

Near-Optimum Soft Decision Equalization for Frequency Selective MIMO Channels

Shoumin Liu, *Student Member, IEEE*, and Zhi Tian, *Member, IEEE*

Abstract—In this paper, we develop soft decision equalization (SDE) techniques for frequency selective multiple-input multiple-output (MIMO) channels in the quest for low-complexity equalizers with error performance competitive to that of maximum likelihood (ML) sequence detection. We demonstrate that decision feedback equalization (DFE) based on soft-decisions, expressed via the posterior probabilities associated with feedback symbols, is able to outperform hard-decision DFE, with a low computational cost that is polynomial in the number of symbols to be recovered and linear in the signal constellation size. Building on the probabilistic data association (PDA) multiuser detector, we present two new MIMO equalization solutions to handle the distinctive channel memory. The first SDE algorithm adopts a zero-padded transmission structure to convert the challenging sequence detection problem into a block-by-block least-square formulation. It introduces key enhancement to the original PDA to enable applications in rank-deficient channels and for higher level modulations. The second SDE algorithm takes advantage of the Toeplitz channel matrix structure embodied in an equalization problem. It processes the data samples through a series of overlapping sliding windows to reduce complexity and, at the same time, performs implicit noise tracking to maintain near-optimum performance. With their low complexity, simple implementations, and impressive near-optimum performance offered by iterative soft-decision processing, the proposed SDE methods are attractive candidates to deliver efficient reception solutions to practical high-capacity MIMO systems. Simulation comparisons of our SDE methods with minimum-mean-square error (MMSE)-based MIMO DFE and sphere decoded quasi-ML detection are presented.

Index Terms—Efficient reception algorithms, equalization for frequency selective MIMO channels, overlapping sliding windowing, soft decision.

I. INTRODUCTION

ENORMOUS increase in bandwidth efficiency is promised by the use of multiple-input-multiple-output (MIMO) systems in wireless radio frequency links [1]. In high-data-rate applications, channel-induced intersymbol interference (ISI) can be mitigated using serial equalization techniques, which unfortunately encounter major challenges in MIMO channels, because of the need for signal detection in the presence of both multiple access interference (MAI) as well as ISI. The BCJR

maximum *a posteriori* (MAP) [2] and Viterbi equalizers perform optimum sequence detection to account for the channel memory but incur prohibitive complexity that is exponential in the number of inputs and the channel memory. Their equalization complexity can be reduced using standard reduced-complexity methods for single-input channels (e.g., [3], [4]), but such methods do not apply to MIMO channels. For complexity considerations, typical equalizers consist of linear processing of the received signal, i.e., linear equalization (LE), and possibly past symbol estimation, e.g., decision feedback equalization (DFE). The optimum MIMO DFE settings in the minimum mean-square error (MMSE) sense have been derived in [5]–[8]. In these schemes, tentative decisions on both past symbols and symbols from MAI sources are made by quantizing properly derived decision statistics. Such a *hard-decision*-based approach may suffer from catastrophic error propagation and, in most cases, incurs nontrivial performance degradation relative to an optimal maximum likelihood (ML) detector, in terms of the bit error rate (BER) performance. Recent advances include turbo detection and equalization [9]–[11], in which low-complexity soft-input-soft-output (SISO) LE and DFE equalizers are devised based on the MMSE criterion [10].

This work accentuates the low-complexity, near-optimum equalization for frequency selective MIMO channels by taking on a *soft-decision* equalization (SDE) approach. We recognize that block transmission by zero padding permits the conversion of sequence detection to block detection; thus, MIMO channel equalization can be viewed as a general high-dimensional integer least-square (LS) problem in the form of $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}$, where \mathbf{x} is an $\bar{N} \times 1$ input vector that takes on finite-alphabet values, and \mathbf{y} , \mathbf{H} , \mathbf{v} are the output vector, the channel matrix, and the noise vector, respectively. Unique to channel equalization, the channel matrix \mathbf{H} entails a special Toeplitz structure, to be discussed in Section II. The exact ML solution to the integer-LS problem unfortunately incurs prohibitively high computational complexity that is exponential in \bar{N} . In the quest for low-complexity implementations of ML block detection, quasi-ML solutions have been developed, such as sphere decoding [12], semi-definite relaxation [13], and probabilistic data association (PDA) filtering [14], [15]. The PDA detector [14] provides near-optimal performance at a low overall complexity of $\mathcal{O}(\bar{N}^3)$. It employs a multistage detection structure and replaces the intermediate finite alphabet symbol decisions by soft decisions, which are expressed via their associated posterior probabilities. Such a soft-decision structure leads to significant computation reduction when MAI is approximated to obey a Gaussian probability distribution: an idea originated from the PDA filter for target tracking

Manuscript received December 18, 2002; revised May 16, 2003. A portion of the manuscript was presented in the *Proceedings of the 2nd IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM'2002)*, Rosslyn, VA, August 2002. The associate editor coordinating the review of this paper and approving it for publication was Prof. Dhananjay A. Gore.

The authors are with the Department of Electrical and Computer Engineering, Michigan Technological University, Houghton, MI 49931 USA (e-mail: ztian@mtu.edu).

Digital Object Identifier 10.1109/TSP.2003.822376

[16]. Despite its impressively low BERs, the PDA method in its current form has limited applications. The original PDA detector was derived for multiuser detection (MUD) within the framework of code-division multiple access (CDMA) in frequency flat-fading channels, in which case, the channel matrices \mathbf{H} are confined to be square matrices made of users' cross-correlation coefficients. Zero-forcing preprocessing via square matrix inverse is performed in PDA, imposing an invertibility constraint on \mathbf{H} .

Symbol detection in practical MIMO systems encounters pronouncedly different system parameters \mathbf{H} and \mathbf{b} compared with those in MUD for single-channel CDMA. The channel matrix is typically nonsquare and possibly rank-deficient. For example, downlink transmissions often face channels that are fat, that is, there are more columns than rows in \mathbf{H} . Such a channel matrix is no longer invertible, even though the system may still be identifiable for digital inputs. A typical case is a BLAST system [17] with a larger number of transmit than receive antennas. Moreover, in high-capacity wireless systems, the input signal constellation size increases to improve spectral efficiency, such as quadrature phase shift keying used in third-generation cellular and 16/64-quadrature amplitude modulation (QAM) in IEEE 802.11a. Higher level modulation is one of the primary reasons that obviate ML detection, and it has not been discussed in the current PDA detection algorithm. In conjunction with Kalman filtering, the PDA method can handle a short channel memory length of 2 induced by asynchronism only [15], but the result cannot be extended to channels with a longer memory length. The PDA MUD is hampered by these obstacles; hence, its remarkable features cannot be fully appreciated by wireless MIMO systems. Constructing a successful PDA-type MIMO channel equalizer calls for a research effort that is beyond the scope of simple generalization of the existing PDA block detectors.

In this paper, we propose MIMO SDE techniques that incorporate the PDA MUD principle. Different from the hard-decision multistage parallel interference cancellation (PIC), the SDE approach generates tentative decisions on ISI and MAI symbols in the form of posterior probabilities instead of quantized bits, and the decision updating is simplified by forcing the composite effect of noise and interference to be Gaussian: a strategy used in the PDA detector. The SDE detection performance improves with additional iterations and stabilizes quickly in three to five iterations for high SNR, and seven to 14 iterations for low SNR. Compared with the suboptimum hard-decision MMSE-DFE [6], the SDE method demonstrates close-to-optimal equalization performance at a comparable low complexity that is polynomial in the number of inputs and only linear in the constellation size of \mathbf{b} . Compared with other quasi-ML methods such as sphere decoding [12], the SDE is not only competitive in both performance and complexity but also applies to situations where sphere decoding does not work well, such as the fat channel case. The SDE is distinct from PDA MUD in several key accounts. First, the SDE approach is tailored to handle near-optimum symbol sequence detection in the presence of channel memory, which is difficult to accommodate in PDA MUD. Our main contribution is to adopt the block transmission structure via zero padding to enable block detection and to

apply sliding windows for ISI cancellation and noise tracking to attain near-MAP detection performance at a low complexity. Second, we propose an alternative implementation of the PDA principle that eliminates the zero-forcing preprocessing. As a result, the restriction on full-rank square channel matrices is lifted. Third, we extend the PDA algorithm to bandwidth-efficient higher level modulation schemes. Through this work, the potential of the iterative soft-decision PDA philosophy can be fully enjoyed by practical wireless systems to achieve near-optimal, low-complexity detection and equalization.

The ensuing paper is organized as follows. Section II describes the input-output model for block transmissions through frequency selective MIMO channels. Section III develops two SDEs: One applies to a general integer-LS problem by enhancing the original PDA algorithm with an alternative implementation that obviates zero-forcing preprocessing; the other focuses on frequency selective channels and capitalizes on the distinctive Toeplitz channel structure to reduce the equalization complexity. Analytical comparisons between our SDE algorithms and the existing iterative methods, including PDA and turbo SISO-MMSE, are discussed. Section IV performs evaluation of the proposed MIMO SDE algorithms in terms of their BER performance and the computational load, with elaboration on the tradeoffs in choosing the sliding window size and on their application in the rank-deficient fat channel case. Comparisons with ML detection, sphere decoding, and MIMO MMSE-DFE are presented via computer simulations, followed by a summary in Section V.

II. SIGNAL MODEL

We consider the discrete-time block transmission equivalent model of a baseband communication system with N_i inputs and N_o outputs. There are a total of $N_i N_o$ links in this MIMO system, wherein the link between each input-output pair is modeled as a linear finite impulse response (FIR) dispersive channel with no greater than $L + 1$ symbol-spaced taps in the channel response. The sampled channel response from the i th input to the j th output, including transmit and receive filters, is denoted by $\mathbf{h}^{(i,j)} := [h^{(i,j)}(0), \dots, h^{(i,j)}(L)]^T$. We adopt a block transmission structure with zero padding to eliminate interblock interference, hence alleviating the performance degradation due to noise enhancement or error propagation [19]. The information-bearing symbols are parsed into N -long frames, with the insertion of $P \geq L$ zeros at the tail of each frame. The k th frame of the input vector is denoted as $\mathbf{b}_k := [\mathbf{b}^T(kN), \dots, \mathbf{b}^T(kN + N - 1)]^T$, where $\mathbf{b}(m) := [b^{(1)}(m), \dots, b^{(N_i)}(m)]^T$ contains the information symbols from all N_i inputs at the m th sampling instant. After zero padding, each N -long information-bearing symbol frame \mathbf{b}_k creates a transmit symbol frame $\bar{\mathbf{b}}_k$ of frame size $K := N + P$, where the first N entries convey messages $\{\bar{b}^{(i)}(kK + n)\}_{n=0}^{N-1} = b^{(i)}(kN + n)$, followed by P trailing zeros $\{\bar{b}^{(i)}(kK + n)\}_{n=N}^{K-1} = 0$, for any frame index k and input $i \in [1, N_i]$. Correspondingly, the received data vector at the k th frame is a concatenation of K noise-contaminated sample vectors $\mathbf{y}_k := [\mathbf{y}^T(kK), \dots, \mathbf{y}^T(kK + K - 1)]^T$, where $\mathbf{y}(m) := [y^{(1)}(m), \dots, y^{(N_o)}(m)]^T$ consists of the m th received signals at all N_o outputs. The redundancy per transmitted frame is measured by the ratio P/K , whereas at the receiver,

the data rate is reduced by the same amount. We set $P = L$ and typically choose a large frame size K (and N) to maintain the transmission rate.

In linear FIR channels, the received $y^{(j)}(kK+n)$ is expressed by

$$y^{(j)}(kK+n) = \sum_{i=1}^{N_i} \sum_{l=0}^{L-1} h^{(i,j)}(l) \bar{b}^{(i)}(kK+n-l) + v^{(j)}(kK+n) \\ n = 0, \dots, K-1; \quad j = 1, \dots, N_o \quad (1)$$

where $v^{(j)}(kK+n)$ denotes the additive zero-mean Gaussian stationary noise received at the j th output. The noise terms at the k th transmission frame are grouped into a $N_o K \times 1$ vector $\mathbf{v}_k := [\mathbf{v}^T(kK), \dots, \mathbf{v}^T(kK+K-1)]^T$, where $\mathbf{v}(m) := [v^{(1)}(m), \dots, v^{(N_o)}(m)]^T$, and the covariance of \mathbf{v}_k is denoted by $\mathbf{R}_v := E\{\mathbf{v}_k \mathbf{v}_k^H\}$. With these definitions, the single-link input-output model (1), when assembled into the vector-matrix format, results in a MIMO channel model in the form of $\mathbf{y}_k = \mathbf{H}\mathbf{b}_k + \mathbf{v}_k$, where the MIMO channel matrix takes on a banded Toeplitz structure [19]:

$$\mathbf{H} := \begin{bmatrix} \mathbf{H}(0) & & \mathbf{0} \\ \vdots & \ddots & \\ \mathbf{H}(L) & \cdots & \mathbf{H}(0) \\ & \ddots & \vdots \\ \mathbf{0} & & \mathbf{H}(L) \end{bmatrix} \quad (2)$$

where

$$\mathbf{H}(l) := \begin{bmatrix} h^{(1,1)}(l) & \cdots & h^{(N_i,1)}(l) \\ \vdots & \cdots & \vdots \\ h^{(1,N_o)}(l) & \cdots & h^{(N_i,N_o)}(l) \end{bmatrix} \\ l = 0, \dots, L. \quad (3)$$

Sequence detection incurred by the channel memory can now be alternatively solved by frame-by-frame symbol detection, for which we will drop the subscript k without raising confusion. This system model subsumes a broad range of MIMO transmission scenarios. The multiple input could result from the combination of three situations:

- i) multiple transmit antennas with a single user, e.g., the single-user space-time coding case;
- ii) single transmit antenna with multiple users, e.g., as encountered in the single-channel multiuser detection problem;
- iii) transmit-induced diversities including orthogonal frequency division multiplexing (OFDM) and CDMA.

Thus, the number of inputs N_i is determined by both the number of transmit antennas and the number of multiple-access users. Meanwhile, multi-output is invoked when there are multiple receive antennas and/or when fractional sampling is used. Thus, N_o is determined by the total number of distinct samplers operating within a symbol period.

With this MIMO model, the integer-LS problem described in the Introduction deals with a channel matrix of size $\bar{K} \times \bar{N}$, where $\bar{K} := N_o(N+L)$ and $\bar{N} := N_i N$. The optimal

ML solution to such a MIMO system faces a major implementation challenge, as its complexity increases exponentially in N_i, N_o, N , and L . Our objective in this paper is to develop soft-decision-based symbol detection and channel equalization schemes that achieve near-optimal BER performance at low polynomial complexity. Throughout, the MIMO channel is assumed to be frequency selective and slowly varying. It is time invariant within each frame of K symbol periods but may change independently from frame to frame. We suppose the receiver has perfect knowledge of the channel state information \mathbf{H} and the noise variance \mathbf{R}_v .

III. NEAR-OPTIMUM SDE

We now develop SDE methods based on the PDA-type soft-decision multistage detection principle. We begin by enhancing the PDA detector to enable its applicability to a generic block transmission system with higher level modulation. A key modification is to associate each symbol with its channel response vector when deriving its posterior probability density function (pdf) and updating its soft decisions. This strategy eliminates the preprocessing of channel matrix inversion, thus relaxing the constraint on channel invertibility. Next, we develop a new SDE algorithm that is specifically tailored to near-optimum symbol detection in channels with memory. Taking advantage of the Toeplitz channel matrix structure, we apply a series of overlapping sliding windows to obtain reduced-complexity local MAP estimation of current symbols and, at the same time, track the ISI and noise components to ensure near-optimum performance. The overall computational load is reduced to a fraction (proportional to the transmission redundancy) of that of the enhanced PDA solution. For each SDE algorithm, we present efficient implementations of the iterative soft-decision updating rule to further reduce the complexity.

A. SDE by PDA Enhancement

In the general MIMO model $\mathbf{y} = \mathbf{H}\mathbf{b} + \mathbf{v}$, we emphasize the i th element b_i of \mathbf{b} by rewriting the received signal as

$$\mathbf{y} = \mathbf{h}_i b_i + \sum_{j=1, j \neq i}^{\bar{N}} \mathbf{h}_j b_j + \mathbf{v}, \quad i = 1, \dots, \bar{N} \quad (4)$$

where \mathbf{h}_i and $\mathbf{h}_j \in \mathcal{R}^{\bar{K} \times 1}$ are the i th and j th columns of \mathbf{H} , respectively, denoting the channel responses of b_i and $b_j, \forall i, j \in [1, \bar{N}]$, as shown in Fig. 1(a). The transmitted bits $\{b_j\}_{j=1}^{\bar{N}}$ take values from a finite alphabet set $\{a_1, \dots, a_M\}$ on M -ary modulation, where the modulation format is typically chosen from phase shift keying (PSK) and QAM. Given \mathbf{y} , there are M posterior probability values associated with each digital input, which we denote as $\eta_{i,m} := \Pr(b_i = a_m | \mathbf{y})$, for $m \in [1, M]$ and $i \in [1, \bar{N}]$.

Finite-alphabet symbol detection on b_i can be alternatively carried out by estimating $\eta_{i,m}$, giving rise to soft decisions. Unfortunately, direct evaluation of $\eta_{i,m}$ via the corresponding likelihood function $p(\mathbf{y} | b_i)$ still incurs exponential complexity in \bar{N} , considering that $p(\mathbf{y} | b_i)$ is a Gaussian mixture with $M^{\bar{N}-1}$ modes [22]. To avoid the combinatorial complexity, we adopt the PDA filtering idea and treat the transmitted

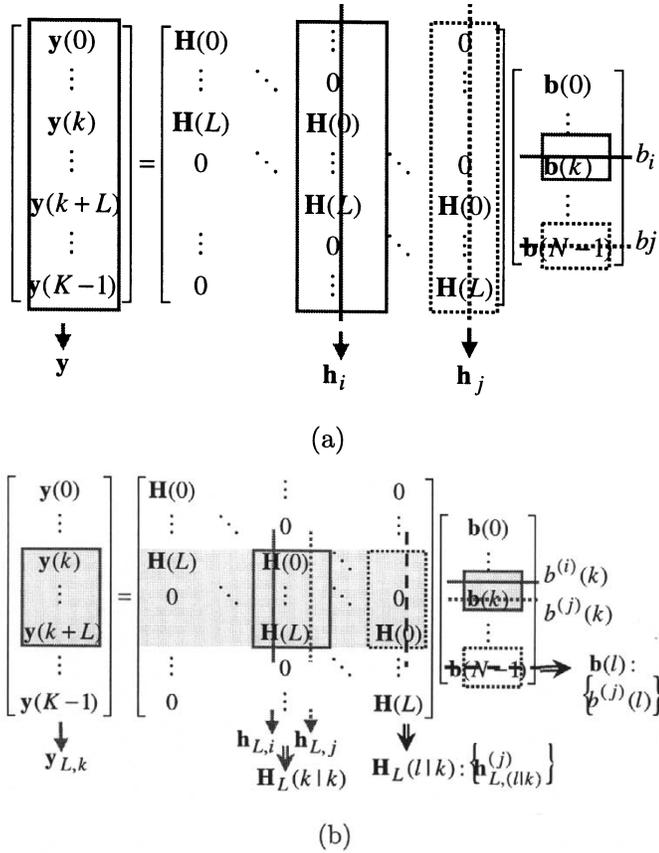


Fig. 1. Channel models. (a) SDE-1. (b) SDE-2. Each vertical line represents the signature vector of the corresponding input denoted by a horizontal line. The length of the signature vector determines the detection complexity for the corresponding input.

symbols $\{b_i\}_{i=1}^{\bar{N}}$ as Gaussian random variables. Since the FIR MIMO channel is a linear system, the posterior pdf of b_i remains Gaussian; thus, it can be fully characterized by its mean and variance, conditioned on the received signal \mathbf{y} . Define $\mathbf{g}_i := E\{\mathbf{h}_i b_i | \mathbf{y}\}$ and $\mathbf{R}_i := \text{cov}\{\mathbf{h}_i b_i | \mathbf{y}\}$ as the conditional mean and covariance of $\mathbf{h}_i b_i$, respectively. These definitions are different from [14], noting that each symbol b_i is now associated with its channel response vector \mathbf{h}_i , instead of a simple unit vector. Such definitions eliminate the need to perform the channel decorrelating preprocessing in [14], which is not applicable for nonsquare channel matrices and rank-deficient channels. When the transmitted symbols are independent and identically distributed (i.i.d.), it follows from (4) that

$$\mathbf{g}_i = \mathbf{y} - \sum_{j=1, j \neq i}^{\bar{N}} E\{b_j | \mathbf{y}\} \mathbf{h}_j \quad (5)$$

$$\mathbf{R}_i = \sum_{j=1, j \neq i}^{\bar{N}} \text{var}\{b_j | \mathbf{y}\} \mathbf{h}_j \mathbf{h}_j^H + \mathbf{R}_v \quad (6)$$

where $(|\cdot|)$ denotes the magnitude of a complex quantity)

$$E\{b_j | \mathbf{y}\} := \sum_{b_j \in \{a_1, \dots, a_M\}} b_j \Pr(b_j | \mathbf{y}) = \sum_{m=1}^M \eta_{j,m} a_m \quad (7)$$

$$\begin{aligned} \text{var}\{b_j | \mathbf{y}\} &:= E\{|b_j|^2 | \mathbf{y}\} - |E\{b_j | \mathbf{y}\}|^2 \\ &= \sum_{m=1}^M \eta_{j,m} |a_m|^2 - |E\{b_j | \mathbf{y}\}|^2. \end{aligned} \quad (8)$$

In antipodal signaling, $b_j \in \{\pm 1\}$, which leads to $E\{b_j | \mathbf{y}\} = 1 \cdot \eta_j + (-1) \cdot (1 - \eta_j) = 2\eta_j - 1$, and $\text{var}\{b_j | \mathbf{y}\} = 4\eta_j(1 - \eta_j)$, which corroborates [14].

When $\mathbf{h}_i b_i | \mathbf{y}$ is approximated as a Gaussian vector with the matched mean and covariance, its pdf can be described by \mathbf{g}_i and \mathbf{R}_i as follows:

$$\begin{aligned} p_{b_i | \mathbf{y}}(\mathbf{h}_i b_i | \mathbf{y}) &= \frac{1}{(2\pi)^{\bar{K}/2} |\mathbf{R}_i|^{1/2}} \\ &\times \exp\left\{-\frac{1}{2} (\mathbf{h}_i b_i - \mathbf{g}_i)^H \mathbf{R}_i^{-1} (\mathbf{h}_i b_i - \mathbf{g}_i)\right\}. \end{aligned} \quad (9)$$

This pdf calibrates the posterior probabilities of all M possible constellation points of b_i , and the soft decisions can be derived from (9) by setting $\eta_{i,m} = c_i p_{b_i | \mathbf{y}}(\mathbf{h}_i a_m | \mathbf{y})$, where c_i is a normalization factor such that $\sum_{m=1}^M \eta_{i,m} = 1, \forall i$. Note that the conditional pdf in (9) is determined by \mathbf{g}_i and \mathbf{R}_i , regardless of the constellation size M . Therefore, the overall complexity of evaluating all M soft decisions $\{\eta_{i,m}\}_{m=1}^M$ is linear in M .

To obtain $\eta_{i,m}$ in an efficient manner, we introduce the ratios $\lambda_{i,m} := \eta_{i,m} / \eta_{i,1}$, for $m = 1, \dots, M$, which can be deduced from (9) as

$$\begin{aligned} \lambda_{i,m} &= \exp\left\{-\frac{1}{2} [(a_m \mathbf{h}_i - \mathbf{g}_i)^H \mathbf{R}_i^{-1} (a_m \mathbf{h}_i - \mathbf{g}_i) \right. \\ &\quad \left. - (a_1 \mathbf{h}_i - \mathbf{g}_i)^H \mathbf{R}_i^{-1} (a_1 \mathbf{h}_i - \mathbf{g}_i)]\right\}. \end{aligned} \quad (10)$$

Using the probability normalization condition $\sum_m \eta_{i,m} = 1$, we obtain

$$\eta_{i,m} = \begin{cases} \frac{1}{1 + \sum_{m=2}^M \lambda_{i,m}}, & m = 1 \\ \lambda_{i,m} \eta_{i,1}, & m = 2, \dots, M. \end{cases} \quad (11)$$

Subsequently, the MAP estimate of \hat{b}_i is decided to be the a_m value that yields the largest $\eta_{i,m}$. The decision rule can be further simplified in the special binary modulation case [14], [18].

For algorithm implementation, we note that computing $\{\eta_{i,m}\}_{m=1}^M$ for one input b_i involves computing \mathbf{g}_i and \mathbf{R}_i , where both depend on $\{\eta_{j,m}\}_{m=1}^M$ of all other inputs $\{b_j\}_{j=1, j \neq i}^{\bar{N}}$. This intertwined relationship among all unknown inputs prompts an iterative multistage procedure [14], where the $(\mathbf{g}_i, \mathbf{R}_i)$ pair is computed from (5) and (6) based on tentative soft decisions $\{\eta_{j,m}\}_{j \neq i}$ obtained at a previous stage. Each $\eta_{i,m}$ can be updated from (11) successively until all $\{\eta_{i,m}\}_{m=1}^M$ converge for all $i \in [1, \bar{N}]$, followed by decision making on this frame of symbols \mathbf{b} via the MAP rule.

In the above iterative soft-decision updating procedure, the most computationally expensive operation is to compute the covariance matrix inverse \mathbf{R}_i^{-1} , which incurs complexity on the third order of the number of outputs \bar{K} . Direct matrix inversion can be avoided using the matrix inversion lemma, which will lower the overall complexity by an order. This speed-up measure is discussed in [14], and it shares the same principle with the widely used recursive least-square (RLS) adaptive filtering

TABLE I
SOFT DECISION EQUALIZATION ALGORITHM: SDE-1

1. Initialization
$\eta_{i,m} = \frac{1}{M}, \forall i, m; \mu_a = \frac{1}{M} \sum_{m=1}^M a_m;$
$\sigma_a^2 = \frac{1}{M} \sum_{m=1}^M a_m - \mu_a ^2;$
$\mathbf{g} = \mathbf{y} - \mu_a \sum_{i=1}^N \mathbf{h}_i; \mathbf{R} = \sigma_a^2 \mathbf{H}\mathbf{H}^H + \mathbf{R}_v;$
2. Soft-decision iterative updating
Do
for $i = 1 : \bar{N}$
– compute \mathbf{g}_i and \mathbf{R}_i^{-1} from (12) and (14);
– compute $\{\eta_{i,m}\}_{m=1}^M$ from (10) – (11),
and update $E\{b_i \mathbf{y}\}$ and $\text{var}\{b_i \mathbf{y}\}$ from (7), (8);
– update \mathbf{g} and \mathbf{R}^{-1} from (12) and (15);
end i
Until $\eta_{i,m}$ converges for all i and m
3. MAP Detection: $\hat{b}_i = \arg \max_{b_i \in \{a_1, \dots, a_M\}} \eta_{i,m}(b_i)$

[21]. We now describe these results for our M -ary modulation case.

First, we form two auxiliary variables: $\mathbf{g} := \mathbf{y} - \sum_{j=1}^{\bar{N}} E\{b_j|\mathbf{y}\}\mathbf{h}_j$ is the conditional mean of the noise term $\mathbf{v} = \mathbf{y} - \mathbf{H}\mathbf{b}$, and $\mathbf{R} := \sum_{j=1}^{\bar{N}} \text{var}\{b_j|\mathbf{y}\}\mathbf{h}_j\mathbf{h}_j^H + \mathbf{R}_v$ is the conditional covariance matrix of \mathbf{y} . Apparently

$$\mathbf{g}_i = \mathbf{g} + E\{b_i|\mathbf{y}\}\mathbf{h}_i \quad (12)$$

$$\mathbf{R}_i = \mathbf{R} - \text{var}\{b_i|\mathbf{y}\}\mathbf{h}_i\mathbf{h}_i^H. \quad (13)$$

Applying the matrix inversion lemma on (13) yields [21]

$$\mathbf{R}_i^{-1} = \mathbf{R}^{-1} + \frac{\text{var}\{b_i|\mathbf{y}\}\mathbf{R}^{-1}\mathbf{h}_i\mathbf{h}_i^H\mathbf{R}^{-1}}{1 - \text{var}\{b_i|\mathbf{y}\}\mathbf{h}_i^H\mathbf{R}^{-1}\mathbf{h}_i} \quad (14)$$

and, conversely

$$\mathbf{R}^{-1} = \mathbf{R}_i^{-1} - \frac{\text{var}\{b_i|\mathbf{y}\}\mathbf{R}_i^{-1}\mathbf{h}_i\mathbf{h}_i^H\mathbf{R}_i^{-1}}{1 + \text{var}\{b_i|\mathbf{y}\}\mathbf{h}_i^H\mathbf{R}_i^{-1}\mathbf{h}_i}. \quad (15)$$

By keeping the updated versions of \mathbf{g} and \mathbf{R}^{-1} , $\{\mathbf{g}_i\}_{i=1}^{\bar{N}}$ and $\{\mathbf{R}_i^{-1}\}_{i=1}^{\bar{N}}$ can be obtained from (12) and (14) at a low complexity of $\mathcal{O}(\bar{K}^2)$ for each input b_i and an overall complexity of $\mathcal{O}(\bar{N}\bar{K}^2)$ in one iteration. The overall soft-decision MIMO equalization algorithm, which is enabled by zero-padded block transmission, is summarized in Table I.

B. SDE by Sliding Windowing

So far, we have not utilized the unique Toeplitz structure of the channel matrix in the equalization problem. When the pdfs in (9) are computed via matrix operations, the complexity of the Gaussian-forcing MAP detection is determined by the length of each input's channel response vector \mathbf{h}_i , which is $\bar{K} = (N+L)N_o$ in the SDE-1 algorithm. On the other hand, when the block size $K = N+L$ is chosen to be much larger than the channel length L to reduce the transmission redundancy, there are a large number of zeros in the channel matrix,

which could be avoided to save computation. Next, we utilize the effective portion of each channel response vector and construct a soft decision approach that is tailored to the equalization problem. The objective is to maintain the near-MAP detection accuracy of SDE-1 and, at the same time, reduce the computational complexity to be proportional to the channel memory L instead of the frame size $K \gg L$.

1) *Algorithm Development*: Revisiting the signal model (4) illustrated in Fig. 1(b), we divide the symbol vector \mathbf{b} into N sub-blocks $\{\mathbf{b}(k)\}_{k=0}^{N-1}$, where each sub-block $\mathbf{b}(k) = [b^{(1)}(k) \dots b^{(N_i)}(k)]^T$ contains the information symbols from all N_i inputs at the k th sampling instant. Due to the finite channel memory length, $\mathbf{b}(k)$ only affects $(L+1)N_o$ output elements, which we group into $\mathbf{y}_{L,k} := [\mathbf{y}^T(k), \dots, \mathbf{y}^T(k+L)]^T$ [see Fig. 1(b)]. This output block $\mathbf{y}_{L,k}$ contains the sufficient statistics of $\mathbf{b}(k)$, as well as contributions from residual ISI elements $\{\mathbf{b}(l)\}_{l \neq k}$. Compared with \mathbf{y} , the reduced-sized vector $\mathbf{y}_{L,k}$ contains all the observations relevant to $\mathbf{b}(k)$. It is thus possible to construct an optimum detector for $\mathbf{b}(k)$ from $\mathbf{y}_{L,k}$ in lieu of \mathbf{y} , provided that the contributions from other symbols to this output block are properly accounted for, possibly through ISI cancellation and noise tracking. Focusing on one input sub-block at a time, we will convert the signal model in (4) into a set of N sub-models, each describing one of the N sufficient statistics $\{\mathbf{y}_{L,k}\}_{k=0}^{N-1}$, such that each sub-model can be expressed by shorter channel response vectors compared with (4). In the k th sub-model depicted by the shaded area in Fig. 1(b), we denote the channel response matrix from the l th input sub-block $\mathbf{b}(l)$ to $\mathbf{y}_{L,k}$ by $\mathbf{H}_L(l|k), \forall l$, which refers to the sub-block matrix of \mathbf{H} at the intersection of the k th row sub-block and the l th column sub-block. To be exact, it is the portion of \mathbf{H} bordered by the (lN_i+1) th to the $(l+1)N_i$ th columns and the (kN_o+1) th to the $(k+L+1)N_o$ th rows. There are N_i columns in $\mathbf{H}_L(l|k) = [\mathbf{h}_{L,(l|k)}^{(1)} \dots \mathbf{h}_{L,(l|k)}^{(N_i)}]$, where $\mathbf{h}_{L,(l|k)}^{(i)}$ is the $(L+1)N_o \times 1$ channel response vector of $b^{(i)}(l)$ during the k -dependent observation window covering the k th to the $(k+L)$ th symbol periods. It is further observed that when $l = k$, $\mathbf{H}_L(k|k)$ represents the nonzero portion of each column block and is independent of k . In fact, $\mathbf{H}_L(k|k) = \mathbf{H}_L := [\mathbf{H}^T(0) \dots \mathbf{H}^T(L)]^T, \forall k$, which happens to be the $(L+1)$ -tap FIR MIMO channel impulse response. With these definitions, $\mathbf{y}_{L,k}$ can be obtained from the general signal model (4) by sliding over \mathbf{y} a k -dependent observation window of $L+1$ symbol periods, yielding

$$\mathbf{y}_{L,k} = \mathbf{H}_L \mathbf{b}(k) + \sum_{l=0, l \neq k}^{N-1} \mathbf{H}_L(l|k) \mathbf{b}(l) + \mathbf{v}_{L,k} \quad k = 0, \dots, N-1 \quad (16)$$

where $\mathbf{v}_{L,k} := [\mathbf{v}^T(k) \dots \mathbf{v}^T(k+L)]^T$ is the zero-mean white Gaussian noise component that falls within the sliding window. Its covariance matrix $\mathbf{R}_{v,k} := E\{\mathbf{v}_{L,k}\mathbf{v}_{L,k}^H\}$ is readily available as a $(L+1) \times (L+1)$ sub-block of \mathbf{R}_v . Hence, we have established N reduced-sized sub-models, resulting from overlapping sliding observation windows.

Next, we will show how to utilize the enhanced PDA MUD detector within each reduced-sized sub-model for symbol-by-symbol detection. Focusing on $\mathbf{b}(k)$ in the k th sliding window described by (16), we further dissect the channel matrix \mathbf{H}_L and represent the nonzero (effective) portion of the channel response vector for $b^{(i)}(k)$ as $\mathbf{h}_{L,i}$, which is the i th column of \mathbf{H}_L and is independent of k . Define $\eta_{i,m}(k) := \Pr(b^{(i)}(k) = a_m | \mathbf{y}_{L,k})$, $m = 1, \dots, M$ as the posterior probabilities of the M -ary modulated symbol $b^{(i)}(k)$. The conditional mean and variance of $b^{(i)}(k)$ are then given by

$$\mu_b^{(i)}(k) := E \left\{ b^{(i)}(k) \middle| \mathbf{y}_{L,k} \right\} = \sum_{m=1}^M a_m \eta_{i,m}(k) \quad (17)$$

$$\begin{aligned} \sigma_b^{(i)}(k) &:= \text{var} \left\{ b^{(i)}(k) \middle| \mathbf{y}_{L,k} \right\} \\ &= \sum_{m=1}^M |a_m|^2 \eta_{i,m}(k) - \left| \mu_b^{(i)}(k) \right|^2 \end{aligned} \quad (18)$$

respectively. To make MAP detection on $b^{(i)}(k)$, the task now is to evaluate its posterior probability distribution. We suppose the posterior pdf is Gaussian after the Gaussian forcing approximation, hence, can be fully characterized by its mean and variance conditioned on $\mathbf{y}_{L,k}$.

Defining $\mathbf{v}_L(k) := \sum_{l \neq k} \sum_{j=1}^{N_i} \mathbf{h}_{L,(l|k)}^{(j)} b^{(j)}(l) + \mathbf{v}_{L,k}$ and emphasizing each individual element $b^{(i)}(k)$ in $\mathbf{b}(k)$, we rewrite (16) as

$$\mathbf{y}_{L,k} = \mathbf{h}_{L,i} b^{(i)}(k) + \sum_{j \neq i}^{N_i} \mathbf{h}_{L,j} b^{(j)}(k) + \mathbf{v}_L(k) \quad i = 1, \dots, N_i. \quad (19)$$

Equation (19) follows the same structure as (4); hence, the SDE-1 algorithm can be applied within this local time window for detecting $\mathbf{b}(k)$. To do so, we express the signal component from $b^{(i)}(k)$ by

$$\mathbf{h}_{L,i} b^{(i)}(k) = \mathbf{y}_{L,k} - \sum_{j \neq i}^{N_i} \mathbf{h}_{L,j} b^{(j)}(k) - \mathbf{v}_L(k). \quad (20)$$

The conditional mean and covariance of $\mathbf{h}_{L,i} b^{(i)}(k)$ are thus given by

$$\begin{aligned} \mathbf{g}_k^{(i)} &:= E \left\{ \mathbf{h}_{L,i} b^{(i)}(k) \middle| \mathbf{y}_{L,k} \right\} \\ &= \mathbf{y}_{L,k} - \sum_{j \neq i}^{N_i} \mathbf{h}_{L,j} \mu_b^{(j)}(k) - E \{ \mathbf{v}_L(k) | \mathbf{y}_{L,k} \} \end{aligned} \quad (21)$$

$$\begin{aligned} \mathbf{R}_k^{(i)} &:= \text{cov} \left\{ \mathbf{h}_{L,i} b^{(i)}(k) \middle| \mathbf{y}_{L,k} \right\} \\ &= \sum_{j \neq i}^{N_i} \mathbf{h}_{L,j} \mathbf{h}_{L,j}^H \sigma_b^{(j)}(k) + \text{cov} \{ \mathbf{v}_L(k) | \mathbf{y}_{L,k} \}. \end{aligned} \quad (22)$$

To obtain the conditional mean and covariance of $\mathbf{v}_L(k)$, we adopt the following approximations for all $l \neq k$ and $j \in [1, N_i]$:

$$E \left\{ b^{(j)}(l) \middle| \mathbf{y}_{L,k} \right\} \approx E \left\{ b^{(j)}(l) \middle| \mathbf{y}_{L,l} \right\} = \mu_b^{(j)}(l) \quad (23)$$

$$\text{var} \left\{ b^{(j)}(l) \middle| \mathbf{y}_{L,k} \right\} \approx \text{var} \left\{ b^{(j)}(l) \middle| \mathbf{y}_{L,l} \right\} = \sigma_b^{(j)}(l). \quad (24)$$

As a result, we have

$$E \{ \mathbf{v}_L(k) | \mathbf{y}_{L,k} \} = \sum_{l \neq k} \sum_{j=1}^{N_i} \mathbf{h}_{L,(l|k)}^{(j)} \mu_b^{(j)}(l) \quad (25)$$

$$\begin{aligned} \text{cov} \{ \mathbf{v}_L(k) | \mathbf{y}_{L,k} \} \\ = \sum_{l \neq k} \sum_{j=1}^{N_i} \mathbf{h}_{L,(l|k)}^{(j)} \mathbf{h}_{L,(l|k)}^{(j)H} \sigma_b^{(j)}(l) + \mathbf{R}_{v,k}. \end{aligned} \quad (26)$$

Now, we are ready to use the enhanced PDA framework to compute the soft-detection on $b^{(i)}(k)$. As a result of the Gaussian forcing approximation, the posterior probabilities of $b^{(i)}(k)$ can be obtained as

$$\begin{aligned} \Pr \left(\mathbf{h}_{L,i} b^{(i)}(k) \middle| \mathbf{y}_{L,k} \right) \\ = \frac{1}{(2\pi)^{N_o(L+1)/2} \left| \mathbf{R}_k^{(i)} \right|^{1/2}} \\ \times \exp \left\{ -\frac{1}{2} \left(\mathbf{h}_{L,i} b^{(i)}(k) - \mathbf{g}_k^{(i)} \right)^H \right. \\ \left. \times \mathbf{R}_k^{(i)-1} \left(\mathbf{h}_{L,i} b^{(i)}(k) - \mathbf{g}_k^{(i)} \right) \right\} \end{aligned} \quad (27)$$

where $b^{(i)}(k) \in \{a_1, \dots, a_M\}$. Once the posterior probabilities $\eta_{i,m}(k) = \Pr(\mathbf{h}_{L,i} a_m | \mathbf{y}_{L,k})$, $m = 1, \dots, M$ are established from (27), the MAP detection on $\{b^{(i)}(k)\}_{i=1}^{N_i}$ can be made in the same manner as described in (10) and (11).

2) *Iterative Implementations:* As discussed in SDE-1, the evaluation of posterior probabilities $\{\eta_{i,m}(k)\}_{m=1}^M$ for each input $b^{(i)}(k)$ involves computing $\mathbf{g}_k^{(i)}$ and $\mathbf{R}_k^{(i)-1}$, which are dependent on not only the unknown $\{\eta_{j,m}(k)\}$ of MAI symbols $\{b^{(j)}(k)\}_{j \neq i}$ at the k th sampling time but, in addition, on the unknown $\{\eta_{j,m}(l)\}$ of ISI symbols $\{b^{(j)}(l)\}_{l \neq k, \forall j}$. Algorithm implementation via iterative multistage processing is, thus, in order. Similar to the speed-up strategy described in (14) and (15), we introduce two auxiliary variables that are instrumental to computational saving in updating the $(\mathbf{g}_k^{(i)}, \mathbf{R}_k^{(i)-1})$ pair: One is the conditional mean $\mathbf{g}(k)$ of the noise term $\mathbf{v}_{L,k} = \mathbf{y}_{L,k} - \mathbf{H}_L \mathbf{b}(k) - \sum_{l \neq k} \mathbf{H}_L(l|k) \mathbf{b}(l)$, given by

$$\begin{aligned} \mathbf{g}(k) &:= \mathbf{y}_{L,k} - \sum_{j=1}^{N_i} \mathbf{h}_{L,j} E \left\{ b^{(j)}(k) \middle| \mathbf{y}_{L,k} \right\} \\ &\quad - \sum_{l \neq k} \sum_{j=1}^{N_i} \mathbf{h}_{L,(l|k)}^{(j)} E \left\{ b^{(j)}(l) \middle| \mathbf{y}_{L,k} \right\}. \end{aligned} \quad (28)$$

The other is the conditional covariance matrix $\mathbf{R}(k)$ of $\mathbf{y}_{L,k}$ in the form of

$$\begin{aligned} \mathbf{R}(k) &:= \sum_{j=1}^{N_i} \mathbf{h}_{L,j} \mathbf{h}_{L,j}^H \text{var} \left\{ b^{(j)}(k) \middle| \mathbf{y}_{L,k} \right\} \\ &+ \sum_{l \neq k} \sum_{j=1}^{N_i} \mathbf{h}_{L,(l|k)}^{(j)} \mathbf{h}_{L,(l|k)}^{(j)H} \text{var} \left\{ b^{(j)}(l) \middle| \mathbf{y}_{L,k} \right\} \\ &+ \mathbf{R}_{v,k}. \end{aligned} \quad (29)$$

Both $\mathbf{g}(k)$ and $\mathbf{R}(k)$ take into account of the channel responses to all the elements in $\mathbf{b}(k)$ and all the ISI components as well. For reasons that will be explained in Section III-B3, we term the updating of $\mathbf{g}(k)$ due to the ISI components as soft ISI cancellation and that of $\mathbf{R}(k)$ as noise tracking.

An obvious contrast between (21) and (22) and (28) and (29) shows that

$$\mathbf{g}(k) = \mathbf{g}_k^{(i)} - \mu_b^{(i)}(k) \mathbf{h}_{L,i} \quad (30)$$

$$\mathbf{R}(k) = \mathbf{R}_k^{(i)} + \sigma_b^{(i)}(k) \mathbf{h}_{L,i} \mathbf{h}_{L,i}^H. \quad (31)$$

Therefore, for $i = 1, \dots, N_i$ within each local window k , we have

$$\mathbf{R}_k^{(i-1)} = \mathbf{R}(k)^{-1} + \frac{\sigma_b^{(i)}(k) \mathbf{R}^{-1}(k) \mathbf{h}_{L,i} \mathbf{h}_{L,i}^H \mathbf{R}^{-1}(k)}{1 - \sigma_b^{(i)}(k) \mathbf{h}_{L,i}^H \mathbf{R}^{-1}(k) \mathbf{h}_{L,i}} \quad (32)$$

$$\mathbf{R}^{-1}(k) = \mathbf{R}_k^{(i-1)} - \frac{\sigma_b^{(i)}(k) \mathbf{R}_k^{(i-1)} \mathbf{h}_{L,i} \mathbf{h}_{L,i}^H \mathbf{R}_k^{(i-1)}}{1 + \sigma_b^{(i)}(k) \mathbf{h}_{L,i}^H \mathbf{R}_k^{(i-1)} \mathbf{h}_{L,i}}. \quad (33)$$

During iterative processing, $\mathbf{g}_i(k)$ and $\mathbf{R}_k^{(i-1)}$ are first updated via (30) and (32), using tentative soft information $\mu_b^{(i)}(k)$ and $\sigma_b^{(i)}(k)$ from the previous stage. Subsequently, the posterior pdf of $b^{(i)}(k)$ can be computed from (27), yielding updated soft information $\mu_{b,\text{new}}^{(i)}(k)$ and $\sigma_{b,\text{new}}^{(i)}(k)$ of the current stage. The auxiliary variables $\mathbf{g}(k)$ and $\mathbf{R}(k)$ are then updated via (30) and (33) using the new values. Bear in mind that $\mathbf{y}_{L,k}$ in each sliding window only yields the soft information of one input block $\mathbf{b}(k)$, but $\mathbf{b}(k)$ also affects L previous overlapping windows $\{\mathbf{y}_{L,l}\}_{l=k-L}^{k-1}$ and L future overlapping windows $\{\mathbf{y}_{L,l}\}_{l=k+1}^{k+L}$. Therefore, there are $\min(2L, N-1)$ additional pairs of auxiliary variables $\{\mathbf{g}(l), \mathbf{R}^{-1}(l)\}_{l \neq k}$ that need to be updated from the new estimates of $\mathbf{b}(k)$. The updating of these ISI pairs can be carried out instantaneously when any $\mu_{b,\text{new}}^{(i)}(k)$ and $\sigma_{b,\text{new}}^{(i)}(k)$ become available or after these soft-information values are updated for all $b^{(i)}(k)$, $i \in [1, N_i]$ and $k \in [0, N-1]$, resulting in two implementation procedures of different computational loads.

In the first procedure, we update the related ISI auxiliary pairs whenever $\mu_{b,\text{new}}^{(i)}(k)$ and $\sigma_{b,\text{new}}^{(i)}(k)$ become available for any i and k . Following (28), an auxiliary ISI mean can be updated by

$$\begin{aligned} \mathbf{g}_{\text{new}}(l) &= \mathbf{g}(l) + \mathbf{h}_{L,(k|l)}^{(i)} \mu_b^{(i)}(k) - \mathbf{h}_{L,(k|l)}^{(i)} \mu_{b,\text{new}}^{(i)}(k) \\ l &= k-L, \dots, k-1, k+1, \dots, k+L. \end{aligned} \quad (34)$$

TABLE II
SOFT DECISION EQUALIZATION ALGORITHM: SDE-2 (I)

1. Initialization
$\eta_{i,m}(k) = \frac{1}{M}$, $\mu_b^{(i)}(k) = \frac{1}{M} \sum_{m=1}^M a_m$,
$\sigma_b^{(i)}(k) = \frac{1}{M} \sum_{m=1}^M \left a_m - \mu_b^{(i)}(k) \right ^2$, $\forall i, m, k$;
compute $\mathbf{g}(k)$, $\mathbf{R}(k)$ and $\mathbf{R}(k)^{-1}$ using (28) and (29).
2. Soft-decision iterative updating
Do
for $k = 0 : N-1$ (each input sub-block)
for $i = 1 : N_i$ (each symbol within a sub-block)
Step a: soft-decision MUD on $\mathbf{b}(k)$
– update $\mathbf{g}_k^{(i)}$ and $\mathbf{R}_k^{(i-1)}$ using (30) and (32);
– compute $\eta_{i,m}(k)$ using (27) and (10)-(11);
– update $\mu_b^{(i)}(k)$ and $\sigma_b^{(i)}(k)$ using (17), (18);
Step b: soft ISI cancellation and noise tracking of auxiliary pairs
– update the related $(2L+1)$ auxiliary pairs $\{\mathbf{g}(l), \mathbf{R}^{-1}(l)\}_{l=k-L}^{k+L}$ using (34), (36), (37) for $l \neq k$, and (30), (33) for $l = k$.
end i
end k
Until $\eta_{i,m}(k)$ converges for all i, m , and k
3. Detection: $\hat{b}^{(i)}(k) = \arg \max_{b^{(i)}(k) \in \{a_1, \dots, a_M\}} \eta_{i,m}(k)$

To update $\mathbf{R}_{\text{new}}^{-1}(l)$, $l = k-L, \dots, k+L$, it is observed from (29) that

$$\mathbf{R}_{\text{new}}(l) = \mathbf{R}_i(l|k) + \mathbf{h}_{L,(k|l)}^{(i)} \mathbf{h}_{L,(k|l)}^{(i)H} \sigma_{b,\text{new}}^{(i)}(k) \quad (35)$$

where $\mathbf{R}_i(l|k) := \mathbf{R}(l) - \mathbf{h}_{L,(k|l)}^{(i)} \mathbf{h}_{L,(k|l)}^{(i)H} \sigma_b^{(i)}(k)$ is independent of $b^{(i)}(k)$. Based on (35), we can apply the matrix inversion lemma twice to update $\mathbf{R}_{\text{new}}^{-1}(l)$ from $\mathbf{R}^{-1}(l)$ and $\sigma_{b,\text{new}}^{(i)}(k)$:

$$\begin{aligned} \mathbf{R}_i^{-1}(l|k) &= \mathbf{R}^{-1}(l) \\ &+ \frac{\sigma_b^{(i)}(k) \mathbf{R}^{-1}(l) \mathbf{h}_{L,(k|l)}^{(i)} \mathbf{h}_{L,(k|l)}^{(i)H} \mathbf{R}^{-1}(l)}{1 - \sigma_b^{(i)}(k) \mathbf{h}_{L,(k|l)}^{(i)H} \mathbf{R}^{-1}(l) \mathbf{h}_{L,(k|l)}^{(i)}}. \end{aligned} \quad (36)$$

$$\begin{aligned} \mathbf{R}_{\text{new}}^{-1}(l) &= \mathbf{R}_i^{-1}(l|k) \\ &- \frac{\sigma_{b,\text{new}}^{(i)}(k) \mathbf{R}_i^{-1}(l|k) \mathbf{h}_{L,(k|l)}^{(i)} \mathbf{h}_{L,(k|l)}^{(i)H} \mathbf{R}_i^{-1}(l|k)}{1 + \sigma_{b,\text{new}}^{(i)}(k) \mathbf{h}_{L,(k|l)}^{(i)H} \mathbf{R}_i^{-1}(l|k) \mathbf{h}_{L,(k|l)}^{(i)}}. \end{aligned} \quad (37)$$

The intermediate matrix inverse $\mathbf{R}_i(l|k)$ plays a similar role to the updating of $\mathbf{R}(l)$ as $\mathbf{R}_k^{(i)}$ to $\mathbf{R}(k)$. The computational load of updating all the relevant auxiliary variables for each input is on the order of $\min(N, 2L+1)(L+1)^2 N_o^2$, which represents an approximate L/K reduction compared with the SDE-1 algorithm. The overall algorithm is summarized in Table II.

In the above procedure, N overlapping time windows are processed in serial, resulting in a total of $\bar{N} \times \min(N, 2L+1)$ times updating the auxiliary pairs inside each stage. To reduce the number of covariance matrix inverse to be processed, we

TABLE III
SOFT DECISION EQUALIZATION ALGORITHM: SDE-2 (II)

1. Initialization, same as in Table II
2. Soft-decision iterative updating
Do
Step a: Soft ISI cancellation and noise tracking of N auxiliary pairs
– update all $\{\mathbf{g}(l), \mathbf{R}^{-1}(l)\}_{l=0}^{N-1}$ using (28), (29).
Step b. Soft-decision MUD over N overlapping truncated windows:
for $k = 0 : N - 1$ (each input sub-block)
for $i = 1 : N_i$ (each symbol within a sub-block)
– update $\mathbf{g}_k^{(i)}$ and $\mathbf{R}_k^{(i)-1}$ using (30), (32);
– compute $\eta_{i,m}(k)$ from (27) and (10)–(11);
– update $\mu_b^{(i)}(k)$ and $\sigma_b^{(i)}(k)$ using (17), (18);
– update $\mathbf{g}(k)$ and $\mathbf{R}^{-1}(k)$ using (30), (33).
end i
end k
Until $\eta_{i,m}(k)$ converges for all i, m , and k
3. Detection: $\hat{b}^{(i)}(k) = \arg \max_{b^{(i)}(k) \in \{a_1, \dots, a_M\}} \eta_{i,m}(k)$

may take an alternative procedure to update the auxiliary matrices $\{\mathbf{R}^{-1}(k)\}_{k=0}^{N-1}$. In each stage, we first process the N local windows in parallel to update the soft decisions of all inputs $\{\mathbf{b}(k)\}_{k=0}^{N-1}$, using tentative decisions and auxiliary variables from the previous stage. When all the new soft-information values are available, we recalculate the N auxiliary pairs $\{\mathbf{g}(k), \mathbf{R}(k)\}_{k=0}^{N-1}$, possibly by using their definitions in (28) and (29). On average, there are only N pairs of auxiliary variables to be updated inside each stage, and the complexity order per symbol is the smaller of $\min(N, 2L+1)(L+1)^2 N_o^2$ (when the matrix inversion lemma is used) and $(L+1)^3 N_o^3 / N_i$ (when direct matrix inversion is used). This procedure is summarized in Table III. The implementations in both Tables II and III are expected to converge to the SDE-1 algorithm, which have been verified in our simulations. The second procedure offers a complexity advantage when the channel length is much less than the frame size, i.e., $L \ll N$.

3) *Comparisons With Existing Algorithms:* The SDE-2 MIMO equalizer resembles a concatenation of a series of enhanced PDA/SDE-1 detectors, each operating on a truncated sub-model to reduce the overall complexity. It is worth emphasizing that the pdf estimators in (9) for SDE-1 and in (27) for SDE-2 are equivalent when both converge to the steady state. The only approximations involved are (23) and (24), which, if at the steady state, does not incur performance loss. The multistage iterative processing nature of the SDE algorithms prompts their links with the PIC [22] and the soft-input-soft-output (SISO) MMSE detector that was developed for turbo detection. Next, we compare SDE with existing interference cancellation techniques.

a) *Comparison With Hard-Decision PIC:* In each truncated output vector $\mathbf{y}_{L,k}$, $\{\mathbf{b}(l)\}_{l \neq k}$ can be viewed as the interference to the desired symbol block $\mathbf{b}(k)$, and the sub-model (19) can be written as $\mathbf{y}_{L,k} - \sum_{l \neq k} \mathbf{H}_L(l|k) \mathbf{b}(l) = \sum_{i=1}^{N_i} \mathbf{h}_{L,i} b^{(i)}(k) + \mathbf{v}_{L,k}$. This appears to be deduced from (16) using the PIC structure [22]. A close examination shows that the

SDE-2 method treats the tentative decisions on $\mathbf{b}(l)$ differently from PIC. An equivalent model for SDE-2 is described as

$$\tilde{\mathbf{y}}_{L,k} = \sum_{i=1}^{N_i} \mathbf{h}_j b^{(i)}(k) + \tilde{\mathbf{v}}_{L,k} \quad (38)$$

where the observation vector is modified to $\tilde{\mathbf{y}}_{L,k}$ in the form of

$$\tilde{\mathbf{y}}_{L,k} := \mathbf{y}_{L,k} - \sum_{l \neq k} \mathbf{H}_L(l|k) E\{\mathbf{b}(l) | \mathbf{y}_{L,k}\}. \quad (39)$$

The noise term $\tilde{\mathbf{v}}_{L,k}$ is independent of $\mathbf{b}(k)$ and is assumed to be Gaussian with zero-mean and covariance $\tilde{\mathbf{R}}_{v,k}$ being equal to $\text{cov}\{\mathbf{v}_L(k) | \mathbf{y}_{L,k}\}$ in (26), i.e.,

$$\tilde{\mathbf{R}}_{v,k} := \mathbf{R}_{v,k} + \sum_{l \neq k} \sum_{j=1}^{N_i} \mathbf{h}_{L,(l|k)}^{(j)} \mathbf{h}_{L,(l|k)}^{(j)H} \sigma_b^{(j)}(l). \quad (40)$$

The efficacy of (38) can be established by the fact that both $\mathbf{g}_k^{(i)}$ and $\mathbf{R}_k^{(i)}$ in (21) and (22) that are required to fully characterize the posterior pdf of $b^{(i)}(k)$ can be equivalently obtained from (38)–(40). Compared with PIC, our SDE-2 method entails three major differences.

- i) The finite-alphabet ISI symbols $\mathbf{b}(l)$ are cancelled out by their soft-decision alternative $E\{\mathbf{b}(l) | \mathbf{y}_{L,k}\}$ instead of tentative hard decisions.
- ii) In addition to the soft-decision interference cancellation, the conditional variances of the soft estimates are tracked and lumped into the variance of noise $\tilde{\mathbf{v}}_{v,k}$, as seen in (40). In contrast, hard-decision PIC does not change the statistics of the noise term $\mathbf{v}_{L,k}$ during iterations. Due to this key noise tracking step, the formulation in (38) retains optimality subject to the Gaussian forcing approximation, whereas the conventional PIC is suboptimum.
- iii) The SDE-2 method performs interference cancellation along overlapping blocks, using a sliding window of size $(L+1)N_o$ outputs, such that all the observations related to each input are retained within the corresponding window. The conventional PIC receiver, on the other hand, operates on a nonoverlapping sliding window of size 1.

b) *Comparisons Between SDE, PDA, and SISO-MMSE-Based Turbo Detection:* The soft-decision iterative processing nature of our SDE algorithms prompts their links to turbo signal processing. In turbo detection [9], soft information in the form of log-likelihood ratios (LLR) is exchanged; interchangeably, the PDA, SDE-1, and SDE-2 methods iteratively feed back the means and variances of Gaussian distributed random symbols. We now compare the multistage Gaussian forcing principle used in PDA, SDE-1, and SDE-2 with the SISO-MMSE-based turbo principle for uncoded systems. For different algorithms, we will explain the posterior probabilities derived for b_i in the general model $\mathbf{y} = \mathbf{h}_i b_i + \sum_{j \neq i} \mathbf{h}_j b_j + \mathbf{v}$. The comparison will be based on binary modulation.

In SDE-1 for binary signaling, the posterior probability distribution of b_i in (9) is reduced to

$$\frac{\Pr(b_i = 1 | \mathbf{y})}{\Pr(b_i = -1 | \mathbf{y})} = \exp\{2\mathbf{h}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i\} \quad (41)$$

where \mathbf{g}_i and \mathbf{R}_i are given by (5) and (6), respectively.

The PDA algorithm focuses on MUD in synchronous CDMA multiple access, where \mathbf{H} is a real-valued, square cross-correlation matrix, and the noise variance is $\mathbf{R}_v = \sigma_n^2 \mathbf{H}$. Let $\mathcal{N}(\mu, \sigma^2)$ represent a Gaussian random variable with mean μ and variance σ^2 . Under the special system setup, the PDA MUD establishes a Gaussian model for b_i in the form of [14]

$$\mathbf{e}_i b_i = \bar{\mathbf{g}}_i + \mathcal{N}(0, \bar{\mathbf{R}}_i) \quad (42)$$

where $\bar{\mathbf{g}}_i = \mathbf{y} - \sum_{j \neq i} E\{b_j | \mathbf{y}\} \mathbf{e}_j$, $\bar{\mathbf{R}}_i = \sum_{j \neq i} \mathbf{e}_j \mathbf{e}_j^T \text{var}\{b_j | \mathbf{y}\} + \sigma_n^2 \mathbf{H}^{-1}$, and \mathbf{e}_i is a column vector whose i th element is 1, whereas all other components are 0. Premultiplying $\mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1}$ on both sides of (42) yields

$$\mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \mathbf{e}_i b_i = \mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \bar{\mathbf{g}}_i + \mathcal{N}(0, \mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \mathbf{e}_i). \quad (43)$$

The soft decision made by PDA is thus given by

$$\frac{\Pr(b_i = 1 | \mathbf{y})}{\Pr(b_i = -1 | \mathbf{y})} = \exp \{2 \mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \bar{\mathbf{g}}_i\}. \quad (44)$$

Using the equality $\mathbf{h}_i = \mathbf{H} \mathbf{e}_i$, it can be established that $\mathbf{h}_i^H \bar{\mathbf{R}}_i^{-1} \bar{\mathbf{g}}_i = \mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \bar{\mathbf{g}}_i$. Therefore, the PDA result in (44) is the same as that of SDE-1 in (41) in this special case. As an enhancement to PDA, our SDE-1 algorithm does not restrict \mathbf{R}_v to be proportional to \mathbf{H}^{-1} , and it applies even when \mathbf{H}^{-1} does not exist.

The SISO-MMSE method generates an MMSE filtered decision statistic z_i for estimating b_i , where z_i boils down to of [9, (41) and (50)]

$$z_k = \mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \bar{\mathbf{g}}_i = \mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \mathbf{e}_i b_i + \mathcal{N}(0, \mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \mathbf{e}_i - \mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \mathbf{e}_i \mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \mathbf{e}_i). \quad (45)$$

The LLR of b_i is thus given by

$$\frac{p(\mathbf{y} | b_i = 1)}{p(\mathbf{y} | b_i = -1)} = \exp \left\{ \frac{2 \mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \bar{\mathbf{g}}_i}{1 - \mathbf{e}_i^T \bar{\mathbf{R}}_i^{-1} \mathbf{e}_i} \right\}. \quad (46)$$

The comparison between SISO-MMSE and PDA (a special case of SDE-1) is clearly illustrated by the similarities and differences between (43) and (44), and (45) and (46).

As to the SDE-2 method, it is specially tailored to the equalization problem with a Toeplitz channel structure; therefore, it is not directly comparable with the existing turbo detectors. Interestingly, the iterative processing in SDE-2 suggests the flowchart in Fig. 2, which interprets the algorithm by a turbo structure in which two major function blocks (MUD and ISI cancellation) exchange information in an iterative manner. SDE-2 builds on the Gaussian forcing idea and incorporates sliding windowing to reduce the overall complexity without sacrificing the detection performance. The key to retaining optimality after data truncation is to carry out the noise tracking step (40) prior to each reduced-dimension local SDE-1 detection on (38). Noise tracking via Kalman filtering has appeared in the context of PDA detection for asynchronous CDMA under frequency flat fading [15]. Such a PDA-Kalman tracker cannot be generalized to track the noise in a channel with a memory length $L > 2$ since the underlying dynamic model is no longer first-order, thus obviating Kalman filtering. In our development, we interpret noise tracking as a means to update the variance of the ISI estimates. This viewpoint allows us to generalize the noise

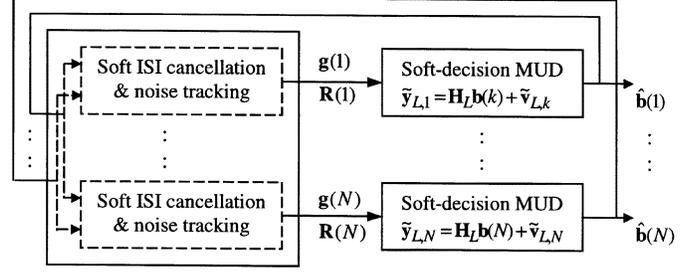


Fig. 2. Turbo-like flow chart of the SDE-2 algorithm.

tracking method easily to channels with long memory length, and there is no need to perform channel Cholesky decomposition and Kalman filtering, as in [15].

IV. ALGORITHM EVALUATION AND SIMULATIONS

In this section, we investigate the characteristics of our two soft-decision equalization methods through computer simulations. In both the SDE-1 and SDE-2 algorithms, the soft decisions are derived to converge to the sequence MAP estimates through multistage iterations; therefore, close-to-optimal symbol detection performance is anticipated. On the other hand, Gaussian forcing explains the low-complexity feature of these SDE methods. These claims will be verified here by comparisons with other competing methods, including the optimum ML detection by brute-force enumeration, quasi-ML by sphere decoding (SD) [12], [20], and the suboptimal hard-decision MIMO FIR MMSE-DFE method [6]. Performance metrics of interest are the BER performance in both full column-rank and rank-deficient channels, as well as the computational complexity in terms of the number of operations versus the frame data size N .

A. BER Performance in Full-Rank MIMO Channels

In the simulated MIMO system, each input-output radio link is generated independently from the broadband wireless high performance European radio LAN (HIPERLAN) model [23], [24]. The channels are complex-valued, and the noise is assumed to be complex white Gaussian. The time-varying FIR channels are generated according to the channel model A specified by ETSI for HiperLAN/2 [23], resulting in a maximum channel memory length of $L = 8$ symbols. Each channel tap varies according to Jakes' model with a maximum Doppler frequency of 52 Hz, corresponding to a typical terminal speed and a carrier frequency of 5.2 GHz.

We study the BER performance versus the signal-to-noise ratio (SNR) of various detectors under different modulation schemes and numbers of antennas. The total transmit power is held constant, irrespective of the number of transmit antennas. For each given SNR, the simulation keeps running until the number of errors for the (near-optimum) sphere decoding algorithm reaches 100 or greater. With this number of errors, the simulated BER is within $\pm 20\%$ of the true BER.

We start with the case of more number of receive antennas than transmit antennas, i.e., $N_o > N_i$. Due to the block transmission structure with a transmit redundancy of L padded zeros, the Toeplitz channel matrix \mathbf{H} in (2) has a full column rank

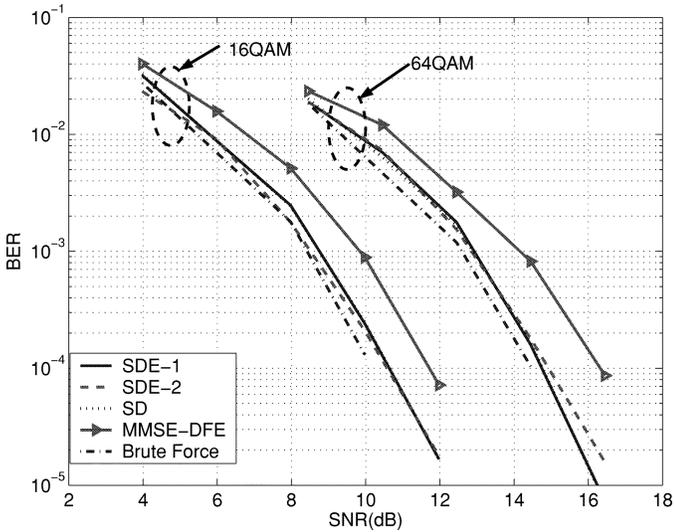


Fig. 3. Performance comparison under 16/64QAM, $N_i = 1$, $N_o = 4$.

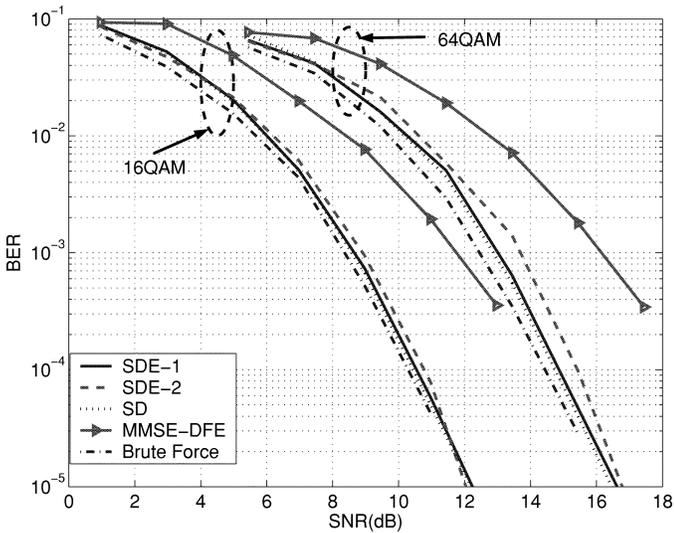


Fig. 4. Performance comparison under 16/64QAM, $N_i = 2$, $N_o = 4$.

of $\bar{N} = NN_i$ and is guaranteed invertibility, irrespective of channel nulls [19].

Fig. 3 illustrates the performance comparison of SDE-1, SDE-2, SD, FIR MMSE-DFE, and ML (by brute-force search) for $N_i = 1$ and $N_o = 4$. The symbol block size N is chosen as 8. The results for both 16- and 64-QAM are presented. The same MIMO setup is considered in Fig. 4, except that the number of transmit antennas is increased to $N_i = 2$. In both figures, it can be seen that the BER curves of SDE-1, SDE-2, and SD are nearly identical for different high-bandwidth-efficiency modulation schemes. They all approach that of the optimum ML detection. This corroborates the near-optimum property of the reduced-complexity SDE-2 technique. FIR MMSE-DFE, however, experiences nontrivial performance degradation in all the above scenarios. The performance gap in the $N_i = 2$ case is more pronounced than that in the $N_i = 1$ case. As N_i increases, the information-theoretic capacity is expected to grow linearly in N_i , given $N_i \leq N_o$ [1]. MMSE-DFE cannot deliver the desired performance as capacity-driven MIMO systems exploit

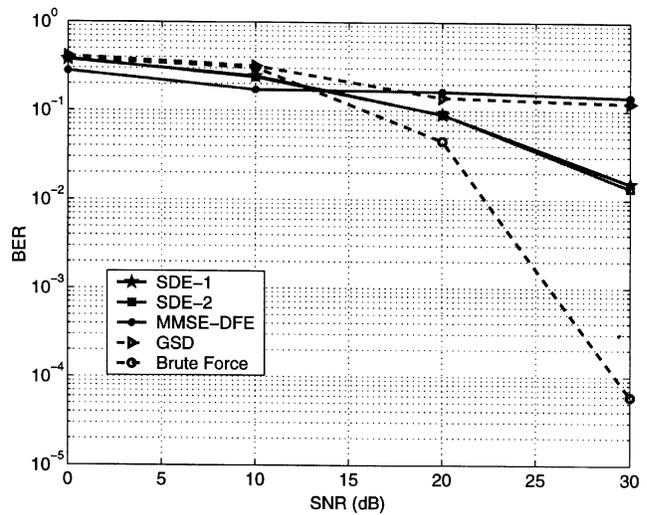


Fig. 5. Performance comparison in fat channel case, $N_i = 3$, $N_o = 1$.

more transmit antennas. On the other hand, our SDE methods, with their near-optimum performance, are very promising candidates to bring the potency of MIMO systems to practice.

B. BER Performance in Rank-Deficient MIMO Channels

Rank-deficient MIMO channels exist in many wireless scenarios, such as in mobile downlink transmission, where there is typically more transmit than receive antennas. A so-called fat channel matrix \mathbf{H} arises when $KN_o < NN_i$, which means that \mathbf{H} has more columns than rows, and its pseudo-inverse \mathbf{H}^\dagger no longer exists. This poses a significant challenge for symbol detection since a $NN_i \times 1$ input data vector is projected onto an output/observation space of a smaller dimension KN_o [20]. Such a channel is identifiable only when each distinct finite-alphabet input \mathbf{b} can be mapped into a distinct and resolvable output \mathbf{y} when free of noise. Even when the system identifiability condition is satisfied, a rank-deficient channel is difficult to process. First, many detection techniques that require channel invertibility do not apply. This includes the linear zero-forcing and MMSE detectors [22] and the original PDA MUD filter [14]. Second, even when a detector does not face implementation difficulty, its detection performance may exhibit an unacceptably large noise floor. Examples include FIR MMSE-DFE and SD. In MMSE-DFE, a fat channel matrix severely reduces the power efficiency of the feedforward filter, which in turn compromises ISI cancellation in the feedback filter design. The original SD judiciously uses the lattice structure of the finite-alphabet input data to perform quasi-ML search at a low complexity. Unfortunately, such a lattice search is infeasible for $NN_i > KN_o$ [12], and the generalized SD (GSD) does not preserve optimality due to a reduced-dimension lattice projection [20].

To investigate the behavior of our SDE methods in the fat-channel case, we choose a MIMO setup with $N_o = 1$ receive antenna, $N_i = 3$ transmit antennas, and a block size of $N = 5$. The corresponding channel matrix \mathbf{H} is thus 13×15 in dimension. In Fig. 5, the performance of channel equalization by brute-force ML is plotted as a baseline, along that of SDE-1, SDE-2, GSD, and MMSE-DFE. The BER values for

both MMSE-DFE and GSD stay above 10^{-1} , even for high SNR. The SDE methods also incur considerable performance degradation compared with the optimum ML but do not seem to exhibit an error floor. Intuitively, soft-decision-based methods with Gaussian forcing track the composite covariance of MAI components as noise. Even when there is a rank reduction, or some MAI components are too close in the signal space, the composite noise effect could still retain full rank under the ill-conditioned channel, thus leading to convergence in symbol detection. The performance of SDE in this case could be potentially improved by input ordering inside the iterative process. Since soft decisions are updated sequentially based on tentative decisions made previously, a good input-ordering mechanism, which starts the soft-decision estimation from stronger and well-identifiable inputs to weaker and less-separable inputs, can be expected to enhance the detection performance, possibly to near-optimum. Finding effective input-ordering methods for SDE is one of our future research topics. Other prominent solutions to tackle rank-deficient channels include complex-field transmitter precoding, which will not be discussed here, as we focus on channel equalization at the receiver end.

C. Complexity Evaluation

As we explained in Section II, the input size $\bar{N} = NN_i$ in a MIMO system may come from multiple access and/or multiple antennas; therefore, it is potentially very large for a high-capacity MIMO system. The computational load of the optimum ML detection is $\mathcal{O}(M^{\bar{N}})$, where M is the alphabet size of the input data. Such complexity is infeasible for a high-capacity (large \bar{N}), high-throughput (large M) system, which motivates the search for near-optimal, low-complexity symbol detection and channel equalization solutions.

Quasi-ML by sphere decoding entails polynomial complexity on the order of $\mathcal{O}((\bar{N})^2 + (1 + (\bar{N} - 1)/(4dC))^{4dC})$, [12], where d^{-1} is a lower bound for the eigenvalues of the Gram matrix $\mathbf{G} := \mathbf{H}^T \mathbf{H}$, and C is the square of the initial searching radius. Choosing a large value for C improves the BER performance but also incurs higher complexity. Typically, close-to-optimal performance can be achieved at a polynomial complexity index between 3–6.

The SDE algorithms are iterative routines that compute the posterior probabilities of each input symbol in a sequential fashion. The number of iterations required for convergence varies from one input to another. Based on our MIMO setups, we have observed from simulations that the posterior probabilities typically converge in three to five iterations for higher SNR (>10 dB), and in seven to 14 iterations for lower SNR (<10 dB). In each iteration, the computational load is mainly composed of two parts: One results from computing $\mathbf{R}_k^{(i)-1}$ and the other from evaluating the posterior probabilities using (9) or (27). If we define m as the sliding window length in terms of the number of sub-blocks, then the number of symbols within the window is defined by $\bar{m} := mN_o$. The \bar{m} value is also the size of the covariance matrix $\mathbf{R}_k^{(i)}$. In the following discussions, we evaluate the performance-complexity tradeoff in choosing the window size \bar{m} .

1) *Full-Size Windowing* $\bar{m} = KN_o$: This case corresponds to the SDE-1 algorithm. There is only one full-size window;

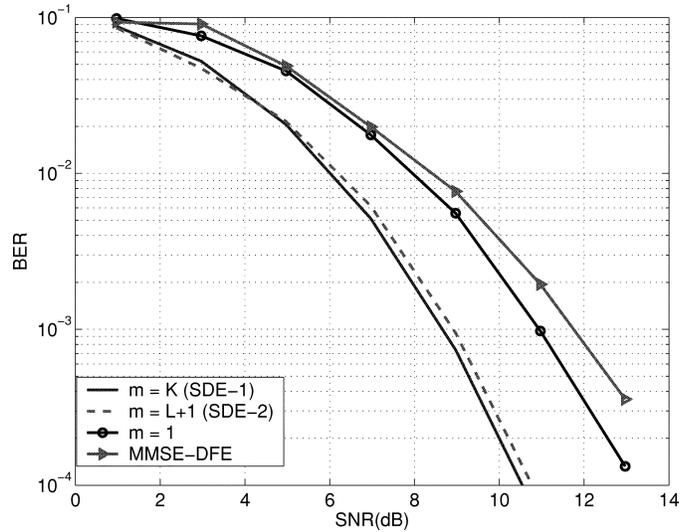
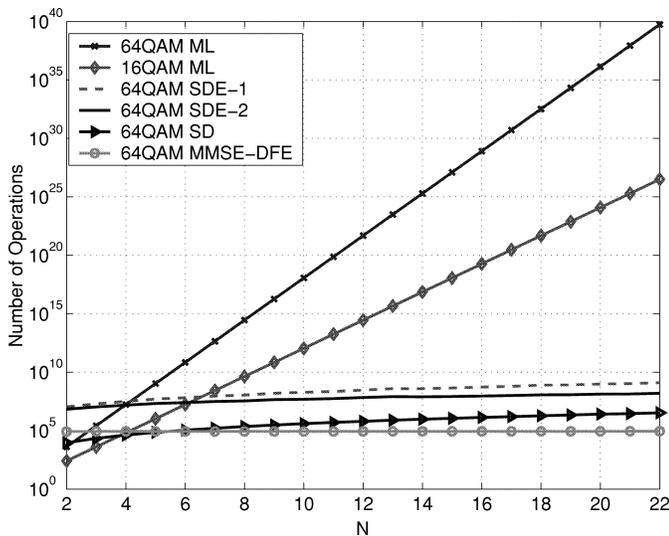
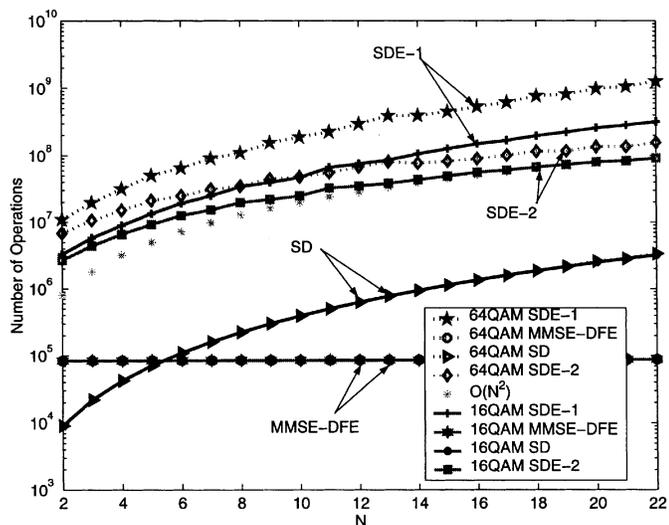


Fig. 6. Performance versus m .

therefore, all the output sub-blocks are processed simultaneously. The dimension of the covariance matrix $\mathbf{R}_k^{(i)}$ and the auxiliary matrix \mathbf{R}_k are both determined by \bar{m} . Because the matrix inversion lemma is used, the complexity of computing $\mathbf{R}_k^{(i)-1}$ is $\mathcal{O}(\bar{m}^2)$. The complexity involved in evaluating M posterior probabilities is $\mathcal{O}(M\bar{m}^2)$ per symbol. Hence, the overall complexity per symbol is on the order of $\mathcal{O}(M\bar{m}^2)$. The complexity for detecting \bar{N} symbols in one iteration is then given by $\mathcal{O}(M\bar{m}^2\bar{N})$. Noting $K = N + L$, the complexity of SDE-1 per symbol is on the second-order polynomial in the input size and is only linear to the constellation size M .

2) *Optimum Window Size* $\bar{m} = (L + 1)N_o$: This case corresponds to the SDE-2 algorithm. SDE-2 takes advantage of the sparse Toeplitz structure of the channel matrix to reduce the equalization complexity. The sliding window only contains the nonzero part of the channel response vector. A symbol can at most affect $L + 1$ output blocks in a L -memory channel. These $L + 1$ output blocks form the sufficient statistics of each symbol. As a result, SDE-2 can retain the near-optimum performance and, at the same time, save the computational cost. In SDE-2, computing $\mathbf{R}_k^{(i)}$ and updating the auxiliary matrix \mathbf{R}_k costs $\mathcal{O}(N\bar{m}^2)$. Finding the pdfs of each symbol takes $\mathcal{O}(M\bar{m}^2)$ operations. The overall complexity per symbol is then $\mathcal{O}(\max(N, M)\bar{m}^2)$, and the complexity per iteration is $\mathcal{O}(\max(N, M)\bar{N}\bar{m}^2)$. Compared with SDE-1, the dimension of each conditional covariance matrix $\mathbf{R}_k^{(i)}$ is reduced from KN_o (corresponding to \mathbf{R}_k in SDE-1) to $N_o(L + 1)$, thus lowering the complexity order of N from 2 to 1 per input. More impressively, such a complexity reduction does not induce noticeable BER performance loss.

3) *Suboptimum Window Size* $\bar{m} < (L + 1)N_o$: The implementation procedures for SDE-2 can be used when the window size $\bar{m} < (L + 1)N_o$. As \bar{m} shrinks, the complexity $\mathcal{O}(\max(N, M)\bar{m}^2)$ decreases. However, since the sliding window only covers a portion of the channel response for the intended input, the algorithm does not make full use of the sufficient statistics, leading to performance degradation. In Fig. 6, we plot the BER curve at a sliding window size of $m = 1$, along

Fig. 7. Complexity versus N .Fig. 8. Complexity versus N .

with SDE-1 ($m = K$) and SDE-2 ($m = L + 1$). It is shown that the small window size yields inferior performance to the other two near-optimum schemes. In fact, $(L + 1)N_o$ is the optimal window length because a longer length does not render better performance, but the complexity increases, whereas a shorter length sacrifices the performance. When $m = 1$, the algorithm becomes a symbol-by-symbol equalization technique, and therefore, the performance cannot match up to that of sequence detection in ISI channels.

The complexity evaluation results are further verified by simulations in Figs. 7 and 8. The system parameters are set to $N_i = 1$, $N_o = 4$, $L = 8$, and $M = 16$ and 64 . The computational load in terms of the number of operations versus the data block size N is depicted for each detection method in Fig. 7. Our SDE methods, along with SD and MMSE-DFE, avoid asymptotic computational explosion suffered by the brute-force ML algorithm at a large data size. More detailed comparison of these low-complexity algorithms are illustrated in Fig. 8. In this simulation setting, the sphere decoder has the same third-order complexity in N as the SDE-1 algorithm. This is not always the case,

as the complexity of SD could be higher if the Gram matrix \mathbf{G} has very small eigenvalues, and the search radius C is chosen to be large. The overall complexity of the SDE-2 algorithm is between SD and SDE-1, but asymptotically, its complexity order in N is only 2 instead of 3, as witnessed by its close match with the function N^2 at large N . Such a reduction in the complexity order will pay off for high-capacity MIMO systems.

V. SUMMARY

Soft-decision based equalization techniques have been developed in this paper for frequency selective MIMO multipath channels. Relying on iterative posterior probability updating and PDA-type Gaussian forcing, the proposed SDE algorithms attain remarkable near-ML performance at low complexity that is polynomial (on the third-order) in the input and output sizes and linear in the modulation constellation size. Unlike existing fast MUD algorithms, our development for MIMO channel equalization relies on zero-padded block transmission to enable block detection for a sequence detection problem and capitalizes on the distinct Toeplitz channel structure to simplify the equalization complexity. Near-optimum symbol detection in the presence of channel memory is attained by virtue of soft-decision MAP multiuser detection, multistage ISI cancellation, and implicit noise tracking. Our algorithms also apply to rank-deficient channels, provided that the channels are identifiable for the signal constellation. This work not only establishes the soft-decision approach as an attractive candidate for near-optimum channel equalization and symbol detection for practical MIMO systems but also enhances the applicability of the original PDA filter to a generic integer-LS problem that is omni-present in many signal processing and wireless communications applications.

REFERENCES

- [1] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, no. 6, pp. 315–335, 1998.
- [2] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.
- [3] M. V. Eyuboglu and S. U. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. 36, pp. 12–20, Jan. 1988.
- [4] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. Commun.*, vol. 37, pp. 428–436, May 1989.
- [5] A. Duel-Hallen, "Equalizers for multiple-input multiple-output channels and PAM systems with cyclostationary input sequences," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 630–639, Sept. 1992.
- [6] N. Al-Dhahir and A. H. Sayed, "The finite-length multi-input multi-output MMSE-DFE," *IEEE Trans. Signal Processing*, vol. 48, pp. 2921–2936, Oct. 2000.
- [7] A. Lozano and C. Papadias, "Layered space-time receivers for frequency selective wireless channels," *IEEE Trans. Commun.*, vol. 50, pp. 65–73, Jan. 2002.
- [8] C. Kominakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Processing*, vol. 50, pp. 1065–1076, May 2002.
- [9] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, pp. 1046–1061, July 1999.
- [10] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, vol. 50, pp. 754–767, May 2002.

- [11] G. Bauch and N. Al-Dahir, "Reduced-complexity space-time turbo-equalization for frequency selective MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 819–828, Oct. 2002.
- [12] U. Finkce and M. Pohst, "Improved methods for calculating vectors of short length in lattice, including a complexity analysis," *Math. Comput.*, vol. 44, pp. 463–471, Apr. 1985.
- [13] W. K. Ma, T. N. Davidson, K. M. Wong, Z. Q. Luo, and P. C. Ching, "Quasimaximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Trans. Signal Processing*, vol. 50, pp. 912–922, Apr. 2002.
- [14] J. Luo, K. R. Pattipati, P. K. Willett, and F. Hasegawa, "Near-optimal multiuser detection in synchronous CDMA using probabilistic data association," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 361–363, 2001.
- [15] D. Pham, J. Luo, K. R. Pattipati, and P. K. Willett, "A PDA-Kalman approach to multiuser detection in asynchronous CDMA," *IEEE Commun. Lett.*, vol. 6, no. 11, pp. 475–477, 2002.
- [16] Y. Bar-Shalom and X. R. Li, *Estimation and Tracking: Principles, Techniques, and Software*. Norwell, MA: Artech House, 1993.
- [17] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 35, no. 1, pp. 14–15, 1999.
- [18] Z. Tian, "Soft decision feedback equalization for pre-coded FIR-MIMO systems," in *Proc. 2nd IEEE Sensor Array Multichannel Signal Process. Workshop*, Rosslyn, VA, Aug. 2002.
- [19] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications—Where Fourier meets Shannon," *IEEE Signal Processing Mag.*, vol. 17, pp. 29–48, May 2000.
- [20] M. O. Damen, K. Abed-Meraim, and J.-C. Belfiore, "A generalized lattice decoder for asymmetrical space-time communication architecture," in *Proc. ICASSP*, 2000, pp. 2581–2584.
- [21] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.
- [22] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [23] *Channel Models for HIPERLAN/2 in Different Indoor Scenarios*, 1998.
- [24] R. D. J. van Nee, G. A. Awater, M. Morikura, H. Takanashi, M. A. Webster, and K. W. Halford, "New high-rate wireless LAN standards," *IEEE Commun. Mag.*, vol. 37, pp. 82–88, Dec. 1999.



Shoumin Liu (S'02) received the B.S. degree in electrical engineering and the M.S. degree in computer engineering from Harbin Institute of Technology, Harbin, China, in 1993 and 1996, respectively. He is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering at Michigan Technological University, Houghton.

His research interests include multiple-input multiple-output communication systems, soft-input soft-output decoding, and near-optimum soft-decision multiuser detection and equalization.



Zhi Tian (M'98) received the B.E. degree in electrical engineering (automation) from the University of Science and Technology of China, Hefei, in 1994 and the M.S. and Ph.D. degrees from George Mason University, Fairfax, VA, in 1998 and 2000, respectively.

From 1995 to 2000, she was a graduate research assistant with the Center of Excellence in Command, Control, Communications, and Intelligence (C3I), George Mason University. Since August 2000, she has been an Assistant Professor with the Department of Electrical and Computer Engineering, Michigan Technological University, Houghton. Her current research focuses on signal processing for wireless communications. She has worked in the areas of digital and wireless communications, statistical array and signal processing, target tracking and data fusion, decision networks, and information fusion.

Dr. Tian serves as an Associate Editor for *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*. She is the recipient of a National Science Foundation CAREER award.