ISIT 2009, Seoul, Korea, June 28 - July 3, 2009

Low-Complexity Near-MAP Decoding of Large Non-Orthogonal STBCs Using PDA

Saif K. Mohammed, A. Chockalingam, and B. Sundar Rajan
Department of ECE, Indian Institute of Science, Bangalore 560012, INDIA

***Abstract*—Non-orthogonal space-time block codes (STBC) from cyclic division algebras (CDA) are attractive because they can simultaneously achieve both high spectral efficiencies (same spectral efficiency as in V-BLAST for a given number of transmit antennas) *as well as* full transmit diversity. Decoding of non-orthogonal STBCs with hundreds of dimensions has been a challenge. In this paper, we present a probabilistic data association (PDA) based algorithm for decoding non-orthogonal STBCs with large dimensions. Our simulation results show that the proposed PDA-based algorithm achieves near SISO AWGN uncoded BER as well as near-capacity coded BER (within 5 dB of the theoretical capacity) for large non-orthogonal STBCs from CDA. We study the effect of spatial correlation on the BER, and show that the performance loss due to spatial correlation can be alleviated by providing more receive spatial dimensions. We report good BER performance when a training-based iterative decoding/channel estimation is used (instead of assuming perfect channel knowledge) in channels with large coherence times. A comparison of the performances of the PDA algorithm and the likelihood ascent search (LAS) algorithm (reported in our recent work) is also presented.**

I. Introduction

Multiple-input multiple-output (MIMO) systems that employ non-orthogonal space-time block codes (STBC) from cyclic division algebras (CDA) for arbitrary number of transmit antennas, N_t, are quite attractive because they can simultaneously provide both *full-rate* (i.e., N_t complex symbols per channel use, which is same as in V-BLAST) as well as *full transmit diversity* [1]. The 2×2 Golden code is a well known non-orthogonal STBC from CDA for 2 transmit antennas [2]. High spectral efficiencies of the order of tens of bps/Hz can be achieved using large non-orthogonal STBCs. For example, a 16×16 STBC from CDA has 256 complex symbols in it with 512 real dimensions; with 16-QAM and rate-3/4 turbo code, this system offers a high spectral efficiency of 48 bps/Hz. Decoding of non-orthogonal STBCs with such large dimensions, however, has been a challenge. Sphere decoder and its low-complexity variants are prohibitively complex for decoding such STBCs with hundreds of dimensions.

In this paper, we present a probabilistic data association (PDA) based algorithm for decoding large non-orthogonal STBCs from CDA. Key attractive features of this algorithm are its low-complexity and near-MAP performance in systems with large dimensions (e.g., hundreds of dimensions). While creating hundreds of dimensions in space alone (e.g., V-BLAST) requires hundreds of antennas, use of non-orthogonal STBCs from CDA can create hundreds of dimensions with just tens of antennas (space) and tens of channel uses (time). Given that 802.11 smart WiFi products with 12 transmit antennas at 2.5 GHz are now commercially available [4][1] (which establishes that issues related to placement of many antennas and RF/IF chains can be solved in large aperture communication terminals like set-top boxes/laptops), large non-orthogonal STBCs (e.g., 16×16 STBC from CDA) in combination with large dimension near-MAP decoding using PDA can enable communications at increased spectral efficiencies of the order of tens of bps/Hz (note that current standards achieve only < 10 bps/Hz using only up to 4 transmit antennas).

PDA, originally developed for target tracking, is widely used in digital communications [5]-[10]. Particularly, PDA algorithm is a reduced complexity alternative to the a posteriori probability (APP) decoder/detector/equalizer. Near-optimal performance has been demonstrated for PDA-based multiuser detection in CDMA systems [5]-[7]. PDA has been used in the detection of V-BLAST signals with small number of dimensions [8]-[10]. To our knowledge, PDA has not been reported for *decoding non-orthogonal STBCs with hundreds of dimensions* so far. Our new contributions in this paper are:

- We adapt the PDA algorithm for decoding non-orthogonal STBCs with large dimensions. With i.i.d fading and perfect channel channel state information at the receiver (CSIR), the algorithm achieves near-SISO AWGN uncoded BER and near-capacity coded BER (within 5 dB of the theoretical capacity) for 12×12 STBC from CDA, 4-QAM, rate-3/4 turbo code, and 18 bps/Hz.
- Relaxing the perfect CSIR assumption, we report results with a training based iterative PDA decoding/channel estimation scheme. The iterative scheme is shown to be effective with large coherence times.
- Relaxing the i.i.d fading assumption by adopting a spatially correlated MIMO channel model (proposed by Gesbert et al in [11]), we show that the performance loss due to spatial correlation is alleviated by using more receive spatial dimensions for a fixed receiver aperture.
- Finally, the performance of the PDA algorithm is compared with that of the likelihood ascent search (LAS) algorithm we recently presented in [12]-[14]. The PDA algorithm is shown to perform better than the LAS algorithm at low SNRs for higher-order QAM (e.g., 16-QAM), and in the presence of spatial correlation.

II. System Model

Consider a STBC MIMO system with multiple transmit and receive antennas. An (n, p, k) STBC is represented by a matrix $\mathbf{X}_c \in \mathbb{C}^{n \times p}$, where n and p denote the number of transmit antennas and number of time slots, respectively, and k denotes the number of complex data symbols sent in one STBC matrix. The (i, j)th entry in $\mathbf{X}_c$ represents the complex number transmitted from the ith transmit antenna in the jth time slot. The rate of an STBC is $\frac{k}{p}$. Let N_r and $N_t = n$ denote

[1] 12 antennas in these products are now used only for beamforming. Single-beam multi-antenna approaches can offer range increase and interference avoidance, but not spectral efficiency increase.

the number of receive and transmit antennas, respectively. Let $\mathbf{H}_c \in \mathbb{C}^{N_r \times N_t}$ denote the channel gain matrix, where the (i,j)th entry in $\mathbf{H}_c$ is the complex channel gain from the jth transmit antenna to the ith receive antenna. We assume that the channel gains remain constant over one STBC matrix duration. Assuming rich scattering, we model the entries of $\mathbf{H}_c$ as $\mathcal{CN}(0,1)$. The received space-time signal matrix, $\mathbf{Y}_c \in \mathbb{C}^{N_r \times p}$, can be written as

$$\mathbf{Y}_c = \mathbf{H}_c \mathbf{X}_c + \mathbf{N}_c, \tag{1}$$

where $\mathbf{N}_c \in \mathbb{C}^{N_r \times p}$ is the noise matrix at the receiver and its entries are modeled as i.i.d $\mathcal{CN}\big(0, \sigma^2 = \frac{N_t E_s}{\gamma}\big)$, where E_s is the average energy of the transmitted symbols, and γ is the average received SNR per receive antenna [3], and the (i,j)th entry in $\mathbf{Y}_c$ is the received signal at the ith receive antenna in the jth time-slot. Consider linear dispersion STBCs, where $\mathbf{X}_c$ can be written in the form [3]

$$\mathbf{X}_c \;=\; \sum_{i=1}^{k} x_c^{(i)} \mathbf{A}_c^{(i)}, \tag{2}$$

where $x_c^{(i)}$ is the ith complex data symbol, and $\mathbf{A}_c^{(i)} \in \mathbb{C}^{N_t \times p}$ is its corresponding weight matrix. The received signal model in (1) can be written in an equivalent V-BLAST form as

$$\mathbf{y}_c \;=\; \sum_{i=1}^{k} x_c^{(i)} \, (\widehat{\mathbf{H}}_c \, \mathbf{a}_c^{(i)}) + \mathbf{n}_c \;=\; \widetilde{\mathbf{H}}_c \mathbf{x}_c + \mathbf{n}_c, \tag{3}$$

where $\mathbf{y}_c \in \mathbb{C}^{N_r p \times 1} = vec\,(\mathbf{Y}_c)$, $\widehat{\mathbf{H}}_c \in \mathbb{C}^{N_r p \times N_t p} = (\mathbf{I}_p \otimes \mathbf{H}_c)$, $\mathbf{I}_p$ is $p \times p$ identity matrix, $\mathbf{a}_c^{(i)} \in \mathbb{C}^{N_t p \times 1} = vec\,(\mathbf{A}_c^{(i)})$, $\mathbf{n}_c \in \mathbb{C}^{N_r p \times 1} = vec\,(\mathbf{N}_c)$, $\mathbf{x}_c \in \mathbb{C}^{k \times 1}$ whose ith entry is the data symbol $x_c^{(i)}$, and $\widetilde{\mathbf{H}}_c \in \mathbb{C}^{N_r p \times k}$ whose ith column is $\widehat{\mathbf{H}}_c \, \mathbf{a}_c^{(i)}$, $i = 1, 2, \cdots, k$. Each element of $\mathbf{x}_c$ is an M-PAM/M-QAM symbol. M-PAM symbols take discrete values from $\mathbb{A} \triangleq \{a_q, q = 1, \cdots, M\}$, where $a_q = (2q-1-M)$, and M-QAM is nothing but two PAMs in quadrature. Let $\mathbf{y}_c$, $\widetilde{\mathbf{H}}_c$, $\mathbf{x}_c$, $\mathbf{n}_c$ be decomposed into real and imaginary parts as:

$$\mathbf{y}_c = \mathbf{y}_I + j\mathbf{y}_Q, \quad \mathbf{x}_c = \mathbf{x}_I + j\mathbf{x}_Q,$$
$$\mathbf{n}_c = \mathbf{n}_I + j\mathbf{n}_Q, \quad \widetilde{\mathbf{H}}_c = \mathbf{H}_I + j\mathbf{H}_Q. \tag{4}$$

Further, we define $\mathbf{H}_r \in \mathbb{R}^{2N_r p \times 2k}$, $\mathbf{y}_r \in \mathbb{R}^{2N_r p \times 1}$, $\mathbf{x}_r \in \mathbb{A}^{2k \times 1}$, and $\mathbf{n}_r \in \mathbb{R}^{2N_r p \times 1}$ as

$$\mathbf{H}_r = \begin{pmatrix} \mathbf{H}_I & -\mathbf{H}_Q \\ \mathbf{H}_Q & \mathbf{H}_I \end{pmatrix}, \quad \mathbf{y}_r = [\mathbf{y}_I^T \;\; \mathbf{y}_Q^T]^T, \tag{5}$$

$$\mathbf{x}_r = [\mathbf{x}_I^T \;\; \mathbf{x}_Q^T]^T, \quad \mathbf{n}_r = [\mathbf{n}_I^T \;\; \mathbf{n}_Q^T]^T. \tag{6}$$

Now, (3) can be written as

$$\mathbf{y}_r \;=\; \mathbf{H}_r \mathbf{x}_r + \mathbf{n}_r. \tag{7}$$

Henceforth, we work with the real-valued system in (7). For notational simplicity, we drop subscripts r in (7) and write

$$\mathbf{y} \;=\; \mathbf{H}' \mathbf{x} + \mathbf{n}, \tag{8}$$

where $\mathbf{H}' = \mathbf{H}_r \in \mathbb{R}^{2N_r p \times 2k}$, $\mathbf{y} = \mathbf{y}_r \in \mathbb{R}^{2N_r p \times 1}$, $\mathbf{x} = \mathbf{x}_r \in \mathbb{A}^{2k \times 1}$, and $\mathbf{n} = \mathbf{n}_r \in \mathbb{R}^{2N_r p \times 1}$. We assume that the channel gains are known at the receiver but not at the transmitter.

A. High-rate Non-orthogonal STBCs from CDA

We focus on the detection of square (i.e., $n = p = N_t$), full-rate (i.e., $k = pn = N_t^2$), circulant (where the weight matrices $\mathbf{A}_c^{(i)}$'s are permutation type), non-orthogonal STBCs from CDA [1], whose construction for arbitrary number of transmit antennas n is given by the matrix in Eqn.(9.a) given at the bottom of this column. In (9.a), $\omega_n = e^{\frac{\mathbf{j}2\pi}{n}}$, $\mathbf{j} = \sqrt{-1}$, and $d_{u,v}$, $0 \leq u, v \leq n-1$ are the n^2 data symbols from a QAM alphabet. When $\delta = e^{\sqrt{5}\mathbf{j}}$ and $t = e^{\mathbf{j}}$, the STBC in (9.a) achieves full transmit diversity (under ML decoding) as well as information-losslessness [1]. When $\delta = t = 1$, the code ceases to be of full-diversity (FD), but continues to be information-lossless (ILL). High spectral efficiencies with large n can be achieved using this code construction. However, since these STBCs are non-orthogonal, MAP/ML detection gets increasingly impractical for large n. Consequently, a key challenge in realizing the benefits of these large STBCs in practice is that of achieving near-MAP/ML performance for large n at low decoding complexities. The PDA-based decoding algorithm we present in the following section essentially addresses this challenge.

III. PROPOSED PDA-BASED DECODING

In this section, we present the proposed PDA-based decoding algorithm for square QAM. The applicability of the algorithm to rectangular QAM is straightforward. In the real-valued system model in (8), each entry of $\mathbf{x}$ belongs to a $\sqrt{M}$-PAM constellation, where M is the size of the original square QAM constellation. Let $b_i^{(0)}, b_i^{(1)}, \cdots, b_i^{(q-1)}$ denote the $q = \log_2(\sqrt{M})$ constituent bits of the ith entry x_i of $\mathbf{x}$. We can write the value of each entry of $\mathbf{x}$ as a linear combination of its constituent bits as

$$x_i \;=\; \sum_{j=0}^{q-1} 2^j \, b_i^{(j)}, \qquad i = 0, 1, \cdots, 2k-1. \tag{9}$$

Let $\mathbf{b} \in \{+1, -1\}^{2qk \times 1}$, defined as

$$\mathbf{b} \;\triangleq\; \Big[b_0^{(0)} \cdots b_0^{(q-1)} b_1^{(0)} \cdots b_1^{(q-1)} \cdots b_{2k-1}^{(0)} \cdots b_{2k-1}^{(q-1)} \Big]^T, \tag{10}$$

denote the transmitted bit vector. Defining $\mathbf{c} \triangleq [2^0 \; 2^1 \cdots 2^{q-1}]$, we can write $\mathbf{x}$ as

$$\mathbf{x} \;=\; (\mathbf{I}_{2k} \otimes \mathbf{c})\mathbf{b}. \tag{11}$$

Using (11), we can rewrite (8) as

$$\mathbf{y} \;=\; \underbrace{\mathbf{H}'(\mathbf{I}_{2k} \otimes \mathbf{c})}_{\triangleq\, \mathbf{H}} \, \mathbf{b} + \mathbf{n}, \tag{12}$$

where $\mathbf{H} \in \mathbb{R}^{2N_r p \times 2qk}$ is the effective channel matrix. The MAP estimate of bit $b_i^{(j)}$ is given by

$$\hat{b}_i^{(j)} \;=\; \mathop{\arg\max}_{a \,\in\, \{\pm 1\}} \; p\big(b_i^{(j)} = a \,|\, \mathbf{y}, \mathbf{H}\big), \tag{13}$$

whose computational complexity is exponential in k. Our goal is to obtain $\widehat{\mathbf{b}}$, an estimate of $\mathbf{b}$, at low complexities. For this, we iteratively update the statistics of each bit of $\mathbf{b}$, as described in the following subsection, for a certain number of iterations, and hard decisions are made on the final statistics to get $\widehat{\mathbf{b}}$.

A. Iterative Procedure

The algorithm is iterative in nature, where $2qk$ statistic updates, one for each of the constituent bits, are performed in each iteration. We start the algorithm by initializing the

$$\begin{bmatrix} \sum_{i=0}^{n-1} d_{0,i}\, t^i & \delta \sum_{i=0}^{n-1} d_{n-1,i}\, \omega_n^i\, t^i & \cdots & \delta \sum_{i=0}^{n-1} d_{1,i}\, \omega_n^{(n-1)i}\, t^i \\ \sum_{i=0}^{n-1} d_{1,i}\, t^i & \sum_{i=0}^{n-1} d_{0,i}\, \omega_n^i\, t^i & \cdots & \delta \sum_{i=0}^{n-1} d_{2,i}\, \omega_n^{(n-1)i}\, t^i \\ \sum_{i=0}^{n-1} d_{2,i}\, t^i & \sum_{i=0}^{n-1} d_{1,i}\, \omega_n^i\, t^i & \cdots & \delta \sum_{i=0}^{n-1} d_{3,i}\, \omega_n^{(n-1)i}\, t^i \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=0}^{n-1} d_{n-2,i}\, t^i & \sum_{i=0}^{n-1} d_{n-3,i}\, \omega_n^i\, t^i & \cdots & \delta \sum_{i=0}^{n-1} d_{n-1,i}\, \omega_n^{(n-1)i}\, t^i \\ \sum_{i=0}^{n-1} d_{n-1,i}\, t^i & \sum_{i=0}^{n-1} d_{n-2,i}\, \omega_n^i\, t^i & \cdots & \sum_{i=0}^{n-1} d_{0,i}\, \omega_n^{(n-1)i}\, t^i \end{bmatrix} \tag{9.a}$$

a priori probabilities as $P(b_i^{(j)}=+1)=P(b_i^{(j)}=-1)=0.5$, $\forall\, i=0,\cdots,2k-1$ and $j=0,\cdots,q-1$. In an iteration, the statistics of the bits are updated sequentially, i.e., the ordered sequence of updates in an iteration is $\{b_0^{(0)},\cdots,b_0^{(q-1)},\cdots\cdots,b_{2k-1}^{(0)},\cdots b_{2k-1}^{(q-1)}\}$. The steps involved in each iteration of the algorithm are derived as follows. The likelihood ratio of bit $b_i^{(j)}$ in an iteration, denoted by $\Lambda_i^{(j)}$, is

$$\Lambda_i^{(j)} \;\triangleq\; \underbrace{\frac{P(\mathbf{y}|b_i^{(j)}=+1)}{P(\mathbf{y}|b_i^{(j)}=-1)}}_{\triangleq\, \beta_i^{(j)}}\;\underbrace{\frac{P(b_i^{(j)}=+1)}{P(b_i^{(j)}=-1)}}_{\triangleq\, \alpha_i^{(j)}}. \qquad (14)$$

Denoting the tth column of $\mathbf{H}$ by $\mathbf{h}_t$, we can write (12) as

$$\mathbf{y} \;=\; \mathbf{h}_{qi+j}\, b_i^{(j)} + \underbrace{\sum_{l=0}^{2k-1}\sum_{\substack{m=0\\ m\neq q(i-l)+j}}^{q-1} \mathbf{h}_{ql+m}\, b_l^{(m)} + \mathbf{n}}_{\triangleq\, \widetilde{\mathbf{n}}}, \qquad (15)$$

where $\widetilde{\mathbf{n}} \in \mathbb{R}^{2N_rp\times 1}$ is the interference plus noise vector. To calculate $\beta_i^{(j)}$, we approximate the distribution of $\widetilde{\mathbf{n}}$ to be Gaussian, and hence $\mathbf{y}$ is Gaussian conditioned on $b_i^{(j)}$. Since there are $2qk-1$ terms in the double summation in (15), this Gaussian approximation gets increasingly accurate for large N_t (note that $k=N_t^2$). Since a Gaussian distribution is fully characterized by its mean and covariance, we evaluate the mean and covariance of $\mathbf{y}$ given $b_i^{(j)}=+1$ and $b_i^{(j)}=-1$. For notational simplicity, let us define $p_i^{j+}\triangleq P(b_i^{(j)}=+1)$ and $p_i^{j-}\triangleq P(b_i^{(j)}=-1)$. It is clear that $p_i^{j+}+p_i^{j-}=1$.

Let $\boldsymbol{\mu}_i^{j+}\triangleq \mathbb{E}(\mathbf{y}|b_i^{(j)}=+1)$ and $\boldsymbol{\mu}_i^{j-}\triangleq \mathbb{E}(\mathbf{y}|b_i^{(j)}=-1)$, where $\mathbb{E}(.)$ denotes the expectation operator. Now, from (15), we can write $\boldsymbol{\mu}_i^{j+}$ as

$$\boldsymbol{\mu}_i^{j+} \;=\; \mathbf{h}_{qi+j} + \sum_{l=0}^{2k-1}\sum_{\substack{m=0\\ m\neq q(i-l)+j}}^{q-1} \mathbf{h}_{ql+m}(2p_l^{m+}-1). \qquad (16)$$

Similarly, we can write $\boldsymbol{\mu}_i^{j-}$ as

$$\boldsymbol{\mu}_i^{j-} = -\mathbf{h}_{qi+j} + \sum_{l=0}^{2k-1}\sum_{\substack{m=0\\ m\neq q(i-l)+j}}^{q-1} \mathbf{h}_{ql+m}(2p_l^{m+}-1) = \boldsymbol{\mu}_i^{j+} - 2\mathbf{h}_{qi+j}. \quad (17)$$

Next, the $2N_rp\times 2N_rp$ covariance matrix $\mathbf{C}_i^j$ of $\mathbf{y}$ given b_i^j is

$$\mathbf{C}_i^j \;=\; \mathbb{E}\Big\{\Big[\mathbf{n} + \sum_{l=0}^{2k-1}\sum_{\substack{m=0\\ m\neq q(i-l)+j}}^{q-1} \mathbf{h}_{ql+m}(b_l^{(m)}-2p_l^{m+}+1)\Big]$$
$$\Big[\mathbf{n} + \sum_{l=0}^{2k-1}\sum_{\substack{m=0\\ m\neq q(i-l)+j}}^{q-1} \mathbf{h}_{ql+m}(b_l^{(m)}-2p_l^{m+}+1)\Big]^T\Big\}. \qquad (18)$$

Assuming independence among the constituent bits, we can simplify $\mathbf{C}_i^j$ in (18) as

$$\mathbf{C}_i^j = \sigma^2\mathbf{I}_{2N_rp} + \sum_{l=0}^{2k-1}\sum_{\substack{m=0\\ m\neq q(i-l)+j}}^{q-1} \mathbf{h}_{ql+m}\,\mathbf{h}_{ql+m}^T\, 4p_l^{m+}(1-p_l^{m+}). \quad (19)$$

Using the above mean and covariance expressions, we can write the distribution of $\mathbf{y}$ given $b_i^{(j)}=\pm 1$ as

$$P(\mathbf{y}|b_i^{(j)}=\pm 1) \;=\; \frac{e^{-(\mathbf{y}-\boldsymbol{\mu}_i^{j\pm})^T(\mathbf{C}_i^j)^{-1}(\mathbf{y}-\boldsymbol{\mu}_i^{j\pm})}}{(2\pi)^{N_rp}|\mathbf{C}_i^j|^{\frac{1}{2}}}. \qquad (20)$$

Using (20), β_i^j can be written as

$$\beta_i^j = e^{-\left((\mathbf{y}-\boldsymbol{\mu}_i^{j+})^T(\mathbf{C}_i^j)^{-1}(\mathbf{y}-\boldsymbol{\mu}_i^{j+})-(\mathbf{y}-\boldsymbol{\mu}_i^{j-})^T(\mathbf{C}_i^j)^{-1}(\mathbf{y}-\boldsymbol{\mu}_i^{j-})\right)}. \quad (21)$$

Using $\alpha_i^{(j)}$ and $\beta_i^{(j)}$, $\Lambda_i^{(j)}$ is computed using (14). Now, using the value of $\Lambda_i^{(j)}$, the statistics of $b_i^{(j)}$ is updated as follows. and Using $P(b_i^{(j)}=+1|\mathbf{y})+P(b_i^{(j)}=-1|\mathbf{y})=1$, we have

$$P(b_i^{(j)}=+1|\mathbf{y}) = \frac{\Lambda_i^{(j)}}{1+\Lambda_i^{(j)}},\quad P(b_i^{(j)}=-1|\mathbf{y}) = \frac{1}{1+\Lambda_i^{(j)}}. \quad (22)$$

This completes one iteration of the algorithm; i.e., each iteration involves the computation of $\alpha_i^{(j)}$ and equations (16), (17), (19), (21), (14), and (22) for all i,j. The updated values of $P(b_i^{(j)}=+1|\mathbf{y})$ and $P(b_i^{(j)}=-1|\mathbf{y})$ in (22) for all i,j are fed back as a priori probabilities to the next iteration[2]. The algorithm terminates after a certain number of such iterations. At the end of the last iteration, hard decision is made on the final statistics to obtain the bit estimate $\widehat{b}_i^{(j)}$ as $+1$ if $\Lambda_i^{(j)}\geq 1$, and -1 otherwise. In coded systems, $\Lambda_i^{(j)}$'s are fed as soft inputs to the decoder.

B. Complexity Reduction

The most computationally expensive operation in computing $\beta_i^{(j)}$ is the evaluation of the inverse of the covariance matrix, $\mathbf{C}_i^j$, of size $2N_rp\times 2N_rp$ which requires $O(N_r^3p^3)$ complexity, which can be reduced as follows. Define matrix $\mathbf{D}$ as

$$\mathbf{D} \;\triangleq\; \sigma^2\mathbf{I}_{2N_rp} + \sum_{l=0}^{2k-1}\sum_{m=0}^{q-1} \mathbf{h}_{ql+m}\mathbf{h}_{ql+m}^T 4p_l^{m+}(1-p_l^{m+}). \quad (23)$$

At the start of the algorithm, with p_i^{j+} and $p_i^{(j)}$ initialized to 0.5 for all i,j, $\mathbf{D}$ becomes $\sigma^2\mathbf{I}_{2N_rp}+\mathbf{H}\mathbf{H}^T$.

Computation of $\mathbf{D}^{-1}$*:* We note that when the statistics of $b_i^{(j)}$ is updated using (22), the $\mathbf{D}$ matrix in (23) also changes. A straightforward inversion of this updated $\mathbf{D}$ matrix would require $O(N_r^3p^3)$ complexity. However, we can obtain the $\mathbf{D}^{-1}$ from the previously available $\mathbf{D}^{-1}$ in $O(N_r^2p^2)$ complexity as follows. Since the statistics of only $b_i^{(j)}$ is updated, the new $\mathbf{D}$ matrix is just a rank one update of the old $\mathbf{D}$ matrix. Therefore, using the matrix inversion lemma, the new $\mathbf{D}^{-1}$ can be obtained from the old $\mathbf{D}^{-1}$ as

$$\mathbf{D}^{-1} \;\leftarrow\; \mathbf{D}^{-1} - \frac{\mathbf{D}^{-1}\mathbf{h}_{ni+j}\mathbf{h}_{ni+j}^T\mathbf{D}^{-1}}{\mathbf{h}_{ni+j}^T\mathbf{D}^{-1}\mathbf{h}_{ni+j}+\frac{1}{\eta}}, \qquad (24)$$

where
$$\eta \;=\; 4p_i^{j+}(1-p_i^{j+}) - 4p_{i,old}^{j+}(1-p_{i,old}^{j+}), \qquad (25)$$

where p_i^{j+} and $p_{i,old}^{j+}$ are the new (i.e., after the update in (22)) and old (before the update) values, respectively. It can be seen that both the numerator and denominator in the 2nd term on the RHS of (24) can be computed in $O(N_r^2p^2)$ complexity. Therefore, the computation of the new $\mathbf{D}^{-1}$ using the old $\mathbf{D}^{-1}$ can be done in $O(N_r^2p^2)$ complexity.

Computation of $(\mathbf{C}_i^j)^{-1}$*:* Using (23) and (19), we can write $\mathbf{C}_i^j$ in terms of $\mathbf{D}$ as

$$\mathbf{C}_i^j \;=\; \mathbf{D} - 4p_i^{j+}(1-p_i^{j+})\,\mathbf{h}_{qi+j}\,\mathbf{h}_{qi+j}^T. \qquad (26)$$

We can compute $(\mathbf{C}_i^j)^{-1}$ from $\mathbf{D}^{-1}$ at a reduced complexity using the matrix inversion lemma, which states that

[2]The computation of the statistics of a current bit in an iteration makes use of the newly computed statistics of its previous bits (as per the ordered sequence of statistic updates) in the same iteration and the statistics of its next bits available from the previous iteration.

$$(\mathbf{P}+\mathbf{QRS})^{-1} = \mathbf{P}^{-1} - \mathbf{P}^{-1}\mathbf{Q}(\mathbf{R}^{-1}+\mathbf{SP}^{-1}\mathbf{Q})^{-1}\mathbf{SP}^{-1}. \quad (27)$$

Substituting $\mathbf{P}_{2N_rp\times 2N_rp} = \mathbf{D}$, $\mathbf{Q}_{2N_rp\times 1} = \mathbf{h}_{qi+j}$, $\mathbf{R}_{1\times 1} = -4p_i^{j+}(1-p_i^{j+})$, and $\mathbf{S}_{1\times 2N_rp} = \mathbf{h}_{qi+j}^T$ in (27), we get

$$(\mathbf{C}_i^j)^{-1} \;=\; \mathbf{D}^{-1} - \frac{\mathbf{D}^{-1}\,\mathbf{h}_{qi+j}\,\mathbf{h}_{qi+j}^T\,\mathbf{D}^{-1}}{\mathbf{h}_{qi+j}^T\,\mathbf{D}^{-1}\,\mathbf{h}_{qi+j} - \frac{1}{4p_i^{j+}(1-p_i^{j+})}}, \quad (28)$$

which can be computed in $O(N_r^2p^2)$ complexity.

Computation of $\boldsymbol{\mu}_i^{j+}$ and $\boldsymbol{\mu}_i^{j-}$: Computation of $\beta_i^{(j)}$ involves the computation of $\boldsymbol{\mu}_i^{j+}$ and $\boldsymbol{\mu}_i^{j-}$ also. From (17), it is clear that $\boldsymbol{\mu}_i^{j-}$ can be computed from $\boldsymbol{\mu}_i^{j+}$ with a computational overhead of only $O(N_rp)$. From (16), it can be seen that computing $\boldsymbol{\mu}_i^{j+}$ would require $O(qN_rpk)$ complexity. However, this complexity can be reduced as follows. Define vector $\mathbf{u}$ as

$$\mathbf{u} \;\stackrel{\Delta}{=}\; \sum_{l=0}^{2k-1}\sum_{m=0}^{q-1} \mathbf{h}_{ql+m}\big(2p_l^{m+}-1\big). \quad (29)$$

Using (16) and (29), we can write

$$\boldsymbol{\mu}_i^{j+} \;=\; \mathbf{u} + 2\big(1-p_i^{j+}\big)\mathbf{h}_{qi+j}. \quad (30)$$

$\mathbf{u}$ can be computed iteratively at $O(N_rp)$ complexity as follows. When the statistics of $b_i^{(j)}$ is updated, we can obtain the new $\mathbf{u}$ from the old $\mathbf{u}$ as

$$\mathbf{u} \;\leftarrow\; \mathbf{u} + 2\big(p_i^{j+} - p_{i,old}^{j+}\big)\mathbf{h}_{ni+j}, \quad (31)$$

whose complexity is $O(N_rp)$. Hence, the computation of $\boldsymbol{\mu}_i^{j+}$ and $\boldsymbol{\mu}_i^{j-}$ needs $O(N_rp)$ complexity.

C. Overall Complexity

We need to compute $\mathbf{HH}^T$ at the start of the algorithm. This requires $O(qkN_r^2p^2)$ complexity. So the computation of the initial $\mathbf{D}^{-1}$ requires $O(qkN_r^2p^2) + O(N_r^3p^3)$. Based on the complexity reduction in Sec. III-B, the complexity in updating the statistics of one constituent bit is $O(N_r^2p^2)$. So, the complexity for the update of all the $2qk$ constituent bits in an iteration is $O(qkN_r^2p^2)$. Since the number of iterations is fixed, the overall complexity of the algorithm is $O(qkN_r^2p^2) + O(N_r^3p^3)$. For $N_t = N_r$, since there are k symbols per STBC and q bits per symbol, the overall complexity per bit is $O(p^2N_t^2)$.

IV. Results and Discussions

In this section, we present the simulated uncoded/coded BER performance of the PDA algorithm in decoding non-orthogonal STBCs from CDA[3]. Number of iterations in the PDA algorithm is set to $m = 10$ in all the simulations.

PDA versus LAS performance with 4-QAM: In Fig. 1, we plot the uncoded BER of the PDA algorithm as a function of average received SNR per receive antenna, γ, in decoding 4×4, 8×8, 16×16 ILL STBCs from CDA with $N_t = N_r$ and 4-QAM. Perfect CSIR and i.i.d fading are assumed. For the same settings, the performance of the LAS algorithm in [12]-[14] with MMSE initial vector are also plotted for comparison. From Fig. 1, it is seen that

- the BER performance of PDA algorithm improves and approaches SISO AWGN performance as $N_t = N_r$ is increased; e.g., performance close to within about 1 dB from SISO AWGN performance is achieved at 10^{-3} uncoded BER in decoding 16×16 STBC from CDA having 512 real dimensions, and this illustrates the ability of the PDA algorithm to achieve excellent performance at low complexities in large non-orthogonal STBC MIMO.
- with 4-QAM, PDA and LAS algorithms achieve almost the same performance.

[3]Our simulation results showed that the performance of FD-ILL ($\delta = e^{\sqrt{5}\mathbf{j}}, t = e^{-\mathbf{j}}$) and ILL ($\delta = t = 1$) STBCs with PDA decoding were almost the same. Here, we present the performance of ILL STBCs.

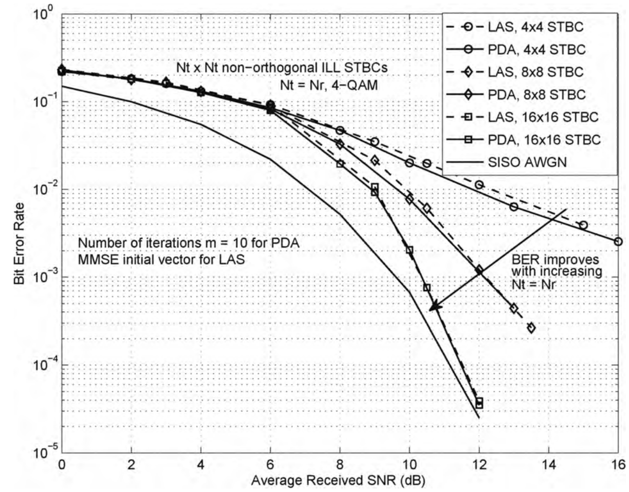

Fig. 1. Comparison of uncoded BER of PDA and LAS algorithms in decoding 4×4, 8×8, 16×16 ILL STBCs. $N_t = N_r$, 4-QAM. *BER improves for increasing STBC sizes. With 4-QAM, PDA and LAS algorithms achieve almost same performance.*

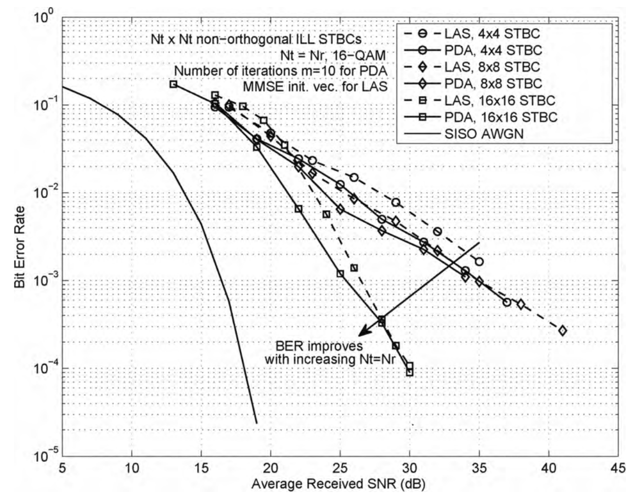

Fig. 2. Comparison of uncoded BER of PDA and LAS algorithms in decoding 4×4, 8×8, 16×16 ILL STBCs. $N_t = N_r$, 16-QAM. *With 16-QAM, PDA performs better than LAS at low SNRs.*

PDA versus LAS performance with 16-QAM: Fig. 2 presents an uncoded BER comparison between PDA and LAS algorithms in decoding ILL STBCs from CDA with $N_t = Nr$ and 16-QAM under perfect CSIR and i.i.d fading. It can be seen that the PDA algorithm performs better at low SNRs than the LAS algorithm. For example, with 8×8 and 16×16 STBCs, at low SNRs (e.g., < 25 dB for 16×16 STBC), PDA algorithm performs better by about 1 dB compared to LAS algorithm at 10^{-2} uncoded BER.

Turbo coded BER performance of PDA: Figure 3 shows the rate-3/4 turbo coded BER of the PDA algorithm under perfect CSIR and i.i.d fading for 12×12 ILL STBC with $N_t = N_r = 12$ and 4-QAM, which corresponds to a spectral efficiency of 18

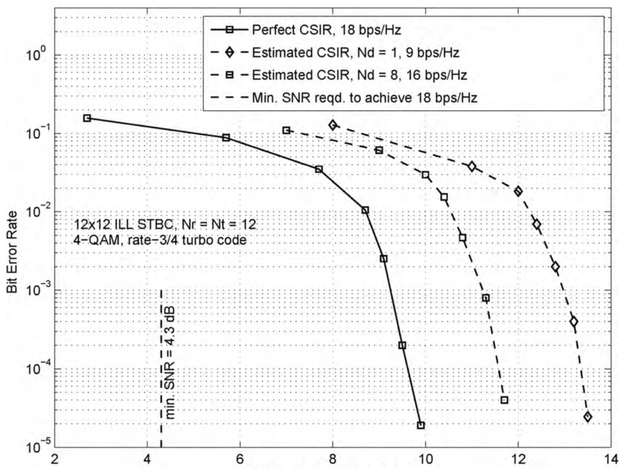

Fig. 3. Turbo coded BER of the PDA algorithm in decoding 12×12 ILL STBC with $N_t = N_r = 12$, 4-QAM, rate-3/4 turbo code, 18 bps/Hz and $m = 10$ for i) perfect CSIR, and ii) estimated CSIR using 2 iterations between PDA decoding/channel estimation. *With perfect CSIR, PDA performs close to within 5 dB from capacity. With estimated CSIR, performance approaches to that with perfect CSIR with increasing coherence times.*

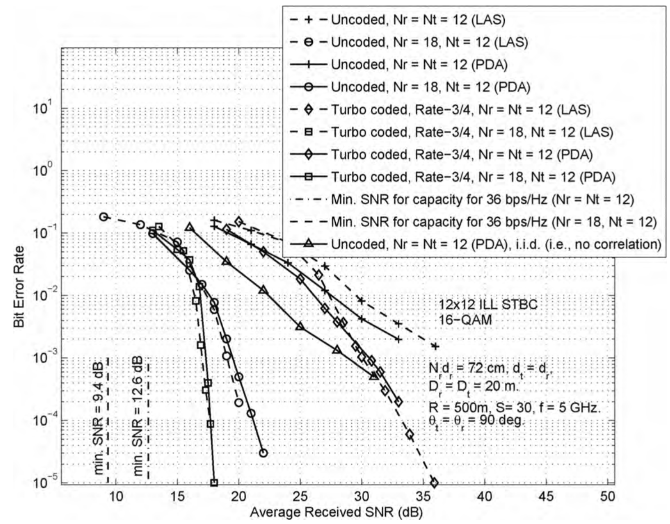

Fig. 4. Effect of spatial correlation on the performance of PDA in decoding 12×12 ILL STBC from CDA. $N_t = 12$, $N_r = 12, 18$, 16-QAM, rate-3/4 turbo code, 36 bps/Hz. Correlated channel parameters: $f_c = 5$ GHz, $R = 500$ m, $S = 30$, $D_t = D_r = 20$ m, $\theta_t = \theta_r = 90°$, $N_r d_r = 72$ cm, $d_t = d_r$. *Spatial correlation degrades performance; using $N_r > N_t$ alleviates the this performance loss.*

bps/Hz. The theoretical minimum SNR required to achieve 18 bps/Hz spectral efficiency on a $N_t{=}N_r{=}12$ MIMO channel with perfect CSIR and i.i.d fading is 4.3 dB (obtained through simulation of the ergodic capacity formula [3]). From Fig. 3, it is seen that the PDA algorithm is able to achieve vertical fall in coded BER within about 5 dB from theoretical minimum SNR, which is a good nearness to capacity performance.

Iterative Decoding/Channel Estimation: We relax the perfect CSIR assumption by considering a training based iterative PDA decoding/channel estimation scheme. Transmission is carried out in frames, where one $N_t \times N_t$ pilot matrix (for training purposes) followed by N_d data STBC matrices are sent in each frame. One frame length, T, (taken to be the channel coherence time) is $T = (N_d + 1)N_t$ channel uses. The proposed scheme works as follows: i) obtain an MMSE estimate of the channel matrix during the pilot phase, ii) use the estimated channel matrix to decode the data STBC matrices using PDA algorithm, and iii) iterate between channel estimation and PDA decoding for a certain number of times. For the 12×12 ILL STBC from CDA, in addition to perfect CSIR performance, Fig. 3 also shows the performance with CSIR estimated using the proposed iterative decoding/channel estimation scheme for $N_d = 1$ and $N_d = 8$. Two iterations between decoding and channel estimation are used. With $N_d = 8$ (which corresponds to large coherence times, i.e., slow fading) the BER and bps/Hz with estimated CSIR get closer to those with perfect CSIR.

Effect of Spatial MIMO Correlation: In Figs. 1 to 3, we assumed i.i.d fading. But spatial correlation at transmit/receive antennas and the structure of scattering and propagation environment can affect the rank structure of the MIMO channel resulting in degraded performance. We relaxed the i.i.d. fading assumption by considering the correlated MIMO channel model in [11], which takes into account carrier frequency (f_c), spacing between antenna elements (d_t, d_r), distance between transmit and receive antennas (R), and scattering environment. In Fig. 4, we plot the BER of the PDA algorithm in decoding 12×12 ILL STBC from CDA with perfect CSIR in i) i.i.d. fading, and ii) correlated MIMO fading model in [11]. It is seen that, compared to i.i.d fading, there is a loss in diversity order in spatial correlation for $N_t = N_r = 12$; further, use of more receive antennas ($N_r = 18, N_t = 12$) alleviates this loss in performance. The proposed PDA algorithm can be used to decode *perfect codes* of large dimensions as well.

REFERENCES

[1] B. A. Sethuraman, B. S. Rajan, and V. Shashidhar, "Full-diversity high-rate space-time block codes from division algebras," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2596-2616, October 2003.

[2] J.-C. Belfiore, G. Rekaya, and E. Viterbo, "The golden code: A 2 × 2 full-rate space-time code with non-vanishing determinants," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1432-1436, April 2005.

[3] H. Jafarkhani, *Space-Time Coding: Theory and Practice*, Cambridge University Press, 2005.

[4] http://www.ruckuswireless.com/technology/beamflex.php

[5] J. Luo, K. Pattipati, P. Willett, and F. Hasegawa, "Near-optimal multiuser detection in synchronous CDMA using probabilistic data association," *IEEE Commun. Letters*, vol. 5, no. 9, pp. 361-363, 2001.

[6] Y. Huang and J. Zhang, "Generalized probabilistic data association multiuser detector," *Proc. ISIT'2004*, June-July 2004.

[7] P. H. Tan and L. K. Rasmussen, "Asymptotically optimal nonlinear MMSE multiuser detection based on multivariate Gaussian approximation," *IEEE Trans. Commun.*, pp. 1427-1438, Aug. 2006.

[8] D. Pham, K. R. Pattipati, P. K. Willet, and J. Luo, "A generalized probabilistic data association detector for multiantenna systems," *IEEE Commun. Letters*, vol. 8, no. 4, pp. 205-207, 2004.

[9] G. Latsoudas and N. D. Sidiropoulos, "A hybrid probabilistic data association-sphere decoding for multiple-input-multiple-output systems," *IEEE Sig. Proc. Letters*, pp. 309-312, April 2005.

[10] Y. Jia *et al*, "Probabilistic data association for symbol detection in MIMO Systems," *Elect. Letters*, vol. 42, no. 1, 5 Jan. 2006.

[11] D. Gesbert, H. Bölcskei, D. A. Gore, A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. on Commun.*, vol. 50, pp. 1926-1934, December 2002.

[12] K. Vishnu Vardhan, Saif K. Mohammed, A. Chockalingam, B. Sundar Rajan, "A low-complexity detector for large MIMO systems and multicarrier CDMA systems," *IEEE JSAC Spl. Iss. on Multiuser Detection for Adv. Commun. Systems and Networks*, pp. 473-485, April 2008.

[13] Saif K. Mohammed, A. Chockalingam, and B. Sundar Rajan "A low-complexity near-ML performance achieving algorithm for large MIMO detection," *Proc. IEEE ISIT'2008*, July 2008.

[14] Saif K. Mohammed, A. Chockalingam, and B. Sundar Rajan, "High-rate space-time coded large MIMO systems: Low-complexity detection and performance," *Proc. IEEE GLOBECOM'2008*, Dec. 2008.