Generalized Belief Propagation Receiver for Near-Optimal Detection of Two-Dimensional Channels with Memory

Ori Shental[1] and Anthony J. Weiss
Faculty of Engineering
Dept. of Electrical Eng.-Systems
Tel Aviv University
Tel Aviv, Israel 69978
{shentalo,ajw}@eng.tau.ac.il

Noam Shental and Yair Weiss
Computer Science & Eng.
Center for Neural Computation
Hebrew University of Jerusalem
Jerusalem, Israel 91904
{fenoam,yweiss}@cs.huji.ac.il

Abstract — **We propose a generalized belief propagation (GBP) receiver for two-dimensional (2-D) channels with memory, which is applicative to 2-D inter-symbol interference (ISI) equalization and multi-user detection (MUD). Our experimental study demonstrates that under non-trivial interference conditions, the performance of this fully tractable GBP receiver is almost identical to the performance of the optimal maximum a-posteriori (MAP) receiver.**

I. Introduction

Two-dimensional (2-D) finite state input channels with memory play a fundamental role in various applications in modern communications. A popular and important instance of this class of channels is 2-D inter-symbol interference (ISI) channels, which appear, e.g., in magnetic and optical recording systems. In ISI, finite state symbols are ordered on a 2-D grid, causing interference in a limited neighborhood.

Another interesting instance of dispersive 2-D channels is multiple-access (MA) channels, which appear in cellular networks. Following Wyner's cellular model [1], a planar uplink model can be viewed as a 2-D channel, where each cell corresponds to a node in the grid, and interference occurs between neighboring cells. This Wyner-like model assumes that most of the multiple-access interference (MAI) is caused by inter-cell effects, rather than intra-cell effects. This assumption is typical to MA systems with long signature codes distinguishing between users within a cell. These codes are long enough to diminish cross-correlations between users within the cell, however, since they are being reused in neighboring tiers, they may cause inter-cell interference. An example is the allocation of scrambling codes in 3rd generation UMTS uplink [2].

The task in both ISI and MA channels is to overcome interference and noise, and detect the transmitted symbols. In ISI this process is termed channel equalization, while in MA it is called multi-user detection (MUD, [3]).

Optimal detection can be achieved by maximum a-posteriori (MAP) joint sequence decision based on the matched filter outputs of all corrupted symbols. The optimal receiver achieves a significant capacity improvement over the conventional matched filter receiver. However, its complexity for 2-D channels is exponential in the grid's size.

Hence, various practical sub-optimal detection methods have been proposed (e.g. [4, 5] and references therein). For example, Marrow and Wolf [4] have recently evaluated the performance of several iterative detection methods for the binary-input 2-D ISI channel. They proposed detection schemes operating iteratively on the rows and columns of the 2-D channel, which approximates the optimal receiver's bit error rate (BER) to within 0.5dB.

In this work we present a detection algorithm based on methods from probabilistic graphical models. The basic observation is that 2-D channels can be viewed as an undirected graphical model, a.k.a. pairwise Markov random fields (MRF, [6]), and detection of the 2-D channel symbols is equivalent to performing inference in this undirected graphical model.

Graphical models provide powerful tools for exact (optimal) or approximate inference. For example, loopy belief propagation (LBP, sum-product algorithm, [7]) is an efficient way to solve inference problems on graphical models, which has been shown to serve remarkably well as a decoding engine in low density parity check (LDPC) codes [8]. In this work we show that, unlike LDPC codes, LBP performs poorly in 2-D channels detection. However, an extension of LBP termed generalized belief propagation (GBP, [9]), is shown to provide a practical method with near-optimal performance both for 2-D ISI equalization and for MUD in uncoded cellular networks.

The paper is organized as follows. Section II introduces the dispersive 2-D channel model, and section III presents its connection to graphical models. Section IV describes the complexity of exact inference in 2-D channels. Section V presents approximate inference using LBP and discusses its shortcomings, and then moves on to GBP. Section VI provides simulation results for the 2-D ISI and MA channels in various topologies, which are discussed in section VII.

We shall use the following notations. The operator $\{\cdot\}^T$ stands for a vector or matrix transpose, $\{\cdot\}_i$ and $\{\cdot\}_{ij}$ denote entries of a vector and matrix, respectively.

II. System Model

Consider a discrete time $N \times N$ 2-D finite state input channel with memory of the form

$$y_{k,l} = d_{k,l} + v_{k,l} + \sum_{(i,j)\in<k,l>} \alpha_{i,j} d_{i,j} \quad \forall k,l = 1,\ldots,N, \quad (1)$$

where $y_{k,l}$, the channel's output observation at symbol (k,l), is the sum of the finite state alphabet input symbol $d_{k,l}$ and two additional terms. The first term, $v_{k,l}$, represents ambient additive white Gaussian noise (AWGN), while the second term is the scaled interference caused by adjacent symbols to (k,l), denoted by $< k,l >$. The parameter $\alpha_{i,j}$ ($|\alpha_{i,j}| \leq 1$)

[1]This work was performed in part during a research visit at the Institute for Telecommunications Research, University of South Australia, supported by the Australian Government under ARC Grant DP0344856.

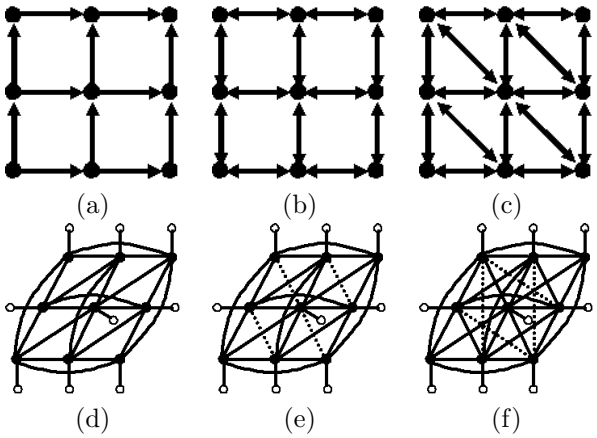

Fig. 1: **Upper pane:** Interference structures for three types of 3×3 2-D channels: (a) ISI grid, (b) rectangular cellular network (c) hexagonal cellular network. The arrows mark the direction of interference. **Lower pane:** The corresponding undirected graphical model representation of the channels in the upper pane: (d) ISI grid, (e) and (f) rectangular and hexagonal cellular networks. Full nodes represent (hidden) transmitted bits, while empty nodes correspond to the observations. Interaction couplings (compatibility function) ψ_{ij} are denoted by a solid line connecting two full nodes, while the external field (evidence) ϕ_i is depicted by a solid line connecting a full node and an empty node. For clarity we use dotted edges in (e) and (f) to represent the extra edges added compared to the graphs (d) and (e), respectively.

controls the interference attenuation and fading. We assume the channel is perfectly known at the receiver's side.

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

$$\mathbf{y} = \mathbf{S}\mathbf{d} + \mathbf{v}, \tag{2}$$

where the matrix $\mathbf{S}$ encapsulates the memory/interference structure.

Different 2-D channels, such as ISI or multi-cell network topologies of the MA system, differ from one another by the interference matrix $\mathbf{S}$. Our basic assumption, which later allows for a graphical model interpretation, is that interference is caused by neighboring symbols, i.e., $\mathbf{S}$ is actually a band matrix. The upper pane in Fig. 1 represents the interference structure of three topologies: ISI, a rectangular cellular network, and the typical hexagonal cellular network.

In the following derivations, we assume real-space data signalling $\mathbf{d}$, interference $\mathbf{S}$ and noise $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ (an extension to the complex domain is straightforward.)

A. Optimal Detection

The individually optimum maximum a-posteriori (MAP) detection of the i'th transmitted symbol, $d_i = \{\mathbf{d}\}_i$, is given by

$$\hat{d}_i = \arg\max_{x_i} \Pr(x_i|\mathbf{y}) = \arg\max_{x_i} \sum_{\mathbf{x}\setminus x_i} \Pr(\mathbf{x}|\mathbf{y}), \tag{3}$$

where $x_i = \{\mathbf{x}\}_i$ represents the possible values of the i'th transmitted symbol.

The joint conditional probability distribution is

$$\Pr(\mathbf{x}|\mathbf{y}) = \Pr(\mathbf{v}) \propto \exp\Big(-\frac{1}{2\sigma^2} ||\mathbf{y} - \mathbf{S}\mathbf{x}||^2 \Big). \tag{4}$$

Hereinafter, for exposition purposes, we consider the case of binary-input alphabet, i.e. $d_i \in \pm 1$. Hence, the sufficient statistics in this case takes the following form

$$\Pr(\mathbf{x}|\mathbf{y}) \propto \exp\Big(-\frac{1}{\sigma^2} \big(\sum_{(i>j)} R_{ij} x_i x_j - \sum_i h_i x_i \big) \Big), \tag{5}$$

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

III. The Connection To Undirected Graphical Models

An undirected graphical model with pairwise potentials (a.k.a. pairwise Markov random fields), consists of a graph G and potential functions $\psi_{ij}(x_i, x_j)$ and $\phi_i(x_i)$ such that the probability of an assignment $\mathbf{x}$ is given by

$$\Pr(\mathbf{x}) \propto \prod_{(i>j)} \psi_{ij}(x_i, x_j) \prod_i \phi_i(x_i). \tag{6}$$

Hence, the probability distribution in (5) defines the following undirected graphical model

$$\Pr(\mathbf{x}|\mathbf{y}) \propto \prod_{(i>j)} \psi_{ij}(x_i, x_j) \prod_i \phi_i(x_i, h_i), \tag{7}$$

where

$$\psi_{ij}(x_i, x_j) = \exp\big(-\frac{R_{ij} x_i x_j}{\sigma^2}\big) \tag{8}$$

is a compatibility function representing the structure of the system. The potential

$$\phi_i(x_i, y_i) = \exp\big(\frac{h_i x_i}{\sigma^2}\big) \tag{9}$$

is the 'evidence' or local likelihood, which describes the statistical dependency between the hidden variable x_i and the observed variable h_i[1]. The lower pane in Fig. 1 presents the resulting graphical model of our three application examples - ISI, a rectangular cellular network and a hexagonal cellular network.

Once the graphical model is defined we turn to the analogue of detection which is termed 'inference', i.e., calculating the probability given in (3). The field of graphical models has developed tools for efficiently calculating (3), i.e., performing exact inference. In cases where exact inference is intractable, approximation algorithms have been developed. In the next section we present the junction tree algorithm for exact inference, and then we turn to approximate inference and discuss our application of generalized belief propagation.

IV. Exact Inference

The junction tree algorithm performs exact inference over a general graph by converting it into an equivalent tree whose nodes contain clusters (cliques) of nodes of the original graph. Then inference can be performed by passing messages forward

[1]Notice that for the non-binary finite state input alphabet case, the MRF modelling is identical, except for an additional external field potential operating on each node which can be absorbed into ϕ_i term. This additional potential arises from the auto-correlations R_{ii}, which can not be dropped out from the sufficient statistics expression as in the binary case.

and backward along neighboring cliques in the tree (for further details see, e.g., [10].)

The complexity of the algorithm is exponential in the size of the largest clique in the derived tree. For $N \times N$ grid-like graphs, such as our 2-D problem, the size of the largest clique is proportional to the grid's dimension length N times the memory depth ν of the channel. Hence, the detection complexity becomes impractical for large grids. Furthermore, even for moderate size graphs (e.g. a small network of tens of cells) when using a non-binary modulation, or for non-trivial memory effect, exact inference is still impossible. Thus, we must resort to approximate inference methods.

V. Approximate Inference

In this section we discuss approximate inference algorithms using belief propagation and its generalizations.

A. Loopy Belief Propagation

Loopy belief propagation (LBP) is equivalent to applying Pearl's local message passing algorithm [7], 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 application, e.g., in decoding Turbo and LDPC codes. The performance of LBP in these applications may be attributed to the sparsity of the graphs. The cycles in the graph are long, hence inference may be performed as if it were a tree.

Can LBP be used for approximating the transmitted symbols over dispersive 2-D channels?

Our studies show that LBP almost always fails to converge and therefore the associated detection performance is poor. This result is not surprising. As opposed to the sparse graphs of LDPC codes, the graphs of 2-D channels consist of many short cycles. As a result LBP's implicit tree-like assumption does not hold, and its approximation is poor. In order to earn back the near-optimal characteristics of message passing inference one must circumvent this problem.

B. Generalized Belief Propagation

The generalized belief propagation algorithm (GBP, [9]) is an extension of LBP that has been shown to provide better approximations in many applications. GBP starts by defining regions (clusters) of nodes which may intersect, and then passes messages between these regions in an analogous way to LBP. Within each such region GBP performs *exact* inference, hence short cycles of nodes which are included in a region cause no problem. To date, no systematic method of choosing these regions in a general graph exists, so as to produce a better approximation. A region that encompasses all nodes along the shortest cycles, might be a desired choice. For example, the regions used for the 2-D channel examples (of all types) consisted of 3×3 squares of nodes, as displayed in Fig. 2. This choice of regions is natural since nodes are connected up to their second nearest neighbor. Notice that the graph 'contained' in a region is identical to the graphs shown in Fig 1(d)–(f).

Surprisingly, the computations required for GBP are only slightly more than the computations required for LBP and the complexity grows exponentially only with the size of the chosen regions. Naturally, enlarging the basic region entails a more accurate inference at the cost of complexity.

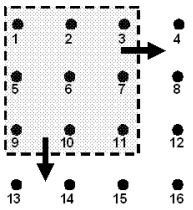

Fig. 2: Covering a 4×4 2-D channel by 3×3 regions used in GBP. Regions are defined by sliding a 3×3 window along the channel. The result is four regions for this 4×4 system:
$\{1, 2, 3, 5, 6, 7, 9, 10, 11\}$,
$\{2, 3, 4, 6, 7, 8, 10, 11, 12\}$,
$\{5, 6, 7, 9, 10, 11, 13, 14, 15\}$,
$\{6, 7, 8, 10, 11, 12, 14, 15, 16\}$.

Can GBP improve the detection of symbols corrupted by the 2-D channel?

Our empirical study shows that GBP gives remarkable results and near-optimal detection performance, even for harsh interference conditions. These close to optimal results are obtained even when small basic clusters are chosen (e.g. 3×3 node regions). Further details of the GBP receiver will be discussed in a future publication. In the following section, we present our experimental results of the GBP receiver for 2-D channels.

VI. Simulation Results

The performance of the proposed GBP receiver 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. In all examples, the size of selected GBP basic region was 3×3, as in Fig. 2.

The performance of GBP was compared to the optimal receiver, and to several other standard receivers. In all our experiments the LBP receiver did not converge in a substantial percentage of the simulations, for both synchronous and asynchronous message scheduling. As a result its performance was worse than a matched filter receiver, and in certain cases was close to a random guess. Hence, LBP's results are omitted from the performance evaluation figures.

A. 2-D ISI Channel

Following [4], we simulated a 6×6 binary ISI channel, under two non-trivial constant interference levels, $\alpha_{i,j} = \alpha = 0.5$ and $\alpha = 0.75$. Outside the grid the symbols were assumed to have value (-1). Fig. 3 compares between the equalization performance of the optimal receiver and GBP, in terms of average bit error rate (BER) per symbol as a function of signal to noise ratio (SNR). As can be seen the GBP error performance coincides with MAP error levels or is extremely close to it. In [4], for similar ISI channel setting, the performance of the best iterative detector was approximately 2/3dB above the optimal BER.

We also evaluated the performance of the GBP receiver in a similar setting, but for a larger 20×20 channel with $\alpha = 0.5$. As MAP equalization is infeasible for such a system, we compared GBP's performance 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 matched filter. Fig. 4 displays the equalization performance of GBP and the lower bound on the optimal error probability. The results show that the GBP receiver's BER is very close to the bound, which implies that its performance is near-optimal for this system also.

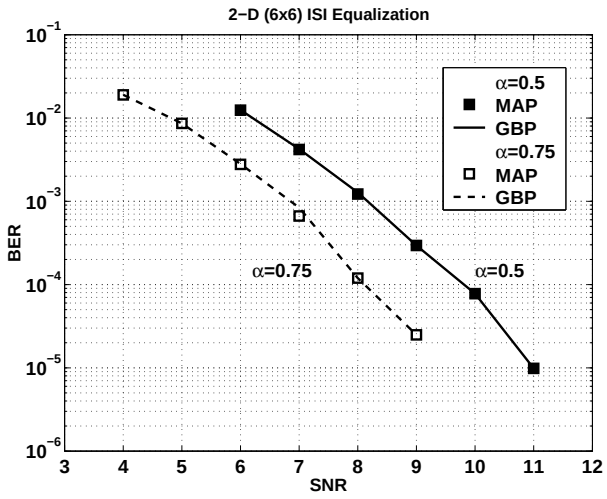

Fig. 3: 2-D ISI equalization in a 6×6 channel for $\alpha_{i,j} = \alpha = 0.5$ and $\alpha = 0.75$. Equalization performance of GBP vs. optimal (MAP) receivers in terms of average bit error rate (BER) per symbol, as a function of signal to noise ratio (SNR). The solid line corresponds to $\alpha = 0.5$, and dashed line corresponds to $\alpha = 0.75$.

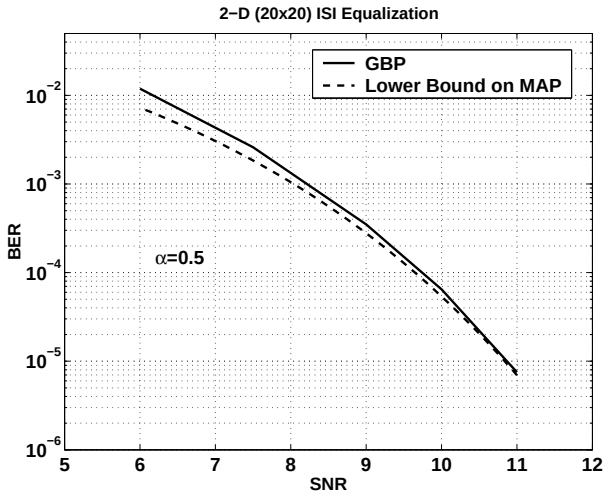

Fig. 4: 2-D ISI equalization in a 20×20 channel for $\alpha = 0.5$. Equalization performance of a GBP receiver (solid line) vs. lower bound (dashed line) on optimal (MAP) receiver error probability in terms of average BER per symbol, as a function of SNR.

B. Rectangular Topology Cellular Network

We simulated a cellular network of 9×9 cells[2] with rectangular topology (Fig. 1-(b)). Inter-cell interference was either constant $\alpha_{i,j} = \alpha = 0.5$, or taken from a zero-mean Gaussian distribution with standard deviation $\alpha_{std} = 0.5$, hard limited at $\alpha = \pm1$. The latter corresponds to a random MAI which may be interpreted as caused by channel fading.

Fig. 5 presents the performance of the optimal MAP receiver, the GBP-MUD receiver, and also two other standard receivers - the linear minimum mean square error (MMSE) and the naive single-user matched filter (MF) receivers. GBP practically coincides with the optimal MUD for both constant and random interference scenarios, while MMSE and MF are substantially inferior.

[2] 9×9 was the largest system for which exact inference was possible.

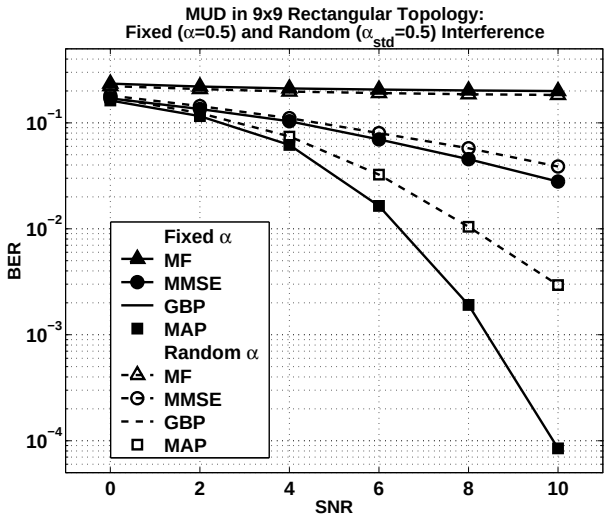

Fig. 5: MUD in a 9×9 rectangular topology. Average BER per cell as a function of SNR, for the optimal (MAP) and GBP receivers. Also shown are the BER for the linear minimum mean square error (MMSE) and single-user matched filter (MF) receivers. The BER was evaluated for two channel profiles of fixed ($\alpha_{i,j} = \alpha = 0.5$, solid lines) and random interference ($\alpha \sim 0.5 \cdot \mathcal{N}(0,1)$, dashed lines).

C. Hexagonal Topology Cellular Network

We empirically studied another instance of the 2-D model, where a 9×9 cellular network is planned according to a hexagonal topology (Fig. 1-(c)). As in the rectangular case the inter-cell interference was either constant or random. Fig. 6 compares the GBP-MUD receiver to the optimal MAP, and to MMSE and MF receivers, showing similar performance as in the rectangular case.

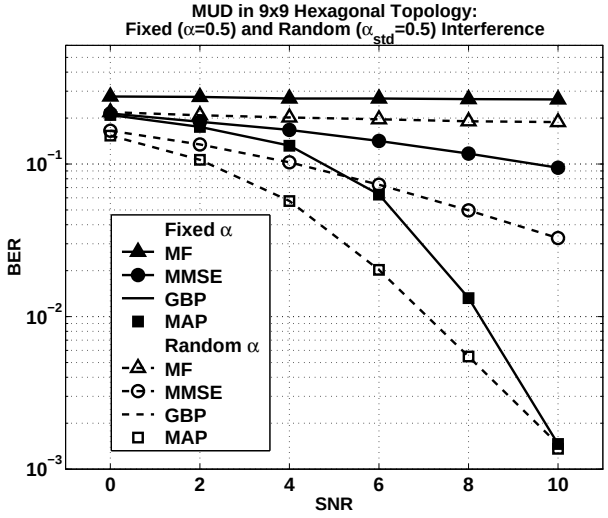

Fig. 6: MUD in 9×9 hexagonal topology. The setting is similar to Fig. 5.

We also evaluated the performance of the GBP-MUD receiver in an additional setting, in which the SNR level was fixed (4dB), and the interference α was varied. We used $\alpha \sim \alpha_{std} \cdot \mathcal{N}(0,1)$, hard limited at $\alpha = \pm1$, where $0 \leq \alpha_{std} \leq 1$. Fig. 7 displays the performance of GBP, MAP, MMSE and MF receivers in this setting. GBP is near-optimal for all interference levels, even for harsh ($\alpha \to 1$) conditions.

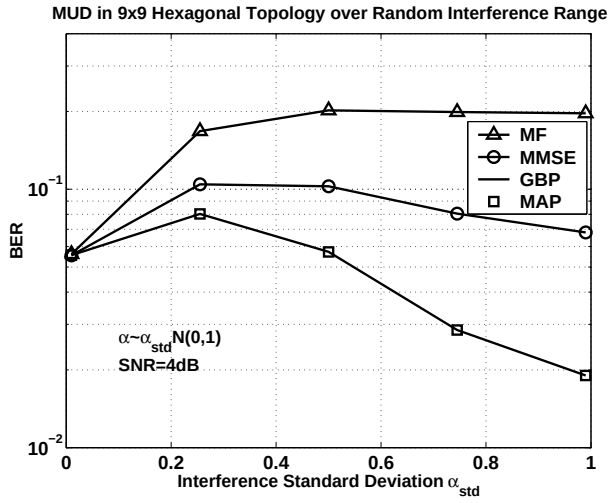

Fig. 7: MUD in 9×9 hexagonal topology over a range of random interferences. Average BER per cell as a function of random interference standard deviation α_{std}, for the optimal (MAP), GBP, MMSE, and MF receivers. The BER was evaluated for a random interference scenario $\alpha \sim \alpha_{std} \cdot \mathcal{N}(0,1)$ and a fixed SNR (= 4dB).

VII. Discussion

In this paper we introduced a GBP receiver for 2-D channels with memory. Simulation results for three different examples show near-optimal error performance of this fully tractable message passing scheme. This behavior is consistent both as a function of the SNR and over the possible interference range.

Other practical instances of the 2-D channel for which the GBP receiver may be suitable are indoor wireless local area networks (W-LAN) and multiple-in multiple-out (MIMO) channels. An example of the latter is a single cell base-station with a planar antenna array and widely spaced antenna elements. The remarkable performance of the proposed detection scheme may also be of interest in designing new ad-hoc wireless networks.

Another attractive property of the GBP receiver in the Wyner-like 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.

Apart from providing the correct hard decisions, GBP receiver may also infer the marginal probabilities which may become useful in coded systems. We empirically observed that the marginal beliefs well approximate the a-posteriori probabilities (APP). Thus, GBP may operate, e.g., as a detection stage in an iterative detection and decoding scheme. Further research concerning the usage of the GBP receiver in coded systems is underway.

Acknowledgments

O. Shental is grateful to S. Shamai, L. Rasmussen and A. Grant for useful discussions. O. Shental also thanks the Institute for Telecommunications Research, University of South Australia, for the warm hospitality during the process of this research, and to the Australia-Israel Chamber of Commerce for its support.

References

[1] A. D. Wyner, "Shannon-theoretic approach to a gaussian cellular multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1713–1727, Nov. 1994.

[2] The 3rd generation partnership project website. [Online]. Available: `http://www.3gpp.org`

[3] S. Verdú, *Multiuser Detection*. Cambridge, UK: Cambridge University Press, 1998.

[4] M. Marrow and J. K. Wolf, "Iterative detection of 2-dimensional channels," in *Inform. Theory Workshop*, Paris, France, Mar. 2003.

[5] Y. Wu, J. A. O'Sullivan, N. Singla, and R. S. Indeck, "Iterative detection and decoding for separable two-dimensional intersymbol interference," *IEEE Trans. Magn.*, vol. 39, pp. 2115–2120, July 2003.

[6] R. Kindermann and J. L. Snell, *Markov Random Fields and Their Applications*. Providence, Rhode Island: American Mathematical Society, 1980.

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

[8] D. J. C. Mackay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Elect. Lett.*, vol. 33, pp. 457–458, 1997.

[9] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free energy approximations and generalized belief propagation algorithms," Mitsubishi Electric Laboratories, Cambridge, MA, Tech. Rep. TR-2004-040, May 2004. [Online]. Available: `http://www.merl.com`

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