A New MIMO detection algorithm based on the Gaussian graphical model

Mohammed Teeti*, Student Member, IEEE, YingZhuang Liu † and Jun Sun ‡
Huazhong University of Science and Technology, Wuhan, China
* moh_libra@hotmail.com, † liuyz@hust.edu.cn, ‡ francissunj@gmail.com

***Abstract*—The graphical models have been proven to be a very powerful and potential framework for addressing the inference problems. In this paper, we propose a new graphical model based algorithm for the detection of MIMO systems. The main feature of the algorithm lies in that it is implemented as a MRF-like graph, when combined with the Gaussian approximation and vector-based inference, our algorithm can lead to very promising performance, especially when the constellation size is small, with just linear complexity per symbol and memory requirement increases linearly with the number of transmit antennas. Simulation results collaborate with the analytical results, hence verifying the appeal of the algorithm for practical applications.**

I. Introduction

The advantage of using spatial-multiplexing techniques lies in the fact that they achieve the capacity of the MIMO channel [1], [2]. What is still a challenging task for achieving such a capacity is the efficient detection by the receiver. Since the complexity of maximum a posteriori (MAP) or maximum likelihood (ML) detectors grows exponentially with the number of transmit antennas, it becomes impractical to use these detectors in high-dimensional MIMO systems. Therefore, many detection techniques have been proposed, which provide suboptimal or near optimal performance with reduced complexity.

Linear detectors such as zero-forcing (ZF) and minimum-mean square error (MMSE) are much less complex than the optimal detector, and achieve suboptimal performance. Since this class of detectors try to nullify or minimize the interference signals rather than performing joint detection, their performance is far from optimal, especially when the number of antennas is high. Different nonlinear techniques, which are basically based on ZF or MMSE methods, are also used in VBLAST (Vertical Bell Labs layered Space-Time) architecture [3]. Indeed, they improve the performance without a significant increase of the complexity. Sphere decoding (SD) [4], on the other hand, can achieve near optimal performance by searching only a restricted space. However, its complexity is variable since the number of lattice points to be scanned in a sphere is unknown.

Other approaches are based on graphical models with belief propagation (BP) and its variants [5]. The graphical model of MIMO systems is fully connected, and can be usually represented as a factor graph (FG) or a pairwise Markov random field (PMRF) [6, Chapter 8], [7]. BP with full and partial graphs was first proposed in [8], where the complexity is of the same order as that of ML for full graphs. The difficulty of BP on loopy graphs is nonconvergence. To overcome this problem, BP can be combined with damped message updates [9], [10]. In message damping, a convex combination of the current message and the previous message is taken to be the new message. In the MIMO example, the success of this technique has been recently reported in [11]–[13].

Finally, the probabilistic data association (PDA) has been addressed in [14] as a marginalization technique for the distributions of variables. This approach is also used in [15], [13]. Ref. [13] uses FG with one-dimensional marginal distributions by using Gaussian approximation of the spatial interference. The performance of the algorithm improves for a large number of antennas. This paper attempts to formulate the MIMO detection problem as a vector-based inference on linear Gaussian MRF-like graphical model, which is related to PDA. Combined with BP algorithm, our algorithm which we dub 'Belief Propagation based Vector Gaussian Inference (BP-VGI)' has simple messages passing on the graph, and hence the complexity is a low order polynomial.

In the next section we review some basic properties of multivariate Gaussians. In Sec.III we extend the results of Sec.II to MIMO systems . Sec.IV presents BP-VGI. The complexity analysis is illustrated in Sec.V. Simulation results are shown in Sec.VI, and conclusion is presented in Sec.VII.

II. Linear Gaussian Model

In this section, we review some basic properties of multivariate Gaussians. Here we also show how to express a multivariate Gaussian as a directed graph, which leads naturally to a linear Gaussian model. Consequently, this framework can be extended to MIMO systems, which makes it easy to do inference on the graph.

Let $f(\mathbf{x})$ be a joint probability distribution of n continuous normal variables $x_1, x_2, \cdots, x_n$. Let $\mathbf{x}$ denote the n vector $(x_1, x_2, \cdots, x_n)^T$ with mean vector $\mu = (\mu_1, \mu_2, \cdots, \mu_n)^T$, and $n \times n$ covariance matrix $\mathbf{C}$. Let $\mathbf{x}_q$ and $\mathbf{x}_l$ represent two subvectors of respective dimensions q and l such that $n = q+l$. Given the mean vectors $\mu_q = \mathrm{E}[\mathbf{x}_q]$ and $\mu_l = \mathrm{E}[\mathbf{x}_l]$, then $f(\mathbf{x})$ can be written as

$$\begin{aligned} f(\mathbf{x}) \quad \propto \quad & \exp\left[-\frac{1}{2}(\mathbf{x}_q - \mu_q)^T \mathbf{C}_{qq}^{-1}(\mathbf{x}_q - \mu_q)\right] \times \\ & \exp\left[-\frac{1}{2}(\mathbf{x}_l - \mathbf{b})^T \mathbf{A}^{-1}(\mathbf{x}_l - \mathbf{b})\right] \qquad (1) \end{aligned}$$

978-1-4673-0990-5/12/$31.00 ©2012 IEEE

where $(.)^T$ denotes the transpose, $\mathbf{C}_{ij}$ is a submatrix of $\mathbf{C}$ with i row partitions and j column partitions, $\mathbf{b}$ is l-dimensional mean vector, and $\mathbf{A}$ is $l \times l$ covariance matrix,

$$\mathbf{b} \triangleq \mu_l + \mathbf{C}_{ql}^T \mathbf{C}_{qq}^{-1} (\mathbf{x}_q - \mu_q) \tag{2a}$$

$$\mathbf{A} \triangleq \mathbf{C}_{ll} - \mathbf{C}_{ql}^T \mathbf{C}_{qq}^{-1} \mathbf{C}_{ql}. \tag{2b}$$

To find the conditional distribution of $\mathbf{x}_l$ given $\mathbf{x}_q$, we apply the chain rule

$$\begin{aligned} f(\mathbf{x}_l|\mathbf{x}_q) &= \frac{f(\mathbf{x})}{f(\mathbf{x}_q)} \\ &\propto \exp\left[-\frac{1}{2}(\mathbf{x}_l - \mathbf{b})^T \mathbf{A}^{-1} (\mathbf{x}_l - \mathbf{b})\right]. \end{aligned} \tag{3}$$

Note that the length of $\mathbf{x}_l$ is specified by the number of variables it has. Now, If $\mathbf{x}_l$ represents an arbitrary single variable x_i, so $\mathbf{x}_q$ consists of $n-1$ variables which are parents of x_i. This implies that the conditional distribution in (3) reduces to one-dimensional Gaussian, specifically,

$$f(x_i|\mathrm{pa}_i) \propto \exp\left[-\frac{1}{2}\frac{(x_i - \eta_i)^2}{s_i^2}\right] \tag{4}$$

where we denote by pa_i all parents of x_i, η_i is the mean and s_i^2 the variance. It can be easily verified from (2a) that η_i can be written as

$$\eta_i \triangleq \sum_{k \in \mathrm{Pa}_i} w_{ik} x_k + a_i \tag{5}$$

with w_{ik} and a_i are parameters governing the mean. According to (5), the mean of x_i is a linear combination of the states of its parents. From (4) and (5), x_i can be expressed as

$$x_i = \sum_{k \in \mathrm{Pa}_i} w_{ik} x_k + a_i + s_i \rho_i \tag{6}$$

where ρ_i is a Gaussian random variable with zero mean and unit variance. The expected value $\mathrm{E}[x_i]$ and the covariance $\mathrm{cov}[x_i, x_j]$ can be determined recursively

$$\mathrm{E}[x_i] = \sum_{k \in \mathrm{pa}_i} w_{ik} \mathrm{E}[x_k] + a_i \tag{7a}$$

$$\begin{aligned} \mathrm{cov}[x_i, x_j] &= \mathrm{E}[(x_i - \mathrm{E}[x_i])(x_j - \mathrm{E}[x_j])] \\ &= \sum_{k \in \mathrm{pa}_j} w_{jk} \mathrm{cov}[x_i, x_k] + s_j^2 [i = j] \end{aligned} \tag{7b}$$

where we denote by $[i = j]$ the Iverson function which has two values; 1 when the proposition "$i = j$" is true, and 0 when "$i = j$" is false.

The main difficulty in carrying out (7b) is that when the resulting graph is fully connected, the number of parameters to be computed is very large, especially for large n. However, the entries on the diagonal (i.e., $\mathrm{cov}[x_i, x_j]$, where $i = j$) can be only updated. This would yield an approximation for the given distribution. In the special case when each pair (x_i, x_j) is independent, the mean vector and covariance matrix will be determined by $[a_1, \cdots, a_n]^T$ and $diag(\sigma_1^2, \cdots, \sigma_n^2)$ respectively.

III. Linear Gaussian Model of MIMO

We consider the uncoded $m \times n$ MIMO communication system with n transmit antennas and m receive antennas. The (j, i) entry of channel matrix $\mathbf{H}$ is the complex gain between transmit antenna i and receive antenna j. The entries of $\mathbf{H}$ are i.i.d complex Gaussians each with zero mean and variance 0.5 per dimension. We denote by $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$ the transmitted vector, and each component of $\mathbf{x}$ is drawn uniformly from a finite set $\mathcal{A}$ with normalized power. For each channel use, the received m-dimensional vector $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v} \tag{8}$$

where $\mathbf{v}$ is an m-dimensional noise vector with all entries are i.i.d complex Gaussians each with zero mean and variance $0.5\sigma^2$ per dimension. It is well known that ML and MAP detection achieve the same optimal performance when all transmitted vectors are equally likely. The ML detection tries to maximize the following distribution

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\left[-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\right], \mathbf{x} \in \mathcal{A}^n \tag{9}$$

which needs to find the minimum Euclidean distance $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$ over all possible transmitted vectors. However, this is prohibitive when n becomes large. By virtue of the inherent properties of multivariate Gaussian, we therefore approximate the distribution in (9). This would enable a simplified implementation of the BP paradigm.

Motivated by the 1D conditional Gaussian distribution in (4), we extend it to the case of MIMO in which the nodes of the graph are now random vectors. Let $\mathbf{h}_i$ represent the i-th column vector of $\mathbf{H}$, then we can rewrite (8) as

$$\mathbf{y} = \mathbf{h}_1 x_1 + \mathbf{h}_2 x_2 + ... + \mathbf{h}_n x_n + \mathbf{v}. \tag{10}$$

Now, consider the MIMO graph having n nodes, and assume that each node is an m-dimensional Gaussian vector. Thus, each vector can be obtained from (10) as follows

$$\begin{aligned} \mathbf{d}_i &= \mathbf{h}_i x_i \\ &= \mathbf{y} - \left[\sum_{k=1, k\neq i}^{n} \mathbf{h}_k x_k + \mathbf{v}\right], i = 1, 2 \cdots, n. \end{aligned} \tag{11}$$

Note that $\mathbf{d}_i$ being dependent on x_i implies that the conditional distribution of $\mathbf{d}_i$ given $\{x_k\}_{k\neq i}$ is proportional to the conditional distribution of x_i given $\{x_k\}_{k\neq i}$, specifically,

$$p(x_i|\mathrm{pa}_i) \propto \exp\left[-\frac{1}{2}(\mathbf{d}_i - \mu_i)^H \Sigma_i^{-1} (\mathbf{d}_i - \mu_i)\right] \tag{12}$$

where $(\cdot)^H$ denotes the Hermitian transpose, $\mu_i = (\mu_{i1}, \cdots, \mu_{in})^T$ is the mean vector, and Σ_i is the $m \times m$ covariance matrix of the distribution.

Instead of dealing with $\mathbf{d}_i$ as a compound quantity, we treat it as a simple random vector. Accordingly, the mean of the resulting conditional distribution in (12) depends on the states of the parents of x_i as a weighted sum of the channel vectors (column vectors) $\{\mathbf{h}_k\}_{k\neq i}$, whereas the covariance matrix Σ_i does not. In addition, the k-th column vector of $\mathbf{H}$ in (11) can be regarded as a vector parameter that captures an implicit correlation between the i-th and the k-th variables. From (12) we see that our approximation approach is, therefore, based on replacing the true distribution in (9) with the following approximation:

$$p(\mathbf{x}|\mathbf{y}) \approx c\prod_{i=1}^{n} \exp\left[-\frac{1}{2}(\mathbf{d}_i - \mu_i)^H \Sigma_i^{-1}(\mathbf{d}_i - \mu_i)\right] \quad (13)$$

where c is a normalizing constant. For simplicity of notation, Σ_i from now on stands for a diagonal covariance matrix. It's worth noting that the vector representation of nodes can benefit from the diversity gain that MIMO provides. This is likely to yield a better performance.

IV. INFERENCE WITH BELIEF PROPAGATION

Consider the fully-connected graph of MIMO, consisting of n nodes with each represents an m-dimensional Gaussian vector, then the a posteriori distribution of $\mathbf{x}$ given $\mathbf{y}$ can be approximated by (13). The i-th factor of (13) is conditioned on all states of the random variables $\{x_k\}_{k\neq i}$. Therefore, we can determine μ_i for each factor recursively as follows

$$\begin{aligned} \mu_i &= \mathrm{E}[\mathbf{d}_i] && \text{(14a)} \\ &= \mathbf{y} - \sum_{k=1,k\neq i}^{n} \mathbf{h}_k \mathrm{E}[x_k]. \end{aligned}$$

Let e_j denotes an m-dimensional vector whose j-th element is 1, and 0 elsewhere. The notation $\mathbf{h}_k^j$ means the j-th element of the k-th channel column, and we will denote by $\mathbf{z}_i$ the column vector with the elements on the diagonal of Σ_i. Then $\mathbf{z}_i$ can be similarly evaluated recursively by

$$\mathbf{z}_i = \sum_{k=1,k\neq i}^{n} \sum_{j=1}^{m} \langle \mathbf{h}_k^j, \mathbf{h}_k^j \rangle \mathrm{Var}(x_k) e_j + \sigma^2 \sum_{j}^{m} e_j \quad \text{(14b)}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product, and $\mathrm{Var}(\cdot)$ is the variance.

We see that (14a) and (14b) are recursive equations, so by starting from an initial solution we can recursively compute μ_i and $\mathbf{z}_i$ $\forall i = 1, \cdots, n$. However, since each factor in (13) is conditioned on $n-1$ variables, this implies that the graph is fully connected. Hence, the direct implementation of (14a) and (14b) fails to converge. We noted that even the use of some loopy graph techniques (i.e., message damping) is not effective. To overcome this problem, we resort to BP algorithm. In the original BP algorithm, each node computes a message for each of its neighbor nodes and passes it to the corresponding node. In the PMRF model of MIMO systems with pairwise cliques, the joint probability distribution can be simply written as

$$P(\mathbf{x}) = \mathrm{Z}^{-1} \prod_{i<j} \phi_{ij}(x_i, x_j) \prod_{i} \phi_i(x_i) \quad (15)$$

where the value of Z^{-1} is chosen to normalize the joint probability distribution. Two random variables x_i and x_j are related by a compatibility function $\phi_{ij}(x_i, x_j)$, and the function $\phi_i(x_i)$ is called the evidence. Since $P(\mathbf{x})$ is expressed as a product of factors, the computation of marginal probabilities is usually done by the BP algorithm. The message sent from a node i to a node j is given by

$$m_{i\to j}(x_j) = \sum_{x_i\in\mathcal{A}} \phi_i(x_i)\phi_{ij}(x_i, x_j) \prod_{k\neq i,j} m_{k\to i}(x_i). \quad (16)$$

After a certain number of iterations, the belief at node i is taken to be the product of its evidence and all the messages coming from all other nodes, specifically,

$$b_i(x_i) = \phi_i(x_i) \prod_{k\neq i} m_{k\to i}(x_i) \quad (17)$$

We should note that this computation is exact for graphs without loops.

Now consider the joint probability distribution in (13) where each factor can be seen as an evidence for the i-th. One can see immediately that the compatibility functions are totally absorbed by the evidences. In other words, all compatibility functions seem to be constants. Unlike the evidences of MRF, the evidences in our case are not fixed and evolve over time.

To compute the marginal probability at a node i, the algorithm works as follows : We use BPSK signaling where $x_i \in \{\pm 1\}$, and for simplicity, computations can be performed in the log-domain. At the start of the algorithm, we initialize the expected values and variances of all variables to 0 and 1 respectively (i.e., $\mathrm{E}[x_i] = 0$, $\mathrm{Var}(x_i) = 1, \forall i = 1, \cdots, n$). Then, the vectors $\mu_1, \mu_2, \cdots, \mu_n$ and $\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_n$ are updated according to (14a) and (14b). The likelihood ratio (LLR) is then calculated at each node. By (12), the LLR at node i is of the form

$$\begin{aligned} \Lambda_i &= \log \frac{p(x_i = +1|\mathbf{y})}{p(x_i = -1|\mathbf{y})} && (18) \\ &= 2\Re\{\mathbf{h}_i{}^H \Sigma_i^{-1} \mu_i\}, i = 1, 2, \cdots, n. \end{aligned}$$

The calculated LLRs are then simultaneously broadcasted. At node i, the LLRs sent from other nodes are collected to update μ_i and Σ_i. To do this, it first computes the prior probabilities of all other variables. Let $p_k \triangleq Pr(x_k = +1)$, then the prior probability of x_k can be computed by

$$p_k = \frac{1}{1 + e^{-\Lambda_k}}, k = 1, \cdots, n. \quad (19)$$

From the definition of the mean and the variance of a discrete random variable, it follows that $\mathrm{E}[x_k]$ and $\mathrm{Var}[x_k]$ are given by

$$\begin{aligned} \mathrm{E}[x_k] &= \sum_{x\in\{\pm 1\}} x \times Pr(x_k = x) && \text{(20a)} \\ &= 2p_k - 1 \end{aligned}$$

$$\begin{aligned} \mathrm{Var}[x_k] &= \sum_{x\in\{\pm 1\}} (x - \mathrm{E}[x_k])^2 \times Pr(x_k = x) \\ &= 1 - \mathrm{E}[x_k]^2. && \text{(20b)} \end{aligned}$$

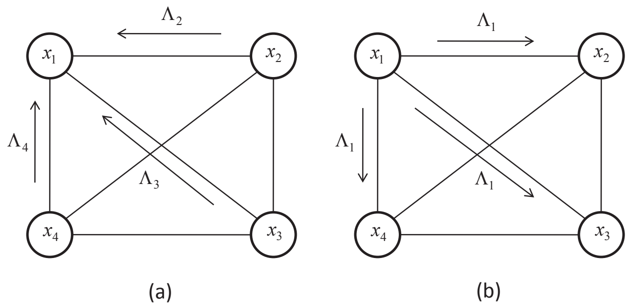

Fig. 1. Belief update for $x_1 : \Lambda_1 = f(\Lambda_2, \Lambda_3, \Lambda_4)$

Having updated the parameters of the density of the i-th node, its LLR can be again recomputed by (18) and broadcasted. This process runs for a certain number of iterations until convergence. Finally, a hard decision on the variable x_i is taken to be in favor of 1 if the LLR is positive, otherwise it is -1. An illustration of the update procedure of belief at node x_1 is shown in Fig. 1.

We emphasize here that the loopy BP does not converge or converges to inexact marginals. However, for a loop-free graph BP and its variants, the messages always converge [5]. In the case of Gaussian graphs, it has been shown [16] that when loopy BP converges, then the means are those of the exact marginal distributions (though the covariances will not be necessary correct). BP also has found success in many applications, i.e., decoding of LDPC codes [17]. In these applications, the girth of the associated graphs is large enough, hence the messages are approximately independent. Unfortunately, the graph of MIMO is fully connected, so the BP algorithm has a problem with convergence. Therefore, the task is now to apply message damping. Since the massages compromise of probabilities, so the new probability at the i-th node can be obtained as

$$\overset{(t)}{p_i} = \beta \overset{(t)}{p_i} + (1-\beta) \overset{(t-1)}{p_i}, 0 \leq \beta \leq 1, i = 1, \cdots, n. \quad (21)$$

We should mention that these probabilities can be computed only once and reused by all nodes. Fig. 2 depicts the effect of belief damping on the bit error rate (BER) in 4×4, 8×8 and 16×16 VBLAST MIMO systems under flat fading. We see that when no damping is used ($\beta = 1$), BER improves slightly as n increases when SNR is fixed. However, when damping is used (i.e., $\beta \approx 0.6, 0.75$ and 0.75 for $n = 4, 8, 16$ respectively, SNR = 12dB) BER is improved significantly, especially for large n. In general, these curves may arise from the fact that different values of noise variance give rise to different optimal damping values. However, we found empirically that the larger the noise variance is, the larger β is needed and vice versa.

V. Complexity Analysis

In order to analyze the complexity of our algorithm, we note that the number of messages required per iteration is n. Each message is in the form of LLR, where the computation of each LLR is linked to the computation of μ_i and Σ_i. So, (14a) and (14b) can be used to more easily analyze the

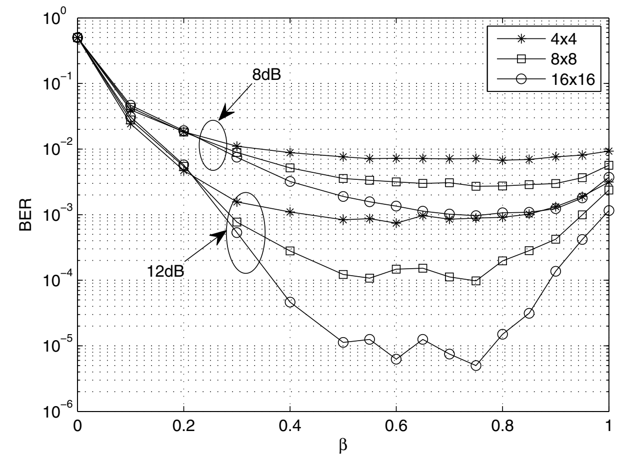

Fig. 2. belief damping effect on BER for different VBLAST MIMO systems under flat fading, # of iterations=10

complexity. Another observation is that the two summations in (14a) and (14b) need to be computed only once by removing the constraint ($i \neq k$), and thus can be reused when μ_i and Σ_i are updated at each node.

For example, the computation of (14a) consists of $n+1$ sums, which requires mn additions. Each time $\mu_i, \forall i = 1, 2, \cdots, n$ is updated, we add its corresponding term $\mathbf{h}_i \, \mathrm{E}[x_i]$, which amounts to mn additions. Thus the overall number of additions of computing the mean vectors is now $2mn$. Similarly, the computation of (14b) can be carried out in the same manner, which yields the same complexity. While the necessary parameters are now available, the computation of all LLRs in (18) is $O(nm)$. It follows that the overall complexity of this algorithm when $n = m$ is $O(n)$ for detecting one symbol. Concerning space complexity, each node only needs to store its past belief for damping purpose. In general, the algorithm requires a memory of size $O(1)$ per symbol which is an improvement by an order over FG-based approach in [13].

VI. Simulation Results and Discussion

In this section, we show some simulation results for the uncoded VBLAST MIMO systems . We consider a flat-fading MIMO channel with BPSK signaling scheme. The damping factor $\beta = 0.6$, and the number of iterations is limited to 10. Since the ML detection is optimal and the performance of the unfaded SISO AWGN channel is a lower bound on the optimal detection, so they are used as references.

Fig 3 shows the BER of BP-VGI, when $n \neq m$. It can be seen that its performance is very close to that of ML, especially when $n < m$. This is clearly illustrated for 8×4 and 8×6 cases. In addition, we also observe that when $m < n$, the performance is still good, but it suffers from a degradation at high SNR, especially for small n (i.e., 4×8). The preceding observations, when compared to the BP performance (i.e., [8], Fig.5 and Fig.6), achieve almost the same performance when $n < m$. In fact, [8] can do even better with the exponential

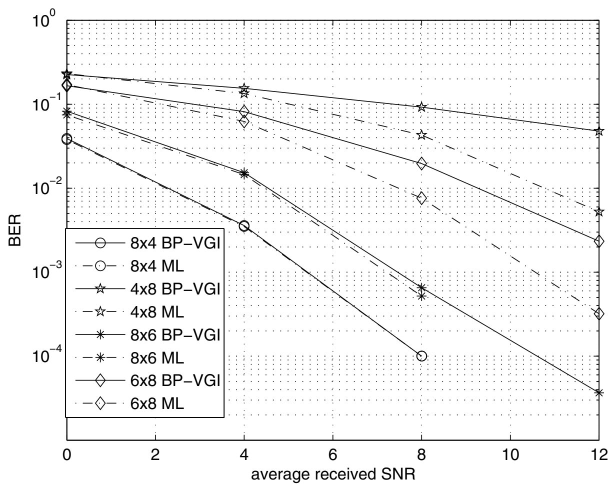

Fig. 3. BER performance of BP-VGI and ML when $n \neq m$

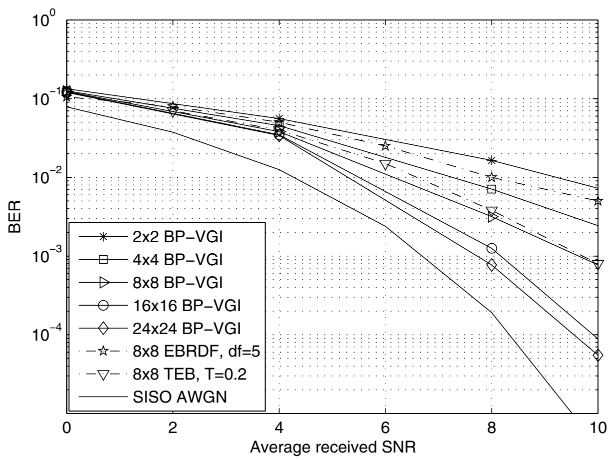

Fig. 4. BER performance of BP-VGI when $n = m$

complexity of BP by reducing the degree of freedoms d_f. This will certainly lead to performance degradation compared to our algorithm, and the complexity is still exponential with d_f.

On the other hand, the BER of BP-VGI with $n = m$ is shown in Fig. 4. We observe that there is always an improvement in the performance when n increases, and therefore the performance approaches that of ML as n increases further. In addition, we present two variants of BP known as edge-based regular-d_f (EBRDF), and threshold and edge based (TEB) algorithms [8] for 8×8 system. The complexity of these two algorithms is exponential with the parameter d_f (i.e., $d_f = 5$ for EBRDF, whereas the average d_f for TEB is equal to $= 8e^{-0.2} \approx 6.5$). It is clear that TEB outperforms EBRDF, and its performance is almost the same as our algorithm. However, its complexity is still exponential with d_f, and hence it is not appropriate for large dimensions.

VII. Conclusion

In this paper, we have proposed a new graphical model based algorithm for MIMO detection in VBLAST architecture with belief propagation. We have shown that the algorithm has $O(n)$ complexity per symbol, the number of messages passing on the graph and memory requirement scale linearly with the number of antennas. The simulation results have shown that the combination of our model and belief propagation is impressive and promising for practical applications.

Acknowledgment

This work is supported by National Natural Science Foundation of China, under grant No. 60972015.

References

[1] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Trans. Telecommun.*, vol. 10, pp. 585–595, November 1999.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[3] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Signals, Systems, and Electronics, 1998. ISSSE 98. 1998 URSI International Symposium on*, sep-2 oct 1998, pp. 295 –300.

[4] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1639–1642, 1999.

[5] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988.

[6] J. S. Yedidia, W. T. Freeman, and Y. Weiss, *Exploring Artificial Intelligence in the New Millennium*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003.

[7] E. Biglieri, A. R. Calderbank, T. Constantinides, A. Goldsmith, and A. Paulraj, *MIMO Wireless Communications*. Cambridge University Press, 2010.

[8] J. Hu and T. M. Duman, "Graph-based detection algorithms for layered space-time architectures," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 2, pp. 269–280, 2008.

[9] M. Pretti, "A message passing algorithm with damping,," *Jl. Stat. Mech.: Theory and Practice*, Nov. 2005.

[10] D. Koller and N. Friedman, *Probabilistic Graphical Models - Principles and Techniques*. MIT Press, 2009.

[11] M. Suneel, P. Som, A. Chockalingam, and B. Rajan, "Belief propagation based decoding of large non-orthogonal STBCs," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, 28 2009-july 3 2009, pp. 2003 –2007.

[12] P. Som and A. Chockalingam, "Damped belief propagation based near-optimal equalization of severely delay-spread UWB MIMO-ISI channels," in *Communications (ICC), 2010 IEEE International Conference on*, may 2010, pp. 1 –5.

[13] P. Som, T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Low-complexity detection in large-dimension MIMO-ISI channels using graphical models," *J. Sel. Topics Signal Processing*, vol. 5, no. 8, pp. 1497–1511, 2011.

[14] D. Pham, K. Pattipati, P. Willett, and J. Luo, "A generalized probabilistic data association detector for multiple antenna systems," *Communications Letters, IEEE*, vol. 8, no. 4, pp. 205 – 207, april 2004.

[15] S. Mohammed, A. Chockalingam, and B. Rajan, "Low-complexity near-MAP decoding of large non-orthogonal STBCs using PDA," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, 28 2009-july 3 2009, pp. 1998 –2002.

[16] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in gaussian graphical models of arbitrary topology," *Neural Computation*, vol. 13, no. 10, pp. 2173–2200, 2001.

[17] R. Gallager, "Low-density parity-check codes," *Information Theory, IRE Transactions on*, vol. 8, no. 1, pp. 21 –28, january 1962.