Multiple-Symbol Differential Detection for Single-Antenna and Multiple-Antenna Systems over Ricean-fading Channels

Tao Cui
Department of Electrical Engineering
California Institute of Technology
Pasadena, CA, USA 91125
Email: taocui@caltech.edu

Chintha Tellambura
Department of Electrical and Computer Engineering
University of Alberta
Edmonton, AB, Canada T6G 2V4
Email: chintha@ece.ualberta.ca

Abstract— **This paper considers multiple symbol differential detection (DD) for both single-antenna and multiple-antenna systems over flat Ricean-fading channels. We derive the optimal multiple symbol detection (MSD) decision rules for both M-ary differential phase-shift keying (MDPSK) and differential unitary space-time modulation (DUSTM). The sphere decoder (SD) is adopted to solve the MSD for MDPSK. As well, an improved SD is proposed by using the Schnorr-Euchner strategy. A suboptimal MSD based decision feedback DD algorithm is proposed for the MSD of DUSTM. We also develop a sphere decoding bound intersection detector (SD-BID) to optimally solve the MSD problem for DUSTM, which still maintains low complexity. Simulation results show that our proposed MSD algorithms for both single-antenna and multiple-antenna systems reduce the error floor of conventional DD but with reasonably low computational complexity.**

I. Introduction

Digital receivers using differential detection (DD) are attractive for flat fading channels, because such receivers do not require channel state information (CSI) and are robust against the carrier phase ambiguity. However, it is well-known that the conventional differential detection (CDD) has an irreducible error floor in flat fading, time selective channels and is 3 dB worse than its coherent counterpart. In single-antenna systems, multiple-symbol detection (MSD) of M-ary differential phase-shift keying (MDPSK) has been proposed in [1], where $N+1$ consecutive received samples are jointly processed to detect N data symbols. MSD reduces the error floor, and when N goes to infinity, the performance of MSD converges to that of coherent detection (CD). But the complexity of MSD is usually high, exponential in $N-1$, which prevents it from practical use. Recently, the sphere decoder (SD) [2] has been applied to reduce the complexity of MSD [3]. On the other hand, decision feedback differential detection (DF-DD) [4] offers reasonable performance while still maintaining low complexity.

DD for singe-antenna systems has recently been generalized to multiple-antenna systems. Hochwald and Sweldens [5] have developed a general framework for differential unitary space-time modulation (DUSTM) via finite group theory [5]. They exist for any number of antennas. Constellation design, search method and performance are treated in detail in [5]. DUSTM performs poorly unless the fading rate is low. Naturally, attempts have been made to extend MSD to DUSTM. In [6], noncoherent receivers for DUSTM based on MSD and DF-DD are derived. We have recently derived, for MSD of DUSTM over quasi-static fading channels, an efficient MSD bound intersection detector (BID) in [7], [8]. Our BID is optimal and can be more efficient than [6] in high signal-to-noise ratio (SNR) regimes.

In certain radio propagation environments, the channel can be described by a Ricean distribution with a Rice factor K. There are only a handful of papers dealing with DD receivers for Ricean channels [9]. For single-antenna systems, a DF-DD scheme for flat Ricean-fading channels based on linear prediction is proposed in [9]. A MSD-based DF-DD decision rule for Ricean fading is also given in [9]. Besides this work, no other paper treats DD in single-antenna Ricean channels. Furthermore, no other paper treats DD and MSD schemes for multiple-antenna systems over Ricean channels.

In this paper, we first investigate the optimal and efficient MSD of MDPSK. A general MSD decision rule is derived for flat Ricean-fading channels. The decision rule reduces to the one in [10] for Rayleigh fading channels when $K=0$ and the one in [1] for AWGN channels when $K\to\infty$. We consider using the SD to solve the MSD problem, which has an integer quadratic form. To further reduce the SD complexity, the Schnorr-Euchner search strategy [11] is extended to PSK constellations. We then generalize the optimal decision rule to multiple-antenna Ricean channels. A quasi-static fading channel is assumed. To the best of our knowledge, the optimal MSD decision rule for DUSTM transmitted over Ricean channels and an efficient detector have not been derived in the open literature. However, the MSD complexity of DUSTM grows exponentially with L^N, where L is the DUSTM constellation size. In order to reduce this detection complexity, we propose a suboptimal MSD-based DF-DD using our BID [7]. Although the proposed DF-DD scheme does not achieve ML performance, it performs substantially better than CDD with complexity only linear in N. Furthermore, we combine the branch and bound (BnB) principle and BID, and give a sphere decoding bound intersection detector (SD-BID), which offers ML performance. Surprisingly, in high SNR, the complexity of SD-BID is even lower than that of the DF-DD scheme.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

II. Multiple symbol differential detection in single-antenna systems

A. Single-antenna system model

Let us consider a communication system employing M-ary DPSK. $l = \log_2(M)$ binary bits are Gray mapped to data-carrying symbols $v[n]$ from an M-ary PSK constellation $\mathcal{V} = \{e^{\jmath 2\pi m/M} | m = 0, 1, \ldots, M-1\}$, where n is the symbol discrete-time index. $v[n]$ is then differentially encoded, and the corresponding channel input symbols $s[n]$ are obtained as

$$s[n] = v[n]s[n-1], \, n \in \mathcal{Z}. \quad (1)$$

We assume that the channel does not change significantly during one symbol interval T, and transmitter and receiver filters with square-root Nyquist characteristics [4]. Therefore, the channel is frequency-nonselective (flat), and the received signal $r[n]$ can be written as

$$r[n] = h[n]s[n] + w[n], \, n \in \mathcal{Z} \quad (2)$$

where $h[n]$ is the Ricean fading gain and $w[n] \sim \mathcal{CN}(0, \sigma_n^2)$ denotes additive white Gaussian noise (AWGN). $h[n]$ is a complex Gaussian random process, and can be expressed as

$$h[n] = h_d[n] + h_s[n] \quad (3)$$

where $h_d[n]$ is the direct component $h_d[n] = E\{h[n]\} = e^{\jmath 2\pi f_D T n} h_m$ (h_m is the magnitude of the fading process, and f_D is the Doppler frequency due to users' mobility). The scattered component $h_s[n]$ is a zero mean Gaussian process with autocovariance function $\varphi_h[k] = E\{(h[n+k] - h_d[n+k])(h[n] - h_d[n])^*\} = \sigma_h^2 J_0(2\pi f_D T k)$ ($J_0(\cdot)$ denotes the zeroth order Bessel function of the first kind, and $\sigma_h^2 = E\{|h_s[n]|^2\}$) according to the Jakes' model. Note, however, that the MSD decision metric derived in the following is not limited to the Jakes' model. The Rice factor K is defined as

$$K = \frac{|h_m|^2}{\sigma_h^2}. \quad (4)$$

B. Multiple-symbol differential detection for MDPSK

To reduce the error floor in time selective channels and bridge the gap between CDD and CD, MSD jointly detects the $N-1$ differentially encoded symbols given N consecutively received symbols. We consider the received symbols from $n = k+1$ to $n = k+N$. Without loss generality, we set $k = 0$ and omit k in the following. The input output relationship (2) can be written in vector form as

$$\mathbf{r} = \mathbf{S}_D \mathbf{h} + \mathbf{w} \quad (5)$$

where $\mathbf{r} = [r[1], \ldots, r[N]]^T$, $\mathbf{w} = [w[1], \ldots, w[N]]^T$, $\mathbf{h} = [h[1], \ldots, h[N]]^T$, and $\mathbf{S}_D$ is a diagonal matrix

$$\mathbf{S}_D = \begin{bmatrix} s[1] & & & \\ & s[2] & & \\ & & \ddots & \\ & & & s[N] \end{bmatrix}. \quad (6)$$

Using (3), (5) can be rewritten as

$$\mathbf{r} = \mathbf{S}_D(\mathbf{h}_s + \mathbf{h}_d) + \mathbf{w} \quad (7)$$

where $\mathbf{h}_s = [h_s[1], \ldots, h_s[N]]^T$, and $\mathbf{h}_d = [h_d[1], \ldots, h_d[N]]^T$. Since both $\mathbf{h}_s$ and $\mathbf{w}$ are complex Gaussian, $\mathbf{S}_D\mathbf{h}_s + \mathbf{w}$ is also complex Gaussian. Therefore, $\mathbf{r}$ is a Gaussian vector, and the conditional probability density function (pdf) given $\mathbf{S}_D$ is

$$f(\mathbf{r}|\mathbf{S}_D) = \frac{1}{\pi^N \det \mathbf{C}_r} \exp\left\{-(\mathbf{r} - \bar{\mathbf{r}})^H \mathbf{C}_r^{-1} (\mathbf{r} - \bar{\mathbf{r}})\right\} \quad (8)$$

where $\bar{\mathbf{r}} = \mathbf{S}_D \mathbf{h}_d$, and $\mathbf{C}_r$ is the covariance matrix of $\mathbf{r}$ and is given by

$$\mathbf{C}_r = \mathbf{S}_D \mathbf{C}_h \mathbf{S}_D^H + \sigma_n^2 \mathbf{I}_N. \quad (9)$$

The $\mathbf{C}_h$ in (9) denotes the covariance matrix of $\mathbf{h}$ and can be represented as

$$\mathbf{C}_h = \begin{bmatrix} \varphi_h[0] & \varphi_h[1] & \cdots & \varphi_h[N-1] \\ \varphi_h[-1] & \varphi_h[0] & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_h[-N+1] & \cdots & \cdots & \varphi_h[0] \end{bmatrix}. \quad (10)$$

Since $s[n]$ are MPSK symbols, $\mathbf{S}_D \mathbf{S}_D^H = \mathbf{I}_N$. Hence, $\det \mathbf{C}_r = \det(\mathbf{S}_D \mathbf{C}_h \mathbf{S}_D^H + \sigma_n^2 \mathbf{I}_N) = \det(\mathbf{C}_h + \sigma_n^2 \mathbf{I}_N)\mathbf{S}_D^H \mathbf{S}_D = \det(\mathbf{C}_h + \sigma_n^2 \mathbf{I}_N)$ is independent of $\mathbf{S}_D$. Therefore, the ML MSD decision rule, from maximizing the pdf (8), is equivalent to minimizing

$$g(\mathbf{S}_D) = (\mathbf{S}_D^H \mathbf{r} - \mathbf{h}_d)^H (\mathbf{C}_h + \sigma_n^2 \mathbf{I}_N)^{-1} (\mathbf{S}_D^H \mathbf{r} - \mathbf{h}_d). \quad (11)$$

Note that $\mathbf{S}_D^H$ is a diagonal matrix. Since the multiplication between a diagonal matrix and a vector is commutative, we can rewrite (11) as

$$g(\mathbf{s}) = (\mathbf{R}_D \mathbf{s}^* - \mathbf{h}_d)^H (\mathbf{C}_h + \sigma_n^2 \mathbf{I}_N)^{-1} (\mathbf{R}_D \mathbf{s}^* - \mathbf{h}_d) \quad (12)$$

where $\mathbf{R}_D$ is a diagonal matrix with diagonal elements from vector $\mathbf{r}$, and $\mathbf{s} = [s[1], s[2], \ldots, s[N]]^T$. The matrix $(\mathbf{C}_h + \sigma_n^2 \mathbf{I}_N)^{-1}$ can be Cholesky factorized as $(\mathbf{C}_h + \sigma_n^2 \mathbf{I}_N)^{-1} = \mathbf{U}^H \mathbf{U}$, where $\mathbf{U}$ is upper triangular. Let $\mathbf{G} = (\mathbf{U}\mathbf{R}_D)^*$ (also upper triangular) and $\mathbf{y} = (\mathbf{U}\mathbf{h}_d)^*$. The MSD decision rule for Ricean channels can be obtained as

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s} \in \mathcal{V}^N} \|\mathbf{y} - \mathbf{G}\mathbf{s}\|^2. \quad (13)$$

If $\hat{\mathbf{s}}$ has been estimated from (13), the transmitted signals can be differentially detected as

$$v[k] = \hat{s}[k]\hat{s}^*[k-1]. \quad (14)$$

Remarks:

- When $K = 0 \Rightarrow \mathbf{h}_d = \mathbf{0}$, (13) reduces to the decision metric in [3], which corresponds to Rayleigh-fading channels. When $K \to \infty$, $\sigma_h^2 \to 0$, and (13) reduces to

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s} \in \mathcal{V}^N} \|\mathbf{r} - \mathbf{S}_D \mathbf{h}\|^2. \quad (15)$$

Eq. (15) corresponds to coherent detection with perfect CSI.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

C. Sphere-decoder based MSD

The Fincke and Phost (FP) [2] is well-known as the SD in communication theory.

Basically, the SD examines the candidate vectors $\mathbf{s}$ that lie within a hypersphere of radius R:

$$\|\mathbf{y}-\mathbf{G}\mathbf{s}\|^2 \leq R^2. \quad (16)$$

Suppose that the initial radius R is large enough so that the hypersphere (16) contains the ML solution. Let the entries of $\mathbf{G}$ be denoted by $g_{i,j}$, $i \leq j$. The diagonal terms of $\mathbf{G}$ are non-zero ($g_{i,i} \neq 0$). Since $\mathbf{G}$ is upper triangular, (16) can be written as

$$\sum_{i=1}^{N}\left|y_i-\sum_{j=i}^{N} g_{i,j}s_j\right|^2 \leq R^2. \quad (17)$$

Note that each term in (17) is nonnegative. A necessary condition for $\mathbf{s}$ to lie inside the hypersphere is

$$|y_N - g_{N,N}\hat{s}_N|^2 \quad \leq R^2 \quad (18)$$

$$\vdots \qquad \vdots$$

$$\sum_{i=1}^{N}\left|y_i-\sum_{j=i}^{N} g_{i,j}\hat{s}_j\right|^2 \quad \leq R^2. \quad (19)$$

Eqs. (18)-(19) can be checked one by one. The candidate set for s_N can be obtained as

$$\mathcal{I}_N = \{s|\, |y_N - g_{N,N}s|^2 \leq R^2,\, s \in \mathcal{V}\}. \quad (20)$$

After $\hat{s}_{k+1}$ has been chosen, we define

$$d_{k+1}^2 = \left|y_{k+1}-\sum_{j=k+1}^{N} g_{k+1,j}\hat{s}_j\right|^2 \quad (21)$$

$$R_N^2 = R^2,\, R_k^2 = R_{k+1}^2 - d_{k+1}^2. \quad (22)$$

We can get the candidate set for s_k as

$$\mathcal{I}_k = \left\{ s \left\| y_k - g_{k,k}s - \sum_{j=k+1}^{N} g_{k,j}\hat{s}_j \right|^2 \leq R_k^2,\, s\in\mathcal{V} \right\}. \quad (23)$$

When a valid candidate vector $\hat{\mathbf{s}}$ is found, all the R_k's are updated according to

$$R_N^2 = \|\mathbf{y}-\mathbf{G}\hat{\mathbf{s}}\|^2,\, R_k^2 = R_{k+1}^2 - d_{k+1}^2,\, k = N-1,\ldots,1. \quad (24)$$

This updating results in a smaller hypersphere with $\hat{\mathbf{s}}$ on its surface. The same process continues until all the candidate points within the hypersphere have been checked. The vector with minimum $g(\hat{\mathbf{s}})$ in (12) is output as the ML solution.

The initial radius R should be chosen according to the statistic of $g(\mathbf{s})$ in (11)

$$g(\mathbf{s}) = (\mathbf{S}_D^H\mathbf{r}-\mathbf{h}_d)^H(\mathbf{C}_h+\sigma_n^2\mathbf{I}_N)^{-1}(\mathbf{S}_D^H\mathbf{r}-\mathbf{h}_d). \quad (25)$$

If $\mathbf{s}$ is the true solution, using (5), $\mathbf{x} = \mathbf{S}_D^H\mathbf{r}-\mathbf{h}_d = \mathbf{h}_s+\mathbf{S}_D^H\mathbf{w}$ is a zero mean complex Gaussian vector with autocovariance matrix $\mathbf{C}_x = \mathbf{C}_h+\sigma_n^2\mathbf{I}_N$. Therefore, $e = \mathbf{x}^H(\mathbf{C}_h+\sigma_n^2\mathbf{I}_N)^{-1}\mathbf{x}$ is a chi-square random variable with $2N$ degrees of freedom. We can choose R^2 to make the probability that e is less than R^2 very high:

$$\int_0^{R^2} \frac{x^{N-1}e^{-x/2}}{\Gamma(N)2^N}\mathrm{d}x = 1-\epsilon, \quad (26)$$

where ϵ is set to a value close to 0 (e.g., $\epsilon = 0.1$), and Γ is the gamma function. If no signal point was found within the hypersphere, we increase the probability $1-\epsilon$ (e.g., $\epsilon = 0.1^2, 0.1^3, \ldots$) until the ML solution is found.

The SD complexity is dependent on the initial radius. To reduce such dependence, Schnorr and Euchner (SE) [11] suggested an important improvement of the SD; the main idea is that the algorithm should first examine the signal points nearest to the center of the hypersphere. While a modified SE principle has been applied by Lampe *et al.* [3] in a zigzag fashion, they do not sort the candidate set $\mathcal{I}_k$ according to d_k^2. Note that d_k^2 (21) can be rewritten as

$$d_k^2 = \quad A_k - B_k\cos\left(\tfrac{2\pi m_k}{M}-\phi_k\right) \quad (27)$$

$$A_k = \quad \left|y_k - \textstyle\sum_{j=k+1}^{N} g_{k,j}\hat{s}_j\right|^2 + g_{k,k}^2 \quad (28)$$

$$B_k = \quad 2\left|g_{k,k}\left(y_k - \textstyle\sum_{j=k+1}^{N} g_{k,j}\hat{s}_j\right)^*\right| \quad (29)$$

$$\phi_k = \quad \arg\left[g_{k,k}\left(y_k - \textstyle\sum_{j=k+1}^{N} g_{k,j}\hat{s}_j\right)\right] \quad (30)$$

Now (23) can be written as, $m_k = 0,1,\ldots,M-1$,

$$\mathcal{I}_k = \left\{ e^{\jmath\frac{2\pi}{M}m_k} \left| A_k - B_k\cos\left(\frac{2\pi m_k}{M}-\phi_k\right) \leq R_k^2 \right.\right\}. \quad (31)$$

For each $m_k \in \{0,1,\ldots,M-1\}$, we compute d_k^2 and test whether it is less than R_k^2. If it is less than R_k^2, m_k is stored in $\mathcal{I}_k$, and d_k^2 is stored in $\mathcal{D}_k$. Finally, $\mathcal{I}_k$ is sorted according to d_k^2. Note that each test in (27) only needs 4 flops but in [3] it needs 7 flops. The sorting and R_k^2 updating increase complexity savings.

D. MSD-based DF-DD

To further reduce the complexity, the MSD (13) can be easily modified to MSD based DF-DD. Assuming correct decisions on $\hat{\mathbf{s}}^{N-1} = [\hat{s}[1], \hat{s}[2], \ldots, \hat{s}[N-1]]^T$, the MSD based DF-DD can be obtained as

$$\hat{s}[N] = \arg\min_{s[N]\in\mathcal{V}} \|\mathbf{y}-\mathbf{G}(:,1:N-1)\hat{\mathbf{s}}^{N-1}-\mathbf{G}(:,N)s[N]\|^2. \quad (32)$$

The V-BLAST detection algorithm [12] for multiple-antenna systems can also be used to solve (13). It is interesting to compare SD, DF-DD and V-BLAST. In V-BLAST, the estimation of $s[i]$ is based on the previous decisions $\hat{s}[j]$, $i+1 \leq j \leq N$, which can be explained as DF-DD with observation size $N-i+1$. Therefore, V-BLAST involves N DF-DD with window size $1,2,\ldots,N$. The average complexity of V-BLAST is only about $1/2$ of that with DF-DD but it results in a large error floor. The SD can generalize V-BLAST.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

In each step, it does not make hard decision. In high SNR, the SD complexity approximates that of V-BLAST and is also less than that of DF-DD. Hence, SD outperforms V-BLAST and DF-DD in both performance and complexity in high SNR.

III. Multiple symbol differential detection in multiple-antenna systems

A. Multiple-antenna system model

We consider a multiple-antenna system with N_t transmit and N_r receive antennas. Each block of the transmitted symbols occupies T time slots with interval T_s, resulting the block interval $T_B = T_s T$. The transmitted symbols during the nth block is denoted by the $T \times N_t$ matrix $\mathbf{S}[n] = [s_{t,i}[n]]$, $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, N_t$, where $s_{t,i}[n]$ is transmitted from the ith antenna in the $t + (n-1)T$ time slot.

We consider a flat Ricean-fading multiple-antenna channel from a rich scattering environment. The complex base-band received signal at the jth receive antenna, $j = 1, 2, \ldots, N_r$, at time slot t in the nth block can be written as

$$r_{t,j}[n] = \sum_{i=1}^{N_t} h_{i,j}[n] s_{t,i}[n] + w_{t,j}[n], \tag{33}$$

where $h_{i,j}[n]$ denotes the channel gain from the ith transmit antenna to the jth receive antenna, and $w_{t,j}[n]$ is the complex additive white Gaussian noise at the jth receive antenna. The additive Gaussian noise at different receive antennas are independent and have equal variance σ_n^2. Similar to single-antenna channels, $h_{i,j}[n]$ is a complex Gaussian random process and can also be expressed as the summation of the direct component $(h_d)_{i,j}[n]$ and the scattered component $(h_s)_{i,j}[n]$

$$h_{i,j}[n] = (h_d)_{i,j}[n] + (h_s)_{i,j}[n]. \tag{34}$$

Assuming the Rice factor K is common to all paths, K is defined as $K = |(h_d)_{i,j}|^2 / E\{|(h_s)_{i,j}|^2\}$ [13]. We assume that all path gains are statistically independent ($E\{h_{i,j}[n] h^*_{i',j'}[n]\} = 0$) and have the same autocorrelation function $\varphi_h(\tau)$. We assume that the fading channel is quasi-static (QS), i.e., channel variations within each block are negligible, whereas the channel changes from block to block[1]. Therefore $h_{i,j}[n]$ has correlation $\varphi_h[k] = \varphi_h(kT_B)$. Typically, when Jakes' model is used, $\varphi_h[k]$ is given by

$$\varphi_h[k] = E\{(h_s)_{i,j}[n](h_s)^*_{i,j}[n+k]\} = \sigma_h^2 J_0(2\pi k f_d T_B), \tag{35}$$

where σ_h^2 denotes the variance of the fading process, and f_D is the Doppler spread due to users' mobility. The QS condition is met when $f_D T_B < 0.03$ [6]. The matrix form of (33) is

$$\begin{aligned} \mathbf{R}[n] =& \mathbf{S}[n]\mathbf{H}[n] + \mathbf{W}[n] \\ =& \mathbf{S}[n](\mathbf{H}_d[n] + \mathbf{H}_s[n]) + \mathbf{W}[n], \end{aligned} \tag{36}$$

where $\mathbf{R}[n] = [r_{t,j}[n]]$ is the $T \times N_r$ receive matrix, $\mathbf{H}[n] = [h_{i,j}[n]]$ is the $N_t \times N_r$ channel matrix, and $\mathbf{W}[n] = [w_{t,j}[n]]$ is the $T \times N_r$ noise matrix. The second equality comes from (34), where $\mathbf{H}_d[n] = [(h_d)_{i,j}[n]]$, and $\mathbf{H}_s[n] = [(h_s)_{i,j}[n]]$.

In [5], the signals are modulated by choosing a matrix from a finite group $\mathcal{V} = \{\mathbf{V}_l, l = 0, 1, \ldots, L-1\}$, where $\mathbf{V}_l$ is a $T \times N_t$ unitary matrix ($\mathbf{V}_l \mathbf{V}_l^H = \mathbf{I}_T$), and $L = 2^{N_t R}$, and R denotes the data rate. To make DUSTM feasible, we assume $T = N_t$ and $\mathbf{V}_0 = \mathbf{I}_{N_t}$. The $N_t R$ binary information bits are first converted to an integer l within $[0, L-1]$, and $\mathbf{V}[n] = \mathbf{V}_l$ is chosen from $\mathcal{V}$. The transmitted symbol at the nth block is encoded as

$$\mathbf{S}[n] = \mathbf{V}[n]\mathbf{S}[n-1]. \tag{37}$$

In the first block, $\mathbf{S}[0] = \mathbf{V}_0$ is sent. The internal composition property of a group ensures that $\mathbf{S}[n] \in \mathcal{V}$ and is unitary for any positive n. Specifically for diagonal constellations, the unitary matrices $\mathbf{V}_l$ are chosen as

$$\mathbf{V}_l = \mathrm{diag}\{e^{j2\pi u_1 l/L}, e^{j2\pi u_2 l/L}, \ldots, e^{j2\pi u_{N_t} l/L}\}, \tag{38}$$

where u_i for $i = 1, 2, \ldots, N_t$ are optimized to achieve the maximum diversity product [5].

B. Multiple-symbol differential space-time detection

We consider the sequence from $n = k+1$ to $n = k+N$. Let $\bar{\mathbf{R}}[k] = [\mathbf{R}^H[k+1], \mathbf{R}^H[k+2], \ldots, \mathbf{R}^H[k+N]]^H$ and $\bar{\mathbf{H}}[k] = [\mathbf{H}^H[k+1], \mathbf{H}^H[k+2], \ldots, \mathbf{H}^H[k+N]]^H$. We ignore the time index k in the following for simplicity. The input-output relationship for the N symbols can be expressed as

$$\bar{\mathbf{R}} = \bar{\mathbf{S}}_D \bar{\mathbf{H}} + \bar{\mathbf{W}} = \bar{\mathbf{S}}_D(\bar{\mathbf{H}}_d + \bar{\mathbf{H}}_s) + \bar{\mathbf{W}}, \tag{39}$$

where $\bar{\mathbf{S}}_D$ is a block diagonal matrix

$$\bar{\mathbf{S}}_D = \begin{bmatrix} \mathbf{S}[1] & & & \\ & \mathbf{S}[2] & & \\ & & \ddots & \\ & & & \mathbf{S}[N] \end{bmatrix} \tag{40}$$

and $\bar{\mathbf{H}}_d = [\mathbf{H}_d^H[1], \mathbf{H}_d^H[2], \ldots, \mathbf{H}_d^H[N]]^H$, $\bar{\mathbf{H}}_s = [\mathbf{H}_s^H[1], \mathbf{H}_s^H[2], \ldots, \mathbf{H}_s^H[N]]^H$, and $\bar{\mathbf{W}} = [\mathbf{W}^H[1], \mathbf{W}^H[2], \ldots, \mathbf{W}^H[N]]^H$. Similar to the argument in single-antenna systems, $\mathrm{vec}(\bar{\mathbf{R}})$ is also a complex Gaussian vector, and the conditional pdf given $\bar{\mathbf{S}}_D$ is

$$f(\bar{\mathbf{R}} | \bar{\mathbf{S}}_D) = A \exp\left\{-\mathrm{tr}\left((\bar{\mathbf{R}} - \bar{\mathbf{S}}_D\bar{\mathbf{H}}_d)^H \mathbf{C}_R^{-1} (\bar{\mathbf{R}} - \bar{\mathbf{S}}_D\bar{\mathbf{H}}_d)\right)\right\}. \tag{41}$$

where $A = \frac{1}{(\pi^{NN_t}\det(\mathbf{C}_R))^{N_r}}$. The autocovariance matrix $\mathbf{C}_R$ is given by

$$\mathbf{C}_R = E\{\bar{\mathbf{R}}\bar{\mathbf{R}}^H\} = \bar{\mathbf{S}}_D \mathbf{C}_H \bar{\mathbf{S}}_D^H + N_r \sigma_n^2 \mathbf{I}_{N_t N}, \tag{42}$$

where $\mathbf{C}_H$ is the covariance matrix of $\bar{\mathbf{H}}$ and can be represented as

$$\mathbf{C}_H = N_r(\mathbf{C}_h \otimes \mathbf{I}_{N_t}), \tag{43}$$

where $\otimes$ denotes the Kronecker product, and $\mathbf{C}_h$ is given by (10). Since $\mathbf{S}[n]$ ($n = 1, \ldots, N$) are unitary matrices, $\bar{\mathbf{S}}_D \bar{\mathbf{S}}_D^H = \mathbf{I}_{N_t N}$. We have

$$\begin{aligned} \mathbf{C}_R =& \bar{\mathbf{S}}_D \mathbf{C}_H \bar{\mathbf{S}}_D^H + N_r \sigma_n^2 \mathbf{I}_{N_t N} \\ =& N_r \bar{\mathbf{S}}_D \left(\mathbf{C} \otimes \mathbf{I}_{N_t}\right) \bar{\mathbf{S}}_D^H, \end{aligned} \tag{44}$$

[1] QS assumption is only used for derivation. However, we do not assume QS in the simulation.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

where $\mathbf{C} = \mathbf{C}_h + \sigma_n^2 \mathbf{I}_{N+1}$. We can show that $\det(\mathbf{C}_R)$ does not depend on $\bar{\mathbf{S}}_D$. Therefore, maximizing (41) is equivalent to minimizing

$$g(\bar{\mathbf{S}}_D) = \mathrm{tr}\left((\bar{\mathbf{R}} - \bar{\mathbf{S}}_D \bar{\mathbf{H}}_d)^H \mathbf{C}_R^{-1} (\bar{\mathbf{R}} - \bar{\mathbf{S}}_D \bar{\mathbf{H}}_d)\right). \quad (45)$$

Note that

$$\mathbf{C}_R^{-1} = \frac{1}{N_r} \bar{\mathbf{S}}_D \left(\mathbf{C}^{-1} \otimes \mathbf{I}_{N_t}\right) \bar{\mathbf{S}}_D^H. \quad (46)$$

We Cholesky factorize $\mathbf{C}^{-1}$ as $\mathbf{C}^{-1} = \mathbf{U}^H \mathbf{U}$, where $\mathbf{U}$ is upper triangular. Using the Kronecker product property $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, $\mathbf{C}^{-1} \otimes \mathbf{I}_{N_t} = (\mathbf{U} \otimes \mathbf{I}_{N_t})^H (\mathbf{U} \otimes \mathbf{I}_{N_t}) = \bar{\mathbf{U}}^H \bar{\mathbf{U}}$, and $\bar{\mathbf{U}}$ is also upper triangular. This factorization needs to be done only once. After several manipulations and ignoring constants, we can simplify (45) as

$$g(\bar{\mathbf{S}}_D) = \left\|\bar{\mathbf{U}}\bar{\mathbf{H}}_d - \bar{\mathbf{U}}\bar{\mathbf{S}}_D^H \bar{\mathbf{R}}\right\|_F^2 = \left\|\bar{\mathbf{Y}} - \bar{\mathbf{U}}\bar{\mathbf{S}}_D^H \bar{\mathbf{R}}\right\|_F^2 \quad (47)$$

where $\bar{\mathbf{Y}} = \bar{\mathbf{U}}\bar{\mathbf{H}}_d = [\mathbf{Y}^H[1], \mathbf{Y}^H[2], \ldots, \mathbf{Y}^H[N]]^H$, and $\mathbf{Y}[n]$ is an $N_t \times N_r$ matrix. The MSD rule for DUSTM over multiple-antenna Ricean channels is given by

$$\left\{\hat{\mathbf{S}}[1], \ldots, \hat{\mathbf{S}}[N]\right\} = \underset{\mathbf{S}[1],\ldots,\mathbf{S}[N] \in \mathcal{V}}{\arg\min} \left\|\bar{\mathbf{Y}} - \bar{\mathbf{U}}\bar{\mathbf{S}}_D^H \bar{\mathbf{R}}\right\|_F^2. \quad (48)$$

The transmitted signals can be differentially detected as

$$\hat{\mathbf{V}}[n] = \hat{\mathbf{S}}[n+1]\hat{\mathbf{S}}^H[n]. \quad (49)$$

When $N_t = N_r = 1$, the MSD rule (48) for multiple-antenna systems reduces to the single-antenna MSD rule (13).

Remarks:

- When $K = 0 \Rightarrow \mathbf{H}_d = \mathbf{0}$, (48) reduces to the decision metric in [7], which corresponds to the case of Rayleigh fading channels. When $K \to \infty$, $\sigma_h^2 \to 0$ and (48) reduces to

$$\left\{\hat{\mathbf{S}}[1], \ldots, \hat{\mathbf{S}}[N]\right\} = \underset{\mathbf{S}[1],\ldots,\mathbf{S}[N] \in \mathcal{V}}{\arg\min} \|\bar{\mathbf{R}} - \bar{\mathbf{S}}_D \bar{\mathbf{H}}_d\|^2. \quad (50)$$

  Eq. (50) is, in fact, coherent detection with perfect CSI.
- When K increases, the MSD performs more like a coherent detector, which has complexity linear in N. However, solving (50) needs exhaustively search over a set size L. We use the previously-derived BID to reduce the complexity for (50).

For brevity, we cannot outline the details of the BID algorithm. For the whole BID algorithm and efficient implementations, the interested readers can refer to [7], [8].

C. Efficient MSD detection

We now present our sphere decoding bound intersection detector (SD-BID) to solve the MSD rule (48). As with the SD, we only examine the candidates that satisfy

$$\|\bar{\mathbf{Y}} - \bar{\mathbf{U}}\bar{\mathbf{S}}_D \bar{\mathbf{R}}\|^2 \leq R^2. \quad (51)$$

Let the entries of $\mathbf{U}$ be denoted by $u_{i,j}$, $i \leq j$. Taking the upper triangular and Kronecker product structure of $\bar{\mathbf{U}}$ into account, (51) can be written as

$$\sum_{i=1}^{N} \left\| \mathbf{Y}[i] - \sum_{j=i}^{N} u_{i,j} \mathbf{S}[j] \mathbf{R}[j] \right\|_F^2 \leq R^2 \quad (52)$$

To proceed, we start from $\mathbf{S}[N]$. Using the BID, we can obtain its candidate set. When all $\hat{\mathbf{S}}[i]$ has found, all the R_i's are updated according to

$$R_N^2 = \|\bar{\mathbf{Y}} - \bar{\mathbf{U}}\bar{\mathbf{S}}_D \bar{\mathbf{R}}\|^2,\ R_i^2 = R_{i+1}^2 - d_{i+1}^2,\ i = N-1, \ldots, 1. \quad (53)$$

The same process continues until all the candidates that meet (51) have been checked. The best candidate is output as the ML solution. The initial radius R can also be obtained according to the statistic of $g(\bar{\mathbf{S}}_D)$ in (45)

$$g(\bar{\mathbf{S}}_D) = \mathrm{tr}\left((\bar{\mathbf{R}} - \bar{\mathbf{S}}_D \bar{\mathbf{H}}_d)^H \mathbf{C}_R^{-1} (\bar{\mathbf{R}} - \bar{\mathbf{S}}_D \bar{\mathbf{H}}_d)\right). \quad (54)$$

If $\bar{\mathbf{S}}_D$ is the true solution, using (36), $\mathbf{X} = \bar{\mathbf{S}}_D^H \bar{\mathbf{R}} - \bar{\mathbf{H}}_d = \bar{\mathbf{H}}_s + \mathbf{S}_D^H \bar{\mathbf{W}}$ is zero mean complex Gaussian with autocovariance matrix $\mathbf{C}_X = \mathbf{C}_h + \sigma_n^2 \mathbf{I}_N$. Therefore $e = \mathrm{tr}\{\mathbf{X}^H (\mathbf{C}_h + \sigma_n^2 \mathbf{I}_N)^{-1} \mathbf{X}\}$ is a chi-square random variable with $2NN_rN_t$ degrees of freedom. Similar to the SD, R^2 can be chosen to make the probability that e is less than R^2 very high. The Schnorr and Euchner (SE) [11] can also be generalized to SD-BID. These details can be found in the journal version of our paper.

D. Reduced-state DD

Assuming correct decisions of $\hat{\mathbf{S}}[1], \ldots, \hat{\mathbf{S}}[N-1]$, the MSD for DUSTM (48) can be readily modified to MSD based DF-DD by replacing $\mathbf{S}[1], \ldots, \mathbf{S}[N-1]$ in (48) with $\hat{\mathbf{S}}[1], \ldots, \hat{\mathbf{S}}[N-1]$. Our BID can be used to solve the DF-DD. We also note that decision feedback sequence estimator is a special case of the reduced-state sequence estimator (RSSE) [14]. Similarly, a reduced-state differential detector (RS-DD) can be used to solve (48) as a generalization of the DF-DD. Instead of assuming $N-1$ correct feedbacks in (48), RS-DD only uses M $(0 \leq M \leq N-1)$ decision feedbacks. $\mathbf{S}[1], \ldots, \mathbf{S}[M]$ in (48) are replaced with $\hat{\mathbf{S}}[1], \ldots, \hat{\mathbf{S}}[M]$, and SD-BID is used for the resulting $N-M$ dimensional problem. If $M = 0$, RS-DD reduces to SD-BID and DF-DD when $M = N-1$. Thus, both the performance and complexity of RS-DD are between SD-BID and DF-DD.

IV. Simulation Results

We now present simulation results for both single-antenna and multiple-antenna systems. We assume that the receiver has perfect knowledge of K, $\mathbf{C}_h$ and σ_n^2. Estimation algorithms for such parameters are available in the literature.

A single-antenna system with 8DPSK and Gray encoding is simulated over a Ricean fading channel. The Jakes' model is assumed for the channel. Fig. 1 shows the BER versus SNR for MSD by the use of SD (or SD for short), MSD based DF-DD (MSD DF-DD), prediction based DF-DD (Prediction DF-DD) with $N = 3, 6$, $f_DT = 0.03$ and Rice factor $K = 3$ dB. They are compared with CDD and CD with perfect CSI. All the schemes SD, MSD DF-DD and Prediction DF-DD significantly reduce the error floor encountered in CDD. In fact, the error floor is virtually eliminated. The performance difference between DF-DD and SD is about 2 dB for $N = 3$ at a BER of 2×10^{-4}. But the gap increases to 4 dB when

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

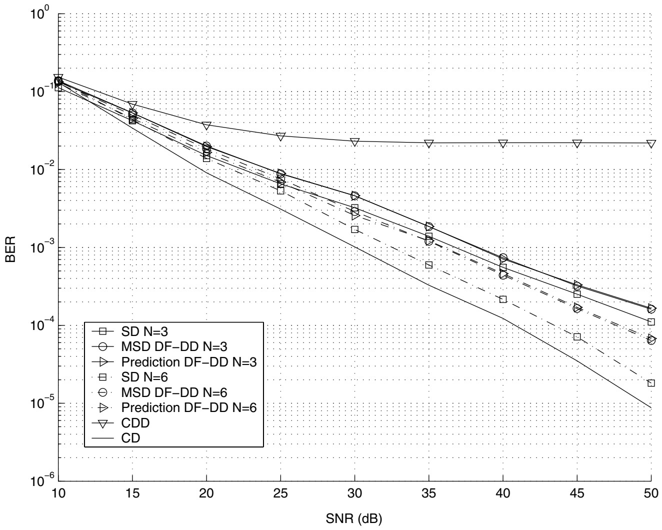

Fig. 1. The performance comparison between SD, MSD based DF-DD, Prediction based DF-DD, CDD and CD with $N = 3, 6$ for 8DPSK over flat Ricean channels ($f_D T = 0.03$ and $K =$ 3dB).

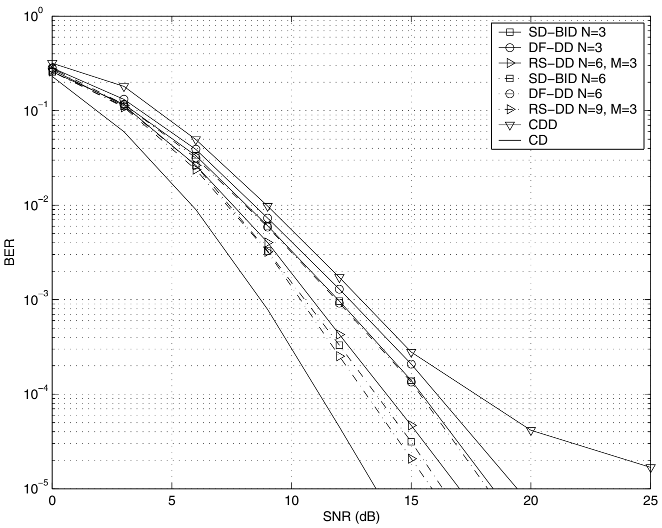

Fig. 2. The performance comparison between SD-BID, MSD based DF-DD, CDD and CD with $N = 3, 6$ for DUSTM ($N_t = 4$, $N_r = 1$ and $R = 1$) over flat Ricean channels ($f_D T_B = 0.03$ and $K =$ 5dB).

$N = 6$. The SD scheme still performs 2 dB worse than CD. This performance loss can be reduced via increasing N.

In the MIMO case, the $N_t = 4$, $N_r = 1$ and rate $R = 1$ DUSTM is used. The code parameters are taken from [5]. The Jakes' model is assumed for each channel. The direct channel matrix is assumed to be $\mathbf{H}_d[n] = \sqrt{K/(K+1)}\mathbf{1}_{N_t \times N_r}$ [13], where $\mathbf{1}_{N_t \times N_r}$ is an all one matrix.

Fig. 2 shows the BER versus SNR for SD-BID, MSD based DF-DD (DF-DD), with $N = 3, 6$, $f_D T = 0.03$ and Rice factor $K = 5$ dB [13]. When $N = 3$, SD-BID has a 2 dB loss over CD at BER $= 5 \times 10^{-4}$. The performance loss of SD-BID over CD reduces as N increases. At a BER of 5×10^{-4}, the DF-DD scheme performs 0.6 dB and 1.2 dB worse than SD-BID. We also show the performance of RS-DD in Fig. 2. When $N = 6$, $M = 3$, RS-DD has about 0.6 dB gain over SD-BID with $N = 3$, where both use 3 dimensional exhaustive search. RS-DD outperforms SD-BID by 0.2 dB when $N = 9$, $M = 3$. RS-DD is a good candidate to achieve good performance while maintaining reasonable complexity.

V. Conclusion

In this paper, we have derived the optimal decision metrics of multiple symbol differential detection for both MDPSK in single-antenna systems and DUSTM in multiple-antenna systems over Ricean fading channels. The SD has been proposed to optimally solved the MSD detection problem for MDPSK. A modification of the Schnorr and Euchner strategy was proposed to remove the complexity dependence on initial radius and reduce the complexity. We have also proposed a SD-BID algorithm to efficiently solve the MSD rule for DUSTM. The SE strategy has also been generalized to the multiple-antenna case. Many details have been omitted for brevity, but will be forthcoming in a journal paper.

References

[1] D. Divsalar and M. K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 38, no. 3, pp. 300 – 308, March 1990.

[2] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, inlcuding a complexity analysis," *Math. Computation*, vol. 44, pp. 463–471, Apr. 1985.

[3] L. Lampe, R. Schober, V. Pauli, and C. Windpassinger, "Multiple-symbol differential sphere decoding," in *Proc. of ICC*, June 2004, pp. 787 – 791.

[4] R. Schober, W. H. Gerstacker, and J. B. Huber, "Decision-feedback differential detection of MDPSK for flat Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1025 – 1035, July 1999.

[5] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2041 – 2052, Dec. 2000.

[6] R. Schober and L. Lampe, "Noncoherent receivers for differential space-time modulation," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 768 – 777, May 2002.

[7] T. Cui and C. Tellambura, "Multiple-symbol differential unitary space-time demodulation with reduced-complexity," in *Proc. of ICC*, May 2005, pp. 16–20.

[8] ——, "Bound intersection detection for multiple-symbol differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 53, no. 12, pp. 2114 – 2123, Dec. 2005.

[9] R. Schober and W. H. Gerstacker, "Decision-feedback differential detection based on linear prediction for MDPSK signals transmitted over Ricean fading channels," *IEEE J. Select. Areas Commun.*, vol. 18, no. 3, pp. 391 – 402, March 2000.

[10] P. Ho and D. Fung, "Error performance of multiple-symbol differential detection of PSK signals transmitted over correlated Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 40, no. 10, pp. 1566 – 1569, Oct. 1992.

[11] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Programming*, vol. 66, pp. 181–191, 1994.

[12] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–15, Jan. 1999.

[13] R. U. Nabar, H. Bolcskei, and A. J. Paulraj, "Outage properties of space-time block codes in correlated Rayleigh or Rician fading environments," in *Proc. of ICASSP*, May 2002, pp. 13–17.

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

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.