Low-Complexity Detection in Large-Dimension MIMO-ISI Channels Using Graphical Models

Pritam Som, Tanumay Datta, *Student Member, IEEE*, N. Srinidhi, *Student Member, IEEE*, A. Chockalingam, *Senior Member, IEEE*, and B. Sundar Rajan, *Senior Member, IEEE*

***Abstract*—In this paper, we deal with low-complexity near-optimal detection/equalization in large-dimension multiple-input multiple-output inter-symbol interference (MIMO-ISI) channels using message passing on graphical models. A key contribution in the paper is the demonstration that near-optimal performance in MIMO-ISI channels with large dimensions can be achieved at low complexities through simple yet effective simplifications/approximations, although the graphical models that represent MIMO-ISI channels are fully/densely connected (loopy graphs). These include 1) use of Markov Random Field (MRF) based graphical model with pairwise interaction, in conjunction with *message damping*, and 2) use of Factor Graph (FG) based graphical model with *Gaussian approximation of interference* (GAI). The per-symbol complexities are $O(K^2n_t^2)$ and $O(Kn_t)$ for the MRF and the FG with GAI approaches, respectively, where K and n_t denote the number of channel uses per frame, and number of transmit antennas, respectively. These low-complexities are quite attractive for large dimensions, i.e., for large Kn_t. From a performance perspective, these algorithms are even more interesting in large-dimensions since they achieve increasingly closer to optimum detection performance for increasing Kn_t. Also, we show that these message passing algorithms can be used in an iterative manner with local neighborhood search algorithms to improve the reliability/performance of M-QAM symbol detection.**

***Index Terms*—MIMO-ISI channels, severe delay spreads, large dimensions, low-complexity detection, graphical models, Markov random fields, pairwise interaction, factor graphs.**

I. Introduction

Signaling in large dimensions can offer attractive benefits in wireless communications. For example, transmission of signals using large spatial dimensions in multiple-input multiple-output (MIMO) systems with large number of transmit/receive antennas can offer increased spectral efficiencies [1]-[3]. The spectral efficiency in a V-BLAST (spatial multiplexing) MIMO system is n_t symbols per channel use, where n_t is the number of transmit antennas [3]. Severely delay-spread inter-symbol interference (ISI) channels can offer opportunities to harness rich diversity benefits [4]. In an L-length ISI channel, each symbol in a frame is interfered by its previous $L-1$ symbols. However, the availability of L copies of the transmitted signal in ISI channels can be exploited to achieve Lth order diversity. A way to achieve this diversity is to organize data into frames, where each frame consists of K channel uses (i.e., K dimensions in time), $K > L$, and carry out joint detection/equalization over the entire frame at the receiver. A MIMO-ISI channel with large Kn_t and L (referred to as large-dimension MIMO-ISI channel) is of interest because of its potential to offer high spectral efficiencies (in large n_t) and diversity orders (in large L[1]). A major challenge, however, is detection complexity. The complexity of optimum detection is exponential in number of dimensions, which is prohibitive for large number of dimensions. Our focus in this paper is to achieve near-optimal detection performance in large dimensions at low complexities. A powerful approach to realize this goal, which we investigate in this paper, is message passing on graphical models.

Graphical models are graphs that indicate inter-dependencies between random variables [10]. Well known graphical models include Bayesian belief networks, factor graphs, and Markov random fields [11]. Belief propagation (BP) is a technique that solves inference problems using graphical models [11]. BP is a simple, yet highly effective, technique that has been successfully employed in a variety of applications including computational biology, statistical signal/image processing, data mining, etc. BP is well suited in several communication problems as well [10]; e.g., decoding of turbo codes and LDPC codes [12],[13], multiuser detection in CDMA [14]-[16], and MIMO detection [17]-[20].

Turbo equalization which performs detection/equalization and decoding in an iterative manner in coded data transmission over ISI channels have been widely studied [21],[22],[23]. More recently, message passing on factor graphs based graphical models [24] have been studied for detection/equalization on ISI channels [25]-[30]. In [27], it has been shown through simulations that application of sum-product algorithm to factor graphs in ISI channels converges to a good approximation of the exact a posteriori probability (APP) of the transmitted

Manuscript received January 22, 2011; revised May 26, 2011 and August 17, 2011; accepted August 17, 2011.
This work in part was presented at IEEE ITW'2010, Cairo, Egypt, January 2010, IEEE ICC'2010, Cape Town, South Africa, May 2010, and National Conference on Communications (NCC'2011), Bangalore, India, January 2011.
This work was supported in part by the DRDO-IISc Program on Advanced Research in Mathematical Engineering and in part by a gift from The Cisco University Research Program, a corporate advised fund of Silicon Valley Community Foundation.
The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore-560012, India. e-mail: {pritamsm, srinidhi, achockal, bsrajan}@ece.iisc.ernet.in, tan.swapnil@gmail.com

[1]A practical example of severely delay-spread ISI channel with large L is an ultra wideband (UWB) channel [5]. UWB channels are highly frequency-selective, and are characterized by severe ISI due to large delay spreads [6]-[9]. The number of multipath components (MPC) in such channels in indoor/industrial environments has been observed to be of the order of several tens to hundreds; number of MPCs ranging from 12 to 120 are common in UWB channel models [6],[9].

symbols. In [28], the problem of finding the linear minimum mean square error (LMMSE) estimate of the transmitted symbol sequence is addressed by employing a factor graph framework. Equalization in MIMO-ISI channels using factor graphs are investigated in [29],[30]. In [29], variable nodes of the factor graph correspond to the transmitted symbols, and each channel use corresponds to a function node. Since the received signal at any channel use depends on the past L symbols transmitted from every transmit antenna, every function node is connected to Ln_t variable nodes. Near-MAP (maximum a posteriori probability) performance was shown through simulations for $n_t = 2$ systems. However, the complexities involved in the computation of messages at the variable and function nodes are exponential in Ln_t, which are prohibitive for large spatial dimensions and delay spreads.

Our key contribution in this paper is the demonstration that graphical models can be effectively used to achieve *near-optimal* detection/equalization performance in *large-dimension* MIMO-ISI channels at *low complexities*. The achieved performance is good because detection is performed jointly over the entire frame of data; i.e., over the full $Kn_t \times 1$ data vector. While simple approximations/simplifications resulted in low complexities, the *large-dimension behavior*[2] natural in message passing algorithms contributed to the near-optimal performance in large dimensions. The graphical models we consider in this paper are Markov random fields (MRF) and factor graphs (FG). We show that these graphical models based algorithms perform increasingly closer to the optimum performance for increasing n_t and increasing values of K and L, keeping L/K fixed.

In the case of MRF BP approach (Section III-A), we show that the use of *damping* of messages, where messages are computed as a weighted average of the messages in the previous iteration and the current iteration (details and associated references given in Section III-A4), is instrumental in achieving good performance. Simulation results show that the MRF BP approach exhibits large-dimension behavior, and that damping significantly improves the bit error performance (details given in Section III-C1). For example, the MRF BP algorithm with message damping achieves close to unfaded single-input single-output (SISO) AWGN performance (which is a lower bound on the optimum detector performance) within 0.25 dB at 10^{-3} bit error rate (BER) in a MIMO-ISI channel with $n_t = n_r = 4$, $K = 100$ channel uses per frame (i.e., problem size is $Kn_t = 400$ dimensions), and $L = 20$ equal-energy multipath components (MPC). Similar performances are shown for large-MIMO systems with $n_t = n_r = 16, 32$ and $K = 64$ (problem size $Kn_t = 1024$ and 2048 dimensions). The per-symbol complexity of the MRF BP approach is $O(K^2 n_t^2)$ (details given in Section III-A5).

In the case of FG BP approach (Section III-B), the Gaussian approximation of interference (GAI) we adopt is found to be effective to further reduce the complexity by an order (Section III-B1); i.e., the per-symbol complexity of our FG-GAI BP approach is just $O(Kn_t)$, which is one order less than that of the MRF BP approach. We note that several Gaussian approximation based approaches are known in the literature in the context of multiuser detection in CDMA and MIMO detection [14],[30]-[34]. For example, MIMO detection algorithms based on probabilistic data association (PDA) principle are studied in [32],[33]. These algorithms approximate the interference-plus-noise *vector* as a Gaussian vector, and require inversion of the covariance matrix resulting in $O(n_t^2)$ per-symbol complexity. The way we do the approximation is different. At each receive node, we approximate the interference(from all other antennas)-plus-noise *scalar* as a Gaussian r.v. which requires the computation of only the mean and variance in closed-form. These approximated Gaussian statistics (mean and variance) are updated in each iteration. Because of the scalar Gaussian approximation, the per-symbol complexity of our algorithm is just linear in n_t. In addition, the algorithm exhibits large-system behavior and near-optimal performance in large dimensions. We note that the algorithm in [34] also uses scalar Gaussian approximation of interference which is different from ours on two major counts. First, the algorithm in [34] parametrizes the number of interference terms in the Gaussian approximation; at each receive node, only $n_t - d_f$, $1 \leq d_f \leq n_t$, weakest interfering terms are approximated as Gaussian, and d_f strongest terms are marginalized through message passing. The implication of this difference is that the complexity of the algorithm in [34] is exponential in d_f, i.e., $O(2^{d_f})$. Performance gets better for increasing d_f, making it prohibitive for large dimensions. The second difference is that the mean and variance in the approximation in [34] are kept fixed and are not updated from iteration to iteration, whereas we update these statistics in each iteration. A performance comparison with the algorithm in [34] shows that our algorithm performs better (Section III-C2). The algorithm in [30] also uses a Gaussian approximation that scales well for detection in large MIMO-ISI channels. But this algorithm exhibits high error floors, whereas our FG-GAI BP approach is seen to avoid flooring and perform significantly better (Fig. 14). We further show that the proposed FG-GAI BP algorithm can be used in an iterative manner with local neighborhood search algorithms, like the reactive tabu search (RTS) algorithm in [35], to improve the performance of M-QAM detection (Section IV).

The proposed algorithms can be extended to coded systems as well, through turbo equalization [21]-[23] (Receiver C in Fig. 1 of [23]) or through joint processing of the entire coded frame using low-complexity graphical models (low-complexity approximations of Receiver A in Fig. 1 of [23]). In [19], we have investigated a scheme with separate MRF BP detection followed by decoding (Receiver B in Fig. 1 of [23]) in a 24×24 V-BLAST MIMO system, and showed that a coded BER performance close to within 2.5 dB of the theoretical ergodic MIMO capacity is achieved. MIMO space-time coding schemes that can achieve separability of detection and decoding without loss of optimality [36] are interesting because they avoid the need for joint processing for optimal detection and decoding. If such detection-decoding separable space-time codes become available for large dimensions, the proposed

[2]We say that an algorithm exhibits 'large-dimension behavior' if its bit error performance improves with increasing number of dimensions. The fact that turbo codes with BP decoding achieve near-capacity performance only when the *frame sizes are large* is an instance of large-dimension behavior.

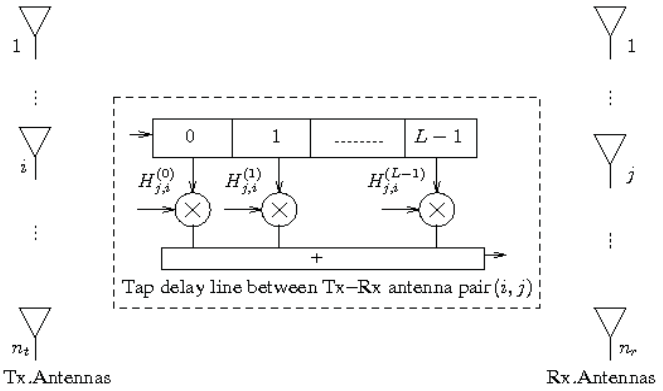

Fig. 1. MIMO-ISI Channel Model.

algorithms can be applicable in their detection/equalization.

The rest of the paper is organized as follows. In Section II, we present the considered MIMO system model in frequency-selective fading. In Section III, we present the proposed detection algorithms based on belief propagation on MRF and FG of the MIMO system. In Section IV, the proposed hybrid RTS-BP algorithm for detection of M-QAM signals and its performance are presented. Conclusions are presented in Section V.

II. System Model

We consider MIMO systems with cyclic prefixed single-carrier (CPSC) signaling, where the overall MIMO channel includes an FFT operation so that the transmitted symbols are estimated from the received frequency-domain signal [37]-[39]. Consider a frequency-selective MIMO channel with n_t transmit and n_r receive antennas as shown in Fig. 1. Let L denote the number of multipath components (MPC). Data is transmitted in frames, where each frame has K' channel uses, out of which data symbol vectors are sent in K channel uses $K \geq L$. These K channel uses are preceded by a cyclic prefix (CP) of length $L-1$ channel uses so that $K' = K+L-1$. In each channel use, an n_t-length data symbol vector is transmitted using spatial multiplexing on n_t transmit antennas. Let $\mathbf{x}_q \in \{\pm 1\}^{n_t}$ denote the data symbol vector transmitted in the qth channel use, $q = 0, 1, \cdots, K-1$. Though the symbol alphabet used here is BPSK, extensions to higher-order alphabet are possible and some are discussed later in the paper. While CP avoids inter-frame interference, there will be ISI within the frame. The received signal vector at time q can be written as

$$\mathbf{y}_q \quad = \quad \sum_{l=0}^{L-1} \mathbf{H}_l \, \mathbf{x}_{q-l} + \mathbf{w}_q, \qquad q = 0, \cdots, K-1, \quad (1)$$

where $\mathbf{y}_q \in \mathbb{C}^{n_r}$, $\mathbf{H}_l \in \mathbb{C}^{n_r \times n_t}$ is the channel gain matrix for the lth MPC such that $H_{j,i}^{(l)}$ denotes the entry on the jth row and ith column of the $\mathbf{H}_l$ matrix, i.e., $H_{j,i}^{(l)}$ is the channel from ith transmit antenna to the jth receive antenna on the lth MPC. The entries of $\mathbf{H}_l$ are assumed to be i.i.d $\mathbb{C}\mathcal{N}(0,1)$. It is further assumed that $\mathbf{H}_l$, $l = 0, \cdots, L-1$ remain constant for one frame duration, and vary i.i.d from one frame to the other. $\mathbf{w}_q \in \mathbb{C}^{n_r}$ is the additive white Gaussian noise vector at time q, whose entries are independent, each with variance $\sigma^2 = n_t L E_s/\gamma$, where γ is the average received SNR per received antenna. The CP will render the linearly convolving channel to a circularly convolving one, and so the channel will be multiplicative in frequency domain. Because of the CP, the received signal in frequency domain, for the ith frequency index ($0 \leq i \leq K-1$), can be written as

$$\mathbf{r}_i \quad = \quad \mathbf{G}_i \; \mathbf{u}_i + \mathbf{v}_i, \quad (2)$$

where $\mathbf{r}_i = \frac{1}{\sqrt{K}} \sum_{q=0}^{K-1} e^{\frac{-2\pi \mathbf{j} qi}{K}} \mathbf{y}_q$, $\mathbf{u}_i = \frac{1}{\sqrt{K}} \sum_{q=0}^{K-1} e^{\frac{-2\pi \mathbf{j} qi}{K}} \mathbf{x}_q$, $\mathbf{v}_i = \frac{1}{\sqrt{K}} \sum_{q=0}^{K-1} e^{\frac{-2\pi \mathbf{j} qi}{K}} \mathbf{w}_q$, $\mathbf{G}_i = \sum_{l=0}^{L-1} e^{\frac{-2\pi \mathbf{j} li}{K}} \mathbf{H}_l$, and $\mathbf{j} = \sqrt{-1}$. Stacking the K vectors $\mathbf{r}_i$, $i = 0, \cdots, K-1$, we write

$$\mathbf{r} \quad = \quad \underbrace{\mathbf{GF}}_{\triangleq \mathbf{H}_{eff}} \; \mathbf{x}_{eff} + \mathbf{v}_{eff}, \quad (3)$$

where

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_0 \\ \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_{K-1} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}_0 & & & \mathbf{0} \\ & \mathbf{G}_1 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{G}_{K-1} \end{bmatrix},$$

$$\mathbf{x}_{eff} = \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{K-1} \end{bmatrix}, \quad \mathbf{v}_{eff} = \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_{K-1} \end{bmatrix},$$

$$\mathbf{F} = \frac{1}{\sqrt{K}} \begin{bmatrix} \rho_{0,0}\mathbf{I}_{n_t} & \rho_{1,0}\mathbf{I}_{n_t} & \cdots & \rho_{K-1,0}\mathbf{I}_{n_t} \\ \rho_{0,1}\mathbf{I}_{n_t} & \rho_{1,1}\mathbf{I}_{n_t} & \cdots & \rho_{K-1,1}\mathbf{I}_{n_t} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{0,K-1}\mathbf{I}_{n_t} & \rho_{1,K-1}\mathbf{I}_{n_t} & \cdots & \rho_{K-1,K-1}\mathbf{I}_{n_t} \end{bmatrix}$$
$$= \frac{1}{\sqrt{K}} \mathbf{D}_K \otimes \mathbf{I}_{n_t},$$

where $\rho_{q,i} = e^{\frac{-2\pi \mathbf{j} qi}{K}}$, $\mathbf{D}_K$ is the K-point DFT matrix and $\otimes$ denotes the Kronecker product. Equation (3) can be written in an equivalent linear vector channel model of the form

$$\mathbf{r} \quad = \quad \mathbf{H}\mathbf{x} + \mathbf{v}, \quad (4)$$

where $\mathbf{H} = \mathbf{H}_{eff}$, $\mathbf{x} = \mathbf{x}_{eff}$, and $\mathbf{v} = \mathbf{v}_{eff}$. Note that the well known MIMO system model for flat fading can be obtained as a special case in the above system model with $L = K = 1$.

We further note that, in the considered system, signaling is done along K dimensions in time and n_t dimensions in space, so that the total number of dimensions involved is Kn_t. We are interested in low-complexity detection/equalization in large dimensions (i.e., for large Kn_t) using graphical models. The goal is to obtain an estimate of vector $\mathbf{x}$, given $\mathbf{r}$ and the knowledge of $\mathbf{H}$. The optimal maximum a posteriori probability (MAP) detector takes the joint posterior distribution

$$p(\mathbf{x} \mid \mathbf{r}, \mathbf{H}) \quad \propto \quad p(\mathbf{r} \mid \mathbf{x}, \mathbf{H}) \, p(\mathbf{x}), \quad (5)$$

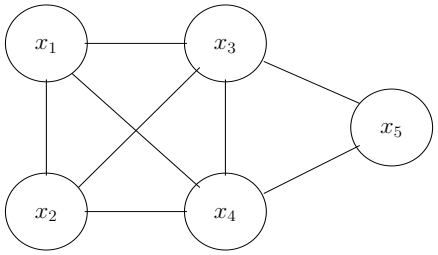

Fig. 2. An example of MRF.

and marginalizes out each variable as $p(x_i|\mathbf{r},\mathbf{H}) = \sum_{x_{-i}} p(\mathbf{x}|\mathbf{r},\mathbf{H})$, where x_{-i} stands for all entries of $\mathbf{x}$ except x_i. The MAP estimate of the bit x_i, $i = 1, \cdots, Kn_t$, is then given by

$$\widehat{x}_i \quad = \quad \underset{a \in \{\pm 1\}}{\arg\max}\ p\big(x_i = a \mid \mathbf{r},\mathbf{H}\big), \tag{6}$$

whose complexity is exponential in Kn_t. In the following sections, we present low-complexity detection algorithms based on graphical models suited for the system model in (4) with large dimensions, i.e., for large K, L, n_t, keeping L/K fixed.

III. Detection Using Belief Propagation on MRFs and FGs

In this section, we present detection algorithms based on message passing on graphical models of the MIMO system in (4). In general, joint distribution of random variables can be represented as graphical models, where nodes represent the variables and edges represent the dependencies among the variables. Message passing can be carried out on these graphs to make inferences [11]. In our MIMO detection problem, the variables of interest are the transmitted symbols x_i's, the joint distribution of interest is given by (5), and as an inference problem we seek to obtain an estimate of the solution given by (6) through message passing. Specifically, we consider message passing on Markov random field (MRF) and factor graph (FG) graphical models of the MIMO system characterized by (5) [20],[40].

A. Detection using BP on MRFs

1) MRFs: An undirected graph is given by $G = (V, E)$, where V is the set of nodes and $E \subseteq \{(i,j) : i,j \in V, i \neq j\}$ is the set of undirected edges. An MRF is an undirected graph whose vertices are random variables [10],[41]. The statistical dependency among the variables are such that any variable is independent of all the other variables, given its neighbors. Usually, the variables in an MRF are constrained by a *compatibility function*, also known as a *clique potential* in literature. A *clique* of an MRF is a fully connected sub-graph, i.e., it is a subset $C \subseteq V$ such that $(i,j) \in E$ for all $i,j \in C$. A clique is *maximal* if it is not a strict subset of another clique. Therefore, a maximal clique does not remain fully connected if any additional vertex of the MRF is included in it. For example, in the MRF shown in Fig. 2, $\{x_1, x_2, x_3, x_4\}$ and $\{x_3, x_4, x_5\}$ are two maximal cliques.

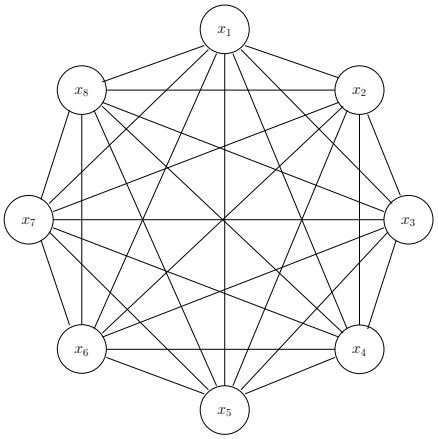

Fig. 3. Fully connected MRF of 8×8 MIMO system.

Let there be N_c maximal cliques in the MRF, and $\mathbf{x}_j$ be the variables in maximal clique j. Let $\psi_j(\mathbf{x}_j)$ be the clique potential of clique j. Then the joint distribution of the variables is given by Hammersley-Clifford theorem [42]

$$p(\mathbf{x}) \quad = \quad \frac{1}{Z}\prod_{j=1}^{N_c}\psi_j(\mathbf{x}_j), \tag{7}$$

where Z is a constant, also known as *partition function*, chosen to ensure the distribution is normalized. In Fig. 2, with two maximal cliques in the MRF, namely, $\{x_1, x_2, x_3, x_4\}$ and $\{x_3, x_4, x_5\}$, the joint probability distribution is given by

$$p(\mathbf{x}) \quad = \quad \frac{1}{Z}\,\psi_1(x_1, x_2, x_3, x_4)\,\psi_2(x_3, x_4, x_5). \tag{8}$$

Pairwise MRF: An MRF is called a *pairwise* MRF if all the maximal cliques in the MRF are of size two. In this case, the clique potentials are all functions of two variables. The joint distribution in such a case takes the form [11]

$$p(\mathbf{x}) \quad \propto \quad \Big(\prod_{(i,j)}\psi_{i,j}(x_i, x_j)\Big)\Big(\prod_i \phi_i(x_i)\Big), \tag{9}$$

where $\psi_{i,j}(x_i, x_j)$ is the clique potential between nodes x_i and x_j denoting the statistical dependence between them, and $\phi_i(x_i)$ is the self potential of node x_i.

2) MRF of MIMO System: The MRF of a MIMO system is a fully connected graph. Figure 3 shows the MRF for a 8×8 MIMO system. We get the MRF potentials for the MIMO system where the posterior probability function of the random vector $\mathbf{x}$, given $\mathbf{r}$ and $\mathbf{H}$, is of the form[3]

[3]In our detection problem, relative values of the distribution for various possibilities of $\mathbf{x}$ are adequate. So, we can omit the normalization constant Z which is independent of $\mathbf{x}$, and replace the equality with proportionality in the distribution.

$$\begin{aligned} p(\mathbf{x} \mid \mathbf{r}, \mathbf{H}) &\propto \exp\Big(\frac{-1}{2\sigma^2}\|\mathbf{r}-\mathbf{H}\mathbf{x}\|^2\Big)\exp\big(\ln p(\mathbf{x})\big) \\ &= \exp\Big(-\frac{1}{2\sigma^2}(\mathbf{r}-\mathbf{H}\mathbf{x})^H(\mathbf{r}-\mathbf{H}\mathbf{x})\Big) \qquad (10) \\ &\quad \cdot\prod_i \exp\big(\ln p(x_i)\big) \\ &\propto \exp\Big(-\frac{1}{2\sigma^2}\big(\mathbf{x}^H\mathbf{H}^H\mathbf{H}\mathbf{x} - 2\Re\{\mathbf{x}^H\mathbf{H}^H\mathbf{r}\}\big)\Big) \\ &\quad \cdot\prod_i \exp\big(\ln p(x_i)\big). \qquad (11) \end{aligned}$$

Now, defining $\mathbf{R} \triangleq \frac{1}{\sigma^2}\mathbf{H}^H\mathbf{H}$ and $\mathbf{z} \triangleq \frac{1}{\sigma^2}\mathbf{H}^H\mathbf{r}$, we can write (11) as

$$\begin{aligned} p(\mathbf{x} \mid \mathbf{r}, \mathbf{H}) &\propto \exp\Big(-\sum_{i<j}\Re\{x_i^* R_{ij} x_j\}\Big) \qquad (12) \\ &\quad \cdot\exp\Big(\sum_i \Re\{x_i^* z_i\}\Big)\prod_i \exp\big(\ln p(x_i)\big) \\ &= \Big(\prod_{i<j}\exp\Big(-x_i\Re\{R_{ij}\}x_j\Big)\Big) \qquad (13) \\ &\quad \cdot\Big(\prod_i \exp\Big(x_i\Re\{z_i\} + \ln p(x_i)\Big)\Big), \end{aligned}$$

where z_i and R_{ij} are the elements of $\mathbf{z}$ and $\mathbf{R}$, respectively. Comparing (14) and (9), we see that the MRF of the MIMO system has only pairwise interactions with the following potentials

$$\begin{aligned} \psi_{i,j}(x_i, x_j) &= \exp\Big(-x_i\Re\{R_{ij}\}x_j\Big), \qquad (14) \\ \phi_i(x_i) &= \exp\Big(x_i\Re\{z_i\} + \ln p(x_i)\Big). \qquad (15) \end{aligned}$$

3) Message Passing: The values of ψ and ϕ given by (14) and (15) define, respectively, the edge and self potentials of an undirected graphical model to which message passing algorithms, such as belief propagation, can be applied to compute the marginal probabilities of the variables. BP attempts to estimate the marginal probabilities of all the variables by way of passing messages between the local nodes.

A *message* from node j to node i is denoted as $m_{j,i}(x_i)$, and belief at node i is denoted as $\mathrm{b}_i(x_i)$, $x_i \in \{\pm 1\}$. The $\mathrm{b}_i(x_i)$ is proportional to how likely x_i was transmitted. On the other hand, $m_{ji}(x_i)$ is proportional to how likely node j thinks x_i was transmitted. The belief at node i is

$$\mathrm{b}_i(x_i) \propto \phi_i(x_i)\prod_{j\in\mathcal{N}(i)} m_{j,i}(x_i), \qquad (16)$$

where $\mathcal{N}(i)$ denotes the neighboring nodes of node i, and the messages are defined as [11]

$$m_{j,i}(x_i) \propto \sum_{x_j}\phi_j(x_j)\,\psi_{j,i}(x_j, x_i)\prod_{k\in\mathcal{N}(j)\setminus i} m_{k,j}(x_j). \qquad (17)$$

Equation (17) actually constitutes an iteration, as the message is defined in terms of the other messages. So, BP essentially involves computing the outgoing messages from a node to each of its neighbors using the local joint compatibility function and the incoming messages and transmitting them. The algorithm terminates after a fixed number of iterations.

4) Improvement through Damping: In systems characterized by fully/highly connected graphical models, BP based algorithms may fail to converge, and if they do converge, the estimated marginals may be far from exact [43],[44]. It may be expected that BP might perform poorly in MIMO graphs due to the high density of connections. However, several methods are known in the literature, including *double loop methods* [45],[46] and *damping* [47],[48],[49], which can be applied to improve things if BP does not converge (or converges too slowly). In this paper, we consider damping method.

Message Damped BP: In message damping, at each step of the algorithm, the evaluation of messages is taken to be a weighted average between the old estimate and the new estimate [47]. Denoting $\widetilde{m}_{i,j}^{(t)}(x_j)$ as the updated message in iteration t obtained by message passing, the new message from node i to node j in iteration t, denoted by $m_{i,j}^{(t)}(x_j)$, is computed as a convex combination of the old message and the updated message as

$$\widetilde{m}_{i,j}^{(t)}(x_j) \propto \sum_{x_i}\phi_i(x_i)\,\psi_{i,j}(x_i, x_j)\prod_{k\in\mathcal{N}(i)\setminus j} m_{k,i}^{(t-1)}(x_i), \qquad (18)$$

$$m_{i,j}^{(t)}(x_j) = \alpha_m\, m_{i,j}^{(t-1)}(x_j) + (1-\alpha_m)\,\widetilde{m}_{i,j}^{(t)}(x_j), \qquad (19)$$

where $\alpha_m \in [0, 1)$ is referred as the *message damping factor*. The proposed MRF BP algorithm employing message damping is listed in Table I.

5) Computation Complexity: The per-symbol complexity of calculating the messages in a single MRF BP iteration is $O(K^2 n_t^2)$ and the per-symbol complexity of calculating the beliefs is $O(Kn_t)$. Likewise, the per-symbol complexities of computing ϕ and ψ are $O(1)$ and $O(Kn_t)$, respectively. The computation of $\mathbf{z}$ can be carried out with $O(Kn_r)$ per-symbol complexity. The computation of $\mathbf{R}$ involves computation of $\mathbf{H}^H\mathbf{H}$, which involves three operations: i) computation of $\mathbf{G}$, ii) calculation of $\mathbf{G}^H\mathbf{G}$, and iii) multiplication of $\mathbf{F}^H$ and $\mathbf{F}$ with $\mathbf{G}^H\mathbf{G}$. The computation i) involves K-point FFT of matrices H_l, $l = 0, \cdots, L-1$, each H_l of dimension $n_r \times n_t$. The complexity associated with this operation is $O(n_t n_r K \log_2 K)$. The total number of symbols transmitted is Kn_t. So, the per-symbol complexity is $O(n_r \log_2 K)$. The computation ii) involves the calculation of ${\mathbf{G}_i}^H\mathbf{G}_i$ for $i = 0, \cdots, K-1$. The computation of each ${\mathbf{G}_i}^H\mathbf{G}_i$ has complexity $O(n_t^3)$. Due to block-diagonal structure of $\mathbf{G}$, K such computations can be done in $O(Kn_t^3)$ complexity, leading to a per-symbol complexity of $O(n_t^2)$. Likewise, due to the block-symmetric structure of $\mathbf{F}$, the per-symbol complexity corresponding to computation iii) is $O(Kn_t^2)$. Since the number of BP iterations is much less than Kn_t, the overall per-symbol complexity is of the proposed MRF BP detection algorithm is given by $O(K^2 n_t^2)$, which scales well for large Kn_t.

The simulated BER performance of the MRF BP detection algorithm is presented in Section III-C1.

B. Detection using BP on FGs with GAI

We next propose a FG based detection approach with Gaussian approximation of interference. Consider the MIMO

Initialization
1. $m_{i,j}^{(0)}(x_j) = 0.5$,
$p(x_i = 1) = p(x_i = -1) = 0.5, \quad \forall i, j = 1, \cdots, Kn_t$
2. $\widetilde{m}_{i,j}^{(0)}(x_j) = 0.5, \quad \forall i, j = 1, \cdots, Kn_t$
3. $\mathbf{z} = \frac{1}{\sigma^2}\mathbf{H}^H\mathbf{r}; \quad \mathbf{R} = \frac{1}{\sigma^2}\mathbf{H}^H\mathbf{H}$
4. for i = 1 to Kn_t
5. $\phi_i(x_i) = \exp\left(x_i \Re\{z_i\} + \ln(p(x_i))\right)$
6. end for
7. for i = 1 to Kn_t
8. for j = 1 to Kn_t, $j \neq i$
9. $\psi_{i,j}(x_i, x_j) = \exp\left(-x_i \Re\{R_{i,j}\} x_j\right)$
10. end for
11. end for
Iterative Update of Messages
12. for t = 1 to num_iter
Damped Message Calculation
13. for i = 1 to Kn_t
14. for j = 1 to Kn_t, $j \neq i$
15. $\widetilde{m}_{i,j}^{(t)}(x_j) \propto \sum_{x_i} \phi_i(x_i)\psi_{i,j}(x_i, x_j) \cdot \prod_{k \in \mathcal{N}(i)\setminus j} m_{k,i}^{(t-1)}(x_i)$
16. $m_{i,j}^{(t)}(x_j) = \alpha_m\, m_{i,j}^{(t-1)}(x_j) + (1-\alpha_m)\, \widetilde{m}_{i,j}^{(t)}(x_j)$
17. end for
18. end for
19. end for; End of for loop starting at line 12
Belief Calculation
20. for i = 1 to Kn_t
21. $\mathrm{b}_i(x_i) \propto \phi_i(x_i) \prod_{j \in \mathcal{N}(i)} m_{j,i}^{(num\_iter)}(x_i)$
22. end for
Detection of Data Bits
23. $\widehat{x}_i = \underset{x_i \in \{\pm 1\}}{\arg\max}\ \mathrm{b}_i(x_i), \quad \forall i = 1, \cdots, Kn_t$
24. Terminate

TABLE I
PROPOSED MRF BASED BP DETECTOR/EQUALIZER ALGORITHM.

system model in (4). We will treat each entry of the observation vector $\mathbf{r}$ as a function node (observation node) in a factor graph, and each transmitted symbol as a variable node. The received signal r_i can be written as

$$\begin{aligned} r_i &= \sum_{j=1}^{Kn_t} h_{ij}x_j + v_i \\ &= h_{ik}x_k + \underbrace{\sum_{j=1, j\neq k}^{Kn_t} h_{ij}x_j}_{Interference} + v_i. \end{aligned} \tag{20}$$

When computing the message from the ith observation node to the kth variable node, we make the following Gaussian approximation of the interference:

$$\begin{aligned} r_i &= h_{ik}x_k + \underbrace{\sum_{j=1, j\neq k}^{Kn_t} h_{ij}x_j + v_i}_{\triangleq\, z_{ik}}, \end{aligned} \tag{21}$$

where the interference plus noise term, z_{ik}, is modeled as $\mathbb{C}\mathcal{N}(\mu_{z_{ik}}, \sigma_{z_{ik}}^2)$ with

$$\mu_{z_{ik}} = \sum_{j=1, j\neq k}^{Kn_t} h_{ij}\mathbb{E}(x_j), \tag{22}$$

$$\sigma_{z_{ik}}^2 = \sum_{j=1, j\neq k}^{Kn_t} |h_{ij}|^2 \,\mathrm{Var}(x_j) + \sigma^2. \tag{23}$$

For BPSK signaling, the log-likelihood ratio (LLR) of the symbol $x_k \in \{+1, -1\}$ at observation node i, denoted by Λ_i^k, can be written as

$$\Lambda_i^k = \log\frac{p(r_i|\mathbf{H}, x_k = 1)}{p(r_i|\mathbf{H}, x_k = -1)} = \frac{4}{\sigma_{z_{ik}}^2}\Re\left(h_{ik}^*(r_i - \mu_{z_{ik}})\right). \tag{24}$$

The LLR values computed at the observation nodes are passed to the variable nodes (Fig. 4a). Using these LLRs, the variable nodes compute the probabilities

$$\begin{aligned} p_i^{k+} &\triangleq p_i(x_k = +1|\mathbf{r}) \\ &= \frac{\exp(\sum_{l=1, l\neq i}^{Kn_r} \Lambda_l^k)}{1 + \exp(\sum_{l=1, l\neq i}^{Kn_r} \Lambda_l^k)}, \end{aligned} \tag{25}$$

and pass them back to the observation nodes (Fig. 4b). This message passing is carried out for a certain number of iterations. Messages can be damped as described in Section III-A4 and then passed. Finally, x_k is detected as

$$\widehat{x}_k = \mathrm{sgn}\Big(\sum_{i=1}^{Kn_r} \Lambda_i^k\Big). \tag{26}$$

The algorithm listing is given in Table II. Note that approximating the interference as Gaussian greatly simplifies the computation of messages (as can be seen from the complexity discussion in the following subsection).

1) Computation Complexity: The computation complexity of the FG-GAI BP algorithm in the above involves i) LLR calculations at the observation nodes as per (24), which has $O(K^2 n_t n_r)$ complexity, and ii) calculation of probabilities at variable nodes as per (25), which also requires $O(K^2 n_t n_r)$ complexity[4]. Hence, the overall complexity of the algorithm is $O(K^2 n_t n_r)$ for detecting Kn_t transmitted symbols. So the per-symbol complexity is just $O(Kn_t)$ for $n_t = n_r$. Note that this complexity is one order less than that of the MRF BP approach in the previous section. Because of its linear complexity in K and n_t, the proposed FG approach with GAI is quite attractive for detection in large-dimension MIMO-ISI channels. In addition, the BER performance achieved by the algorithm in large dimensions is very good (as shown in the

[4]A naive implementation of (24) would require a summation over $Kn_t - 1$ variable nodes for each message, amounting to a complexity of order $O(K^3 n_t^2 n_r)$. However, the summation over $Kn_t - 1$ variables in (22) can be written in the form $\sum_{j=1}^{Kn_t} h_{ij}\mathbb{E}(x_j) - h_{ik}\mathbb{E}(x_k)$, where the computation of the full summation from $j = 1$ to Kn_t (which is independent of the variable index k) requires $Kn_t - 1$ additions. In addition, one subtraction operation for each k is required. The makes the complexity order for computing (22) to be only $O(K^2 n_t n_r)$. A similar argument holds for computation of the variance in (23), and hence the complexity of computing the LLR in (24) becomes $O(K^2 n_t n_r)$. Likewise, a similar rewriting of the summation in (25) leads to a complexity of $O(K^2 n_t n_r)$.

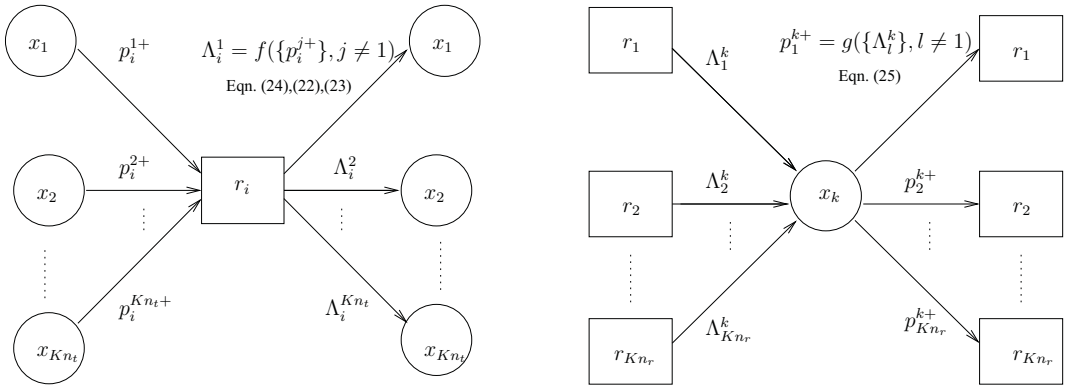

Fig. 4. Message passing between variable nodes and observation nodes in FG-GAI BP algorithm.

Initialization

1. $\Lambda_i^k = 0,\ p_i^{k+} = 0.5,\ s_{\Lambda^k} = 0,$
   $\mu_{z_{ik}} = \sigma_{z_{ik}}^2 = s_{\mu_{z_i}} = 0,$
   $s_{\sigma_{z_i}^2} = 0, \quad \forall i = 1, \cdots, Kn_r,\ \forall k = 1, \cdots, Kn_t$
2. for $t = 1$ to num_iter

   Computation of LLRs at observation nodes

3. for $i = 1$ to Kn_r
4. $s_{\mu_{z_i}} = \sum_{j=1}^{Kn_t} h_{ij}\left(2p_i^{j+} - 1\right)$
5. $s_{\sigma_{z_i}^2} = 4\sum_{j=1}^{Kn_t} |h_{ij}|^2 p_i^{j+}\left(1 - p_i^{j+}\right)$
6. for $k = 1$ to Kn_t
7. $\mu_{z_{ik}} = s_{\mu_{z_i}} - h_{ik}\left(2p_i^{k+} - 1\right)$
8. $\sigma_{z_{ik}}^2 = s_{\sigma_{z_i}^2} - 4|h_{ik}|^2 p_i^{k+}\left(1 - p_i^{k+}\right) + \sigma^2$
9. $\Lambda_i^k = \frac{4}{\sigma_{z_{ik}}^2}\Re\left(h_{ik}^*(r_i - \mu_{z_{ik}})\right)$
10. end for
11. end for

    Computation of probabilities at variable nodes

12. for $k = 1$ to Kn_t
13. $s_{\Lambda^k} = \sum_{l=1}^{Kn_r} \Lambda_l^k$
14. for $i = 1$ to Kn_r
15. $p_i^{k+} = \frac{exp(s_{\Lambda^k} - \Lambda_i^k)}{1 + exp(s_{\Lambda^k} - \Lambda_i^k)}$
16. end for
17. end for
18. end for; End of for loop starting at line 2

    Detection of data bits

19. for $k = 1$ to Kn_t
20. $\widehat{x}_k = \text{sgn}\left(\sum_{i=1}^{Kn_r} \Lambda_i^k\right)$
21. end for
22. Terminate

TABLE II
PROPOSED FG-GAI BASED BP DETECTOR/EQUALIZER ALGORITHM.

BER performance results in the Section III-C2).

C. Simulation Results

We now present the simulated BER performance of the proposed MRF BP and FG-GAI BP algorithms. In the simulations, i.i.d. Rayleigh fading channel realizations are generated for different data frames, and channel gains are held constant over one data frame of K channel uses. In all the simulations of MIMO-ISI channels, we have taken uniform power delay profile (i.e., all the L paths have equal energy). Also, perfect channel knowledge is assumed at the receiver.

1) Performance of MRF BP Algorithm in Sec. III-A: Performance in Flat-Fading with Large n_t: In Fig. 5, we illustrate the effect of damping and the large-dimension behavior of the MRF BP algorithm for large number (tens) of transmit and receive antennas with BPSK modulation on flat fading channels (i.e., $L = K = 1$). The number of MRF BP iterations is 5. Figure 5 shows the variation of the achieved BER as a function of the message damping factor, α_m, in 16×16 and 24×24 V-BLAST MIMO systems at an average received SNR per receive antenna, γ, of 8 dB. Note that $\alpha_m = 0$ corresponds to the case of undamped BP. In Fig. 5, we have also plotted the ZF and MMSE performances for comparison. In addition, maximum-likelihood (ML) performance evaluated through sphere decoding simulation are also plotted. It can be observed from Fig. 5 that, depending on the choice of the value of α_m, message damping can significantly improve the BER performance of the MRF BP algorithm. There is an optimum value of α_m at which the BER improvement over no damping case is maximum. For the chosen set of system parameters in Fig. 5, the optimum value of α_m is observed to be about 0.2. For this optimum value of $\alpha_m = 0.2$, it is observed that about an order of BER improvement is achieved with message damping compared to that without damping. As can also be seen, this performance is significantly better than ZF/MMSE performance and close to ML performance. It can be further seen that the performance improves for increasing $n_t = n_r$ (i.e., performance of the $n_t = n_r = 24$ system is better that of the $n_t = n_r = 16$ system).

In Fig. 6, we plot the BER performance of MRF BP in V-BLAST MIMO as a function of SNR for different $n_t = n_r = 4, 8, 16, 24, 32$ for $\alpha_m = 0.2$. MMSE and ML performance are plotted for comparison. While MMSE performance remains poor for large n_t, MRF BP performance shows large-dimension behavior whereby the performance improves and moves closer to unfaded SISO AWGN performance for increasing n_t. It is seen that for 32×32 V-BLAST MIMO, the

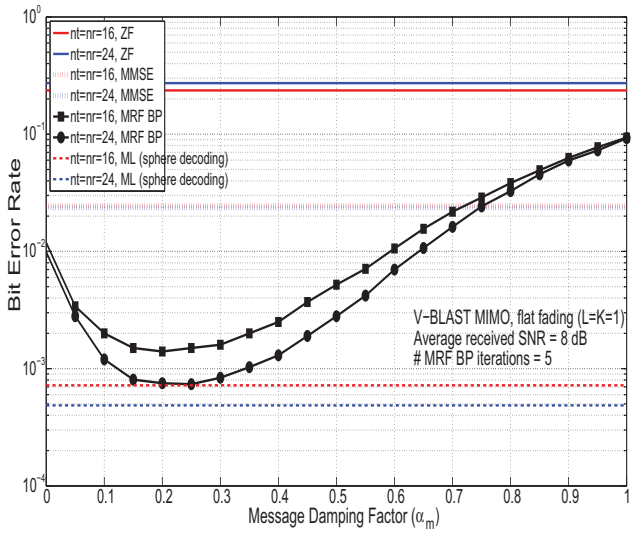

Fig. 5. BER performance of the MRF BP algorithm as a function of message damping factor, α_m, in V-BLAST MIMO with $n_t = n_r = 16, 24$ on flat fading ($L = K = 1$) at 8 dB SNR. # MRF BP iterations = 5.

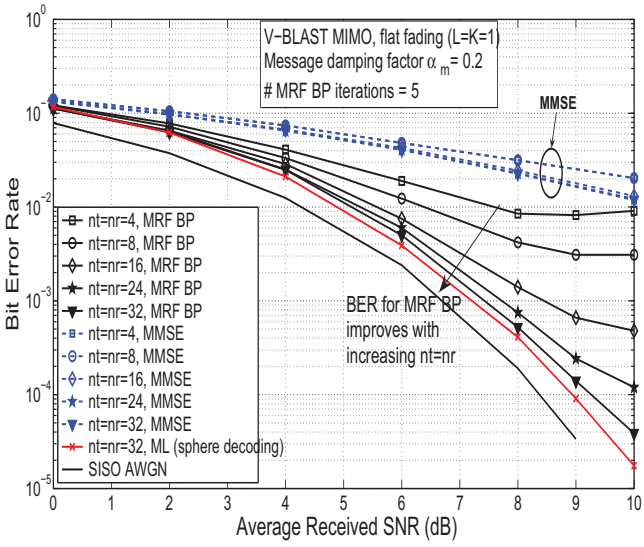

Fig. 6. BER performance of the MRF BP algorithm as a function of SNR in V-BLAST MIMO for different $n_t = n_r$ on flat fading ($L = K = 1$) with message damping $\alpha_m = 0.2$ and # MRF BP iterations = 5.

MRF BP performance is close to ML performance predicted by sphere decoding simulation. For more than 32 dimensions, sphere decoder simulation is prohibitively complex. In such cases, unfaded SISO AWGN performance can serve as a lower bound on ML performance[5]. The nearness to sphere decoder and SISO AWGN performances shown in Figs. 5 and 6 illustrates the near-optimal performance of MRF BP in large dimensions.

Performance in MIMO-ISI Channels with Large Kn_t: In Fig. 7, we explore the effect of message damping on the BER

[5]Henceforth, since sphere decoding is prohibitive for more than 32 dimensions, for performances in more than 32 dimensions we will give SISO AWGN performance as a lower bound on ML performance.

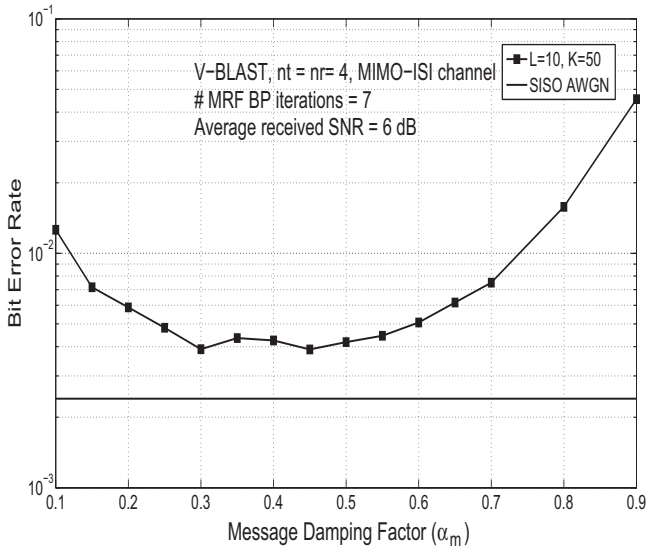

Fig. 7. BER performance of the MRF BP algorithm as a function of the message damping factor, α_m, in MIMO-ISI channels. $n_t = n_r = 4$, $[L = 10, K = 50]$, uniform power delay profile, average received SNR = 6 dB, # MRF BP iterations = 7.

performance of the MRF BP detector/equalizer in MIMO-ISI channels. Figure 7 shows the variation of the achieved BER as a function of the message damping factor, α_m, for $n_t = n_r = 4$, BPSK, $[L = 10, K = 50]$, at an average received SNR of 6 dB. The number of dimensions Kn_t is 200. The number of MRF BP iterations used is 7. From Fig. 7, it is can be seen that damping can significantly improve the BER performance of the MRF BP algorithm. For the chosen set of system parameters in Fig. 7, the optimum value of α_m is observed to be about 0.45, which gives about an order of BER improvement. This point of the damping benefit in terms of BER performance (and also in terms of convergence) is even more clearly brought out in Fig. 8, where we have compared the BER performance without damping ($\alpha_m = 0$) and with damping ($\alpha_m = 0.45$) in severely delay-spread MIMO-ISI channel with $[L = 20, K = 100]$, $n_t = n_r = 4$ (number of dimensions = 400) at an SNR of 7 dB as a function of the number of MRF BP iterations. It is interesting to see that without damping (i.e., with $\alpha_m = 0$), the algorithm indeed shows 'divergence' behavior, i.e., BER increases as number of iterations is increased beyond 4. Such divergence behavior is effectively removed by damping, as can be seen from the BER performance achieved with $\alpha_m = 0.45$. Indeed, the algorithm with damping ($\alpha_m = 0.45$) is seen to converge smoothly. It is also interesting to note that the algorithm converges to a BER which is quite close to the unfaded SISO AWGN BER; BER on SISO AWGN channel at 7 dB SNR is about 7.8×10^{-4} and the converged BER using damped MRF BP is about 1×10^{-3}. It is noted that damping (as per Eqn. (19)) does not increase the order of complexity of the algorithm without damping, and that the order of complexity without and with damping remains the same.

Comparison with MIMO-OFDM Performance: In Fig. 9, we present a performance comparison between the considered

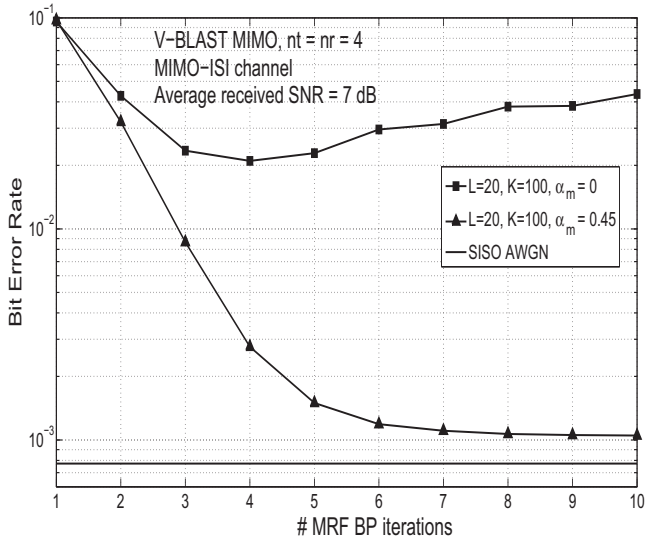

Fig. 8. Comparison of the BER performance of message damped and undamped MRF BP detector/equalizer as a function of number of BP iterations in MIMO-ISI channels. $n_t = n_r = 4$, $[L = 20, K = 100]$, uniform power delay profile, average received SNR = 7 dB, $\alpha_m = 0$ (undamped), $\alpha_m = 0.45$ (damped).

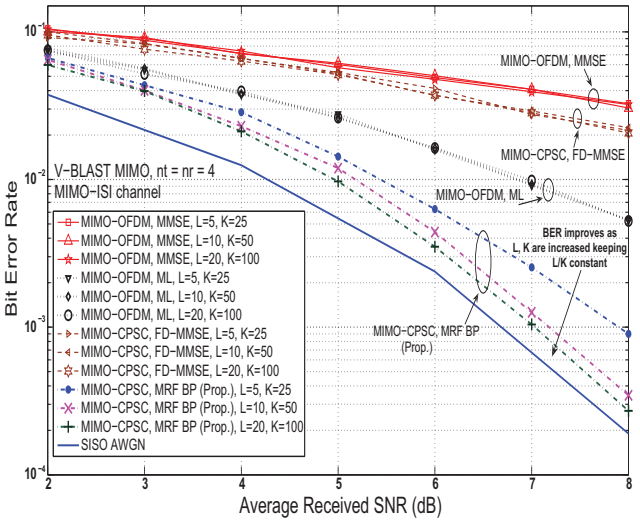

Fig. 9. BER performance of message damped MRF BP detector/equalizer as a function of average received SNR in MIMO-ISI channels with $n_t = n_r = 4$ for different values of L and K keeping L/K constant: $[L = 5, K = 25]$, $[L = 10, K = 50]$, and $[L = 20, K = 100]$. Uniform power delay profile. # MRF BP iterations = 10, $\alpha_m = 0.45$.

MIMO-CPSC scheme and a MIMO-OFDM scheme for the same system/channel parameters in both cases; for $n_t = n_r = 4$ and following combinations of L and K: $[L = 5, K = 25]$, $[L = 10, K = 50]$, $[L = 20, K = 100]$. For MIMO-CPSC, two detection schemes are considered: frequency-domain MMSE (FD-MMSE) and proposed MRF BP. For the MRF BP, number of BP iterations used is 10 and the value of α_m used is 0.45. For MIMO-OFDM, two detection schemes, namely, MMSE and ML detection on each subcarrier are considered. We have also plotted the unfaded SISO AWGN performance that serves as a lower bound on the optimum detection performance. The following observations can be made from Fig. 9: i) MIMO-OFDM with MMSE detection performs the worst among all the considered system/detection configurations, ii) MIMO-CPSC with FD-MMSE performs better than MIMO-OFDM with MMSE (this better performance in CPSC is in line with other reported comparisons between OFDM and CPSC, e.g., [37],[38],[39]), iii) at the expense of increased detection complexity, MIMO-OFDM with ML detection performs better than both MIMO-OFDM with MMSE and MIMO-CPSC with FD-MMSE, and iv) more interestingly, MIMO-CPSC with the proposed low-complexity MRF BP detection significantly outperforms MIMO-OFDM even with ML detection. Indeed, the performance of the MIMO-CPSC with MRF BP detection gets increasingly closer to the SISO AWGN performance for increasing L, K, keeping L/K constant. For example, the gap between the MRF BP performance and the SISO AWGN performance is only about 0.25 dB for $L = 20$ at a BER of 10^{-3}. This illustrates the ability of the MRF BP algorithm to achieve near-optimal performance for severely delay spread MIMO-ISI channels (i.e., large L) as witnessed in UWB systems.

A comparison of the turbo coded BER performance of

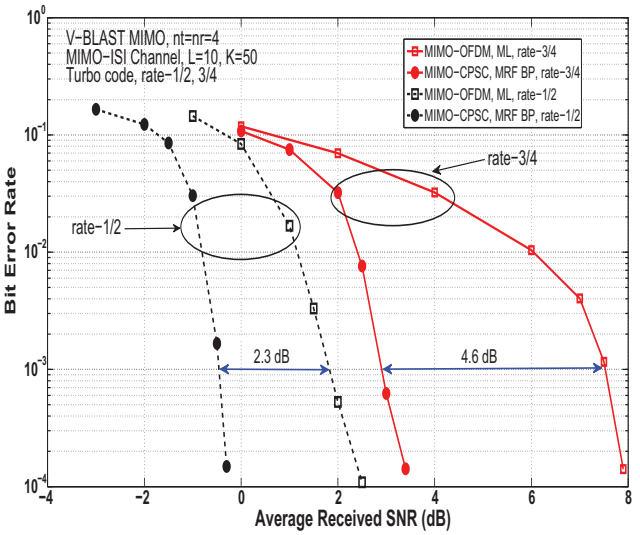

Fig. 10. Turbo coded BER performance of MIMO-CPSC with MRF BP detection and MIMO-OFDM with ML detection in MIMO-ISI channel with $n_t = n_r = 4$, $[L = 10, K = 50]$, uniform power delay profile. # MRF BP iterations = 10, $\alpha_m = 0.45$.

MIMO-CPSC with MRF BP detection and MIMO-OFDM with ML detection is shown in Fig. 10 for $n_t = n_t = 4$, $[L = 10, K = 50]$, and BPSK. Rate-1/2 and rate-3/4 turbo codes are considered. In Fig. 10, we observe that the turbo coded BER of MIMO-CPSC with MRF-BP is significantly better than that of MIMO-OFDM with ML. At a coded BER of 10^{-3}, MIMO-CPSC with MRF-BP outperforms MIMO-OFDM with ML by about 2.3 dB for rate-1/2 coding and by about 4.6 dB for rate-3/4 coding.

2) Performance of FG-GAI BP Algorithm in III-B: Figure 11 shows the simulated BER performance of the FG-GAI BP algorithm in $n_t \times n_r$ V-BLAST MIMO with $n_t = n_r =$

$8, 16, 32$ and BPSK on flat fading ($L = K = 1$). The number of FG-GAI BP iterations and message damping factor used are 10 and 0.4, respectively. We observe that, like the MRF BP approach, the FG-GAI BP approach also exhibits large-dimension behavior. For example, in 32×32 V-BLAST MIMO, the FG-GAI BP performance is close to sphere decoding performance showing FG-GAI BP's near-optimality in large dimensions. Figure 11 also shows the performance of the edge-based regular-d_f (EBRDF) algorithm in [34], which used a different Gaussian approximation as described in Section I. BER plots of EBRDF for 8×8 V-BLAST MIMO with $d_f = 7$ and 5 are shown. It is seen that FG-GAI BP performance is better than EBRDF performance. In addition, the complexity of FG-GAI BP is linear in n_t, whereas EBRDF complexity is exponential in d_f making EBRDF not attractive for large dimensions.

Next, in Fig. 12, we plot the performance of FG-GAI BP in MIMO-ISI channel with $L = 6$ and $K = 64$ for $n_t = n_r = 4, 8, 16$. Since the number of dimensions Kn_t here is more than 32 ($Kn_t = 256, 512, 1024$), we plot SISO AWGN performance for comparison instead of sphere decoding performance. It can be seen that the performance is quite close to SISO AWGN performance illustrating FG-GAI BP's near-optimality in MIMO-ISI channels with large dimensions. Figure 13 presents a comparison of the performances achieved by the MRF BP and FG-GAI BP approaches for the following system settings: $n_t = n_r = 4$, $[L = 5, K = 25]$ and $[L = 20, K = 100]$. It can be seen that, for these system settings, the FG-GAI BP approach performs almost the same as the MRF BP approach, at one order lesser complexity than that of the MRF BP approach.

Figure 14 presents a comparison of the performances achieved by the proposed FG-GAI BP scheme and the scheme in [30] for $n_t = n_r = 4$ and $[L = 4, K = 400]$. It can be seen that while the scheme in [30] exhibits an error floor, the proposed FG-GAI BP avoids flooring and achieves much better performance. Such good performance is achieved because equalization is done jointly on all the Kn_t symbols in a frame. The complexity of the scheme in [30] is $O(Ln_t)$, whereas the complexity of the proposed scheme is $O(Kn_t)$. Though $K > L$, the linear complexity of the proposed scheme in K is still very attractive. Also, as in the case of MRF BP, MIMO-CPSC with FG-GAI BP detection performs significantly better than MIMO-OFDM with ML detection.

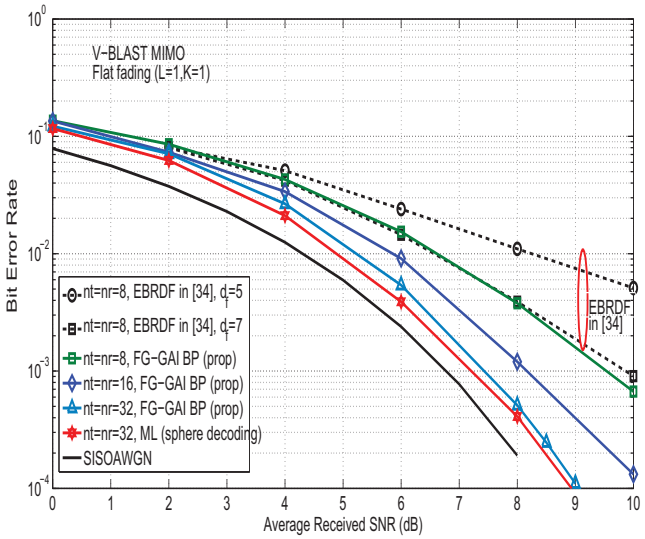

Fig. 11. BER performance of the FG-GAI BP algorithm in V-BLAST MIMO systems with $n_t = n_r = 8, 16, 32$ on flat fading ($L = K = 1$). # FG-GAI BP iterations = 10, $\alpha_m = 0.4$.

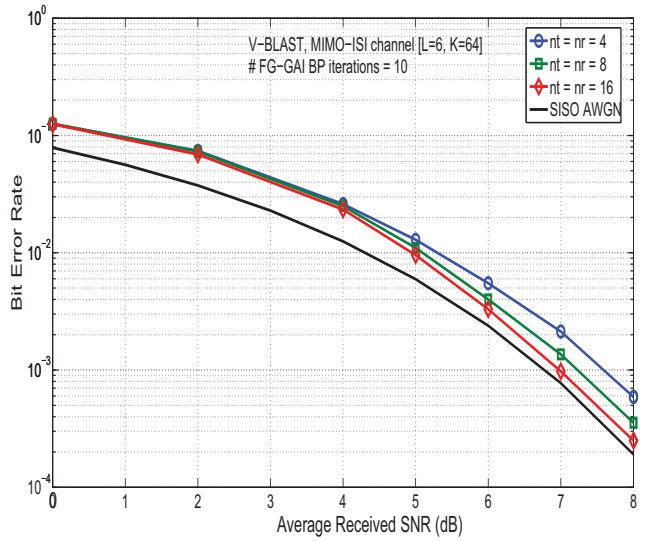

Fig. 12. BER performance of the FG-GAI BP algorithm in MIMO-ISI channels with for $[L = 6, K = 64]$ and $n_t = n_r = 4, 8, 16$. Uniform power delay profile, # FG-GAI BP iterations = 10, $\alpha_m = 0.4$.

IV. Hybrid Algorithms Using BP and Local Neighborhood Search for M-QAM

The MRF and FG-GAI BP algorithms proposed in the previous section work well for BPSK modulation, i.e., for $\mathbf{x} \in \{\pm 1\}^{Kn_t}$. They can work for 4-QAM also by viewing the transmit symbol vector to be in $\{\pm 1\}^{2Kn_t}$. Low-complexity algorithms for detection/equalization for higher-order M-QAM, $M > 4$, over large dimension MIMO-ISI channels are of interest. A BP based algorithm that is suited for higher-order QAM in MIMO has been reported recently in [50], which uses a Gaussian tree approximation (GTA) to convert the fully-connected graph representing the MIMO system into a tree and carries out BP on the resultant approximate tree. We refer to this algorithm in [50] as the GTA algorithm. In this section, we take an alternate hybrid approach for efficient detection of M-QAM signals, where the proposed FG-GAI BP algorithm for BPSK is used to improve the M-QAM detection performance of local neighborhood search algorithms, including tabu search algorithm [51],[52].

Local Neighborhood Search Based Detection: Low complexity search algorithms that attempt to minimize the maximum-likelihood (ML) cost $\|\mathbf{r} - \mathbf{Hx}\|^2$, by limiting the search space to local neighborhood have been proposed for detection of M-QAM signals in MIMO – e.g., tabu search (TS) algorithm [35]. Such local neighborhood search algo-

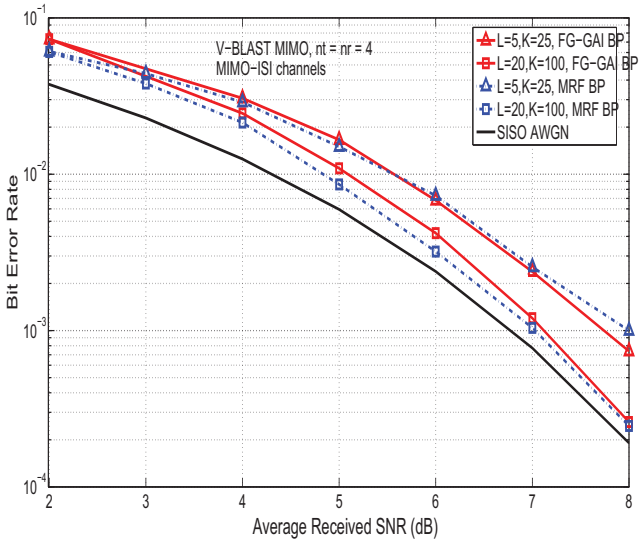

Fig. 13. Comparison of the BER performances of the MRF BP and FG-GAI BP algorithms in MIMO-ISI channels with $n_t = n_r = 4$, $[L = 5, K = 25]$, $[L = 20, K = 100]$, uniform power delay profile.

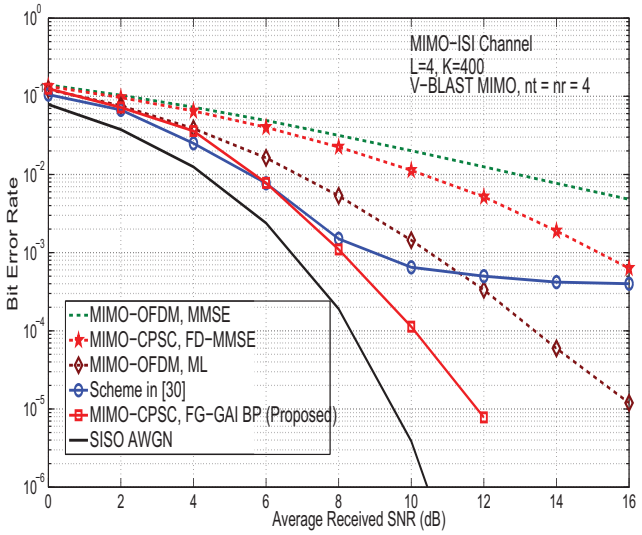

Fig. 14. Comparison of the BER performances of the FG-GAI BP scheme and the scheme in [30] in MIMO-ISI channels with $n_t = n_r = 4$, $[L = 4, K = 400]$, uniform power delay profile.

rithms have the advantage of low-complexity (e.g., TS algorithms, like the proposed MRF BP algorithm, has quadratic complexity in Kn_t), making them suited for large dimensions. However, their higher-order QAM performance is away from optimal performance. Here, we propose to improve the M-QAM performance of these search algorithms through the application of the proposed BP algorithms on the search algorithm outputs. This approach essentially improves the reliability of the output symbols from the local neighborhood search, thereby improving the overall BER performance. We apply this hybrid approach to the reactive tabu search (RTS) algorithm in [35]. Though other low-complexity search algorithms can be considered as a pre-processor, we chose RTS algorithm as the pre-processor because of its better performance than other local search algorithms with same order of complexity (e.g., LAS algorithm in [53],[54]).

Hybrid RTS-BP Approach: In the following subsections, we first present a brief summary of the RTS algorithm in [35] and the motivation behind the proposed hybrid approach. Next, we present the proposed hybrid RTS-BP algorithm [55] and its BER performance. Finally, we present a method to reduce complexity based on the knowledge of the simulated pdf of the RTS algorithm output.

A. Reactive Tabu Search (RTS) Algorithm

Here, we present a brief summary of the RTS algorithm in [35]. The RTS algorithm starts with an initial solution vector, defines a neighborhood around it (i.e., defines a set of neighboring vectors based on a neighborhood criteria), and moves to the best vector among the neighboring vectors (even if the best neighboring vector is worse, in terms of ML cost $\|\mathbf{r} - \mathbf{Hx}\|^2$, than the current solution vector); this allows the algorithm to escape from local minima. This process is continued for a certain number of iterations, after which the algorithm is terminated and the best among the solution vectors in all the iterations is declared as the final solution vector. In defining the neighborhood of the solution vector in a given iteration, the algorithm attempts to avoid cycling by making the moves to solution vectors of the past few iterations as 'tabu' (i.e., prohibits these moves), which ensures efficient search of the solution space. The number of these past iterations is parametrized as the 'tabu period,' which is dynamically changed depending on the number of repetitions of the solution vectors that are observed in the search path (e.g., increase the tabu period if more repetitions are observed). The per-symbol complexity of the RTS algorithm is quadratic in Kn_t for $n_t = n_r$.

B. Motivation for Hybrid RTS-BP Algorithm

The proposed hybrid RTS-BP approach is motivated by the following two observations we made in our BER simulations of the RTS algorithm: i) the RTS algorithm performed very close to optimum performance in large dimensions for 4-QAM; however, its higher-order QAM performance is far from optimal, and ii) at moderate to high SNRs, when an RTS output vector is in error, the least significant bits (LSB) of the data symbols are more likely to be in error than other bits. An analytical reasoning for the second observation can be given as follows.

Let the transmitted symbols take values from M-QAM alphabet $\mathbb{A}$, so that $\mathbf{x} \in \mathbb{A}^{n_t}$ is the transmitted vector. Consider the real-valued system model corresponding to (4), given by

$$\mathbf{r}' = \mathbf{H}'\mathbf{x}' + \mathbf{v}', \tag{27}$$

where

$$\mathbf{H}' = \begin{bmatrix} \Re(\mathbf{H}) & -\Im(\mathbf{H}) \\ \Im(\mathbf{H}) & \Re(\mathbf{H}) \end{bmatrix}, \quad \mathbf{r}' = \begin{bmatrix} \Re(\mathbf{r}) \\ \Im(\mathbf{r}) \end{bmatrix},$$
$$\mathbf{x}' = \begin{bmatrix} \Re(\mathbf{x}) \\ \Im(\mathbf{x}) \end{bmatrix}, \quad \mathbf{v}' = \begin{bmatrix} \Re(\mathbf{v}) \\ \Im(\mathbf{v}) \end{bmatrix}.$$

$\mathbf{x}'$ is a $2Kn_t \times 1$ vector; $[x'_1, \cdots, x'_{Kn_t}]$ can be viewed to be from an underlying M-PAM signal set, and so is $[x'_{Kn_t+1}, \cdots, x'_{2Kn_t}]$. Let $\mathbb{B} = \{a_1, a_2, \cdots, a_M\}$ denote the M-PAM alphabet that x'_i takes its value from.

Let $\widehat{\mathbf{x}}'$ denote the detected output vector from the RTS algorithm corresponding to the transmitted vector $\mathbf{x}'$. Consider the expansion of the M-PAM symbols in terms of ± 1's, where we can write the value of each entry of $\widehat{\mathbf{x}}'$ as a linear combination of ± 1's as

$$\widehat{x}'_i \;=\; \sum_{j=0}^{N-1} 2^j \widehat{b}_i^{(j)}, \quad i = 1, \cdots, 2Kn_t, \tag{28}$$

where $N = \log_2 M$ and $\widehat{b}_i^{(j)} \in \{\pm 1\}$. We note that the RTS algorithm outputs a local minima as the solution vector. So, $\widehat{\mathbf{x}}'$, being a local minima, satisfies the following conditions:

$$\|\mathbf{r}' - \mathbf{H}'\widehat{\mathbf{x}}'\|^2 \leq \|\mathbf{r}' - \mathbf{H}'(\widehat{\mathbf{x}}' + \lambda_i \mathbf{e}_i)\|^2, \; \forall i = 1, .., 2Kn_t, \tag{29}$$

where $\lambda_i = (a_q - \widehat{x}'_i)$, $q = 1, \cdots, M$, and $\mathbf{e}_i$ denotes the ith column of the identity matrix. Defining $\mathbf{F}' \triangleq \mathbf{H}'^T\mathbf{H}'$ and denoting the ith column of $\mathbf{H}'$ as $\mathbf{h}_i$, the conditions in (29) reduce to

$$2\lambda_i \mathbf{r}'^T \mathbf{h}_i \;\leq\; 2\lambda_i (\mathbf{H}'\widehat{\mathbf{x}}')^T \mathbf{h}_i + \lambda_i^2 f_{ii}, \tag{30}$$

where f_{ij} denotes the (i,j)th element of $\mathbf{F}'$. Substituting for $\mathbf{r}'$ in the above equation, and ignoring noise under moderate to high SNR conditions, (30) can be written as

$$2\lambda_i (\mathbf{x}' - \widehat{\mathbf{x}}')^T \mathbf{f}_i \;\leq\; \lambda_i^2 f_{ii}, \tag{31}$$

where $\mathbf{f}_i$ denotes the ith column of $\mathbf{F}'$. If $\lambda_i > 0$,

$$2(\mathbf{x}' - \widehat{\mathbf{x}}')^T \mathbf{f}_i \;\leq\; \lambda_i f_{ii}. \tag{32}$$

If $\lambda_i \leq 0$,

$$2(\mathbf{x}' - \widehat{\mathbf{x}}')^T \mathbf{f}_i \;\geq\; \lambda_i f_{ii}. \tag{33}$$

Combining the above two equations, we can write

$$2(\mathbf{x}' - \widehat{\mathbf{x}}')^T \mathbf{f}_i \,\mathrm{sgn}(\lambda_i) \;\leq\; \lambda_i f_{ii} \,\mathrm{sgn}(\lambda_i). \tag{34}$$

For Rayleigh fading, f_{ii} is chi-square distributed with $2KN_t$ degrees of freedom with mean KN_t. Approximating the distribution of f_{ij} to be normal with mean zero and variance $\frac{KN_t}{4}$ for $i \neq j$ by central limit theorem, we can drop the $\mathrm{sgn}(\lambda_i)$ in (34). Using the fact that the minimum value of $|\lambda_i|$ is 2, (34) can be simplified as

$$\sum_{x'_j \neq \widehat{x}'_j} \Delta_j f_{ij} \;\leq\; f_{ii}, \tag{35}$$

where $\Delta_j = x'_j - \widehat{x}'_j$. Also, if $x'_i = \widehat{x}'_i$, by the normal approximation in the above

$$\sum_{x'_j \neq \widehat{x}'_j} \Delta_j f_{ij} \;\sim\; \mathcal{N}\Big(0, \frac{KN_t}{4} \sum_{x'_j \neq \widehat{x}'_j} \Delta_j^2\Big). \tag{36}$$

The LHS of (35) is normal with variance proportional to Δ_j^2, and the RHS is positive. If Δ_j increases, then the variance of the LHS of (35) increases, and hence the probability that LHS $\leq$ RHS decreases. Then, for the inequality in (35), Δ_j takes smaller values with higher probability. Hence, the symbols of $\widehat{\mathbf{x}}'$ are nearest Euclidean neighbors of their corresponding symbols of the transmitted vector with high probability[6]. Now, because of the symbol-to-bit mapping in (28), $\widehat{x}'_i$ will differ from its nearest Euclidean neighbors certainly in the LSB position, and may or may not differ in other bit positions. Consequently, the LSBs of the symbols in the RTS output $\widehat{\mathbf{x}}'$ are least reliable.

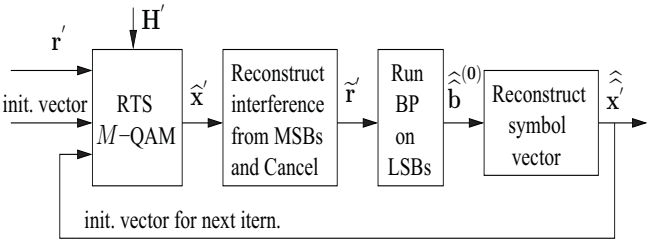

Fig. 15. Hybrid RTS-BP algorithm.

The above observation then led us to consider improving the reliability of the LSBs of the RTS output using the proposed FG-GAI BP algorithm presented in Section III-B, and iterate between RTS and FG-GAI BP as follows.

C. Proposed Hybrid RTS-BP Algorithm

Figure 15 shows the block schematic of the proposed hybrid RTS-BP algorithm. The following four steps constitute the proposed algorithm.

- *Step 1:* Obtain $\widehat{\mathbf{x}}'$ using the RTS algorithm. Obtain the output bits $\widehat{b}_i^{(j)}$, $i = 1, \cdots, 2Kn_t$, $j = 0, \cdots, N-1$, from $\widehat{\mathbf{x}}'$ and (28).
- *Step 2:* Using the $\widehat{b}_i^{(j)}$'s from Step 1, reconstruct the interference from all bits other than the LSBs (i.e., interference from all bits other than $\widehat{b}_i^{(0)}$'s) as

$$\widetilde{\mathbf{I}} \;=\; \sum_{j=1}^{N-1} 2^j \,\mathbf{H}' \,\widehat{\mathbf{b}}^{(j)}, \tag{37}$$

  where $\widehat{\mathbf{b}}^{(j)} = \big[\widehat{b}_1^{(j)}, \widehat{b}_2^{(j)}, \ldots, \widehat{b}_{2Kn_t}^{(j)}\big]^T$. Cancel the reconstructed interference in (37) from $\mathbf{r}$ as

$$\widetilde{\mathbf{r}}' \;=\; \mathbf{r}' - \widetilde{\mathbf{I}}. \tag{38}$$

- *Step 3:* Run the FG-GAI BP algorithm in Section III-B on the vector $\widetilde{\mathbf{r}}'$ in Step 2, and obtain an estimate of the LSBs. Denote this LSB output vector from FG-GAI BP as $\widehat{\widehat{\mathbf{b}}}^{(0)}$. Now, using $\widehat{\widehat{\mathbf{b}}}^{(0)}$ from the BP output, and the $\widehat{\mathbf{b}}^{(j)}$, $j = 1, \cdots, N-1$ from the RTS output in Step 1, reconstruct the symbol vector as

$$\widehat{\widehat{\mathbf{x}}}' \;=\; \widehat{\widehat{\mathbf{b}}}^{(0)} + \sum_{j=1}^{N-1} 2^j \,\widehat{\mathbf{b}}^{(j)}. \tag{39}$$

- *Step 4:* Repeat Steps 1 to 3 using $\widehat{\widehat{\mathbf{x}}}'$ as the initial vector to the RTS algorithm.

[6]Because x'_i's and $\widehat{x}'_i$'s take values from M-PAM alphabet, $\widehat{x}'_i$ is said to be the Euclidean nearest neighbor of x_i if $|x'_i - \widehat{x}'_i| = 2$.

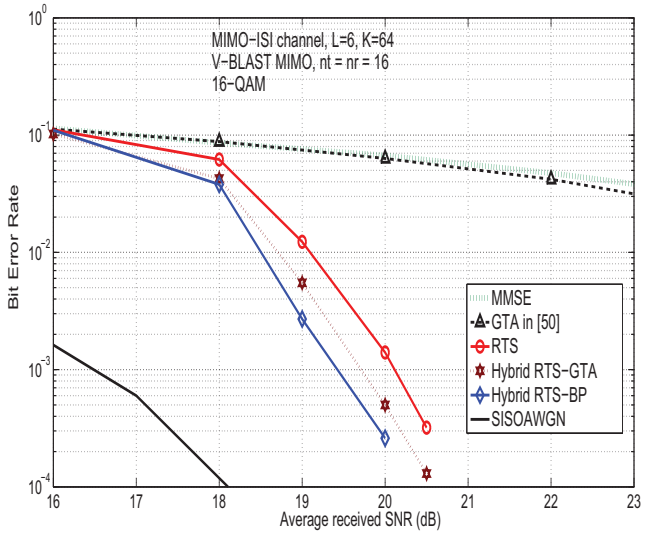

Fig. 16. BER performance comparison between proposed RTS-BP, MMSE, GTA in [50], RTS, and RTS-GTA in 16×16 V-BLAST MIMO with 16-QAM in MIMO-ISI channel with $L = 6$, $K = 64$, uniform power-delay profile.

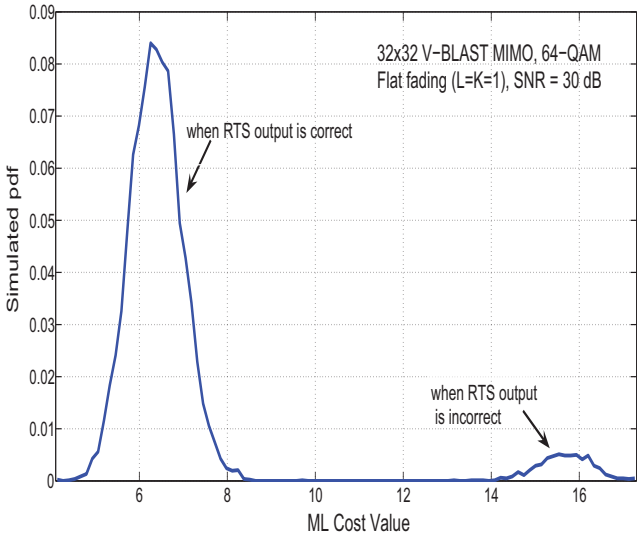

Fig. 17. Simulated pdfs of the ML cost of the RTS output vector (M_1) in a 32×32 V-BLAST MIMO system with 64-QAM and SNR = 30 dB on flat fading ($L = K = 1$).

The algorithm is stopped after a certain number of iterations between RTS and BP. Our simulations showed that two iterations between RTS and BP are adequate to achieve good improvement; more than two iterations resulted in only marginal improvement for the system parameters considered in the simulations. Since the complexity of BP part of RTS-BP is less than that of the RTS part, the order of complexity of RTS-BP is same as that of RTS, $O(K^2 n_t^2)$.

D. Simulation Results

Figure 16 shows the BER performance of the proposed hybrid RTS-BP algorithm in 16×16 V-BLAST MIMO with 16-QAM on a frequency-selective fading channel with $L = 6$ equal energy multipath components and $K = 64$ data vectors per frame. Two iterations between RTS and FG-GAI BP are used. For comparison, we plot SISO AWGN performance, MMSE performance, and performance of the GTA algorithm in [50]. Though the GTA algorithm in [50] scales well for large dimensions, its performance is near to MMSE performance and is far from SISO AWGN performance. On the other hand, the proposed hybrid RTS-BP algorithm performs much closer to SISO AWGN performance. This is because of the improvement in the reliability of LSBs due to FG-GAI BP algorithm run on them. We also show the performance of a hybrid algorithm using RTS and GTA (termed as hybrid RTS-GTA), where RTS is used as a pre-processor to generate priors for GTA (instead of uniform priors used in [50]) and multiple iterations between RTS and GTA are carried out. Two iterations between RTS and GTA are used in the simulations. It can be seen that the RTS-GTA performs better than both RTS and GTA, and RTS-BP performs better than RTS-GTA.

E. Complexity Reduction Using Selective BP

In the proposed RTS-BP algorithm, the use of BP at the RTS output was done unconditionally. Whereas the use of BP can improve performance only when the RTS output is erroneous. So, the additional complexity due to BP can be avoided if BP is not carried out whenever the RTS output is error-free. To decide whether to use BP or not, we can use the knowledge of the simulated pdf of the ML cost of the RTS output vector, i.e., the pdf of $M_1 \stackrel{\triangle}{=} \|\mathbf{r}' - \mathbf{H}'\widehat{\mathbf{x}}'\|$. Figure 17 shows the simulated pdf of M_1 for a 32×32 V-BLAST MIMO system with 64-QAM at an SNR of 30 dB on flat fading ($L = K = 1$). From Fig. 17, it is seen that a comparison of the value of M_1 with a suitable threshold can give an indication of the reliability of the RTS output. For example, the output is more likely to be erroneous if $M_1 > 12$ in Fig. 17.

Based on the above observation, we modify the RTS-BP algorithm as follows. If $M_1 > \theta$, only then BP algorithm is used; otherwise, the RTS output is taken as the final output. The threshold θ has to be carefully chosen to achieve good performance. It is seen that $\theta = 0$ corresponds to the case of unconditional RTS-BP, and $\theta = \infty$ corresponds to the case of RTS without BP. For $\theta = \infty$, there is no additional complexity due to BP, but there is no performance gain compared to RTS. For $\theta = 0$, performance gain is possible compared to RTS, but BP complexity will be there for all realizations. So there exits a performance-complexity trade off as a function of θ. We illustrate this trade-off in Fig. 18 for a 32×32 V-BLAST MIMO system with 64-QAM in flat fading. For this purpose, we define 'SNR gain' in dB for a given threshold θ as the improvement in SNR achieved by RTS with selective BP using threshold θ to achieve an uncoded BER of 10^{-3} compared to RTS without BP. Likewise, we define 'complexity gain' for a given θ as $10 \log_{10}(\beta)$, where β is the ratio of the average number of computations required to achieve 10^{-3} uncoded BER in unconditional RTS-BP and that in RTS with selective BP using threshold θ. In Fig. 18, we plot these two gains on the y-axis as a function of the threshold θ. From this figure, we can observe that for θ values less than 4, there is not much

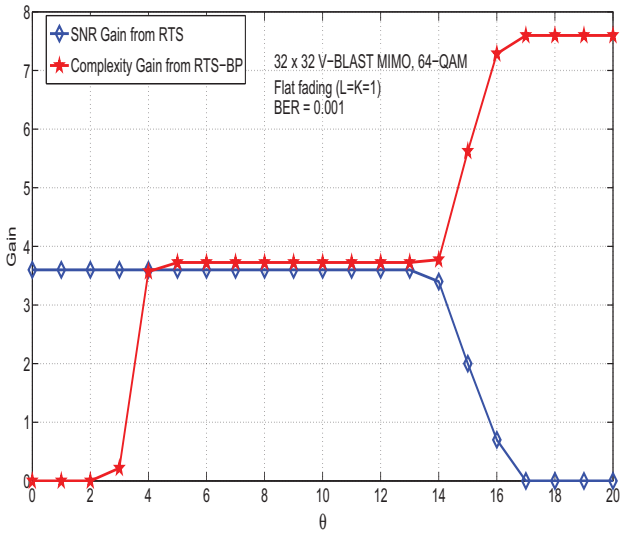

Fig. 18. SNR gain versus complexity gain trade-off in selectively using BP as a function of θ in a 32×32 V-BLAST MIMO system with 64-QAM at a BER of 0.001 on flat fading ($L = K = 1$).

complexity gain since such small threshold values invoke BP more often (i.e., the system behaves more like unconditional RTS-BP). Similarly, for θ values greater than 14, the system behaves more like RTS without BP; i.e., the complexity gain is maximum but there is no SNR gain. Interestingly, for θ values in the range 4 to 14, maximum SNR gain is retained while achieving significant complexity gain as well.

V. Conclusions

In this paper, we demonstrated that belief propagation on graphical models including Markov random fields and factor graphs can be efficiently used to achieve near-optimal detection in large-dimension MIMO-ISI channels at quadratic and linear complexities in Kn_t. It was shown through simulations that damping of messages in the MRF BP algorithm can significantly improve the BER performance and convergence behavior. The Gaussian approximation of interference we adopted in the factor graph approach is novel, which offered the attractive linear complexity in number of dimensions while achieving near-optimal performance in large dimensions. In higher-order QAM, iterations between a tabu search algorithm and the proposed FG-GAI BP algorithm was shown to improve the bit error performance of the basic tabu search algorithm. Although we have demonstrated the proposed algorithms in uncoded systems, they can be extended to coded systems as well, using either turbo equalization or joint processing of the entire coded symbol frame based on low-complexity graphical models. Finally, a theoretical analysis of the convergence behavior and the bit error performance of the proposed BP algorithms is challenging, and remains to be studied.

References

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.,* vol. 6, pp. 311-335, March 1998.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommun.,* vol. 10, no. 6, pp. 585-595, November 1999.

[3] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.

[4] J. G. Proakis, *Digital Communications*, 4th Ed., Mc-Graw Hill, 2001.

[5] X. Shen, M. Guizani, R. C. Qiu, and T. Le-Ngoc, *Ultra-wideband Wireless Communications and Networks,* John Wiley & Sons, 2006.

[6] A. F. Molisch, J. R. Foerster, M. Pendergrass, "Channel models for ultrawideband personal area networks," *IEEE Wireless Commun.*, vol. 10, no. 6, pp. 14-21, December 2003.

[7] A. F. Molisch, "Ultrawideband propagation channels - Theory, measurement, and modeling," *IEEE Trans. on Veh. Tech.,* vol. 54, no. 5, pp. 1528-1545, September 2005.

[8] J. Karedal, S. Wyne, P. Almers, F. Tufvesson, and A. F. Molisch, "Statistical analysis of the UWB channel in an industrial environment," *Proc. IEEE VTC'2004-Fall,*, pp. 81-85, September 2004.

[9] R. Saadane and A. M. Hayar, "DRB1.3 third report on UWB channel models," http://www.eurecom.fr/util/publidownload.fr.htm?id=2112, Newcom, November 2006.

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

[11] J. S. Yedidia, W. T. Freeman, Y. Weiss, "Understanding belief propagation and its generalizations," *MERL Tech. Rep. TR-2001-22*, January 2002.

[12] R. J. McEliece and D. J. C. MacKay, and J-F. Cheng, "Turbo decoding as an instance of Pearl's belief propagation algorithm," *IEEE Jl. Sel. Areas in Commun.,* vol. 16, no.2, pp. 140-152, February 1998.

[13] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. on Inform. Theory*, vol. 45, no. 2, pp. 399-431, March 1999.

[14] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *Journal of Physics A: Mathematical and General*, pp. 11111-11121, October 2003.

[15] A. Montanari, B. Prabhakar, and D. Tse, "Belief propagation based multiuser detection," Online arXiv:cs/0510044v2 [cs.IT] 22 May 2006.

[16] D. Guo and C-C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE JSAC Spl. Iss. on Multiuser Detection, for Adv. Commun. Systems and Networks*, vol. 26, no. 3, pp. 421-431, April 2008.

[17] J. Soler-Garrido, R. J. Piechocki, K. Maharatna, and D. McNamara, "Analog MIMO detection on the basis of belief propagation," *Proc. IEEE Mid-West Symp. on Circuits and Systems, 2006*.

[18] X. Yang, Y. Xiong, and F. Wang, "An adaptive MIMO system based on unified belief propagation detection," *Proc. IEEE ICC'2007*, Glasgow, June 2007.

[19] M. Suneel, P. Som, A. Chockalingam, and B. S. Rajan, "Belief propagation based decoding of large non-orthogonal STBCs," *Proc. IEEE ISIT'2009*, Seoul, July 2009.

[20] P. Som, T. Datta, A. Chockalingam, and B. S. Rajan, "Improved large-MIMO detection based on damped belief propagation," *Proc. IEEE Inform. Theory Workshop (ITW'2010)*, Cairo, January 2010.

[21] C. Douillard, M. Jezequel, and C. Berrou, "Iterative correction of intersymbol interference: Turbo equalization," *European Trans. on Telecommunications*, vol. 6, pp. 507-511, September-October 1995.

[22] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. on Commun.,* vol. 50, no. 5, pp. 754-767, May 2002.

[23] R. Koetter, A. C. Singer, and M. Tuchler, "Turbo equalization," *IEEE Sig. Process. Mag.,* pp. 67-80, January 2004.

[24] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 498-519, February 2001

[25] M. Tutchler, R. Koetter, and A. C. Singer, "Graphical models for coded data transmission over inter-symbol interference channels," *European Trans. on Telecommunications*, vol. 5, no. 4, July/August 2004.

[26] O. Shental, A. J. Weiss, N. Shental, and Y. Weiss, "Generalized belief propagation receiver for near-optimal detection of two-dimensional channels with memory," *Proc. IEEE Inform. Theory Workshop*, pp. 225-229, October 2004.

[27] G. Colavolpe and G. Germi, "On the application of factor graphs and the sum-product algorithm to ISI channels," *IEEE Trans. on Commun.,* vol. 53, no. 5, pp. 818-825, May 2005.

[28] R. J. Drost and A. C. Singer, "Factor graph algorithms for equalization," *IEEE Trans. on Sig. Process.*, vol. 55, no. 5, pp. 2052-2065, May 2007.

[29] M. N. Kaynak, T. M. Duman, and E. M. Kurtas, "Belief propagation over MIMO frequency selective fading channels," *Proc. Joint Intl. Conf. on Autonomic and Autonomous Systems and Intl. Conf. on Networking and Services*, Papeete, Tahiti, October 2005.

[30] T. Wo and P. A. Hoeher, "A simple iterative Gaussian detector for severely delay-spread MIMO channels," *Proc. IEEE ICC'2007*, Glasgow, June 2007.

[31] J. P. Neirotti and D. Saad, "Improved message passing for inference in densely connected systems," *Europhys. Lett.*, 2005. arXiv:cs/0503070v1 [cs.IT] 24 Mar 2005.

[32] Y. Jia, C. Andrieu, R. J. Piechocki, and M. Sandell, "Gaussian approximation based mixture reduction for near optimum detection in MIMO systems," *IEEE Commun. Letters*, vol. 9, no. 11, pp. 997-999, November 2005.

[33] J. C. Fricke, M. Sandell, J. Mietzner, and P. A. Hoeher, "Impact of the Gaussian approximation on the performance of the probabilistic data association MIMO decoder," *EURASIP Jl. Wireless Commun. & Networking*, vol. 5, pp. 796-800, 2005.

[34] J. Hu and T. M. Duman, "Graph-based detector for BLAST architecture," *Proc. IEEE ICC'2007*, Glasgow, June 2007.

[35] N. Srinidhi, S. K. Mohammed, and A. Chockalingam, "A reactive tabu search based equalizer for severely delay-spread UWB MIMO-ISI channels," *Proc. IEEE GLOBECOM'2009*, Honolulu, December 2009.

[36] Y. Jiang, R. Koetter, and A. C. singer, "On the separability of demodulation and decoding for communications over multiple-antenna block-fading channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2709-2713, October 2003.

[37] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.,* vol. 33, no. 2, pp. 100-109, February 1995.

[38] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.,* pp. 58-66, April 2002.

[39] B. Devillers, J. Louveaux, and L. Vandendorpe, "About the diversity in cyclic prefixed single-carrier systems," *Physical Communications*, pp. 266-276, 2008.

[40] P. Som and A. Chockalingam, "Damped belief propagation based near-optimal equalization of severely delay-spread UWB MIMO-ISI channels," *Proc. IEEE ICC'2010*, Cape Town, May 2010.

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

[42] D. Griffeath, *Introduction to Markov Random Fields*, Springer, 1976.

[43] J. M. Mooij, *Understanding and Improving Belief Propagation*, Ph.D Thesis, Radboud University Nijmegen, May 2008.

[44] J. M. Mooij and H. J. Kappen, "Sufficient conditions for convergence of the sum-product algorithm," *IEEE Trans. on Inform. Theory,* vol. 53, no. 12, pp. 4422-4437, December 2007.

[45] T. Heskes, K. Albers, and B. Kappen, "Approximate inference and constrained optimization," *Proc. Uncertainty in AI*, August 2003.

[46] A. L. Yuille, "A double-loop algorithm to minimize Bethe and Kikuchi free energies," *Neural Computation*, 2002.

[47] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," *Proc. Uncertainty in AI*, 1999.

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

[49] T. Heskes, "On the uniqueness of loopy belief propagation fixed points," *Neural Computation*, vol. 16, no. 11, pp. 2379-2413, November 2004.

[50] J. Goldberger and A. Leshem, "MIMO detection for high-order QAM based on a Gaussian tree approximation," *arXiv:1001.5364v1[cs.IT] 29 Jan 2010*.

[51] F. Glover, "Tabu Search - Part I," *ORSA Jl. of Computing,* vol. 1, no. 3, Summer 1989, pp. 190-206.

[52] F. Glover, "Tabu Search - Part II," *ORSA Jl. of Computing,* vol. 2, no. 1, Winter 1990, pp. 4-32.

[53] K. V. Vardhan, S. K. Mohammed, A. Chockalingam, and B. Sundar Rajan, "A low-complexity detector for large MIMO systems and multicarrier CDMA systems," *IEEE Jl. Sel. Areas in Commun. (JSAC): Spl. Iss. on Multiuser Detection for Advanced Commun. Systems and Networks*, vol. 26, no. 3, pp. 473-485, April 2008.

[54] S. K. Mohammed, A. Zaki, A. Chockalingam, and B. S. Rajan, "High-rate space-time coded large-MIMO systems: Low-complexity detection and channel estimation," *IEEE Jl. Sel. Topics in Signal Processing (JSTSP): Spl. Iss. on Managing Complexity in Multiuser MIMO Systems*, vol. 3, no. 6, pp. 958-974, December 2009.

[55] T. Datta, N. Srinidhi, A. Chockalingam, and B. Sundar Rajan, "A hybrid RTS-BP algorithm for improved detection of large-MIMO M-QAM signals," *Proc. National Conf. on Commun. (NCC'2011)*, Bangalore, January 2011.

Pritam Som received the B.E. degree in electronics and telecommunication engineering from Jadavpur University, Kolkata, India, in 2007, and the M.Sc.(Engg.) degree in electrical communication engineering from the Indian Institute of Science, Bangalore, in 2010. He is currently pursuing the Ph.D. degree in electrical communication engineering at the Indian Institute of Science, Bangalore, India. His research interests include low-complexity receiver design for MIMO and multiuser systems.

Tanumay Datta received the B.E. degree in electronics and telecommunication engineering from Jadavpur University, Kolkata, India, in 2008. He is currently pursuing the Ph.D. degree in electrical communication engineering at the Indian Institute of Science, Bangalore, India. His current research interests include low-complexity near-optimal detection algorithms for large-MIMO systems and wireless relay networks.

N. Srinidhi obtained his B.E. degree in electronics and communication engineering from Bapuji Institute of Engineering and Technology affiliated to Visvesvaraya Technological University, Belgaum, Karnataka, India. He is currently pursuing his Ph.D. degree in electrical communication engineering at the Indian Institute of Science, Bangalore, India. His research interests include low complexity detection in MIMO systems, MIMO and multiuser precoding, space-time coding, detection and estimation, and application of machine learning algorithms to communication problems.

A. Chockalingam (S'92-M'93-SM'98) was born in Rajapalayam, Tamil Nadu, India. He received the B.E. (Honors) degree in electronics and communication engineering from the P. S. G. College of Technology, Coimbatore, India, in 1984, the M.Tech. degree in electronics and electrical communications engineering (with specialization in satellite communications) from the Indian Institute of Technology, Kharagpur, India, in 1985, and the Ph.D. degree in electrical communication engineering (ECE) from the Indian Institute of Science (IISc), Bangalore, India, in 1993.

During 1986 to 1993, he worked with the Transmission R&D division of the Indian Telephone Industries Limited, Bangalore. From December 1993 to May 1996, he was a Postdoctoral Fellow and an Assistant Project Scientist at the Department of Electrical and Computer Engineering, University of California, San Diego. From May 1996 to December 1998, he served Qualcomm, Inc., San Diego, CA, as a Staff Engineer/Manager in the systems engineering group. In December 1998, he joined the faculty of the Department of ECE, IISc, Bangalore, India, where he is a Professor, working in the area of wireless communications and networking.

Dr. Chockalingam is a recipient of the Swarnajayanti Fellowship from the Department of Science and Technology, Government of India. He served as an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from May 2003 to April 2007. He currently serves as an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He served as a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (*Special Issue on Multiuser Detection for Advanced Communication Systems and Networks*). Currently, he serves as a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (*Special Issue on Soft Detection on Wireless Transmission*). He is a Fellow of the Institution of Electronics and Telecommunication Engineers and the Indian National Academy of Engineering.

B. Sundar Rajan received the B.Sc. degree in Mathematics from Madras University, India, the B.Tech degree in electronics from Madras Institute of Technology, and the M.Tech and Ph.D. degrees in electrical engineering from Indian Institute of Technology, Kanpur. He was a faculty member with the Department of Electrical Engineering at Indian Institute of Technology, Delhi, from 1990 to 1997. Since 1998, he has been with the Department of Electrical Communication Engineering at Indian Institute of Science, Bangalore, where, currently, he is the INAE Chair Professor. His primary research interests include space-time coding for MIMO channels, distributed space-time coding and cooperative communication, coding for multi-user and relay channels, and network coding.

Dr. Rajan is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Editor of IEEE WIRELESS COMMUNICATIONS LETTERS and an Editorial Board Member of INTERNATIONAL JOURNAL of INFORMATION and CODING THEORY. He was an Associate Editor of the IEEE TRANSACTIONS on INFORMATION THEORY during 2008-2011. He served as Technical Program Co-Chair of the IEEE Information Theory Workshop (ITW'02), held in Bangalore, in 2002.

Dr. Rajan is a recipient of IEEE Wireless Communications & Networking Conference (WCNC'2011) Best Academic Paper Award, recipient of Prof. Rustum Choksi award by Indian Institute of Science, for excellence in research in engineering in 2009, recipient of Khosla National Award in 2010, from Indian Institute of Technology, Roorkee and recipient of the IETE Pune Center's S.V.C. Aiya Award for Telecom Education in 2004. He is a Fellow of Indian National Academy of Engineering (INAE), Fellow of the National Academy of Sciences, India (NASI) and Fellow of Institution of Electronics and Telecommunication Engineers (IETE) India. He is a Senior Member of IEEE and a Member of American Mathematical Society.