Trans. Commun.*, vol. 42, no. 2, pp. 386–395, Feb. 1994.

[9] W. Weeks, IV, "Full Surface Data Storage," Ph.D. dissertation, Univ. Illinois at Urbana-Champaign, Urbana, IL, 1996.

[10] K. M. Chugg, X. Chen, and M. A. Neifeld, "Two-dimensional equalization in coherent and incoherent page-oriented optical memory," *J. Opt. Soc. Amer. A*, vol. 16, pp. 549–562, Mar. 1999.

[11] N. Singla, J. A. O'Sullivan, R. S. Indeck, and Y. Wu, "Iterative decoding and equalization for 2-D recording channels," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2328–2330, Sep. 2002.

[12] Y. Wu, J. A. O'Sullivan, R. S. Indeck, and N. Singla, "Iterative detection and decoding for separable two-dimensional intersymbol interference," *IEEE Trans. Magn.*, vol. 39, no. 4, pp. 2115–2120, Jul. 2003.

[13] A. H. J. Immink *et al.*, "Signal processing and coding for two-dimensional optical storage," in *Proc. IEEE Global Conf. Communications (GLOBECOM)*, Piscataway, NJ, Dec. 2003, pp. 3904–3908.

[14] M. Marrow and J. K. Wolf, "Iterative detection of 2-dimensional ISI channels," in *Proc. IEEE Information Theory Workshop (ITW)*, Paris, France, Mar. 2003, pp. 131–134.

[15] N. Singla and J. A. O'Sullivan, "Joint equalization and decoding for nonlinear two-dimensional intersymbol interference channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Adelaide, Australia, Sep. 2005, pp. 1353–1357.

[16] W. Hirt, "Capacity and Information Rates of Discrete-Time Channels with Memory," Ph.D. dissertation, Swiss Federal Inst. Technol. (ETH), Zurich, Switzerland, 1998.

[17] S. Shamai (Shitz), L. H. Ozarow, and A. D. Wyner, "Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1527–1539, Nov. 1991.

[18] J. Chen and P. H. Siegel, "Markov processes asymptotically achieve the capacity of finite-state intersymbol interference channels," *IEEE Trans. Inf. Theory*, accepted for publication.

[19] A. Kavčić, "On the capacity of Markov sources over noisy channels," in *IEEE Global Conf. Communications (GLOBECOM)*, San Antonio, TX, Nov. 2001, pp. 2997–3001.

[20] S. Yang and A. Kavčić, "Markov sources achieve feedback capacity of finite-state machine channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Lausanne, Switzerland, Jun./Jul. 2002, p. 361.

[21] S. Shamai (Shitz) and R. Laroia, "The intersymbol interference channel: Lower bounds on capacity and precoding loss," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1388–1404, Sep. 1996.

[22] 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. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[23] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 284–287, Mar. 1974.

[24] J. Chen and P. H. Siegel, "On the symmetric information rate of two-dimensional finite state ISI channels," in *Proc. IEEE Information Theory Workshop (ITW)*, Paris, France, Mar. 2003.

[25] J. Chen and P. H. Siegel, "On the symmetric information rate of two-dimensional finite state ISI channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 227–236, Jan. 2006.

[26] J. Chen and P. H. Siegel, "Information rates of two-dimensional finite state ISI channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Yokohama, Japan, Jun. 2003, p. 118.

[27] J. Chen and P. H. Siegel, "Information rates of two-dimensional finite state ISI channels," *IEEE Trans. Inform. Theory*, submitted for publication.

[28] M. I. Jordan, Ed., Learning in Graphical Models   Cambridge, MA, MIT Press, 1999.

[29] M. I. Jordan and C. M. Bishop, Eds., An Introduction to Graphical Models, 2000, draft.

[30] B. J. Frey, *Graphical Models for Machine Learning and Digital Communication*.   Cambridge, MA: MIT Press, 1998.

[31] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*.   Singapoer: World Scientific, 1987, vol. 9, Lecture Notes in Physics.

[32] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing*.   Oxford, U.K.: Oxford Univ. Press, 2001.

[33] O. Shental, N. Shental, A. J. Weiss, and Y. Weiss, "Generalized belief propagation receiver for near-optimal detection of two-dimensional channels with memory," in *Proc. IEEE Information Theory Workshop (ITW)*, San Antonio, TX, Oct. 2004.

[34] O. Shental, N. Shental, and S. Shamai (Shitz), "On the achievable information rates of two-dimensional channels with memory," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Adelaide, Australia, Sep. 2005, pp. 2354–2358.

[35] R. Kindermann and J. L. Snell, *Markov Random Fields and their Applications*.   Providence, RI: Amer. Math. Soc., 1980.

[36] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Generalized belief propagation," in *Advances in Neural Information Processing Systems*, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds.   Cambridge, MA: MIT Press, 2001, vol. 13.

[37] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, Jul. 2005.

[38] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*.   San Francisco, CA: Morgan Kaufmann, 1988.

[39] D. J. C. Mackay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electron. Lett.*, vol. 33, no. 6, pp. 457–458, Mar. 1997.

[40] R. G. Gallager, "Low density parity check codes," *IRE Trans. Inf. Theory*, vol. IT-8, no. 1, pp. 21–28, Jan. 1962.

[41] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding," in *Proc. IEEE Int. Confe. Communications (ICC)*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[42] C. Berrou and A. Glavieux, "Near optimum error-correcting coding and decoding: Turbo codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

[43] R. J. McEliece, D. J. C. MacKay, and J. F. Cheng, "Turbo decoding as an instance of Pearl's 'belief propagation' algorithm," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 140–152, Feb. 1998.

[44] P. Pakzad and V. Anantharam, "Kikuchi approximation method for joint decoding of LDPC codes and partial-response channels," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1149–1153, Jul. 2006.

[45] M. Opper and D. Saad, *Advanced Mean Field Methods: Theory and Practice*.   Cambridge, MA: MIT Press, 2001.

[46] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

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

[48] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems*.   Berlin, Germany: Springer, 1999.

[49] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

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

[51] D. Ornstein and B. Weiss, "Entropy and recurrence rates for stationary random fields," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1694–1697, Jun. 2002.

[52] D. Ornstein and B. Weiss, "The Shannon–McMillan–Breiman theorem for a class of amenable groups," *Israel J. Math.*, vol. 44, pp. 53–60, 1983.

[53] E. Lindenstrauss, "Pointwise theorems for amenable groups," *Invent. Math.*, vol. 146, no. 2, pp. 259–295, Jan. 2001.

[54] B. G. Leroux, "Maximum-likelihood estimation for hidden Markov models," *Stochastic Processes and their Applications*, vol. 40, pp. 127–143, 1992.

[55] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[56] M. A. Iqbal and W. Weeks, IV, "Simulation-based estimation of the capacity of full-surface channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Adelaide, Australia, Sep. 2005, pp. 1778–1782.

[57] I. Kanter and H Kfir, "Statistical mechanical aspects of joint source-channel coding," *Europhys. Lett.*, vol. 63, no. 2, pp. 310–316, 2003.

[58] L. Onsager, "Crystal statistics: A two-dimensional model with order-disorder transition," *Phys. Rev.*, vol. 65, pp. 117–149, 1944.

[59] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*.   London, U.K.: Academic, 1982.

[60] G. Blatter, M. V. Feigel'man, V. B. Geshkenbein, A. I. Larkin, and V. M. Vinokur, "Vortices in high-temperature superconductors," *Rev. Mod. Phys.*, vol. 66, no. 4, pp. 1125–1388, Oct. 1994.