High-Rate Space–Time Coded Large-MIMO Systems: Low-Complexity Detection and Channel Estimation

Saif K. Mohammed, Ahmed Zaki, A. Chockalingam, *Senior Member, IEEE*, and B. Sundar Rajan, *Senior Member, IEEE*

***Abstract*—In this paper, we present a low-complexity algorithm for detection in high-rate, non-orthogonal space–time block coded (STBC) large-multiple-input multiple-output (MIMO) systems that achieve high spectral efficiencies of the order of tens of bps/Hz. We also present a training-based iterative detection/channel estimation scheme for such large STBC MIMO systems. Our simulation results show that excellent bit error rate and nearness-to-capacity performance are achieved by the proposed multistage *likelihood ascent search* (M-LAS) detector in conjunction with the proposed iterative detection/channel estimation scheme at low complexities. The fact that we could show such good results for large STBCs like 16×16 and 32×32 STBCs from Cyclic Division Algebras (CDA) operating at spectral efficiencies in excess of 20 bps/Hz (even after accounting for the overheads meant for pilot based training for channel estimation and turbo coding) establishes the effectiveness of the proposed detector and channel estimator. We decode perfect codes of large dimensions using the proposed detector. With the feasibility of such a low-complexity detection/channel estimation scheme, large-MIMO systems with tens of antennas operating at several tens of bps/Hz spectral efficiencies can become practical, enabling interesting high data rate wireless applications.**

***Index Terms*—Channel estimation, high spectral efficiencies, large-multiple-input multiple-output (MIMO) systems, low-complexity detection, non-orthogonal space–time block codes.**

I. Introduction

CURRENT wireless standards (e.g., IEEE 802.11n and 802.16e) have adopted multiple-input multiple-output (MIMO) techniques [1]–[3] to achieve the benefits of transmit diversity (using space–time coding) and high data rates (using spatial multiplexing). They, however, harness only a limited potential of MIMO benefits since they use only a small number of transmit antennas (e.g., two to four antennas). Significant benefits can be realized if large number of antennas are used; e.g., large-MIMO systems with tens of antennas in communication terminals can enable multi-giga bit rate transmissions at high spectral efficiencies of the order of *several tens of bps/Hz*.[1] Key challenges in realizing such large-MIMO systems include low-complexity detection and channel estimation, RF/IF technologies, and placement of large number of antennas in communication terminals.[2] Our focus in this paper is on low-complexity detection and channel estimation for large-MIMO systems.

Spatial multiplexing (V-BLAST) with large number of transmit antennas can offer high spectral efficiencies, but it does not give transmit diversity. On the other hand, well known orthogonal space–time block codes (STBC) have the advantages of full transmit diversity and low decoding complexity, but they suffer from rate loss for increasing number of transmit antennas [3], [5], [6]. However, *full-rate, non-orthogonal STBCs from Cyclic Division Algebras (CDA)* [7] are attractive to achieve high spectral efficiencies in addition to achieving full transmit diversity, using large number of transmit antennas. For example, a 32×32 STBC matrix from CDA has 1024 symbols (i.e., 32 complex symbols per channel use), and using this STBC along with 16-QAM and rate-3/4 turbo code offers a spectral efficiency of 96 bps/Hz. While maximum-likelihood (ML) decoding of orthogonal STBCs can be achieved in linear complexity, ML or near-ML decoding of non-orthogonal STBCs with large number of antennas at low complexities has been a challenge. Channel estimation is also a key issue in large-MIMO systems. In this paper, we address these two challenging problems; our proposed solutions can potentially enable realization of large-MIMO systems in practice. Sphere decoding and several of its low-complexity variants are known in the literature [8]–[11]. These detectors, however, are prohibitively complex for large number of antennas. Recent approaches to low-complexity multiuser/MIMO detection involve application of techniques from belief propagation [12], Markov Chain Monte-Carlo methods [13], neural networks [14]–[16], etc. In particular, in [15] and [16], we presented a powerful Hopfield neural network based low-complexity search algorithm for detecting large-MIMO V-BLAST signals,

Manuscript received September 01, 2008; revised August 29, 2009. Current version published January 13, 2010. This work in part was presented at IEEE ISIT'2008, Toronto, ON, Canada, July 2008, at IEEE GLOBECOM'2008, New Orleans, LA, December 2008, and at IEEE ICC'2009, Dresden, Germany, June 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Robert Calderbank.

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India. (e-mail: saifind2007@yahoo.com; zakismail@gmail.com; achockal@ece.iisc.ernet.in; bsrajan@ece.iisc.ernet.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2009.2035862

[1]Spectral efficiencies achieved in current MIMO wireless standards are only about 10 bps/Hz or less.

[2]WiFi products in 2.5-GHz band which use 12 transmit antennas for beamforming purposes are becoming commercially available [4]. With such RF and antenna technologies for placing large number of antennas in medium/large aperture communication terminals (like set-top boxes/laptops) getting increasingly matured, low-complexity high-performance MIMO baseband receiver techniques (e.g., detection and channel estimation) are crucial to enable practical implementations of high spectral efficiency large-MIMO systems, which, in turn, can enable high data rate applications like wireless IPTV/HDTV distribution.

and showed that it performs quite close to (within 4.6 dB of) the theoretical capacity, at high spectral efficiencies of the order of tens to hundreds of bps/Hz using tens to hundreds of antennas, at an average per-symbol detection complexity of just $O(N_tN_r)$, where N_t and N_r denote the number of transmit and receive antennas, respectively.

In this paper, we present 1) a low-complexity near-ML achieving detector, and 2) an iterative detection/channel estimation scheme for large non-orthogonal STBC MIMO systems having tens of transmit and receive antennas. Our key contributions here can be summarized as follows.

1) We generalize the 1-symbol update-based *likelihood ascent search* (LAS) algorithm we proposed in [15], [16], by employing a low-complexity multistage multi-symbol update based strategy; we refer to this new algorithm as multistage LAS (M-LAS) algorithm. We show that the M-LAS algorithm outperforms the basic LAS algorithm with some increase in complexity.
2) We propose a method to generate soft outputs from the M-LAS output vector. Soft outputs generation was not considered in [15] and [16]. The proposed soft outputs generation for the individual bits results in about 1 to 1.5 dB improvement in coded bit error rate (BER) compared to hard decision M-LAS outputs.
3) Assuming independent and identically distributed (i.i.d.) fading and perfect channel state information at the receiver (CSIR), our simulation results show that the proposed M-LAS algorithm is able to decode large non-orthogonal STBCs (e.g., 16×16 and 32×32 STBCs) and achieve near single-input single-output (SISO) AWGN uncoded BER performance as well as near-capacity (within 4 dB from theoretical capacity) coded BER performance.
4) Using the proposed detector, we decode and report the simulated BER performance of "perfect codes" [17]–[21] of large dimensions.
5) Presenting a BER performance and complexity comparison of the proposed CDA STBC/M-LAS detection approach with other large-MIMO/detector approaches (e.g., stacked Alamouti codes/QOSTBCs and associated interference canceling receivers reported in [22]), we show that the proposed approach outperforms the other considered approaches, both in terms of performance as well as complexity.
6) We present simulation results that quantify the loss in BER performance due to spatial correlation in large-MIMO systems, by considering a more realistic spatially correlated MIMO fading channel model proposed by Gesbert *et al.* in [23]. We show that this loss in performance can be alleviated by providing more receive dimensions (i.e., more receive antennas than transmit antennas).
7) Finally, we present a training-based iterative detection/channel estimation scheme for large STBC MIMO systems. We report BER and nearness-to-capacity results when the channel matrix is estimated using the proposed iterative scheme and compare these results with those obtained using perfect CSIR assumption.

The rest of the paper is organized as follows. In Section II, we present the STBC MIMO system model considered. The proposed detection algorithm is presented in Section III. BER performance results with perfect CSIR are presented in Section IV. This section includes the results on the effect of spatial correlation, BER performance of large perfect codes, and comparison of the proposed scheme with other large-MIMO architecture/detector combinations. The proposed iterative detection/channel estimation scheme and the corresponding performance results are presented in Section V. Conclusions are presented in Section VI.

II. System Model

Consider a STBC MIMO system with multiple transmit and multiple 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 r is given by $r \triangleq (k)/(p)$. Let N_r and $N_t = n$ denote 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 i.i.d $\mathcal{CN}(0, 1)$.[3] 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 $\mathcal{CN}(0, \sigma^2 = (N_tE_s)/(\gamma))$, 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. In a linear dispersion (LD) STBC, $\mathbf{X}_c$ can be decomposed into a linear combination of weight matrices corresponding to each data symbol and its conjugate as [3]

$$\mathbf{X}_c = \sum_{i=1}^{k} x_c^{(i)}\mathbf{A}_c^{(i)} + \left(x_c^{(i)}\right)^* \mathbf{E}_c^{(i)} \tag{2}$$

where $x_c^{(i)}$ is the ith complex data symbol, and $\mathbf{A}_c^{(i)}, \mathbf{E}_c^{(i)} \in \mathbb{C}^{N_t\times p}$ are its corresponding weight matrices. The detection algorithm we propose in this paper can decode general LD STBCs of the form in (2). For the purpose of simplicity in exposition, here we consider a subclass of LD STBCs, where $\mathbf{X}_c$ can be written in the form

$$\mathbf{X}_c = \sum_{i=1}^{k} x_c^{(i)}\mathbf{A}_c^{(i)}. \tag{3}$$

From (1) and (3), applying the $\mathrm{vec}(\cdot)$ operation[4] we have

$$\mathrm{vec}\,(\mathbf{Y}_c) = \sum_{i=1}^{k} x_c^{(i)}\mathrm{vec}\,\left(\mathbf{H}_c\mathbf{A}_c^{(i)}\right) + \mathrm{vec}\,(\mathbf{N}_c). \tag{4}$$

[3] $\mathcal{CN}(0, \sigma^2)$ denotes a circularly symmetric complex Gaussian distribution with mean zero and variance σ^2.

[4] For a $p \times q$ matrix $\mathbf{M} = [\mathbf{m}_1\mathbf{m}_2\cdots\mathbf{m}_q]$, where $\mathbf{m}_i$ is the ith column of $\mathbf{M}$, $\mathrm{vec}(\mathbf{M})$ is a $pq \times 1$ vector defined as $\mathrm{vec}(\mathbf{M}) = [\mathbf{m}_1^T\mathbf{m}_2^T\cdots\mathbf{m}_q^T]^T$, where $[\cdot]^T$ denotes the transpose operation.

If $\mathbf{U},\mathbf{V},\mathbf{W},\mathbf{D}$ are matrices such that $\mathbf{D}=\mathbf{UWV}$, then it is true that $\text{vec}\,(\mathbf{D})=(\mathbf{V}^T\otimes\mathbf{U})\,\text{vec}\,(\mathbf{W})$, where $\otimes$ denotes tensor product of matrices [24]. Using this, we can write (4) as

$$\text{vec}\,(\mathbf{Y}_c)=\sum_{i=1}^{k}x_c^{(i)}(\mathbf{I}\otimes\mathbf{H}_c)\,\text{vec}\,\left(\mathbf{A}_c^{(i)}\right)+\text{vec}\,(\mathbf{N}_c) \quad (5)$$

where $\mathbf{I}$ is the $p\times p$ identity matrix. Further, define $\mathbf{y}_c\triangleq\text{vec}\,(\mathbf{Y}_c)$, $\widehat{\mathbf{H}}_c\triangleq(\mathbf{I}\otimes\mathbf{H}_c)$, $\mathbf{a}_c^{(i)}\triangleq\text{vec}\,(\mathbf{A}_c^{(i)})$, and $\mathbf{n}_c\triangleq\text{vec}\,(\mathbf{N}_c)$. From these definitions, it is clear that $\mathbf{y}_c\in\mathbb{C}^{N_rp\times1}$, $\widehat{\mathbf{H}}_c\in\mathbb{C}^{N_rp\times N_tp}$, $\mathbf{a}_c^{(i)}\in\mathbb{C}^{N_tp\times1}$, and $\mathbf{n}_c\in\mathbb{C}^{N_rp\times1}$. Let us also define a matrix $\widetilde{\mathbf{H}}_c\in\mathbb{C}^{N_rp\times k}$, whose ith column is $\widehat{\mathbf{H}}_c\,\mathbf{a}_c^{(i)}, i=1,\ldots,k$. Let $\mathbf{x}_c\in\mathbb{C}^{k\times1}$, whose ith entry is the data symbol $x_c^{(i)}$. With these definitions, we can write (5) as

$$\mathbf{y}_c=\sum_{i=1}^{k}x_c^{(i)}\left(\widehat{\mathbf{H}}_c\,\mathbf{a}_c^{(i)}\right)+\mathbf{n}_c=\widetilde{\mathbf{H}}_c\mathbf{x}_c+\mathbf{n}_c. \quad (6)$$

Each element of $\mathbf{x}_c$ is an $\mathcal{M}$-PAM or $\mathcal{M}$-QAM symbol. $\mathcal{M}$-PAM symbols take discrete values from $\{A_m, m=1,\ldots,\mathcal{M}\}$, where $A_m=(2m-1-\mathcal{M})$, and $\mathcal{M}$-QAM is nothing but two PAMs in quadrature. Let $\mathbf{y}_c,\widetilde{\mathbf{H}}_c,\mathbf{x}_c$, and $\mathbf{n}_c$ be decomposed into real and imaginary parts as

$$\begin{aligned}&\mathbf{y}_c=\mathbf{y}_I+j\mathbf{y}_Q,\mathbf{x}_c=\mathbf{x}_I+j\mathbf{x}_Q,\\&\mathbf{n}_c=\mathbf{n}_I+j\mathbf{n}_Q,\widetilde{\mathbf{H}}_c=\widetilde{\mathbf{H}}_I+j\widetilde{\mathbf{H}}_Q.\end{aligned} \quad (7)$$

Further, we define $\mathbf{x}_r\in\mathbb{R}^{2k\times1}$, $\mathbf{y}_r\in\mathbb{R}^{2N_rp\times1}$, $\mathbf{H}_r\in\mathbb{R}^{2N_rp\times2k}$, and $\mathbf{n}_r\in\mathbb{R}^{2N_rp\times1}$ as

$$\begin{aligned}&\mathbf{x}_r=\left[\mathbf{x}_I^T\mathbf{x}_Q^T\right]^T,\mathbf{y}_r=\left[\mathbf{y}_I^T\mathbf{y}_Q^T\right]^T,\\&\mathbf{H}_r=\begin{pmatrix}\widetilde{\mathbf{H}}_I & -\widetilde{\mathbf{H}}_Q\\ \widetilde{\mathbf{H}}_Q & \widetilde{\mathbf{H}}_I\end{pmatrix},\mathbf{n}_r=\left[\mathbf{n}_I^T\mathbf{n}_Q^T\right]^T.\end{aligned} \quad (8)$$

Now, (6) can be written as

$$\mathbf{y}_r=\mathbf{H}_r\mathbf{x}_r+\mathbf{n}_r. \quad (9)$$

Henceforth, we work with the real-valued system in (9). For notational simplicity, we drop subscripts r in (9) and write

$$\mathbf{y}=\mathbf{Hx}+\mathbf{n} \quad (10)$$

where $\mathbf{H}=\mathbf{H}_r\in\mathbb{R}^{2N_rp\times2k}$, $\mathbf{y}=\mathbf{y}_r\in\mathbb{R}^{2N_rp\times1}$, $\mathbf{x}=\mathbf{x}_r\in\mathbb{R}^{2k\times1}$, and $\mathbf{n}=\mathbf{n}_r\in\mathbb{R}^{2N_rp\times1}$. The channel coefficients are assumed to be known only at the receiver but not at the transmitter. Let $\mathbb{A}_i$ denote the $\mathcal{M}$-PAM signal set from which x_i (ith entry of $\mathbf{x}$) takes values $i=1,\ldots,2k$. Now, define a $2k$-dimensional signal space $\mathbb{S}$ to be the Cartesian product of $\mathbb{A}_1$ to $\mathbb{A}_{2k}$. The ML solution is given by

$$\begin{aligned}\mathbf{d}_{\text{ML}}&=\underset{\mathbf{d}\in\mathbb{S}}{\text{argmin}}\ \|\mathbf{y}-\mathbf{Hd}\|^2\\&=\underset{\mathbf{d}\in\mathbb{S}}{\text{argmin}}\ \mathbf{d}^T\mathbf{H}^T\mathbf{Hd}-2\mathbf{y}^T\mathbf{Hd}\end{aligned} \quad (11)$$

whose complexity is exponential in k [25].

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 [26], whose construction for arbitrary number of transmit antennas n is given by the matrix in (11a) shown at the bottom of the page [7].

In (11a), $\omega_n=e^{(\mathbf{j}2\pi)/(n)}$, $\mathbf{j}=\sqrt{-1}$, and $x_{u,v}, 0\le u,v\le n-1$ are the data symbols from a QAM alphabet. When $\delta=e^{\sqrt{5}\mathbf{j}}$ and $t=e^{\mathbf{j}}$, the STBC in (11a) achieves full transmit diversity (under ML decoding) as well as information-losslessness [7]. When $\delta=t=1$, the code ceases to be of full-diversity (FD), but continues to be information-lossless (ILL) [27], [52]. High spectral efficiencies with large n can be achieved using this code construction. For example, with $n=32$ transmit antennas, the 32×32 STBC from (11a) with 16-QAM and rate-3/4 turbo code achieves a spectral efficiency of 96 bps/Hz. This high spectral efficiency is achieved along with the full-diversity of order nN_r. 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 at low detection complexities. Our proposed detector, termed as the *multistage likelihood ascent search (M-LAS) detector*, presented in the following section essentially addresses this challenging issue.

III. Proposed Multistage LAS Detector

The proposed M-LAS algorithm consists of a sequence of likelihood-ascent search stages, where the likelihood increases monotonically with every search stage. Each search stage consists of several substages. There can be at most M substages, each consisting of one or more iterations (the first substage can have one or more iterations, whereas all the other substages can have at most one iteration). In the first substage, the algorithm updates one symbol per iteration such that the likelihood mono-

$$\begin{bmatrix}\sum_{i=0}^{n-1}x_{0,i}t^i & \delta\sum_{i=0}^{n-1}x_{n-1,i}\omega_n^it^i & \delta\sum_{i=0}^{n-1}x_{n-2,i}\omega_n^{2i}t^i & \cdots & \delta\sum_{i=0}^{n-1}x_{1,i}\omega_n^{(n-1)i}t^i\\ \sum_{i=0}^{n-1}x_{1,i}t^i & \sum_{i=0}^{n-1}x_{0,i}\omega_n^it^i & \delta\sum_{i=0}^{n-1}x_{n-1,i}\omega_n^{2i}t^i & \cdots & \delta\sum_{i=0}^{n-1}x_{2,i}\omega_n^{(n-1)i}t^i\\ \sum_{i=0}^{n-1}x_{2,i}t^i & \sum_{i=0}^{n-1}x_{1,i}\omega_n^it^i & \sum_{i=0}^{n-1}x_{0,i}\omega_n^{2i}t^i & \cdots & \delta\sum_{i=0}^{n-1}x_{3,i}\omega_n^{(n-1)i}t^i\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \sum_{i=0}^{n-1}x_{n-2,i}t^i & \sum_{i=0}^{n-1}x_{n-3,i}\omega_n^it^i & \sum_{i=0}^{n-1}x_{n-4,i}\omega_n^{2i}t^i & \cdots & \delta\sum_{i=0}^{n-1}x_{n-1,i}\omega_n^{(n-1)i}t^i\\ \sum_{i=0}^{n-1}x_{n-1,i}t^i & \sum_{i=0}^{n-1}x_{n-2,i}\omega_n^it^i & \sum_{i=0}^{n-1}x_{n-3,i}\omega_n^{2i}t^i & \cdots & \sum_{i=0}^{n-1}x_{0,i}\omega_n^{(n-1)i}t^i\end{bmatrix} \quad (11a)$$

tonically increases from one iteration to the next until a local minima is reached. Upon reaching this local minima, the algorithm initiates the second substage.

In the second substage, a two-symbol update is tried to further increase the likelihood. If the algorithm succeeds in increasing the likelihood by two-symbol update, it starts the next search stage. If the algorithm does not succeed in the second substage, it goes to the third substage where a three-symbol update is tried to further increase the likelihood. Essentially, in the Kth substage, a K-symbol update is tried to further increase the likelihood. This goes on until a) either the algorithm succeeds in the Kth substage for some $K \leq M$ (in which case a new search stage is initiated), or b) the algorithm terminates.

The M-LAS algorithm starts with an initial solution $\mathbf{d}^{(0)}$, given by $\mathbf{d}^{(0)} = \mathbf{B}\mathbf{y}$, where $\mathbf{B}$ is the initial solution filter, which can be a matched filter (MF) or zero-forcing (ZF) filter or MMSE filter. The index m in $\mathbf{d}^{(m)}$ denotes the iteration number in a substage of a given search stage. The ML cost function after the kth iteration in a given search stage is

$$C^{(k)} = \mathbf{d}^{(k)^T}\mathbf{H}^T\mathbf{H}\mathbf{d}^{(k)} - 2\mathbf{y}^T\mathbf{H}\mathbf{d}^{(k)}. \tag{12}$$

A. One-Symbol Update

Let us assume that we update the pth symbol in the $(k+1)$th iteration; p can take value from $1, \ldots, N_t$ for $\mathcal{M}$-PAM and $1, \ldots, 2N_t$ for $\mathcal{M}$-QAM. The update rule can be written as

$$\mathbf{d}^{(k+1)} = \mathbf{d}^{(k)} + \lambda_p^{(k)}\mathbf{e}_p \tag{13}$$

where $\mathbf{e}_p$ denotes the unit vector with its pth entry only as one, and all other entries as zero. Also, for any iteration k, $\mathbf{d}^{(k)}$ should belong to the space $\mathbb{S}$, and therefore $\lambda_p^{(k)}$ can take only certain integer values. For example, in case of 4-PAM or 16-QAM (both have the same signal set $\mathbb{A}_p = \{-3, -1, 1, 3\}$), $\lambda_p^{(k)}$ can take values only from $\{-6, -4, -2, 0, 2, 4, 6\}$. Using (12) and (13), and defining a matrix $\mathbf{G}$ as

$$\mathbf{G} \triangleq \mathbf{H}^T\mathbf{H} \tag{14}$$

we can write the cost difference as

$$\begin{aligned}\Delta C_p^{k+1} &\triangleq C^{(k+1)} - C^{(k)} \\ &= \lambda_p^{(k)^2}(\mathbf{G})_{p,p} - 2\lambda_p^{(k)}z_p^{(k)}\end{aligned} \tag{15}$$

where $\mathbf{h}_p$ is the pth column of $\mathbf{H}$, $\mathbf{z}^{(k)} = \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{d}^{(k)})$, $z_p^{(k)}$ is the pth entry of the $\mathbf{z}^{(k)}$ vector, and $(\mathbf{G})_{p,p}$ is the (p,p)th entry of the $\mathbf{G}$ matrix. Also, let us define a_p and $l_p^{(k)}$ as

$$a_p = (\mathbf{G})_{p,p}, \qquad l_p^{(k)} = \left|\lambda_p^{(k)}\right|. \tag{16}$$

With the above variables defined, we can rewrite (15) as

$$\Delta C_p^{k+1} = l_p^{(k)^2}a_p - 2l_p^{(k)}\left|z_p^{(k)}\right| \operatorname{sgn}\left(\lambda_p^{(k)}\right) \operatorname{sgn}\left(z_p^{(k)}\right) \tag{17}$$

where $\operatorname{sgn}(\cdot)$ denotes the signum function. For the ML cost function to reduce from the kth to the $(k+1)$th iteration, the cost difference should be negative. Using this fact and that a_p and $l_p^{(k)}$ are non-negative quantities, we can conclude from (17) that the sign of $\lambda_p^{(k)}$ must satisfy

$$\operatorname{sgn}\left(\lambda_p^{(k)}\right) = \operatorname{sgn}\left(z_p^{(k)}\right). \tag{18}$$

Using (18) in (17), the ML cost difference can be rewritten as

$$\mathcal{F}\left(l_p^{(k)}\right) \triangleq \Delta C_p^{k+1} = l_p^{(k)^2}a_p - 2l_p^{(k)}\left|z_p^{(k)}\right|. \tag{19}$$

For $\mathcal{F}(l_p^{(k)})$ to be non-positive, the necessary and sufficient condition from (19) is that

$$l_p^{(k)} < \frac{2\left|z_p^{(k)}\right|}{a_p}. \tag{20}$$

However, we can find the value of $l_p^{(k)}$ which satisfies (20) and at the same time gives the largest descent in the ML cost function from the kth to the $(k+1)$th iteration (when symbol p is updated). Also, $l_p^{(k)}$ is constrained to take only certain integer values, and therefore the brute-force way to get optimum $l_p^{(k)}$ is to evaluate $\mathcal{F}(l_p^{(k)})$ at all possible values of $l_p^{(k)}$. This would become computationally expensive as the constellation size $\mathcal{M}$ increases. However, for the case of one-symbol update, we could obtain a closed-form expression for the optimum $l_p^{(k)}$ that minimizes $\mathcal{F}(l_p^{(k)})$, which is given by (corresponding theorem and proof are given in the Appendix)

$$l_{p,\text{opt}}^{(k)} = 2\left\lfloor\frac{\left|z_p^{(k)}\right|}{2a_p}\right\rceil \tag{21}$$

where $\lfloor\cdot\rceil$ denotes the rounding operation, where for a real number x, $\lfloor x\rceil$ is the integer closest to x. If the pth symbol in $\mathbf{d}^{(k)}$, i.e., $d_p^{(k)}$, were indeed updated, then the new value of the symbol would be given by

$$\tilde{d}_p^{(k+1)} = d_p^{(k)} + l_p^{(k)} \operatorname{sgn}\left(z_p^{(k)}\right). \tag{22}$$

However, $\tilde{d}_p^{(k+1)}$ can take values only in the set $\mathbb{A}_p$, and therefore we need to check for the possibility of $\tilde{d}_p^{(k+1)}$ being greater than $(\mathcal{M}-1)$ or less than $-(\mathcal{M}-1)$. If $\tilde{d}_p^{(k+1)} > (\mathcal{M}-1)$, then $l_p^{(k)}$ is adjusted so that the new value of $\tilde{d}_p^{(k+1)}$ with the adjusted value of $l_p^{(k)}$ using (22) is $(\mathcal{M}-1)$. Similarly, if $\tilde{d}_p^{(k+1)} < -(\mathcal{M}-1)$, then $l_p^{(k)}$ is adjusted so that the new value of $\tilde{d}_p^{(k+1)}$ is $-(\mathcal{M}-1)$. Let $\tilde{l}_{p,\text{opt}}^{(k)}$ be obtained from $l_{p,\text{opt}}^{(k)}$ after these adjustments. It can be shown that if $\mathcal{F}(l_{p,\text{opt}}^{(k)})$ is non-positive, then $\mathcal{F}(\tilde{l}_{p,\text{opt}}^{(k)})$ is also non-positive. We compute $\mathcal{F}(\tilde{l}_{p,\text{opt}}^{(k)})$, $\forall\ p = 1, \ldots, 2N_t^2$. Now, let

$$s = \underset{p}{\operatorname{argmin}}\ \mathcal{F}\left(\tilde{l}_{p,\text{opt}}^{(k)}\right). \tag{23}$$

If $\mathcal{F}(\tilde{l}_{s,\text{opt}}^{(k)}) < 0$, the update for the $(k+1)$th iteration is

$$\mathbf{d}^{(k+1)} = \mathbf{d}^{(k)} + \tilde{l}_{s,\text{opt}}^{(k)} \operatorname{sgn}\left(z_s^{(k)}\right)\mathbf{e}_s, \tag{24}$$

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} - \tilde{l}_{s,\text{opt}}^{(k)} \operatorname{sgn}\left(z_s^{(k)}\right)\mathbf{g}_s \tag{25}$$

where $\mathbf{g}_s$ is the sth column of $\mathbf{G}$. The update in (25) follows from the definition of $\mathbf{z}^{(k)}$ in (15). If $\mathcal{F}(\tilde{l}^{(k)}_{s,\text{opt}}) \geq 0$, then the one-symbol update search terminates. The data vector at this point is referred to as "one-symbol update local minima." After reaching the one-symbol update local minima, we look for a further decrease in the cost function by updating multiple symbols simultaneously.

B. Why Multiple Symbol Updates?

The motivation for trying out multiple symbol updates can be explained as follows. Let $\mathbb{L}_K \subseteq \mathbb{S}$ denote the set of data vectors such that for any $\mathbf{d} \in \mathbb{L}_K$, if a K-symbol update is performed on $\mathbf{d}$ resulting in a vector $\mathbf{d}'$, then $\|\mathbf{y} - \mathbf{H}\mathbf{d}'\| \geq \|\mathbf{y} - \mathbf{H}\mathbf{d}\|$. We note that $\mathbf{d}_{\text{ML}} \in \mathbb{L}_K, \forall K = 1, 2, \ldots, 2N_t$, because any number of symbol updates on $\mathbf{d}_{\text{ML}}$ will not decrease the cost function. We define another set $\mathbb{M}_K = \bigcap_{j=1}^{K} \mathbb{L}_j$. Note that $\mathbf{d}_{\text{ML}} \in \mathbb{M}_K, \forall K = 1, 2, \ldots, 2N_t$, and $\mathbb{M}_{2N_t} = \{\mathbf{d}_{\text{ML}}\}$, i.e., $\mathbb{M}_{2N_t}$ is a singleton set with $\mathbf{d}_{\text{ML}}$ as the only element. It is noted that if the updates are done optimally, then the output of the K-LAS algorithm converges to a vector in $\mathbb{M}_K$. Also, $|\mathbb{M}_{K+1}| \leq |\mathbb{M}_K|, K = 1, 2, \ldots, 2N_t - 1$. For any $\mathbf{d} \in \mathbb{M}_K, K = 1, 2, \ldots, 2N_t$ and $\mathbf{d} \neq \mathbf{d}_{\text{ML}}$, it can be seen that $\mathbf{d}$ and $\mathbf{d}_{\text{ML}}$ will differ in $K + 1$ or more locations. The probability that $\mathbf{d}_{\text{ML}} = \mathbf{x}$ increases with increasing SNR, and so the separation between $\mathbf{d} \in \mathbb{M}_K$ and $\mathbf{x}$ will monotonically increase with increasing K. Since $\mathbf{d}_{\text{ML}} \in \mathbb{M}_K$, and $|\mathbb{M}_K|$ decreases monotonically with increasing K, there will be lesser non-ML data vectors to which the algorithm can converge to for increasing K. Therefore, the probability of the noise vector $\mathbf{n}$ inducing an error would decrease with increasing K. This indicates that K-symbol updates with large K could get near to ML performance with increasing complexity for increasing K.

C. K-Symbol Update, $1 < K \leq 2N_t^2$

In this subsection, we present the update algorithm for the general case where K symbols, $1 < K \leq 2N_t^2$, are updated simultaneously in one iteration. K-symbol updates can be done in $\binom{2N_t^2}{K}$ ways, among which we seek to find that update which gives the largest reduction in the ML cost. Assume that in the $(k+1)$th iteration, K symbols at the indices $i_1, i_2, \ldots, i_K$ of $\mathbf{d}^{(k)}$ are updated. Each $i_j, j = 1, 2, \ldots, K$, can take values from $1, 2, \ldots, N_t^2$ for $\mathcal{M}$-PAM and $1, 2, \ldots, 2N_t^2$ for $\mathcal{M}$-QAM. Further, define the set of indices, $\mathcal{U} \triangleq \{i_1, i_2, \ldots, i_K\}$. The update rule for the K-symbol update can then be written as

$$\mathbf{d}^{(k+1)} = \mathbf{d}^{(k)} + \sum_{j=1}^{K} \lambda_{i_j}^{(k)} \mathbf{e}_{i_j}. \tag{26}$$

For any iteration $k, \mathbf{d}^{(k)}$ belongs to the space $\mathbb{S}$, and therefore $\lambda_{i_j}^{(k)}$ can take only certain integer values. In particular, $\lambda_{i_j}^{(k)} \in \mathbb{A}_{i_j}^{(k)}$, where $\mathbb{A}_{i_j}^{(k)} \triangleq \{x | (x + d_{i_j}^{(k)}) \in \mathbb{A}_{i_j}, x \neq 0\}$. For example, for 16-QAM, $\mathbb{A}_{i_j} = \{-3, -1, 1, 3\}$, and if $d_{i_j}^{(k)}$ is -1, then $\mathbb{A}_{i_j}^{(k)} = \{-2, 2, 4\}$. Using (12), we can write the cost difference function $\Delta C_{\mathcal{U}}^{k+1}(\lambda_{i_1}^{(k)}, \lambda_{i_2}^{(k)}, \ldots, \lambda_{(i_K)}^{(k)}) \triangleq C^{(k+1)} - C^{(k)}$ as

$$\Delta C_{\mathcal{U}}^{k+1}\left(\lambda_{i_1}^{(k)}, \lambda_{i_2}^{(k)}, \ldots, \lambda_{i_K}^{(k)}\right) = \sum_{j=1}^{K} \lambda_{i_j}^{(k)^2} (\mathbf{G})_{i_j, i_j} + 2\sum_{q=1}^{K} \sum_{p=q+1}^{K} \lambda_{i_p}^{(k)} \lambda_{i_q}^{(k)} (\mathbf{G})_{i_p, i_q} - 2\sum_{j=1}^{K} \lambda_{i_j}^{(k)} z_{i_j}^{(k)} \tag{27}$$

where $\lambda_{i_j}^{(k)} \in \mathbb{A}_{i_j}^{(k)}$, which can be compactly written as $(\lambda_{i_1}^{(k)}, \lambda_{i_2}^{(k)}, \ldots, \lambda_{i_K}^{(k)}) \in \mathbb{A}_{\mathcal{U}}^{(k)}$, where $\mathbb{A}_{\mathcal{U}}^{(k)}$ denotes the Cartesian product of $\mathbb{A}_{i_1}^{(k)}, \mathbb{A}_{i_2}^{(k)}$ through to $\mathbb{A}_{i_K}^{(k)}$.

For a given $\mathcal{U}$, in order to decrease the ML cost, we would like to choose the value of the K-tuple $(\lambda_{i_1}^{(k)}, \lambda_{i_2}^{(k)}, \ldots, \lambda_{i_K}^{(k)})$ such that the cost difference given by (27) is negative. If multiple K-tuples exist for which the cost difference is negative, we choose the K-tuple which gives the most negative cost difference.

Unlike for one-symbol update, for K-symbol update we do not have a closed-form expression for $(\lambda_{i_1,\text{opt}}^{(k)}, \lambda_{i_2,\text{opt}}^{(k)}, \ldots, \lambda_{i_K,\text{opt}}^{(k)})$ which minimizes the cost difference over $\mathbb{A}_{\mathcal{U}}^{(k)}$, since the cost difference is a function of K discrete valued variables. Consequently, a brute-force method is to evaluate $\Delta C_{\mathcal{U}}^{k+1}(\lambda_{i_1}^{(k)}, \lambda_{i_2}^{(k)}, \ldots, \lambda_{i_K}^{(k)})$ over all possible values of $(\lambda_{i_1}^{(k)}, \lambda_{i_2}^{(k)}, \ldots, \lambda_{i_K}^{(k)})$. Approximate methods can be adopted to solve this problem using lesser complexity. One method based on zero-forcing is as follows. The cost difference function in (27) can be rewritten as

$$\Delta C_{\mathcal{U}}^{k+1}\left(\lambda_{i_1}^{(k)}, \lambda_{i_2}^{(k)}, \ldots, \lambda_{i_K}^{(k)}\right) = \mathbf{\Lambda}_{\mathcal{U}}^{(k)^T} \mathbf{F}_{\mathcal{U}} \mathbf{\Lambda}_{\mathcal{U}}^{(k)} - 2\mathbf{\Lambda}_{\mathcal{U}}^{(k)^T} \mathbf{z}_{\mathcal{U}}^{(k)}, \tag{28}$$

where $\mathbf{\Lambda}_{\mathcal{U}}^{(k)} \triangleq [\lambda_{i_1}^{(k)} \lambda_{i_2}^{(k)} \cdots \lambda_{i_K}^{(k)}]^T$, $\mathbf{z}_{\mathcal{U}}^{(k)} \triangleq [z_{i_1}^{(k)} z_{i_2}^{(k)} \cdots z_{i_K}^{(k)}]^T$, and $\mathbf{F}_{\mathcal{U}} \in \mathbb{R}^{K \times K}$, where $(\mathbf{F}_{\mathcal{U}})_{p,q} = (\mathbf{G})_{i_p, i_q}$ and $p, q \in \{1, 2, \ldots, K\}$. Since $\Delta C_{\mathcal{U}}^{k+1}(\lambda_{i_1}^{(k)}, \lambda_{i_2}^{(k)}, \ldots, \lambda_{i_K}^{(k)})$ is a strictly convex quadratic function of $\mathbf{\Lambda}_{\mathcal{U}}^{(k)}$ (the Hessian $\mathbf{F}_{\mathcal{U}}$ is positive definite with probability 1), a unique global minima exists, and is given by

$$\tilde{\mathbf{\Lambda}}_{\mathcal{U}}^{(k)} = \mathbf{F}_{\mathcal{U}}^{-1} \mathbf{z}_{\mathcal{U}}^{(k)}. \tag{29}$$

However, the solution given by (29) need not lie in $\mathbb{A}_{\mathcal{U}}^{(k)}$. So, we first round-off the solution as

$$\widehat{\mathbf{\Lambda}}_{\mathcal{U}}^{(k)} = 2\left\lfloor 0.5\tilde{\mathbf{\Lambda}}_{\mathcal{U}}^{(k)} \right\rceil \tag{30}$$

where the operation in (30) is done element-wise, since $\tilde{\mathbf{\Lambda}}_{\mathcal{U}}^{(k)}$ is a vector. Further, let $\widehat{\mathbf{\Lambda}}_{\mathcal{U}}^{(k)} \triangleq [\widehat{\boldsymbol{\lambda}}_{i_1}^{(k)} \widehat{\boldsymbol{\lambda}}_{i_2}^{(k)} \cdots \widehat{\boldsymbol{\lambda}}_{i_K}^{(k)}]^T$. It is still possible that the solution $\widehat{\mathbf{\Lambda}}_{\mathcal{U}}^{(k)}$ in (30) need not lie in $\mathbb{A}_{\mathcal{U}}^{(k)}$. This would result in $d_{i_j}^{(k+1)} \notin \mathbb{A}_{i_j}$ for some j. For example, if $\mathbb{A}_{i_j}$

is $\mathcal{M}$-PAM, then $d_{i_j}^{(k+1)} \notin \mathbb{A}_{i_j}$ if $d_{i_j}^{(k)} + \widehat{\boldsymbol{\lambda}}_{i_j}^{(k)} > (\mathcal{M}-1)$ or $d_{i_j}^{(k)} + \widehat{\boldsymbol{\lambda}}_{i_j}^{(k)} < -(\mathcal{M}-1)$. In such cases, we propose the following adjustment to $\widehat{\boldsymbol{\lambda}}_{i_j}^{(k)}$ for $j = 1, 2, \ldots, K$:

$$\widehat{\lambda}_{i_j}^{(k)} = \begin{cases} (\mathcal{M}-1) - d_{i_j}^{(k)}, & \text{when } \widehat{\lambda}_{i_j}^{(k)} + d_{i_j}^{(k)} > (\mathcal{M}-1) \\ -(\mathcal{M}-1) - d_{i_j}^{(k)}, & \text{when } \widehat{\lambda}_{i_j}^{(k)} + d_{i_j}^{(k)} < -(\mathcal{M}-1). \end{cases} \tag{31}$$

After these adjustments, we are guaranteed that $\widehat{\Lambda}_{\mathcal{U}}^{(k)} \in \mathbb{A}_{\mathcal{U}}^{(k)}$. Therefore, the new cost difference function value is given by $\Delta C_{\mathcal{U}}^{k+1}(\widehat{\lambda}_{i_1}^{(k)}, \widehat{\lambda}_{i_2}^{(k)}, \ldots, \widehat{\lambda}_{i_K}^{(k)})$. It is noted that the complexity of this approximate method does not depend on the size of the set $\mathbb{A}_{\mathcal{U}}^{(k)}$, i.e., it has constant complexity. Through simulations, we have observed that this approximation results in a performance close to that of the brute-force method for $K = 2$ and 3. Defining the optimum $\mathcal{U}$ for the approximate method as $\hat{\mathcal{U}}$, we can write

$$\begin{aligned} \hat{\mathcal{U}} &\triangleq (\hat{i}_1, \hat{i}_2, \ldots, \hat{i}_K) \\ &= \underset{\mathcal{U}}{\operatorname{argmin}} \, \Delta C_{\mathcal{U}}^{k+1}\left(\widehat{\lambda}_{i_1}^{(k)}, \widehat{\lambda}_{i_2}^{(k)}, \ldots, \widehat{\lambda}_{i_K}^{(k)}\right). \end{aligned} \tag{32}$$

The K-update is successful and the update is done only if $\Delta C_{\hat{\mathcal{U}}}^{k+1}(\widehat{\lambda}_{\hat{i}_1}^{(k)}, \widehat{\lambda}_{\hat{i}_2}^{(k)}, \ldots, \widehat{\lambda}_{\hat{i}_K}^{(k)}) < 0$. The update rules for the $\mathbf{z}^{(k)}$ and $\mathbf{d}^{(k)}$ vectors are given by

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} - \sum_{j=1}^{K} \widehat{\lambda}_{\hat{i}_j}^{(k)} \mathbf{g}_{\hat{i}_j}, \tag{33}$$

$$\mathbf{d}^{(k+1)} = \mathbf{d}^{(k)} + \sum_{j=1}^{K} \widehat{\lambda}_{\hat{i}_j}^{(k)} \mathbf{e}_{\hat{i}_j}. \tag{34}$$

D. Computational Complexity of the M-LAS Algorithm

The complexity of the proposed M-LAS algorithm comprises of three components, namely, 1) computation of the initial vector $\mathbf{d}^{(0)}$, 2) computation of $\mathbf{H}^T\mathbf{H}$, and 3) the search operation. Fig. 1 shows the per-symbol complexity plots as a function of $N_t = N_r$ for 4-QAM at an SNR of 6 dB using MMSE initial vector. Two good properties of the STBCs from CDA are useful in achieving low orders of complexity for the computation of $\mathbf{d}^{(0)}$ and $\mathbf{H}^T\mathbf{H}$. They are: 1) the weight matrices $\mathbf{A}_c^{(i)}$'s are *permutation type*, and 2) the $N_t^2 \times N_t^2$ matrix formed with $N_t^2 \times 1$-sized $\mathbf{a}_c^{(i)}$ vectors as columns is a *scaled unitary matrix*. These properties allow the computation of MMSE/ZF initial solution in $O(N_t^3 N_r)$ complexity, i.e., in $O(N_t N_r)$ per-symbol complexity since there are N_t^2 symbols in one STBC matrix. Likewise, the computation of $\mathbf{H}^T\mathbf{H}$ can be done in $O(N_t^3)$ per-symbol complexity.

The average per-symbol complexities of the 1-LAS and 2-LAS search operations are $O(N_t^2)$ and $O(N_t^2 \log N_t)$, respectively, which can be explained as follows. The average search complexity is the complexity of one search stage times the mean number of search stages till the algorithm terminates. For 1-LAS, the number of search stages is always one. There are multiple iterations in the search, and in each iteration all possible $\binom{2N_t^2}{1}$ 1-symbol updates are considered. So, the per-iteration complexity in 1-LAS is $O(N_t^2)$, i.e., $O(1)$ complexity per symbol. Further, the mean number of iterations before the algorithm terminates in 1-LAS was found to be $O(N_t^2)$ through simulations. So, the overall per-symbol complexity of 1-LAS is $O(N_t^2)$. In 2-LAS, the complexity of the two-symbol update dominates over the one-symbol update. Since there are $\binom{2N_t^2}{2}$ possible two-symbol updates, the complexity of one search stage is $O(N_t^4)$, i.e., $O(N_t^2)$ complexity per symbol. The mean number of stages till the algorithm terminates in 2-LAS was found to be $O(\log N_t)$ through simulations. Therefore, the overall per-symbol complexity of 2-LAS is $O(N_t^2 \log N_t)$. These can be observed from Fig. 1, where it can be seen that the per-symbol complexity in the initial vector computation plus the 1-LAS/2-LAS search operation is $O(N_t^2)/O(N_t^2 \log N_t)$; i.e., 1-LAS and 2-LAS complexity plots run parallel to the $c_1 N_t^2$ and $c_2 N_t^2 \log N_t$ lines, respectively. With the computation of $\mathbf{H}^T\mathbf{H}$ included, the complexity order is more than N_t^2. From the slopes of the plots in Fig. 1, we find that the overall complexities for $N_t = 16$ and 32 are proportional to $N_t^{2.5}$ and $N_t^{2.7}$, respectively.

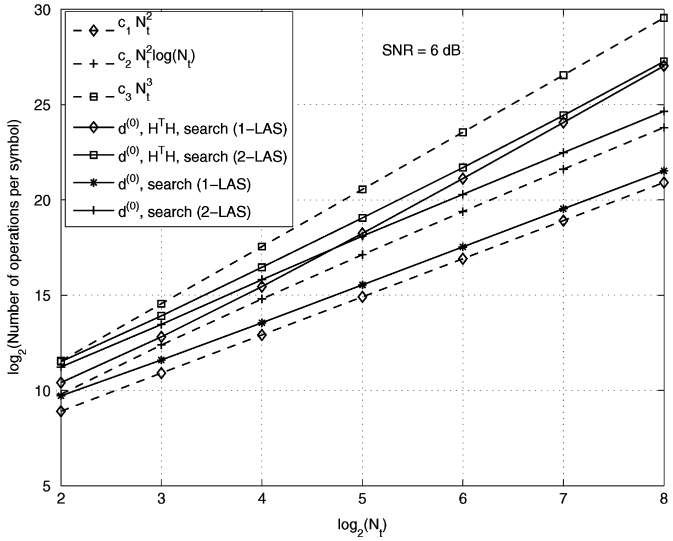

Fig. 1. . Computational complexity of the proposed M-LAS algorithm in decoding non-orthogonal STBCs from CDA. MMSE initial vector, 4-QAM, SNR = 6 dB.

For the special case of ILL-only STBCs (i.e., $\delta = t = 1$), the complexity involved in computing $\mathbf{d}^{(0)}$ and $\mathbf{H}^T\mathbf{H}$ can be reduced further. This becomes possible due to the following property of ILL-only STBCs. Let $\mathbf{V}_a$ be the complex $N_t^2 \times N_t^2$ matrix with $\mathbf{a}_c^{(i)}$ as its ith column. The computation of $\mathbf{d}^{(0)}$ (or $\mathbf{H}^T\mathbf{H}$) involves multiplication of $\mathbf{V}_a^H$ with another vector (or matrix). The columns of $\mathbf{V}_a^H$ can be permuted in such a way that the permuted matrix is block-diagonal, where each block is a $N_t \times N_t$ DFT matrix for $\delta = t = 1$. So, the multiplication of $\mathbf{V}_a^H$ by any vector becomes equivalent to a N_t-point DFT operation, which can be efficiently computed using fast Fourier transform (FFT) in $O(N_t \log N_t)$ complexity. Using this simplification, the per-symbol complexity of computing $\mathbf{H}^T\mathbf{H}$ is reduced from $O(N_t^3)$ to $O(N_t^2 \log N_t)$. Computing $\mathbf{d}^{(0)}$ using MMSE filter involves the computation of $(1)/(N_t)\mathbf{V}_a^H(\mathbf{I} \otimes ((\mathbf{H}_c^H\mathbf{H}_c + (1)/(\gamma N_t)\mathbf{I})^{-1}\mathbf{H}_c^H))\mathbf{y}_c$.

The complexity of computing the vector $(\mathbf{I} \otimes ((\mathbf{H}_c^H\mathbf{H}_c + (1)/(\gamma N_t)\mathbf{I})^{-1}\mathbf{H}_c^H))\mathbf{y}_c$ is $O(N_t^2N_r)$, and the complexity of computing $\mathbf{V}_a^H(\mathbf{I} \otimes ((\mathbf{H}_c^H\mathbf{H}_c + (1)/(\gamma N_t)\mathbf{I})^{-1}\mathbf{H}_c^H))\mathbf{y}_c$ is $O(N_t^3N_r)$. In the case of ILL-only STBC, because of the above-mentioned property, the complexity of computing $\mathbf{V}_a^H(\mathbf{I} \otimes ((\mathbf{H}_c^H\mathbf{H}_c + (1)/(\gamma N_t)\mathbf{I})^{-1}\mathbf{H}_c^H))\mathbf{y}_c$ gets reduced to $O(N_t^2 \log N_t)$ from $O(N_t^3N_r)$. So the total complexity for computing $\mathbf{d}^{(0)}$ in ILL-only STBC is $O(N_t^2N_r) + O(N_t^2 \log N_t)$, which gives a per-symbol complexity of $O(N_r) + O(\log N_t)$. So, the overall per-symbol complexity for 1-LAS detection of ILL-STBCs is $O(N_t^2 \log N_t)$.

E. Generation of Soft Outputs

We propose to generate soft values at the M-LAS output for all the individual bits that constitute the $\mathcal{M}$-PAM/$\mathcal{M}$-QAM symbols as follows. These output values are fed as soft inputs to the decoder in a coded system. Let $\mathbf{d} = [\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_{2N_t^2}]$, $\widehat{x}_i \in \mathbb{A}_i$ denote the detected output symbol vector from the M-LAS algorithm. Let the symbol $\widehat{x}_i$ map to the bit vector $\mathbf{b}_i = [b_{i,1}, b_{i,2}, \ldots, b_{i,K_i}]^T$, where $K_i = \log_2 |\mathbb{A}_i|$, and $b_{i,j} \in \{+1, -1\}$, $i = 1, 2, \ldots, 2N_t^2$, and $j = 1, 2, \ldots, K_i$. Let $\tilde{b}_{i,j} \in \mathbb{R}$ denote the soft value for the jth bit of the ith symbol. Given $\mathbf{d}$, we need to find $\tilde{b}_{i,j}, \forall (i,j)$.

Note that the quantity $\|\mathbf{y} - \mathbf{Hd}\|^2$ is inversely related to the likelihood that $\mathbf{d}$ is indeed the transmitted symbol vector. Let the $\mathbf{d}$ vector with its jth bit of the ith symbol forced to $+1$ be denoted as vector $\mathbf{d}_i^{j+}$. Likewise, let $\mathbf{d}_i^{j-}$ be the vector $\mathbf{d}$ with its jth bit of the ith symbol forced to -1. Then the quantities $\|\mathbf{y} - \mathbf{Hd}_i^{j+}\|^2$ and $\|\mathbf{y} - \mathbf{Hd}_i^{j-}\|^2$ are inversely related to the likelihoods that the jth bit of the ith transmitted symbol is $+1$ and -1, respectively. So, if $\|\mathbf{y} - \mathbf{Hd}_i^{j-}\|^2 - \|\mathbf{y} - \mathbf{Hd}_i^{j+}\|^2$ is $+$ve (or $-$ve), it indicates that the jth bit of the ith transmitted symbol has a higher likelihood of being $+1$ (or -1). So, the quantity $\|\mathbf{y} - \mathbf{Hd}_i^{j-}\|^2 - \|\mathbf{y} - \mathbf{Hd}_i^{j+}\|^2$, appropriately normalized to avoid unbounded increase for increasing N_t, can be a good soft value for the jth bit of the ith symbol. With this motivation, we generate the soft output value for the jth bit of the ith symbol as

$$\tilde{b}_{i,j} = \frac{\left\|\mathbf{y} - \mathbf{Hd}_i^{j-}\right\|^2 - \left\|\mathbf{y} - \mathbf{Hd}_i^{j+}\right\|^2}{\|\mathbf{h}_i\|^2} \tag{35}$$

where the normalization by $\|\mathbf{h}_i\|^2$ is to contain unbounded increase of $\tilde{b}_{i,j}$ for increasing N_t. The RHS in the above can be efficiently computed in terms of $\mathbf{z}$ and $\mathbf{G}$ as follows. Since $\mathbf{d}_i^{j+}$ and $\mathbf{d}_i^{j-}$ differ only in the ith entry, we can write

$$\mathbf{d}_i^{j-} = \mathbf{d}_i^{j+} + \lambda_{i,j}\mathbf{e}_i. \tag{36}$$

Since we know $\mathbf{d}_i^{j-}$ and $\mathbf{d}_i^{j+}$, we know $\lambda_{i,j}$ from (36). Substituting (36) in (35), we can write

$$\begin{aligned}\tilde{b}_{i,j}\|\mathbf{h}_i\|^2 &= \left\|\mathbf{y} - \mathbf{Hd}_i^{j+} - \lambda_{i,j}\mathbf{h}_i\right\|^2 - \left\|\mathbf{y} - \mathbf{Hd}_i^{j+}\right\|^2 \\ &= \lambda_{i,j}^2\|\mathbf{h}_i\|^2 - 2\lambda_{i,j}\mathbf{h}_i^T\left(\mathbf{y} - \mathbf{Hd}_i^{j+}\right) \qquad (37) \\ &= -\lambda_{i,j}^2\|\mathbf{h}_i\|^2 - 2\lambda_{i,j}\mathbf{h}_i^T\left(\mathbf{y} - \mathbf{Hd}_i^{j-}\right). \qquad (38)\end{aligned}$$

If $b_{i,j} = 1$, then $\mathbf{d}_i^{j+} = \mathbf{d}$ and substituting this in (37) and dividing by $\|\mathbf{h}_i\|^2$, we get

$$\tilde{b}_{i,j} = \lambda_{i,j}^2 - 2\lambda_{i,j}\frac{z_i}{(\mathbf{G})_{i,i}}. \tag{39}$$

If $b_{i,j} = -1$, then $\mathbf{d}_i^{j-} = \mathbf{d}$ and substituting this in (38) and dividing by $\|\mathbf{h}_i\|^2$, we get

$$\tilde{b}_{i,j} = -\lambda_{i,j}^2 - 2\lambda_{i,j}\frac{z_i}{(\mathbf{G})_{i,i}}. \tag{40}$$

It is noted that $\mathbf{z}$ and $\mathbf{G}$ are already available upon the termination of the M-LAS algorithm, and hence the complexity of computing $\tilde{b}_{i,j}$ in (39) and (40) is constant. Hence, the overall complexity in computing the soft values for all the bits is $O(N_t \log_2 \mathcal{M})$. We also see from (39) and (40) that the magnitude of $\tilde{b}_{i,j}$ depends upon $\lambda_{i,j}$. For large-size signal sets, the possible values of $\lambda_{i,j}$ will also be large in magnitude. We therefore have to normalize $\tilde{b}_{i,j}$ for the turbo decoder to function properly. It has been observed through simulations that normalizing $\tilde{b}_{i,j}$ by $((\lambda_{i,j})/(2))^2$ resulted in good performance. In [28], we have shown that this soft decision output generation method, when used in large V-BLAST systems, offers about 1 to 1.5 dB improvement in coded BER performance compared to that achieved using hard decision outputs from the M-LAS algorithm. We have observed similar improvements in STBC MIMO systems also. In all coded BER simulations in this paper, we use the soft outputs proposed here as inputs to the decoder.

IV. BER Performance With Perfect CSIR

In this section, we present the uncoded/turbo coded BER performance of the proposed M-LAS detector in decoding non-orthogonal STBCs from CDA, assuming perfect knowledge of CSI at the receiver.[5] In all the BER simulations in this section, we have assumed that the fade remains constant over one STBC matrix duration and varies i.i.d. from one STBC matrix duration to the other. We consider two STBC designs; 1) "FD-ILL" STBCs where $\delta = e^{\sqrt{5}\mathbf{j}}, t = e^{\mathbf{j}}$ in (11a), and 2) "ILL-only" STBCs where $\delta = t = 1$. The SNRs in all the BER performance figures are the average received SNR per received antenna, γ, defined in Section II [3]. We have used MMSE filter as the initial filter in all the simulations.

A. Uncoded BER as a Function of Increasing $N_t = N_r$

In Fig. 2, we plot the uncoded BER performance of the proposed 1-, 2-, and 3-LAS algorithms in decoding ILL-only STBCs (4×4, 8×8, 16×16, 32×32 STBCs) for $N_t = N_r = 4, 8, 16, 32$ and 4-QAM. SISO AWGN performance (without fading) and MMSE-only performance (i.e., without the search using LAS) are also plotted for comparison. It can be seen that MMSE-only performance does not improve with increasing STBC size (i.e., increasing $N_t = N_r$). However, it is interesting to see that, when the proposed search using LAS is performed following the MMSE operation, the performance improves for increasing $N_t = N_r$, illustrating

[5]We will relax this perfect channel knowledge assumption in the next section, where we present an iterative detection/channel estimation scheme for the considered large STBC MIMO system.

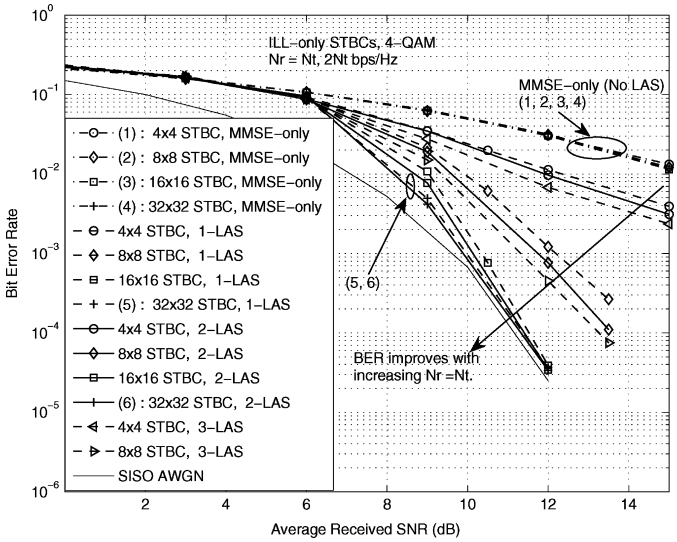

Fig. 2. Uncoded BER of the proposed 1-LAS, 2-LAS and 3-LAS detectors for **ILL-only** STBCs for different $N_t = N_r$. 4-QAM, $2N_t$ bps/Hz. *BER improves as $N_t = N_r$ increases and approaches SISO AWGN performance for large $N_t = N_r$.*

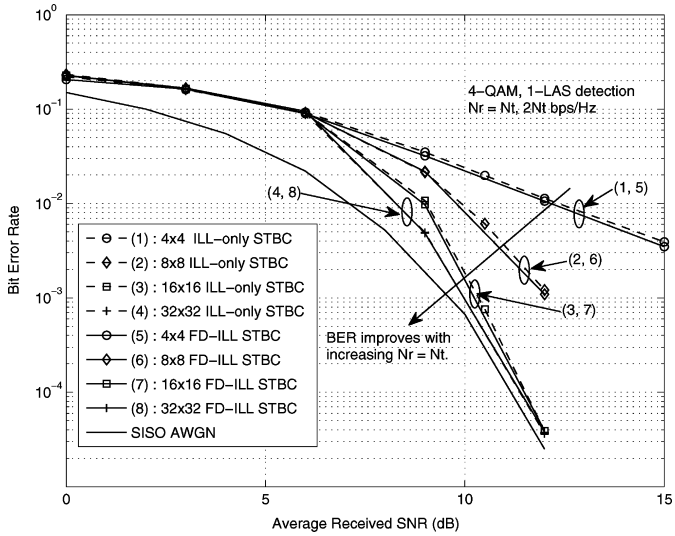

Fig. 3. Uncoded BER comparison between **FD-ILL** and **ILL-only** STBCs for different $N_t = N_r$. 4-QAM, $2N_t$ bps/Hz, 1-LAS detection. *ILL-only STBCs perform almost same as FD-ILL STBCs.*

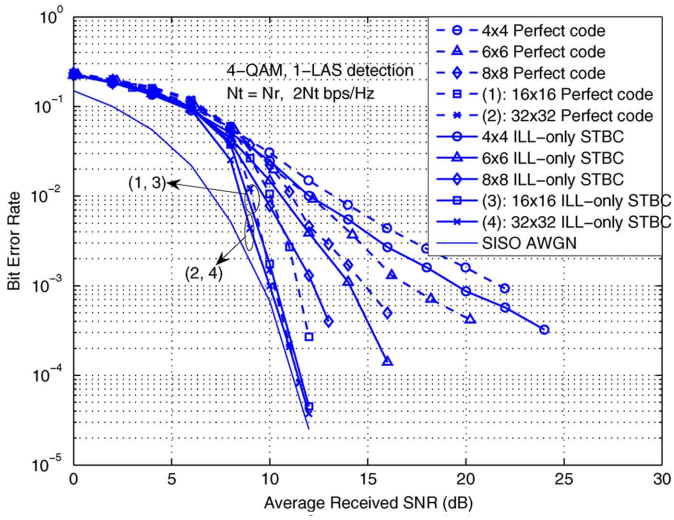

Fig. 4. Uncoded BER comparison between **perfect codes** and **ILL-only** STBCs for different $N_t = N_r$, 4-QAM, $2N_t$ bps/Hz, 1-LAS detection. *For small dimensions* (e.g., 4×4, 6×6, 8×8), *perfect codes with 1-LAS detection perform worse than ILL-only STBCs. For large dimensions (e.g.*, 16×16, 32×32), *ILL-only STBCs and perfect codes perform almost same.*

the performance benefit due to the proposed search strategy. For example, though the LAS detector performs far from SISO AWGN performance for small number of dimensions (e.g., 4×4, 8×8 STBCs with 32 and 128 real dimensions, respectively), its large system behavior at increased number of dimensions (e.g., 16×16 and 32×32 STBCs with 512 and 2048 real dimensions, respectively) effectively renders near SISO AWGN performance; e.g., with $N_t = N_r = 16, 32$, for BERs better than 10^{-3}, the LAS detector performs very close to SISO AWGN performance. We also observe that 3-LAS performs better than 2-LAS for $N_t = N_r = 4, 8$, and 2-LAS performs better than 1-LAS. Since close to SISO AWGN performance is achieved with one-, two-, or three-symbol update itself, the cases of more than three-symbol update, which will result in increased complexity with diminishing returns in performance gain, are not considered in the performance evaluation.

B. Performance of FD-ILL Versus ILL-Only STBCs

In Fig. 3, we present uncoded BER performance comparison between FD-ILL versus ILL-only STBCs for 4-QAM at different $N_t = N_r$ using 1-LAS detection. The BER plots in Fig. 3 illustrate that the performance of ILL-only STBCs with 1-LAS detection for $N_t = N_r = 4, 8, 16, 32$ and 4-QAM are almost as good as those of the corresponding FD-ILL STBCs. A similar closeness between the performance of ILL-only and FD-ILL STBCs is observed in the turbo coded BER performance as well, which is shown in Fig. 8 for a 16×16 STBC with 4-QAM and turbo code rates of 1/3, 1/2 and 3/4. This is an interesting observation, since this suggests that, in such cases, the computational complexity advantage with $\delta = t = 1$ in ILL-only STBCs can be taken advantage of without incurring much performance loss compared to FD-ILL STBCs.

C. Decoding and BER of Perfect Codes of Large Dimensions

While the STBC design in (11a) offers both ILL and FD, *perfect codes*[6] under ML decoding can provide coding gain in addition to ILL and FD [17]–[21]. Decoding of perfect codes has been reported in the literature for only up to five antennas using sphere/lattice decoding [20]. The complexity of these decoders are prohibitive for decoding large-sized perfect codes, although large-sized codes are of interest from a high spectral efficiency view point. We note that, because of its low-complexity attribute, the proposed M-LAS detector is able to decode

[6]We note that the definition of perfect codes differ in [19] and [20]. The perfect codes covered by the definition in [20] includes the perfect codes of [19] as a proper subclass. However, for our purpose of illustrating the performance of the proposed detector in large STBC MIMO systems, we refer to the codes in [19] as well as [20] as perfect codes.

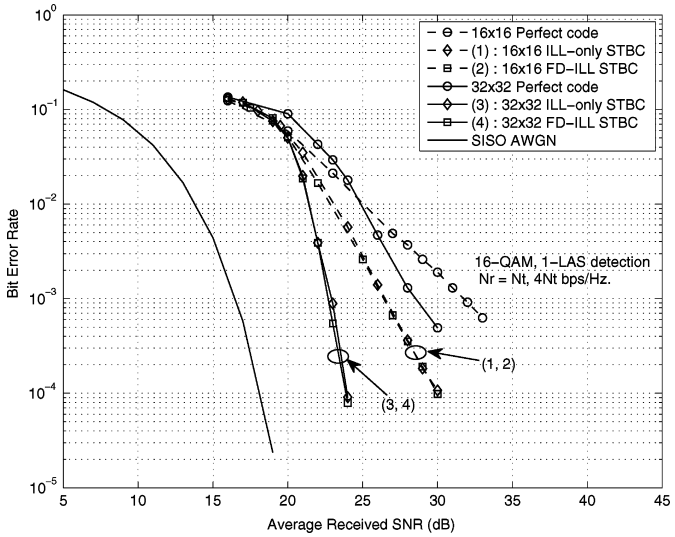

Fig. 5. Uncoded BER comparison between **perfect codes**, **ILL-only**, and **FD-ILL** STBCs for $N_t = N_r = 16, 32$, **16-QAM**, $4N_t$ bps/Hz, 1-LAS detection. *For larger modulation alphabet sizes (e.g., 16 QAM), perfect codes with 1-LAS detection perform poorer than ILL-only and FD-ILL STBCs.*

perfect codes of large dimensions. In Figs. 4 and 5, we present the simulated BER performance of perfect codes in comparison with those of ILL-only and FD-ILL STBCs for up to 32 transmit antennas using 1-LAS detector.

In Fig. 4, we show uncoded BER comparison between perfect codes and ILL-only STBCs for different $N_t = N_r$ and 4-QAM using 1-LAS detection. The 4×4 and 6×6 perfect codes are from [19], and the 8×8, 16×16 and 32×32 perfect codes are from [20]. From Fig. 4, it can be seen that the 1-LAS detector achieves better performance for ILL-only STBCs than for perfect codes, when codes with small number of transmit antennas are considered (e.g., $N_t = 4, 6, 8$). While perfect codes are expected to perform better than ILL-only codes under ML detection for any N_t, we observe the opposite behavior under 1-LAS detection for small N_t (i.e., ILL-only STBCs performing better than perfect codes for small dimensions). This behavior could be attributed to the nature of the LAS detector, which achieves near-optimal performance only when the number of dimensions is large,[7] and it appears that, in the detection process, LAS is more effective in disentangling the symbols in STBCs when $\delta = t = 1$ (i.e., in ILL-only STBCs) than in perfect codes. The performance gap between perfect codes and ILL-only STBCs with 1-LAS detection diminishes for increasing code sizes such that the performance for 32×32 perfect code and ILL-only STBC with 4-QAM are almost same and close to the SISO AWGN performance. In Fig. 5, we show a similar comparison between perfect codes, ILL-only and FD-ILL only STBCs when larger modulation alphabet sizes (e.g., 16-QAM) are used in the case of 16×16 and 32×32 codes. It can be seen that with higher-order QAM like 16-QAM, perfect codes with 1-LAS detection perform poorer than ILL-only and FD-ILL STBCs, and that ILL-only and FD-ILL STBCs perform almost same and close to the SISO AWGN performance. The results in Figs. 4 and 5 suggest that, with 1-LAS detection, owing to the complexity advantage and good performance in using $\delta = t = 1$, ILL-only STBCs can be a good choice for practical large STBC MIMO systems [27], [52].

[7]In [29], we have presented an analytical proof that the bit error performance of 1-LAS detector for V-BLAST with 4-QAM in i.i.d. Rayleigh fading converges to that of the ML detector as $N_t, N_r \to \infty$, keeping $N_t = N_r$.

D. Comparison With Other Large-MIMO Architecture/Detector Combinations

In [30], Choi *et al.* have presented an iterative soft interference cancellation (ISIC) scheme for multiple antenna systems, derived based on maximum *a posteriori* (MAP) criterion. We compared the performance of the ISIC scheme in [30] with that of the proposed 1-LAS algorithm in detecting $4 \times$,4, 8×8 and 16×16 ILL-only STBCs with $N_t = N_r$ and 4-QAM. Fig. 6 shows this performance comparison. In [30], zero-forcing vector was used as the initial vector in the ISIC scheme. However, performance is better with MMSE initial vector. Since we used MMSE initial vector for 1-LAS, we have used MMSE initial vector for the ISIC algorithm as well. Also, in [30], four to five iterations were shown to be good enough for the ISIC algorithm to converge. In our simulations of the ISIC algorithm, we used ten iterations. Two key observations can be made from Fig. 6: 1) like the 1-LAS algorithm, the ISIC algorithm also shows large system behavior (i.e., improved BER for increasing $N_t = N_r$), and 2) the proposed 1-LAS algorithm outperforms the ISIC algorithm by about 3 to 5 dB at 10^{-3} uncoded BER. In addition, the complexity of the ISIC scheme is higher than the proposed scheme (see the complexity comparison in Table I).

Next, we compare the proposed large-MIMO architecture using STBCs from CDA and M-LAS detection with other large-MIMO architectures and associated detectors reported in the literature. Large-MIMO architectures that use stacking of multiple small-sized STBCs and interference cancellation (IC) detectors for these schemes have been investigated in [22], [31], [32]. Here, we compare different architecture/detector combinations, fixing the total number of transmit/receive antennas and spectral efficiency to be same in all the considered combinations. Specifically, we fix $N_t = N_r = 16$ and a spectral efficiency of 32 bps/Hz for all the combinations. We compare the following seven different architecture/detector combinations which use the same $N_t = N_r = 16$ and achieve 32 bps/Hz spectral efficiency (see Table I): 1) proposed scheme using 16×16 ILL-only STBC (rate-16) with 4-QAM and 1-LAS detection, 2) 16×16 ILL-only STBC (rate-16) with 4-QAM and ISIC algorithm in [30] with ten iterations, 3) four 4×4 stacked QOSTBCs (rate-1) with 256-QAM and IC algorithm presented in [22], 4) eight 2×2 stacked Alamouti codes (rate-1) with 16-QAM and IC algorithm in [22], 5) 16×16 V-BLAST scheme (rate-16) with 4-QAM and sphere decoding (SD) algorithm in [53], 6) 16×16 V-BLAST scheme (rate-16) with 4-QAM and ZF-SIC detector, and 7) 16×16 V-BLAST scheme (rate-16) with 4-QAM and ISIC algorithm in [30]. We present the BER performance comparison of these different combinations in Fig. 7. We also obtained the complexity numbers (in number of real operations per bit) from simulations for these different combinations at an uncoded BER of 5×10^{-2}; these numbers are presented in Table I, along with the SNRs at which 5×10^{-2} uncoded BER is achieved. The following interesting observations can be made from Fig. 7 and Table I:

- the proposed scheme [combination 1)] significantly outperforms the stacked architecture/IC detector combinations presented in [22] [combinations 3) and 4)]; e.g., at 5×10^{-2}

TABLE I
COMPLEXITY AND PERFORMANCE COMPARISON OF DIFFERENT LARGE-MIMO ARCHITECTURE/DETECTOR COMBINATIONS, ALL WITH $N_t = N_r = 16$ AND ACHIEVING 32 bps/Hz SPECTRAL EFFICIENCY. *Proposed Scheme Outperforms the Other Considered Architectures/Detectors Both in Terms of Performance as Well as Complexity*

No.	Large-MIMO Architecture/Detector Combinations (fixed $N_t = N_r = 16$ and 32 bps/Hz for all combinations)	Complexity (in # real operations per bit) at 5×10^{-2} uncoded BER	SNR required to achieve 5×10^{-2} uncoded BER (from Fig. 7)
$i)$	**16×16 ILL-only CDA STBC (rate-16), 4-QAM and 1-LAS detection [Proposed scheme]**	$\mathbf{3.473 \times 10^3}$	**6.8 dB**
$ii)$	16×16 ILL-only CDA STBC (rate-16), 4-QAM and ISIC algorithm in [30]	1.187×10^5	11.3 dB
$iii)$	Four 4×4 stacked rate-1 QOSTBCs, 256-QAM and IC algorithm in [22]	5.54×10^6	24 dB
$iv)$	Eight 2×2 stacked rate-1 Alamouti codes, 16-QAM and IC algorithm in [22]	8.719×10^3	17 dB
$v)$	16×16 V-BLAST (rate-16) scheme, 4-QAM and sphere decoding algorithm in [53]	4.66×10^4	7 dB
$vi)$	16×16 V-BLAST (rate-16) scheme, 4-QAM and V-BLAST detector (ZF-SIC)	1.75×10^4	13 dB
$vii)$	16×16 V-BLAST (rate-16) scheme, 4-QAM and ISIC algorithm in [30]	7.883×10^3	10.6 dB

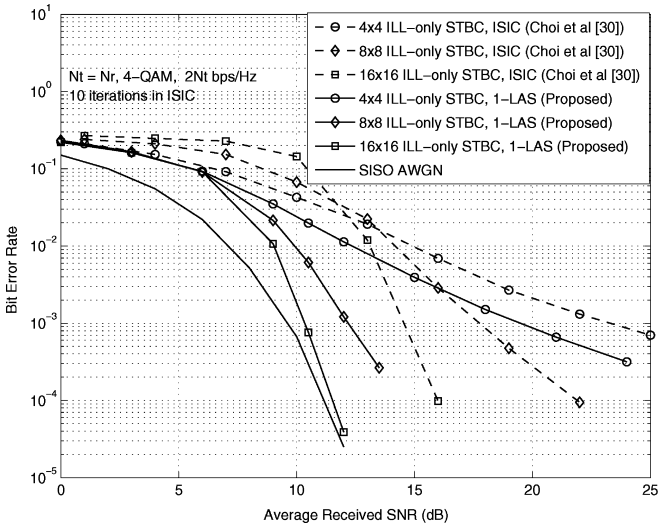

Fig. 6. Uncoded BER comparison between the proposed 1-LAS algorithm and the ISIC algorithm in [30] for **ILL-only** STBCs for different $N_t = N_r$. 4-QAM, $2N_t$ bps/Hz. MMSE initial vectors for both 1-LAS and ISIC. *1-LAS performs significantly better than ISIC in* [30].

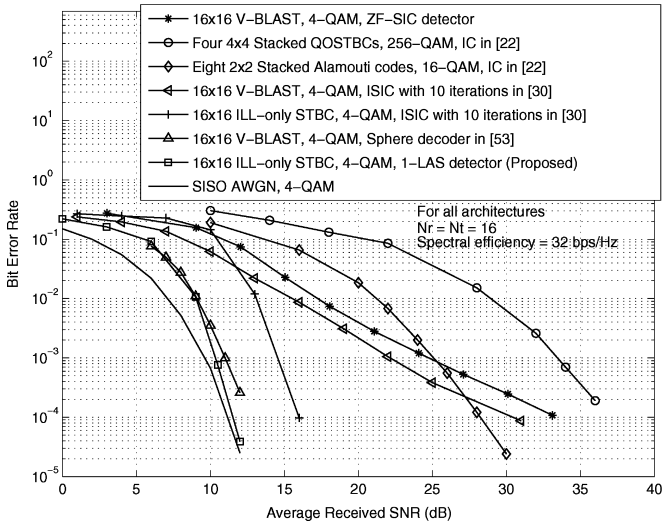

Fig. 7. Uncoded BER comparison between different large-MIMO architecture/detector combinations for given number of transmit/receive antennas ($N_t = N_r = 16$) and spectral efficiency (32 bps/Hz). *Proposed scheme performs better than other architecture/detector combinations considered. It outperforms them in complexity as well (see* Table I).

uncoded BER, the proposed scheme performs better than the stacked architecture/IC in [22] by 17 dB (for four 4×4 QOSTBCs) and 10 dB (for eight 2×2 Alamouti codes). Also, the proposed scheme achieves this significant performance advantage at a much lesser complexity than those of the stacked architecture/IC combinations (see Table I).

- the proposed scheme performs slightly better than the V-BLAST/sphere decoder combination [combination 5)]; 6.8 dB in proposed scheme versus 7 dB in V-BLAST with sphere decoding at 5×10^{-2} uncoded BER. Importantly, the proposed scheme enjoys a significant complexity advantage (by more than an order) over the V-BLAST/sphere decoder combination.
- the ISIC algorithm in [30] applied to ILL-only STBC detection [combination 2)] is inferior to the proposed scheme in both performance (by about 4.5 dB at 5×10^{-2} uncoded BER) as well as complexity (by about two orders).
- the ISIC algorithm in [30] applied to 16×16 V-BLAST detection [combination 7)] is also inferior to the proposed scheme in BER performance (by about 3.8 dB at 5×10^{-2} uncoded BER) as well as complexity (by about a factor of 2).
- comparing the stacked architecture/IC combinations with V-BLAST/ZF-SIC [combination 6)] and V-BLAST/ISIC combinations, we see that although the diversity orders achieved in stacked architecture/IC combinations are high (see their slopes at high SNRs in Fig. 7), V-BLAST with ZF-SIC and ISIC detectors perform much better at low and medium SNRs.

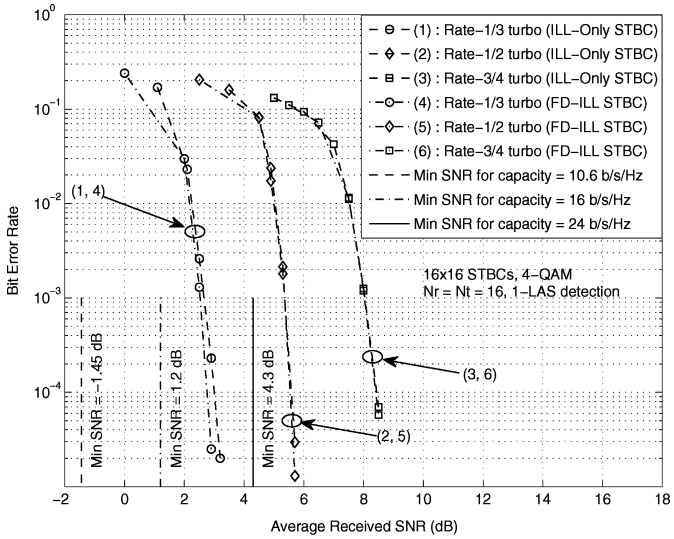

Fig. 8. Turbo coded BER of 1-LAS detector for 16×16 **FD-ILL** and **ILL-only** STBCs. $N_t = N_r = 16$, 4-QAM, turbo code rates: 1/3, 1/2, 3/4 (**10.6, 16, 24 bps/Hz**). *1-LAS detector performs close to within 4 dB from capacity ILL-only STBCs preform as good as FD-ILL STBCs.*

In summary, the proposed scheme outperforms the other considered architecture/detector combinations both in terms of performance as well as complexity.

E. Turbo Coded BER and Nearness-to-Capacity Results

Next, we evaluated the turbo coded BER performance of the proposed scheme. In all the coded BER simulations, we fed the soft outputs presented in Section III-E as input to the turbo decoder. In Fig. 8, we plot the turbo coded BER of the 1-LAS detector in decoding 16×16 FD-ILL and ILL-only STBCs, with $N_t = N_r = 16$, 4-QAM and turbo code rates 1/3 (10.6 bps/Hz), 1/2 (16 bps/Hz), 3/4 (24 bps/Hz). The minimum SNRs required to achieve these capacities in a 16×16 MIMO channel (obtained by evaluating the ergodic capacity expression in [1] through simulation) are also shown. It can be seen that the 1-LAS detector performs close to within just about 4 dB from capacity, which is very good in terms of nearness-to-capacity considering the high spectral efficiencies achieved. It can also be seen that the coded BER performance of FD-ILL and ILL-only STBCs are almost the same for the system parameters considered.

F. Effect of MIMO Spatial Correlation

In generating the BER results in Figs. 2–8, we have assumed i.i.d. fading. However, MIMO propagation conditions witnessed in practice often render the i.i.d. fading model as inadequate. More realistic MIMO channel models that take into account the scattering environment, spatial correlation, etc., have been investigated in the literature [23], [33]. For example, spatial correlation at the transmit and/or receive side can affect the rank structure of the MIMO channel resulting in degraded MIMO capacity [33]. The structure of scattering in the propagation environment can also affect the capacity [23]. Hence, it is of interest to investigate the performance of the M-LAS detector in more realistic MIMO channel models. To this end, we use the non-line-of-sight (NLOS) correlated MIMO channel model proposed by Gesbert *et al.*[8] in [23], and evaluate the effect of spatial correlation on the BER performance of the M-LAS detector [34].

We consider the following parameters[9] in the simulations: $f_c = 5$ GHz, $R = 500$ m, $S = 30$, $D_t = D_r = 20$ m, $\theta_t = \theta_r = 90^\circ$, and $d_t = d_r = 2\lambda/3$. For $f_c = 5$ GHz, $\lambda = 6$ cm and $d_t = d_r = 4$ cm. In [34, Fig. 3], we plot the BER performance of the 1-LAS detector in decoding 16×16 ILL-only STBC with $N_t = N_r = 16$ and 16-QAM. Uncoded BER as well as rate-3/4 turbo coded BER (48-bps/Hz spectral efficiency) for i.i.d. fading as well as correlated fading are shown. In addition, from the MIMO capacity formula in [1], we evaluated the theoretical minimum SNRs required to achieve a capacity of 48 bps/Hz in i.i.d. as well as correlated fading, and plotted them also in [34, Fig. 3]. It is seen that the minimum SNR required to achieve a certain capacity (48 bps/Hz) gets increased for correlated fading compared to i.i.d. fading. From the BER plots in [34, Fig. 3], it can be observed that at an uncoded BER of 10^{-3}, the performance in correlated fading degrades by about 7 dB compared that in i.i.d. fading. Likewise, at a rate-3/4 turbo coded BER of 10^{-4}, a performance loss of about 6 dB is observed in correlated fading compared to that in i.i.d. fading. In terms of nearness to capacity, the vertical fall of the coded BER for i.i.d. fading occurs at about 24-dB SNR, which is about 13 dB away from theoretical minimum required SNR of 11.1 dB. With correlated fading, the detector is observed to perform close to capacity within about 18.5 dB. One way to alleviate such degradation in performance due to spatial correlation can be by providing more number of dimensions at the receive side, which is highlighted in Fig. 9.

Fig. 9 illustrates that the 1-LAS detector can achieve substantial improvement in uncoded as well as coded BER performance in decoding 12×12 ILL-only STBC by increasing N_r beyond N_t for 16-QAM in correlated fading. In the simulations, we have maintained $N_r d_r = 72$ cm and $d_t = d_r$ in both the cases of symmetry (i.e., $N_t = N_r = 12$) as well as asymmetry (i.e., $N_t = 12, N_r = 18$). By comparing the 1-LAS detector performance with $[N_t = N_r = 12]$ versus $[N_t = 12, N_r = 18]$, we observe that the uncoded BER performance with $[N_t = 12, N_r = 18]$ improves by about 17 dB compared to that of $[N_t = N_r = 12]$ at 2×10^{-3} BER. Even the uncoded BER performance with $[N_t = 12, N_r = 18]$ is significantly better than the coded BER performance with $[N_t = N_r = 12]$ by about 11.5 dB at 10^{-3} BER. This improvement is essentially due to the ability of the 1-LAS detector to effectively pick up the additional diversity orders provided by the increased number of receive antennas. With a rate-3/4 turbo code (i.e., 36 bps/Hz), at a coded BER of 10^{-4}, the 1-LAS detector achieves a significant performance improvement of about 13 dB with

[8]Please see [23] for more elaborate details of the spatially correlated MIMO channel model. We note that this model can be appropriate in application scenarios like high data rate wireless IPTV/HDTV distribution using high spectral efficiency large-MIMO links, where large N_t and N_r can be placed at the base station (BS) and customer premises equipment (CPE), respectively.

[9]The parameters used in the model in [23] include: N_t, N_r : # transmit and receive (omni-directional) antennas; d_t, d_r: spacing between antenna elements at the transmit side and at the receive side; R: distance between transmitter and receiver, D_t, D_r: transmit and receive scattering radii; S: number of scatterers on each side; θ_t, θ_r: angular spread at the transmit and receiver sides, and f_c, λ: carrier frequency, wavelength.

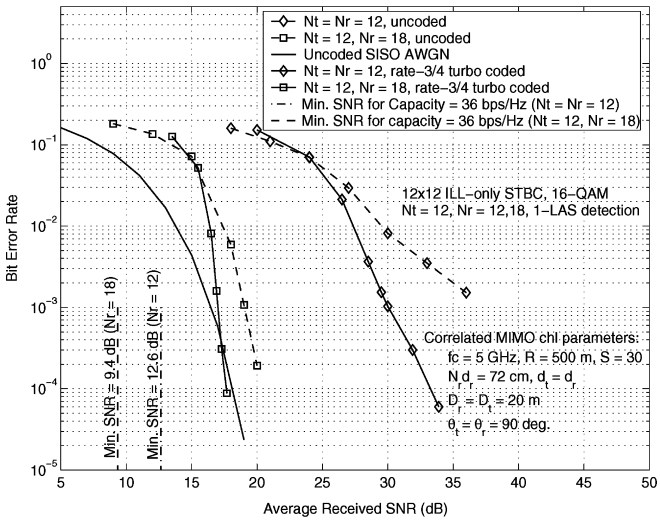

Fig. 9. Effect of $N_r > N_t$ in correlated MIMO fading channel model in [23] keeping $N_r d_r$ constant and $d_t = d_r$. $N_r d_r = 72$ cm, $f_c = 5$ GHz, $R = 500$ m, $S = 30, D_t = D_r = 20$ m, $\theta_t = \theta_r = 90^\circ, 12 \times 12$ **ILL-only** STBC, $N_t = 12, N_r = 12, 18$, **16-QAM**, rate-3/4 turbo code, **36 bps/Hz**. *Increasing # receive dimensions alleviates the loss due to spatial correlation.*

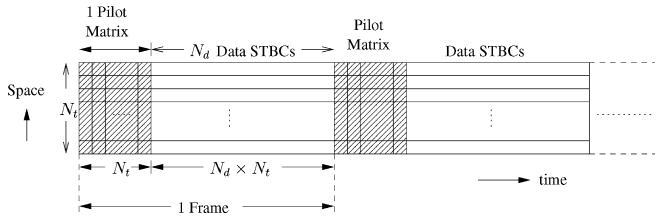

Fig. 10. Transmission scheme with one pilot matrix followed by N_d data STBC matrices in each frame.

$[N_t = 12, N_r = 18]$ compared to that with $[N_t = N_r = 12]$. With $[N_t = 12, N_r = 18]$, the vertical fall of coded BER is such that it is only about 8 dB from the theoretical minimum SNR needed to achieve capacity. This points to the potential for realizing high spectral efficiency multi-gigabit large-MIMO systems that can achieve good performance even in the presence of spatial correlation. We further remark that transmit correlation in MIMO fading can be exploited by using non-isotropic inputs (precoding) based on the knowledge of the channel correlation matrices [35]–[37]. While [35]–[37] propose precoders in conjunction with orthogonal/quasi-orthogonal small MIMO systems in correlated Rayleigh/Ricean fading, design of precoders for large-MIMO systems can be investigated as future work.

V. Iterative Detection/Channel Estimation

In this section, we relax the perfect CSIR assumption made in the previous section, and estimate the channel matrix based on a training-based iterative detection/channel estimation scheme [38]. Training-based schemes, where a pilot signal known to the transmitter and the receiver is sent to get a rough estimate of the channel (training phase) has been studied for STBC MIMO systems in [39]–[42]. Here, we adopt a training-based approach for channel estimation in large STBC MIMO systems. In the considered training-based channel estimation scheme, transmission is carried out in frames, where one $N_t \times N_t$ pilot matrix, $\mathbf{X}_c^{(\mathrm{P})} \in \mathbb{C}^{N_t \times N_t}$, for training purposes, followed by N_d data STBC matrices, $\mathbf{X}_c^{(i)} \in \mathbb{C}^{N_t \times N_t}$, $i = 1, 2, \ldots, N_d$, are sent in each frame as shown in Fig. 10. One frame length, T, (taken to be the channel coherence time) is $T = (N_d+1)N_t$ channel uses. A frame of transmitted pilot and data matrices is of dimension $N_t \times N_t(1 + N_d)$, which can be written as

$$\boldsymbol{\mathcal{X}}_c = \left[\mathbf{X}_c^{(\mathrm{P})}\, \mathbf{X}_c^{(1)}\, \mathbf{X}_c^{(2)}\, \cdots \mathbf{X}_c^{(N_d)}\right]. \tag{41}$$

As in [43], let γ_p and γ_d denote the average SNR during pilot and data phases, respectively, which are related to the average received SNR γ as $\gamma(N_d + 1) = \gamma_p + N_d\gamma_d$. Define $\beta_p \triangleq (\gamma_p)/(\gamma)$, and $\beta_d \triangleq (\gamma_d)/(\gamma)$. Let E_s denote the average energy of the transmitted symbol during the data phase. The average received signal power during the data phase is given by $\mathbb{E}[\mathrm{tr}(\mathbf{X}_c^{(\mathrm{i})}\mathbf{X}_c^{(\mathrm{i})H})] = N_t^2 E_s$, and the average received signal power during the pilot phase is $\mathbb{E}[\mathrm{tr}(\mathbf{X}_c^{(\mathrm{P})}\mathbf{X}_c^{(\mathrm{P})H})] = (N_t^2 E_s \beta_p)/(\beta_d) = \mu N_t$, where $\mu \triangleq (N_t E_s \beta_p)/(\beta_d)$. For optimal training, the pilot matrix should be such that $\mathbf{X}_c^{(\mathrm{P})}\mathbf{X}_c^{(\mathrm{P})H} = \mu \mathbf{I}_{N_t}$ [43]. As in Section II, let $\mathbf{H}_c \in \mathbb{C}^{N_r \times N_t}$ denote the channel matrix, which we want to estimate. We assume block fading, where the channel gains remain constant over one frame consisting of $(1 + N_d)N_t$ channel uses, which can be viewed as the channel coherence time. This assumption can be valid in slow fading fixed wireless applications (e.g., as in possible applications like BS-to-BS backbone connectivity and BS-to-CPE wireless IPTV/HDTV distribution). For this training-based system and channel model, Hassibi and Hochwald presented a lower bound on the capacity in [43]; we will illustrate the nearness of the performance achieved by the proposed iterative detection/estimation scheme to this bound. The received frame is of dimension $N_r \times N_t(1 + N_d)$, and can be written as

$$\boldsymbol{\mathcal{Y}}_c = \left[\mathbf{Y}_c^{(\mathrm{P})}\, \mathbf{Y}_c^{(1)}\, \mathbf{Y}_c^{(2)}\, \cdots\, \mathbf{Y}_c^{(N_d)}\right] = \mathbf{H}_c \boldsymbol{\mathcal{X}}_c + \boldsymbol{\mathcal{N}}_c \tag{42}$$

where $\boldsymbol{\mathcal{N}}_c = [\mathbf{N}_c^{(\mathrm{P})}\ \mathbf{N}_c^{(1)}\ \mathbf{N}_c^{(2)}\ \cdots\ \mathbf{N}_c^{(N_d)}]$ is the $N_r \times N_t(1 + N_d)$ noise matrix and its entries are modeled as i.i.d. $\mathcal{CN}(0, \sigma^2 = (N_t E_s)/(\gamma\beta_d))$. Equation (42) can be decomposed into two parts, namely, the pilot matrix part and the data matrices part, as

$$\mathbf{Y}_c^{(\mathrm{P})} = \mathbf{H}_c \mathbf{X}_c^{(\mathrm{P})} + \mathbf{N}_c^{(\mathrm{P})}, \tag{43}$$

$$\begin{aligned}\mathbf{Y}_c^{(\mathrm{D})} &= \left[\mathbf{Y}_c^{(1)}\, \mathbf{Y}_c^{(2)}\, \cdots\, \mathbf{Y}_c^{(N_d)}\right] \\ &= \mathbf{H}_c \left[\mathbf{X}_c^{(1)}\, \mathbf{X}_c^{(2)}\, \cdots\, \mathbf{X}_c^{(N_d)}\right] \\ &\quad + \left[\mathbf{N}_c^{(1)}\, \mathbf{N}_c^{(2)}\, \cdots\, \mathbf{N}_c^{(N_d)}\right].\end{aligned} \tag{44}$$

A. MMSE Estimation Scheme

A straight-forward way to achieve detection of data symbols with estimated channel coefficients is as follows:

1) Estimate the channel gains via an *MMSE estimator* from the signal received during the first N_t channel uses (i.e.,

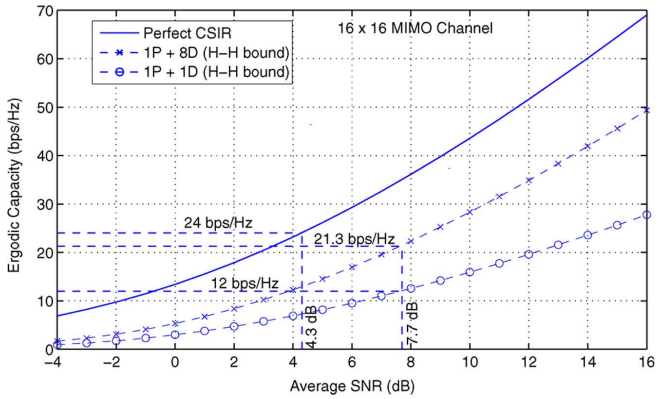

Fig. 11. Hassibi–Hochwald (H-H) capacity bound for 1P + 8D ($T = 144, \tau = 16, \beta_p = \beta_d = 1$) and 1P + 1D ($T = 32, \tau = 16, \beta_p = \beta_d = 1$) training for a 16 × 16 MIMO channel. Perfect CSIR capacity is also shown.

during pilot transmission); i.e., given $\mathbf{Y}_c^{(\mathrm{P})}$ and $\mathbf{X}_c^{(\mathrm{P})}$, an estimate of the channel matrix $\mathbf{H}_c$ is found as

$$\mathbf{H}_c^{\mathrm{est}} = \mathbf{Y}_c^{(\mathrm{P})} \left(\mathbf{X}_c^{(\mathrm{P})}\right)^H \left[\sigma^2 \mathbf{I}_{N_t} + \mathbf{X}_c^{(\mathrm{P})} \left(\mathbf{X}_c^{(\mathrm{P})}\right)^H\right]^{-1}. \tag{45}$$

2) Use the above $\mathbf{H}_c^{\mathrm{est}}$ in place of $\mathbf{H}_c$ in the LAS algorithm (as described in Sections II and III) and detect the transmitted data symbols.

We refer to the above scheme as the *"MMSE estimation scheme."* In the absence of the knowledge of σ^2, a zero-forcing estimate can be obtained at the cost of some performance loss compared to the MMSE estimate. The performance of the estimator can be improved by using a cyclic minimization technique for minimizing the ML metric [44].

B. Proposed Iterative Detection/Estimation Scheme

Techniques that employ iterations between channel estimation and detection can offer improved performance. Iterative receiver algorithms are attractive to achieve a good tradeoff between performance and complexity [45]–[51]. In [45]–[47], receivers that iterate between channel estimation, multiuser detection and channel decoding in coded CDMA systems are presented. Similar iterative techniques in the context of MIMO and MIMO-OFDM systems are presented in [48]–[51]. Here, we propose an iterative scheme, where we iterate between channel estimation and detection in the considered large STBC MIMO system. The proposed scheme works as follows.

1) Obtain an initial estimate of the channel matrix using the MMSE estimator in (45) from the pilot part.
2) Using the estimated channel matrix, detect the data STBC matrices $\mathbf{X}_c^{(i)}, i = 1, 2, \ldots, N_d$ using the LAS detector. Substituting these detected STBC matrices into (41), form $\boldsymbol{\mathcal{X}}_c^{\mathrm{est}}$.
3) Re-estimate the channel matrix using $\boldsymbol{\mathcal{X}}_c^{\mathrm{est}}$ from the previous step, via

$$\mathbf{H}_c^{\mathrm{est}} = \boldsymbol{\mathcal{Y}}_c \left(\boldsymbol{\mathcal{X}}_c^{\mathrm{est}}\right)^H \left[\sigma^2 \mathbf{I}_{N_t} + \boldsymbol{\mathcal{X}}_c^{\mathrm{est}} \left(\boldsymbol{\mathcal{X}}_c^{\mathrm{est}}\right)^H\right]^{-1}. \tag{46}$$

4) Iterate steps 2 and 3 for a specified number of iterations.

The total complexity of obtaining the MMSE estimate of the channel matrix $\mathbf{H}_c^{\mathrm{est}}$ in (45) and (46) is $O(N_t^2 N_r) + O(N_t^3)$, which is less than the total complexity of 1-LAS detection of $O(N_t^4 \log N_t)$ for ILL-only STBCs.

C. BER Performance With Estimated CSIR

We evaluated the BER performance of the 1-LAS detector using estimated CSIR, where we estimate the channel gain matrix through the training-based estimation schemes described in the previous two subsections. We consider the BER performance under three scenarios, namely, 1) under perfect CSIR, 2) under CSIR estimated using the MMSE estimation scheme in Section V-A, and 3) under CSIR estimated using the iterative detection/estimation scheme in Section V-B. In the case of estimated CSIR, we show plots for 1P+N_dD training, where by 1P+N_dD training we mean a training scheme with a frame size of $1 + N_d$ matrices, with one pilot matrix followed N_d data STBC matrices from CDA. For this 1P + N_dD training scheme, a lower bound on the capacity is given by [43]

$$C \geq \frac{T - \tau}{T} \times \mathbb{E}\left[\mathrm{logdet}\left(\mathbf{I}_{N_t} + \frac{\gamma^2 \beta_d \beta_p \tau}{N_t(1 + \gamma\beta_d) + \gamma\beta_p\tau} \frac{\hat{\mathbf{H}}_c \hat{\mathbf{H}}_c^H}{N_t \sigma^2_{\hat{\mathbf{H}}_c}}\right)\right] \tag{47}$$

where T and τ, respectively, are the frame size (i.e., channel coherence time) and pilot duration in number of channel uses, and $\sigma^2_{\hat{\mathbf{H}}_c} = (1)/(N_t N_r)\ \mathbb{E}[\mathrm{tr}\{\hat{\mathbf{H}}_c \hat{\mathbf{H}}_c^H\}]$, where $\hat{\mathbf{H}}_c = \mathbb{E}[\mathbf{H}_c \,|\, \mathbf{X}_c^{(\mathrm{P})}, \mathbf{Y}_c^{(\mathrm{P})}]$ is the MMSE estimate of the channel gain matrix. We computed the capacity bound in (47) through simulations for 1P + 8D and 1P + 1D training for a 16 × 16 MIMO channel. For 1P + 8D training $T = (1 + 8)16 = 144$, $\tau = 16$, and for 1P + 1D training $T = (1 + 1)16 = 32, \tau = 16$. In computing the bounds (shown in Fig. 11) and in BER simulations (in Figs. 12 and 13), we have used $\beta_p = \beta_d = 1$. In Fig. 11, we plot the computed capacity bounds, along with the capacity under perfect CSIR [1]. We obtain the minimum SNR for a given capacity bound in (47) from the plots in Fig. 11, and show (later in Fig. 11) the nearness of the coded BER of the proposed scheme to this SNR limit. We note that improved capacity and BER performance can be achieved if optimum pilot/data power allocation derived in [43] is used instead of the allocation used in Figs. 11 to 13 (i.e., $\beta_p = \beta_d = 1$). We have used the optimum power allocation in [43] for generating the BER plots in Figs. 14 and 15. In all the BER simulations with training, $\sqrt{\mu}\mathbf{I}_{N_t}$ is used as the pilot matrix. ILL-only STBCs and 1-LAS detection are used.

First, in Fig. 12, we plot the uncoded BER performance of 1-LAS detector when 1P + 1D and 1P + 8D training are used for channel estimation in a 16 × 16 STBC MIMO system with $N_t = N_r = 16$ and 4-QAM. BER performance with perfect CSIR is also plotted for comparison. From Fig. 12, it can be observed that, as expected, the BER degrades with estimated CSIR compared to that with perfect CSIR. With MMSE estimation scheme, the performance with 1P + 1D and 1P + 8D are same because of the one-shot estimation. Also, with 1P+1D training, both the MMSE estimation scheme as well as the iterative detection/estimation scheme (with four iterations between detection and estimation) perform almost the same, which is about 3 dB

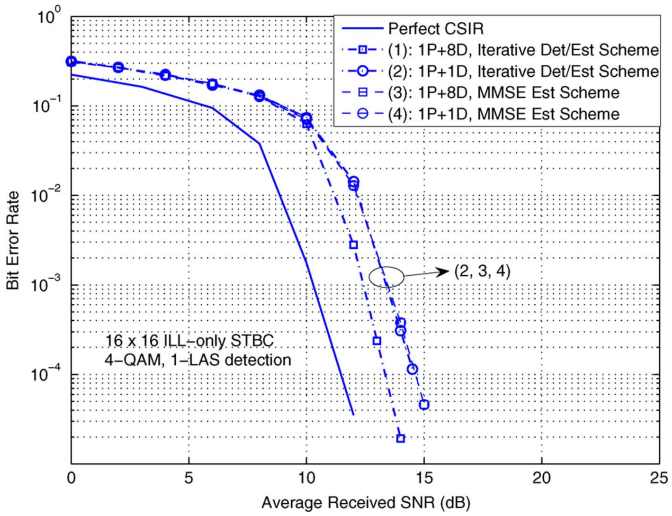

Fig. 12. Uncoded BER of 1-LAS detector for 16×16 **ILL-only** STBC with 1) perfect CSIR, 2) CSIR using MMSE estimation scheme, and iii) CSIR using iterative detection/channel estimation scheme (four iterations). $N_t = N_r = 16$, 4-QAM, $1\mathrm{P} + 1\mathrm{D}$ ($T = 32, \tau = 16, \beta_p = \beta_d = 1$) and $1\mathrm{P} + 8\mathrm{D}$ ($T = 144, \tau = 16, \beta_p = \beta_d = 1$) training.

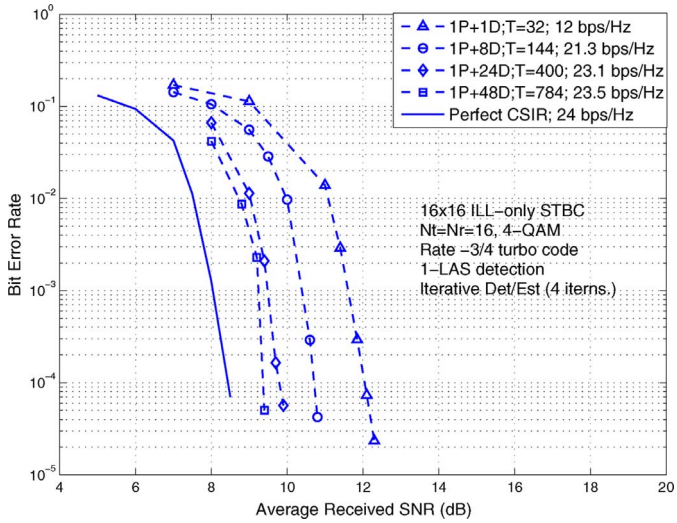

Fig. 14. Turbo coded BER performance of 1-LAS detection and iterative estimation/detection as a function of coherence time, $T = 32, 144, 400, 784$, for a given $N_t = N_r = 16$, 16×16 **ILL-only** STBC, 4-QAM, rate-3/4 turbo code. *Spectral efficiency and BER performance with estimated CSIR approaches to those with perfect CSIR in slow fading (i.e., large T).*

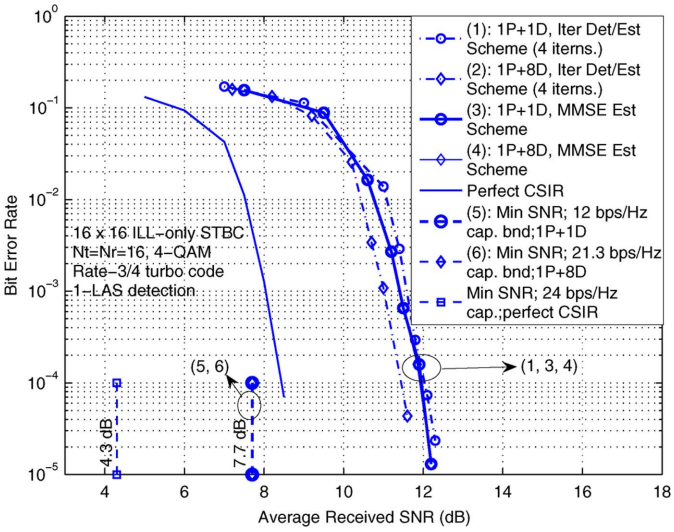

Fig. 13. Turbo coded BER performance of 1-LAS detector for 16 $16\times$ 16 **ILL-only** STBC with *i*) perfect CSIR, *ii*) CSIR using MMSE estimation, and iii) CSIR using iterative detection/channel estimation (four iterations). $N_t = N_r = 16$, 4-QAM, rate-3/4 turbo code, $1\mathrm{P} + 1\mathrm{D}$ ($T = 32, \tau = 16, \beta_p = \beta_d = 1$) and $1\mathrm{P} + 8\mathrm{D}$ ($T = 144, \tau = 16, \beta_p = \beta_d = 1$) training.

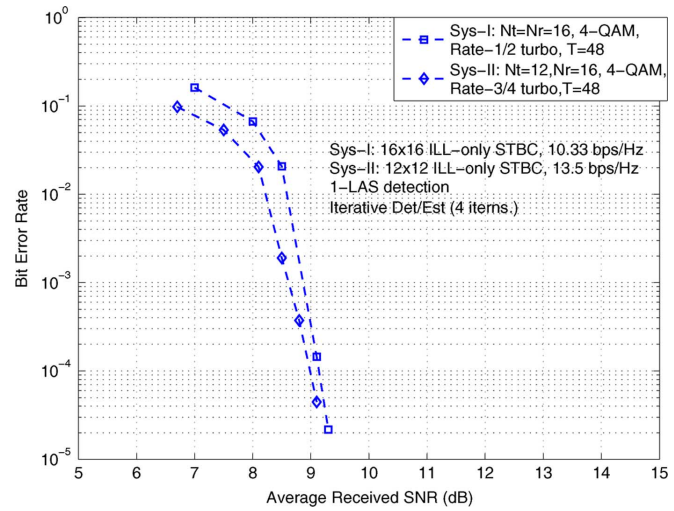

Fig. 15. Comparison between two $1\mathrm{P} + N_d\mathrm{D}$ training-based systems, one with a larger N_t than the other for a given N_r and $T$with $N_r = 16, T = 48$ *and optimum power allocation in both systems, System-II with $N_t = 12$ achieves a higher spectral efficiency (13.5 versus 10.33 bps/Hz) while achieving 10^{-3} coded* BER *at a lesser* SNR (8.6 *versus* 8.9 dB) *than System-I with $N_t = 16$.*

worse compared to that of perfect CSIR at an uncoded BER of 10^{-3}. This indicates that with $1\mathrm{P} + N_d\mathrm{D}$ training, iteration between detection and estimation does not improve performance much over the non-iterative scheme (i.e., the MMSE estimation scheme) for small N_d. With large N_d (e.g., slow fading), however, the iterative scheme outperforms the non-iterative scheme; e.g., with $1\mathrm{P} + 8\mathrm{D}$ training, the performance of the iterative detection/estimation improves by about 1 dB compared to the MMSE estimation.

Next, in Fig. 13, we present the rate-3/4 turbo coded BER of 1-LAS detector using estimated CSIR for the cases of $1\mathrm{P} + 8\mathrm{D}$ and $1\mathrm{P} + 1\mathrm{D}$ training. From Fig. 13, it can be seen that, compared to that of perfect CSIR, the estimated CSIR performance is worse by about 3 dB in terms of coded BER for $1\mathrm{P} + 8\mathrm{D}$ training. With MMSE estimation scheme, 10^{-4} coded BER occurs at about $12 - 7.7 = 4.3$ dB away from the capacity bound for $1\mathrm{P} + 1\mathrm{D}$ and $1\mathrm{P} + 8\mathrm{D}$ training. This nearness to capacity bound improves by about 0.6 dB for the iterative detection/estimation scheme. We note that for the system in Fig. 13 with parameters 16×16 STBC, 4-QAM, rate-3/4 turbo code, and $1\mathrm{P} + 8\mathrm{D}$ training with $T = 144, \tau = 16$, we achieve a high spectral efficiency of $16 \times 2 \times (3)/(4) \times (8)/(9) = 21.3$ bps/Hz even after accounting for the overheads involved in channel estimation (i.e., pilot matrix) and channel coding, while achieving good near-capacity performance at low complexity. This points to the suitability of the proposed approach of using LAS detection along with iterative detection/estimation in practical implementation of large STBC MIMO systems.

Finally, in Fig. 14, we illustrate the coded BER performance of 1-LAS detection and iterative detection/estimation scheme for different coherence times, T, for a fixed $N_t = N_r = 16$, 16×16 STBC, 4-QAM, and rate-3/4 turbo code. The various values of T considered and the corresponding spectral efficiencies are 1) $T = 32$, $1\mathrm{P} + 1D$, 12 bps/Hz, 2) $T = 144$, $1\mathrm{P} + 8\mathrm{D}$, 21.3 bps/Hz, 3) $T = 400$, $1\mathrm{P} + 24D$, 23.1 bps/Hz, and 4) $T = 784$, $1\mathrm{P} + 48\mathrm{D}$, 23.5 bps/Hz. In all these cases, the corresponding optimum pilot/data power allocations in [43] are used. From Fig. 14, it can be seen that for these four cases, 10^{-4} coded BER occurs at around 12 dB, 10.6 dB, 9.7 dB, and 9.4 dB, respectively. The 10^{-4} coded BER for perfect CSIR happens at around 8.5 dB. This indicates that the performance with estimated CSIR improves as T is increased, and that a performance loss of less than 1 dB compared to perfect CSIR can be achieved with large T (i.e., slow fading). For example, with $1\mathrm{P} + 48\mathrm{D}$ training ($T = 784$), the performance with estimated CSIR gets close to that with perfect CSIR both in terms of spectral efficiency (23.5 vs 24 bps/Hz) as well as SNR at which 10^{-4} coded BER occurs (8.5 vs 9.4 dB). This is expected, since the channel estimation becomes increasingly accurate in slow fading (large coherent times) while incurring only a small loss in spectral efficiency due to pilot matrix overhead. This result is significant because T is typically large in fixed/low-mobility wireless applications, and the proposed system can effectively achieve high spectral efficiencies as well as good performance in such applications.

D. On Optimum N_t for A Given N_r and T

In [43], through theoretical capacity bounds it has been shown that, for a given N_r, T and SNR, there is an optimum value of N_t that maximizes the capacity bound (refer Figs. 5 and 6 in [43], where the optimum N_t is shown to be greater than N_r in Fig. 5 and less than N_r in Fig. 6). For example, for $N_r = 16, T = 48$, and SNR $= 10$ dB, the capacity bound evaluated using (47) with optimum power allocation for $N_t = 12$ is 19.73 bps/Hz, whereas for $N_t = 16$ the capacity bound reduces to 17.53 bps/Hz showing that the optimum N_t in this case will be less than N_r. We demonstrate such an observation in practical systems by comparing the simulated coded BER performance of two systems, referred to as System-I and System-II, using 1-LAS detection and iterative detection/estimation scheme. The parameters of System-I and System-II are listed in Table II. N_r and T are fixed at 16 and 48, respectively, in both systems. System-I uses 16 transmit antennas and 16×16 STBC, whereas System-II uses 12 transmit antennas and 12×12 STBC. Since the pilot matrix is $\sqrt{\mu}\mathbf{I}_{N_t}$, the pilot duration τ is 16 and 12, respectively, for System-I and System-II. Optimum pilot/data power allocation and 4-QAM modulation are employed in both systems. System-I uses rate-1/2 turbo code and system-II uses rate-3/4 turbo code. With the above system parameters, the spectral efficiency achieved in System-I is $16 \times 2 \times (1)/(2) \times (2)/(3) = 10.33$ bps/Hz, whereas System-II achieves a higher spectral efficiency of $12 \times 2 \times (3)/(4) \times (3)/(4) = 13.5$ bps/Hz. In Fig. 15, we plot the coded BER of both these systems using 1-LAS detection and iterative detection/estimation. From the simulation points shown in Fig. 15, it can be observed that System-II with a smaller N_t and higher spectral efficiency in fact achieves a certain coded BER performance at a lesser SNR compared to System-I. For example, to achieve 10^{-3} coded BER, System-I requires an SNR of about 8.9 dB, whereas System-II requires only 8.6 dB. This implies that because of the reduction of throughput due to pilot symbols (by a factor of $(T - \tau)/(T)$ for a given T and $\tau = N_t$), a larger N_t does not necessarily mean a higher spectral efficiency. Such an observation has also been made in [43] based on theoretical capacity bounds. The proposed detection/channel estimation scheme allows the prediction of such behavior through simulations, which, in turn, allows system designers to find optimum N_t and STBC size to achieve a certain spectral efficiency in large STBC MIMO systems.

TABLE II
ON OPTIMUM N_t FOR A GIVEN N_r AND T. SYSTEM-II WITH A SMALLER N_t ACHIEVES A HIGHER SPECTRAL EFFICIENCY WHILE ACHIEVING 10^{-3} CODED BER AT A LESSER SNR THAN SYSTEM-I WITH A LARGER N_t

Parameters	System-I	System-II
# Rx antennas, N_r	16	16
Coherence time, T	48	48
# Tx antennas, N_t	**16**	**12**
STBC from CDA	16×16	12×12
Pilot duration, τ	16	12
Training	1P+2D	1P+3D
β_p^{opt}	1.2426	1.4641
β_d^{opt}	0.8786	0.8453
Modulation	4-QAM	4-QAM
Turbo code rate	$1/2$	$3/4$
Spectral efficiency	**10.33** bps/Hz	**13.5** bps/Hz
SNR at 10^{-3} coded BER	**8.9** dB	**8.6** dB

VI. CONCLUSION

We presented a low-complexity algorithm for the detection of high-rate, non-orthogonal STBC large-MIMO systems with tens of antennas that achieve high spectral efficiencies of the order of several tens of bps/Hz. We also presented a training-based iterative detection/channel estimation scheme for such large STBC MIMO systems. Our simulation results showed that the proposed 1-LAS detector along with the proposed iterative detection/channel estimation scheme achieved very good performance at low complexities. With the feasibility of low-complexity high-performance receivers, like the proposed detection/channel estimation scheme, large-MIMO systems with tens of antennas at high spectral efficiencies can become practical, enabling interesting high data rate wireless applications (e.g., wireless IPTV/HDTV distribution). This can motivate the inclusion of large-MIMO architectures (e.g., 12×12, 16×16, 24×24, 32×32 MIMO systems, including those using STBCs from CDA) into wireless standards like IEEE 802.11n/VHT and IEEE 802.16/LTE-A in their evolution to achieve high data rates at increased spectral efficiencies.

APPENDIX

Theorem 1: The $l_p^{(k)}$ in (21) minimizes $\mathcal{F}(l_p^{(k)})$ in (19) and this minimum value is non-positive.

Proof: Let $r \triangleq \lfloor (|z_p^{(k)}|)/(2a_p) \rfloor$. Then $(|z_p^{(k)}|)/(2a_p) = r + f$, where $0 \leq f < 1$, and so we can write

$$\frac{\left|z_p^{(k)}\right|}{a_p} = 2r + 2f. \tag{48}$$

If $l_p^{(k)}$ were unconstrained to be any real number, then the optimal value of $l_p^{(k)}$ is $(|z_p^{(k)}|)/(a_p)$, which would lie between $2r$ and $2r+2$ (as per (48)). Since $\mathcal{F}(l_p^{(k)})$ is quadratic in $l_p^{(k)}$, it is unimodular, and hence the optimal point (with $l_p^{(k)}$ constrained) would be either $2r$ or $2r+2$. Using (19) and (48), we can evaluate $\mathcal{F}(2r+2) - \mathcal{F}(2r)$ to be

$$\mathcal{F}(2r+2) - \mathcal{F}(2r) = 4a_p(1-2f). \tag{49}$$

Since a_p is a positive quantity, the sign of $\mathcal{F}(2r+2) - \mathcal{F}(2r)$ depends upon the sign of $(1-2f)$. If $f \geq 0.5$, then $\mathcal{F}(2r+2) \leq \mathcal{F}(2r)$, and therefore $2r+2$ is the optimal value of $l_p^{(k)}$. Similarly, when $f < 0.5, 2r$ is the optimal value of $l_p^{(k)}$. Therefore, it follows that indeed the rounding solution given by (21) is optimal. $\mathcal{F}(l_p^{(k)})$ is non-positive for all values of $l_p^{(k)}$ between zero and $(2|z_p^{(k)}|)/(a_p)$. If $f < 0.5$, then $2r$ is optimal, and, from (48), we know that $2r \leq (|z_p^{(k)}|)/(a_p)$, and therefore $2r < 2(|z_p^{(k)}|)/(a_p)$. Hence, $\mathcal{F}(2r) = \mathcal{F}^{(\mathrm{opt})}$ is non-positive. Similarly, if $f \geq 0.5$, then $2r+2$ is optimal, and $\mathcal{F}(2r+2) \leq \mathcal{F}(2r)$. However, since $2r$ is always less than $2(|z_p^{(k)}|)/(a_p)$, $\mathcal{F}(2r)$ is non-positive and therefore $\mathcal{F}(2r+2) = \mathcal{F}^{(\mathrm{opt})}$ is non-positive. □

ACKNOWLEDGMENT

The authors would like to thank the Editor, Prof. R. Calderbank, for handling the review process. We would like to thank the anonymous reviewers for their critical and useful comments, and for motivating us to compare the performance and complexity of the proposed scheme with those of other large-MIMO architectures/detectors.

REFERENCES

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

[2] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space–Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[3] H. Jafarkhani, *Space–Time Coding: Theory and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

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

[5] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[6] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space–time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

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

[8] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1639–1242, Jul. 1999.

[9] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2401, Oct. 2003.

[10] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2806–2818, Aug. 2005.

[11] L. Azzam and E. Ayanoglu, "Reduced complexity sphere decoding for square QAM via a new lattice representation," May 16, 2007, arXiv:0705.2435v1 [cs.IT].

[12] X. Yang, Y. Xiong, and F. Wang, "An adaptive MIMO system based on unified belief propagation detection," in *Proc. IEEE ICC'2007*, Jun. 2007, pp. 4156–4161.

[13] B. Farhang-Boroujeny, H. Zhu, and Z. Shi, "Markov chain Monte Carlo algorithms for CDMA and MIMO communication systems," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1896–1908, May 2006.

[14] Y. Sun, "A family of linear complexity likelihood ascent search detectors for CDMA multiuser detection," in *Proc. IEEE Int. Symp. Spread Spectrum Tech. Applicat.*, Sep. 2000, pp. 713–717.

[15] K. V. Vardhan, S. K. Mohammed, A. Chockalingam, and B. S. Rajan, "A low-complexity detector for large MIMO systems and multicarrier CDMA systems," *IEEE JSAC Special Iss. Multiuser Detection, Adv. Commun. Syst. Netw.*, vol. 26, no. 3, pp. 473–485, Apr. 2008.

[16] S. K. Mohammed, K. V. Vardhan, A. Chockalingam, and B. S. Rajan, "Large MIMO systems: A low-complexity detector at high spectral efficiencies," in *Proc. IEEE ICC'2008*, May 2008, pp. 3839–3945.

[17] 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. Inf. Theory*, vol. 51, no. 4, pp. 1432–1436, Apr. 2005.

[18] P. Dayal and M. K. Varanasi, "An optimal two transmit antenna space–time code and its stacked extensions," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2003.

[19] F. E. Oggier, G. Rekaya, J.-C. Belfiore, and E. Viterbo, "Perfect space–time block codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3885–3902, Sep. 2006.

[20] P. Elia, B. A. Sethuraman, and P. V. Kumar, "Perfect space–time codes for any number of antennas," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 3853–3868, Nov. 2007.

[21] F. Oggier, J.-C. Belfiore, and E. Viterbo, "Cyclic division algebras: A tool for space–time coding," *Foundations and Trends in Commun. Inf. Theory*, vol. 4, no. 1, pp. 1–95, 2007, Now Publishers.

[22] J. Kazemitabar and H. Jafarkhani, "Multiuser interference cancellation and detection for users with more than two transmit antennas," *IEEE Trans. Commun.*, vol. 56, no. 4, pp. 574–583, Apr. 2008.

[23] D. Gesbert, H. Bölcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[24] G. H. Golub and C. F. Vanloan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1989.

[25] S. Verdu, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[26] S. K. Mohammed, A. Chockalingam, and B. S. Rajan, "High-rate space–time coded large MIMO systems: Low-complexity detection and performance," in *Proc. IEEE GLOBECOM'2008*, Nov.–Dec. 2008, pp. 1–5.

[27] B. Hassibi and B. Hochwald, "High rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1804–1824, Jul. 2002.

[28] S. K. Mohammed, A. Chockalingam, and B. S. Rajan, "A low-complexity near-ML performance achieving algorithm for large MIMO detection," in *Proc. IEEE ISIT'2008*, Jul. 2008, pp. 2012–2016.

[29] S. K. Mohammed, A. Chockalingam, and B. S. Rajan, "Asymptotic analysis of the performance of LAS algorithm for large MIMO detection," Jun. 16, 2008, Online arXiv:0806.2533v1 [cs.IT].

[30] W. J. Choi, K. W. Cheong, and J. M. Cioffi, "Iterative soft interference cancellation for multiple antenna systems," in *Proc. IEEE WCNC'2000*, Sep. 2000, vol. 1, pp. 304–309.

[31] N. Prasad, M. K. Varanasi, L. Venturino, and X. Wang, "An analysis of the MIMO-SDMA channel with space–time orthogonal and quasi-orthogonal user transmissions and efficient successive cancellation decoders," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5427–5446, Dec. 2008.

[32] C. W. Tan and A. R. Calderbank, "Multiuser detection of Alamouti signals," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 2080–2089, Jul. 2009.

[33] 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. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[34] S. K. Mohammed, A. Chockalingam, and B. S. Rajan, "Low-complexity detection and performance in multi-gigabit high spectral efficiency large MIMO systems," in *Proc. IEEE PIMRC'2008*, Sept. 2008, pp. 1–5.

[35] M. Vu and A. Paulraj, "Optimal linear precoders for MIMO wireless correlated channels with nonzero mean in space time coded systems," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2318–2332, Jun, 2006.

[36] H. R. Bahrami and T. Le-Ngoc, "Precoder design based on correlation matrices for MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3579–3587, Dec. 2006.

[37] K. T. Phan, S. A. Vorobyov, and C. Tellambura, "Precoder design for space–time coded systems with correlated Rayleigh fading channels using convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 814–819, Feb. 2009.

[38] A. Zaki, S. K. Mohammed, A. Chockalingam, and S. S. Rajan, "A training-based iterative detection/channel estimation scheme for large non-orthogonal STBC MIMO systems," in *Proc. IEEE ICC'2009*, Jun. 2009, pp. 1–5.
[39] M. Brehler and M. K. Varanasi, "Training-codes for non-coherent multi-antenna block-Rayleigh fading channel," in *Proc. CISS'2003*, Mar. 2003.
[40] H. El Gamal and M. O. Damen, "Universal space–time coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1097–1119, May 2003.
[41] H. El Gamal, H. Aktas, and M. O. Damen, "Coherent space–time codes for noncoherent channels," in *Proc. IEEE GLOBECOM'2003*, Dec. 2003, pp. 1915–1918.
[42] J.-C. Belfiore and A. M. Cipriano, "Space–time coding for noncoherent channels," in *Book Chapter in Space–Time Wireless Systems: From Array Processing to MIMO Communications*, H. Bölcskei, D. Gesbert, C. B. Papadias, and A.-J. van der Veen, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2006.
[43] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
[44] P. Stoica and G. Ganesan, "Space–time block codes: Trained, blind and semi-blind detection," in *Proc. IEEE ICASSP'2002*, 2002, pp. 1609–1612.
[45] A. Lampe, "Iterative multiuser detection with integrated channel estimation for coded DS-CDMA," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1217–1223, Aug. 2002.
[46] H. Li, S. M. Betz, and H. V. Poor, "Performance analysis of iterative channel estimation and multiuser detection in multipath DS-CDMA channels," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1981–1993, May 2007.
[47] B. Hu, I. Land, L. Rasmussen, R. Piton, and B. H. Fleury, "A divergence minimization approach to joint multiuser decoding for coded CDMA," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 432–445, Apr. 2008.
[48] M. Loncar, R. R. Muller, J. Wehinger, C. F. Mecklenbrauker, and T. Abe, "Iterative channel estimation and data detection in frequency-selective fading MIMO channels," *Eur. Trans. Telecomm.*, vol. 15, no. 5, pp. 459–470, Sep./Oct. 2004.
[49] H. Zhu, B. Farhang-Boroujeny, and C. Schlegel, "Pilot embedding for joint channel estimation and data detection in MIMO communication systems," *IEEE Commun. Lett.*, vol. 7, no. 1, pp. 30–32, Jan. 2003.
[50] J. Akhtman and L. Hanzo, "Iterative receiver architectures for MIMO-OFDM," in *Proc. IEEE WCNC'2007*, Mar. 2007, pp. 825–829.
[51] P. S. Rossi and R. R. Muller, "Joint twofold-iterative channel estimation and multiuser detection for MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4719–4729, Nov. 2008.
[52] V. Shashidhar, B. S. Rajan, and B. A. Sethuraman, "Information-lossless space–time block codes from crossed-product algebras," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3913–3935, Sep. 2006.
[53] Y. Wang and K. Roy, "A new reduced complexity sphere decoder with true lattice boundary awareness for multi-antenna systems," *IEEE ISCAS'2005*, vol. 5, pp. 4963–4966, May 2005.

Saif K. Mohammed received the B.Tech. degree in computer science and engineering from the Indian Institute of Technology, New Delhi, India, in 1998. He is currently pursuing the Ph.D. degree in electrical and communications engineering at the Indian Institute of Science, Bangalore, India

From 1998 to 2000, he was employed with Philips, Inc., Bangalore, as an ASIC Design Engineer. From 2000 to 2003, he worked with Ishoni Networks, Inc., Santa Clara, CA, as a Senior Chip Architecture Engineer. From 2003 to 2007, he was employed with Texas Instruments, Bangalore, as a Systems and Algorithms Designer in the wireless systems group. His research interests include low-complexity detection, estimation, and coding for wireless communications systems.

Ahmed Zaki received the B.E. degree in electronics and communication engineering from Osmania University, Hyderabad, India, in 2007, and the M.E. degree in telecommunication from the Indian Institute of Science, Bangalore, India, in 2009.

His research interest lies in the area of wireless communications, including receiver design and channel estimation for large-MIMO systems, MIMO-OFDM, multiuser communications, and algorithm design.

A. Chockalingam (S'92–M'95–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 with specialization in satellite communications from the Indian Institute of Technology, Kharagpur, India, in 1985, and the Ph.D. degree in electrical communication engineering from the Indian Institute of Science (IISc), Bangalore, India, in 1993.

From 1986 to 1993, he worked with the Transmission R&D Division, 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 was with 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 Electrical and Computer Engineering, IISc, Bangalore, 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 for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He served as a Guest Editor for the IEEE JSAC Special Issue on Multiuser Detection for Advanced Communication Systems and Networks. He is a Fellow of the Institution of Electronics and Telecommunication Engineers and a Fellow of the Indian National Academy of Engineering.

B. Sundar Rajan (S'84–M'91–SM'98) was born in Tamil Nadu, India. He received the B.Sc. degree in mathematics from Madras University, Madras, India, the B.Tech degree in electronics from Madras Institute of Technology, Madras, and the M.Tech and Ph.D. degrees in electrical engineering from the Indian Institute of Technology, Kanpur, India, in 1979, 1982, 1984, and 1989 respectively.

He was a faculty member with the Department of Electrical Engineering, Indian Institute of Technology, Delhi, India, from 1990 to 1997. Since 1998, he has been a Professor in the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. His primary research interests include space–time coding for MIMO channels, distributed space–time coding and cooperative communication, coding for multiple-access, relay channels, and network coding with emphasis on algebraic techniques.

Dr. Rajan is an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY, an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and an Editorial Board Member of *International Journal of Information and Coding Theory*. He served as Technical Program Co-Chair of the IEEE Information Theory Workshop (ITW'02), held in Bangalore, in 2002. He is a Fellow of Indian National Academy of Engineering and recipient of the IETE Pune Center's S.V.C Aiya Award for Telecom Education in 2004. Dr. Rajan is also a Member of the American Mathematical Society.