

# Pseudo Prior Belief Propagation for Densely Connected Discrete Graphs

Jacob Goldberger  
School of Engineering  
Bar-Ilan University, Ramat-Gan 52900, Israel  
goldbej@eng.biu.ac.il

Amir Leshem  
School of Engineering  
Bar-Ilan University, Ramat-Gan 52900, Israel  
leshema@eng.biu.ac.il

**Abstract**—This paper proposes a new algorithm for the linear least squares problem where the unknown variables are constrained to be in a finite set. The factor graph that corresponds to this problem is very loopy; in fact, it is a complete graph. Hence, applying the Belief Propagation (BP) algorithm yields very poor results. The Pseudo Prior Belief Propagation (PPBP) algorithm is a variant of the BP algorithm that can achieve near maximum likelihood (ML) performance with low computational complexity. First, we use the minimum mean square error (MMSE) detection to yield a pseudo prior information on each variable. Next we integrate this information into a loopy Belief Propagation (BP) algorithm as a pseudo prior. We show that, unlike current paradigms, the Belief Propagation (BP) algorithm can be advantageous even for dense graphs with many short loops. The performance of the proposed algorithm is demonstrated on the MIMO detection problem based on simulation results.

## I. INTRODUCTION

Finding a linear least squares fit to data is a well-known problem, with applications in almost every field of science. When there are no restrictions on the variables, the problem has a closed form solution. In many cases, a-priori knowledge on the values of the variables is available. One example is the existence of priors, which leads to Bayesian estimators. Another example of great interest in many applications is when the variables are constrained to a discrete finite set. This problem has many diverse applications such as decoding of multi-input-multi-output (MIMO) digital communication systems, and lattice problems in computer science, such as finding the closest vector in a lattice to a given point in  $R^n$ . In contrast to the continuous linear least squares problem, this problem is known to be NP hard.

The graphical model induced by the constrained linear problem is a complete graph where each variable is connected to all the other variables. Previous attempts to apply loopy Belief Propagation (BP) algorithms to this problem have all yielded poor results. The goal of this paper is to show that certain BP schemes can be effective on fully connected graphs. In this study we introduce a modified BP based on pseudo prior information that is computed in a pre-processing step. This algorithm, which we dub *Pseudo Prior Belief Propagation* (PPBP), results in a near-optimal and low complexity algorithm. The algorithm is demonstrated on the MIMO application. It should be noted, however, that the method is general and can be applied to any integer linear least-square problem. The main focus of this paper is the effectiveness of certain BP

schemes to fully connected graphical models.

A MIMO system is a communication system with  $n$  transmit antennas and  $m$  receive antennas. The tap gain from transmit antenna  $i$  to receive antenna  $j$  is denoted by  $\mathbf{H}_{ij}$ . In each use of the MIMO channel a signal vector  $x = (x_1, \dots, x_n)^\top$  is independently selected from a set of constellation points  $\mathcal{A}$  according to the data to be transmitted, so that  $x \in \mathcal{A}^n$ . The received vector  $y$  is given by:

$$y = \mathbf{H}x + \epsilon \quad (1)$$

The vector  $\epsilon$  is an additive noise in which the noise components are assumed to be zero mean, statistically independent Gaussians with a known variance  $\sigma^2$ . The channel matrix which is assumed to be known, comprises iid elements drawn from a zero-mean normal distribution of unit variance. The MIMO detection problem then becomes one of finding the transmitted vector  $x$  given  $\mathbf{H}$  and  $y$ . The maximum likelihood (ML) solution is:

$$\hat{x} = \arg \min_{x \in \mathcal{A}^n} \|\mathbf{H}x - y\|^2 \quad (2)$$

However, ML decoding is unfeasible when either the number of transmit antennas or the constellation size is large. A simple approximation is the zero-forcing (ZF) algorithm which is based on a linear decision that ignores the finite constellation constraint:

$$\hat{x} = (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top y \quad (3)$$

and then, neglects the correlation between the symbols, to find the closest constellation point for each symbol independently. This scheme performs poorly due to its inability to handle ill-conditioned channel matrix realizations. Somewhat better performance can be obtained by using a minimum mean square error (MMSE) filter instead of ZF on the unconstrained linear system:

$$\hat{x} = (\mathbf{H}^\top \mathbf{H} + \frac{\sigma^2}{e} I)^{-1} \mathbf{H}^\top y \quad (4)$$

( $e$  is the transmit signal power) and then finding the closest lattice point in each component independently. A vast improvement over the linear approach can be achieved by using sequential decoding. The MMSE-SIC detection algorithm [3] is based on the MMSE solution, but detects the signals one after another and not in parallel. Recently several alternative structures have been proposed to approach the ML detection

performance such as the sphere decoding algorithm [5] and methods based on semidefinite relaxation [13].

## II. THE LOOPY BELIEF PROPAGATION APPROACH

Given the constrained linear system  $y = \mathbf{H}x + \epsilon$ , and a uniform prior distribution on  $x$ , the posterior probability function of the discrete random vector  $x$  given  $y$  is:

$$p(x|y) \propto \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{H}x - y\|^2\right), \quad x \in \mathcal{A}^n \quad (5)$$

The notation  $\propto$  stands for equality up to a normalization constant. Observing that  $\|\mathbf{H}x - y\|^2$  is a quadratic expression, it can be easily verified that  $p(x|y)$  is factorized into a product of two- and single-variable potentials:

$$p(x_1, \dots, x_n|y) \propto \prod_i \psi_i(x_i) \prod_{i < j} \psi_{ij}(x_i, x_j) \quad (6)$$

such that

$$\psi_i(x_i) = \exp\left(-\frac{1}{2\sigma^2}y^\top \mathbf{h}_i x_i\right) \quad (7)$$

$$\psi_{ij}(x_i, x_j) = \exp\left(-\frac{1}{\sigma^2}\mathbf{h}_i^\top \mathbf{h}_j x_i x_j\right) \quad (8)$$

where  $\mathbf{h}_i$  is the  $i$ -th column of the matrix  $\mathbf{H}$ . Since the obtained factors are just a function of pairs we obtain a Markov-Random-Field (MRF) representation [14]. In the MIMO application the (known) matrix  $\mathbf{H}$  is randomly selected and therefore, the MRF graph is usually a completely connected graph.

In a loop-free MRF graph the max-product variant of the BP algorithm always converges to the most likely configuration (which corresponds to ML decoding in our case). For loop-free graphs, BP is essentially a distributed variant of dynamic programming. The BP message update equations only involve passing messages between neighboring nodes. Computationally, it is thus straightforward to apply the same local message updates in graphs with cycles. In most such models, however, this loopy BP algorithm will not compute exact marginal distributions; hence, there is almost no theoretical justification for applying the BP algorithm. However, the BP algorithm applied to loopy graphs has been found to have outstanding empirical success in many applications, e.g., in decoding LDPC codes [4]. The performance of BP in this application may be attributed to the sparsity of the graphs. The cycles in the graph are long, hence the graph have tree-like properties, so that messages are approximately independent and inference may be performed as though the graph was loop-free. The BP algorithm has also been used successfully in image processing and computer vision where the image is represented using a grid-structured MRF based on local connections between neighbor nodes.

However, when the graph is not sparse, and is not based on local grid connections, loopy BP almost always fails to converge. (One exception is that, for Gaussian graphs, if BP converges, then the means are correct [12]). Unlike the sparse graphs of LDPC codes, the MRF graphs of MIMO channels are *completely connected graphs* and therefore the associated

detection performance is poor. This has prevented the BP from being an asset for the MIMO problem. Fig. 1 shows an example of a MIMO real-valued system based on an  $8 \times 8$  matrix and  $\mathcal{A} = \{-1, 1\}$  (see the simulation results section for a detailed description of the simulation set up). As can be seen in Fig. 1, the BP decoder based on the MRF representation (6) has very poor results (standard techniques such as damping do not help here). In recent year there were several attempts to applying BP for densely connected graphs (mainly Gaussian graphs) [9], [2], [1], [7], [10]. In this study we apply a novel variant of the BP algorithm to the MIMO fully connected bipartite graph.

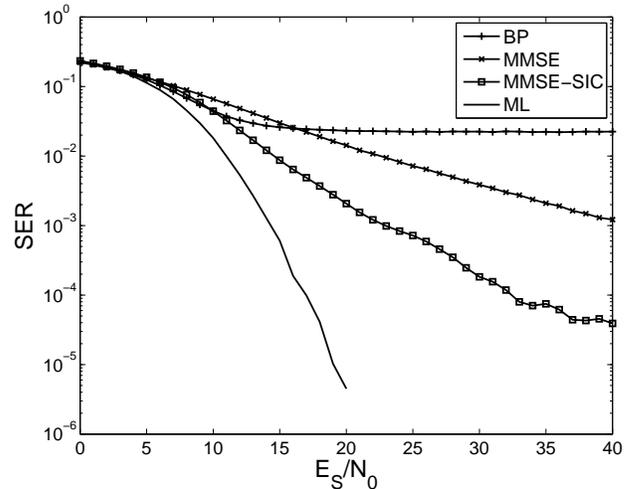


Fig. 1. Decoding results for  $8 \times 8$  system,  $\mathcal{A} = \{-1, 1\}$ .

There have been several recent attempts to apply BP to the MIMO detection problem with good results (e.g. [6], [8]). However in these methods the factorization of the probability function is done in such a way that each factor corresponds to a single linear equation. This leads to a partition of the probability function into factors, each of which is a function of all the unknown variables. This leads to exponential computational complexity in computing the BP messages. Shental et. al [11] analyzed the case where the matrix  $\mathbf{H}$  is relatively sparse (and has a grid structure). They showed that even under this restricted assumption the BP still does not perform well. As an alternative method they proposed the generalized belief propagation (GBP) algorithm that does work well on a sparse matrix if the algorithm regions are carefully chosen. There are situations where the sparsity assumption makes sense (e.g. 2D intersymbol interference (ISI) channels). However, in the MIMO channel model we assume that the channel matrix elements are iid and Gaussian; hence we cannot assume that the channel matrix  $\mathbf{H}$  is sparse.

## III. THE PSEUDO PRIOR BELIEF PROPAGATION DETECTOR

Our approach is based on an approximation of the exact probability function (5) that enables a successful implementation of the Belief Propagation paradigm.

To assist the BP algorithm to concentrate on the right target, we suggest combining the result in the ZF detector (4) in the BP iterations as a pseudo prior. Ignoring the finite-set constraints on the input, we can solve the MIMO system as a standard linear system. As such we can obtain not just the unbiased estimator  $z = (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top y$  but also its covariance matrix which is  $\text{Var}(z) = \sigma^2 (\mathbf{H}^\top \mathbf{H})^{-1}$ . Using the variance information we can compute the posterior probability of each symbol (soft-decision) as follows:

$$p_{zf}(x_i|y) \propto \exp\left(-\frac{(x_i - z_i)^2}{2\text{Var}(z_i)}\right) \quad \forall x_i \in \mathcal{A} \quad (9)$$

We next integrate the ZF posterior information into the BP scheme as additional single-variable factors. The additional potentials are redundant and do not contribute any new information. They can be viewed as a regularization term that prevents the BP from drifting away from the optimal solution. To control the influence of the prior, we can also add a weight (denoted by  $\lambda$ ) to the additional factors. The proposed approach, therefore, is based on applying the BP on the factors of the following modified probability function:

$$\begin{aligned} \hat{p}(x|y) &\propto \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{H}x - y\|^2 - \lambda \sum_i \frac{(x_i - z_i)^2}{2\text{Var}(z_i)}\right) \\ &= \prod_i \psi_i(x_i) \prod_{i \neq j} \psi_{ij}(x_i, x_j) \prod_i p_{zf}^\lambda(x_i|y) \end{aligned}$$

It can be easily verified that  $p(x|y)$  (5) can also be written as:

$$p(x|y) \propto \exp\left(-\frac{1}{2\sigma^2} (x - z)^\top \mathbf{H}^\top \mathbf{H} (x - z)\right) \quad (10)$$

and  $p_{zf}(x_i|y)$  (9) can be written as:

$$p_{zf}(x_i|y) \propto \exp\left(-\frac{1}{2\sigma^2} ((x - z)^\top (\text{diag}((\mathbf{H}^\top \mathbf{H})^{-1})^{-1} (x - z))\right)$$

Hence we can rewrite the modified probability function  $\hat{p}(x|y)$  as:

$$\hat{p}(x|y) \propto \exp\left(-\frac{1}{2\sigma^2} (x - z)^\top A (x - z)\right) \quad (11)$$

where

$$A = \mathbf{H}^\top \mathbf{H} + \lambda (\text{diag}(\mathbf{H}^\top \mathbf{H})^{-1})^{-1} \quad (12)$$

The proposed approach is based on applying the BP algorithm to  $\hat{p}(x|y)$  instead of the true probability  $p(x|y)$ . As can be seen from equation (12), the difference between the two distributions is that in  $\hat{p}(x|y)$  the diagonal of the 'inverse-covariance' element in the exponent is more dominant. We show empirically that this makes the BP algorithm work much better. This empirical observation resembles the behavior of the BP algorithm in densely connected Gaussian graphs. In Gaussian MRF if the inverse-covariance (precision) matrix  $A$  is strictly diagonally dominant (i.e.,  $|A_{ii}| > \sum_{j \neq i} |A_{ij}|$ ) then the BP converges and the marginal means converge to the true means [12].

The MMSE Bayesian approach (4) is known to be better than the linear based solution (3). In a similar way, by using

the MMSE solution instead of ZF we can consider a Bayesian version of the proposed modified distribution.

$$\hat{p}(x|y) \propto \exp\left(-\frac{1}{2\sigma^2} (x - u)^\top B (x - u)\right) \quad (13)$$

such that  $u = (\mathbf{H}^\top \mathbf{H} + \frac{\sigma^2}{e} I)^{-1} \mathbf{H}^\top y$  and

$$B = (\mathbf{H}^\top \mathbf{H} + \frac{\sigma^2}{e} I)^{-1} + \lambda (\text{diag}((\mathbf{H}^\top \mathbf{H} + \frac{\sigma^2}{e} I)^{-1}))^{-1} \quad (14)$$

The scalar parameter  $\lambda$  controls the relative importance of the redundant factors. If  $\lambda = 0$  the modified BP is reduced to the standard BP. When  $\lambda$  tends to infinity, the modified BP is reduced to the original MMSE solution. We found empirically that the need for the pseudo prior increases monotonically with the SNR. For higher SNR we need a larger  $\lambda$  to obtain good results. This may stem from the fact that for larger SNR levels the distribution is less smooth and there are more local minima. We have found that setting  $\lambda = \frac{1}{\sigma^2}$  works well for all SNR values.

Thus the proposed method is a combination of two known MIMO detection methods, MMSE and BP. Each one of these methods has a poor detection performance. In the experiment section we show that by combining these methods using a weight of  $\lambda = \frac{1}{\sigma^2}$  we can achieve near maximum likelihood performance. The proposed Pseudo Prior Belief Propagation (PPBP) algorithm for MIMO detection is summarized (written in the log domain) in the algorithm-box in Fig. 2.

There are a few implementation issues regarding the proposed PPBP algorithm that require discussion. Applying the BP algorithm exactly as it is described above can lead to problems of numerical stability. To avoid this we can work in the log domain. Here we used the max-product variant since it avoids complicated exponent-log computations and it can be carried out in the log domain just by using addition operations, which is useful for practical implementations.

The complexity of a straightforward implementation of the BP algorithm is  $O(n^3|\mathcal{A}|)$  where  $\mathcal{A}$  is the constellation set and  $n$  is the number of transmit antennas. The complexity is  $n^3$  because on each iteration we need to compute a message between each two variables and computing a single message is linear in  $n$ . The following efficient implementation (that is based on a certain message passing scheduling) can reduce the BP complexity to be just a quadratic function of the number of transmit antennas:

$$\begin{aligned} &\text{for } j = 1, \dots, n \\ &\quad \text{sum}_j(x_j) = \sum_{k \neq j} m_{k \rightarrow j}(x_j) \\ &\quad \text{for } i = 1, \dots, n \\ &\quad \quad m_{j \rightarrow i}(x_i) = \max_{x_j \in \mathcal{A}} (\text{prior}_j(x_j) + \varphi_j(x_j) \\ &\quad \quad \quad + \varphi_{ij}(x_i, x_j) + \text{sum}_j(x_j) - m_{i \rightarrow j}(x_j)) \end{aligned}$$

where

$$\begin{aligned} \varphi_i(x_i) &= -\frac{1}{2\sigma^2} y^\top \mathbf{H}_i x_i \\ \varphi_{ij}(x_i, x_j) &= -\frac{1}{\sigma^2} \mathbf{H}_i^\top \mathbf{H}_j x_i x_j \end{aligned}$$

and  $\text{prior}_j(x_j)$  is the weighted log posterior probability of the MMSE detector. The complexity of the PPBP algorithm

### The Pseudo Prior Belief Propagation (PPBP) algorithm:

Goal: Find (approx. to)  $\arg \min_{x \in \mathcal{A}^n} \|\mathbf{H}x - y\|^2$

Algorithm:

- Perform MMSE linear detection.

$$u = (\mathbf{H}^\top \mathbf{H} + \frac{\sigma^2}{e} I)^{-1} \mathbf{H}^\top y$$

$$\text{Var}(u_i) = \sigma^2 ((\mathbf{H}^\top \mathbf{H} + \frac{\sigma^2}{e} I)^{-1})_{ii}$$

$$\text{prior}_i(x_i) = \frac{1}{\sigma^2} \times \frac{-(x_i - u_i)^2}{2\text{Var}(u_i)}$$

- Max-Product initialization

Denote:

$$\varphi_i(x_i) = -\frac{1}{2\sigma^2} y^\top \mathbf{h}_i x_i$$

$$\varphi_{ij}(x_i, x_j) = -\frac{1}{\sigma^2} \mathbf{h}_i^\top \mathbf{h}_j x_i x_j$$

Initial messages:  $m_{j \rightarrow i}(x_i) = \text{prior}_i(x_i)$

- Loop until convergence:

$$m_{j \rightarrow i}(x_i) = \max_{x_j \in \mathcal{A}} (\text{prior}_j(x_j) + \varphi_j(x_j) + \varphi_{ij}(x_i, x_j) + \sum_{k \neq i, j} m_{k \rightarrow j}(x_j))$$

$$\text{belief}_i(x_i) = \text{prior}_i(x_i) + \varphi_i(x_i) + \sum_{j \neq i} m_{j \rightarrow i}(x_i)$$

- Hard decision:  $\hat{x}_i = \arg \max_{x_i \in \mathcal{A}} \text{belief}_i(x_i)$

Fig. 2. The PPBP algorithm

therefore is cubic in the number of transmit antennas and quadratic in the constellation order.

#### IV. SIMULATION RESULTS

In this section we provide simulation results for the proposed PPBP detector over various MIMO systems. We assume a quasi-static fading channel with a frame length of 100. Under the assumption of a block-fading channel model, the channel matrix  $\mathbf{H}$  is constant for 100 channel uses. The channel matrix comprised of iid elements drawn from a zero-mean normal distribution of unit variance. We used 10,000 realizations of the channel matrix. This resulted in  $10^6$  vector messages. The SNR is defined as  $10 \log_{10}(E_b/N_0)$  where  $E_b$  is the average received energy per symbol at each receiver antenna.

Fig. 3 shows the symbol error rate (SER) versus SNR for an  $8 \times 8$  BPSK MIMO system. The performance of the proposed PPBP method is compared to ML detection and to other linear suboptimal algorithms: the linear MMSE and the MMSE-SIC. In all our experiments the max-product variant described in Fig. 2 was used. It can be seen that the PPBP algorithm is significantly better than the MMSE-SIC.

Fig. 4 presents the performance of two possible variants of the BP algorithm using the same setup used in Fig. 3. The first is a standard BP without any pseudo-prior term where the MMSE factors are just used to initialize the messages at the first iteration. We showed above in Fig. 1. that the BP yields very poor results for the MIMO detection problem. We show here that using the MMSE results solely for initializing the messages does not help and in order to obtain good performance the MMSE factors should be used as additional pseudo prior factors. The second variant is a PPBP with a pseudo prior weight based on zero-forcing instead of the MMSE estimator. It is well known that the MMSE estimator outperforms the zero forcing estimator. Fig. 4. shows that MMSE based factors yield better results than factors that are based on the zero forcing estimator.

Fig. 5 shows the symbol error rate of the PPBP algorithm versus the regularization term  $\lambda$  for an  $8 \times 8$  BPSK MIMO system with SNR=11 ( $\sigma^2 = 0.8$ ). The scale of the x-axis is logarithmic and centered at  $\log(\frac{1}{\sigma^2})$  (e.g.  $x = 3$  means  $\log(\lambda) = 3 + \log(\frac{1}{\sigma^2})$ ). As shown in Fig. 5, setting  $\lambda = \frac{1}{\sigma^2}$  yields the best results. It is almost obvious that it is not a coincidence that the optimal value of  $\lambda$  is  $\frac{1}{\sigma^2}$ . This is the same constant that appears in the MMSE estimator. Analytical justification for the optimal value of  $\lambda$  is left for future research.

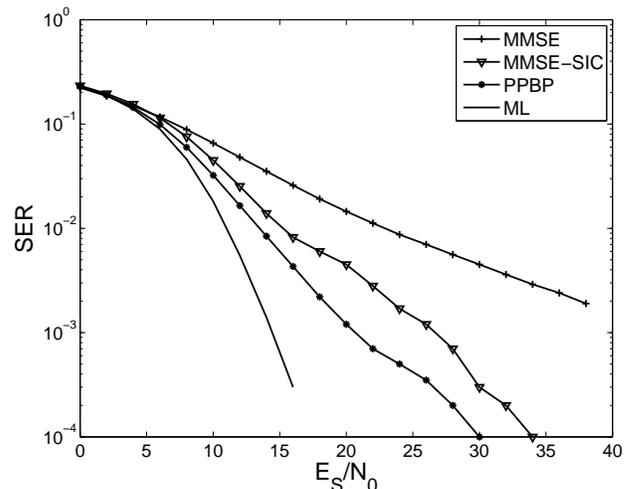


Fig. 3. SER versus SNR for an  $8 \times 8$  system using BPSK. Performance of PPBP compared to MMSE, MMSE-SIC and ML detection methods.

Fig. 6 shows the symbol error rate (SER) versus SNR for a  $6 \times 6$  4QAM MIMO system. It can be seen that the performance of the PPBP algorithm in this case is comparable to MMSE-SIC and is significantly better than the MMSE that is incorporated in the PPBP as a pseudo-prior.

#### V. CONCLUSIONS

We proposed a new BP approach for MIMO detection which is both endowed with low computational complexity and exhibits high decoding performance. The main contribution of this paper is introducing the concept of a pseudo prior into the

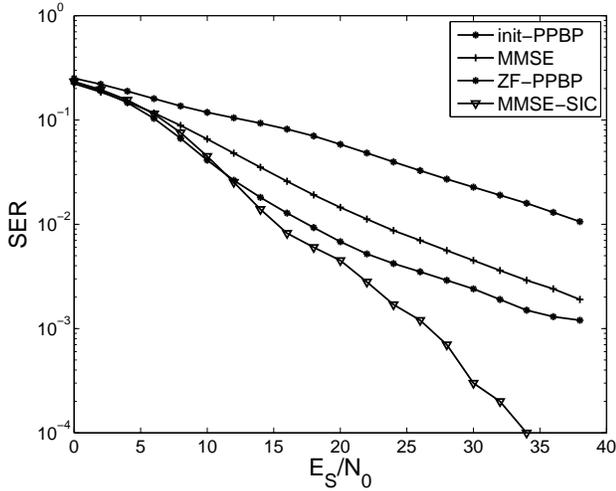


Fig. 4. SER versus SNR for an  $8 \times 8$  system using BPSK. Performance of variants of PPBP compared to MMSE and MMSE-SIC detection methods.

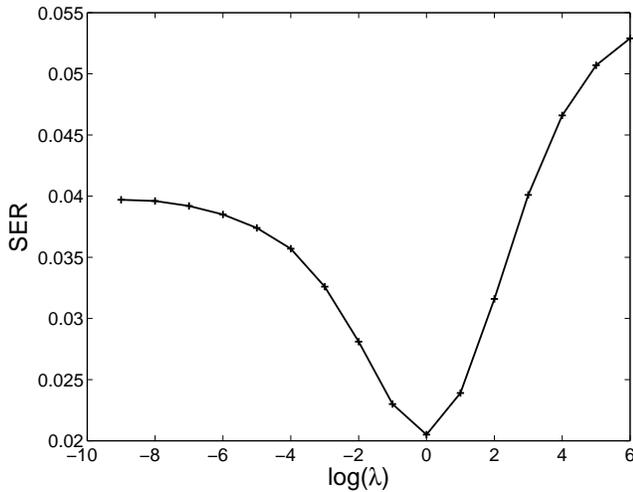


Fig. 5. SER versus the regularization term  $\lambda$  for an  $8 \times 8$  BPSK system with SNR = 11.

framework of loopy BP. We showed that there is a variant of the BP algorithm that can converge in dense graphs. We by no means claim that the proposed method is the optimal way to combine pseudo priors into the loopy BP. There are certainly other ways to extract pseudo priors and other ways to control the relative importance of their weights. In a more general perspective we show that we can slightly modify the true dense graph distribution such that the BP can be successfully applied on the modified model. It is left for future research to find other such modifications that on one hand are still close to the original distribution and on the other are suitable for iterative message passing algorithms.

#### REFERENCES

- [1] M. Bayati, D. Shah, and M. Sharma. Maximum weight matching via max-product belief propagation. *ISIT*, 2005.
- [2] D. Bickson, O. Shental, P. H. Siegel, J. K. Wolf, and D. Dolev. Gaussian belief propagation based multiuser detection. *ISIT*, 2008.

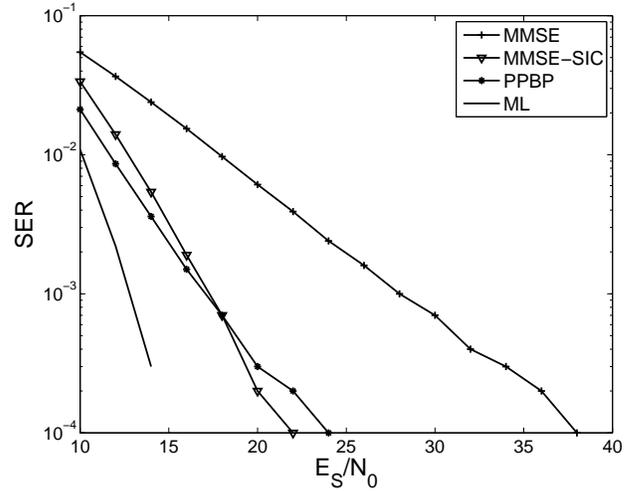


Fig. 6. SER versus SNR for a  $6 \times 6$  complex system using 4-QAM. Performance of PPBP compared to MMSE, MMSE-SIC and ML detection methods.

- [3] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Tech. J.*, pages 41–59, 1996.
- [4] R. G. Gallager. Low density parity check codes. *IRE Trans. Inform. Theory*, pages 21–28, 1962.
- [5] B. M. Hochwald and S. ten Brink. Achieving near-capacity on a multiple antenna channel. *IEEE Trans. Commun.*, pages 389–399, 2003.
- [6] J. Hu and T. M. Duman. Graph-based detector for BLAST architecture. *IEEE Communications Society ICC 2007*, 2007.
- [7] Y. Kabashima. A CDMA multiuser detection algorithm on the basis of belief propagation. *J. Phys. A*, pages 11111–11121, 2003.
- [8] M. Kaynak, T. Duman, and E. Kurtas. Belief propagation over SISO/MIMO frequency selective fading channels. *IEEE Transactions on Wireless Communications*, pages 2001–5, 2007.
- [9] A. Montanari, B. Prabhakar, and D. Tse. Belief propagation based multiuser detection. *Allerton Conference on Communication, Control and Computing*, 2005.
- [10] J.P. Neirotti and D. Saad. Improved message passing for inference in densely connected systems. *Europhys. Lett.*, pages 866–872, 2005.
- [11] O. Shental, N. Shental, S. Shamai (Shitz), I. Kanter, A.J. Weiss, and Y. Weiss. Discrete-input two-dimensional Gaussian channels with memory: Estimation and information rates via graphical models and statistical mechanics. *Information Theory, IEEE Transactions on*, pages 1500–1513, 2008.
- [12] Y. Weiss and W.T. Freeman. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Computation*, pages 2173–2200, 2001.
- [13] A. Wiesel, Y. C. Eldar, and S. Shamai. Semidefinite relaxation for detection of 16-QAM signaling in MIMO channels. *IEEE Signal Processing Letters*, 2005.
- [14] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. *IJCAI*, 2001.