Low-Complexity Near-ML 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) having large dimensions 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 about 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.

Keywords — Non-orthogonal STBCs, large dimensions, high spectral efficiency, low-complexity near-ML decoding, probabilistic data association.

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 \( 2 \times 2 \) 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 \( 16 \times 16 \) 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-ML 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 \(^1\) at 2.5 GHz are now commercially available [4] (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., \( 16 \times 16 \) STBC from CDA) in combination with large dimension near-ML 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]-[12]. 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]-[8]. PDA has been used in the detection of V-BLAST signals with small number of dimensions [10]-[12]. To our knowledge, PDA has not been reported for decoding non-orthogonal STBCs with hundreds of dimensions so far. Our results in this paper can be summarized as follows:

1. We adapt the PDA algorithm for decoding non-orthogonal STBCs with large dimensions. With i.i.d fading and perfect CSIR, the algorithm achieves near-SISO AWGN uncoded BER and near-capacity coded BER (within about 5 dB of the theoretical capacity) for \( 12 \times 12 \) STBC from CDA, 4-QAM, rate-3/4 turbo code, and 18 bps/Hz.

2. 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.

3. Relaxing the i.i.d fading assumption by adopting a spatially correlated MIMO channel model (proposed by Gesbert et al in [18]), we show that the performance loss due to spatial correlation is alleviated by using more receive spatial dimensions for a fixed receiver aperture.

4. Finally, the performance of the PDA algorithm is compared with that of the likelihood ascent search (LAS) algorithm we recently presented in [13]-[15]. 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

\(^1\)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.
its entries are modeled as i.i.d. random variables. Let $N_r$ and $N_t = n$ denote the number of receive and transmit antennas, respectively. Let $H_c \in \mathbb{C}^{N_r \times N_t}$ denote the channel gain matrix, where the $(i,j)$th entry in $H_c$ is the complex channel gain from the $i$th transmit antenna to the $j$th receive antenna. We assume that the channel gains remain constant over the duration of one STBC matrix and vary (i.i.d) from one STBC matrix to the other. Assuming rich scattering, we model the entries of $H_c$ as i.i.d $\mathcal{CN}(0,1)$.

The received space-time signal matrix, $Y_c \in \mathbb{C}^{N_r \times p}$, can be written as

$$Y_c = H_c X_c + N_c,$$

where $N_c \in \mathbb{C}^{N_r \times p}$ is the noise matrix at the receiver and its entries are modeled as i.i.d $\mathcal{CN}(0,\sigma^2) = \frac{N_c \mathcal{N}(0,1)}{\sqrt{n}}$, where $E_\gamma$ is the average energy of the transmitted symbols, and $\gamma$ is the average received SNR per receive antenna. Assume and (i,j)th entry in $Y_c$ is the received signal at the $i$th receive antenna in the $j$th time-slot. Consider linear dispersion STBCs, where $X_c$ can be written in the form:

$$X_c = \sum_{i=1}^{k} x_c(i) A_c(i),$$

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

$$y_c = \sum_{i=1}^{k} x_c(i) (H_c a_c(i)) + n_c = \widetilde{H}_c x_c + n_c,$$

where $y_c \in \mathbb{C}^{N_r \times p}$ is the received signal, $H_c \in \mathbb{C}^{N_r \times N_t} = (I \otimes H_c), a_c(i) \in \mathbb{C}^{N_r \times 1} = v \mathcal{CN}(0,\sigma^2)$, $n_c \in \mathbb{C}^{N_r \times 1}$ is the Gaussian noise, $x_c \in \mathbb{C}^{N_x \times 1}$ whose $i$th entry is the data symbol $x_c(i)$, and $H_c \in \mathbb{C}^{N_r \times N_t}$ whose $i$th column is $H_c(i)$, $i = 1,2,\ldots,k$. Each element of $x_c$ is an $M$-PAM/M-QAM symbol. Let $y_c$, $H_c$, $x_c$, $n_c$ be decomposed into real and imaginary parts as:

$$y_c = y_1 + jy_Q$$

$$n_c = n_1 + jn_Q$$

$$H_c = H_I + jH_Q.$$ 

Further, we define $H_r \in \mathbb{R}^{2N_r \times 2k}$, $y_r \in \mathbb{R}^{2N_r \times 1}$, $x_r \in \mathbb{R}^{2k \times 1}$, and $n_r \in \mathbb{R}^{2N_r \times 1}$ as:

$$H_r = \begin{pmatrix} H_I & -H_Q \\ H_Q & H_I \end{pmatrix},$$

$$y_r = [y_1^T \ y_Q^T]^T,$$

$$x_r = [x_1^T \ x_Q^T]^T,$$

$$n_r = [n_1^T \ n_Q^T]^T.$$ 

Now, (3) can be written as:

$$y_r = H_r x_r + n_r.$$ 

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

$$y = H' x + n,$$

where $H' = H_r \in \mathbb{R}^{2N_r \times 2k}$, $y = y_r \in \mathbb{R}^{2N_r \times 1}$, $x = x_r \in \mathbb{R}^{2k \times 1}$, and $n = n_r \in \mathbb{R}^{2N_r \times 1}$. We assume that the channel coefficients are known at the receiver but not at the transmitter. Let $\lambda_i$ denote the $i$-th eigenvalue of $n_i$-th signal set from which $x_i$ (ith entry of $x$) takes values, $i = 0, \ldots, 2k-1$.

$$\text{P} = \arg \min_{d \in \mathbb{S}} |d^T (H')^T H'd - 2y^T H'd|,$$

whose complexity is exponential in $k$.

A. Full-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 $A_c(i)$’s are permutation type), non-orthogonal STBCs from CDA, whose construction for arbitrary number of transmit antennas $n$ is given by the matrix in Eqn.(9.a) given at the bottom of this page. In (9.a), $\omega_{u,v} = \frac{e^{\pi i u/v}}{\sqrt{n}}$, $j = \sqrt{-1}$, and $d_{u,v}$, $0 \leq u, v \leq n - 1$ are the $n^2$ data symbols from a QAM alphabet. When $\delta = 1$, the code in (9.a) is information lossless (ILL), and when $\delta = e^{\pi i u/v}$ and $t = e^\theta$, it is of full-diversity and information lossless (FD-ILL) [1]. High spectral efficiencies with large $n$ can be achieved using this code construction. However, since these STBCs are non-orthogonal, 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-ML performance for large $n$ in low decoding complexities. The BER performance results we report in Sec. IV show that the PDA-based decoding algorithm we propose in the following section essentially meets 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 any rectangular QAM is straightforward. In the real-valued system model in (8), each entry of $y$ belongs to a $\sqrt{M}$-PAM constellation, where $M$ is the size of the original square QAM constellation. Let $b_i^{(0)}$, $b_i^{(1)}$, \ldots, $b_i^{(q-1)}$ denote the $q = \log_2(\sqrt{M})$ constituent bits of the $i$th entry $x_i$ of $x$.

$$[\begin{array}{c} \sum_{i=0}^{N_t-1} d_{1,i} t_i^1 \\ \sum_{i=0}^{N_t-1} d_{1,i} t_i^2 \\ \cdots \\ \sum_{i=0}^{N_t-1} d_{1,i} t_i^{N_t} \end{array}, \begin{array}{c} \sum_{i=0}^{N_t-1} d_{2,i} t_i^1 \\ \sum_{i=0}^{N_t-1} d_{2,i} t_i^2 \\ \cdots \\ \sum_{i=0}^{N_t-1} d_{2,i} t_i^{N_t} \end{array}, \ldots, \begin{array}{c} \sum_{i=0}^{N_t-1} d_{N_t-1,i} t_i^1 \\ \sum_{i=0}^{N_t-1} d_{N_t-1,i} t_i^2 \\ \cdots \\ \sum_{i=0}^{N_t-1} d_{N_t-1,i} t_i^{N_t} \end{array}]$$

(9.a)
We can write the value of each entry of $x$ as a linear combination of its constituent bits as

$$x_i = \sum_{j=0}^{q-1} 2^j b_i^{(j)}, \quad i = 0, 1, \ldots, 2k - 1. \quad (10)$$

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

$$b = b_0^{(0)} \ldots b_0^{(q-1)} b_1^{(0)} \ldots b_1^{(q-1)} \ldots b_{2k-1}^{(0)} \ldots b_{2k-1}^{(q-1)} T,$$

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

$$x = (I \otimes c)b,$$  

where $I$ is the $2k \times 2k$ identity matrix. Using (12), we can rewrite (8) as

$$y = H'(I \otimes c) b + n,$$  

where $H \in \mathbb{R}^{2N_r p \times 2qk}$ is the effective channel matrix. Our goal is to obtain $\hat{b}$, an estimate of the $b$ vector. For this, we iteratively update the statistics of each bit of $b$, as described in the following subsection, for a certain number of iterations, and hard decisions are made on the final statistics to get $\hat{b}$.

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 a priori probabilities as $P(b_i^{(j)} = +1) = P(b_i^{(j)} = -1) = 0.5$, $\forall i = 0, \ldots, 2k - 1$ and $j = 0, \ldots, q - 1$. In an iteration, the statistics of the bits are updated sequentially, i.e., the ordered sequence of updates in an iteration is $\{b_0^{(0)}, \ldots, b_0^{(q-1)}, \ldots, b_{2k-1}^{(0)}, \ldots, b_{2k-1}^{(q-1)}\}$. 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 given by

$$\Lambda_i^{(j)} \triangleq \frac{P(y\mid b_i^{(j)} = +1)}{P(y\mid b_i^{(j)} = -1)} = \frac{P(y\mid b_i^{(j)} = +1) P(b_i^{(j)} = +1)}{P(y\mid b_i^{(j)} = -1) P(b_i^{(j)} = -1)} \triangleq \alpha_i^{(j)} - \beta_i^{(j)}.$$  

Denoting the $t$th column of $H$ by $h_t$, we can write (13) as

$$y = h_{q+j} b_i^{(j)} + \sum_{l=0}^{2k-1} \sum_{m=0}^{q-1} h_{ql+m} b_i^{(m)} + n,$$  

where $n \in \mathbb{R}^{2N_r p \times 1}$ is the interference plus noise vector. To calculate $\beta_i^{(j)}$, we approximate the distribution of $n$ to be Gaussian, and hence $y$ 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 $y$ given $b_i^{(j)} = +1$ and $b_i^{(j)} = -1$. For notational simplicity, let us define $p_{i}^{+} = P(b_i^{(j)} = +1)$ and $p_{i}^{-} = P(b_i^{(j)} = -1)$. It is clear that $p_{i}^{+} + p_{i}^{-} = 1$.

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

$$\mu_i^{+} = h_{q+i} + \sum_{l=0}^{2k-1} \sum_{m=0}^{q-1} h_{ql+m} (2p_{i}^{m+} - 1).$$  

Similarly, we can write $\mu_i^{-}$ as

$$\mu_i^{-} = -h_{q+i} + \sum_{l=0}^{2k-1} \sum_{m=0}^{q-1} h_{ql+m} (2p_{i}^{m-} - 1) = \mu_i^{+} - 2h_{q+i}.$$  

Next, the $2N_r p \times 2N_r p$ covariance matrix $C_i^{j}$ of $y$ given $b_i^{(j)}$ is given by

$$C_i^{j} = \sigma^2 I + \sum_{l=0}^{2k-1} \sum_{m=0}^{q-1} h_{ql+m} (b_i^{(m)} - 2p_{i}^{m+} + 1) (18)$$

Assuming independence among the constituent bits, we can simplify $C_i^{j}$ in (13) as

$$C_i^{j} = \sigma^2 I + \sum_{l=0}^{2k-1} \sum_{m=0}^{q-1} h_{ql+m} (b_i^{(m)} - 2p_{i}^{m+} + 1) T. \quad (19)$$

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

$$P(y\mid b_i^{(j)} = \pm 1) = \frac{e^{-(y - \mu_i^{(j)} \pm)T (C_i^{j})^{-1} (y - \mu_i^{(j)} \pm)}}{(2\pi)^{N_r p} |C_i^{j}|^{\frac{1}{2}}}.$$  

Similarly, $P(y\mid b_i^{(j)} = -1)$ is given by

$$P(y\mid b_i^{(j)} = -1) = \frac{e^{-y - (2p_{i}^{m+} - 1) T (C_i^{j})^{-1} (y - \mu_i^{(j)} \pm)}}{(2\pi)^{N_r p} |C_i^{j}|^{\frac{1}{2}}}.$$  

Using (20) and (21), $\beta_i^{(j)}$ can be written as

$$\beta_i^{(j)} = P(y\mid b_i^{(j)} = +1) P(y\mid b_i^{(j)} = -1) = e^{-(y - \mu_i^{(j)} \pm)T (C_i^{j})^{-1} (y - \mu_i^{(j)} \pm)}. \quad (22)$$

Using (20) and (21), $\beta_i^{(j)}$ can be written as

$$\beta_i^{(j)} = P(y\mid b_i^{(j)} = +1) P(y\mid b_i^{(j)} = -1) = e^{-(y - \mu_i^{(j)} \pm)T (C_i^{j})^{-1} (y - \mu_i^{(j)} \pm)}. \quad (22)$$
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. From (14), and using $P(b_i^{(j)} = +1|y) + P(b_i^{(j)} = -1|y) = 1$, we have

$$P(b_i^{(j)} = +1|y) = \frac{\Lambda_i^{(j)}}{1 + \Lambda_i^{(j)}},$$ (23)

and

$$P(b_i^{(j)} = -1|y) = \frac{1}{1 + \Lambda_i^{(j)}}.$$ (24)

As an approximation, dropping the conditioning on $y$,

$$P(b_i^{(j)} = +1) \approx \frac{\Lambda_i^{(j)}}{1 + \Lambda_i^{(j)}},$$ (25)

and

$$P(b_i^{(j)} = -1) \approx \frac{1}{1 + \Lambda_i^{(j)}}.$$ (26)

Using the above procedure, we update $P(b_i^{(j)} = +1)$ and $P(b_i^{(j)} = -1)$ for all $i = 0, \ldots, 2k-1$ and $j = 0, \ldots, q-1$ sequentially. This completes one iteration of the algorithm; i.e., each iteration involves the computation of $\alpha_i^{(j)}$ and equations (16), (17), (19), (22), (14), (25), and (26) for all $i, j$. The updated values of $P(b_i^{(j)} = +1)$ and $P(b_i^{(j)} = -1)$ in (25) and (26) for all $i, j$ are fed back to the next iteration.

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 $\hat{b}_i^{(j)}$ 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, $C_i^j$, of size $2N,p \times 2N,p$ which requires $O(N^2_p p^3)$ complexity, which can be reduced as follows. Define matrix $D$ as

$$D \triangleq \sigma^2 I + \sum_{l=0}^{2k-1} \sum_{m=0}^{q-1} h_{q_{l+m}} h_{q_{l+m}}^T 4p_l^{m+} (1 - p_l^{m+}).$$ (27)

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

**Computation of $D^{-1}$:** We note that when the statistics of $b_i^{(j)}$ is updated using (25) and (26), the $D$ matrix in (27) also changes. A straightforward inversion of this updated $D$ matrix would require $O(N^2_p p^3)$ complexity. However, we can obtain the $D^{-1}$ from the previously available $D^{-1}$ in $O(N^2_p p^3)$ complexity as follows. Since the statistics of only $b_i^{(j)}$ is updated, the new $D$ matrix is just a rank one update of the old $D$ matrix. Therefore, using the matrix inversion lemma, the new $D^{-1}$ can be obtained from the old $D^{-1}$ as

$$D^{-1} \leftarrow D^{-1} - \frac{D^{-1} h_{n_{i+j}} h_{n_{i+j}}^T D^{-1}}{h_{n_{i+j}}^T D^{-1} h_{n_{i+j}} + \eta^{-1}}$$ (28)

where

$$\eta = 4p_i^{i+} (1 - p_i^{i+}) - 4p_{i,old}^{i+} (1 - p_{i,old}^{i+})$$ (29)

where $p_i^{i+}$ and $p_{i,old}^{i+}$ are the new (i.e., after the update in (25)) 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 (28) can be computed in $O(N^2_p p^2)$ complexity. Therefore, the computation of the new $D^{-1}$ using the old $D^{-1}$ can be done in $O(N^2_p p^2)$ complexity.

**Computation of $(C_i^j)^{-1}$:** Using (27) and (19), we can write $C_i^j$ in terms of $D$ as

$$(C_i^j)^{-1} = D - 4p_i^{i+} (1 - p_i^{i+}) h_{q_{i+j}} h_{q_{i+j}}^T.$$ (30)

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

$$(P + QRS)^{-1} = P^{-1} - P^{-1} Q (R^{-1} + SP^{-1} Q)^{-1} SP^{-1}.$$ (31)

Substituting $P_{2N,p \times 2N,p} = D$, $Q_{2N,p \times 1} = h_{q_{i+j}}$, $R_{1 \times 1} = -4p_i^{i+} (1 - p_i^{i+})$, and $S_{1 \times 2N,p} = h_{q_{i+j}}^T$ in (31), we get

$$(C_i^j)^{-1} = D^{-1} - \frac{D^{-1} h_{q_{i+j}} h_{q_{i+j}}^T D^{-1}}{h_{q_{i+j}}^T D^{-1} h_{q_{i+j}} - 4p_i^{i+} (1 - p_i^{i+})^{-1}}.$$ (32)

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

**Computation of $\mu_i^{i+}$ and $\mu_i^{i-}$:** Computation of $\beta_i^{(j)}$ involves the computation of $\mu_i^{i+}$ and $\mu_i^{i-}$ also. From (17), it is clear that $\mu_i^{i-}$ can be computed from $\mu_i^{i+}$ with a computational overhead of only $O(N,p)$. From (16), it can be seen that computing $\mu_i^{i+}$ would require $O(qN,pk)$ complexity. However, this complexity can be reduced as follows. Define vector $u$ as

$$u \triangleq \sum_{l=0}^{2k-1} \sum_{m=0}^{q-1} h_{q_{l+m}} (2p_l^{m+} - 1).$$ (33)

Using (16) and (33), we can write

$$\mu_i^{i+} = u + 2 (1 - p_i^{i+}) h_{q_{i+j}}.$$ (34)

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

$$u \leftarrow u + 2 (p_i^{i+} - p_i^{i+old}) h_{n_{i+j}},$$ (35)

whose complexity is $O(N,p)$. Hence, the computation of $\mu_i^{i+}$ in (33) and $\mu_i^{i-}$ in (34) needs $O(N,p)$ complexity. The listing of the proposed PDA algorithm is summarized in the Table-I in the next page.
Table-I: Proposed PDA-based Algorithm Listing

**Initialization**
1. \( p_{i}^{+} = p_{i}^{-} = 0.5 \), \( \Lambda_{i}^{(j)} = 1 \),
   \[ \forall i = 0, 1, \ldots, 2k - 1, j = 0, 1, \ldots, q - 1. \]
2. \( u = 0, \ D^{-1} = \left(HH^{T} + \sigma^{2}I\right)^{-1}. \)
3. num\_\text{iter}: number of iterations
4. \( \kappa = 1; \ \kappa \) is the iteration number
   
**Statistics update in the \( i \)th iteration**
5. for \( i = 0 \) to \( 2k - 1 \) \( \quad \) \( \text{End of for loop starting at line 5} \)
6. for \( j = 0 \) to \( q - 1 \)
   
   **Update of statistics of bit \( b_{i}^{(j)} \)**
7. \( \mu_{i}^{+} = u + 2(1 - p_{i}^{+})h_{i}^{+} \)
8. \( \mu_{i}^{-} = 2h_{i}^{-} \)
9. \( (C_{i}^{(j)})^{-1} = D^{-1} - \frac{h_{i}^{+} h_{i}^{+ T} + h_{i}^{-} h_{i}^{- T}}{4\eta_{i}^{+} (1 - p_{i}^{+}) - (y - \mu_{i}^{+})^{T} (C_{i}^{(j)})^{-1} (y - \mu_{i}^{+})} \)
10. \( p_{i,\text{old}}^{+} = p_{i}^{+}, \quad p_{i,\text{old}}^{-} = p_{i}^{-} \)
11. \( \alpha_{i}^{(j)} = \frac{p_{i}^{+}}{p_{i,\text{old}}^{+}}, \quad \alpha_{i}^{(j)} = \frac{p_{i}^{-}}{p_{i,\text{old}}^{-}} \)
12. \( \Lambda_{i}^{(j)} = \frac{p_{i}^{+}}{1 + \Lambda_{i}^{(j)}}, \quad \Lambda_{i}^{(j)} = \frac{p_{i}^{-}}{1 + \Lambda_{i}^{(j)}} \)
13. \( u \leftarrow u + 2(p_{i}^{+} - p_{i,\text{old}}^{+})h_{i}^{+} \)
14. \( \eta = 4p_{i}^{+} (1 - p_{i}^{+} - 4p_{i,\text{old}}^{+} (1 - p_{i}^{+})) \)
15. \( D^{-1} \leftarrow D^{-1} - \frac{h_{i}^{+} h_{i}^{+ T} + h_{i}^{-} h_{i}^{- T}}{4\eta_{i}^{+} (1 - p_{i}^{+}) - (y - \mu_{i}^{+})^{T} (C_{i}^{(j)})^{-1} (y - \mu_{i}^{+})} \)
16. \( \text{End; End of for loop starting at line 5} \)
17. if (\( \kappa = \text{num\_\text{iter}} \)) goto line 21
18. \( \kappa = \kappa + 1, \)goto line 5
19. \( \beta_{i}^{(j)} = \text{sgn}\left(\log\left(\Lambda_{i}^{(j)}\right)\right) \)
   \[ \forall i = 0, 1, \ldots, 2k - 1, j = 0, 1, \ldots, q - 1 \]
20. \( \tilde{\alpha}_{i} = \sum_{j=1}^{q} 2^{j} \beta_{i}^{(j)}, \quad \forall i = 0, 1, \ldots, 2k - 1 \)
21. \( \text{Terminate} \)

C. **Overall Complexity**

We need to compute \( HH^{T} \) at the start of the algorithm. This requires \( O(qkN_{t}^{2}p^{2}) \) complexity. So the computation of the initial \( D^{-1} \) in line 2 requires \( O(qkN_{t}^{2}p^{2}) \). Based on the complexity reduction in Sec. II-B, the complexity in updating the statistics of one constituent bit (lines 7 to 17) is \( O(N_{t}^{2}p^{3}) \). So, the complexity for the update of all the \( 2qk \) constituent bits in an iteration is \( O(N_{t}^{2}p^{3}) \). Since the number of iterations is fixed, the overall complexity of the algorithm is \( O(qkN_{t}^{2}p^{4}) + O(N_{t}^{2}p^{3}) \). For \( N_{t} = N_{r} \), since there are \( k \) symbols per STBC and \( q \) bits per symbol, the overall complexity per bit is \( O(p^{2}N_{t}^{2}) \).

IV. **Results and Discussions**

In this section, we present the simulated uncoded/coded BER of the PDA algorithm in decoding non-orthogonal STBCs from CDA. 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 rx antenna, \( \gamma \), in decoding \( 4 \times 4, 8 \times 8, 16 \times 16 \) STBCs from CDA with \( N_{t} = N_{r} = 4 \) and 4-QAM. Perfect channel state information at the receiver (CSIR) and i.i.d fading are assumed. For the same settings, the performance of the LAS algorithm in [13]-[15] 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 \( 16 \times 16 \) 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.

PDA versus LAS performance with 16-QAM: Figure 2 presents an uncoded BER comparison between PDA and LAS algorithms for \( 16 \times 16 \) STBC from CDA with \( N_{t} = N_{r} = 16 \) 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 \( 8 \times 8 \) and \( 16 \times 16 \) STBCs, at low SNRs (e.g., \( < 25 \) dB for \( 16 \times 16 \) STBC), PDA algorithm performs better by about 1 dB compared to LAS algorithm at \( 10^{-3} \) 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 \( 12 \times 12 \) ILL STBC with \( N_{t} = N_{r} = 12 \) and 4-QAM, which corresponds to a spectral efficiency of 18 bps/Hz. The theoretical minimum SNR required to achieve 18 bps/Hz spectral efficiency on a \( N_{t} = N_{r} = 12 \)
tering, and $d_i$ is the distance between $i$th tx and $i$th rx antennas ($R$), and scattering environment. In

Effect of Spatial MIMO Correlation: In Figs. 1 to 5 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 [17]. We relaxed the i.i.d. fading assumption by considering the correlated MIMO channel model in [18], which takes into account carrier frequency ($f_c$), spacing between antenna elements ($d_i, d_r$), distance between tx and rx antennas ($R$), and scattering environment. In

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 the theoretical minimum SNR, which is a good nearness to capacity performance.

Iterative PDA 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 $N_t \times N_r$ pilot matrix (for training purposes) followed by $N_d$ data STBC matrices are sent in each frame as shown in Fig. 4. 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 [10]: 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 $12 \times 12$ 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 $N_d = 8$. 2 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 5 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 [17]. We relaxed the i.i.d. fading assumption by considering the correlated MIMO channel model in [18], which takes into account carrier frequency ($f_c$), spacing between antenna elements ($d_i, d_r$), distance between tx and rx antennas ($R$), and scattering environment. In

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 $2 \times 2$ 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 multi-user 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.
Effect of spatial correlation on the performance of PDA in decoding $12 \times 12$ 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, $\theta_t = \theta_r = 90^\circ$, $Nrd_r = 72$ cm, $d_t = d_r$. Correlation degrades performance; using $N_r > N_t$ alleviates this performance loss.

[7] X. Wang, H. V. Poor, “Iterative (Turbo) soft interference cancellation and decoding for coded CDMA,” IEEE Trans. on Commun., vol. 47, no. 7, pp. 1046-1061, July 1999.

[8] P. H. Tan and L. K. Rasmussen, “Asymptotically optimal nonlinear MMSE multiuser detection based on multivariate Gaussian approximation,” IEEE Trans. on Commun., vol. 54, no. 8, pp. 1427-1438, August 2006.

[9] Y. Yin, Y. Huang, and J. Zhang, “Turbo equalization using probabilistic data association,” Proc. IEEE GLOBECOM’2004, vol. 4, pp. 2535-2539, November-December 2004.

[10] 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.

[11] G. Latsoudas and N. D. Sidiropoulos, “A hybrid probabilistic data association-sphere decoding for multiple-input-multiple-output systems,” IEEE Sig. Proc. Letters, vol. 12, no. 4, pp. 309-312, April 2005.

[12] Y. Jia et al, C. M. Vithanage, C. Andrieu, and R. J. Piechocki, “Probabilistic data association for symbol detection in MIMO Systems,” Electronic Letters, vol. 42, no. 1, 5 Jan. 2006.

[13] 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, vol. 26, no. 3, pp. 473-485, April 2008.

[14] 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, Toronto, Canada, July 2008.

[15] 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, New Orleans, USA, Dec. 2008.

[16] Ahmed Zaki, Saif K. Mohammed, A. Chockalingam, and B. Sundar Rajan, “A training-based iterative detection/channel estimation scheme for large non-orthogonal STBC MIMO systems,” accepted in IEEE ICC’2009, Dresden, Germany, April 2009. Also see arXiv:0809.2446v2 [cs.IT] 4 Oct 2008.

[17] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Khan, “Fading correlation and its effect on the capacity of multi-antenna systems,” IEEE Trans. on Commun., vol. 48, pp. 502-513, March 2000.

[18] 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.

[19] F. E. Oggier, G. Rekaya, J.-C. Belfiore, and E. Viterbo, “Perfect space-time block codes,” IEEE Trans. on Inform. Theory, vol. 52, no. 9, September 2006.