Reduced-State MIMO Sequence Detection With Application to EDGE Systems

Jianzhong Zhang, *Member, IEEE*, Akbar M. Sayeed, *Senior Member, IEEE*, and Barry D. Van Veen, *Fellow, IEEE*

Abstract—In this paper we propose a joint reduced-state sequence detector (JRSSD) for a multiple-input multiple-output (MIMO) system. The proposed JRSSD incorporates the set-partitioning principle to obtain a reduced-state trellis and is the space-time extension of the RSSD proposed in [1] for a single-input single-output (SISO) equalization problem. We show that two elements are essential in achieving the desired complexity and performance tradeoff for the proposed JRSSD algorithm: 1) a proper multisymbol set-partition and 2) an efficient space-time structure that effectively decouples the spatial and temporal processing. We propose a simple multisymbol uniform set-partition (USP) that retains certain geometric symmetry in the partitioned subsets. We also develop two suboptimal algorithms, namely the decorrelating (DC) and ordered successive (OS) algorithms, to decouple the spatial multisymbol detection problem inherent in the processing of parallel transitions, into single-symbol detection problems. The symmetry in USP, together with the DC or OS algorithm, guarantees efficient processing of parallel transitions and leads to a low-complexity JRSSD. Furthermore, numerical simulations show that the performance of JRSSD is near that of the more complex delayed decision feedback sequence estimator (DDFSE).

Index Terms—Delayed decision feedback sequence estimator (DDFSE), DFE, enhanced data rates for GSM evolution (EDGE), multiple-input multiple-output (MIMO), minimum mean square error (MMSE), prefilter, receiver, reduced-state sequence detector (RSSD), reduced-state sequence estimator (RSSE), set partitioning.

I. INTRODUCTION

THE throughput of a M transmit, N receiver antenna multiple-input multiple-output (MIMO) system can be greatly improved by transmitting independent data streams across M different transmit antennas. However, as shown in Fig. 1, multiple access interference (MAI) exists at the receiver since each receive antenna receives signals from all the transmit antennas. This self-introduced MAI complicates the MIMO receiver design, especially in a system where the propagation channel is frequency selective and inter-symbol interference (ISI) is present, such as the enhanced data rates for GSM evolution (EDGE) system [2]. For example, if the vector Viterbi algorithm (VVA) [3] is applied to jointly estimate the
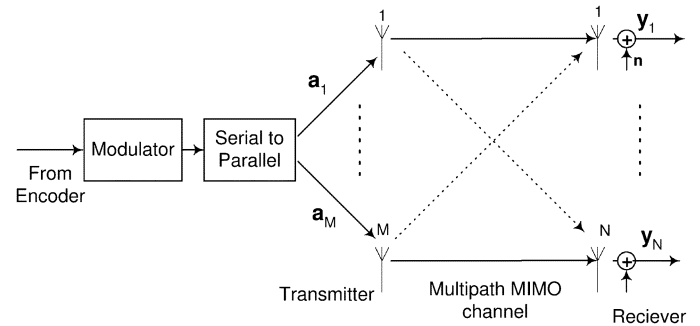
J. Zhang is with the Nokia Research Center, Irving, TX 75039 USA (e-mail: charlie.zhang@nokia.com).

A. M. Sayeed and B. D. Van Veen are with the Electrical and Computer Engineering Department, University of Wisconsin, Madison, WI 53706 USA (e-mail: akbar@ece.wisc.edu; vanveen@engr.wisc.edu).

Fig. 1.   Air interface of a M transmit, N receive antenna MIMO system.

M data sequences in an EDGE system with 8 PSK modulation, the receiver complexity is $\mathcal{O}(8^{M(L+1)})$ where L is the channel memory length. This complexity is far beyond current practical implementation limits.

The minimum mean square error—delayed decision feedback sequence estimator (MMSE-DDFSE) is a well-studied structure for reducing the complexity of the ISI equalization problem in a single-input single-output (SISO) system [4], [5]. The extension of this structure to a MIMO receiver is found in [6], [7]. The MIMO MMSE-DDFSE receiver includes a prefilter and a DDFSE equalizer (which includes a delayed feedback filter and a VVA joint sequence estimator), as shown in Fig. 2. The prefilter attempts to reshape and shorten the channel impulse response by shifting most of the channel energy into the first $L_S + 1$ taps. Consequently, the VVA only needs to equalize the first $L_S + 1$ taps since the remaining low-energy channel taps are accounted for by the feedback filter. Therefore, the overall MIMO receiver complexity is reduced to $\mathcal{O}(8^{M(L_S+1)})$. In order to further reduce the complexity, the optimal joint sequence detector, the VVA, can be replaced by suboptimal MIMO sequence detectors such as the partitioned Viterbi algorithm (PVA) [8] or the decoupled Viterbi algorithm (DVA) [9] where the M sequences are estimated separately. While operating at a much lower complexity, the performance penalty is significant for these suboptimal approaches.

In this paper, we propose another suboptimal MIMO sequence detector, namely the set-partition based joint reduced-state sequence detector (JRSSD), to replace the optimal VVA in the MMSE-DDFSE receiver shown in Fig. 2. We show that a JRSSD provides near-optimal performance at very low complexity, and is a space-time extension of a SISO RSSD,[1] which is a systematic complexity reduction approach first proposed in [1] for a temporal ISI equalization problem. Fur-

[1]Note that in the original paper it was called a reduced-state sequence estimator (RSSE).

thermore, we show that in the proposed JRSSD, two elements are essential in achieving the desired complexity and performance tradeoff: 1) a proper multisymbol set-partition; and 2) a suboptimal algorithm solving the multisymbol detection problem associated with each edge in the reduced-state trellis.

The maximum minimum square subset distance (MSSD) partition [10] is known to be geometrically symmetric for single-symbol set-partitions and, therefore, allows a simple threshold detector within a subset, which is a key aspect of overall complexity reduction [1]. The extension of the maximum MSSD criteria to the multisymbol set-partitioning results in the multidimensional set partitioning (MDSP) [11], [12], where the partition is performed on the vectors that are Cartesian products of multiple symbols. However, an MDSP partition is not desirable in a JRSSD since it does not retain the aforementioned geometric symmetry. Instead, we propose an alternative multisymbol set-partitioning rule named uniform set-partitioning (USP) [13]. The same single-symbol set-partitioning is performed uniformly on each symbol to form symbol partitions in a USP, where the overall multisymbol vector partition is obtained by taking a Cartesian product of these symbol partitions. Note that USP is also where the proposed JRSSD differs from [14], where a biased nonuniform set-partition is used in solving a similar joint equalization problem for a multiuser time division-synchronous code division multiple access (TD-SCDMA) system.

The spatial and temporal processing is effectively decoupled in the space-time JRSSD structure with the help of USP. A USP-based reduced-state trellis forms the backbone of temporal ISI equalization in this structure, whereas an auxiliary spatial multisymbol detection problem is solved for each edge in the reduced-state trellis. In each spatial multisymbol detection problem, so-called "early-decisions" on the parallel transitions are made and passed to the reduced-state trellis. Furthermore, we propose two low-complexity solutions, the decorrelating () and ordered successive (OS) algorithms, to decouple the multisymbol detection problem into M single-symbol detection problems, where a simple threshold detector may be applied due to the geometric symmetry inherent in USP. Note that the OS solution is similar to the approach proposed for the V-BLAST system [15].

The rest of the paper is organized as follows. Section II presents the MIMO transmit/receive signal model. The overall MIMO receiver structure is summarized in Section III. It is followed by the detailed description of the JRSSD in Sections IV and V. The analytical and numerical evaluation of JRSSD performance is found in Sections VI and VII, respectively.

II. MIMO SIGNAL MODEL

We consider a M transmit, N receive antenna MIMO system as shown in Fig. 1. For ease of exposition, we use a vector-matrix notation similar to the one used in [16]. Stacking up the received samples across all the receive antennas for the kth symbol interval, we get

$$\mathbf{y}_k = \sum_{l=0}^{L} \boldsymbol{\mathcal{H}}_l \mathbf{a}_{k-l} + \mathbf{n}_k. \tag{1}$$

Note that $\mathbf{y}_k = [\mathbf{y}_{k,1}^T, \ldots, \mathbf{y}_{k,N}^T]^T$ where each small vector $\mathbf{y}_{k,n}$ includes all the temporal samples within the kth symbol interval and n denotes receive antenna index. We assume Δ samples per symbol and, therefore, $\mathbf{y}_k$ is of length $N\Delta$. Meanwhile, L is the channel memory length, $\mathbf{a}_{k-l} = [a_{k-l,1}, \ldots, a_{k-l,M}]^T$ is the transmitted symbol vector at time $k-l$, and $\mathbf{n}_k$ is the $(N\Delta) \times 1$ dimensional white Gaussian noise vector with $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. Note that σ^2 denotes noise variance and $\mathbf{I}$ is the identity matrix. The channel coupling between the transmitter and receiver is denoted by $\boldsymbol{\mathcal{H}}_0, \ldots, \boldsymbol{\mathcal{H}}_L$ where each matrix is of dimension $N\Delta \times M$. In this paper these channel coefficients are assumed to be known at the receiver. However, the numerical results include the effect of estimated channel coefficients. Furthermore, in order to facilitate the discussion on the MMSE-DDFSE receiver, we stack up a block of $L_f + 1$ received vectors

$$\mathbf{y}_{k+L_f:k} = \mathbf{H}\mathbf{a}_{k+L_f:k-L} + \mathbf{n}_{k+L_f:k} \tag{2}$$

where $L_f + 1$ is the length of the prefilter of the MMSE-DDFSE receiver and

$$\mathbf{y}_{k+L_f:k} = \left[\mathbf{y}_{k+L_f}^T, \ldots, \mathbf{y}_k^T\right]^T, \quad ((L_f+1)N\Delta \times 1)$$

$$\mathbf{n}_{k+L_f:k} = \left[\mathbf{n}_{k+L_f}^T, \ldots, \mathbf{n}_k^T\right]^T, \quad ((L_f+1)N\Delta \times 1)$$

$$\mathbf{a}_{k+L_f:k-L} = \left[\mathbf{a}_{k+L_f}^T, \ldots, \mathbf{a}_{k-L}^T\right]^T, \quad ((L_f+L+1)M \times 1)$$

$$\mathbf{H} = \begin{bmatrix} \boldsymbol{\mathcal{H}}_0 & \cdots & \boldsymbol{\mathcal{H}}_L & & \\ & \ddots & & \ddots & \\ & & \boldsymbol{\mathcal{H}}_0 & \cdots & \boldsymbol{\mathcal{H}}_L \end{bmatrix}$$

is of size $((L_f+1)N\Delta \times (L_f+L+1)M)$. where the dimensions of the matrices are given next to them. Note that to keep the notation more intuitive, we keep the subscripts at a "block" level. For instance, $\mathbf{y}_{k+L_f:k}$ is the vector that contains blocks $\mathbf{y}_{k+L_f}, \ldots, \mathbf{y}_k$ where each block is a vector of size $N\Delta \times 1$.

III. MIMO RECEIVER STRUCTURE

The overall receiver structure is shown in Fig. 2. The MIMO prefilter and feedback filter combine to shorten and reshape the channel impulse response. Meanwhile, we assume that the VVA block is replaced by the proposed low-complexity JRSSD block as in Fig. 2. Note that if we keep the VVA block, we return to a conventional MMSE-DDFSE receiver.

In this section, we summarize the conventional MMSE-DDFSE receiver, including the MIMO FIR prefilter/feedback filter design and the VVA equalizer. First, we define an error vector $\tilde{\mathbf{n}}_k = \mathbf{W}^H \mathbf{y}_{k+L_f:k} - \mathbf{B}^H \mathbf{a}_{k:k-L}$, where the superscript ^{H} denotes Hermitian transpose operation, $\mathbf{W}$ is the prefilter matrix and $\mathbf{B}$ is the feedback filter matrix. To be specific, $\mathbf{W}$ is of dimension $(N\Delta(L_f + 1)) \times M$ and $\mathbf{B}$ is of dimension $((L + 1)M) \times M$. We can also write $\mathbf{B}^H = [\mathbf{B}_0, \ldots, \mathbf{B}_L]$ where each submatrix is of size $M \times M$. Similar to the treatment in [16], the filters $\mathbf{W}$ and $\mathbf{B}$ are optimized according to the MMSE criteria

$$\mathbf{W}^{\text{opt}}, \mathbf{B}^{\text{opt}} = \arg \min_{\mathbf{W}, \mathbf{B}} \text{Trace}(\mathbf{R}) \quad \text{s.t. } \mathbf{B}_0 = \mathbf{I} \tag{3}$$
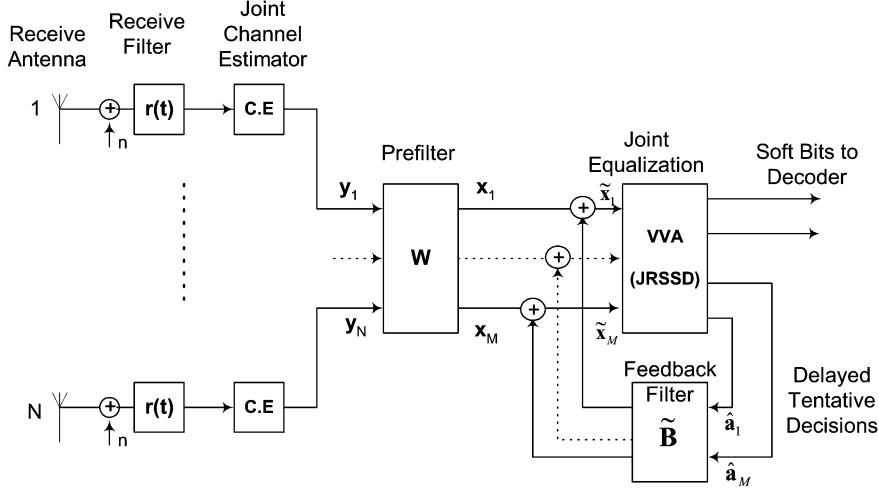
Fig. 2.   Overall MIMO MMSE-DDFSE receiver structure. Note that the VVA block is replaced by a proposed JRSSD block in this paper.

where $\mathbf{R} = E\{\tilde{\mathbf{n}}_k \tilde{\mathbf{n}}_k^H\}$ is the error correlation matrix. Note that $E\{\cdot\}$ denotes the expectation operation. The solution to (3) is obtained from Lagrange multipliers: $\mathbf{W}^{\mathrm{opt}} = \mathbf{V}^{-1}\mathbf{H}_{k:k}(\mathbf{H}_{k:k}^H \mathbf{V}^{-1}\mathbf{H}_{k:k})^{-1}$ and $\mathbf{B}^{\mathrm{opt}} = \mathbf{H}_{k:k-L}^H \mathbf{W}^{\mathrm{opt}}$, where $\mathbf{V} \triangleq (\mathbf{H}_{k+L_f:k+1}\mathbf{H}_{k+L_f:k+1}^H + \sigma^2 \mathbf{I})$ and the resulting error correlation matrix is: $\mathbf{R} = (\mathbf{H}_{k:k}^H \mathbf{V}^{-1}\mathbf{H}_{k:k})^{-1}$. Note that here we use the notation $\mathbf{H}_{k+L_f:k+1}$, $\mathbf{H}_{k:k}$ and $\mathbf{H}_{k-1:k-L}$ to represent the submatrices of $\mathbf{H}$ that are associated with $\mathbf{a}_{k+L_f:k+1}$, $\mathbf{a}_k$ and $\mathbf{a}_{k-1:k-L}$ in the expansion of the matrix-vector product $\mathbf{H}\mathbf{a}_{k+L_f:k-L} = \mathbf{H}_{k+L_f:k+1}\mathbf{a}_{k+L_f:k+1} + \mathbf{H}_{k:k}\mathbf{a}_k + \mathbf{H}_{k-1:k-L}\mathbf{a}_{k-1:k-L}$. Furthermore, $\mathbf{H}_{k:k-L}$ is defined as $\mathbf{H}_{k:k-L} \triangleq [\mathbf{H}_{k:k}, \mathbf{H}_{k-1:k-L}]$. After MIMO prefiltering, the effective signal model is described by

$$\mathbf{x}_k = \mathbf{W}^H \mathbf{y}_{k+L_f:k} = \mathbf{B}^H \mathbf{a}_{k:k-L} + \tilde{\mathbf{n}}_k = \sum_{l=0}^{L} \mathbf{B}_l \mathbf{a}_{k-l} + \tilde{\mathbf{n}}_k. \tag{4}$$

Furthermore, denoting L_S as the length of shortened channel memory, the feedback filter is formed as $\tilde{\mathbf{B}}^H = [\mathbf{B}_{L_S+1}, \ldots, \mathbf{B}_L]$ and the input to the JRSSD block is obtained assuming perfect feedbacks $\hat{\mathbf{a}}_{k-L_S-1:k-L} = \mathbf{a}_{k-L_S-1:k-L}$

$$\tilde{\mathbf{x}}_k = \mathbf{x}_k - \tilde{\mathbf{B}}^H \hat{\mathbf{a}}_{k-L_s-1:k-L} = \sum_{l=0}^{L_S} \mathbf{B}_l \mathbf{a}_{k-l} + \tilde{\mathbf{n}}_k. \tag{5}$$

Note that to arrive at (5), we have assumed *common* tentative decision feedback symbols $\hat{\mathbf{a}}_{k-L_S-1:k-L}$ associated with the tail part of the channel impulse response. That is, the feedback symbols are the same for all the states in the trellis. This is an alternative to another approach that assumes *state-dependent* feedback symbols where each state in the trellis keeps track of its own feedback symbols. The *state-dependent* approach is optimal but is more complex because it requires extra memory to store the feedback symbols for each state, and extra computations for decision feedback subtractions at each state. Furthermore, simulation results show that in the presence of a prefilter, the performance loss associated with the *common* feedback symbols is negligible. Therefore, in the scope of this paper we assume *common* feedback for simplicity of exposition. Consequently,

the optimal joint equalization of (5), known as the VVA, is obtained by the following maximum likelihood (ML) optimization problem that effectively maximizes the probability density function $f_{\tilde{\mathbf{x}}_1,\ldots,\tilde{\mathbf{x}}_K}(\cdot)$

$$[\mathbf{a}_1^{\mathrm{opt}}, \ldots \mathbf{a}_K^{\mathrm{opt}}] = \arg \max_{\mathbf{a}_1,\ldots,\mathbf{a}_K} f_{\tilde{\mathbf{x}}_1,\ldots,\tilde{\mathbf{x}}_K}$$
$$\times (\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_K | \mathbf{a}_1, \ldots, \mathbf{a}_K)$$
$$= \arg \max_{\mathbf{a}_1,\ldots,\mathbf{a}_K} c \cdot e^{-\sum_{k=1}^{K} d^2(k)} \tag{6}$$

where K is the burst length, c is a multiplicative constant that does not depend on $\mathbf{a}_1, \ldots, \mathbf{a}_K$, and $d^2(k)$ is the Euclidean metric at time k: $d^2(k) = (\tilde{\mathbf{x}}_k - \sum_{l=0}^{L_S} \mathbf{B}_l \mathbf{a}_{k-l})^H \mathbf{R}^{-1}(\tilde{\mathbf{x}}_k - \sum_{l=0}^{L_S} \mathbf{B}_l \mathbf{a}_{k-l})$.

IV. USP FOR JRSSD

A VVA solution is optimal to (6) but requires a prohibitively high complexity. To achieve a better complexity and performance tradeoff, we extend the RSSD method proposed for SISO equalization [1], [17] and propose a JRSSD structure as illustrated in Fig. 3.

In this structure, the USP-based reduced-state trellis forms the backbone of temporal ISI equalization whereas an auxiliary V-BLAST like spatial multisymbol detection problem is solved for each edge of the reduced-state trellis.

To facilitate the discussions on the USP, we follow a set of notation that is similar to that of [1]. In the presence of a channel shortening prefilter described in the previous section, the trellis states for a VVA are defined as $\mathbf{p}_k = [\mathbf{a}_{k-1}, \ldots, \mathbf{a}_{k-L_S}]$ and the transition symbol vector is $\mathbf{a}_k$. Denoting Q as the modulation size, the number of states in the VVA trellis is Q^{ML_S} and the number of transitions to and from each state is Q^M. To reduce the number of states, we partition the possible signals of each element $\mathbf{a}_{k-l}$ in the state $\mathbf{p}_k$ into J_l subsets (each of cardinality Q^M/J_l), where a subset is denoted as $\mathcal{V}(\alpha_{k-l})$ with $\alpha_{k-l} : 0 \leq \alpha_{k-l} \leq J_l - 1$ being the index to the subset. Overall, all possible signals of the state $\mathbf{p}_k$ are partitioned into $\prod_{l=1}^{L_S} J_l$ composite subsets, where each composite subset is a Cartesian product of element subsets and is denoted by $\mathcal{V}(\mathbf{t}_k) = \mathcal{V}(\alpha_{k-1}) \times \cdots \times \mathcal{V}(\alpha_{k-L_S})$. Note that $\mathbf{t}_k = [\alpha_{k-1}, \ldots, \alpha_{k-L_S}]$ is the index to
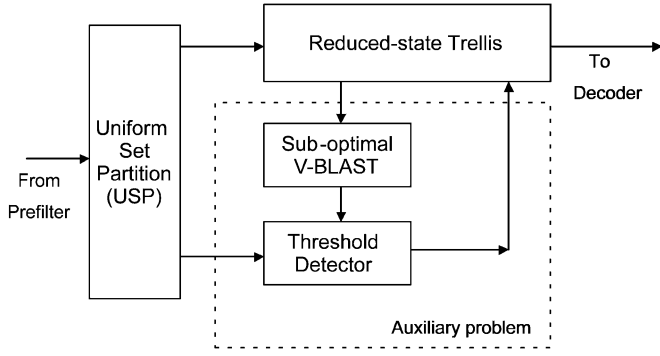
Fig. 3. Block diagram of JRSSD structure.

Fig. 4. JRSSD trellis, $E_1 = 2$, $J_1 = 2^2$, $Q = 8$, $M = 2$, $L_S = 1$.

such a composite subset (also known as a *subset state*). The cardinality of each composite subset is $Q^{ML_S} / \prod_{l=1}^{L_S} J_l$.

We proceed to discuss the methods of set partitioning. For a SISO case, the RSSD in [1] uses a partitioning method that is based on maximum Minimum Square Subset Distance (MSSD) criteria, or known as Ungerboeck set partitioning criteria [10]. The Ungerboeck maximum MSSD set-partitioning tree for a single 8 PSK symbol is geometrically symmetric. In a MIMO case, however, direct extension of the maximum MSSD criteria leads to Multidimensional Set Partitioning (MDSP), which maximizes the MSSD between different elements within a composite subset, as shown in [11] in the context of multidimensional trellis coded modulation. One major disadvantage of MDSP is that the geometric symmetry in the partitioned subset is lost once $M > 1$. To see this, consider the example where $Q = 8$, $M = 2$, $L_S = 1$, if we choose $J_1 = 16$, then the cardinality of a composite subset (subset state) is $8^2/16 = 4$. Note that in this case $\mathbf{V}(\mathbf{t}_k) = \mathcal{V}(\alpha_{k-1})$ since $L_S = 1$. Applying the MSSD criteria to the problem, we encounter a subset that looks like: $\mathcal{V}(\alpha_{k-1}) = \{e_1 e_5, e_4 e_3, e_5 e_2, e_7 e_7\}$ where $e_1, \ldots, e_8$ denote the constellation points of an 8PSK modulation. Denoting $\mathbf{a}_{k-1} = [a_{k-1,1}, a_{k-1,2}]$ where $a_{k-1,1}$ and $a_{k-1,2}$ are the two transmitted symbols at time $k-1$, it is observed that in this particular subset $a_{k-1,1} \in \{e_1, e_4, e_5, e_7\}$ and $a_{k-1,2} \in \{e_5, e_3, e_2, e_7\}$. Due to the apparent loss of geometric symmetry, simple threshold detectors, a key factor in the overall complexity reduction of a RSSD [1], cannot be used in the processing of parallel transitions.

To retain the geometric symmetry in the partitioned subsets, we propose an alternative set-partitioning rule named uniform set-partitioning (USP). The idea of USP is somewhat different from MDSP. In MDSP, each symbol vector $\mathbf{a}_{k-l}$ is first formed as the Cartesian product of M symbols (across the spatial dimension) and set-partitioning is applied on the M-dimensional vectors according to the maximum MSSD criteria. In a USP, on the other hand, the same single-symbol set-partitioning is performed uniformly on each symbol $a_{k-l,m}$ to form single-symbol subsets. After that, the overall set-partition for the multisymbol vector is obtained by taking a Cartesian product of the single-symbol partitions. To this end, we isolate the symbol sequence from the mth transmit antenna and let $p_k = [a_{k-1,m}, \ldots, a_{k-L_S,m}]$ denote the trellis state for the associated SISO VA trellis. The possible signals of each symbol $a_{k-l,m}$ are partitioned into $E_{l,m}$ subsets each of cardinality
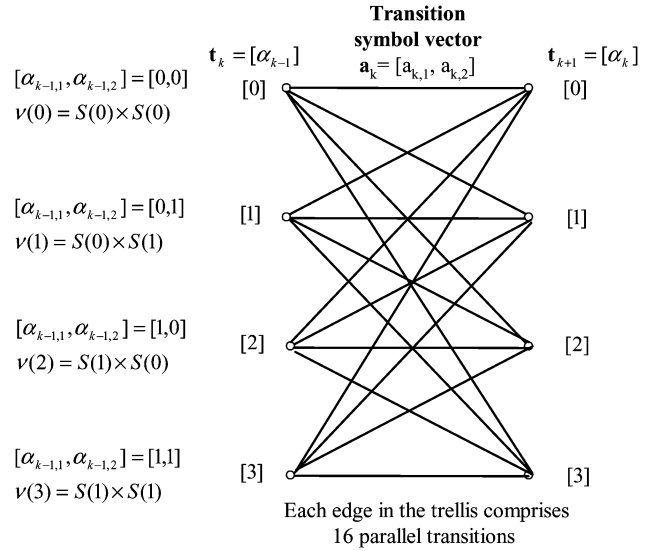
$Q/E_{l,m}$, and the symbol subsets are denoted as $\mathcal{S}(\alpha_{k-l,m})$ where $\alpha_{k-1,m}$ is the index to the symbol subset. In a USP, the partitions are kept the same for all the symbols within a symbol vector $\mathbf{a}_{k-l}$, meaning that $\mathcal{S}(\alpha_{k-l,1}) = \cdots = \mathcal{S}(\alpha_{k-l,M})$ if $\alpha_{k-l,1} = \cdots = \alpha_{k-l,M}$. Note that $0 \leq \alpha_{k-l,m} \leq E_{l,m} - 1$ for $m = 1, \ldots, M$. Consequently, the partition of the multisymbol vector $\mathbf{a}_{k-l} = [a_{k-l,1}, \ldots, a_{k-l,M}]$ is the Cartesian product of the symbol partitions

$$\mathcal{V}(\alpha_{k-l}) = \mathcal{S}(\alpha_{k-l,1}) \times \cdots \times \mathcal{S}(\alpha_{k-l,M}) \qquad (7)$$

where the number of subsets is $J_l = E_l^M$ and $E_l = E_{l,1} = \cdots = E_{l,M}$. Furthermore, in (7) the index α_{k-1} to the vector subset is related to the indices to the symbol subsets by

$$\alpha_{k-l} = \sum_{m=1}^{M} \alpha_{k-l,m} E_l^{M-m}. \qquad (8)$$

A JRSSD-USP trellis with $Q = 8$, $M = 2$, $L_S = 1$, $E_1 = 2$, $J_1 = 2^2$ is illustrated in Fig. 4.

Unlike the MDSP, the USP does not attempt to maximize the MSSD within a subset. For the example where $Q = 8$, $L_S = 1$ and $E_1 = 2$, it can be shown that as the number of transmit antennas M increases, the MSSD for USP stays at $d_{\min}^{USP}(J_1 = 2^M) = \sqrt{2}$ and does not change as M increases. Therefore, if MSSD is the only concern, the MDSP is more favorable since in a MDSP the MSSD in general increases with M. However, MSSD is only one of the contributing factors in the error probability performance of a JRSSD. And, as we show in a later section, near-optimal bit-error rate (BER) performance can be achieved with the simple USP.

V. SPATIAL MULTISYMBOL DETECTION

Once the reduced-state trellis is set-up according to the USP, the sequence detection is a two-step process. First, a spatial multisymbol detection problem is solved for each edge in the trellis to make the so-called "early-decisions" on the parallel transitions associated with this edge; second, the ISI equalization is carried out in the trellis with the help of these early decisions. Parallel transitions are the transitions in a reduced-state trellis

that originate and end in the same subset state. For the example $Q = 8$, $M = 2$, $L_S = 1$, $E_1 = 2$, $J_1 = 2^2$ as shown in Fig. 4, if we zoom in on the edge that connects subset states $\mathbf{t}_k = [\alpha_{k-1}] = [0]$ and $\mathbf{t}_{k+1} = [\alpha_k] = [0]$, it is comprised of $8^2/2^2 = 16$ parallel transitions, each representing a pair of transition symbols $[a_{k,1}, a_{k,2}] \in \mathcal{V}(0) = \mathcal{S}(0) \times \mathcal{S}(0)$. Note that the symbol partition in this case is given by $\mathcal{S}(0) = \{e_0, e_2, e_4, e_6\}$ and $\mathcal{S}(1) = \{e_1, e_3, e_5, e_7\}$.

We now derive the signal model for the spatial multisymbol detection problem. For ease of exposition, we use the example where $Q = 8$, $E_1 = 2$ but leave both L_S and M as variables. In this case the number of subset states is given by $J_1 = E_1^M = 2^M$ and the number of parallel transitions is $8^M/2^M = 4^M$. We start with the signal model in (4). Since all the parallel transitions originate from the same subset state, they share the same path ISI history. Therefore, subtracting the ISI terms from the receiver signal, we get

$$\mathbf{z}_k = \tilde{\mathbf{x}}_k - \sum_{i=1}^{L_s} \mathbf{B}_i \mathbf{a}_{i-k} = \mathbf{B}_0 \mathbf{a}_k + \tilde{\mathbf{n}}_k \qquad (9)$$

where the transition symbol vector $\mathbf{a}_k \in \mathcal{V}(\alpha_k) = \mathcal{S}(\alpha_{k,1}) \times \cdots \times \mathcal{S}(\alpha_{k,M})$. Note that here we have assumed perfect decision feedback for notational simplicity. The effects of decision error propagation are included in the numerical results.

A. Maximum Likelihood Solution (JRSSD-ML)

From (9), the optimal transition $\mathbf{a}_k^{\text{opt}}$ is chosen among all the 4^M parallel transitions, by solving the following ML optimization problem

$$\mathbf{a}_k^{\text{opt}} = \arg \min_{\mathbf{a}_k \in \mathcal{V}(\alpha_k)} (\mathbf{z}_k - \mathbf{B}_0 \mathbf{a}_k)^H \mathbf{R}^{-1} (\mathbf{z}_k - \mathbf{B}_0 \mathbf{a}_k) \quad (10)$$

which involves 4^M Euclidean distance (ED) computations. Since there are a total of 2^M states and $(2^M 2^M)$ edges between two neighboring stages, the overall complexity of the JRSSD-ML method for this example is $\mathcal{O}((2^M 2^M)4^M)$ ED computations. Although a significant reduction compared to a DDFSE without set-partitioning and state reduction, which requires $\mathcal{O}(8^M 8^M)$ ED computations, it does not achieve the full potential of complexity reduction. In the ensuing subsections, we propose suboptimal DC and OS algorithms that decouple the spatial multisymbol detection problem associated with each edge, into M single-symbol detection problems. Due to the geometric symmetry inherent in the USP, for each single-symbol detection problem, the symbol to be detected belongs to either set $\mathcal{S}(0) = \{e_0, e_2, e_4, e_6\}$ or set $\mathcal{S}(1) = \{e_1, e_3, e_5, e_7\}$. As a result, the symbol decision is made with an efficient threshold detector that does not require any ED computation. Therefore, after all the symbol decisions are available, only one ED computation is necessary to provide a distance metric for this edge. The number of ED computations needed for JRSSD-DC and JRSSD-OS is thus reduced to $\mathcal{O}(2^M 2^M)$ for this example. It is important to note, however, that the ED computations are only part of overall complexity needed for DC and OS methods, as we discuss in detail in Sections V-B and C.

B. Decorrelating Solution (JRSSD-DC)

Due to the $\mathbf{B}_0 = \mathbf{I}$ constraint in (3), the transmitted symbols $\mathbf{a}_k$ are spatially decoupled in the received signal model (9). This suggests a simple suboptimal DC solution to (10) by detecting $a_{k,m}$ individually from

$$z_{k,m} = a_{k,m} + \tilde{n}_{k,m} \text{ for } m = 1, \ldots, M \qquad (11)$$

using the SISO slicer that takes advantage of the geometric symmetry in the subset. Although it can be shown that the DC solution is optimal if $\mathbf{R}$ is diagonal, a closer examination reveals that this condition cannot be simultaneously satisfied with the $\mathbf{B}_0 = \mathbf{I}$ constraint in the optimization problem (3). Therefore, the DC solution is always suboptimal in a MMSE-JRSSD receiver. The performance of the DC solution is characterized by the SNR of each data stream. For example, γ_m^{DC} denotes the SNR of m_{th} data stream for the DC solution

$$\gamma_m^{DC} = \frac{E\{|a_{k,m}|^2\}}{E\{|\tilde{n}_{k,m}|^2\}} = \frac{1}{r_{m,m}} \qquad (12)$$

where $r_{m,m}$ is the mth diagonal element of the matrix $\mathbf{R}$.

The complexity of the JRSSD-DC comprises of two parts: the ED computations and the slicer operations which we denote as $\mathcal{E}$ and $\mathcal{C}$, respectively. The overall complexity of the JRSSD-DC is, therefore, given by $\mathcal{O}(2^M 2^M \mathcal{E} + M 2^M 2^M \mathcal{C})$, where we have taken into account the fact that M slicer operations are needed for each parallel transition in the trellis. Note that the notation for complexity is meant symbolically, since the symbols $\mathcal{E}$ and $\mathcal{C}$ are not numbers and cannot be simply added.

C. Ordered Successive Solution (JRSSD-OS)

In (9), notice $\mathbf{B}_0 = \mathbf{I}$ and let $\mathbf{R} = \mathbf{L}\mathbf{D}\mathbf{U}$ be the LU decomposition of $\mathbf{R}$ where

$$\mathbf{L} = \begin{bmatrix} 1 & & \\ \vdots & \ddots & \\ L_{M,1} & \ldots & 1 \end{bmatrix} \text{ and } \mathbf{D} = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_M \end{bmatrix} \qquad (13)$$

are lower triangular and diagonal matrices, respectively. Note $\mathbf{L} = \mathbf{U}^H$ since $\mathbf{R}$ is Hermitian symmetric, and in this case the LU decomposition is equivalent to the Cholesky decomposition. We multiply both sides of (9) by $\mathbf{L}^{-1}$ to whiten the noise in $\mathbf{z}_k$

$$\check{\mathbf{z}}_k = \mathbf{L}^{-1}\mathbf{z}_k = \mathbf{L}^{-1}\mathbf{a}_k + \check{\mathbf{n}}_k \qquad (14)$$

where $\check{\mathbf{n}}_k = \mathbf{L}^{-1}\tilde{\mathbf{n}}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$. Note that $\mathbf{L}^{-1}$ is also lower triangular with unit diagonal elements. The structure suggests a successive detection algorithm that uses simple SISO slicer to detect the symbols in the order of $a_{k,1}, \ldots, a_{k,M}$. At stage m, the previous decisions on symbols $\hat{a}_{k,1}, \ldots, \hat{a}_{k,m-1}$ are feedback and subtracted from $\check{z}_{k,m}$

$$\check{z}_{k,m} - \sum_{j=1}^{m-1} (L^{-1})_{m,j} \hat{a}_{k,j} = a_{k,m} + \check{n}_{j,m}. \qquad (15)$$

Now the simple SISO slicer can be used to detect $a_{k,m}$. The SNR characterization of this successive algorithm is more difficult due to possible error propagation. However, if we consider the limiting performance where all the feedback decisions are

correct, then we can calculate the SNR of each data stream similar to the DC solution. With the assumption of perfect feedback, one can show that the SNR of the mth data stream is given by

$$\gamma_m^{OS} = \frac{E\{|a_{k,m}|^2\}}{E\{|\check{n}_{k,m}|^2\}} = \frac{1}{d_m}. \tag{16}$$

In the successive detection algorithm above, we have fixed the detection order to be $(1, 2, \ldots, M)$ rather arbitrarily. However, the performance of the successive detection can be improved by optimizing the detection order [15]. In [15], the detection order for a similar V-BLAST problem is chosen so that the worst SNR among M data streams is maximized. In particular, let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_M)$ denote an arbitrary ordering and let Ω be the set of all possible orderings. The cardinality of the set is $|\Omega| = M!$ and the following problem is solved to obtain the optimal ordering $\boldsymbol{\beta}$

$$\boldsymbol{\beta} = \arg\max_{\boldsymbol{\omega} \in \Omega} \min_m \gamma_m^{OS}(\boldsymbol{\omega}) \tag{17}$$

where $\gamma_m^{OS}(\boldsymbol{\omega})$ denotes the SNR of the mth data stream given the detection order $\boldsymbol{\omega}$. The ordered successive algorithm is now complete by combining the optimal ordering of (17) with successive detection (15)

```
Ordered Successive Algorithm
1)    Solve (17) for the optimal ordering
      and denote it β = (β₁,...,β_M).
2)    Construct a permutation matrix P
      such that Pa_k = a'_k = [a_{k,β₁},...,a_{k,β_M}]^T.
3)    In (9), multiply both sides by P
      and we get z'_k = a'_k + n'_k where z'_k = Pz_k,
      n'_k = Pñ_k and n'_k ∼ N(0, PRP^H).
      Now perform the LU decomposition
      R' = PRP^H = L'D'U' and follow the
      treatment from (13) to (15) to de-
      tect a'_k successively.
```

1) Optimal Ordering: We focus on solving the optimal ordering $\boldsymbol{\beta}$ in Step 1. The direct solution to (17) involves $M!$ LU decompositions and is not practical when M is large. A localized optimization process proposed in [15] for a related V-BLAST problem can be adopted here with some modification. In the localized optimization process, the optimum ordering is obtained by selecting the symbol with the best SNR at each stage of detection. We start the local optimization process by finding first symbol to be detected, i.e., β_1. To this end, let $(\boldsymbol{\omega}, \mathbf{P}(\boldsymbol{\omega}))$ be the pair of an arbitrary ordering and its corresponding permutation matrix. Let $\mathbf{R}'(\boldsymbol{\omega}) = \mathbf{P}(\boldsymbol{\omega})\mathbf{R}\mathbf{P}^H(\boldsymbol{\omega})$ be the "ordered" noise correlation matrix, and $\mathbf{R}'(\boldsymbol{\omega}) = \mathbf{L}'(\boldsymbol{\omega})\mathbf{D}'(\boldsymbol{\omega})\mathbf{U}'(\boldsymbol{\omega})$ be its LU decomposition. We try to find the $\boldsymbol{\omega}$ that maximizes the SNR of the first symbol to be detected

$$\boldsymbol{\omega}^{\text{opt}} = \arg\max_{\boldsymbol{\omega} \in \Omega} \gamma_1^{OS}(\boldsymbol{\omega}) = \arg\max_{\boldsymbol{\omega} \in \Omega} \frac{1}{d_1'(\boldsymbol{\omega})} \tag{18}$$

where $d_1'(\boldsymbol{\omega})$ is the first diagonal element of $\mathbf{D}'(\boldsymbol{\omega})$ and we have used (16) here. Once $\boldsymbol{\omega}^{\text{opt}}$ is obtained, β_1 is given by $\beta_1 \triangleq \omega_1^{\text{opt}}$. Since $d_1'(\boldsymbol{\omega})$ is a function of the overall ordering $\boldsymbol{\omega}$, it looks like

the enumeration of all $M!$ possible $\boldsymbol{\omega}$ is inevitable, even if we only want to find the first symbol β_1. However, borrowing the results of Proposition 1 from next section, one can show that $\gamma_1^{OS} = 1/(R'(\boldsymbol{\omega}))_{1,1} = 1/r_{\omega_1,\omega_1}$ where $r_{\omega_1,\omega_1} = (\mathbf{R})_{\omega_1,\omega_1}$, and reduce (18) into

$$\beta_1 = \arg\max_{\omega_1} \frac{1}{r_{\omega_1,\omega_1}} \tag{19}$$

which is solved by simply picking the ω_1 that gives the smallest r_{ω_1,ω_1}. Once β_1 is obtained, we fix the first element of $\boldsymbol{\omega}$ as $\omega_1 = \beta_1$, and obtain β_2 by maximizing the second post-detection SNR $\gamma_2^{OS}(\boldsymbol{\omega}) = 1/d_2'(\boldsymbol{\omega})$ for all $\boldsymbol{\omega} \in \Omega(\beta_1)$ where $\Omega(\beta_1) \triangleq \{\boldsymbol{\omega} : \omega_1 = \beta_1\}$. To avoid searching over the whole space of $\Omega(\beta_1)$, we again invoke Proposition 1 and get

$$\gamma_2^{OS}(\boldsymbol{\omega}) = ((\mathbf{R}_2'(\boldsymbol{\omega}))^{-1})_{1,1} = \frac{r_{1,1}'}{r_{1,1}'r_{2,2}' - r_{1,2}'r_{2,1}'}$$
$$= \frac{r_{\beta_1,\beta_1}}{r_{\beta_1,\beta_1}r_{\omega_2,\omega_2} - r_{\beta_1,\omega_2}r_{\omega_2,\beta_1}} \tag{20}$$

where $\mathbf{R}_2'(\boldsymbol{\omega})$ is the 2×2 primary submatrix of $\mathbf{R}'(\boldsymbol{\omega})$, $r_{i,j}' \triangleq (\mathbf{R}'(\boldsymbol{\omega}))_{i,j}$ and $r_{i,j} \triangleq (\mathbf{R})_{i,j}$. It follows that β_2 is obtained by solving the following single-variable optimization problem

$$\beta_2 = \arg\max_{\omega_2 \neq \beta_1} \frac{1}{r_{\omega_2,\omega_2} - \frac{|r_{\beta_1,\omega_2}|^2}{r_{\beta_1,\beta_1}}} \tag{21}$$

where the denominator is the Schur complement of the 2×2 first primary submatrix of $\mathbf{R}'(\boldsymbol{\omega})$. The solution to β_2 involves $M - 1$ computations of the Schur complement of 2×2 submatrices. In fact, if we extend this algorithm to solve for $\beta_3, \ldots, \beta_M$, then the solution of β_m will involve $M - m + 1$ computations of Schur complements of $m \times m$ submatrices.

In a JRSSD-OS, in addition to the ED computations and the slicer operations, a total of $(1/2)M(M-1)$ additions are needed for each parallel transition in the trellis. If we denote the addition as $\mathcal{A}$, then the overall complexity for a JRSSD-OS is given by $\mathcal{O}(2^M 2^M \mathcal{E} + M 2^M 2^M \mathcal{C} + (1/2)M(M-1)2^M 2^M \mathcal{A})$.

VI. SNRS OF THE MULTISYMBOL DETECTION SOLUTIONS

In this section, we compare the performance of the ML, OS and DC solution for the optimization problem shown in (10). Without loss of generality, we assume a detection order of $\boldsymbol{\beta} = 1, 2, \ldots, M$ for the OS method. The SNR of the DC solution is given by $\gamma_m^{DC} = 1/r_{m,m}$ from earlier discussion. We show in the following proposition that the SNR of the ordered successive solution is $\gamma_m^{OS} = (\mathbf{R}_m^{-1})_{m,m}$ where $\mathbf{R}_m$ is the $m \times m$ primary submatrix of $\mathbf{R}$.

Proposition 1: The SNR of the ordered successive solution is given by $\gamma_m^{OS} = (\mathbf{R}_m^{-1})_{m,m}$ where $\mathbf{R}_m$ is the $m \times m$ primary submatrix of $\mathbf{R}$.

Proof: See Appendix A.

We now compute the SNR of the ML solution. In ML, all the symbols $a_{k,1}, \ldots, a_{k,M}$ are detected jointly and, therefore, it is not obvious how to calculate the SNR associated with the detection of a single symbol. Here we again look at the SNR in a limiting situation: while computing the SNR for a given symbol, for example $a_{k,M}$, we assume all the other symbols $a_{k,1}, \ldots, a_{k,M-1}$ are correctly detected. With this assumption,

it is easy to see that for $a_{k,M}$ the SNR of ML solution is the same as that of the ordered successive solution

$$\gamma_M^{ML} = \gamma_M^{OS} = (\mathbf{R}_M^{-1})_{M,M} = (\mathbf{R}^{-1})_{M,M}. \qquad (22)$$

Furthermore, one can show that in general the SNR of ML solution is $\gamma_m^{ML} = (\mathbf{R}^{-1})_{m,m}$ for any $1 \le m \le M$. To see this, let $\mathbf{P}_m$ be the left permutation matrix that swaps the mth and Mth element for a vector of size $M \times 1$, then multiply both sides of (9) by $\mathbf{P}_m$ and we get $\mathbf{z}_k' = \mathbf{a}_k' + \mathbf{n}_k'$ where $\mathbf{z}_k' = \mathbf{P}_m \mathbf{z}_k$, $\mathbf{n}_k' = \mathbf{P}_m \tilde{\mathbf{n}}_k$ and $\mathbf{n}_k' \sim \mathcal{N}(\mathbf{0}, \mathbf{R}' = \mathbf{P}_m \mathbf{R} \mathbf{P}_m^H)$. Furthermore, let $(\gamma')_M^{ML}$ denote the Mth SNR for the new vector $\mathbf{a}_k'$, then $\gamma_m^{ML} = (\gamma')_M^{ML}$ by definition. It follows from (22) that

$$\gamma_m^{ML} = (\gamma')_M^{ML} = ((\mathbf{R}')^{-1})_{M,M} = ((\mathbf{P}_m \mathbf{R} \mathbf{P}_m^H)^{-1})_{M,M}$$
$$= (\mathbf{P}_m \mathbf{R}^{-1} \mathbf{P}_m^H)_{M,M} = (\mathbf{R}^{-1})_{m,m}.$$

Note that we have implicitly used the property that $\mathbf{P}_m = \mathbf{P}_m^H = \mathbf{P}_m^{-1}$. We have now obtained the SNR of the three different solutions as functions of the noise correlation matrix $\mathbf{R}$

$$\gamma_m^{ML} = (\mathbf{R}^{-1})_{m,m}, \quad \gamma_m^{OS} = (\mathbf{R}_m^{-1})_{m,m} \text{ and } \gamma_m^{DC} = \frac{1}{r_{m,m}}. \qquad (23)$$

It remains to be seen how the SNRs, or equivalently the performance of the three methods, are related. An interesting observation is that $\gamma_1^{DC} = \gamma_1^{OS}$ and $\gamma_M^{OS} = \gamma_M^{ML}$ which shows that these SNRs are closely related. Moreover, among the three methods, one would expect ML to be the best, and the decorrelating solution to be the worst due to the diagonal approximations in that solution. This intuition is verified in the following proposition.

Proposition 2: $\gamma_m^{DC} \le \gamma_m^{OS} \le \gamma_m^{ML}$ for any $1 \le m \le M$.
Proof: See Appendix B.

VII. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed JRSSD algorithm with numerical simulations. The simulations are carried out in a practical link-level EDGE simulator. We focus on a two transmit, two receive antenna MIMO system, i.e, $M = N = 2$, with 8PSK modulation. We assume that each antenna encounters independent fading channels. Two channel models are used in the simulations: the 6-tap ($L = 6$) typical urban (TU) model and the 6-tap rural area (RA) model as defined in European Telecommunications Standard Institute (ETSI) Global System for Mobile Communication (GSM) standard [18]. Note that in our simulations, we fix the mobile station speed at 50 km/h for TU channels and 100 km/h for RA channels. Note that an over-sampling factor $\Delta = 2$ is used throughout our simulations, and the numerical results include the effect of estimated channel coefficients.

In the JRSSD, the shortened channel memory length is chosen to be $L_S = 1$. Meanwhile, the number of subsets for each symbol is $E_1 = 2$ in the simulations, and the number of subset states in the trellis is $J_1 = 2^M = 2^2$. The three algorithms concerning the processing of parallel transitions, the ML, OS and DC algorithms, are simulated and compared. The complexity of these different algorithms is compared in Table I. Note we have also included the complexity of the DDFSE with trellis memory length $L_S = 1$ for comparison. For the complexity

TABLE I
COMPARISON OF ALGORITHM COMPLEXITIES. HERE $\mathcal{E}$ DENOTES EUCLIDEAN DISTANCE COMPUTATION, $\mathcal{C}$ DENOTES SLICER OPERATION AND $\mathcal{A}$ DENOTES COMPLEX ADDITION

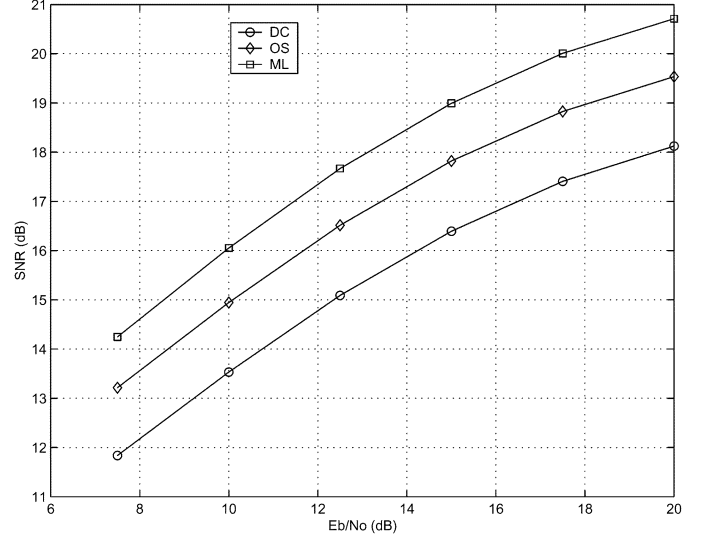|	# of States	Complexity
DDFSE	8^2	$\mathcal{O}(8^2 \cdot 8^2 \mathcal{E})$
JRSSD-ML	2^2	$\mathcal{O}(2^2 \cdot 8^2 \mathcal{E})$
JRSSD-OS	2^2	$\mathcal{O}(2^2 \cdot 2^2 \mathcal{E} + 2 \cdot 2^2 \cdot 2^2 \mathcal{C} + 2^2 \cdot 2^2 \mathcal{A})$
JRSSD-DC	2^2	$\mathcal{O}(2^2 \cdot 2^2 \mathcal{E} + 2 \cdot 2^2 \cdot 2^2 \mathcal{C})$

Fig. 5.   Minimum symbol SNR for a TU50 channel.

contributed by the ED computations, two levels of complexity reduction are observed: 1) a 16–fold reduction from DDFSE to JRSSD-ML, which is solely attributed to the multisymbol USP; and 2) another 16–fold reduction from JRSSD-ML to JRSSD-DC/JRSSD-OS, which results from the suboptimal decoupling of the spatial multisymbol detection problem. Note that in general the numerical complexity of an ED computation is much higher than a slicer operation or an addition. Therefore, we emphasize the reduction in the ED computations.

A. Minimum Symbol SNR Performance

In Fig. 5, the minimum symbol SNRs resulting from various algorithms are compared for a TU 50 channel, as a function of the transmit SNR denoted by Energy per Bit versus Noise (E_b/N_0). The SNR gaps between the ML, OS and DC methods are fairly uniform for the transmit SNR region simulated. The gap between ML and DC is about 3.5–4 dB in the TU channel. On the other hand, the gap between OS and ML is a little less than half of the gap between ML and DC. Note that, although the SNR results partially characterize the BER behavior of the algorithms of interest, the SNR gaps do not directly translate into gaps in the BER curves, as we show in Section VII-B. Furthermore, we note that the SNR results for the RA channel are similar to those for the TU channel.

B. BER Performance in Noise-Limited Environment

In Figs. 6 and 7, the BER comparisons are presented for a noise-limited TU 50 and RA 100 channel, respectively. A noise-limited situation arises when a mobile station is near the center
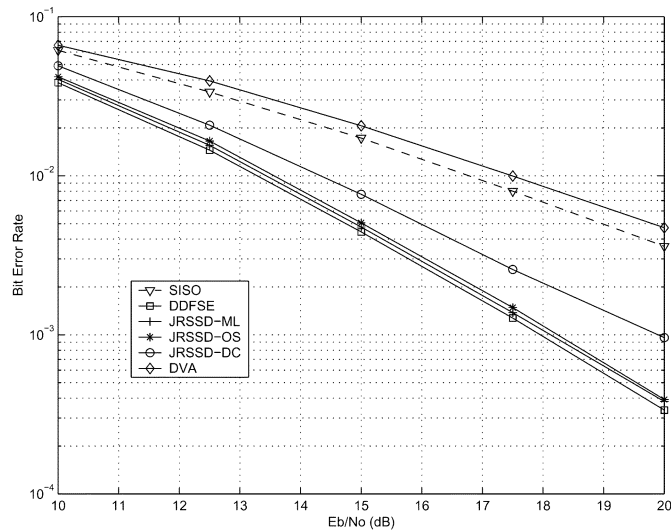
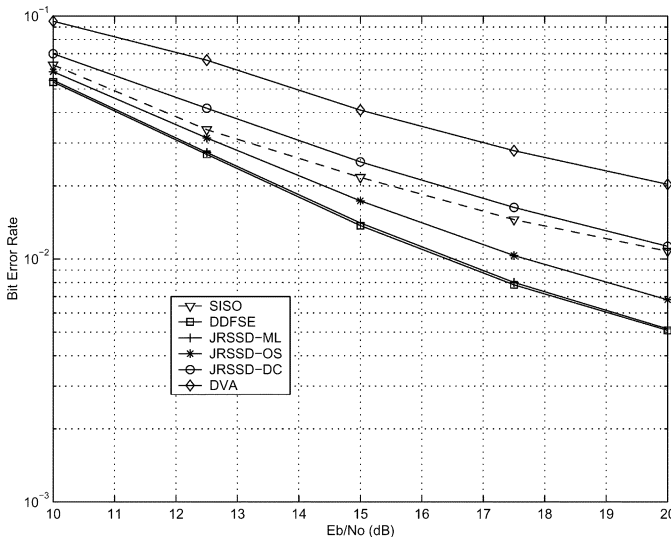Fig. 6. BER comparison for a noise-limited TU50 channel.

Fig. 7. BER comparison for a noise-limited RA100 channel.

of a cell, where we are able to model the sum of co-channel interferences from all other base stations as white Gaussian noise. Note that in addition to the JRSSD algorithm results, we have included in the comparison the DDFSE results, the interference suppression based suboptimal DVA [9] results, and results of a SISO EDGE system assuming conventional MMSE-RSSD with $Q = 8$, $L_S = 1$ and the number of subsets $E_1 = 2$.

Under both channel conditions, the performance difference between the JRSSD-ML and the optimal DDFSE is negligible, meaning that almost no performance penalty is associated with the ED reduction due to USP alone (a reduction from $\mathcal{O}(4096\mathcal{E})$ to $\mathcal{O}(256\mathcal{E})$ in this case). On the other hand, if OS or DC algorithm is adopted in the processing of parallel transitions, the ED computations are further reduced to $\mathcal{O}(16\mathcal{E})$ at the cost of some performance loss. For example, the loss of the DC method (against ML) is significant at about 1 and 4 dB for TU and RA channels (at 1 % BER operation point), respectively. On the other and, the OS is almost lossless in TU channel and loses about 0.7 dB in RA channel (at 1 percent BER operation point).

The reason for OS to behave differently in TU and RA is that in an RA channel, most of the channel energy is in the first tap, therefore, the overall BER performance is dominated by the performance of the spatial multisymbol detection problem, whose SNR gap (between OS and ML) is directly reflected in the BER performance; whereas in a much-longer TU channel, the performance of the spatial multisymbol detection problem dominates less in the overall BER performance and the signal-to-noise ratio (SNR) gap is less visible in the BER curves. Note that, both OS and DC outperform the interference suppression based DVA method, where the transmit sequences are decoupled before the equalization step.

In the noise-limited situations, both JRSSD-ML and JRSSD-OS results are significantly better than the SISO RSSD results for the full E_b/N_0 range. This implies that in a noise-limited situation, for both TU and RA channel, a MIMO configuration with $M = N = 2$ is essentially able to double the data rate of a SISO configuration, provided that either the JRSSD-ML or JRSSD-OS algorithm is used. Note that, although we have not shown the block error rate (BLER) results, the numerical results based on BLER show similar trends as the BER curves.

C. BER Performance in Interference-Limited Environment

Since the cochannel interference from other-cell base stations experience different fading and shadowing effects, the interference from one or a few base stations may become much stronger than the others at any given moment. In these so-called interference-limited situations, a simple white Gaussian noise model is not appropriate for the sum of all interference. In [19], a more realistic interference model is proposed where the overall interference is broken into two parts, where the first component remains white Gaussian noise and the second component is modeled as a dominant interfering base station. A parameter called dominant to rest of interference power ratio (DIR) is introduced to describe the relative power ratio of these two components. In this paper we adopt this interference model and set $\mathrm{DIR} = 15$ dB in our simulations to focus on the situation where the interference is highly structured.

Since the noise is no longer white in an interference-limited environment, some form of noise whitening (interference suppression) capability is necessary in the receiver. In our approach, interference suppression is achieved by incorporating the second-order noise correlation structure in the MMSE filter optimization. Specifically, the white Gaussian noise $\mathbf{n}_{k+L_f:k}$ in (2) becomes a zero-mean random noise vector with a second-order correlation matrix $\mathbf{R}_{nn} \triangleq E\{\mathbf{n}_{k+L_f:k}\mathbf{n}_{k+L_f:k}^H\}$. Consequently, the $\mathbf{V}$ matrix, which is defined after (3), becomes $\mathbf{V} \triangleq (\mathbf{H}_{k+L_f:k+1}\mathbf{H}_{k+L_f:k+1}^H + \mathbf{R}_{nn})$. Note that in practice, the noise correlation matrix $\mathbf{R}_{nn}$ is estimated together with channel matrices $\mathcal{H}_l$, using the training sequences embedded in the middle of data bursts.

In Figs. 8 and 9, the BER comparisons are presented for an interference-limited TU 50 and a RA 100 channel, respectively. Note that in the figures the term $C/I(dB)$ is read as "carrier to interference ratio in decibels," and we have assumed only one transmit antenna at the interfering base station. It is observed
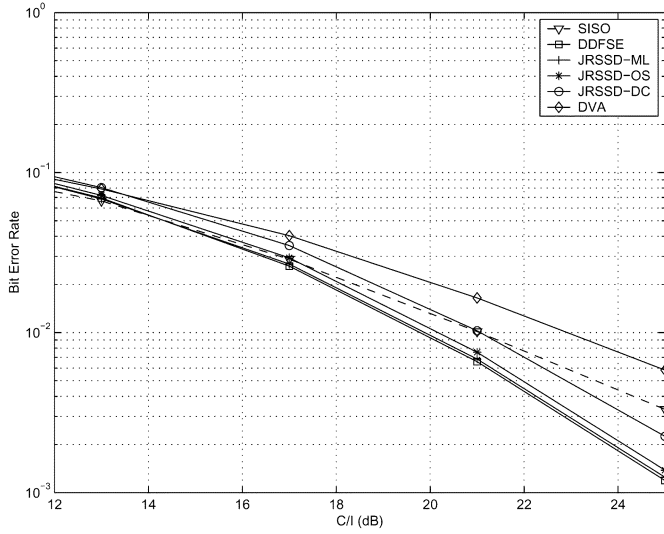
Fig. 8.   BER comparison for an interference-limited TU50 channel, $\mathrm{DIR} = 15$ dB.
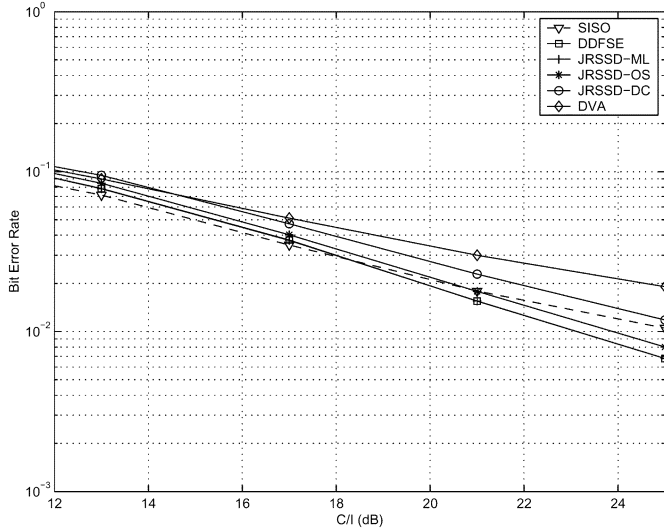

Fig. 9.   BER comparison for an interference-limited RA100 channel, $\mathrm{DIR} = 15$ dB.

that the relative performance comparisons for the MIMO algorithms resemble those of the noise-limited case. However, it is observed that in the interference-limited situation, the MIMO results actually slightly under-perform the SISO result at the low C/I range for both TU and RA channels, and there is a cross-over point after which MIMO starts to outperform. One major reason for this behavior might be due to the fact that in the presence of strong interference, the degradation in the quality of both the channel estimate and the noise correlation estimate is greater for a MIMO configuration. Consequently, the benefit of the interference suppression algorithm described in the previous paragraph is less significant. One possible means of improving these estimates is to increase the length of the training sequences in a MIMO configuration.

VIII. CONCLUSION

In this paper, we propose a MIMO JRSSD joint equalizer that achieves near-optimal performance at a very low complexity. We show that two elements are essential in achieving the desired

complexity and performance tradeoff in the proposed JRSSD: 1) a multisymbol USP set-partition that retains the geometric symmetry; and 2) a space-time structure that effectively decouples the spatial and temporal processing. In an example EDGE system with a set-partition of $J_1 = 2^2$, the ED computations required by the JRSSD-OS is a mere $\mathcal{O}(16\mathcal{E})$, which is a dramatic reduction from $\mathcal{O}(4096\mathcal{E})$, the complexity of a conventional DDFSE receiver. Meanwhile, the performance penalty associated with JRSSD-OS ranges from 0–0.7 dB for the channel conditions studied.

APPENDIX A
PROOF OF PROPOSITION 1

We first define a partition of $\mathbf{R}$ and the corresponding partitioned LU decomposition

$$
\mathbf{R} \equiv \begin{bmatrix} \mathbf{R}_m & \mathbf{R}_t \\ \mathbf{R}_t^H & \mathbf{R}_s \end{bmatrix}
$$
$$
= \begin{bmatrix} \mathbf{L}_m & \\ \mathbf{L}_t & \mathbf{L}_s \end{bmatrix} \begin{bmatrix} \mathbf{D}_m & \\ & \mathbf{D}_s \end{bmatrix} \begin{bmatrix} \mathbf{U}_m & \mathbf{U}_t \\ & \mathbf{U}_s \end{bmatrix} \quad (24)
$$

which provides the LU decomposition of the primary submatrix $\mathbf{R}_m$: $\mathbf{R}_m = \mathbf{L}_m \mathbf{D}_m \mathbf{U}_m$. From (16), we have $\gamma_m^{OS} = 1/d_m = 1/(\mathbf{D}_m)_{m,m}$. Now all we need to do is to find $d_m = (\mathbf{D}_m)_{m,m}$. To this end, notice

$$
\mathbf{R}_m^{-1} = \mathbf{U}_m^{-1} \mathbf{D}_m^{-1} \mathbf{L}_m^{-1} \quad (25)
$$

where $\mathbf{U}_m^{-1}$ and $\mathbf{L}_m^{-1}$ are upper and lower triangular matrices with unit diagonals, respectively. Therefore, from (25) we get $(\mathbf{D}_m^{-1})_{m,m} = (\mathbf{R}_m^{-1})_{m,m}$ after some algebra. Furthermore, we know $1/d_m = (\mathbf{D}_m^{-1})_{m,m}$ since $\mathbf{D}_m$ is diagonal. Therefore, we have proved that

$$
\gamma_m^{OS} = \frac{1}{d_m} = (\mathbf{R}_m^{-1})_{m,m}. \quad (26)
$$

APPENDIX B
PROOF OF PROPOSITION 2

The first inequality is easy to show since by the matrix inversion theorem [20], we have

$$
\gamma_m^{OS} = (\mathbf{R}_m^{-1})_{m,m} = \frac{1}{r_{m,m} - \mathbf{c}^H \mathbf{G}^{-1} \mathbf{c}} \quad (27)
$$

where

$$
\mathbf{R}_m = \begin{bmatrix} \mathbf{G} & \mathbf{c} \\ \mathbf{c}^H & r_{m,m} \end{bmatrix} \quad (28)
$$

and $\mathbf{G}$ is a positive semidefinite matrix. It follows that $\gamma_m^{DC} \leq \gamma_m^{OS}$. On the other hand, the second inequality is more difficult to prove as we need to show $\gamma_m^{OS} = (\mathbf{R}_m^{-1})_{m,m} \leq (\mathbf{R}^{-1})_{m,m} = \gamma_m^{ML}$. To this end, we obtain the following partition for both $\mathbf{R}$ and $\mathbf{R}^{-1}$:

$$
\mathbf{R} \equiv \begin{bmatrix} \mathbf{R}_m & \mathbf{R}_t \\ \mathbf{R}_t^H & \mathbf{R}_s \end{bmatrix} \quad \mathbf{R}^{-1} \equiv \begin{bmatrix} \mathbf{V}_m & \mathbf{V}_t \\ \mathbf{V}_t^H & \mathbf{V}_s \end{bmatrix}. \quad (29)
$$

Furthermore, we construct the following "block" LU decomposition of $\mathbf{R}^{-1}$ where $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are unknown submatrices

$$
\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{A} \\ & \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{B} & \\ & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \\ \mathbf{A}^H & \mathbf{C}^H \end{bmatrix}
$$
$$
= \begin{bmatrix} \mathbf{B} - \mathbf{A}\mathbf{A}^H & \mathbf{A}\mathbf{C}^H \\ \mathbf{C}\mathbf{A}^H & \mathbf{C}\mathbf{C}^H \end{bmatrix}. \quad (30)
$$

Considering (29) and (30) together, we get $\mathbf{C} = \mathbf{C}^H = \mathbf{V}_s^{1/2}$, $\mathbf{A} = \mathbf{V}_t \mathbf{C}^{-1/2}$ and $\mathbf{B} = \mathbf{V}_m - \mathbf{A}\mathbf{A}^H = \mathbf{V}_m - \mathbf{V}_t \mathbf{V}_s^{-1} \mathbf{V}_t^H$. Furthermore, by inverting both side of (30) we get

$$\mathbf{R} = (\mathbf{R}^{-1})^{-1} = \begin{bmatrix} \mathbf{I} & \\ \tilde{\mathbf{A}} & \tilde{\mathbf{C}} \end{bmatrix} \begin{bmatrix} \mathbf{B}^{-1} & \\ & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \tilde{\mathbf{A}}^H \\ & \tilde{\mathbf{C}}^H \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{B}^{-1} & \mathbf{B}^{-1}\tilde{\mathbf{A}}^H \\ \tilde{\mathbf{A}}\mathbf{B}^{-1} & \tilde{\mathbf{A}}\mathbf{B}^{-1}\tilde{\mathbf{A}}^H + \tilde{\mathbf{C}}\tilde{\mathbf{C}}^H \end{bmatrix} \quad (31)$$

where $\tilde{\mathbf{A}}$, $\tilde{\mathbf{C}}$ can be solved as functions of $\mathbf{A}$ and $\mathbf{C}$. Comparing (31) and (29) and we have $\mathbf{R}_m = \mathbf{B}^{-1}$, which leads to $\mathbf{R}_m^{-1} = \mathbf{B} = \mathbf{V}_m - \mathbf{V}_t \mathbf{V}_s^{-1} \mathbf{V}_t^H$. Note $\mathbf{V}_t \mathbf{V}_s^{-1} \mathbf{V}_t^H$ is a positive semi-definite matrix and has nonnegative diagonals. Therefore, we get $(\mathbf{R}_m^{-1})_{m,m} \leq (\mathbf{V}_m)_{m,m} = (\mathbf{R}^{-1})_{m,m}$ and proved $\gamma_m^{OS} \leq \gamma_m^{ML}$.

REFERENCES

[1] M. V. Eyuboglu and S. U. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. 36, no. 1, pp. 13–20, Jan. 1988.

[2] P. Schramm *et al.*, "Radio interface performance of EDGE, a proposal for enhanced data rates in existing digital cellular system," in *Proc. IEEE Vehicular Technological Conf.*, vol. 2, May 1998, pp. 1064–1068.

[3] W. V. Etten, "Maximum-likelihood receiver for multiple channel transmission systems," *IEEE Trans. Commun.*, vol. COM-24, no. 2, pp. 276–284, Feb. 1976.

[4] N. Al-Dhahir and J. M. Cioffi, "MMSE decision feedback equalizers: Finite-length results," *IEEE Trans. Inf. Theory*, vol. 41, no. 7, pp. 961–975, Jul. 1995.

[5] A. Duel-Hallen and C. Hegaard, "Delayed decision feedback sequence estimation," *IEEE Trans. Commun.*, vol. 37, no. 5, pp. 428–436, May 1989.

[6] C. Tidestav, M. Sternad, and A. Ahlen, "Reuse within a cell—Interference rejection or multiuser detection?," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1511–1522, Oct. 1999.

[7] N. Al-Dhahir and A. H. Sayed, "The finite-length multiinput multioutput MMSE-DFE," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2921–2936, Oct. 2000.

[8] C. Miller, D. Taylor, and P. Gough, "Estimation of co-channel signals with linear complexity," *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1997–2005, Nov. 2001.

[9] J. Zhang, A. Sayeed, and B. VanVeen, "Low complexity MIMO receiver with decoupled detection," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2002, pp. 313–317.

[10] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 1, pp. 55–66, Jan. 1982.

[11] S. S. Pietrobon, R. H. Deng, A. Lafanechere, G. Ungerboeck, and D. J. Costello, "Trellis-coded multidimensional phase modulation," *IEEE Trans. Inf. Theory*, vol. 36, no. 1, pp. 63–89, Jan. 1990.

[12] L. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 7, pp. 483–500, Jul. 1987.

[13] J. Zhang, H. Berg, A. Sayeed, and B. VanVeen, "Reduced-state MIMO sequence estimation for EDGE systems," in *Proc. 36th Asilomar Conf. Signals, Systems and Computers*, 2002, pp. 541–545.

[14] N. Benvenuto, R. Sandre, and G. Sostrato, "Reduced-state maximum-likelihood multiuser detection for down-link TD-CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 1, pp. 264–272, Feb. 2002.

[15] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: Architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. ISSSE*, 1998, pp. 295–300.

[16] N. Al-Dhahir, "FIR channel-shortening equalizers for MIMO ISI channels," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 213–218, Feb. 2001.

[17] W. H. Gerstacker and R. Schober, "Equalization concepts for EDGE," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 190–199, Jan. 2002.

[18] "GSM 05.05 Radio Transmission and Reception, Version 6.3.0," ETSI, 1997.

[19] P. A. Ranta and A. Lappetelainen, "Application of dominant interference cancellation in street microcells," in *Proc. ICC*, 1997, pp. 946–948.

[20] L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.

Jianzhong Zhang (S'96–M'04) received the B.S. degrees in electrical engineering and applied physics from Tsinghua University, Beijing, China in 1995, the M.S. degree in electrical engineering from Clemson University, Clemson, SC, in 1998, and the Ph.D. degree in electrical engineering from University of Wisconsin at Madison in 2003.

His research has focused on the application of statistical signal processing methods to wireless communication problems. Since 2001, he has been with Nokia research Center, Irving, TX, where he is currently a Senior Research Engineer. He has worked on the transceiver designs for both GSM/EDGE and CDMA2000/WCDMA cellular systems from 2001 to 2004. Since July 2004, he has been leading Nokia's physical layer contributions for the IEEE 802.16e standard.

Akbar M. Sayeed (S'89–M'97–SM'02) received the B.S. degree from the University of Wisconsin at Madison in 1991, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign, in electrical and computer engineering in 1993 and 1996, respectively.

While at the University of Illinois, he was a Research Assistant in the Coordinated Science Laboratory and was also the Schlumberger Fellow in signal processing from 1992 to 1995. During 1996–1997, he was a Postdoctoral Fellow at Rice University, Houston, TX. Since August 1997, he has been with the University of Wisconsin at Madison, where he is currently Associate Professor of Electrical and Computer Engineering. His research interests are in wireless communications, statistical signal processing, communication theory, information theory and time-frequency analysis.

Dr. Sayeed received the NSF CAREER Award in 1999 and the ONR Young Investigator Award in 2001. He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1999 to 2002.

Barry D. Van Veen (S'81–M'86–SM'97–F'02) was born in Green Bay, WI. He received the B.S. degree from the Michigan Technological University, Houghton, in 1983, and the Ph.D. degree from the University of Colorado at Boulder, in 1986, both in electrical engineering. He was an Office of Naval Research (ONR) Fellow while working toward the Ph.D. degree.

In Spring 1987, he was with the Department of Electrical and Computer Engineering, University of Colorado at Boulder. Since August 1987, he has been with the Department of Electrical and Computer Engineering, University of Wisconsin–Madison, where he is currently a Professor. He coauthored *Signals and Systems* (New York: Wiley, 1999, 1st ed.; 2003, 2nd ed.). His research interests include signal processing for sensor arrays, wireless communications, and biomedical applications of signal processing.

Dr. Van Veen was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE Signal Processing Society's Statistical Signal and Array Processing Technical Committee and the Sensor Array and Multichannel Technical Committee. He was a recipient of a 1989 Presidential Young Investigator Award presented by the National Science Foundation and a 1990 IEEE Signal Processing Society Paper Award. He was also the recipient of the Holdridge Teaching Excellence Award presented by the Department of Electrical and Computer Engineering, University of Wisconsin–Madison, in 1997.