Reduced-Complexity Soft MIMO Detection Based on Causal and Noncausal Decision Feedback

Yong Li and Jaekyun Moon, *Fellow, IEEE*

***Abstract*—We present a reduced-complexity soft detection (RCSD) scheme geared to multiple-input multiple-output (MIMO) systems with spatial domain multiplexing leading to layered space-time or space-frequency architecture. The proposed algorithm relies on a trellis representation of the MIMO signals and a subsequent formulation of constrained-depth maximum a posterior (MAP) detection in conjunction with soft decision feedback (SDF). Decision feedback is broken into causal and noncausal parts in an effort to maximize the observation window while maintaining a reasonable computational load. Two variations of RCSD are proposed to utilize the noncausal information in different ways. Analysis based on mean squared error (MSE) and log-likelihood ratio (LLR) associated with symbols in different stages of the trellis is done to develop insight into error propagation in the RCSD algorithm, as well as to compare the quality of the soft information obtained by the two RCSD variants. Error-rate simulations are conducted in the context of turbo-like iterative demapping and decoding (IDD). The resulting performance and required complexity are compared with those of maximum a posterior (MAP) detection, soft sphere detection (SD), as well as the Turbo-BLAST processing scheme. We observe favorable performance/complexity tradeoffs with the proposed soft detection scheme for a number of modulation/channel scenarios.**

***Index Terms*—Iterative Decoding and Demodulation (IDD), multiple-input multiple-output (MIMO), reduced complexity soft detection (RCSD), soft decision feedback (SDF), spatial multiplexing.**

I. Introduction

It is well known that in rich-scattering environments multiple-input multiple-output (MIMO) communication systems provide much higher spectral efficiency than single-input single-output systems [1]. However, with a large number of antennas and/or high-order modulation methods needed to achieve very high data rates, optimal maximum a posterior (MAP) detection [or maximum likelihood (ML) detection if equally probably symbols are assumed] is often impractical due to high complexity. Therefore, developing a reduced-complexity detection algorithm without significant performance degradation is a critical issue in the receiver design for MIMO systems.

A popular early MIMO architecture is the Bell Laboratories' layered space-time (BLAST) system. The original architecture was proposed in [1] and is known as diagonal BLAST (D-BLAST). A simplified version based on spatial multiplexing known as vertical BLAST (V-BLAST) was proposed in [2], and a detection algorithm has also been provided based on three elements: linear interference nulling, successive interference cancellation (SIC), and optimized ordering. The idea of SIC is also well established in multiuser detection (see [3] and [4], for example) and can be viewed as decision feedback equalization (DFE) in spatial domain [5]. Some variations of this algorithm have been proposed utilizing either the zero-forcing (ZF) [6] or minimum mean-square error (MMSE) [7] nulling criterion. Other approaches that alleviate the computational burden of optimized ordering have also been proposed [8]–[10].

Although an appropriate ordering assures that the detection process starts from the "best" symbol in terms of signal-to-noise ratio (SNR), in BLAST-type detection the decisions that are fed back always come from "weaker" symbols with a lower order of diversity, often resulting in catastrophic error propagation. Different strategies have been investigated to alleviate this effect, such as parallel detection [11], two-directional QR detection [12], and group detection [13], [14]. Making use of the lattice property of the quaternary phase-shift keying (QPSK) and quadrature-amplitude modulation (QAM) modulations, sphere decoding ([15]–[20]) performs the closest-point search to approximate MAP detection, and a lattice reduction technique [21], [22] can be employed to aid the detection process.

Most techniques described above are not directly applicable to coded systems composed of a bit-to-symbol mapper/demapper and the outer encoder/decoder. Such coded systems usually employ iterative demapping and decoding (IDD) [23], sometimes referred to as Turbo-BLAST [24], which requires the detector or demapper to provide reliable soft information rather than simple hard decisions to the outer decoder. Accordingly, a need for soft-input soft-output (SISO) detectors arises. One approach is to modify the sphere decoding to keep a list of candidate output sequences instead of just the "nearest" word in order to generate the soft information [25], [26]. The soft outputs will be sufficiently reliable when the search radius is chosen large enough, although a larger radius implies higher decoding complexity. Some other suboptimal approaches, based on tree search [27], semidefinite programming (SDP) [28], probabilistic data association (PDA) [29], etc., are shown to achieve similar performance to sphere decoding with comparable complexity. Another possible approach, which has not been widely applied to MIMO detection, comes from the idea of reduced-state sequence estimation (RSSE) for intersymbol interference (ISI) channels. These complexity reduction algorithms are based on either a truncated channel impulse [30] or set partitioning of the symbol constellation [31] or both [32]; these schemes all rely on decision

Manuscript received August 14, 2006; revised August 2, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Javier Garcia-Frias.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: moon@umn.edu; lixx0266@umn.edu).

Digital Object Identifier 10.1109/TSP.2007.909048

feedback to resolve ambiguity caused by the residual channel impulse response or the presence of multiple symbols within a partition. A "survivor path" [33] of past hard-decisions needs to be kept for each state in the reduced trellis. There also exist SISO versions of reduced-state algorithms, as discussed in [34] and [35], for example.

In this paper, we shall focus on the derivation of a reduced-complexity soft detection (RCSD) algorithm well suited to MIMO detection, based on the constrained-depth[1] approximation of the MAP detection [36]. The proposed suboptimal, low-complexity soft detection scheme is based on trellis representation of the MIMO signals and a subsequent formulation of constrained-depth MAP detection as well as the use of soft decision feedback (SDF). Decision feedback is applied based on decomposition of the MIMO interference into causal and noncausal parts. The causal interference is in turn broken into the state-dependent portion and the past contribution. The complexity of the proposed algorithm depends mainly on the size of the trellis, which is in turn set by the span of interference associated with a prescribed, parameterized number of preceding interference terms. The overall decision depth, on the other hand, coincides with the interference span covering both the state-dependent portion and the noncausal part. Two different variations are suggested for handling the noncausal interference, based on *a priori* information or hard decisions, respectively. Error rate simulations are conducted based on the IDD procedure on different modulation schemes. The resulting performance and required complexity are compared with those of MAP detection, soft Turbo-BLAST processing, as well as sphere detection (SD). Excellent performance/complexity tradeoffs were observed with the proposed soft detection scheme for both suggested RCSD variations.

This paper is organized as follows. Section II formulates the MIMO signals in a trellis form and derives the new RCSD algorithm with two variations. Section III conducts analysis based on mean-square error (MSE) and log-likelihood ratio (LLR) for symbols at different stages of the trellis. In Section IV simulations are done to verify the performance behavior of RCSD expected from the analysis, and the tradeoff of the resulting performance and required complexity is compared between the proposed RCSD and some existing detection schemes. Finally, conclusions are drawn made in Section V.

II. Proposed RDSD

The system block diagram is shown in Fig. 1, which depicts a bit-interleaved coded modulation (BICM) scheme with spatial multiplexing (SM) and IDD processing. The OFDM operations at transmitter (TX) and receiver (RX) are composed of inverse fast Fourier transmission (IFFT) and FFT as well as cyclic prefix (CP) insertion/removal. In flat fading channels, the OFDM operations will be skipped. The preprocessing filter is to shape the MIMO channel to be lower-triangular (or upper-triangular) for the convenience of layer-based detection. The

[1]"Depth" refers to the number of adjacent layers that the detector looks into for the current-layer detection. It is in the sense of the spatial domain versus the time-domain. The same will be true for any reference of "causal" or "noncausal" part of interference.

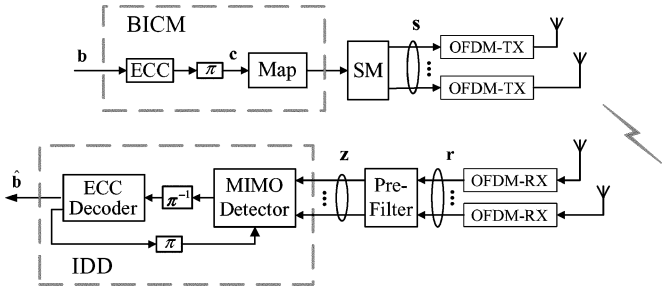

Fig. 1. Bit-interleaved spatial multiplexing scheme with IDD processing.

IDD process is performed with the extrinsic information exchanged between the MIMO detector and the error correction code (ECC) decoder.

Consider an $N \times N$ antenna configuration (N transmit and N receive antennas). While the techniques discussed in this paper can definitely be applied to unbalanced cases, we assume that after proper signal decomposition,[2] we effectively have an $N \times N$ antenna configuration, where N is the minimum of the numbers of transmit and receive antennas. The well-known vector MIMO channel can be rewritten in a lower-triangular form as

$$z_i = H_{ii}s_i + H_{i,i-1}s_{i-1} + \cdots + H_{i1}s_1 + n_i, \quad 1 \le i \le N \quad (1)$$

where z_i is the observation signal at the ith receive antenna, s_j is the modulated symbol transmitted by the jth transmit antenna, and H_{ij} represents the effective channel response of the communication link between the jth transmit antenna and the ith receive antenna. n_i denotes the post-filtering noise for the ith receive antenna, and is assumed to be Gaussian with zero-mean and variance σ^2. Note that n_i could include some residual interferences after MMSE filtering, but for simplicity here σ^2 is always assumed to be the noise variance associated with the system SNR. The received signals in (1) are accurately represented by a tree, since there is no merging of signal paths until the last layer. Optimal MAP/ML detection is based on processing all M^N paths of the tree (M is the symbol constellation size), and is not practical when M and/or N is large. However, as will be shown in the following, when the interference length is truncated, low-complexity algorithms can be applied based on a trellis representation of the received signals in a MIMO system.

For a given layer j, we can make a soft estimation on the symbol τ-layer back, $s_{j-\tau}$, based on z_1^j, the collection of observation samples z_1 through z_j, plus $z_{j+1}^{j+\lambda}$, the observation samples through λ-layer forward. In other words, we aim to compute or estimate the constrained-depth *a posteriori* probability (APP) $P(s_{j-\tau}/z_1^{j+\lambda})$.

The state and state transition of the trellis representing MIMO signals can be defined respectively as

$$\begin{aligned} v_j &\triangleq s_{j-\tau}^{j-1}, \quad j \ge \tau + 1 \\ \eta_j &\triangleq \{v_j, v_{j+1}\} = s_{j-\tau}^{j}, \quad j \ge \tau + 1. \end{aligned} \quad (2)$$

[2]For ZF, the QR decomposition will result in the exact form of (1); for MMSE, there is noncausal interference, but the causal form of (1) can still be assumed at the receiver by considering the residual interference as part of the noise.

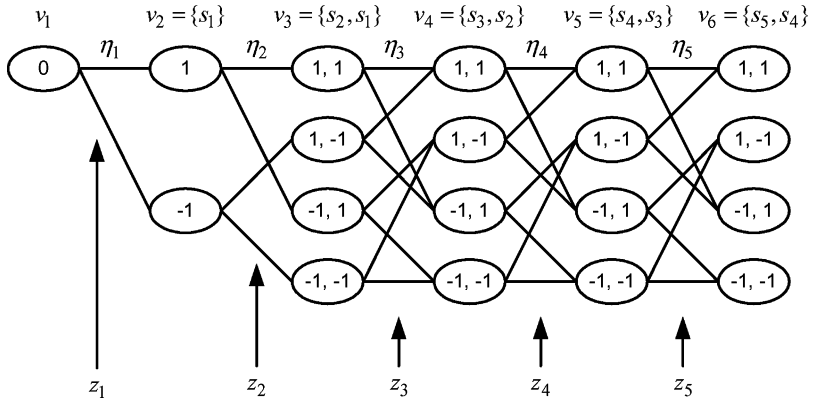

Fig. 2. Reduced trellis for BPSK, $\tau = 2$ and 5×5 antenna configuration.

where $s_{j-\tau}^{j-1}$ represents the collection of symbols from layer $(j-\tau)$ to layer $(j-1)$. For $1 \le j \le \tau$, we have a spatially varying trellis with the state at the jth layer given by $v_j \triangleq s_1^{j-1}$ (v_1 is a dummy node, where all trellis paths originate). See Fig. 2 for the resulting trellis corresponding to BPSK, $\tau = 2$ and $N = 5$. It is seen that the state transition is defined by extending the state for one more layer, thus all transitions originating from the same state differ only in one constellation symbol.

The joint probability density function (pdf) $f(s_{j-\tau}, z_1^{j+\lambda})$ is obtained as

$$f\left(s_{j-\tau}, z_1^{j+\lambda}\right) = \sum_{\eta_j(s_{j-\tau})} f\left(\eta_j, z_1^{j+\lambda}\right) = \sum_{\eta_j(s_{j-\tau})} f\left(\eta_j, z_1^{j}\right) f\left(z_{j+1}^{j+\lambda}/\eta_j, z_1^{j}\right) \quad (3)$$

for $1+\tau \le j \le N+\tau$, where the sum is over all state transitions associated with the given $s_{j-\tau}$. In computing the joint probability density for the last symbols $s_{N-\tau+1}$ through s_N using (3) (i.e., when $N+1 \le j \le N+\tau$), it is understood that $\eta_{N+m} = \eta_N$ and $z_1^{N+m} = z_1^N$ whenever $m > 0$. The joint probability density $f(\eta_j, z_1^j)$, the "causal part" which only depends on the observation samples z_1^j, can be computed recursively, as will be described next. Exact computation of the "noncausal part" $f(z_{j+1}^{j+\lambda}/\eta_j, z_1^j)$ is hard, but can be reasonably approximated based on the available *a priori* information as well as decision feedback of the already detected symbols.

A. Computing $f(\eta_j, z_1^j)$

Based on the definitions in (2), we have

$$f\left(\eta_j, z_1^j\right) = P_a(s_j) \cdot f\left(z_j/\eta_j, z_1^{j-1}\right) \cdot \sum_{\eta_{j-1}(\eta_j)} f\left(\eta_{j-1}, z_1^{j-1}\right) \quad (4)$$

where the sum is over all transitions η_{j-1} that are connected to the particular transition η_j, $\sum_{\eta_0(\eta_1)} f(\eta_0, z_1^0)$ is initialized to 1, and the *a priori* probability $P_a(s_j)$ is to be updated at each iterative stage in the case of IDD. Let us now focus on the likelihood function defined as $l(\eta_j) = f(z_j/\eta_j, z_1^{j-1})$. While a state transition η_j is designed to collect only $s_{j-\tau}^j$, the observation z_j does in general depend on all symbols s_1 through s_j. This issue can be resolved by writing the likelihood function as follows:

$$l(\eta_j) = \sum_{s_1^{j-\tau-1}} f\left(s_1^{j-\tau-1}, z_j/\eta_j, z_1^{j-1}\right) = \sum_{s_1^{j-\tau-1}} P\left(s_1^{j-\tau-1}/\eta_j, z_1^{j-1}\right) f\left(z_j/s_1^{j-\tau-1}, \eta_j\right) \quad (5)$$

for $\tau+1 \le j \le N$; for $1 \le j \le \tau$, we simply have $l(\eta_j) = f(z_j/s_1^j)$ (see Fig. 2). It can be seen that the likelihood function is a weighted sum of Gaussian pdfs. Exact computation of $l(\eta_j)$ need to enumerate all combinations of s_1^j, which will considerably increase the computational load.

One effective way of reducing computation is to force an approximation

$$l(\eta_j) \approx f_\sigma\left(z_j/\bar{s}_1^{j-\tau-1}, \eta_j\right) \quad (6)$$

where $\bar{s}_1^{j-\tau-1}$ represent soft-decision symbols (average decisions as will be discussed shortly) and the subscript σ emphasizes that the noise variance is now adjusted to account for the quality of the soft decisions. To be more clear, we are making an assumption that $l(\eta_j)$ is a Gaussian pdf with some mean μ_{T_j} and variance $\sigma_{T_j}^2$, which depend on the soft decisions. Following derivation steps similar to those in [37], it can be shown that the "total mean" and the "total variance" are

$$\mu_{T_j} = \sum_{k=\tau+1}^{j-1} H_{j,j-k}\bar{s}_{j-k} + \sum_{k=0}^{\tau} H_{j,j-k}s_{j-k}$$
$$\sigma_{T_j}^2 = \sigma^2 + \sum_{k=\tau+1}^{j-1} |H_{j,j-k}|^2 \sigma_{j-k}^2 \quad (7)$$

for $1 \le j \le N$, where $\bar{s}_l = \sum_{s_l} P(s_l)s_l$, σ^2 is the variance of additive noise and $\sigma_l^2 = \sum_{s_l} P(s_l)|s_l - \bar{s}_l|^2$. It is implied that $\bar{s}_l = 0$ and $\sigma_l = 0$ for $l < 1$ or $l > N$. Since the probabilities of s_ls are not available, we make approximations $\bar{s}_l \approx \sum_{s_l} P(s_l/z_1^{l+\tau+\lambda}) \cdot s_l$, and $\sigma_l^2 \approx \sum_{s_l} P(s_l/z_1^{l+\tau+\lambda}) \cdot |s_l - \bar{s}_l|^2$ for $1 \le l \le j-\tau-1$ based on a posteriori probabilities available from the previous detection layers during the current IDD iteration. Note that another underlying approximation here is that $P(s_1^{j-\tau-1}/\eta_j, z_1^{j-1}) \approx \prod_{k=1}^{j-\tau-1} P(s_k/z_1^{k+\tau})$. The first

sum in the right hand side of the first equation in (7) represents interference from the distant past, whereas the second sum corresponds to the more recent interference terms implied in the given state. The state is not defined when $\tau = 0$ (i.e., no trellis), in which case the scheme simply reduces to essentially symbol-by-symbol detection even when $\lambda > 0$.

We also note that for $\lambda = 0$, the proposed scheme is to some extent similar to the SIC algorithm in [38] which also relies on the Gaussian approximation of the residual interference plus noise. The SIC algorithm of [38] conducts full MAP detection for the first p layers and performs layer-by-layer detection for subsequent layers. In contrast, our scheme treats all layers more evenly with a consistent depth of τ layers, based on which a trellis is constructed with a constant number of state transitions. We also extend the depth to λ layers ahead to enhance performance.

B. Computing $f(z_{j+1}^{j+\lambda}/\eta_j, z_1^j)$

Similar to the idea implied in (6), which assumes that the summation over all combinations of possible symbols can be approximated by a single term as long as a particular decision (soft or hard) for this symbol set has been made, the noncausal part can be expressed as

$$\begin{aligned} f\left(z_{j+1}^{j+\lambda}/\eta_j, z_1^j\right) &= \sum_{s_1^{j-\tau-1}, s_{j+1}^{j+\lambda}} f\left(s_1^{j-\tau-1}, s_{j+1}^{j+\lambda}, z_{j+1}^{j+\lambda}/\eta_j, z_1^j\right) \\ &= \sum_{s_1^{j-\tau-1}, s_{j+1}^{j+\lambda}} f\left(z_{j+1}^{j+\lambda}/s_1^{j-\tau-1}, \eta_j, s_{j+1}^{j+\lambda}\right) \\ &\quad \times P\left(s_1^{j-\tau-1}, s_{j+1}^{j+\lambda}/\eta_j, z_1^j\right) \\ &\approx f_\sigma\left(z_{j+1}^{j+\lambda}/\bar{s}_1^{j-\tau-1}, \eta_j, \tilde{s}_{j+1}^{j+\lambda}\right) \\ &= \prod_{i=1}^{\lambda} f_\sigma\left(z_{j+i}/\bar{s}_1^{j-\tau-1}, \eta_j, \tilde{s}_{j+1}^{j+i}\right) \end{aligned} \tag{8}$$

where the approximation is the result of replacing the random variables $s_1^{j-\tau-1}$ and $s_{j+1}^{j+\lambda}$ with the deterministic variables $\bar{s}_1^{j-\tau-1}$ and $\tilde{s}_{j+1}^{j+\lambda}$, and each term in the product in the last line is a single Gaussian pdf, i.e.

$$f_\sigma\left(z_{j+i}/\bar{s}_1^{j-\tau-1}, \eta_j, \tilde{s}_{j+1}^{j+i}\right) \sim N\left(\mu_{T_{j+i}}, \sigma_{T_{j+i}}\right), \quad 1 \le i \le \lambda$$

where the mean and variance can be obtained as

$$\begin{aligned} \mu_{T_{j+i}} &= \sum_{k=\tau+1}^{j-1} H_{j,j-k}\bar{s}_{j-k} + \sum_{k=0}^{\tau} H_{j,j-k}s_{j-k} \\ &\quad + \sum_{k=1}^{i} H_{j,j+k}\tilde{s}_{j+k} \\ \sigma_{T_{j+i}}^2 &= \sigma^2 + \sum_{k=\tau+1}^{j-1} |H_{j,j-k}|^2\sigma_{j-k}^2 \\ &\quad + \sum_{k=1}^{i} |H_{j,j+k}|^2\sigma_{j+k}^2. \end{aligned} \tag{9}$$

Note that $\bar{s}_1^{j-\tau-1}$ are the soft-decision symbols obtained from detection of the previous layers as in (7), while $\tilde{s}_{j+1}^{j+\lambda}$ represent either hard or soft decisions that are not yet available. In the following, we discuss two possible ways to obtain $\tilde{s}_{j+1}^{j+\lambda}$ based on the available information.

1) Based on a Priori Information: In this scheme $\tilde{s}_{j+1}^{j+\lambda}$ are obtained in a way similar to how $\bar{s}_1^{j-\tau-1}$ is obtained: we average over all possible $s_{j+1}^{j+\lambda}$s according to the *a priori* probabilities for $s_{j+1}^{j+\lambda}$, i.e.

$$\begin{aligned} \tilde{s}_{j+i} &\approx \sum_{s_{j+i}} P_a(s_{j+i}) \cdot s_{j+i} \\ \sigma_{j+i}^2 &\approx \sum_{s_{j+i}} P_a(s_{j+i}) \cdot |s_{j+i} - \tilde{s}_{j+i}|^2 \end{aligned} \tag{10}$$

for $1 \le i \le \lambda$, where $P_a(s_{j+i})$ represents the *a priori* probability estimation from the last iteration stage in IDD. This method shall be referred to as RCSD1.

2) Based on Hard Decision: Another way to obtain $\tilde{s}_{j+1}^{j+\lambda}$ is to choose a particular pattern $\tilde{s}_{j+1}^{j+\lambda}$ that maximizes (8), i.e.

$$\tilde{s}_{j+1}^{j+\lambda} = \arg\max_{s_{j+1}^{j+\lambda}} f_\sigma\left(z_{j+1}^{j+\lambda}/\bar{s}_1^{j-\tau-1}, \eta_j, s_{j+1}^{j+\lambda}\right). \tag{11}$$

When $\lambda = 1$, $\tilde{s}_{j+1}$ can be easily obtained by "quantizing" z_{j+1} to its nearest neighboring symbol, requiring little arithmetic operations; As λ grows, the maximization in (11) over all possible $s_{j+1}^{j+\lambda}$s can get computationally intense for high-order modulations. However, since only hard decisions for $\tilde{s}_{j+1}^{j+\lambda}$ are required, it can be conveniently found by the conventional SD method (for example, see [16]). Although this represents added complexity and entails same implementation issues of conventional hard-decision SD [18], we are only interested in small λs (in practice $\lambda \le 2$) for which the expected complexity is $O(\lambda^3)$ [17], much lower than the overall complexity of soft-decision SD. We shall call the resulting scheme RCSD2, which becomes identical to RCSD1 for $\lambda = 0$.

In RCSD2, $\tilde{s}_{j+1}^{j+\lambda}$ is chosen based on η_j and may take on different values for different η_j, while in RCSD1 $\tilde{s}_{j+1}^{j+\lambda}$ is predetermined by *a priori* probabilities and so is uniform irrespective of η_js. Furthermore, $\tilde{s}_{j+1}^{j+\lambda}$s in RCSD2 are hard decisions which are chosen from the symbol constellation, so they are not random variables and we have $\sigma_{j+k} = 0$ for $1 \le k \le \lambda$ in (9). As will be shown below, RCSD2 is better than RCSD1 in the cases of high-order constellations (e.g., 16/64-QAM) that possess a large number of distinct states.

C. Summary of the Algorithm

- Initialization ($j = 0$): $\eta_0 = v_1 = 0$,
$$f\left(\eta_0, z_1^0\right) = 1.$$
- For $1 \le j \le \tau$: $v_j \triangleq s_1^{j-1}$, $\eta_j \triangleq s_1^j$,
$$f\left(\eta_j, z_1^j\right) = P(s_j) \cdot f\left(z_j/s_1^j\right) \cdot f\left(\eta_{j-1}, z_1^{j-1}\right).$$

- For $\tau+1 \le j \le N+\tau$: $v_j \triangleq s_{j-\tau}^{j-1}$, $\eta_j \triangleq s_{j-\tau}^{j}$,
  1) Compute

$$f\left(\eta_j, z_1^{j+\lambda}\right) = P(s_j) \cdot \sum_{\eta_{j-1}(\eta_j)} f\left(\eta_{j-1}, z_1^{j-1}\right) \cdot \prod_{i=1}^{\lambda} f_\sigma\left(z_{j+i}/\bar{s}_1^{j-\tau-1}, \eta_j, \tilde{s}_{j+1}^{j+i}\right)$$

  with $\eta_{N+m} = \eta_N$, $z_1^{N+m} = z_1^N$, for $m > 0$.
  2) Compute $f(s_{j-\tau}, z_1^{j+\lambda})$ according to (3), so obtain $P(s_{j-\tau}/z_1^{j+\lambda})$, update $\{\bar{s}_{j-\tau}, \sigma_{j-\tau}^2\}$, and then go to 1).

Note that the RCSD algorithm is derived based on the signal form of (1), regardless of whether it is obtained based on the ZF or MMSE criterion. Although MMSE requires higher arithmetic complexity due to extra matrix inversion or suboptimal scalar inversion [24], it usually achieves better performance than ZF. In the sequel, we shall assume MMSE design.

III. Analysis and Comparison

A. Error Propagation Analysis

In this section, the MSE of symbols in each layer will be calculated for RCSD with different τ parameters, to visualize the error propagation effect due to decision feedback. Based on the signal model in (1), the MSE associated with each layer is defined as

$$\Delta_j = E\left[\left|s_j^\circ - \hat{s}_j\right|^2\right] + \sum_{i=1}^{r} E\left[\left|s_i^\circ - \bar{s}_i\right|^2\right], \quad 1 \le j \le N \tag{12}$$

where $r = \min(j, N-\tau) - 1$, s_j° denotes the actually transmitted jth-layer symbol, and $\bar{s}_i$s are the soft symbol decisions for previously detected layers. $\hat{s}_j$ is the jth-layer symbol that would be chosen if hard decisions had been made at this stage, based on the pdf obtained in (3), i.e.

$$\hat{s}_j = \arg\max_{s_j} f\left(s_j, z_1^{j+\tau}\right) \tag{13}$$

assuming $\lambda = 0$ for the sake of simplicity here. It is seen that the MSE at each layer is composed of the detection error of the current-layer symbol, and the error due to feedback of incorrect previous-layer symbols. Note that the last τ layer symbols are jointly detected, so the detection of s_j does not rely on decision feedback of $s_{N-\tau}^{j-1}$. That is why in (12) the errors associated with $s_{N-\tau}^{j-1}$ are not included in Δ_j for $N-\tau < j \le N$.

If all past decisions in (12) are precise, i.e., $\bar{s}_i = s_i^\circ$, we will have $\Delta_j = E[|s_j^\circ - \hat{s}_j|^2]$, so the MSE of all layers will be roughly the same. But with the presence of fading and noise, there is always a chance that $s_i^\circ \neq \bar{s}_i$, and the error on previous symbols will "propagate" to subsequent layers and affect detection at those layers. In Fig. 3, the average MSEs of RCSD schemes with different τs are shown for QPSK and 16QAM. Each plot is obtained for a particular E_b/N_0, which is the ratio of average energy per information bit to one-sided noise power spectral density. The average MSEs of BLAST and MAP detection are also shown as references. First, compared to BLAST

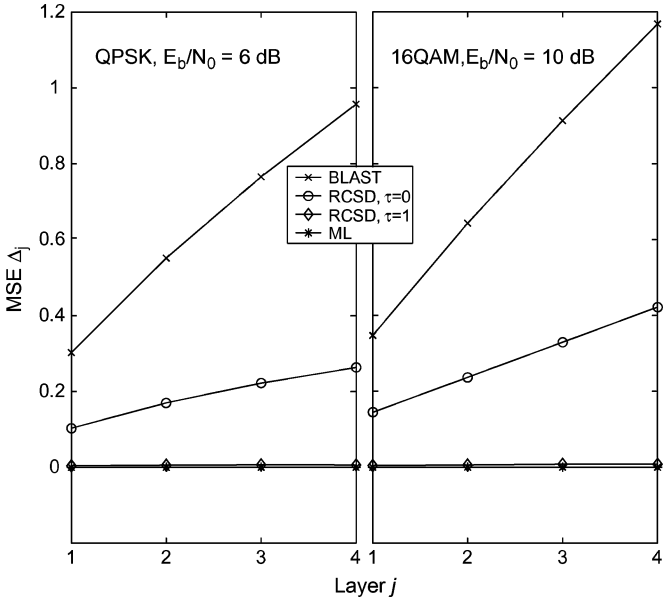

Fig. 3. MSEs of different layers for QPSK and 16 QAM, 4 × 4.

detection, error propagation in RCSD with $\tau = 0$ is considerably less; its MSE accumulates much more slowly as the detection process moves from layer-1 to layer-N. This is due to the improvement of SDF over hard decision feedback (HDF). Secondly, for a larger τ error propagation is reduced further, due to the fact that both $\hat{s}_j$ and $\bar{s}_1^{j-1}$ become more reliable when more layers are jointly considered. For RCSD with $\tau = 1$, error propagation is already negligible, and the resulting MSE is nearly indistinguishable from that of MAP detection.

B. Comparison of RCSD1 and RCSD2 Based on Mutual Information

It has been seen in Section II-B that two RCSD schemes proposed in this paper are distinguished in the way they utilize the non-causal layer information for $\lambda > 0$. This could result in different reliability of the soft outputs.

In general, the reliability of the soft bit information at demapper's output can be characterized by the mutual information (MI) between the coded bit c and its corresponding LLR Λ_c:

$$\begin{aligned} I(\Lambda_c; c) &= \frac{1}{2} \sum_{b=0,1} \int_{-\infty}^{+\infty} p_{\Lambda_c}(\xi|c=b) \\ &\quad \times \log_2 \frac{2p_{\Lambda_c}(\xi|c=b)}{p_{\Lambda_c}(\xi|c=0) + p_{\Lambda_c}(\xi|c=1)} d\xi \\ &\approx 1 - \frac{1}{P} \sum_{n=1}^{P} \log_2\left(1 + e^{-(2c_n-1)\cdot\Lambda_{c_n}}\right) \end{aligned} \tag{14}$$

where P is the number of coded bits, i.e., the codeword length, and the approximation is usually used in practice without knowing the actual distribution of Λ_c, as argued in [39].

MI has been utilized in the extrinsic information transfer (EXIT) chart [40], where MI at the input/output of both demapper and decoder is plotted versus the number of iterations for a given SNR, to analyze the convergence behavior

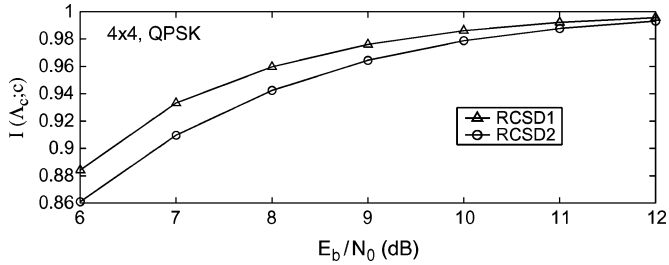

Fig. 4. Bitwise MI $I(\Lambda_c; c)$ at the demapper's output for QPSK, 4×4, RCSD with $\tau = 0, \lambda = 2$.

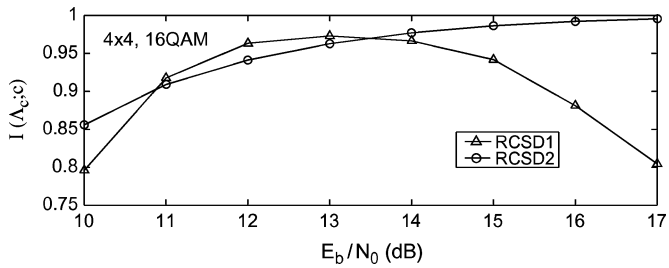

Fig. 5. Bitwise MI $I(\Lambda_c; c)$ at the demapper's output for 16 QAM, 4×4, RCSD with $\tau = 0, \lambda = 2$.

of some concatenation schemes. In our case, the EXIT chart is less useful, since the convergence behavior for a particular SNR is not much different between the two RCSD schemes, at least not in a very obvious way. To better serve our purposes, namely, to better visualize the evolution of soft bit information in two RCSD schemes, in Fig. 4 and Fig. 5 we depict the MI at demapper's output versus SNR with a sufficient number of iterations (5 iterations here), for both QPSK and 16QAM cases, respectively. These figures reflect the overall reliability of the soft information generated by RCSD1 and RCSD2.

It is seen that in QPSK RCSD1 consistently generates more reliable soft information than RCSD2, while in 16QAM the reliability of soft information generated by RCSD1 starts to drop beyond some SNR value. This is because RCSD1 heavily relies on the accuracy of the noncausal symbol estimates, which is harder to achieve for higher-order modulations. RCSD2, on the other hand, can still work with less accurate estimates as long as (11) is satisfied. This will also be verified by error rate simulation results next. It is safe to say that RCSD2 is expected to achieve better performance than RCSD1 with high-order modulation methods.

C. Complexity Analysis

In this section, we consider the proposed RCSD scheme, as well as Turbo-BLAST, sphere detection/decoding (SD). Basically, Turbo-BLAST employs a layer-by-layer detection with soft cancellation based on *a priori* information feedback from the outer decoder. Therefore, it has a similar complexity as RCSD with $\tau = 0$, since for each layer soft bit information is generated in the same manner, i.e., symbol estimation followed by symbol-to-bit conversion. However, in Turbo-BLAST MMSE filtering needs to be performed every iteration in order to perform the soft cancellation [24]. Unless some low-complexity schemes are used [41], [42], MMSE processing requires a matrix inversion per iteration and therefore is quite complex. On the other hand, in RCSD prefilter processing (based on ZF or MMSE) only needs to be done once at the first iteration, since the purpose here is to shape a lower-triangular signal model as (1), instead of trying to cancel the interfering symbols directly.

It is well known that the arithmetic complexity associated with the closest-point search in hard SD is asymptotically subject to an expected complexity requirement of $O(N^3)$ at high SNRs [17], regardless of M. However, in coded modulations, soft SD cannot just do the closest-point search, but must also keep a sufficiently large list of candidate symbol sequences to generate soft information for the decoder [25]. As a result, the search radius can not be reduced during the search as in hard SD, and so the search complexity of soft SD usually grows exponentially with N. Although the search effort is needed only once at the first iteration, the entire candidate list still has to be examined through to update the soft bit information at each following iteration. Therefore, the "enumeration complexity", defined as the candidate list size, denoted N_{cand}, in the case of SD or the number of transition edges per layer in the RCSD trellis in the case of RCSD, is a more relevant complexity measure than the one-time search complexity in coded systems with IDD, and is more likely to become the bottleneck especially for high M and/or N.

In summary, the enumeration size per-layer per-iteration in RCSD is $M^{\tau+1}$ (for $\tau = 0$, enumeration simply involves M constellation symbols). The MAP has an enumeration complexity of M^N, and is obviously a special case of RCSD when $\tau = N - 1$, while Turbo-BLAST has an enumeration complexity of M, the same as that of layer-by-layer detection of RCSD when $\tau = 0$. A similar tradeoff can be achieved by SD by appropriately choosing the search radius and N_{cand}.

IV. Simulated Performance

Simulations are based on a coded system composed of a rate-3/4 (133, 171) convolutional code (CC), specified in the 802.11a Wireless LAN (WLAN) standard [43], and QPSK/16-QAM/64-QAM modulation with Gray mapping. Th modulated symbols are spatially demultiplexed onto $N = 4$ transmit antennas. At the receiver, $N = 4$ receive antennas are also assumed and the IDD procedure with 5 iterations is applied to improve the performance. The demapper performs a typical symbol-to-bit probability conversion, which enumerates all symbols associated with a particular bit, as discussed, for example, in [23]. Two channel scenarios are investigated: for QPSK/16-QAM,a flat, fast-fading (fully interleaved) channel is assumed with Rayleigh distribution; for 64-QAM, a frequency-selective, quasi-static Rayleigh fading channel is assumed, which has an exponential-power profile with an rms delay spread of 150 ns. In the latter case, OFDM is also assumed with specifications as given in the 802.11a WLAN standard [43]. The packet-error-rate (PER) is chosen as the figures of merit, with packet size fixed at 1024 bits.

A. QPSK, Flat Fast-Fading Channel

In Fig. 6, the PER performance of the QPSK-based system is shown. As a reference, the performance of MAP, as well as

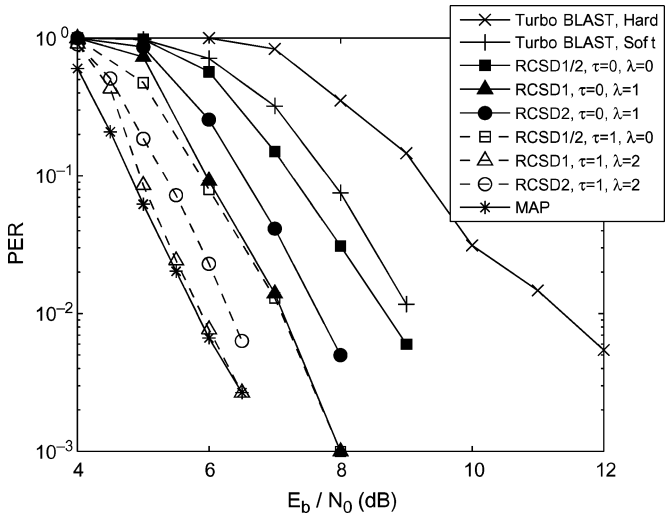

Fig. 6. MMSE-RCSD for QPSK on flat fast fading channels.

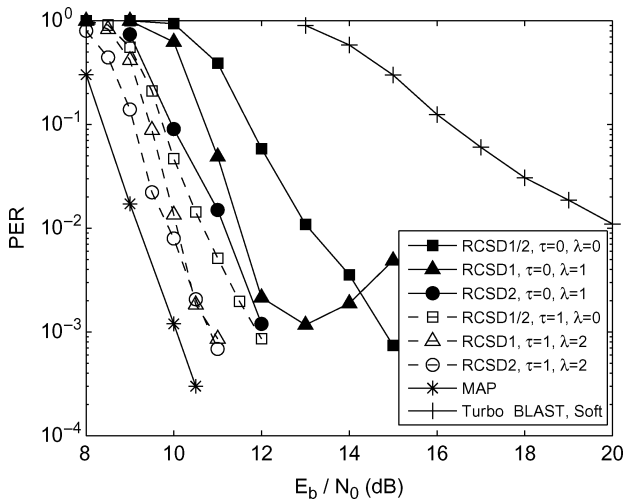

Fig. 7. MMSE-RCSD for 16-QAM on flat fast fading channels.

that of Turbo-BLAST detection with both hard and soft decision feedback, is also shown. When $\tau = \lambda = 0$, RCSD1 and RCSD2 are identical and they are different from BLAST detection only in their use of soft-decision feedback in interference cancellation. It is seen that RCSD with even $\tau = \lambda = 0$ outperforms hard Turbo-BLAST with a 2.5 dB gain at the PER of $10^{(-2)}$, which comes strictly from the performance advantage of SDF over HDF. Note that a hard output decoder (e.g., Viterbi) would be sufficient for HDF, which could result in reduced decoding complexity compared to SDF. For QPSK, RCSD performs somewhat better than soft Turbo-BLAST with about a 0.5 dB performance advantage.

Note that the main computational load of the proposed scheme is proportional to $M^{\tau+1}$, whereas the overall layer depth is $\tau + \lambda$. Given $\tau = 0$, increasing the overall depth by allowing $\lambda = 1$ causes the performances of both RCSD1 and RCSD2 to improve, although RCSD1 seems to benefit more from the increased λ. Note that the performance of RCSD1 with $\lambda = 1$ approaches the performance with $\tau = 1$, $\lambda = 0$, which has four times the trellis complexity of the former.

For the larger trellis associated with $\tau = 1$, when increasing the noncausal depth parameter from $\lambda = 0$ to $\lambda = 2$, a similar performance improvement is observed, with RCSD1 still being somewhat superior to RCSD2. Finally, RCSD1 with $\tau = 1$, $\lambda = 2$ performs asymptotically the same as MAP, but with the trellis complexity reduced by a factor of 16.

B. 16-QAM, Flat Fast-Fading Channel

In Fig. 7, the PER performance of the 16-QAM-based system is shown. We will no longer show the performance of the hard Turbo-BLAST scheme, as its performance is considerably worse than its soft counterpart and including its PER curve would overly clutter the plots. It is seen that RCSD with $\tau = 0$, $\lambda = 0$ has a 7–dB gain over soft Turbo-BLAST detection at the PER of 10^{-2}. Note that the detection portion of soft Turbo-BLAST [24] is a linear operation, which is fundamentally inferior to the nonlinear RCSD performance wise. The latter utilizes both *a priori* information from the decoder and previously detected symbols to successively cancel interference, whereas Turbo-BLAST utilizes *a priori* decoder feedback only once for interference cancellation. It is obvious that for higher-order modulation schemes the symbol estimate purely based on *a priori* information is not sufficiently accurate.

On the other hand, both RCSD schemes with $\tau = 1$ consistently outperform the corresponding scheme with $\tau = 0$ regardless of the value of λ. Furthermore, for a given τ, when increasing the depth parameter λ, the performance in general improves. For a given nonzero λ, RCSD2 outperforms RCSD1 when $\tau = 0$ but two are comparable when $\tau = 1$. In particular, for $\tau = 0$ an error floor appears in RCSD1 at high E_b/N_0, but at the practically meaningful PER level of 10^{-2}, there is still about a 1.5-dB gain compared to $\lambda = 0$. This is consistent with the analysis in Section III-B and implies that for 16-QAM and when the complexity requirement is stringent (e.g., receiver implementation for $\tau > 0$ is not allowed), RCSD2 is preferred to RCSD1.

C. 64-QAM, Frequency-Selective Quasi-Static Channel

In Fig. 8 the PER performance of the 64-QAM-based system is shown. There is a 6-dB gain associated with RCSD ($\tau = 0$, $\lambda = 0$) over soft Turbo-BLAST detection at the 10^{-2} PER. Also, as in the 16-QAM case, the performance plots of RCSD1 with $\lambda > 0$ clearly show error floors at high SNRs. In contrast, for a given τ, RCSD2 with $\lambda > 0$ outperforms RCSD1 and does not exhibit an error floor down to the simulation point. As shown in Section III-B, if the *a priori* information from the outer decoder is utilized for noncausal symbol estimation, as done in RCSD1 with $\lambda > 0$, the exchanged soft information might be degraded at high SNRs, especially for large constellations like 64-QAM. However, down to the practically meaningful PER of 10^{-2}, attempting noncausal estimation still outperforms RCSD with $\lambda = 0$, although the latter does not show any tendency for an error floor.

We realize that $\tau > 0$ is essential to achieve near-MAP performance, requiring a higher complexity level, especially for a larger constellation such as 64-QAM. However, we find that

TABLE I
COMPLEXITY COMPARISON: RCSD VERSUS SOFT SD

	RCSD	soft SD
Candidate/Enumeration Size	16	100
Search Complexity (in avg. no. of visited nodes)	705	1988
Symbol APP Computation (in real multiplications)	2816	5200
Symbol-to-bit Conversion (enumeration size)	16	100

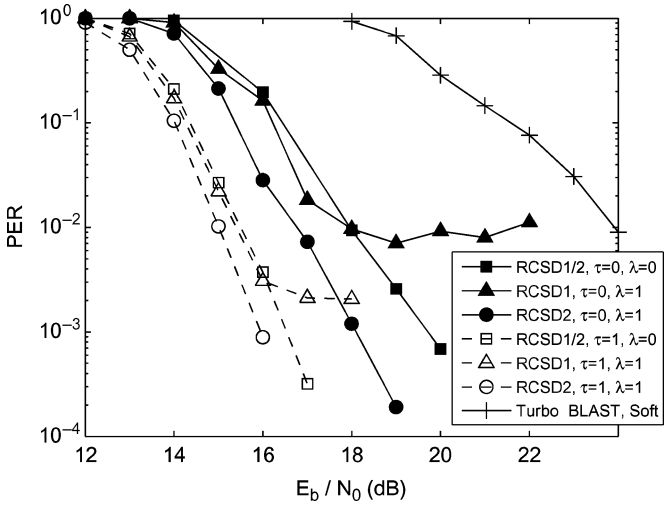

Fig. 8. MMSE-RCSD for 64-QAM on frequency-selective quasi-static fading channels.

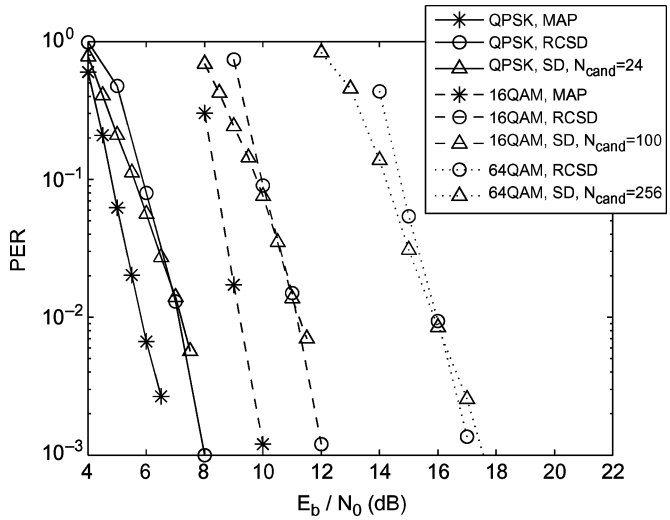

Fig. 9. Comparison of ML, soft SD, and MMSE-RCSD ($\tau = 0$), for QPSK/16-QAM/64-QAM.

RCSD can still do reasonably well with a practical level of complexity. To be more specific, for $\tau = 0$, $\lambda = 1$, RCSD2 has its complexity comparable to the canonical layer-by-layer approach while having a significant gain over BLAST detection, and is only 1.5 dB away from the performance of RCSD with $\tau = 1$, $\lambda = 0$, which requires 64 times higher complexity.

In the remaining portion of the paper, we will compare RCSD with existing schemes based on only $\tau = 0$, which is considered more practical for low-complexity systems.

D. Comparison of Complexity Requirements

In Fig. 9, we show the performance of RCSD with $\tau = 0$, $\lambda = 2$ (either RCSD1 or RCSD2, whichever is better), and that of soft SD with a particular value of $N_{\rm cand}$ that require a similar E_b/N_0 for the given PER of 10^{-2}. The Turbo-BLAST performance is not included, since from previous figures it is clear that RCSD with $\tau = 0$ consistently has better performance. The performance of MAP (when simulation is feasible) is also shown as a reference. We set the search radius of SD sufficiently large to cover $N_{\rm cand}$ points (symbol sequences) with high probability, while keeping the number of candidate symbol sequences in the list to be at most $N_{\rm cand}$. The channels are the same as those used in Figs. 6, 7, and 8, for QPSK, 16-QAM, and 64-QAM, respectively.

We are interested in comparing complexity requirements of RCSD and soft SD. Detailed complexity comparison in general is difficult as the complexity level depends highly on specific implementation strategies and design goals. However, as implied earlier and as is to be shown shortly through a representative test case, the performance and complexity level of RCSD and soft SD depend largely on the enumeration size.

Let us focus on a test case. We choose the 4×4 16QAM case from Fig. 9, at $E_b/N_0 = 11$ dB, where RCSD2 with $\tau = 0$, $\lambda = 2$ and SD with $N_{\rm cand} = 100$ have very similar PERs at about 10^{-2}. To obtain soft bit information, required operations include a search for candidate symbol sequences, symbol a posteriori probability (APP) computation and symbol-to-bit APP conversion. For convenience, we assume no extra memory is available to SD, so that its search procedure and symbol APP computation need to be conducted for each iteration. Therefore, we focus on per iteration complexity level.

1) Search: We use the total number of nodes (in the generic tree representation of interfering signals) visited by the end of search as the complexity measure of the search process. Again, we emphasize that, although the search for noncausal symbols in RCSD2 is conducted for each candidate symbol sequence, it only needs to reach the closest-point without maintaining a candidate list as SD, so that eventually a much smaller number of nodes are visited. We record the average number of visited nodes through simulations, since it is difficult to find this number analytically. The result is summarized in Table I, which shows RCSD has much smaller search complexity.

2) Symbol APP Computation: We use the number of real multiplications as the measure of complexity in APP computation, since it is usually the most demanding requirement in implementation.

According to the signal model (1), for each candidate symbol sequence soft SD needs to compute the log-likelihood function $\sum_{i=1}^{4} |z_i - \sum_{j=1}^{i} H_{ij} \cdot s_j|^2/\sigma^2$, which requires $\sum_{i=1}^{4}(4n_i+2+1)$ real multiplications for the operations of $(H \cdot s)$, $(|\cdot|^2)$, and $(\cdot/\sigma^2)$, respectively, assuming the value of σ^2 is known. The

variable $n_i = i$ denotes the number of involved symbols at the ith layer. There are altogether 52 real multiplications needed for each candidate in the list.

For RCSD, the symbol APP computation of a particular layer is based on 16 candidate symbols since we are focusing on $\tau = 0$, and for each possible symbol computation involves up to $(\tau + \lambda + 1 = 3)$ layers, along with the soft symbol cancellation.

- Soft Symbol Estimation: For the first three layers, after the APP computation of each layer, a soft symbol estimate is obtained and will be used in subsequent computations, as described right after (7). For example, to obtain $\bar{s}_1$ and σ_1^2, it requires $16 \cdot 2$ and $16 \cdot 3$ real multiplications, respectively. The same complexity is required to obtain $(\bar{s}_2, \sigma_2^2)$ and $(\bar{s}_3, \sigma_3^2)$. Therefore, the total number of real multiplications needed for all layers is $3 \cdot 16 \cdot (2 + 3) = 240$.
- APP Computation: Based on (6) $\sim$ (9), the APP for each candidate symbol can be computed. For example, consider computation of the APP for a symbol in layer-2. First, using (6) and (7), the log-likelihood function component associated with the current-layer observation requires $(4 \cdot 2 + 2 + 4)$ real multiplications, to perform the $(H \cdot s)$, $(|\cdot|^2)$ and $(\cdot/\sigma_T^2)$ operations. Secondly, $(4 \cdot 3 + 2 + 4)$ and $(4 \cdot 4 + 2 + 4)$ real multiplications are needed according to (9) for the log-likelihood function accumulations due to the two noncausal observations. Note that in (7) the effort of computing σ_1^2 for the detected layer-1 has been accounted for in the soft symbol estimation part, while in (9) the computations associated with noncausal σ_2^2 and σ_3^2 are skipped since these variances are zeros for RCSD2, as explained at the end of Section II-B-2. So 54 real multiplications are needed for each layer-2 symbol APP computation. Repeating this, we see that 33, 46, and 28 real multiplications are needed for a symbol APP computation in layer-1, layer-3, and layer-4, respectively. Thus, the number of real multiplications required for all candidate symbols is $16 \cdot (33 + 54 + 46 + 28) = 2576$.

The total number of real multiplications required for all symbol APP computations is $(2576 + 240 = 2816)$, as summarized in Table I.

3) Symbol-to-Bit Conversion: The computational load associated with symbol-to-bit conversion (in terms of the APP) can be examined by going through all candidates in the enumeration list, possibly with max-log approximation. Regardless of the particular conversion method used, the arithmetic operations associated with each candidate is the same. Therefore, we only make the comparison based on the candidate-list size, which is also given in Table I.

This test case shows definite complexity advantage of RCSD over SD. In general, however, complexity comparison at the quantitative level is tricky and depends on the implementation strategies employed. For example, if the particular implementation strategy allows a generous use of memory, then the candidate list of SD and its symbol APP of each candidate member can be stored at the first iteration, and no symbol estimation needs to be done for iterations afterwards [25], in which case the computational burden of SD is substantially reduced. On the other hand, it should be noted that the die area of the chip in very large scale integration (VLSI) depends heavily on the use of buffer memory, and the memory use is generally minimized in VLSI designs targeting at volume production. Therefore, our suggestion of required complexity levels is meant to be largely qualitative and we leave specific quantitative interpretation of the required complexity of RCSD and SD to the implementation engineers. Nevertheless, we feel that it is safe to say that the proposed RCSD schemes are advantageous over SD in certain modulation/channel scenarios and implementation environments that are of practical interest.

V. Conclusion

We presented a suboptimal, RCSD scheme well suited to MIMO systems with layered space-time (or space-frequency) architecture. The proposed algorithm is based on trellis representation of the MIMO signals and a subsequent formulation of constrained-depth MAP detection, as well as the use of soft decision feedback (SDF). Decision feedback has been applied based on decomposition of the MIMO interference in causal and non-causal parts in the spatial sense. Two different variations have been suggested for the processing of non-causal interference. The MSE of each layer is computed to analyze the error propagation effect of the RCSD scheme. The mutual information between the bit-wise LLR and corresponding coded bits are compared for the two RCSD variants. Error rate simulations were conducted based on the IDD procedure on different modulation constellations, and the resulting performance validated the analysis. Compared with MAP detection, soft SD and Turbo-BLAST, the proposed RCSD scheme is shown to exhibit favorable performance and complexity tradeoffs in certain practical scenarios.

References

[1] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

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

[3] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[4] A. L. C. Hui and K. B. Letaief, "Successive interference cancellation for multiuser asynchronous DS/CDMA detectors in multipath fading links," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 384–391, Mar. 1998.

[5] R. Fischer, *Precoding and Signal Shaping for Digital Transmission*. Erlangen, Germany: Habilitationsschrift, Universität Erlangen-Nürnberg, 2001.

[6] D. Wübben, R. Böhnke, J. Rinas, V. Kühn, and K. D. Kammeyer, "Efficient algorithm for decoding layered space-time codes," *Inst. Elect. Eng. Elect. Lett.*, vol. 37, no. 22, pp. 1348–1350, Oct. 2001.

[7] A. Benjebbour, H. Murata, and S. Yoshida, "Comparison of ordered successive receivers for space-time transmission," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, Fall, 2001, vol. 4, pp. 2053–2057.

[8] W. Zha and S. Blostein, "Modified decorrelating decision-feedback detection of BLAST space-time system," in *IEEE Int. Conf. Commun. (ICC)*, May 2002, vol. 1, pp. 335–339.

[9] J. Benesty, Y. Huang, and J. Chen, "A fast recursive algorithm for optimum sequential signal detection in a BLAST system," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1722–1730, Jul. 2003.

[10] B. Hassibi, "A fast square-root implementation for BLAST," in *Conf. Rec. 34th Asilomar Conf. Signal Syst. Comput.*, 2000, pp. 1255–1259.

[11] Y. Li and Z. Luo, "Parallel detection for V-BLAST system," in *IEEE Int. Conf. Commun. (ICC)*, May 2002, vol. 1, pp. 340–344.

[12] O. Damen, K. Abed-Meraim, and S. Burykh, "Iterative QR detection for BLAST," *Wireless Pers. Commun.*, vol. 19, no. 3, pp. 179–192, Dec. 2001.

[13] W. Choi, R. Negi, and J. Cioffi, "Combined ML and DFE decoding for the V-BLAST system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2000, vol. 3, pp. 1243–1248.
[14] X. Li, H. C. Huang, A. Lozano, and G. J. Foschini, "Reduced-complexity detection algorithms for systems using multi-element arrays," in *IEEE GLOBECOM'00*, Nov. 2000, no. 1, pp. 1072–1076.
[15] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comput.*, vol. 44, pp. 463–471, Apr. 1985.
[16] O. Damen, A. Chkeif, and J. Belfiore, "Lattice code decoder for space-time codes," *IEEE Commun. Lett.*, vol. 4, no. 5, pp. 161–163, May 2000.
[17] B. Hassibi and H. Vikalo, "On the expected complexity of sphere decoding," in *Proc. 35th Asilomar Conf. Signals, Syst., Comp.*, Pacific Grove, CA, Nov. 2001, pp. 1051–1055.
[18] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.
[19] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1804–1824, Jul. 2002.
[20] A. Chan and I. Lee, "A new reduced-complexity sphere decoder for multiple antenna systems," in *IEEE Int. Conf. Commun. (ICC)*, New York, Apr. 2002, vol. 1, pp. 460–464.
[21] H. Yao and G. Wornell, "Lattice-feduction-aided detectors for MIMO communication systems," in *Proc. IEEE GLOBECOM*, Nov. 2002, pp. 424–428.
[22] C. Windpassinger and R. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proc. IEEE Info. Theory Workshop*, Apr. 2003, pp. 345–348.
[23] S. T. Brink, J. Speidel, and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc. IEEE GLOBECOM*, Sydney, Australia, 1998, pp. 579–584.
[24] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: Theory and experiments," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2538–2546, Oct. 2002.
[25] B. M. Hochwald and S. T. Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.
[26] R. Wang and G. B. Giannakis, "Approaching MIMO channel capacity with reduced-complexity soft sphere decoding," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*, Mar. 2004, vol. 3, pp. 1620–1625.
[27] Y. L. C. de Jong and T. J. Willink, "Iterative tree search detection for MIMO wireless systems," in *IEEE Veh. Technol. Conf. (VTC)—Fall*, Sep. 2002, vol. 2, pp. 1041–1045.
[28] B. Steingrimsson, Z. Q. Luo, and K. M. Wang, "Soft quasi-maximum likelihood detection for multiple-antenna channels," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2710–2719, Nov. 2003.
[29] D. Pham, K. R. Pattipati, P. K. Willett, and J. Luo, "A generalized probabilistic data association detector for multiple antenna systems," *IEEE Commun. Lett.*, vol. 8, no. 4, pp. 205–207, April 2004.
[30] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. Commun.*, vol. 37, no. 4, pp. 428–436, May 1989.
[31] M. V. Eyuboglu and S. U. H. Qureshi, "Reduced-state sequence estimation for coded modulation on intersymbol interference chanels," *IEEE J. Sel. Areas Commun.*, vol. 7, pp. 989–995, Aug. 1989.
[32] R. E. Kamel and Y. Bar-Ness, "Reduced-complexity sequence estimation using state partitioning," *IEEE Trans. Commun.*, vol. 44, no. 9, pp. 1057–1063, Sep. 1996.
[33] R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 354–364, Feb.–Mar.–Apr. 1995.
[34] X. Chen and K. M. Chugg, "Reduced-state soft-input/soft-output algorithms for complexity reduction in iterative and non-iterative data detection," in *IEEE Int. Conf. Commun. (ICC)*, Jun. 2000, pp. 6–10.
[35] G. Colavolpe, G. Ferrari, and R. Raheli, "Reduced-state BCJR-type algorithms," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 848–859, May 2001.
[36] K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proc. IEEE*, vol. 58, no. 50, pp. 779–785, May 1970.
[37] J. Moon and F. R. Rad, "Turbo equalization via constrained-delay APP estimation with decision feedback," *IEEE Trans. Commun.*, vol. 53, no. 12, pp. 2102–2113, Dec. 2005.
[38] W. Choi, K. Cheong, and J. M. Cioffi, "Iterative soft interference cancellation for multiple antenna systems," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*, Sep. 2000, vol. 1, pp. 304–309.
[39] M. Tuchler and J. Hagenauer, "EXIT charts of irregular codes," presented at the 2002 Conf. Inf. Sci. Syst., Princeton Univ., Princeton, NJ, Mar. 2002.
[40] S. T. Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.
[41] A. Tomasoni *et al.*, "A low complexity turbo MMSE receiver for W-LAN MIMO systems," *IEEE Int. Conf. Commun. (ICC)*, vol. 9, pp. 4119–4124, Jun. 2006.
[42] D. N. Liu *et al.*, "Low complexity affine MMSE detector for iterative detection-decoding MIMO OFDM system," in *IEEE Int. Conf. Commun. (ICC)*, Jun. 2006, vol. 10, pp. 4654–4659.
[43] *LAN/MAN Specific Requirements 3/4 Part 11: Wireless MAC and PHY Specifications: High Speed Physical Layer in the 5 GHz Band*, IEEE Std. 802.11a, May 1999.

Yong Li received the B.S. degree from Tsinghua University, Beijing, P.R. China, in 1997, and the M.Eng. degree from Nanyang Technological University, Singapore, in 1999, and Ph.D. degree from the University of Minnesota, Minneapolis, in 2005, all in electrical engineering.

Since 2005, he has been with Qualcomm Inc., San Diego, CA, as a senior engineer to develop the system design for the next generation wireless communications. His concentration area is in communications/DSP algorithms, including iterative coding/decoding and MIMO receiver design.

Jaekyun Moon (S'89–M'90–SM'97–F'07) received the B.S.E.E. degree from the State University of New York at Stony Brook and the M.S. and Ph.D. degrees from the Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, PA, in 1987 and 1990, respectively.

Since 1990, he has been on the faculty of the Department of Electrical and Computer Engineering, University of Minnesota, Twin Cities. His research interests are in the area of channel characterization, signal processing and coding for data storage, and digital communication.

Dr. Moon received the 1994-1996 McKnight Land-Grant Professorship from the University of Minnesota. He also received the IBM Faculty Development Awards, as well as the IBM Partnership Awards. He was awarded the National Storage Industry Consortium (NSIC) Technical Achievement Award for the invention of the maximum transition run (MTR) code, a widely used error-control/modulation code in commercial storage systems. He served as Program Chair for the 1997 IEEE Magnetic Recording Conference. He is also Past Chair of the Signal Processing for Storage Technical Committee of the IEEE Communications Society. In 2001, he cofounded Bermai, Inc., a fabless semiconductor start-up, and served as founding President and CTO. He served as a guest Editor for the 2001 IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS issue on Signal Processing for High Density Recording. He also served as an Editor for the IEEE TRANSACTIONS ON MAGNETICS in the area of signal processing and coding for 2001–2006.