Maximum Likelihood Detection for Differential Unitary Space-Time Modulation with Carrier Frequency Offset

Feifei Gao, Tao Cui, Arumugam Nallanathan, *Senior Member, IEEE*, and Chinthananda Tellambura, *Senior Member, IEEE*

***Abstract*—Can conventional differential unitary space time modulation (DUSTM) be applied when there is an unknown carrier frequency offset (CFO)? This paper answers this question affirmatively and derives the necessary maximum likelihood (ML) detection rule. The asymptotic performance of the proposed ML rule is analyzed, leading to a code design criterion for DUSTM by using the modified diversity product. The resulting proposed decision rule is a new differential modulation scheme in both the temporal and spatial domains. Two sub-optimal multiple-symbol decision rules with improved performance are also proposed. For the efficient implementation of these, we derive a modified bound intersection detector (BID), a generalization of the previously derived optimal BID for the conventional DUSTM. The simulation results show that the proposed differential modulation scheme is more robust against CFO drifting than the existing double temporal differential modulation.**

***Index Terms*—Carrier frequency offset, differential unitary space time code, maximum likelihood, MIMO, bound intersection, wireless communications.**

I. Introduction

WIRELESS multiple-input multiple-output (MIMO) systems and space time coding (STC) techniques have received much attention due to their potential to increase the transmission rate and to combat fading [1]. Reliable symbol detection in traditional STC requires accurate channel estimation at the receiver and is a difficult task as the fading rate and/or the number of transmit antennas increases. Consequently, single differential STC (DSTC) schemes have been proposed in [2]- [4], and the full diversity gain is achieved without requiring the channel state information (CSI).

However, the performance of DSTC degrades greatly if the carrier frequency offset (CFO)[1] is not compensated for at the receiver before symbol detection. To the best of our knowledge, most DSTC papers [2]- [11], however, focus on either the code design or the receiver design based on the assumption of perfect synchronization. Conventionally popular CFO-compensation methods involve training-sequence based estimation [12], [13]. However, this approach clearly contradicts the original motivation for using DSTC. Recently, Liu, Giannakis and Hughes [14] and Ma [15] proposed a new modulation scheme[2] that can handle both unknown CSI and CFO. This method is based on unitary matrix group [3], [4] and uses time-domain double differential modulation, where the first(second) differentiation removes the unknown CSI(CFO).

Although DDUSTM successfully deals with CFO, it is of interest to investigate whether the conventional DUSTM itself can handle the unknown CFO, which was believed to be impossible for single antenna transmission [14]. This paper investigates this issue and provides an affirmative answer to the following question: *Can conventional DUSTM thus be applied in the presence of an unknown CFO?* We show that by sacrificing one degree of freedom in transmit antennas, the ML single-symbol detector (SSD), which processes two consecutive blocks, can be derived for DUSTM even in the presence of an unknown CFO. This loss of degree of freedom can be viewed as the result of applying spatial differential modulation across transmit antennas, which helps mitigating the impact of the unknown CFO. The proposed modulation scheme is named as modified DUSTM (MDUSTM) in this paper. A multiple-symbol detector (MSD), which performs better than the SSD, for MDUSTM is developed. However, since a simple decision rule for this case appears to be impossible, two suboptimal MSD rules are derived instead (they have the advantage of the ease of implementation). The asymptotic performance of MDUSTM with SSD is analyzed, and the modified diversity product instead of the conventional diversity product for DUSTM [3] is defined. The modified diversity product enables the derivation of an

Paper approved by H. Jafarkhani, the Editor for Wireless Communications of the IEEE Communications Society. Manuscript received December 7, 2006; revised May 14, 2007 and August 23, 2007. This work was supported in part by the National University of Singapore under Grant R-263-000-421-112. This paper was presented in part at the IEEE Global Telecommunications Conference, Washington, DC, USA, November 2007.

F. Gao is with the Institute for Infocomm Research, 1 Fusionopolis Way, #21-01 Connexis, South Tower, Singapore 138632 (e-mail: feifeigao@ieee.org).

A. Nallanathan is with the Division of Engineering, King's College London, London WC2R 2LS, U.K. (e-mail: arumugam.nallanathan@kcl.ac.uk).

T. Cui is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA (e-mail: taocui@caltech.edu).

C. Tellambura is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: chintha@ece.ualberta.ca).

Digital Object Identifier 10.1109/TCOMM.2008.060672

[1] In wireless communications, CFO arises due to the transceiver oscillator mismatch or the Doppler shift caused by the relative motion between the transmitter and the receiver.

[2] Throughout this paper, the original differential scheme proposed in [3], [4] is named as differential unitary space time modulation (DUSTM) and the double differential scheme in [14], [15] is named as double differential unitary space time modulation (DDUSTM).

0090-6778/08$25.00 © 2008 IEEE

asymptotically-optimal code design criterion. Although the asymptotic design is predicated on a large number of antennas, the simulation results show that it is nevertheless effective in practical scenarios where the number of antennas is finite. To avoid the naive exhaustive search for the SSD and MSD, we generalize the bound intersection detector (BID) [16] to optimally detect MDUSTM, resulting in the reduction of the detection complexity. Since MDUSTM requires constant CFO and channel for two consecutive blocks, the simulation results show that MDUSTM is more robust to the CFO drifting than DDUSTM, which requires a constant CFO and channel for three consecutive blocks.

This paper is organized as follows. Section II presents the DUSTM system model. Section III provides the proposed ML decision rules of the DUSTM in the presence of an unknown CFO, i.e., the MDUSTM algorithm, including both the SSD and MSD cases. In Section IV, we analyze the pairwise symbol error rate (SER) and derive the asymptotically optimal code design criterion. Section V develops the fast decoding algorithm for MDUSTM, i.e., the modified BID algorithm. Section VI displays various simulation results and finally, conclusions are drawn in Section VII. The related proofs are given in the appendices.

Notation: Vectors and matrices are represented in boldface small and capital letters; the transpose, complex conjugate, Hermitian, inverse, and pseudo-inverse of matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$ and $\mathbf{A}^\dagger$, respectively; $tr(\mathbf{A})$ and $\|\mathbf{A}\|_F$ are the trace and the Frobenius norm of $\mathbf{A}$; $\mathrm{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with the diagonal element constructed from $\mathbf{a}$; $\Re\{\mathbf{A}\}$ and $\Im\{\mathbf{A}\}$ represent the real and imaginary part of a matrix; $\delta(\cdot)$ denotes the Kronecker Delta function; $gcd(a,b)$ is the greatest common divisor of a and b; $\mathbf{I}$ is the identity matrix; $\mathrm{E}\{\cdot\}$ denotes the statistical expectation; $|x|$ is the absolute value of a scalar x; $\lfloor x \rfloor$ denotes the closest integer to x to minus infinity, and the imaginary unit $\mathrm{j} = \sqrt{-1}$.

II. System Model

Consider a MIMO system with N_t transmit and N_r receive antennas over a Rayleigh flat fading environment. Each time slot occupies an interval of T_s seconds, and each transmit-signal block consists of N_t time slots. The transmit signals are generated by choosing a unitary matrix from a finite group $\mathcal{V} = \{\mathbf{V}_l, l = 0, 1, \ldots, L-1\}$, where $L = 2^{N_t R}$, and R denotes the data rate. As in [14] and [15], we consider diagonal constellations based on the cyclic group whose unitary matrices $\mathbf{V}_l$ are written as

$$\mathbf{V}_l = \mathrm{diag}\left\{e^{\mathrm{j}2\pi u_1 l/L}, e^{\mathrm{j}2\pi u_2 l/L}, \ldots, e^{\mathrm{j}2\pi u_{N_t} l/L}\right\}, \quad (1)$$

where $u_i\,(i = 1, \ldots, N_t)$ are integers and can be optimized, for instance, to minimize the upper bound of the pairwise error probability (PEP) [3]. The transmitted symbol during the kth block is denoted by the $N_t \times N_t$ matrix

$$\mathbf{S}[k] = \mathbf{V}[k]\mathbf{S}[k-1], \quad (2)$$

and the (t,i)th element of $\mathbf{S}[k]$ is transmitted from the ith antenna in the $t+(k-1)N_t$ time slot. The initial transmission matrix $\mathbf{S}[0]$ can be chosen as an arbitrary matrix from $\mathcal{V}$.

The received signals from all N_r antennas through time slot $1+(k-1)N_t$ to kN_t may be expressed in an $N_t \times N_r$ matrix form

$$\mathbf{X}[k] = \sqrt{\rho}\mathbf{S}[k]\mathbf{H}[k] + \mathbf{W}[k], \quad (3)$$

where ρ is the received signal-to-noise ratio (SNR) and $\mathbf{W}[k]$ is the $N_t \times N_r$ noise matrix whose entries are complex Gaussian random variables with unit variance and are independent across the receiver and time indices. The (i,j)th entry of $\mathbf{H}[k]$, denoted by $h_{i,j}[k]$, is the channel gain from the ith transmit antenna to the jth receive antenna, which remains constant during one block interval. All path gains, with unit variances, are statistically independent ($\mathrm{E}\{h_{i,j}[k]h^*_{i',j'}[k]\} = \delta(i-i')\delta(j-j')$) and have the same autocorrelation function $\varphi_h[n] = \mathrm{E}\{(h_{i,j}[k+n]h_{i,j}[k]^*\}$. Typically, when Jakes' model [17] is used, $\varphi_h[n]$ is given by

$$\varphi_h[n] = J_0\left(2\pi f_D N_t T_s n\right), \quad (4)$$

where f_D is the Doppler spread due to user mobility, and $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind.

Suppose the CFO is constant in one block but may vary from block to block due to the CFO drifting [15]. Let ε_k denote the normalized CFO by the sampling period T_s during the kth block. Then, the phase distortion on the $i+(k-1)N_t$ time slot is

$$\theta_{i,k} = 2\pi\varepsilon_k, \quad i = 1, ..., N_t, \ k = 0, 1, \ldots. \quad (5)$$

Therefore, the overall phase rotation before the $i+(k-1)N_t$ time slot, denoted as $\vartheta_{i,k}$, is that accumulated from all the previous time slots:

$$\vartheta_{i,k} = \sum_{\tau<i}\sum_{c<k}\theta_{\tau,c} = 2\pi N_t \sum_{c=0}^{k-1}\varepsilon_c + 2\pi(i-1)\varepsilon_k. \quad (6)$$

If CFO does not vary, as is usually the case, the accumulated phase $\vartheta_{i,k}$ becomes $2\pi(N_t k + i - 1)\varepsilon_0$. The receive signal block $\mathbf{X}[k]$ with the unknown CFO is then modelled as

$$\mathbf{X}[k] = \sqrt{\rho}e^{j\vartheta_{1,k}}\mathbf{\Gamma}(\varepsilon_k)\mathbf{S}[k]\mathbf{H}[k] + \mathbf{W}[k], \quad (7)$$

where $\mathbf{\Gamma}(\varepsilon_k)$ is the diagonal matrix with the form

$$\mathbf{\Gamma}(\varepsilon_k) = \mathrm{diag}\left\{1, e^{\mathrm{j}2\pi\varepsilon_k}, \ldots, e^{\mathrm{j}2\pi(N_t-1)\varepsilon_k}\right\}. \quad (8)$$

III. ML Detection of DUSTM with CFO

A. Single Symbol Detection

For the ease of exposition, we assume that both the CFO and the channel remain constant over two blocks, say the $(k-1)$th block and the kth block. Hence, the indices $k-1$ and k are dropped in both the channel matrix and the CFO. The received signal blocks may be rewritten as

$$\mathbf{X}[k-1] = \sqrt{\rho}e^{\mathrm{j}\vartheta_{1,k-1}}\mathbf{\Gamma}(\varepsilon)\mathbf{S}[k-1]\mathbf{H} + \mathbf{W}[k-1], \quad (9)$$

$$\begin{aligned}\mathbf{X}[k] &= \sqrt{\rho}e^{\mathrm{j}(\vartheta_{1,k-1}+2\pi N_t\varepsilon)}\mathbf{\Gamma}(\varepsilon)\mathbf{S}[k]\mathbf{H} + \mathbf{W}[k],\\ &= \left(e^{\mathrm{j}2\pi N_t\varepsilon}\mathbf{V}[k]\right)\left(\sqrt{\rho}e^{\mathrm{j}\vartheta_{1,k-1}}\mathbf{\Gamma}(\varepsilon)\mathbf{S}[k-1]\mathbf{H}\right)\\ &\quad + \mathbf{W}[k]. \end{aligned} \quad (10)$$

$$\{\hat{\varepsilon}, \hat{\bar{\mathbf{V}}}[k]\} = \arg\max_{\varepsilon, \bar{\mathbf{V}}[k]} \sum_{i=0}^{N-1} \sum_{j=i+1}^{N} b_{i,j} \Re\left\{ e^{\mathrm{j}2\pi N_t (j-i)\varepsilon} tr\left(\left(\prod_{m=i+k}^{j+k-1} \mathbf{V}[m] \right) \mathbf{X}[i+k-1]\mathbf{X}^H[j+k-1] \right) \right\}. \tag{16}$$

$$\hat{\bar{\mathbf{V}}}[k] = \arg\max_{\bar{\mathbf{V}}[k]} \sum_{i=0}^{N-1} \sum_{j=i+1}^{N} b_{i,j} \left| tr\left(\left(\prod_{m=i+k}^{j+k-1} \mathbf{V}[m] \right) \mathbf{X}[i+k-1]\mathbf{X}^H[j+k-1] \right) \right|. \tag{18}$$

$$\hat{\bar{\mathbf{V}}}[k] = \arg\max_{\bar{\mathbf{V}}[k]} \sum_{i=0}^{N-1} \sum_{j=i+1}^{N} b_{i,j} \left| tr\left(\left(\prod_{m=i+k}^{j+k-1} \mathbf{V}[m] \right) \mathbf{X}[i+k-1]\mathbf{X}^H[j+k-1] \right) \right|^2. \tag{19}$$

The probability density function (PDF) of $\bar{\mathbf{X}}[k] \triangleq [\mathbf{X}^T[k-1], \mathbf{X}^T[k]]^T$ is expressed as

$$f(\bar{\mathbf{X}}[k] \big| \mathbf{V}[k], \varepsilon, \vartheta_{1,k-1}) = \frac{\exp\left(-tr\left(\mathbf{\Lambda}^{-1}\bar{\mathbf{X}}[k]\bar{\mathbf{X}}^H[k]\right)\right)}{\pi^{2N_t N_r} \det^{N_r}\{\mathbf{\Lambda}\}}, \tag{11}$$

where

$$\mathbf{\Lambda} = \mathbf{I}_{2N_t} + \rho \begin{bmatrix} \mathbf{I}_{N_t} & e^{-\mathrm{j}2\pi N_t \varepsilon} \mathbf{V}^H[k] \\ e^{\mathrm{j}2\pi N_t \varepsilon} \mathbf{V}[k] & \mathbf{I}_{N_t} \end{bmatrix}. \tag{12}$$

Since $f(\bar{\mathbf{X}}[k] \big| \mathbf{V}[k], \varepsilon, \vartheta_{1,k-1})$ does not depend on $\vartheta_{1,k-1}$, it may be rewritten as $f(\bar{\mathbf{X}}[k] \big| \mathbf{V}[k], \varepsilon)$. Furthermore, it can be easily shown that $\det\{\mathbf{\Lambda}\}$ is a constant by applying the formula $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$. Using the matrix inversion formula [18], the ML estimate can be obtained as

$$\begin{aligned} \left\{\hat{\varepsilon}, \hat{\mathbf{V}}[k]\right\} &= \arg\max_{\varepsilon, \mathbf{V}[k]} f(\bar{\mathbf{X}}[k] \big| \mathbf{V}[k], \varepsilon) \qquad (13) \\ &= \arg\max_{\varepsilon, \mathbf{V}[k]} K \exp\left[\Re\left\{e^{\mathrm{j}2\pi N_t \varepsilon} tr(\mathbf{V}[k]\mathbf{X}[k-1]\mathbf{X}^H[k])\right\}\right] \\ &= \arg\max_{\varepsilon, \mathbf{V}[k]} K \exp\left[\left| tr(\mathbf{V}[k]\mathbf{X}[k-1]\mathbf{X}^H[k])\right|\right] \\ &\qquad \times \exp\left[\cos(\psi_k + 2\pi N_t \varepsilon)\right], \end{aligned}$$

where ψ_k is the phase of $tr(\mathbf{V}[k]\mathbf{X}[k-1]\mathbf{X}^H[k])$, and K contains all the factors of $f(\bar{\mathbf{X}}[k] \big| \mathbf{V}[k], \varepsilon)$ that are independent of both $\mathbf{V}[k]$ and ε. Obviously, the maximization of the argument in (13) occurs when $\cos(\psi_k + 2\pi N_t \varepsilon) = 1$, which requires $\hat{\varepsilon}$ to be

$$\hat{\varepsilon} = -\frac{\psi_k}{2\pi N_t} + \frac{n}{N_t}, \quad n = 0, ..., N_t - 1, \tag{14}$$

and n is the integer ambiguity related to the estimation of ε. Substituting (14) back into (13), the ML[3] detection (MLD) for $\mathbf{V}[k]$ is obtained as

$$\hat{\mathbf{V}}[k] = \arg\max_{\mathbf{V}[k]} \left| tr(\mathbf{V}[k]\mathbf{X}[k-1]\mathbf{X}^H[k]) \right|. \tag{15}$$

Clearly, the ambiguity n does not affect the MLD of $\mathbf{V}[k]$.

Remarks:

- If $N_t = 1$, then $|tr(\mathbf{V}[k]\mathbf{X}[k-1]\mathbf{X}^H[k])| = |tr(\mathbf{X}[k-1]\mathbf{X}^H[k])|$. Therefore, for single-input single-output (SISO) or single-input multiple-output (SIMO) DPSK with unknown CFO, MLD (15) fails. Unlike a matrix noncoherent channel, where an unknown unitary matrix prevents any detection without further (i.e., double) differential encoding, the problem at hand involves only a scalar unknown phase (i.e., the unknown frequency offset), and thus, only one degree of freedom is lost.
- Since DDUSTM in [14] is derived for relatively slow fading or for the channels static over three blocks, we expect that our method is more robust to CFO drifting than that in [14], as confirmed by the simulation results.

[3]The more precise terminology is "generalized likelihood testing" (GLRT) [19] but the ML terminology is used in [20].

B. Multiple Symbol Detection

If CFO remains constant across more than two blocks or varies very slowly among several blocks, multiple symbol detection (MSD) can also be applied for DUSTM as in [10], [16]. In MSD, the transmitted symbols in N consecutive intervals are estimated using $N+1$ received blocks. Let us define $\bar{\mathbf{V}}[k] = [\mathbf{V}^H[k], \ldots, \mathbf{V}^H[k+N-1]]^H$. By using the same approach as in [10], [16], the joint ML estimates of $\bar{\mathbf{V}}[k]$ and ε can be derived as in (16), where $b_{i,j}$ is the (i,j)-th entry of $\mathbf{B} = (\mathbf{C}_h + \sigma_n^2 \mathbf{I}_{N+1})^{-1}$, and

$$\mathbf{C}_h = \begin{bmatrix} \varphi_h[0] & \varphi_h[1] & \cdots & \varphi_h[N] \\ \varphi_h[-1] & \varphi_h[0] & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_h[-N] & \cdots & \cdots & \varphi_h[0] \end{bmatrix} \tag{17}$$

is the channel covariance matrix. Since the phase $(j-i)\varepsilon$ varies for different terms in (16), obtaining a closed-form solution of ε by solving (16) with fixed $\bar{\mathbf{V}}[k]$, as we did in (13), is impossible. Therefore, we propose a suboptimal MSD algorithm instead. Clearly, without the noise, the true CFO ε maximizes each summand in (16). We thus consider that ε in each summand of (16) is different. Specifically, we replace ε with $\varepsilon_{i,j}$ in the (i,j)-th summand of (16). Given $\bar{\mathbf{V}}[k]$, each summand in (16) is maximized individually in $\varepsilon_{i,j}$, and the resulting $\varepsilon_{i,j}$ is substituted into (16). Similar to (15), the suboptimal MSD rule is obtained as (18)

We denote (18) as MSD1 in the following. However, it is difficult to design efficient detection algorithms for (18) directly because the sum of absolute terms is difficult to handle in practice. As an approximation of (18), we propose a new estimation criterion as shown in (19) for designing efficient MSD algorithms. We denote (19) as MSD2. The validity of this approximation is also justified by our simulation results in Section VI, where MSD1 and MSD2 are shown to perform similarly.

Remarks:

- Note that (18) reduces to (15) and is the true MLD when $N = 1$. Moreover, since the derivation of (18) does not assume a constant channel, (15) works even for time-varying channels.
- One must resort to the suboptimal rule (18) because the exact multiple-symbol MLD appears intractable. For the same reason, only a suboptimal MLD algorithm is derived in [14].

IV. Performance Analysis and Code Design

The performance analysis and code design are derived for the case of SSD. The application of the codes to MSD will be numerically verified in our simulations.

A. Pairwise Error Probability Analysis

Assuming the codeword $\mathbf{V}_l \in \mathcal{V}$ was transmitted, the probability that the MLD (15) decides $\mathbf{V}_{l'} \in \mathcal{V}$ rather than $\mathbf{V}_l$, when only these two codewords are possible, is given by

$$P_e\left(\mathbf{V}_l \to \mathbf{V}_{l'}\right) = P\Big\{ \left|tr\left(\mathbf{V}_l \mathbf{X}[k-1]\mathbf{X}^H[k]\right)\right| < \left|tr\left(\mathbf{V}_{l'} \mathbf{X}[k-1]\mathbf{X}^H[k]\right)\right| \ \Big| \ \mathbf{V}_l \text{ transmitted}\Big\}. \quad (20)$$

A similar definition holds for $P_e\left(\mathbf{V}_{l'} \to \mathbf{V}_l\right)$. It is proved in Appendix I that $P_e\left(\mathbf{V}_{l'} \to \mathbf{V}_l\right) = P_e\left(\mathbf{V}_l \to \mathbf{V}_{l'}\right)$. Therefore, the PEP is denoted as P_e for brevity. The high SNR approximation of P_e is proved to be

$$\begin{aligned} P_e &= \mathrm{E}_{\mathbf{H}}\left\{P_{e|\mathbf{H}}\right\} \\ &= \mathrm{E}_{\mathbf{H}}\left\{\frac{1}{2}\left[1 + Q(\sqrt{a},\sqrt{b}) - Q(\sqrt{b},\sqrt{a})\right]\right\}, \end{aligned} \quad (21)$$

where

$$Q(\alpha,\beta) = \int_{\beta}^{\infty} \exp\left(-\frac{\alpha^2+x^2}{2}\right) I_0(\alpha x) x dx \quad (22)$$

is the Marcum Q function, and a, b are defined in Appendix I. The P_e derived from high SNR approximation is sufficient for the later code design based on optimizing the coding gain and the diversity gain (which are defined at a high SNR only). The expression (21) is, in fact, related to the fourth order Gaussian statistics and currently has no solution. The Marcum Q function can be approximated in several ways, as it has been, for example, in [21] and [22]. However, we will adopt a more general way by utilizing the following bounds of the Marcum Q function. From [23], we know

$$\begin{cases} Q(\alpha,\beta) \le \exp\left(-\frac{(\beta-\alpha)^2}{2}\right), & \beta > \alpha \ge 0, \\ Q(\alpha,\beta) \ge 1 - \frac{1}{2}\left[\exp\left(-\frac{(\alpha-\beta)^2}{2}\right) - \exp\left(-\frac{(\alpha+\beta)^2}{2}\right)\right], & \alpha > \beta \ge 0. \end{cases}$$

Then

$$\begin{aligned} P_{e|\mathbf{H}} &\le \frac{1}{4}\left[3\exp\left(-\frac{(\sqrt{b}-\sqrt{a})^2}{2}\right) - \exp\left(-\frac{(\sqrt{a}+\sqrt{b})^2}{2}\right)\right] \\ &\le \frac{3}{4}\exp\left(-\frac{(\sqrt{b}-\sqrt{a})^2}{2}\right). \end{aligned} \quad (23)$$

Substituting the specific forms of a, b into (23) results in

$$P_{e|\mathbf{H}} \le \frac{3}{4}\exp\left(-\frac{\rho}{4}\left(\|\mathbf{H}\|_F^2 - |tr(\mathbf{H}^H\mathbf{Q}\mathbf{H})|\right)\right), \quad (24)$$

where $\mathbf{Q} \triangleq \mathbf{V}_l^H \mathbf{V}_{l'}$. The unfriendly factor $|tr(\mathbf{H}^H\mathbf{Q}\mathbf{H})|$ forbids the further derivation of the conditional PEP bound (24). To simplify the analysis, we use the same asymptotic performance as that used in [24], [25]. The following two cases are considered:

1) $N_r \to \infty$ and N_t is finite: As in [24], we assume that the variance of h_{ij} is $1/N_r$, because the normalizing factor $1/N_r$ ensures that the total power received by N_r antennas from each transmit antenna remains constant as N_r approaches ∞. By the strong law of large numbers, we have $\mathbf{H}\mathbf{H}^H \approx \mathbf{I}_{N_t}$ and $|tr(\mathbf{H}^H\mathbf{Q}\mathbf{H})| \approx |tr(\mathbf{Q})|$. Then

$$P_{e|\mathbf{H}} \le \frac{3}{4}\exp\left(-\frac{\rho}{4}(N_t - |tr(\mathbf{Q})|)\right). \quad (25)$$

Now that $P_{e|\mathbf{H}}$ is independent from $\mathbf{H}$, the upper bound of P_e is

$$P_e \le \frac{3}{4}\exp\left(-\frac{\rho}{4}(N_t - |tr(\mathbf{Q})|)\right). \quad (26)$$

2) $N_t \to \infty$ and N_r is finite: It is noted that

$$\begin{aligned} \|\mathbf{H}\|_F^2 - |tr(\mathbf{H}^H\mathbf{Q}\mathbf{H})| &= \sum_{j=1}^{N_r}\sum_{i=1}^{N_t}|h_{ij}|^2 - \left|\sum_{j=1}^{N_r}\sum_{i=1}^{N_t} q_i |h_{ij}|^2\right| \\ &\ge \sum_{j=1}^{N_r}\left(\sum_{i=1}^{N_t}|h_{ij}|^2 - \left|\sum_{i=1}^{N_t} q_i |h_{ij}|^2\right|\right). \end{aligned} \quad (27)$$

Since h_{ij} are independent and identically distributed random variables over (i,j), the upperbound on P_e may be expressed as

$$\begin{aligned} P_e &= \mathrm{E}\{P_{e|\mathbf{H}}\} \\ &\le \frac{3}{4}\mathrm{E}^{N_r}\left\{\exp\left(-\frac{\rho}{4}\left(\sum_{i=1}^{N_t}|h_{i1}|^2 - \left|\sum_{i=1}^{N_t} q_i |h_{i1}|^2\right|\right)\right)\right\}. \end{aligned} \quad (28)$$

From the free probability theory [24, Fact A.2 or Eq. (24)], we know

$$\sum_{i=1}^{N_t} q_i |h_{i1}|^2 \longrightarrow \frac{1}{N_t} tr(\mathbf{Q}) \sum_{i=1}^{N_t} |h_{i1}|^2. \quad (29)$$

Then, P_e can be asymptotically bounded by

$$\begin{aligned} P_e &\le \frac{3}{4}\mathrm{E}^{N_r}\left\{\exp\left(-\frac{\rho}{4}\left(\sum_{i=1}^{N_t}\left(1 - \frac{1}{N_t}|tr(\mathbf{Q})|\right)|h_{i1}|^2\right)\right)\right\} \\ &= \frac{3}{4}\mathrm{E}^{N_t N_r}\left\{\exp\left(-\frac{\rho}{4}\left(1 - \frac{1}{N_t}|tr(\mathbf{Q})|\right)|h_{11}|^2\right)\right\} \\ &= \frac{3}{4}\left(\frac{1}{1 + \frac{\rho}{4}\left(1 - \frac{1}{N_t}|tr(\mathbf{Q})|\right)}\right)^{N_t N_r}. \end{aligned} \quad (30)$$

B. Code Design

To minimize P_e in both cases, it is clear from (26) and (30) that $|tr(\mathbf{Q})|$ should be minimized. As in [3], we define the *modified diversity product* ζ as

$$\zeta = \max_{0 \le l < l' \le L-1} \left|tr\left(\mathbf{V}_l^H \mathbf{V}_{l'}\right)\right|. \quad (31)$$

TABLE I
CYCLIC GROUP CODE DESIGN FOR MDUSTM WITH $L = 2^{RN_t}$.

N_t	R	L	$[u_1, u_2, ..., u_{N_t}]$	N_t	R	L	$[u_1, u_2, ..., u_{N_t}]$
2	1	4	[0, 3]	2	1.5	8	[0, 5]
3	1	8	[0, 1, 3]	4	1.5	64	[0, 11, 55, 59]
4	1	16	[0, 2, 11, 15]	2	2	16	[0, 13]
5	1	32	[0, 15, 25, 26, 28]	4	2	256	[0, 131, 135, 186]

TABLE II
CYCLIC GROUP CODE DESIGN FOR MDUSTM BY USING ODD/PRIME L.

N_t	L	R	$[u_1, u_2, ..., u_{N_t}]$	N_t	L	R	$[u_1, u_2, ..., u_{N_t}]$
2	3	0.7925	[0, 1]	3	11	1.1531	[0, 7, 10]
2	5	1.1610	[0, 4]	4	13	0.9251	[0, 1, 3, 9]
3	7	0.9358	[0, 2, 3]	4	15	0.9767	[0, 1, 3, 7]
3	9	1.0566	[0, 3, 8]	4	17	1.0219	[0, 2, 7, 16]

Therefore, the design criterion is to minimize the modified diversity product. For cyclic group codes, the unitary matrices $\mathbf{V}_l$ are chosen as in (1), and u_i for $i = 1, \ldots, N_t$ has to be optimized to achieve the minimum modified diversity product (31). Denote $l - l^{'}$ as Δl. The modified diversity product ζ can be simplified as

$$\zeta = \max_{0<\Delta l \le L-1} \left| \sum_{i=1}^{N_t} e^{\mathrm{j}2\pi u_i \Delta l/L} \right|. \tag{32}$$

As in [3], we consider only the integer choices of u_i and use an exhaustive computer search for the best set of $u_1, \ldots, u_{N_t} \in \{0, ..., L-1\}$. The search space can be reduced by using the following rules:

1) Since $|tr(\mathbf{Q})| = |e^{-\mathrm{j}2\pi u_1 \Delta l/L} tr(\mathbf{Q})|$, we can always set $u_1 = 0$.
2) The order of u_i does not affect the metric, so we always assume that $u_{i_2} \ge u_{i_1}$, if $i_2 > i_1$.
3) In (32), Δl and $L - \Delta l$ give the same ζ. Hence, we need to search for Δl only from 1 to $\lfloor \frac{L}{2} \rfloor$.

Table I shows the results of our search for parameters that minimize ζ.

Normally, $L = 2^{RN_t}$ is an even number. Then u_i cannot all be odd/even numbers since this condition would cause $\mathbf{V}_l$ and $\mathbf{V}_{l+L/2}$ to be indistinguishable. Intuitively, this effect may reduce the flexibility of the search space. This effect can be eliminated by taking L as an odd number or a prime number. The design method is described as follows. Suppose we prefer a transmission with roughly rate R. We can choose L as an odd or prime number around 2^{RN_t}. Whether L is greater than or less than 2^{RN_t} depends on the exact rate requirement. Similar searching for u_i could then be carried out. Table II provides some search results for u_i with the rate around $R = 1$.

Remarks:

- We can show that $|tr(\mathbf{Q})| = N_t$ leads to the worst PEP, which occurs when $\mathbf{Q} = \gamma \mathbf{I}_{N_t}$ where γ is a complex scalar with a unit norm. The traditionally designed cyclic group codes [3] are among those who yield the worst PEP. For example, if we choose $l = l_1$, $l^{'} = l_1 + L/2$, then $|tr(\mathbf{Q})| = \left| \sum_{i=1}^{N_t} e^{\mathrm{j}2\pi u_i \Delta l/L} \right| = \left| \sum_{i=1}^{N_t} e^{\mathrm{j}\pi u_i} \right|$. From [3], all u_i are odd integers. Therefore, we have $|tr(\mathbf{Q})| = N_t$, and $\zeta = N_t$.
- In Section III, we showed that the proposed method cannot work for single transmit antenna systems, e.g., SISO, SIMO. From the design criterion above, we know that the constant transmission will be adopted on one antenna. This fact suggests that the MLD rule (15) would not achieve the full transmit diversity because handling the additional unknown CFO sacrifices one degree of freedom in transmit antennas.
- The previous remark also suggests that the MLD rule (15) can be considered as employing differential modulation across the transmit antennas, resulting in a new differential modulation scheme. Our scheme applies differential modulation in both the spatial and temporal domains as opposed to only the temporal used in DDUSTM [14].
- The design criterion obtained by maximizing the modified diversity product is not limited to cyclic group codes, but can also be used in other diagonal constellation designs.

V. EFFICIENT DETECTION ALGORITHMS

A. Single Symbol Detection

The detection of $\mathbf{V}[k]$ by solving (15) requires an exhaustive search over the whole constellation $\mathcal{V}$. The complexity of this naive algorithm increases exponentially with the increase of the transmission rate or the number of the transmit antennas, and prohibits the use of this algorithm in practice. Therefore, we adapt the BID proposed in [16] for DUSTM detection to the proposed SSD rule (15).

We consider detecting cyclic group codes [3] in this section, where the unitary matrices $\mathbf{V}_l$ are chosen as (1).

Denote the (i, i)-th entry of $\mathbf{X}[k-1]\mathbf{X}^H[k]$ as d_i. The MLD

(15) is equivalent to

$$\begin{aligned}\hat{l} &= \arg\max_{l}\left|\sum_{k=1}^{N_t} d_k e^{\mathrm{j}2\pi u_k l/L}\right| \\ &= \arg\max_{l}\sum_{i=1}^{N_t}\sum_{j=1}^{N_t} d_i d_j^* e^{\mathrm{j}2\pi u_i l/L} e^{-\mathrm{j}2\pi u_j l/L} \\ &= \arg\min_{l}\sum_{i=1}^{N_t-1}\sum_{j=i+1}^{N_t} -2\Re\left\{d_i d_j^* e^{\mathrm{j}2\pi u_i l/L} e^{-\mathrm{j}2\pi u_j l/L}\right\} \\ &= \arg\min_{l}\sum_{i=1}^{N_t-1}\sum_{j=i+1}^{N_t} A_{i,j} - B_{i,j}\cos\left(\left(u_{i,j}l-\phi_{i,j}\right)2\pi/L\right) \\ &= \arg\min_{l}\lambda(l), \end{aligned} \tag{33}$$

where

$$\begin{aligned} A_{i,j} &= |d_i|^2 + |d_j|^2, & B_{i,j} &= 2\left|d_i d_j\right|, \\ u_{i,j} &= u_j - u_i, & \phi_{i,j} &= \angle(d_i/d_j)L/2\pi, \end{aligned}$$

and $\angle(d_i/d_j)$ is the phase of d_i/d_j. The approximated PDF of $\lambda(l_0)$ corresponding to the true transmitting code $\mathbf{V}_{l_0}$ is derived in Appendix II. Instead of searching all of the $0 \le l < L$, we can search only the values of l subject to $\lambda(l) < C$, where C is the initial searching radius such that the probability $P(\lambda(l_0) < C)$ is no less than ϵ. From Appendix II, we know that C should be selected from

$$F_\chi\left(N_t\sum_{i=1}^{N_t}|d_i|^2 - C\right) \le 1-\epsilon, \tag{34}$$

where F_χ is defined in Appendix II.

To find all the l's that satisfy $\lambda(l) < C$, we note that $\lambda(l)$ in (33) consists of $\frac{N_t(N_t-1)}{2}$ non-negative terms. Thus, a necessary condition for $\lambda(l) < C$ is that each term of (33) is less than C, or equivalently

$$A_{i,j} - B_{i,j}\cos\left(\left(\mathrm{mod}(u_{i,j}l, L) - \phi_i\right)2\pi/L\right) < C, \quad 1 \le i < j \le N_t, \tag{35}$$

where $\mathrm{mod}\,(x, L)$ reduces x to an integer between 0 and $L-1$. Let us define the candidate set $\mathcal{L}_{i,j} = \{l \,|\, A_{i,j} - B_{i,j}\cos\left(\left(\mathrm{mod}(u_{i,j}l, L) - \phi_i\right)2\pi/L\right) < C, 0 \le l < L\}$. To find $\mathcal{L}_{i,j}$, we first show how, given a number n, to find an l such that $\mathrm{mod}(u_{i,j}l, L) = n$. In [16], an efficient algorithm by using the well-known *Extended Euclidean Algorithm* [26] is given by considering the fact that $u_{i,j}$ and L are coprime. However, for our problem, $u_{i,j}$ and L may not be coprime as shown in Tables I and II.

A new algorithm is developed here. Let $\xi_{i,j} = gcd(u_{i,j}, L)$, $u'_{i,j} = u_{i,j}/\xi_{i,j}$, and $L' = L/\xi_{i,j}$. Clearly, we have $\mathrm{mod}(u_{i,j}l, L) = \xi_{i,j} \cdot \mathrm{mod}(u'_{i,j}l, L') = n$. Therefore, there exists such an l if and only if $\mathrm{mod}(n, \xi_{i,j}) = 0$. If $\mathrm{mod}(n, \xi_{i,j}) = 0$ holds, let $n' = n/\xi_{i,j}$. Since now $u'_{i,j}$ is relatively prime to L', the gcd of $u'_{i,j}$ and L' is 1. The Extended Euclidean Algorithm computes the gcd of $u'_{i,j}$ and L', as well as the numbers $\mu_{i,j}$ and κ such that

$$u'_{i,j}\mu_{i,j} + \kappa L' = 1, \tag{36}$$

where 1 is the gcd of $u'_{i,j}$ and L'. For the details of the Extended Euclidean Algorithm, see [26]. To find $\mathrm{mod}(u'_{i,j}l, L') = n'$, we multiply both sides of (36) by n', to obtain

$$u'_{i,j}(n'\mu_{i,j}) + \kappa n' L' = n'. \tag{37}$$

Therefore, $\tilde{l} = \mathrm{mod}(n'\mu_{i,j}, L')$ satisfies $\mathrm{mod}(u'_{i,j}\tilde{l}, L') = n'$, or, equivalently, $\mathrm{mod}(u_{i,j}\tilde{l}, L) = n$. Since $0 \le \tilde{l} < L'$, a total of $\xi_{i,j}$ different l satisfy $\mathrm{mod}(u_{i,j}l, L) = n$ (They are $l = \tilde{l} + kL', k = 0, \ldots, \xi_{i,j} - 1$, respectively). We can now determine $\mathcal{L}_{i,j}$. Define

$$UB_{i,j} = \lfloor \phi_i + \rho_i \rfloor, \quad LB_{i,j} = \lceil \phi_i - \rho_i \rceil, \tag{38}$$

where

$$\rho_{i,j} = \frac{L}{2\pi}\cos^{-1}\left(\frac{A_{i,j}-C}{B_{i,j}}\right). \tag{39}$$

Let $\mathcal{S}_{i,j} = \{n/\xi_{i,j} | n \in \mathbb{Z}, LB_i \le n \le UB_i, \text{and}\, \mathrm{mod}(n, \xi_{i,j}) = 0\}$. Following the same arguments as in [16], we find $\mathcal{L}_{i,j}$ is given by

$$\mathcal{L}_{i,j} = \bigcup_{0 \le k < \xi_{i,j}} \left(\mathrm{mod}(\mu_{i,j}\mathcal{S}_{i,j}, L') + kL'\right), 1 \le i < j \le N_T, \tag{40}$$

where both the mod operation and addition are performed component-wise. The candidate set for all l's such that $\lambda(l) < C$ is the intersection of all of the $\frac{N_t(N_t-1)}{2}$ sets $\mathcal{L}_{i,j}$:

$$\mathcal{L} = \bigcap_{1 \le i < j \le N_T} \mathcal{L}_{i,j}. \tag{41}$$

As with the BID in [16], we first choose l^* from $\mathcal{L}$. C is then replaced by the new cost $\lambda(l^*)$, and l^* is deleted from the set $\mathcal{L}$ ($\mathcal{L} = \mathcal{L} - \{l^*\}$). All $\mathcal{L}_{i,j}$ $1 \le i < j \le N_T$ are updated by using the new bound C. In later iterations, (41) is replaced by

$$\mathcal{L} = \mathcal{L} \cap \left(\bigcap_{1 \le i < j \le N_T} \mathcal{L}_{i,j}\right). \tag{42}$$

The process continues until $\mathcal{L}$ becomes the null set. The l with the minimum cost is then the optimal solution. This optimal detection algorithm is called the "modified BID".

If $u_{i,j}$ is coprime to L or $\xi_{i,j} = 1$, as shown in Table II, the modified BID reduces to the original BID [16]. Alternatively, (15) can also be suboptimally solved by using the lattice reduction algorithm proposed in [27] for the detection of DUSTM.

B. Multiple Symbol Detection

Note that (19) has the same form as (8) in [16]. By combining our modified BID, (19) can be efficiently solved by using the MSDs proposed in [16]. Moreover, (19) can also be solved by using the sphere decoding BID proposed in [28] by replacing the BID in [28] with the modified BID. With these variants of MSD, (19) can be solved efficiently.

VI. Simulation Results

In this section, the proposed MDUSTM scheme is simulated. The SER is the figure of merit. The signal transmitted in the first block is chosen as $\mathbf{S}[0] = \mathbf{I}_{N_t}$.

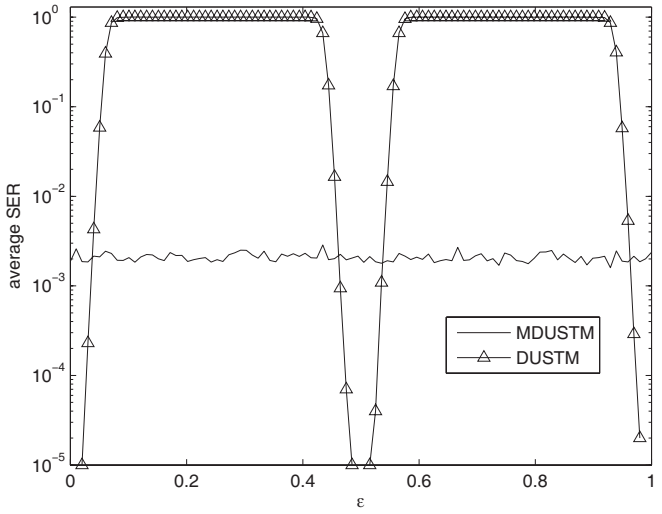


Fig. 1. Average SER versus CFO for $N_t = N_r = 2$, $R = 1$ (SNR=12dB).

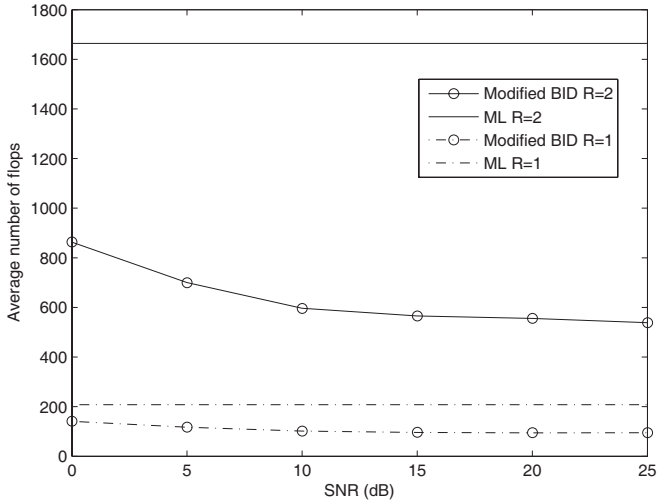


Fig. 2. Average number of flops versus SNR for $N_t = 3$, $N_r = 4$ with the rate $R = 1$ and $R = 2$.

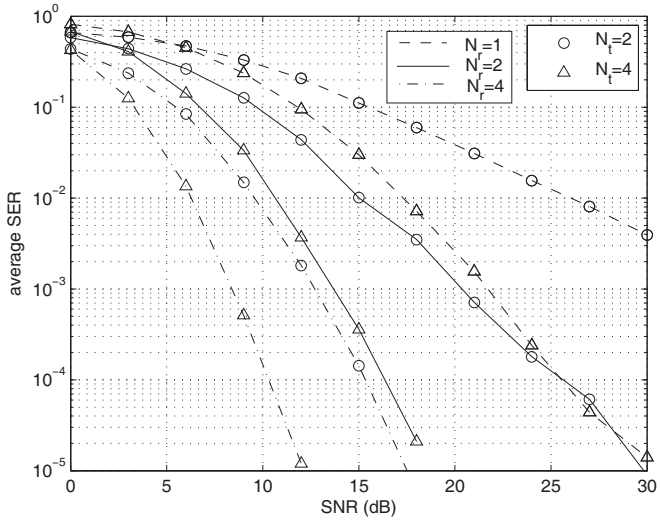


Fig. 3. Average SER versus SNR for different N_t and N_r with $R = 1$. The frequency offset is $0.4/N_t$.

A. Performance Comparison with DUSTM under Frequency Errors

This experiment investigates the performance of a 2×2 MIMO system with the MDUSTM scheme and the DUSTM without compensating for the frequency errors. The maximum range of ε, $[0, 1]$, is examined. For both schemes, their respective optimal codes with $R = 1$ ($\mathbf{u} = [0, 3]$, and $\mathbf{u} = [1, 1]$) are employed. The simulation results are shown in Fig. 1 for SNR$= 12$ dB. We note that, when the CFO is small, the conventional DUSTM without CFO compensation performs better. The SER plot of DUSTM is periodic over ε. However, for a large range of CFO values, DUSTM totally fails while MDUSTM is effective. A similar phenomenon has been observed in [14] and [15].

B. Complexity Reduction by Using the Modified BID

Fig. 2 shows the complexity of the modified BID in flops for $N_t = 3$ and $N_r = 4$ with different rates $R = 1$ and $R = 2$, respectively. The complexity exhibited includes only that from the detection process. We use the flops function in MATLAB for both the modified BID and the ML detection (33) (this function provides an estimate of the number of floating-point operations performed by a sequence of MATLAB statements). No parallelization is considered. For $R = 2$, the modified BID is about 2 times less complex than the ML. Thus, the use of the modified BID is more computationally efficient than that of the conventional ML searching. However, the complexity of the modified BID is not reduced as quickly as that of the traditional BID [16] when SNR increases.

C. Performance for Different N_t, N_r

We now examine the performance of the proposed scheme for different numbers of antennas. The asymptotically designed codes shown in Table I with $R = 1$ are used. The SER performance as a function of SNR is depicted in Fig. 3. The proposed code, although designed for an asymptotically large N_t or N_r, performs well for a small number of transceiver antennas. Moreover, Fig. 3 also shows that, increasing N_t or N_r would enhance the diversity order of the proposed MDUSTM algorithm.

D. Performance Comparison of the Designed Optimal Codes with All Other Codes

Next, we investigate how the asymptotically optimal codes perform with a finite number of transceiver antennas. In this example, we choose $N_t = 3$, $N_r = 4$, $R = 1$. The performance curves of our designed code $\mathbf{u} = [0, 1, 3]$, as well as those of all other possible $\mathbf{u}$'s are presented in Fig. 4. Basically, different choices of $\mathbf{u}$ will provide different performance. All codes could be divided into several groups, and within each group, the codes would provide similar performance. However, the performance from group to group are quite different. One group in particular totally fails the detection. Our asymptotically optimal codes belong to the group that gives the best performance. In fact, only one choice of $\mathbf{u} = [0, 1, 6]$ in this example outperforms the designed code. This result is not unexpected since the optimal code for a finite

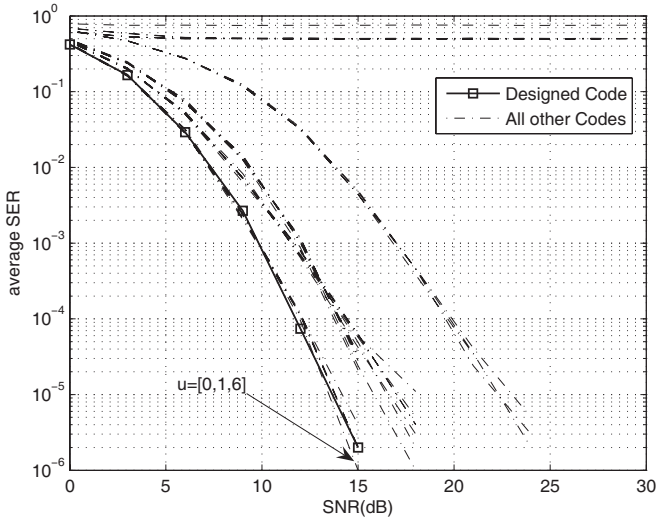


Fig. 4. Demonstration of the optimality of the designed code under regular transceiver antennas, $N_t = 3$, $N_r = 4$, $R = 1$. The frequency offset is $0.4/N_t$.

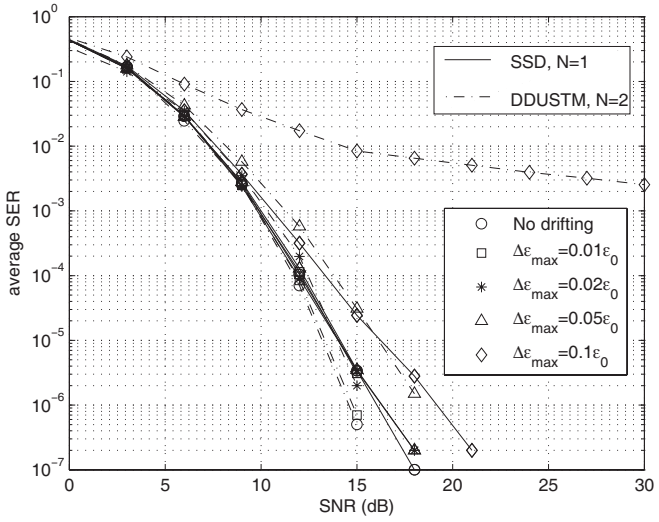


Fig. 5. Performance comparison between SSD and DDUSTM under various CFO drifting, $N_t = 3$, $N_r = 4$, $R = 1$.

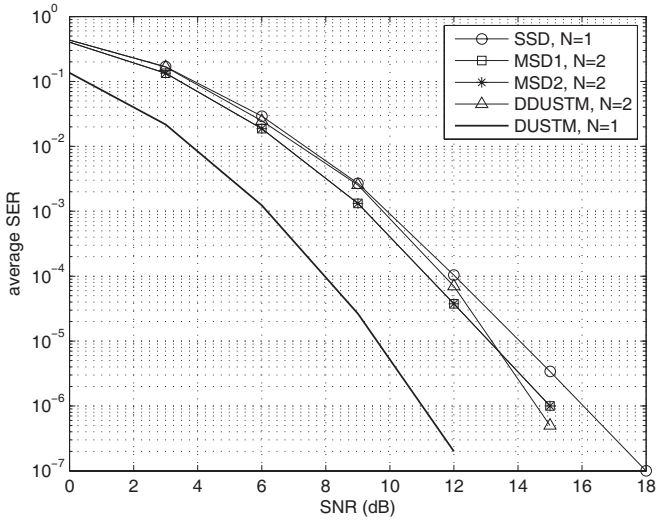


Fig. 6. Performance comparison of SSD, MSD1, MSD2 and DDUSTM, $N_t = 3$, $N_r = 4$, $R = 1$ with the frequency offset $0.4/N_t$. The DUSTM performance with zero frequency offset is also included as reference.

number of antennas may not necessarily be the asymptotic codes. Nevertheless, we might safely claim that our codes give near optimal performance even for a finite number of antennas.

E. Performance Comparison with Normal DDUSTM under Frequency Drifting

The optimal cyclic group code for both MDUSTM and DDUSTM is tested in this experiment. Since the optimal cyclic group codes are not necessarily the universal optimal codes, we cannot claim one method is better based only on the numerical figures. However, since MDUSTM sacrifices one degree of freedom in transmit antennas to cope with CFO, it is expected that the DDUSTM would outperform MDUSTM at high SNR due to its higher diversity order.

Nevertheless, it is found that our proposed scheme is more robust than DDUSTM against frequency drifting. Similar to [15], the CFO drifting is modeled as a random process from block to block, but the drifting is zero inside each block. Suppose the CFO at the kth block is ε_k. For the $(k+1)$st block, the CFO becomes

$$\varepsilon_{k+1} = \varepsilon_k + \Delta\varepsilon_k, \tag{43}$$

where $\Delta\varepsilon_k$ is drawn uniformly from $[-0.5\Delta\varepsilon_{\max}, 0.5\Delta\varepsilon_{\max}]$. Four different values of $\Delta\varepsilon_{\max}$ are chosen, and the performance comparisons between MDUSTM and DDUSTM are plotted in Fig. 5. The parameters are selected as $N_t = 3$, $N_r = 4$, $R = 1$, $\varepsilon_0 = 0.4/N_t$. Both schemes employ the optimal code designs $\mathbf{u} = [0, 1, 3]$ and $\mathbf{u} = [1, 1, 3]$, respectively. With a smaller CFO drifting, both schemes are affected slightly. However, when $\Delta\varepsilon_{\max}$ goes beyond $0.05\varepsilon_0$, the performance of DDUSTM degrades significantly. Especially for very large drifting[4] $\Delta\varepsilon_{\max} = 0.1\Delta\varepsilon_0$, the performance of DDUSTM is almost unacceptable. This evidence suggests that MDUSTM is more robust than the DDUSTM for mitigating CFO drifting.

[4]The large value is examined for theoretical study

F. Performance of Multi-Symbol Detection

This experiment examines the performance of our proposed MSD1 and MSD2 and compares them with SSD and DDUSTM. The parameters are chosen as $N_t = 3$, $N_r = 4$, $R = 1$ and $N = 2$. No CFO drifting is considered. The average SER as a function of SNR for these four algorithms is displayed in Fig. 6. Both MSD1 and MSD2 perform similarly and, at the same time, are better than SSD. The diversity order of MSD1 and MSD2 is similar to that of SSD, an observation consistent with those in traditional MSD [16], [22]. Intuitively, we expect that the SNR gain by applying MSD goes to 3 dB when N is sufficiently large. Note that MSD1 and MSD2 outperform DDUSTM at lower SNR, although all these three algorithms use the same number of received blocks. However, since DDUSTM has higher diversity, it is superior to MSD1 and MSD2 at higher SNR.

In addition, we include a performance curve for the traditional DUSTM [3] with zero CFO in Fig, 6. Bearing in mind

the performance of DDUSTM or MDUSTM is not related with the value of the CFO, we see that DUSTM with no CFO performs around 4 dB better than both DDUSTM and MDUSTM. Therefore, the receiver benefits from knowing whether CFO exists or not. However, if wrong information is used, DUSTM may fail to detect any symbol as is seen from Fig. 1.

VII. Conclusions

In this paper, we developed transmit-symbol detection algorithms for DUSTM in the presence of an unknown CFO. The proposed MDUSTM removes the effect of the unknown CFO by spatial differentiation, whereas the conventional method relies on temporal differentiation. We have derived the asymptotic pairwise symbol error rate and, accordingly, designed the asymptotically optimal codes for a large number of antennas. Nevertheless, these asymptotic codes also perform well when the number of transceiver antennas is small. In order to reduce the detection complexity, we developed a modified BID algorithm for both single-symbol and multiple-symbol cases of the proposed MDUSTM scheme. Since the proposed MDUSTM requires a slowly varying channel and constant CFO during only two blocks, it is more robust than DDUSTM against CFO drifting. Finally, all the results provided in this paper can be readily generalized to Ricean MIMO channels.

Acknowledgement

We would like to express our thanks to Dr. Rong Li from National University of Singapore for suggesting reference [23]. We also thank the editor H. Jafarkhani and anonymous reviewers for their helpful comments to improve the quality of this work.

Appendix I
High SNR Approximation of PEP and Its Upper Bound

Consider a specific realization of the channel matrix $\mathbf{H}$. It is known that the error event $\mathbf{V}_l \to \mathbf{V}_{l'}$ happens if

$$\left|tr\left(\mathbf{V}_l\mathbf{X}[k-1]\mathbf{X}^H[k]\right)\right| < \left|tr\left(\mathbf{V}_{l'}\mathbf{X}[k-1]\mathbf{X}^H[k]\right)\right|, \tag{44}$$

when $\mathbf{V}_l$ is remitted. The inequality (44) could be further converted to

$$\left|tr\left((\sqrt{\rho}\mathbf{H}+\tilde{\mathbf{W}}[k])^H(\sqrt{\rho}\mathbf{H}+\tilde{\mathbf{W}}[k-1])\right)\right| < \left|tr\left((\sqrt{\rho}\mathbf{H}+\tilde{\mathbf{W}}[k])^H\mathbf{Q}(\sqrt{\rho}\mathbf{H}+\tilde{\mathbf{W}}[k-1])\right)\right|, \tag{45}$$

where

$$\tilde{\mathbf{W}}[k-1] = e^{-\mathrm{j}\vartheta_{1,k-1}}\mathbf{\Gamma}^H(\varepsilon)\mathbf{S}^H[k-1]\mathbf{W}[k], \tag{46}$$

$$\tilde{\mathbf{W}}[k] = e^{-\mathrm{j}(\vartheta_{1,k-1}+2\pi N_t\varepsilon)}\mathbf{\Gamma}^H(\varepsilon)\mathbf{S}^H[k-1]\mathbf{V}_l^H\mathbf{W}[k-1]. \tag{47}$$

All entries of $\tilde{\mathbf{W}}[k-1]$ or $\tilde{\mathbf{W}}[k]$ are independent and identically distributed (i.i.d) Gaussian random variables with unit variance. Another observation is that the probability of (45) does not change if we replace $\mathbf{Q}$ by $\mathbf{Q}^H$. Therefore, $P_{e|\mathbf{H}}(\mathbf{V}_{l'} \to \mathbf{V}_l) = P_{e|\mathbf{H}}(\mathbf{V}_l \to \mathbf{V}_{l'})$ and $P_e(\mathbf{V}_{l'} \to \mathbf{V}_l) = P_e(\mathbf{V}_l \to \mathbf{V}_{l'})$.

At the high SNR region, the second-order noise term can be ignored and then (45) becomes

$$\left|\underbrace{\sum_{i=1}^{N_t}\sum_{j=1}^{N_r}\left(\sqrt{\rho}|h_{ij}|^2 + h_{ij}^*\tilde{w}_{ij}[k-1] + h_{ij}\tilde{w}_{ij}^*[k]\right)}_{x}\right| < \left|\underbrace{\sum_{i=1}^{N_t}\sum_{j=1}^{N_r}\left(\sqrt{\rho}|h_{ij}|^2 q_i + h_{ij}^*\tilde{w}_{ij}[k-1]q_i + h_{ij}\tilde{w}_{ij}^*[k]q_i\right)}_{y}\right|, \tag{48}$$

where q_i denotes the (i,i)th entry of the diagonal matrix $\mathbf{Q}$. Since all h_{ij} are currently deterministic values, x, y are Gaussian random variables. The probability of $|x| < |y|$ has been derived in [21]. Following [21], we obtain

$$\begin{aligned}
&\bar{x} = \sum_{i=1}^{N_t}\sum_{j=1}^{N_r}\sqrt{\rho}|h_{ij}|^2 = \sqrt{\rho}tr(\mathbf{H}^H\mathbf{H}) = \sqrt{\rho}\|\mathbf{H}\|_F^2,\\
&\theta_{1f} = 0,\\
&\mathrm{cov}(x) = 2\sum_{i=1}^{N_t}\sum_{j=1}^{N_r}|h_{ij}|^2 = 2tr(\mathbf{H}^H\mathbf{H}) = 2\|\mathbf{H}\|_F^2,\\
&\bar{y} = \sum_{i=1}^{N_t}\sum_{j=1}^{N_r}\sqrt{\rho}|h_{ij}|^2 q_i = \sqrt{\rho}tr(\mathbf{H}^H\mathbf{Q}\mathbf{H}),\\
&\theta_{2f} = \arg\{tr(\mathbf{H}^H\mathbf{Q}\mathbf{H})\},\\
&\mathrm{cov}(y) = 2\sum_{i=1}^{N_t}\sum_{j=1}^{N_r}|h_{ij}|^2 = 2\|\mathbf{H}\|_F^2,\\
&\mathrm{cov}(x,y) = 2tr(\mathbf{H}^H\mathbf{Q}\mathbf{H}),\\
&\phi = arg\{tr(\mathbf{H}^H\mathbf{Q}\mathbf{H})\}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&S_{1f} = |\bar{x}|^2/2 = \rho\|\mathbf{H}\|_F^4/2,\\
&N_{1f} = \mathrm{cov}(x)/2 = \|\mathbf{H}\|_F^2,\\
&S_{2f} = |\bar{y}|^2/2 = \rho|tr(\mathbf{H}^H\mathbf{Q}\mathbf{H})|^2/2,\\
&N_{2f} = \mathrm{cov}(y)/2 = \|\mathbf{H}\|_F^2,\\
&\rho_f = tr(\mathbf{H}^H\mathbf{Q}\mathbf{H})/\|\mathbf{H}\|_F^2.
\end{aligned}$$

Since $\rho_f = \sqrt{\frac{S_{2f}}{S_{1f}}}\exp[\mathrm{j}(\theta_{2f}-\theta_{1f})]$, and $N_{1f} = N_{2f} = N_f$, we define

$$\begin{aligned}
\left\{\begin{array}{c} a \\ b \end{array}\right\} &= \frac{S_{1f}}{2N_f}(1 \mp \sqrt{1-|\rho_f|^2})\\
&= \frac{S_{1f}}{2N_f}\left(1 \mp \sqrt{1-\frac{S_{2f}}{S_{1f}}}\right).
\end{aligned} \tag{49}$$

The PEP conditioned on $\mathbf{H}$ is [21]

$$P_{e|\mathbf{H}} = \frac{1}{2}\left[1 + Q(\sqrt{a},\sqrt{b}) - Q(\sqrt{b},\sqrt{a})\right], \tag{50}$$

and the PEP is

$$P_e = \mathrm{E}_{\mathbf{H}}\left\{\frac{1}{2}\left[1 + Q(\sqrt{a},\sqrt{b}) - Q(\sqrt{b},\sqrt{a})\right]\right\}. \tag{51}$$

APPENDIX II
ON THE RADIUS OF THE MODIFIED BID ALGORITHM

From (33), we have the following equation:

$$\lambda(l) = N_t \sum_{i=1}^{N_t} |d_i|^2 - \left| tr(\mathbf{V}_l[k]\mathbf{X}[k-1]\mathbf{X}^H[k]) \right|^2. \quad (52)$$

For the true solution l_0, $|tr(\mathbf{V}_l\mathbf{X}[k-1]\mathbf{X}^H[k])|$ could be represented by the left-hand side of (45), which is independent from both code group $\mathcal{V}$ and the CFO ε. The radius $\lambda(l_0)$ corresponding to the true solution l_0 is related to SNR in the modified BID. As we did in Appendix I, we consider the high SNR approximation such that $\sqrt{\rho}x$ could be used to represent $tr(\mathbf{V}_l\mathbf{X}[k-1]\mathbf{X}^H[k])$. Consequently, $\chi = \rho|x|^2$ is a non-central chi-square distributed random variable with mean $\rho\|\mathbf{H}\|_F^2$ and variance $\rho\|\mathbf{H}\|_F^2$ per dimension, conditioned on fixed $\mathbf{H}$. The conditioned PDF of χ is then expressed as

$$p_\chi(\chi|\mathbf{H}) = \frac{1}{2\rho\|\mathbf{H}\|_F^2} e^{-\frac{\rho^2\|\mathbf{H}\|_F^4 + \chi}{2\rho\|\mathbf{H}\|_F^2}} I_0(\sqrt{\chi}), \quad \chi \geq 0, \quad (53)$$

where $I_\nu(\cdot)$ is the νth order modified Bessel functions of the first kind. As well, $\iota = \|\mathbf{H}\|_F^2$ is a chi-square random variable with degree $2N_tN_r$ whose PDF is

$$p_\iota(\iota) = \frac{2^{N_tN_r}}{(N_tN_r-1)!} \iota^{N_tN_r-1} e^{-\iota}, \quad \iota \geq 0. \quad (54)$$

Therefore, $p_\chi(\chi)$ can be calculated as

$$\begin{aligned} p_\chi(\chi) &= \int_0^\infty p_\chi(\chi|\mathbf{H}) p_\iota(\iota) d\iota \\ &= \frac{2^{N_tN_r-1} I_0(\sqrt{\chi})}{\rho(N_tN_r-1)!} \int_0^\infty \iota^{N_tN_r-2} e^{-(\frac{\rho}{2}+1)\iota - \frac{\chi}{2\rho\iota}} d\iota. \end{aligned} \quad (55)$$

From [29, pp. 384, and pp.364], the above PDF could be written into a closed form as

$$p_\chi(\chi) = \begin{cases} \frac{2^{N_tN_r}}{\rho(N_tN_r-1)!} \left(\frac{\chi}{\rho^2+2\rho}\right)^{\frac{N_tN_r-1}{2}} \\ \qquad \times I_0(\sqrt{\chi}) K_{(N_tN_r-1)} \left(2\sqrt{\left(\frac{1}{4}+\frac{1}{2\rho}\right)\chi}\right), & \chi > 0, \\ \frac{2^{N_tN_r-1}}{\rho(N_tN_r-1)} (\frac{\rho}{2}+1)^{1-N_tN_r}, & \chi = 0. \end{cases} \quad (56)$$

where $K_\nu(\cdot)$ is the νth order modified Bessel functions of the second kind. One observation is that $p_\chi(\chi)$ is related to N_tN_r and SNR only. Then the cumulative distribution function (CDF) of χ, denoted as $F_\chi(\chi)$, can be expressed as

$$F_\chi(\chi) = \int_0^\chi p_\chi(t) dt. \quad (57)$$

The CDF of both $|tr(\mathbf{V}_l\mathbf{X}[k-1]\mathbf{X}^H[k])|^2$ and $\rho|x|^2$ are shown in Fig. 7 for different pairs of (N_t, N_r) at SNR$= 0$ dB. The statistics of $\rho|x|^2$ fit those of $|tr(\mathbf{V}_l\mathbf{X}[k-1]\mathbf{X}^H[k])|^2$ very well even at a low SNR. Finally, the CDF of the radius $\lambda(l_0)$ is given by

$$F_\lambda(C) = P(\lambda(l_0) < C) = 1 - F_\chi\left(N_t \sum_{i=1}^{N_t} |d_i|^2 - C\right). \quad (58)$$

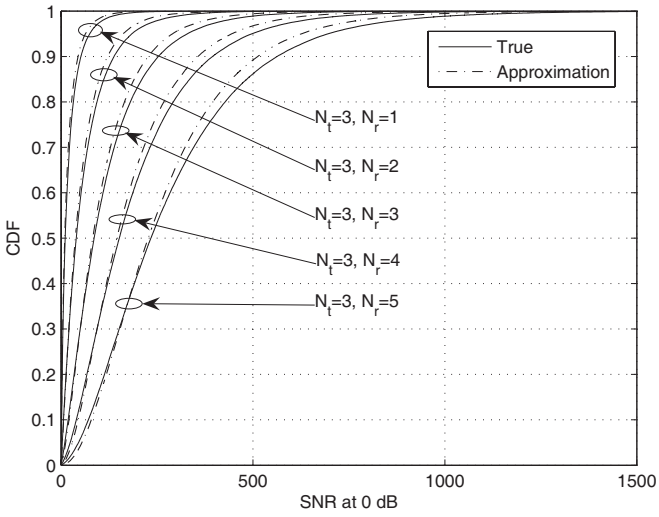


Fig. 7. The CDFs of $\rho|x|^2$ and $|tr(\mathbf{V}_l\mathbf{X}[k-1]\mathbf{X}^H[k])|^2$ at SNR$= 0$ dB.

REFERENCES

[1] N. Al-Dhahir, C. Fragouli, A. Stamoulis, W. Younis, and R. Calderbank, "Space-time processing for broadband wireless access communications," *IEEE Commun. Mag.*, vol. 40, pp. 136-142, Sept. 2002.

[2] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1169-1174, July 2000.

[3] B. M. Hochwald and W. Sweldens, "Differential unitary space time modulation," *IEEE Trans. Commun.*, vol. 48, pp. 2041-2052, Dec. 2000.

[4] B. L. Hughes,"Differential space-time modulation," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2567-2578, Nov. 2000.

[5] A. Shokrollahi, B. Hassibi, B. M. Hochwald, and W. Sweldens, "Representation theory for high-rate multiple antenna code design," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2335-2367, Sept. 2001.

[6] X. B. Liang and X. G. Xia, "Unitary signal constellations for differential space time modulation with two transmit antennas: parametric codes, optimal designs, and bounds," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2291-2322, Aug. 2002.

[7] C. Shan, A. Nallanathan, and P. Y. Kam, "A new class of signal constellations for differential unitary space-time modulation (DUSTM)," *IEEE Commun. Lett.,* vol. 8, pp. 1-3, Jan. 2004.

[8] T. P. Soh, C. S. Ng, and P. Y. Kam, "Improved signal constellations for differential unitary space-time modulations with more than two transmit antennas," *IEEE Commun. Lett.,* vol. 9, pp. 7-9, Jan. 2005.

[9] S. N. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, "Differential space-time coding for frequency-selective channels," *IEEE Commun. Lett.*, vol. 6, pp. 253-255, June 2002.

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

[11] ——, "Differential modulation diversity," *IEEE Trans. Veh. Technol.*, pp. 1431-1444, Nov. 2002.

[12] O. Besson and P. Stoica, "On parameter estimation of MIMO flat-fading channels with frequency offsets," *IEEE Trans. Signal Processing*, vol. 51, pp. 602-613, Mar. 2003.

[13] H. Minn, N. Al-Dhahir, and Y. Li, "Optimal training signals for MIMO OFDM channel estimation in the presence of frequency offset and phase noise," *IEEE Trans. Commun.*, vol. 54, pp. 1754-1759, Oct. 2006.

[14] Z. Liu, G. B. Giannakis, and B. L. Hughes, "Double differential space-time block coding for time selective fading channels," *IEEE Trans. Commun.*, vol. 49, pp. 1529-1539, Sept. 2001.

[15] X. Ma, "Low complexity block double differential design for OFDM with carrier frequency offset," *IEEE Trans. Commun.*, vol. 53, pp. 2129-2138, Dec. 2005.

[16] T. Cui and C. Tellambura, "Bound-intersection detection for multiple-symbol differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 53, pp. 2114-2123, Dec. 2005.

[17] G. L. Stuber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA: Kluwer Academic, 2001.

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, New York: Cambridge University Press, 1985.

[19] H. L. van Tree, *Detection, Estimation, and Modulation Theory*. Wiley Interscience, 2001.
[20] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, pp. 1580-1589, Sept. 2000.
[21] S. Stein, "Unified analysis of certain coherent and noncoherent binary communications systems," *IEEE Trans. Inform. Theory*, pp. 43-51, Jan. 1964.
[22] D. Divsalar and M. K. Simon, "Multiple symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 38, pp. 300-308, Mar. 1990.
[23] M. K. Simon and M. S. Alouini, "Exponential-type bounds on the generalized Marcum Q-function with application to error probability analysis over fading channels," *IEEE Trans. Commun.*, vol. 48, pp. 359-366, Mar. 2000.
[24] E. Biglieri, G. Taricco, and A. Tulino, "Performance of space-time codes for a large number of antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1794-1803, July 2002.
[25] J. Wang, M. P. Fitz, and K. Yao, "Differential unitary space-time modulation for a large number of receive antennas," in *Proc. 36th Asilomar Conf.*, Pacific Grove, USA, Nov. 2002, pp. 565-569.
[26] E. Bach and J. Shallit, *Algorithmic Number Theory, vol. 1: Efficient Algorithms*. Cambridge, MA: MIT Press, 1996.
[27] K. L. Clarkson, W. Sweldens, and A. Zheng, "Fast multiple-antenna differential decoding," *IEEE Trans. Commun.,* vol. 49, pp. 253-261, Feb. 2001.
[28] T. Cui and C. Tellambura, "Multiple-symbol differential detection for single-antenna and multiple-antenna systems over Ricean-fading channels," in *Proc. IEEE ICC'06*, Istanbul, Turkey, June 2006.
[29] I. S. Gradshteny and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. San Diego: Academic Press. 1994.



Feifei Gao (S'05) received the B.Eng. degree in information engineering from Xi'an Jiaotong University, Xi'an, Shaanxi China, in 2002, the M.Sc. degree from the McMaster University, Hamilton, ON, Canada in 2004, and is currently working toward the Ph.D. degree at the Department of Electrical Engineering, National University of Singapore. His research interests are in communication theory, broadband wireless communications, signal processing for communications, MIMO systems, and array signal processing. Mr. Gao was a recipient of the president scholarship from the National University of Singapore.



Tao Cui (S'04) received the M.Sc. degree in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, in 2005, and the M.S. degree from the Department of Electrical Engineering, California Institute of Technology, Pasadena, USA, in 2006. He is currently working toward the Ph.D. degree at the Department of Electrical Engineering, California Institute of Technology, Pasadena. His research interests are in the interactions between networking theory, communication theory, and information theory.



Arumugam Nallanathan (S'97−M'00−SM'05) received the B.Sc. with honors from the University of Peradeniya, Sri-Lanka, in 1991, the CPGS from the Cambridge University, United Kingdom, in 1994 and the Ph.D. from the University of Hong Kong, Hong Kong, in 2000, all in Electrical Engineering. He was an Assistant Professor in the Department of Electrical and Computer Engineering, National University of Singapore, Singapore from August 2000 to December 2007. Currently, he is a Senior Lecturer in the Department of Electronic Engineering at King's College London, United Kingdom. His research interests include cooperative communications, cognitive radio, MIMO-OFDM systems, ultra-wide bandwidth (UWB) communication and localization. In these areas, he has published over 130 journal and conference papers. He is a co-recipient of the Best Paper Award presented at 2007 IEEE International Conference on Ultra-Wideband (ICUWB'2007).

He currently serves on the Editorial Board of IEEE Transactions on Wireless Communications and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY as an Associate Editor. He served as a Guest Editor for EURASIP JOURNAL OF WIRELESS COMMUNICATIONS AND NETWORKING: Special issue on UWB Communication Systems–Technology and Applications. He served as a technical program committee member for more than 30 IEEE international conferences. He also served as the General Track Chair for the IEEE VTC'2008-Spring. He currently serves as the Co-Chair for the IEEE GLOBECOM'2008 Signal Processing for Communications Symposium, and IEEE ICC'2009 Wireless Communications Symposium.



Chintha Tellambura (SM'02) received the B.Sc. degree (with first-class honors) from the University of Moratuwa, Moratuwa, Sri Lanka, in 1986, the M.Sc. degree in electronics from the University of London, London, U.K., in 1988, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 1993.

He was a Postdoctoral Research Fellow with the University of Victoria (1993-1994) and the University of Bradford (1995-1996). He was with Monash University, Melbourne, Australia, from 1997 to 2002. Presently, he is a Professor with the Department of Electrical and Computer Engineering, University of Alberta. His research interests include coding, communication theory, modulation, equalization, and wireless communications.

Prof. Tellambura is an Associate Editor for both the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was Chair of the Communication Theory Symposium in Globecom'05 held in St. Louis, MO.