Differential Modulation for Two-Way Wireless Communications: A Perspective of Differential Network Coding at the Physical Layer

Tao Cui, Feifei Gao, *Member, IEEE*, and Chintha Tellambura, *Senior Member, IEEE*

Abstract—This work considers two-way relay channels (TWRC), where two terminals transmit simultaneously to each other with the help of a relay node. For single antenna systems, we propose several new transmission schemes for both *amplify-and-forward* (AF) protocol and *decode-and-forward* (DF) protocol where the channel state information is not required. These new schemes are the counterpart of the traditional non-coherent detection or differential detection in point-to-point communications. Differential modulation design for TWRC is challenging because the received signal is a mixture of the signals from both source terminals. We derive maximum likelihood (ML) detectors for both AF and DF protocols, where the latter can be considered as performing differential network coding at the physical layer. As the exact ML detector is prohibitively complex, we propose several suboptimal alternatives including decision feedback detectors and prediction-based detectors. All these strategies work well as evidenced by the simulation results. The proposed protocols are especially useful when the required average data rate is high. In addition, we extend the protocols to the multiple-antenna case and provide the design criterion of the differential unitary space time modulation (DUSTM) for TWRC.

Index Terms—Differential modulation, physical layer network coding, two-way relay networks.

I. INTRODUCTION

TWO-WAY communication is a popular type of modern communications, where two source terminals simultaneously communicate. Recently, relay-aided two-way transmissions have attracted a great deal of research interest [1]–[7]. For example, both AF (amplify-and-forward) and DF (decode-and-forward) relaying schemes under one-way relay channels were extended to the two-way relay channels (TWRC) in [1]. In [2], network coding from network layer [8] was extended

to physical layer, but unfortunately, was only effective in an additive white Gaussian noise (AWGN) environment. In [3], a decode and forward scheme was proposed which works for fading channels by using error detection codes at the relay. In [4], a new type of relaying scheme, called *partial decode-and-forward* was designed for TWRC with fading, and the space-time codes that can achieve full spatial diversity were also proposed in the same paper. In [5], [6], the relay function was optimized to attain the minimum error probability at both source terminals, and a new relaying scheme, called *estimate-and-forward* (EF), was developed. The capacity region of TWRC is analyzed in [7]. Most works on TWRC [1], [4]–[6] are based on the assumption of knowledge of channel state information (CSI), which for example can be obtained from the method developed in [9], and the corresponding data recovery relies on coherent detection.

In practice, accurate CSI is hard to obtain in a rapidly changing mobile environment or when multiple transmit antennas are employed, especially in TWRC where two channel coefficients are required to be estimated each way. In these cases, non-coherent schemes, which do not rely on instant CSI, become a preferred choice [10]–[12]. The earliest such scheme is the so-called differential phase-shift keying (DPSK), which has long been used in single-antenna links. The receiver in this case decodes the information by comparing the phase of the current symbol to that of the previous symbol. This differential scheme was extended to the multiple-antenna scenario in [11], which uses matrix signal group modulations. The resulting scheme is called differential unitary space time modulation (DUSTM).

However, the development of such a differential strategy for TWRC is a new and challenging problem as the received signal at one terminal is a mixture of its transmitted signal and the signal from the other terminal. If the self-signal component is known, it can be subtracted from the received signal, and the conventional differential scheme [10]–[12] can then be applied on the residual signal. However, when both channels are unknown, the mixture of the two unknown parts destroys the phase rotation property and prevents the use of the traditional differential scheme.

In this work, we first consider the AF relaying protocol and derive the probability density function (pdf) of the received signal for a single antenna system. Since the pdf cannot be expressed in a simple form, we propose a suboptimal criterion, where the pdf of the received signal conditioned

on the desired information can be exhibited by the modified bessel function. As the maximum likelihood (ML) detector may suffer from high complexity and other implementation issues, we propose a decision feedback scheme that uses three consecutive received signals. Moreover, a prediction-based detector is proposed for the AF protocol over time varying channels. In the conventional DF protocol, the relay first decodes the signals from the two source terminals and re-encodes the information before broadcasting. In this case, the signal from the first terminal, denoted by s_1, and the signal from the second terminal, denoted by s_2, are detected separately. In contrast, we propose to directly detect $s_1 + s_2$ and re-encode $s_1 + s_2$ according to a sophisticatedly designed code book. The new strategy in fact performs no-coherent network coding at the physical layer and is named physical layer differential network coding (PLDNC). Finally, all the proposed schemes are extended to the multiple antennas environment. The design criterion of DUSTM that is applicable to PLDNC is also given through the performance analysis of the pairwise error probability. The simulations show that the proposed strategies are especially useful when the required average data rate is high.

Notations: Vectors and matrices are boldface small and capital letters, respectively; the transpose, complex conjugate, Hermitian, Frobenius norm and inverse of the matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\|\mathbf{A}\|_F$ and $\mathbf{A}^{-1}$, respectively; $\|\mathbf{a}\|_2$ is the l_2 norm of $\mathbf{a}$; $\mathrm{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with the diagonal element constructed from $\mathbf{a}$; $\mathbf{I}_T$ is the $T \times T$ identity matrix; $E\{\cdot\}$ denotes the statistical expectation. $\jmath = \sqrt{-1}$ is the imaginary unit.

II. DIFFERENTIAL MODULATION IN SINGLE ANTENNA SYSTEMS

A. System Model

Consider a network with two source nodes $\mathbb{T}_i$, $i = 1, 2$ and one relay node $\mathbb{R}$, where $\mathbb{T}_i$'s exchange information with the help of $\mathbb{R}$. The half-duplex system is assumed throughout this paper, i.e., each node cannot transmit and receive at the same time. Nonetheless, the results in this paper can be readily generalized to the full duplex systems. The system is operated in time slots. In [4], the 2-, 3-, 4-time slot protocols are proposed. By the "2-time slot protocol", we mean that one time slot is divided into 2 phases and both phases will be indexed by n for the n-th time slot. Similar definitions hold for the 3- and 4-time slot protocols. For the 4-time slot protocols, the communication between $\mathbb{T}_i$'s actually applies the one-way relay channel twice, and the differential modulation in [13] can be directly used. Moreover, the conventional differential modulation can also be used for the 3-time slot protocols, as shown in [14]. This work mainly focuses on the 2-time slot protocol, where the first phase is used for uplink transmission from $\mathbb{T}_i$'s to $\mathbb{R}$ and the second phase is used for downlink transmission from $\mathbb{R}$ to $\mathbb{T}_i$'s. Note that $\mathbb{T}_i$'s are active simultaneously in the first phase, which can provide a higher spectral efficiency [4] than 3- and 4-time slot protocols.

Suppose $\mathbb{T}_i$ wishes to transmit $s_i[n] \in \mathcal{Q}$ in time slot n, where $\mathcal{Q}$ is the signal constellation. Similar to the conventional DPSK modulation, the M-PSK constellation is used,

i.e., $\mathcal{Q} = \{ e^{\jmath \theta} | \theta = \frac{2\pi}{M} l, l = 0, 1, \ldots, M-1 \}$. Suppose that the transmitted signals by $\mathbb{T}_i$ and $\mathbb{R}$ at time slot n are $x_i[n]$ and $x_r[n]$, respectively. The transmitted signal is modulated as $x_i[n] = x_i[n-1] s_i[n]$. The received signals at $\mathbb{R}$ and $\mathbb{T}_i$ are

$$y_r[n] = h_1 x_1[n] + h_2 x_2[n] + w_r[n], \quad y_i[n] = h_i x_r[n] + w_i[n], \quad (1)$$

respectively, where h_i is the channel gain between $\mathbb{T}_i$ and $\mathbb{R}$; $w_r[n]$ and $w_i[n]$ are zero-mean AWGNs at $\mathbb{R}$ and $\mathbb{T}_i$ with variances σ_r^2 and σ_s^2, respectively. In (1), a reciprocal channel is assumed for notational simplicity, and h_i remains static over at least two time slots, unless otherwise stated. The protocols proposed in this paper can also be applied to the case without reciprocal channel.

To realize differential modulation, we do not assume the availability of knowledge of h_1 and h_2 at any node. However, the statistics of h_i's are assumed known, which is fixed as $\mathcal{CN}(0,1)$ in this paper. The differential schemes under the two protocols, AF and DF, are proposed in the following subsections.

B. Amplify-and-Forward

With the AF protocol, the transmit signal from $\mathbb{R}$ is a linear transformation of its received signal, i.e., $x_r[n] = \beta y_r[n]$ where $\beta > 0$ is a constant to keep the average power constraint at $\mathbb{R}$. The received signal at $\mathbb{T}_i$ is

$$y_i[n] = \beta h_i (h_1 x_1[n] + h_2 x_2[n]) + \beta h_i w_r[n] + w_i[n]. \quad (2)$$

Exploiting symmetry, we focus only on $\mathbb{T}_1$, and $\mathbb{T}_2$ can be treated similarly. The signal (2) at $\mathbb{T}_1$ can be written as

$$y_1[n] = \beta h_1 h_2 x_2[n] + \beta h_1^2 x_1[n] + z_1[n], \quad (3)$$

where $z_1[n]$ is zero-mean AWGN with variance $\sigma_z^2 = \beta^2 |h_1|^2 \sigma_r^2 + \sigma_s^2$ conditioned on a deterministic h_1. When h_1^2 is known at $\mathbb{T}_1$, the contribution of $x_1[n]$ on $y_1[n]$ can be eliminated as $\mathbb{T}_1$ knows the signal it has sent. In this case, conventional differential modulation follows naturally on the remaining signal, which resembles the differential modulation in conventional one-way communication. However, the problem becomes complicated when h_1^2 is unknown at $\mathbb{T}_1$ since $\beta h_1^2 x_1[n]$ can no longer be canceled, even though $x_1[n]$ is known. Due to the mixture of the two signal components, the conventional differential detection fails here.

1) Optimal ML Detector: We first consider the optimal ML detector, which should be derived from the pdf of $y_1[n], y_1[n-1]$ conditioned on $s_1[n], s_2[n]$. Here, we slightly modify the system model by requiring $\mathbb{R}$ to transmit $\beta y_r^*[n]$, the conjugate version of $\beta y_r[n]$. Define $\mathbf{y}_1 = [y_1[n], y_1[n-1]]^T$, which can be written in matrix form as

$$\mathbf{y}_1 = \beta h_1 h_2^* \mathbf{X}_2^H \mathbf{1} + \beta |h_1|^2 \mathbf{x}_1^* + \mathbf{z}_1, \quad (4)$$

where $\mathbf{1} = [1, 1]^T$, $\mathbf{x}_1 = [x_1[n-1], x_1[n]]^T$, $\mathbf{x}_2 = [x_2[n-1], x_2[n]]^T$, $\mathbf{X}_2 = \mathrm{diag}\{\mathbf{x}_2\}$ and $\mathbf{z}_1 = [z_1[n-1], z_1[n]]^T$. Clearly, $\mathbf{y}_1$ is complex Gaussian given h_1 and $\mathbf{x}_1, \mathbf{x}_2$; i.e.[1],

$$p(\mathbf{y}_1 | h_1, \mathbf{x}_1, \mathbf{x}_2)$$
$$= \frac{1}{(2\pi)^2 \det(\mathbf{C})} \exp \left(- \left(\mathbf{y}_1 - \beta |h_1|^2 \mathbf{x}_1^* \right)^H \mathbf{C}^{-1} \left(\mathbf{y}_1 - \beta |h_1|^2 \mathbf{x}_1^* \right) \right), \quad (5)$$

[1]This expression is only valid for reciprocal channels. When the channel is not reciprocal, $|h_1|^2$ should be replaced with $h_1^{(1)} h_1^{*(2)}$, where $h_1^{(1)}$ and $h_1^{(2)}$ are uplink and downlink channels, respectively.

$$p(\mathbf{y}_1|h_1,\mathbf{x}_1,\mathbf{x}_2) = \frac{1}{(2\pi)^2\left((\beta^2 t(1+\sigma_r^2)+\sigma_s^2)^2 - (\beta^2 t)^2\right)}\exp\left(-\frac{\left(|y_1[n-1]-\beta t x_1^*[n-1]|^2 + |y_1[n]-\beta t x_1^*[n]|^2\right)\left(\beta^2 t(1+\sigma_r^2)+\sigma_s^2\right)}{(\beta^2 t(1+\sigma_r^2)+\sigma_s^2)^2 - (\beta^2 t)^2}\right.$$
$$\left. + \frac{\beta^2 t(y_1[n-1]-\beta t x_1^*[n-1])s_2[n](y_1^*[n]-\beta t x_1[n])}{(\beta^2 t(1+\sigma_r^2)+\sigma_s^2)^2 - (\beta^2 t)^2} + \frac{\beta^2 t(y_1^*[n-1]-\beta t x_1[n-1])s_2^*[n](y_1[n]-\beta t x_1^*[n])}{(\beta^2 t(1+\sigma_r^2)+\sigma_s^2)^2 - (\beta^2 t)^2}\right), \quad (7)$$

$$p(\mathbf{y}_1|\mathbf{x}_1,\mathbf{x}_2) \approx \frac{1}{2\pi^2\beta^4(2\sigma_r^2+\sigma_r^4)}\exp\left(\frac{B}{\beta(2\sigma_r^2+\sigma_r^4)}\right)\sqrt{\frac{(C+2\sigma_r^2+\sigma_r^4)\beta^2}{A}}K_{-1}\left(2\sqrt{\left(\frac{C}{2\sigma_r^2+\sigma_r^4}+1\right)\frac{A}{\beta^2(2\sigma_r^2+\sigma_r^4)}}\right), \quad (8)$$

$$p(u_1[n]|s_2[n],s_1[n]) = \frac{1}{\pi\beta^2(|s_2[n]-s_1[n]|^2+2\sigma_r^2)}\exp\left(\frac{2\sigma_s^2}{\beta^2(|s_2[n]-s_1[n]|^2+2\sigma_r^2)}\right)K_0\left(2\sqrt{\frac{|u_1[n]|^2}{\beta^2(|s_2[n]-s_1[n]|^2+2\sigma_r^2)}}\right), \quad (12)$$

where

$$\mathbf{C}=\beta^2|h_1|^2\mathbf{X}_2^H\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\mathbf{X}_2+\left(\beta^2|h_1|^2\sigma_r^2+\sigma_s^2\right)\mathbf{I}_2. \quad (6)$$

The pdf can be further expanded as (7) at the top of this page where $t\triangleq|h_1|^2$. To obtain the optimal ML detector, we need to derive $p(\mathbf{y}_1|\mathbf{x}_1,\mathbf{x}_2)$ by integrating $p(\mathbf{y}_1|h_1,\mathbf{x}_1,\mathbf{x}_2)$ over t. Note that the integral of (7) actually depends only on $s_2[n]$. Hence, only a single variable search is needed. Thus, the complexity of the search is greatly reduced. Unfortunately, obtaining a closed-form solution of $p(\mathbf{y}_1|\mathbf{x}_1,\mathbf{x}_2)$ appears intractable, and numerical integration must be used.

We therefore consider the case when $\sigma_s^2\to0$, in which case $p(\mathbf{y}_1|\mathbf{x}_1,\mathbf{x}_2)$ can be approximated by (8) at the top of this page [14], where $K_{-1}(\cdot)$ is the -1th order modified Bessel function of the second kind, and

$$A = \left(1+\sigma_r^2\right)\left(|y_1[n-1]|^2+|y_1[n]|^2\right)$$
$$- y_1[n-1]s_2[n]y_1^*[n]-y_1^*[n-1]s_2^*[n]y_1[n],$$
$$B = \left(1+\sigma_r^2\right)(y_1[n-1]x_1[n-1]+y_1^*[n-1]x_1^*[n-1]+y_1[n]x_1[n]$$
$$+y_1^*[n]x_1^*[n])-y_1[n-1]s_2[n]x_1[n]-y_1^*[n]s_2[n]x_1^*[n-1]$$
$$-y_1^*[n-1]s_2^*[n]x_1^*[n]-y_1[n]s_2^*[n]x_1[n-1], \quad (9)$$
$$C = 2\left(1+\sigma_r^2\right)-x_1^*[n-1]s_2[n]x_1[n]-x_1[n-1]s_2^*[n]x_1^*[n].$$

The approximate ML detector is thus obtained by maximizing (8).

2) Suboptimal ML Detector: Since the closed-form optimal ML detector is not available, we consider a heuristic and suboptimal ML detector. Assuming $x_i[n]=x_i[n-1]s_i[n]$, the contribution of $x_1[n]$ on $y_1[n]$ can be eliminated as

$$u_1[n] = y_1[n]-y_1[n-1]s_1[n]$$
$$= \beta h_1 h_2 x_2[n-1](s_2[n]-s_1[n])+z_1[n]-s_1[n]z_1[n-1]. \quad (10)$$

Given h_1 and $s_2[n]-s_1[n]$, $u_1[n]$ is complex Gaussian with pdf

$$p(u_1[n]|h_1,s_2[n],s_1[n])$$
$$= \frac{1}{2\pi(\beta^2|h_1|^2|s_2[n]-s_1[n]|^2+2\beta^2|h_1|^2\sigma_r^2+2\sigma_s^2)}$$
$$\times\exp\left(-\frac{|u_1[n]|^2}{\beta^2|h_1|^2|s_2[n]-s_1[n]|^2+2\beta^2|h_1|^2\sigma_r^2+2\sigma_s^2}\right). \quad (11)$$

Integrating (11) over the pdf of h_1, we obtain $p(u_1[n]|s_2[n],s_1[n])$ (12) at the top of this page, where equality comes from [15, 3.471], and $K_0(\cdot)$ is the zeroth order modified Bessel function of the second kind. The suboptimal ML detector can then be obtained as

$$\hat{s}_2[n] = \underset{s_2[n]}{\arg\max}\, p(u_1[n]|s_2[n],s_1[n]). \quad (13)$$

This detector has the following interpretation. In the noise-free case, it is easily seen that $\frac{|u_1[n]|^2}{\beta^2|s_2[n]-s_1[n]|^2}=|h_1h_2|^2$. Therefore, the solution of (13) should make $\beta^2|s_2[n]-s_1[n]|^2$ close to $|u_1[n]|^2$. From (13), if $\beta^2|s_2[n]-s_1[n]|^2\ll|u_1[n]|^2$, $s_2[n]$ is unlikely to be the optimal solution as $K_0(x)$ is a decreasing function. On the other hand, when $\beta^2|s_2[n]-s_1[n]|^2\gg|u_1[n]|^2$, $\frac{1}{\pi\beta^2(|s_2[n]-s_1[n]|^2+2\sigma_r^2)}$ penalizes this choice. Therefore, the ML detector forces $\beta^2|s_2[n]-s_1[n]|^2$ to stay around $|u_1[n]|^2$. However, the suboptimal detector requires $|s_2[n]-s_1[n]|^2\neq|s_2'[n]-s_1[n]|^2$, $\forall s_1[n]\in\mathcal{Q}$ and $s_2[n]\neq s_2'[n]$, $\forall s_2[n],s_2'[n]\in\mathcal{Q}$. Clearly, BPSK satisfies this condition. For general M-PSK, the constellation for $s_2[n]$ can be rotated to satisfy this condition.

3) Decision Feedback Detector: Eq. (13) uses only two received signals. By computing $u_1[n],...,u_1[n-K]$, we can derive the joint pdf of $u_1[n],...,u_1[n-K]$ by following (12). If the already detected symbols $s_2[n-1],...,s_2[n-K]$ are substituted into the resulting equation, a decision feedback detector can be obtained. However, this detector requires the computation of $K_0(\cdot)$.

By assuming error-free detection in the previous time slots, we construct

$$d[n] = x_2[n-2](s_2[n-1]-s_1[n-1])u_1[n]$$
$$\quad -x_2[n-1](s_2[n]-s_1[n])u_1[n-1]$$
$$= x_2[n-2](s_2[n-1]-s_1[n-1])(z_1[n]-s_1[n]z_1[n-1])$$
$$\quad -x_2[n-1](s_2[n]-s_1[n])(z_1[n-1]-s_1[n-1]z_1[n-2]), \quad (14)$$

that only contains noise. The pdf of $d[n]$ can be obtained as

$$p(d[n]|s_1[n-1],s_1[n],s_2[n-1],s_2[n]) = \frac{1}{\pi\hat{\sigma}^2}\exp\left(-\frac{|d[n]|^2}{\hat{\sigma}^2}\right), (15)$$

where

$$\hat{\sigma}^2 = \left(|s_2[n-1]-s_1[n-1]|^2+|s_2[n]-s_1[n]|^2\right.$$
$$\left.+|s_2[n]s_2[n-1]-s_1[n]s_1[n-1]|^2\right)\sigma_z^2. \quad (16)$$

The decision feedback detector is used to maximize the pdf (15) over $s_2[n]$. As σ_z^2 depends on $|h_1|^2$, one way to remove such dependence is to integrate (15) over h_1. Another way is to simply replace $|h_1|^2$ with $E\{|h_1|^2\}=1$. Note that $s_2[n]$ also appears in $\hat{\sigma}^2$.

To reduce the complexity, we propose to minimize $|d[n]|^2$ directly since $d[n]$ should be zero in the noise-free environment. This detector requires $s_2[n]-s_1[n]\neq0$, $\forall s_2[n],s_1[n]\in\mathcal{Q}$. Hence, the two source terminals should transmit with different constellations. For example, we could choose $\mathcal{Q}_1=\{-1,1\}$ and $\mathcal{Q}_2=\{-\jmath,\jmath\}$ for $\mathbb{T}_1$ and $\mathbb{T}_2$, respectively. When M-PSK is used, $\mathcal{Q}_2$ can be a rotation of $\mathcal{Q}_1$.

$$E\{|w_e|^2\}=E\left\{\left|\beta h_1^2[n]x_1[n]+w_1[n]-\sum_{k=1}^{K}p_{n,k}\left(\beta h_1^2[n-k]x_1[n-k]+w_1[n-k]\right)\prod_{i=0}^{k-1}s_1[n-i]\right|^2\right\}. \quad (18)$$

$$p(\mathbf{y}_r|s[n])=\frac{1}{(2\pi)^2((2+\sigma_r^2)^2-|s[n]|^2)}\exp\left(-\frac{(1+\sigma_r^2)|y_r[n]|^2+(2+\sigma_r^2-|s[n]^2|)|y_r[n-1]|^2+|y_r[n]-y_r[n-1]s[n]|^2}{(2+\sigma_r^2)^2-|s[n]|^2}\right). \quad (26)$$

To improve the performance and reduce the error propagation due to incorrectly decoded symbols, (15) can be readily extended to multiple symbol detection. Note that $d[n]$ depends on $z_1[n], z_1[n-1], z_1[n-2]$. Thus, $d[n], d[n-1], d[n-2]$ are correlated. For simplicity, we do not give the explicit form of the multiple symbol detector here.

4) Prediction Based Detector: When channel gains h_1 and h_2 vary over time, the contribution of $x_1[n]$ cannot be completely canceled from (10). Moreover, $p(\mathbf{y}_1|\mathbf{x}_1,\mathbf{x}_2)$ is difficult to compute in time-varying channels. Motivated by [10], where the use of the prediction-based decision feedback differential detection is proposed for the point-to-point communication, we develop a similar TWRC differential detector. To this end, instead of subtracting $y_1[n-1]s_1[n]$ from $y_1[n]$, we consider canceling the effect of $x_1[n]$ by using K previously received symbols; i.e.,

$$u_1[n]=y_1[n]-\sum_{k=1}^{K}p_{n,k}y_1[n-k]\prod_{i=0}^{k-1}s_1[n-i]. \quad (17)$$

We need to find $p_{n,k}$ to minimize the expected noise variance and estimation error w_e given in (18) at the top of this page. Hence, $p_{n,k}$ can be determined from the Yule-Walker equations as [16]

$$\mathbf{p}_n=\left(\check{\mathbf{C}}_{h1}+\left(\sigma_r^2+\frac{\sigma_s^2}{\beta^2}\right)\mathbf{I}_K\right)^{-1}\mathbf{b}, \quad (19)$$

where $\check{\mathbf{C}}_{h1}=[E\{h_1^2[n-i](h_1^2[n-j])^*\}]$ and $\mathbf{b}=[E\{h_1^2[n](h_1^2[n-1])^*\},\cdots,E\{h_1^2[n](h_1^2[n-K])^*\}]^T$. With this $\mathbf{p}_n$, we can write $u_1[n]$ in (17) as

$$\begin{aligned}u_1[n]=&w_e+\beta h_1[n]h_2[n]x_2[n]\\&-\beta\sum_{k=1}^{K}p_{n,k}h_1[n-k]h_2[n-k]x_2[n-k]\prod_{i=0}^{k-1}s_1[n-i].\end{aligned} \quad (20)$$

Since the joint distribution of $h_1[n]h_2[n],...,h_1[n-K]h_2[n-K]$ is unknown, we approximate $u_1[n]$ as a complex Gaussian random variable with zero mean and variance

$$\begin{aligned}\sigma_{u1}^2(s_2[n],...,s_2[n-K])&=E\{|u_1[n]|^2\}\\&=\beta^2\check{\mathbf{s}}^H\check{\mathbf{C}}_h\check{\mathbf{s}}+\beta^2(2+\sigma_r^2)+\sigma_s^2-\mathbf{b}^H\left(\check{\mathbf{C}}_{h1}+\left(\sigma_r^2+\frac{\sigma_s^2}{\beta^2}\right)\mathbf{I}_K\right)^{-1}\mathbf{b},\end{aligned} \quad (21)$$

where $\mathbf{s}=[1,s_1[n]s_2^*[n],...,\prod_{i=0}^{K-1}s_1[n-i]s_2^*[n-i]]^T$ and $\check{\mathbf{C}}_h=[E\{h_1[n-i]h_1^*[n-j]\}E\{h_2[n-i]h_2^*[n-j]\}]$. The pdf of $u_1[n]$ conditioned on $s_2[n],...,s_2[n-K]$ can be approximated by

$$\begin{aligned}&p(u_1[n]|s_2[n],...,s_2[n-K])\\&=\frac{1}{2\pi\sigma_{u1}^2(s_2[n],...,s_2[n-K])}\exp\left(-\frac{|u_1[n]|^2}{\sigma_{u1}^2(s_2[n],...,s_2[n-K])}\right).\end{aligned} \quad (22)$$

We assume that $s_2[n-1]...,s_2[n-K]$, which are denoted as $\hat{s}_2[n-1]...,\hat{s}_2[n-K]$, have been detected correctly. The detector for $s_2[n]$ is then obtained by maximizing $p(u_1[n]|s_2[n],\hat{s}_2[n-1],...,\hat{s}_2[n-K])$.

This prediction-based detector can also be extended to multiple symbol detection. Suppose that $u_1[n-1],...,u_1[n-K]$ are obtained similarly as $u_1[n]$ in (17). By assuming that $u_1[n],u_1[n-1],...,u_1[n-K]$ are jointly Gaussian, it is easy to obtain the pdf $p(u_1[n],u_1[n-1],...,u_1[n-K]|s_2[n],...,s_2[n-K])$, maximizing which gives a multiple symbol detector.

C. Decode-and-Forward

1) Single Symbol Detector: The DF protocol requires the relay to decode its received signal. From (1), the received signals at the relay from the n-th and $n-1$-th time slots can be combined as

$$\mathbf{y}_r=h_1\mathbf{X}_1\mathbf{1}+h_2\mathbf{X}_2\mathbf{1}+\mathbf{w}_r, \quad (23)$$

where $\mathbf{y}_r=[y_r[n-1],y_r[n]]^T$, $\mathbf{X}_i=\mathrm{diag}\{x_i[n-1],x_i[n]\}$, $i=1,2$, $\mathbf{1}=[1,1]^T$ and $\mathbf{w}_r=[w_r[n-1],w_r[n]]^T$. Clearly, $\mathbf{y}_r$ is complex Gaussian given $\mathbf{x}_1,\mathbf{x}_2$; i.e.,

$$p(\mathbf{y}_r|\mathbf{x}_1,\mathbf{x}_2)=\frac{1}{\pi^2\det(\mathbf{C})}\exp\left(-\mathbf{y}_r^H\mathbf{C}^{-1}\mathbf{y}_r\right), \quad (24)$$

where

$$\mathbf{C}=\begin{bmatrix}2 & s_1^*[n]+s_2^*[n]\\s_1[n]+s_2[n] & 2\end{bmatrix}+\sigma_r^2\mathbf{I}_2. \quad (25)$$

Let $s[n]=s_1[n]+s_2[n]$. We can rewrite (24) as (26) at the top of this page. As several $s_1[n]$ and $s_2[n]$ pairs may give the same value of $s_1[n]+s_2[n]$, we should use the maximum a posteriori (MAP) detector by maximizing $p(s[n]|\mathbf{y}_r)=\frac{p(\mathbf{y}_r|s[n])p(s[n])}{p(\mathbf{y}_r)}$, and $s[n]$ at the relay can be obtained from

$$\hat{s}[n]=\underset{s[n]=s_1[n]+s_2[n],s_1[n],s_2[n]\in\mathcal{Q}}{\arg\max}p(\mathbf{y}_r|s[n])p(s[n]). \quad (27)$$

After obtaining $\hat{s}[n]=\hat{s}_1[n]+\hat{s}_2[n]$ from (27), we do not require $\mathbb{R}$ to transmit a scaled version of $\hat{s}[n]$ as was the case in the conventional DF protocol. To assist the differential scheme, we will apply an idea similar to network coding [8]. Recall that in the basic network coding, source $\mathbb{T}_i$ transmits $b_i\in\{0,1\}$. The relay decodes b_1 and b_2 separately and broadcasts $b_r=(b_1+b_2)\mod 2$. As $\mathbb{T}_i$ already knows b_i, it can decode the signal from $(b_r-b_i)\mod 2$. In the conventional network coding, $\mathbb{R}$ is required to detect b_1 and b_2 separately at the physical layer and to perform the network coding at the networking layer.

To apply physical layer network coding, we first solve $\hat{s}[n]=\hat{s}_1[n]+\hat{s}_2[n]$ from (27) by searching through different $s_1[n]+s_2[n]=e^{j\frac{2\pi}{M}l_1}+e^{j\frac{2\pi}{M}l_2}$, $l_1,l_2\in\{0,1,...,M-1\}$. We then use a mapping function $\mathcal{M}$ to map $\hat{s}[n]$ to $s_r[n]=\mathcal{M}(\hat{s}[n])$. By differential modulation, the relay then broadcasts $x_r[n]=x_r[n-1]s_r[n]$, where we have assumed the average power constraint of the relay is 0.5 and $x_r[0]=1$ for notational simplicity. As $\mathbb{T}_i$ knows $s_i[n]$, it can decode signals from the other terminal if it receives the correct $s_1[n]+s_2[n]$. Therefore, the mapping function should satisfy the condition $\mathcal{M}(s_1[n]+s_2[n])\neq\mathcal{M}(s_1[n]+s_2'[n])$ and $\mathcal{M}(s_1[n]+s_2[n])\neq$

$$\Pr(y_1[n], y_1[n-1]|s_1[n], s_2[n]) = \frac{1}{4\pi\sigma_s^2} \sum_{\tilde{s}_1[n], \tilde{s}_2[n]} P_e(\{s_1[n], s_2[n]\} \to \{\tilde{s}_1[n], \tilde{s}_2[n]\}) \exp\left(-\frac{|y_1[n] - y_1[n-1]\tilde{s}_r[n]|^2}{2\sigma_s^2}\right),\qquad(32)$$

$\mathcal{M}(s_1'[n] + s_2[n])$, $\forall s_1[n] \neq s_1'[n], s_2[n] \neq s_2'[n]$. We then provide the following lemma.

Lemma 1: For any $p_1, p_2, q_1, q_2 \in \{0, 1, ..., M-1\}$, $e^{j\frac{2\pi}{M}p_1} + e^{j\frac{2\pi}{M}p_2} = e^{j\frac{2\pi}{M}q_1} + e^{j\frac{2\pi}{M}q_2}$ holds if

1) $((p_1 - p_2) \mod M) = \frac{M}{2}$ and $((q_1 - q_2) \mod M) = \frac{M}{2}$ or,

2) $(p_1 \mod M) = (q_1 \mod M)$ and $(p_2 \mod M) = (q_2 \mod M)$ or,

3) $(p_1 \mod M) = (q_2 \mod M)$ and $(p_1 \mod M) = (q_2 \mod M)$.

Proof: From the condition in the lemma, we have $\left|e^{j\frac{2\pi}{M}p_1} + e^{j\frac{2\pi}{M}p_2}\right|^2 = \left|e^{j\frac{2\pi}{M}q_1} + e^{j\frac{2\pi}{M}q_2}\right|^2$ which gives $\cos\left(\frac{2\pi}{M}(p_1 - p_2)\right) = \cos\left(\frac{2\pi}{M}(q_1 - q_2)\right)$, and $\left(e^{j\frac{2\pi}{M}p_1} + e^{j\frac{2\pi}{M}p_2}\right)^2 = \left(e^{j\frac{2\pi}{M}q_1} + e^{j\frac{2\pi}{M}q_2}\right)^2$ which gives

$$e^{j\frac{2\pi}{M}(p_1+p_2)}\left(1 + \cos\left(\frac{2\pi}{M}(p_1 - p_2)\right)\right)$$
$$= e^{j\frac{2\pi}{M}(q_1+q_2)}\left(1 + \cos\left(\frac{2\pi}{M}(q_1 - q_2)\right)\right). \qquad (28)$$

Therefore, $e^{j\frac{2\pi}{M}p_1} + e^{j\frac{2\pi}{M}p_2} = e^{j\frac{2\pi}{M}q_1} + e^{j\frac{2\pi}{M}q_2}$ holds if $1 + \cos\left(\frac{2\pi}{M}(p_1 - p_2)\right) = 0$ and $1 + \cos\left(\frac{2\pi}{M}(q_1 - q_2)\right) = 0$ or $e^{j\frac{2\pi}{M}(p_1+p_2)} = e^{j\frac{2\pi}{M}(q_1+q_2)}$, where the latter condition implies $((p_1 + p_2) \mod M) = ((q_1 + q_2) \mod M)$. Substituting it into the condition in the lemma, we obtain $\cos\left(p_1 - \frac{p_1+p_2}{2}\right) = \cos\left(q_1 - \frac{p_1+p_2}{2}\right)$. The lemma then follows. $\square$

2) Mapping function $\mathcal{M}()$: To find the mapping, we borrow some concepts from graph theory. We first construct a graph $\mathcal{G}$ as in [5], where each node corresponds to $s_1[n] + s_2[n]$ and there exists an edge between $s_1[n] + s_2[n]$ and $s_1[n] + s_2'[n]$ and another edge between $s_1[n] + s_2[n]$ and $s_1'[n] + s_2[n]$, $\forall s_1[n] \neq s_1'[n], s_2[n] \neq s_2'[n]$. Then, the mapping $\mathcal{M}$ corresponds to a coloring of graph $\mathcal{G}$. We assume that M colors are used. When M is odd, there do not exist l_1 and l_2 such that $((l_1 - l_2) \mod M) = \frac{M}{2}$. From Lemma 1, each node in $\mathcal{G}$ corresponds only to $s_1[n] + s_2[n]$ and $s_2[n] + s_1[n]$. Therefore, the degree of each node is $2M - 1$, and each node lies exactly in two cliques, where each clique contains M nodes. We can show that the graph can be colored greedily using M colors. We first pick an arbitrary node and assign a color to this node. Then, the remaining $M-1$ colors are assigned to $M-1$ nodes in the two cliques, respectively. Note that all the nodes in the first clique corresponds to $s_1[n] + s_2'[n]$ while they corresponds to $s_1'[n] + s_2[n]$ in the second clique. There is no edge between nodes in the first clique and nodes in the second clique since $s_1[n] \neq s_1'[n], s_2[n] \neq s_2'[n]$. The greedy coloring does not void the coloring condition. We then apply the same process to each node. We can obtain a valid coloring with M colors in the end. When M is even, all $s_1[n]$ and $s_2[n]$ such that $(l_1 - l_2) \mod M = \frac{M}{2}$ corresponds to the same node, i.e., $s_1[n] + s_2[n]$. We color this node using a color out of the M colors and remove this node from $\mathcal{G}$, which gives a graph $\mathcal{G}'$. In $\mathcal{G}'$, each node belongs to two cliques of size $M-1$. From the same greedy coloring algorithm as when M

is odd, $\mathcal{G}'$ can be colored with $M-1$ colors. Finally, M colors are sufficiently to color $\mathcal{G}$. Therefore, we can obtain a valid mapping $\mathcal{M}$ for any M.

For example, when $\mathcal{Q} = \{-1, 1\}$, we can select

$$\mathcal{M}(\hat{s}[n]) = \begin{cases} 1, & \text{if } \hat{s}[n] = 0, \\ -1, & \text{otherwise}, \end{cases} \qquad (29)$$

The conventional network coding operates on a finite field at the networking layer. In contrast, our new type of coding is designed from the mapping $\mathcal{M}$, which is defined on real signals rather than finite filed. In fact, the mapping $\mathcal{M}$ defines a group with an equivalent additive operation on the indices as

$$l_r = l_1 \oplus l_2. \qquad (30)$$

The signal transmitted by the relay is thus $s_r[n] = e^{j\frac{2\pi}{M}l_r}$. This operation (30) is called physical layer differential network coding. In [3], a DF scheme is proposed for coherent detection by assuming the use of error detection codes at the relay. An XOR type mapping is used at the relay for mapping the decoded symbols to the transmit symbol. Except that we focus on non-coherent systems, error detection codes are not required at the relay. Moreover, we provide a systematic way to design the mapping at the relay rather than to choose it intuitively.

In the downlink, the received signal at $\mathbb{T}_1$ is

$$y_1[n] = h_1 x_r[n] + w_1[n] = y_1[n-1]s_r[n] + z_1[n], \qquad (31)$$

where $z_1[n] = w_1[n] - w_1[n-1]\hat{s}_1[n]\hat{s}_r^*[n]$ is zero mean complex Gaussian with variance $2\sigma_s^2$. Let $P_e(\{s_1[n], s_2[n]\} \to \{\tilde{s}_1[n], \tilde{s}_2[n]\})$ be the pairwise error probability (PEP) when $s_1[n] = e^{j\frac{2\pi}{M}l_1}$ and $s_2[n] = e^{j\frac{2\pi}{M}l_2}$ are transmitted, but $\tilde{s}_1[n] = e^{j\frac{2\pi}{M}\tilde{l}_1}$ and $\tilde{s}_2[n] = e^{j\frac{2\pi}{M}\tilde{l}_2}$ are decoded by $\mathbb{R}$ and $l_1 \oplus l_2 \neq \tilde{l}_1 \oplus \tilde{l}_2$. The probability that $y_1[n]$ and $y_1[n-1]$ are received conditioned on $s_1[n]$ and $s_2[n]$ is (32) given at the top of this page, where $\tilde{s}_r[n] = e^{j\frac{2\pi}{M}\tilde{l}_1 \oplus \tilde{l}_2}$. By using the ML detector (26) at $\mathbb{R}$, the Chernoff bound of PEP can be derived as (49) in Section III-D.

The ML detector for $s_2[n]$ can thus be obtained by maximizing $\Pr(y_1[n], y_1[n-1]|s_1[n], s_2[n])$ given $s_1[n]$. Implementing the true ML detector directly might be complicated. In high signal-to-noise ratio (SNR), (32) is dominated by the term corresponding to the event that $s_1[n] + s_2[n]$ is decoded correctly. Hence, the suboptimal ML detector for $s_r[n]$ can be obtained as [12]

$$\hat{s}_r[n] = \underset{s_r[n] \in \mathcal{Q}}{\arg\min} |y_1[n] - y_1[n-1]s_r[n]|^2. \qquad (33)$$

Let $\hat{s}_r[n] = e^{j\frac{2\pi}{M}\hat{l}_r}$. Then l_2 can be decoded from $\hat{l}_2 = l_r \ominus l_1$, where $\ominus$ follows naturally from (30). The ML detector at $\mathbb{T}_2$ can be obtained similarly.

3) Multiple Symbol Detector: Next, multiple symbol (K symbols) differential detection is considered at both the relay and source terminals. The channel gains are assumed to vary from time slot to time slot. Let $\mathbf{y}_r = [y_r[n], ..., y_r[n-K]]^T$, $\mathbf{h}_i = [h_i[n], ..., h_i[n-K]]^T$, $\mathbf{x}_i = [x_i[n], ..., x_i[n-K]]^T$, $\mathbf{X}_i = \text{diag}\{\mathbf{x}_i\}$, $i = 1, 2$, and $\mathbf{w}_r = [w_r[n], ..., w_r[n-K]]^T$. Then

$$\mathbf{y}_r = \mathbf{X}_1\mathbf{h}_1 + \mathbf{X}_2\mathbf{h}_2 + \mathbf{w}_r. \qquad (34)$$

Note that $\mathbf{y}_r$ is a Gaussian vector with pdf

$$p(\mathbf{y}_r|\mathbf{x}_1,\mathbf{x}_2)=\frac{1}{(2\pi)^{K+1}\det\mathbf{C}}\exp\left(-\mathbf{y}_r^H\mathbf{C}^{-1}\mathbf{y}_r\right), \quad (35)$$

where

$$\begin{aligned}\mathbf{C}&=\mathbf{X}_1\mathbf{C}_{h1}\mathbf{X}_1^H+\mathbf{X}_2\mathbf{C}_{h2}\mathbf{X}_2^H+\sigma_s^2\mathbf{I}_{K+1}\\&=\mathbf{X}_1\left(\mathbf{C}_{h1}+\mathbf{X}_1^H\mathbf{X}_2\mathbf{C}_{h2}\mathbf{X}_2^H\mathbf{X}_1+\sigma_s^2\mathbf{I}_{K+1}\right)\mathbf{X}_1^H,\end{aligned} \quad (36)$$

with $\mathbf{C}_{h1}=E\{\mathbf{h}_1\mathbf{h}_1^H\}$ and $\mathbf{C}_{h2}=E\{\mathbf{h}_2\mathbf{h}_2^H\}$. Let $\mathbf{x}_3=[x_1[n]x_2^*[n], \quad ...,x_1[n-K]x_2^*[n-K]]^T$ and $\mathbf{X}_3=\mathbf{X}_2^H\mathbf{X}_1=\mathrm{diag}\{\mathbf{x}_3\}$. We can rewrite (35) as

$$\begin{aligned}p(\mathbf{y}_r|\mathbf{x}_1,\mathbf{x}_3)&=\frac{1}{(2\pi)^{K+1}\det(\mathbf{C}_{h1}+\mathbf{X}_3^H\mathbf{C}_{h2}\mathbf{X}_3+\sigma_s^2\mathbf{I}_{K+1})}\\&\times\exp\left(-\mathbf{x}_1^T\mathbf{Y}_r^H\left(\mathbf{C}_{h1}+\mathbf{X}_3^H\mathbf{C}_{h2}\mathbf{X}_3+\sigma_s^2\mathbf{I}_{K+1}\right)^{-1}\mathbf{Y}_r\mathbf{x}_1^*\right),\end{aligned} \quad (37)$$

where $\mathbf{Y}_r\triangleq\mathrm{diag}\{\mathbf{y}_r\}$. By maximizing (37), we obtain the ML detector for $\mathbf{x}_1$ and $\mathbf{x}_3$. One remarkable property of (37) is that the detection for $\mathbf{x}_1$ and $\mathbf{x}_3$ is separable. For a given $\mathbf{x}_3$, $\mathbf{x}_1$ can be solved from

$$\hat{\mathbf{x}}_1=\underset{\mathbf{x}_1\in\mathcal{Q}^{K+1}}{\arg\min}\,\mathbf{x}_1^T\mathbf{Y}_r^H\left(\mathbf{C}_{h1}+\mathbf{X}_3^H\mathbf{C}_{h2}\mathbf{X}_3+\sigma_s^2\mathbf{I}_{K+1}\right)^{-1}\mathbf{Y}_r\mathbf{x}_1^*. \quad (38)$$

Note that (38) is an example of the so-called integer least squares programming and can be solved efficiently using the sphere decoder (SD) for M-PSK [17] or the semidefinite programming (SDP) for M-PSK [18]. However, the complexity is still high since the solution of (38) is needed for every possible $\mathbf{x}_3$, which may be infeasible when K is large. To further reduce the complexity, we decouple the detection of $\mathbf{x}_1$ and $\mathbf{x}_3$. We first solve a relaxed problem of (38) by replacing the finite constellation constraint with $\mathbf{x}_1^H\mathbf{x}_1=K+1$, whose minimum value can be easily obtained by using the Rayleigh quotient theory as $(K+1)\lambda_{\min}(\mathbf{A})$, where $\mathbf{A}=\mathbf{Y}_r^H\left(\mathbf{C}_{h1}+\mathbf{X}_3^H\mathbf{C}_{h2}\mathbf{X}_3+\sigma_s^2\mathbf{I}_{K+1}\right)^{-1}\mathbf{Y}_r$ and $\lambda_{\min}(\mathbf{A})$ is the minimum eigenvalue of $\mathbf{A}$. Then $\mathbf{x}_3$ can be obtained from

$$\hat{\mathbf{x}}_3=\underset{\mathbf{x}_3\in\mathcal{Q}^{K+1}}{\arg\min}\,\frac{\exp(-(K+1)\lambda_{\min}(\mathbf{A}))}{(2\pi)^{K+1}\det(\mathbf{C}_{h1}+\mathbf{X}_3^H\mathbf{C}_{h2}\mathbf{X}_3+\sigma_s^2\mathbf{I}_{K+1})}. \quad (39)$$

After solving (39), we substitute $\hat{\mathbf{x}}_3$ into (38) to obtain $\hat{\mathbf{x}}_1$. The transmitted signal is thus $\mathbf{s}_r=\mathcal{M}((\mathbf{I}_{K+1}+\mathrm{diag}\{\hat{\mathbf{x}}_3\})\hat{\mathbf{x}}_1)$, where $\mathcal{M}$ is defined in (29).

By assuming that $s_i[n],...,s_i[n-K]$, $i=1,2$, are decoded correctly at the relay, the multiple symbol differential detection for $\mathbf{s}_r$ at the source terminals can be obtained as in [10], [12]. Finally, $\mathbb{T}_1$ recovers the desired signal via $\hat{l}_2=l_r\ominus l_1$.

Remarks:

- Although the channel gains are assumed identically distributed, both proposed AF and DF strategies can be readily extended to the non identically distributed case. Particularly in DF, the relay can still use a mapping $\mathcal{M}(\hat{s}_1[n],\hat{s}_2[n])$ with slight modification to the mapping in (29) by following the same method used in [5].
- Traditional network coding [8] is performed on a finite field. In this paper, network coding operates on a finite group in the physical layer in DF protocols. Hence, the DF differential protocols can be considered to be differential network coding on the physical layer.

III. DIFFERENTIAL MODULATION IN MULTIPLE ANTENNA SYSTEMS

A. System Model

This section considers differential space time modulation (DSTM) when each node has multiple antennas. For simplicity, both $\mathbb{T}_i$'s are assumed to have N antennas and $\mathbb{R}$ has M antennas. As in Section II, each time slot is partitioned into two phases. Both $\mathbb{T}_i$'s transmit in the first phase while $\mathbb{R}$ broadcasts its received signals in the second phase. Different from Section II, the first and the second phases contain T and K time intervals, respectively.

The input-output relationship at the relay at time slot n can be written as

$$\mathbf{Y}_r[n]=\mathbf{X}_1[n]\mathbf{H}_1[n]+\mathbf{X}_2[n]\mathbf{H}_2[n]+\mathbf{W}_r[n], \quad (40)$$

where $\mathbf{X}_i[n]$ is the transmitted $T\times N$ matrix from $\mathbb{T}_i$; $\mathbf{Y}_r[n]$ is the received $T\times M$ matrix at $\mathbb{R}$; $\mathbf{H}_i[n]$ is the $N\times M$ channel matrix between $\mathbb{T}_i$ and $\mathbb{R}$; the entries of $\mathbf{H}_i[n]$ follow $\mathcal{CN}(0,1)$; and $\mathbf{W}_r[n]$ is the $T\times N$ noise matrix with i.i.d. entries, each following $\mathcal{CN}(0,\sigma_r^2)$, $i=1,2$. Differential modulation is applied to $\mathbf{X}_i[n]$, e.g., $\mathbf{X}_i[n]=\mathbf{S}_i[n]\mathbf{X}_i[n-1],i=1,2$, where $\mathbf{S}_i[n]$ contains the information to be sent from $\mathbb{T}_i$. In this paper, $\mathbf{S}_i[n]$ is assumed to be a finite unitary group $\mathcal{S}=\{\mathbf{S}_l,l=0,...,L-1\}$, where $\mathbf{S}_l$ is a $T\times T$ unitary matrix ($\mathbf{S}_l^H\mathbf{S}_l=\mathbf{I}_N$), and $L=2^{NR}$ with R denoting the data rate. We assume that $T=N$, as in [11]. Similarly, the received signal at $\mathbb{T}_i$ at time slot n is

$$\mathbf{Y}_i[n]=\mathbf{X}_r[n]\mathbf{H}_i^T[n]+\mathbf{W}_i[n], \quad (41)$$

where $\mathbf{X}_r[n]$, $\mathbf{Y}_i[n]$ and $\mathbf{W}_i[n]$ are $K\times M$, $K\times N$ and $K\times N$ matrices, respectively. The definitions of these matrices are similar to those in (40).

We assume that $\mathbf{H}_1[n]$ remains constant during two time slots. The case with time selective channel and multiple symbol detection can be discussed as in Section II.

B. Amplify-and-Forward

The AF relay simply transmits $\mathbf{X}_r[n]=\beta\mathbf{Y}_r[n]$, where β is a scalar to keep the average power constraint at $\mathbb{R}$. When the entries of the channel matrix are non-i.i.d., we may use a precoding matrix to decorrelate the entries of $\mathbf{H}_i[n]$. We have implicitly assumed that $K=T$. All the detectors for DUSTM, e.g., the ML detector, suboptimal ML detector, decision feedback detector, and prediction-based detector, are similar to those in Section II-B for single antenna systems. We omit them here for brevity, but details can be found in [14]. We focus on the decision feedback detector in the following to explain its main idea.

As in (10), the contribution of $\mathbf{X}_1[n]$ in $\mathbf{Y}_1[n]$ is subtracted via

$$\begin{aligned}\mathbf{U}_1[n]&=\mathbf{Y}_1[n]-\mathbf{S}_1[n]\mathbf{Y}_1[n-1]\\&=\beta(\mathbf{S}_2[n]-\mathbf{S}_1[n])\mathbf{X}_2[n-1]\mathbf{H}_2\mathbf{H}_1^T+\mathbf{Z}_1[n]-\mathbf{S}_1[n]\mathbf{Z}_1[n-1].\end{aligned} \quad (42)$$

We can derive $p(\mathbf{U}_1[n]|\mathbf{S}_2[n],\mathbf{S}_1[n])$ as in (12), and the suboptimal ML detector follows.

When $\mathcal{S}$ is a commutative group, i.e., $\mathbf{S}_1\mathbf{S}_2=\mathbf{S}_2\mathbf{S}_1$, $\forall\mathbf{S}_1,\mathbf{S}_2\in\mathcal{S}$, the decision feedback detector can be obtained by maximizing the pdf of

$$\begin{aligned}\mathbf{D}[n]&=\mathbf{X}_2[n-2](\mathbf{S}_2[n-1]-\mathbf{S}_1[n-1])\mathbf{U}_1[n]\\&\quad-\mathbf{X}_2[n-1](\mathbf{S}_2[n]-\mathbf{S}_1[n])\mathbf{U}_1[n-1],\end{aligned} \quad (43)$$

$$
\underbrace{\begin{bmatrix} \mathbf{Y}_r[n-1] \\ \mathbf{Y}_r[n] \end{bmatrix}}_{\breve{\mathbf{Y}}_r} = \underbrace{\begin{bmatrix} \mathbf{X}_1[n-1] & \\ & \mathbf{X}_1[n] \end{bmatrix}}_{\bar{\mathbf{X}}_1} \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_1 \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{X}_2[n-1] & \\ & \mathbf{X}_2[n] \end{bmatrix}}_{\bar{\mathbf{X}}_2} \begin{bmatrix} \mathbf{H}_2 \\ \mathbf{H}_2 \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{W}_r[n-1] \\ \mathbf{W}_r[n] \end{bmatrix}}_{\bar{\mathbf{W}}_r}. \tag{44}
$$

$$
\begin{aligned}
p(\breve{\mathbf{Y}}_r|\mathbf{S}[n]) ={}& \frac{1}{(2M\pi)^{2MN}\det((2+\sigma_r^2)^2\mathbf{I}_N - \mathbf{S}[n]\mathbf{S}^H[n])} \\
&\times \exp\Bigg(-\frac{1}{M}\mathrm{tr}\Bigg(\Big((2+\sigma_r^2)\mathbf{Y}_r^H[n-1]-\mathbf{Y}_r^H[n]\mathbf{S}[n]\Big)\Big((2+\sigma_r^2)^2\mathbf{I}_N-\mathbf{S}^H[n]\mathbf{S}[n]\Big)^{-1}\mathbf{Y}_r[n-1] \\
&\qquad\qquad\quad + \Big((2+\sigma_r^2)\mathbf{Y}_r^H[n]-\mathbf{Y}_r^H[n-1]\mathbf{S}^H[n]\Big)\Big((2+\sigma_r^2)^2\mathbf{I}_N-\mathbf{S}[n]\mathbf{S}^H[n]\Big)^{-1}\mathbf{Y}_r[n]\Bigg)\Bigg).
\end{aligned} \tag{47}
$$

$$
\begin{aligned}
& p(\breve{\mathbf{Y}}_r|\check{\mathbf{S}}[n],\mathbf{S}_2[n]) \\
&= \frac{1}{(2M\pi)^{2MN}\det\Big((2+\sigma_r^2)^2\mathbf{I}_N - \big(\check{\mathbf{S}}[n]+\mathbf{I}_N\big)\big(\check{\mathbf{S}}[n]+\mathbf{I}_N\big)^H\Big)} \\
&\times \exp\Bigg(-\frac{1}{M}\mathrm{tr}\Bigg(\Big((2+\sigma_r^2)\mathbf{Y}_r^H[n-1]\mathbf{S}_2^H[n]-\mathbf{Y}_r^H[n]\big(\check{\mathbf{S}}[n]+\mathbf{I}_N\big)\Big)\Big((2+\sigma_r^2)^2\mathbf{I}_N-\big(\check{\mathbf{S}}[n]+\mathbf{I}_N\big)^H\big(\check{\mathbf{S}}[n]+\mathbf{I}_N\big)\Big)^{-1}\mathbf{S}_2[n]\mathbf{Y}_r[n-1] \\
&\qquad\qquad\quad + \Big((2+\sigma_r^2)\mathbf{Y}_r^H[n]-\mathbf{Y}_r^H[n-1]\mathbf{S}_2^H[n]\big(\check{\mathbf{S}}[n]+\mathbf{I}_N\big)^H\Big)\Big((2+\sigma_r^2)^2\mathbf{I}_N-\big(\check{\mathbf{S}}[n]+\mathbf{I}_N\big)\big(\check{\mathbf{S}}[n]+\mathbf{I}_N\big)^H\Big)^{-1}\mathbf{Y}_r[n]\Bigg)\Bigg).
\end{aligned} \tag{48}
$$

as in Section II-B. For example, when diagonal constellations are used, the unitary matrices $\mathbf{S}_l$'s are chosen as $\mathbf{S}_l = \mathrm{diag}\{e^{j2\pi\mu_1 l/L},\cdots,e^{j2\pi\mu_N l/L}\}$, where μ_i for $i=1,\dots,N$ are optimized to achieve maximum diversity product [11].

C. Decode-and-Forward

The DF relay must decode its received signal. The received signals at the relay in the n-th and $n-1$-th time slots can be written in vector form as (44) at the top of this page. Therefore, $\breve{\mathbf{Y}}_r$ is complex Gaussian given $\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$, and the pdf is

$$
p(\breve{\mathbf{Y}}_r|\bar{\mathbf{X}}_1,\bar{\mathbf{X}}_2) = \frac{1}{(2\pi)^{2MT}\det(\mathbf{C}_Y)}\exp\Big(-\mathrm{tr}\big(\breve{\mathbf{Y}}_r^H\mathbf{C}_Y^{-1}\breve{\mathbf{Y}}_r\big)\Big), \tag{45}
$$

where

$$
\mathbf{C}_Y = M\begin{bmatrix} 2\mathbf{I}_N & \mathbf{S}_1^H[n]+\mathbf{S}_2^H[n] \\ \mathbf{S}_1[n]+\mathbf{S}_2[n] & 2\mathbf{I}_N \end{bmatrix} + M\sigma_r^2\mathbf{I}_{2N}. \tag{46}
$$

Let $\mathbf{S}[n]=\mathbf{S}_1[n]+\mathbf{S}_2[n]$. By using several results of the block matrices [19], we have (47) at the top of this page. The ML detector for $\mathbf{S}[n]$ can be obtained by maximizing (47) over $\mathbf{S}[n]$, which requires searching over all possible $\mathbf{S}_1[n]$ and $\mathbf{S}_2[n]$. To reduce the complexity of searching, we note that $\mathbf{S}[n]=\big(\mathbf{S}_1[n]\mathbf{S}_2^H[n]+\mathbf{I}_N\big)\mathbf{S}_2[n]=\big(\check{\mathbf{S}}[n]+\mathbf{I}_N\big)\mathbf{S}_2[n]$, where $\check{\mathbf{S}}[n]=\mathbf{S}_1[n]\mathbf{S}_2^H[n]$. Then (47) can be rewritten as (48) at the top of this page. Solving $\check{\mathbf{S}}[n]$ and $\mathbf{S}_2[n]$ by maximizing (48) or using an MAP detector as (27) can be separated by solving $\mathbf{S}_2[n]$ for a given $\check{\mathbf{S}}[n]$ and then searching over all $\check{\mathbf{S}}[n]\in\mathcal{S}$. Given $\check{\mathbf{S}}[n]$, (48) can be solved by using the bound intersection detector (BID) in [12] if $\mathcal{S}$ is a diagonal constellation. In high SNR, the separation-based detector with BID reduces the detection complexity from $O(L^2)$ to $O(L)$, where $O(L^2)$ is the complexity by performing exhaustive search over $\mathbf{S}_1[n]$ and $\mathbf{S}_2[n]$ in (47).

In the broadcast phase, as in Section II-C, we consider a mapping $\mathcal{M}$ such that the output signal $\mathbf{V}_r[n]=\mathcal{M}(\hat{\mathbf{S}}[n])$, where $\hat{\mathbf{S}}[n]=\big(\check{\mathbf{S}}[n]+\mathbf{I}_N\big)\mathbf{S}_2[n]$, and $\check{\mathbf{S}}[n]$ and $\mathbf{S}_2[n]$ are obtained from maximizing (48). Since $\mathbb{R}$ has M antennas, and M is different from N, the number of antennas on $\mathbb{T}_i$, $\mathbb{R}$ should not use the same constellation as $\mathbb{T}_i$. Instead, another unitary constellation $\mathcal{V}=\{\mathbf{V}_{l_r}, l_r=0,\dots,L-1\}$, where $\mathbf{V}_{l_r}$ is an $M\times M$ unitary matrix and L is 2^{NR}, should be used. We

have implicitly assumed that $K=M$. The mapping $\mathcal{M}$ may also define a group as in (30). Specifically, when $\mathbf{S}_1[n]=\mathbf{S}_{l_1}$ and $\mathbf{S}_2[n]=\mathbf{S}_{l_2}$, we have $\mathbf{V}_r[n]=\mathbf{V}_{l_r}$, where $l_r=l_1\oplus l_2$, and $\oplus$ is defined by $\mathcal{M}$. For example, when diagonal constellations are used, the mapping (29) still works and can be individually applied to each diagonal entry. The group operation $\oplus$ thus follows that in (30), and $\mathbb{R}$ will send $\mathbf{X}_r[n]=\mathbf{V}_{l_r}\mathbf{X}_r[n-1]$. The conventional DUSTM detector is used at $\mathbb{T}_2$ to decode $\hat{l}_r$, e.g., by using BID [12]. As $\mathbb{T}_2$ knows l_2, it can decode the signal from $\mathbb{T}_1$ from $\hat{l}_r\ominus l_2$. A similar operation takes place at $\mathbb{T}_1$.

D. Constellation Design

The closed-form performance analysis for the AF protocol appears intractable to obtain since the ML detector is related to both $\mathbf{H}_2\mathbf{H}_1^T$ and $K_v(\cdot)$. In the following, we focus on the performance analysis and constellation design for the DF protocol. Nonetheless, our design principle is also applicable to the AF protocol. By using the same strategy in Section III-C, the broadcast phase from $\mathbb{R}$ to $\mathbb{T}_i$'s is found similar to the conventional point-to-point DUSTM operation. Hence, the performance analysis and signal constellation design directly follow those in [11]. If $\mathbb{T}_i$'s have different channel statistics and noise variances, the constellation should be designed for the source terminal that is weaker than the other.

In this subsection, we focus on the first phase of the 2-time slot protocol. We are interested in the pairwise error probability $P_e((l_1,l_2)\to(l_1',l_2'))$; i.e., $\mathbf{S}_1[n]=\mathbf{S}_{l_1}$ and $\mathbf{S}_2[n]=\mathbf{S}_{l_2}$ are transmitted but $\mathbb{R}$ decodes $\mathbf{S}_1'[n]=\mathbf{S}_{l_1'}$ and $\mathbf{S}_2'[n]=\mathbf{S}_{l_2'}$ with $l_1\oplus l_2\neq l_1'\oplus l_2'$. An exact analysis of the PEP for the ML detector is available in [14]. To shed the light on the constellation design, we apply the Chernoff bound for the ML detector [20]:

$$
\begin{aligned}
& P_e((l_1,l_2)\to(l_1',l_2')) \\
&\leq \frac{1}{2}E\Big\{\exp\Big[\lambda\big(\log p(\breve{\mathbf{Y}}_r|\mathbf{S}'[n])-\log p(\breve{\mathbf{Y}}_r)|\mathbf{S}[n]\big)\Big]\Big\} \\
&= \frac{1}{2}E\Big\{\exp\Big[\mathrm{tr}\big(\breve{\mathbf{Y}}_r^H\big(\mathbf{C}_Y^{-1}-(\mathbf{C}_Y')^{-1}\big)\breve{\mathbf{Y}}_r\big)-\log\frac{\det(\mathbf{C}_Y')}{\det(\mathbf{C}_Y)}\Big]\Big\},
\end{aligned} \tag{49}
$$

where $p(\breve{\mathbf{Y}}_r|\mathbf{S}[n])$ is defined in (47) and $\mathbf{S}[n]=\mathbf{S}_{l_1}+\mathbf{S}_{l_2}$, $\mathbf{S}'[n]=\mathbf{S}_{l_1'}+\mathbf{S}_{l_2'}$, and $\mathbf{C}_Y'$ is obtained by substituting $\mathbf{S}'[n]$ into

$$\tilde{\mathbf{D}} = \mathbf{C}_Y \left(\mathbf{C}_Y^{-1} - (\mathbf{C}_Y')^{-1} \right) = \mathbf{I}_{2N} - \mathbf{C}_Y (\mathbf{C}_Y')^{-1}$$

$$= \mathbf{I}_{2N} - \underbrace{\left[\begin{array}{cc} \left((2+\sigma_r^2)^2 \mathbf{I}_N - \mathbf{S}^H \mathbf{S}' \right) \left((2+\sigma_r^2)^2 \mathbf{I}_N - (\mathbf{S}')^H \mathbf{S}' \right)^{-1} & (2+\sigma_r^2)(\mathbf{S} - \mathbf{S}')^H \left((2+\sigma_r^2)^2 \mathbf{I}_N - \mathbf{S}'(\mathbf{S}')^H \right)^{-1} \\ (2+\sigma_r^2)(\mathbf{S} - \mathbf{S}') \left((2+\sigma_r^2)^2 \mathbf{I}_N - (\mathbf{S}')^H \mathbf{S}' \right)^{-1} & \left((2+\sigma_r^2)^2 \mathbf{I}_N - \mathbf{S}(\mathbf{S}')^H \right) \left((2+\sigma_r^2)^2 \mathbf{I}_N - \mathbf{S}'(\mathbf{S}')^H \right)^{-1} \end{array} \right]}_{\boldsymbol{\Phi}}. \tag{51}$$

$$P_e((l_1, l_2) \to (l_1', l_2')) \le \frac{1}{2} E \left\{ \exp\left[\lambda \left(\sum_{i=1}^{2N} (1-\mu_i) t_i - \sum_{i=1}^{2N} \log \mu_i \right) \right] \middle| \mathbf{S}[n] \right\} \le 2^{M-1} \prod_{i=1}^{2N} \frac{1}{\mu_i^{-\frac{1}{2}} (1+\mu_i)^M}, \tag{52}$$

(46). As the entries of $\vec{\mathbf{Y}}_r$ are correlated Gaussian, we first decorrelate $\vec{\mathbf{Y}}_r$ by left multiplying $\mathbf{C}_Y^{-\frac{1}{2}}$; i.e., the entries of $\mathbf{G} = \mathbf{C}_Y^{-\frac{1}{2}} \vec{\mathbf{Y}}_r$ are i.i.d. Gaussian. We can write

$$\text{tr} \left(\vec{\mathbf{Y}}_r^H \left(\mathbf{C}_Y^{-1} - (\mathbf{C}_Y')^{-1} \right) \vec{\mathbf{Y}}_r \right)$$

$$= \text{tr} \left(\mathbf{G}^H \left(\mathbf{C}_Y^{\frac{1}{2}} \right)^H \left(\mathbf{C}_Y^{-1} - (\mathbf{C}_Y')^{-1} \right) \mathbf{C}_Y^{\frac{1}{2}} \mathbf{G} \right) \tag{50}$$

$$= \text{tr} \left(\mathbf{G}^H \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^H \mathbf{G} \right) = \sum_{i=1}^{2N} \lambda_i t_i,$$

where $\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^H$ is the singular value decomposition (SVD) of $\mathbf{D} = \left(\mathbf{C}_Y^{\frac{1}{2}} \right)^H \left(\mathbf{C}_Y^{-1} - (\mathbf{C}_Y')^{-1} \right) \mathbf{C}_Y^{\frac{1}{2}}$, λ_i is the i-th singular value of $\mathbf{D}$, and t_i is a chi-square random variable with $2M$ degrees of freedom. Note that t_i's are independent from each other. As the polynomial $\det(\mathbf{AB} - \lambda\mathbf{I})$ is exactly the same as $\det(\mathbf{BA} - \lambda\mathbf{I})$ for any non-singular $\mathbf{A}$ [19], the non-zero singular values of $\mathbf{D}$ are the same as those of (51) at the top of this page. Let the singular value of $\boldsymbol{\Phi}$ be μ_i, $i = 1, \dots, 2N$. We can write (49) as (52) at the top of this page, where we choose $\lambda = \frac{1}{2}$ in the last inequality, and t_i is defined in (50). Hence, the modulation design criterion for the two-way DUSTM is to maximize the modified diversity product

$$\xi = \min_{\substack{0 \le l_1 < l_1' \le L-1, \\ 0 \le l_2 < l_2' \le L-1, \\ l_1 \oplus l_2 \ne l_1' \oplus l_2'}} \prod_{i=1}^{2N} \mu_i^{\frac{1}{2}} (1+\mu_i)^M. \tag{53}$$

To simplify the design, we could choose σ_r^2 to be small in $\boldsymbol{\Phi}$. To optimize the performance, $\mathbb{T}_1$ and $\mathbb{T}_2$ may use different constellations. Nevertheless, (53) is still the criterion to find these constellations. One simple choice is that the constellation of one source terminal is a phase rotation of that of the other one; i.e., $\mathbf{S}_{l_1} = \boldsymbol{\Theta}\mathbf{S}_{l_2}$, where $\mathbf{S}_{l_1} \in \mathcal{S}_1, \mathbf{S}_{l_2} \in \mathcal{S}_2$, and $\mathcal{S}_i$ is $\mathbb{T}_i$'s constellation and $\boldsymbol{\Theta}$ is a rotation matrix.

Remarks:

- The AF protocol uses $2N$ time intervals and can achieve a diversity order at most $\min\{N^2, MN\}$. The DF protocol, on the other hand, consumes $M+N$ time intervals but attains a diversity order at most MN. Actually, DF does not necessarily need M time slots in the downlink. It can use $\min\{N, M\}$ time intervals to trade diversity for throughput.

- We do not discuss multiple symbol detection for DUSTM, which can be readily derived by following the approach in Section II. The search required by multiple symbol detection can be solved by using the variants of BID as in [12].

- Only a single relay network is considered. The proposed schemes can be readily extended to the multiple-relay case by following [4], [13].

- We have considered two-way relay networks without direct communication between two source terminals. This case may occur when the source terminals are far away from each other. When the source terminals are close to each other, direct transmission can occur. If we still partition each time slot into two phases, the source terminal cannot benefit from the direct communication link. We could partition each time slot into 3 phases. In the first phase, only $\mathbb{T}_1$ transmits, while $\mathbb{T}_2$ transmits in the second phase. In the third phase, $\mathbb{R}$ broadcasts to both $\mathbb{T}_i$'s after processing its received signals in the first two phases. This situation corresponds to the 3 time-slot protocol in [4]. More discussion of this aspect can be found in [14].

IV. SIMULATION RESULTS

In this section, we compare the performance of the proposed differential decoding strategies with that by applying the conventional differential decoding twice over one-way relay protocol. The proposed strategies are denoted as 2-AF and 2-DF, respectively, while the conventional strategies are denoted as 4-AF and 4-DF since each time slot is partitioned into 4 phases. The transmission power of each node is fixed at 1. In addition, we set $\sigma_r^2 = \sigma_s^2$. The SNR is defined as $1/\sigma_r^2$. In the following, the symbol error rate (SER) is used as the figure of merit.

A. Single Antenna Systems

Fig. 1 compares the SER of 2-AF strategies with that of 4-AF strategies. The detector (13) is denoted as "Suboptimal ML". Decision feedback using (16) and using $|d[n]|^2$ only with perfectly decoded previous symbols are denoted as "Genie Decision Feedback" and "Genie Suboptimal Decision Feedback", respectively, while decision feedback using $|d[n]|^2$ only and decoded previous symbols is denoted as "Non-Genie Decision Feedback". In the suboptimal ML case, BPSK is used at both terminals, while $\{-1, 1\}$ and $\{-\jmath, \jmath\}$ are used at each terminal in the case of the decision feedback detectors. Also included is 2-AF with perfect h_i^2 (i.e, perfect CSI) such that the self-interference signal can be canceled and conventional differential decoding can be used. If each symbol is assumed to take the same time, 4-AF consumes twice as much time as 2-AF. Thus, we compare 2-AF using BPSK with 4-AF using QPSK, where both strategies have the same
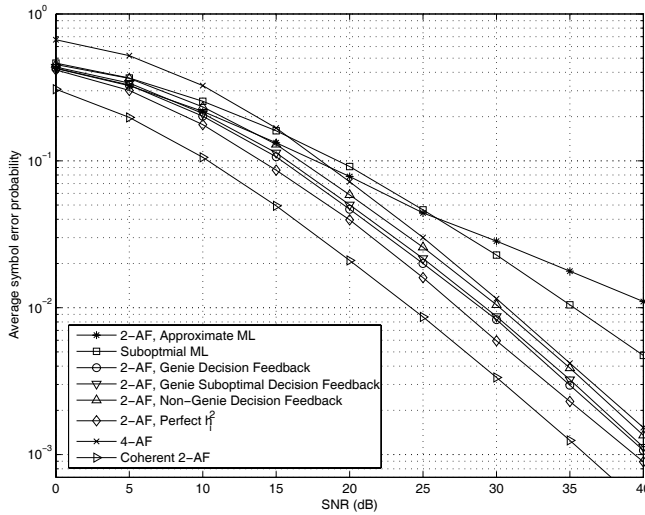
Fig. 1. Average symbol error rate comparison between different 2-AF strategies using BPSK and 4-AF protocol using QPSK in a single antenna system.
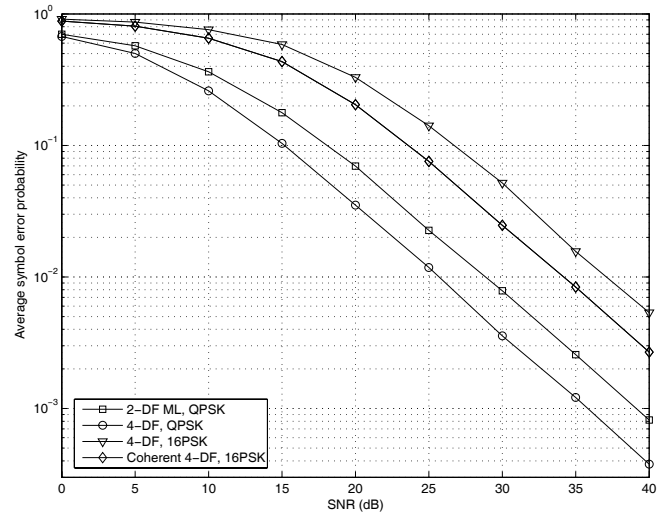

Fig. 3. Average symbol error rate comparison between 2-DF strategy using QPSK and 4-DF protocol using QPSK and 16PSK in a single antenna system.
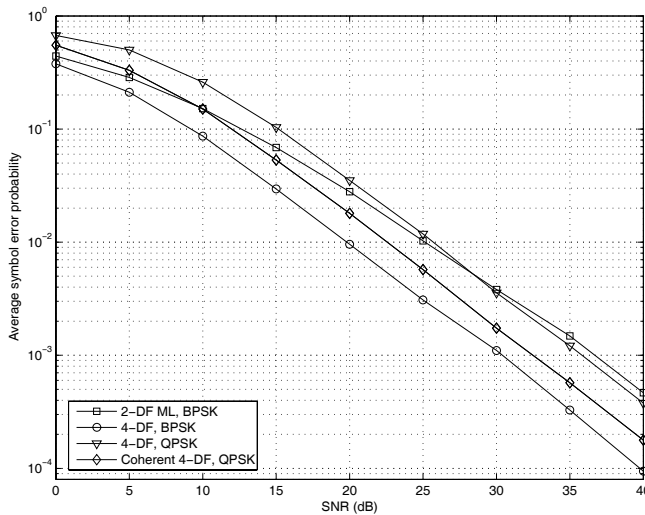

Fig. 2. Average symbol error rate comparison between 2-DF strategy using BPSK and 4-DF protocol using BPSK and QPSK in a single antenna system.
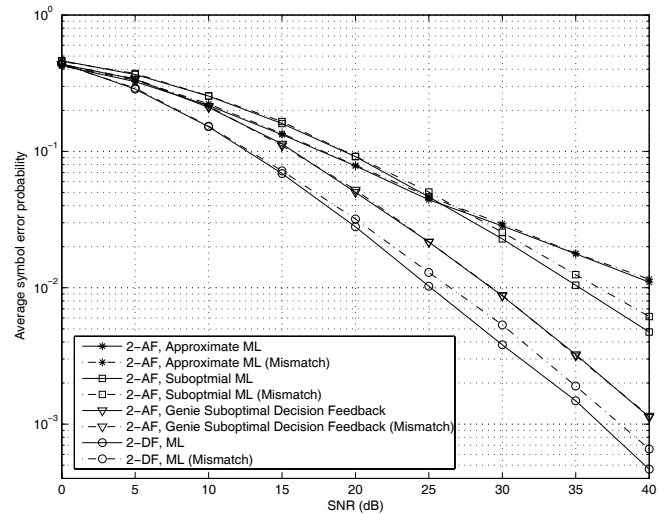

Fig. 4. Average symbol error rate comparison of different strategies when there is noise variance mismatch. 10 symbols are used to estimate the noise variance.

average transmission rate. The approximation (8) is poor as the performance of the resulting detector is inferior to that of other detectors in high SNR. The suboptimal ML detector performs worse than decision feedback detectors because 2 received signals are used in it while 3 received signals are used in decision feedback detectors. The decision feedback detectors perform close to 2-AF with perfect h_i^2 even though h_i^2 is unknown in the former case. With the genie-aided detection, we observe that compared with decision feedback using (16) the performance degradation by using $|d[n]|^2$ only is small. This finding suggests that using $|d[n]|^2$ is a good choice in practice due to its simplicity. Without the genie, an additional 1-dB loss occurs at SER=10^{-2} by using $|d[n]|^2$. At SER=10^{-2}, non-genie decision feedback has a 0.5-dB gain over 4-AF, which shows the advantage of 2-AF. We also include the performance of coherent 2-AF which assume perfect CSI at the terminals. Coherent 2-AF has about 3-dB gain over 2-AF with perfect h_i^2 at SER=10^{-2}.

Fig. 2 compares the SER of 2-DF with that of 4-DF. We also include the performance of coherent 4-DF. Both the terminals in 2-DF use BPSK. We compare 2-DF with 4-DF by using BPSK and QPSK. 2-DF BPSK and 4-DF QPSK have the same average data rate. We find that 2-DF BPSK performs better than 4-DF QPSK in low SNR, but performs worse in high SNR. However, when the average data rate increases, 2-DF always performs better than 4-DF as shown in Fig. 3. In 2-DF, one terminal uses $\{e^{j\frac{\pi}{2}i}, i=0,1,2,3\}$ and the other terminal uses $\{e^{j(\frac{\pi}{2}i+\frac{\pi}{4})}, i=0,1,2,3\}$. We maximize (37) by choosing $K=4$ and choose $s_i[n-1],...,s_i[n-K]$ to be previously decoded symbols (37). 2-DF using QPSK performs better than 4-DF using 16QPSK because the spectral efficiency of M-PSK reduces as M increases. Non-coherent 2-DF using QPSK even performs better than coherent 4-DF using 16QPSK. Hence, 2-DF is preferable when the required data rate is high.

Complexity comparison of proposed algorithms for each

TABLE I
COMPLEXITY COMPARISON OF PROPOSED ALGORITHMS FOR EACH EVALUATION OF THE DECODING METRIC.

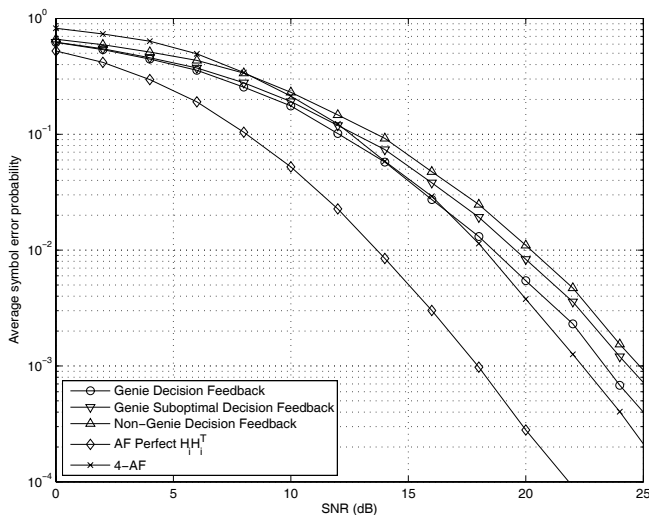| Algorithm	Real Multiplication	Real Addition	Evaluation of Bessel function
Approximate ML	28	18	1
Suboptimal ML	9	2	1
Genie Decision Feedback	41	36	0
Genie Suboptimal Decision Feedback	28	23	0
DF	22	12	0

Fig. 5. Average symbol error rate comparison between 2-AF protocols and 4-AF protocol in a multiple antenna system with $N=M=2$.
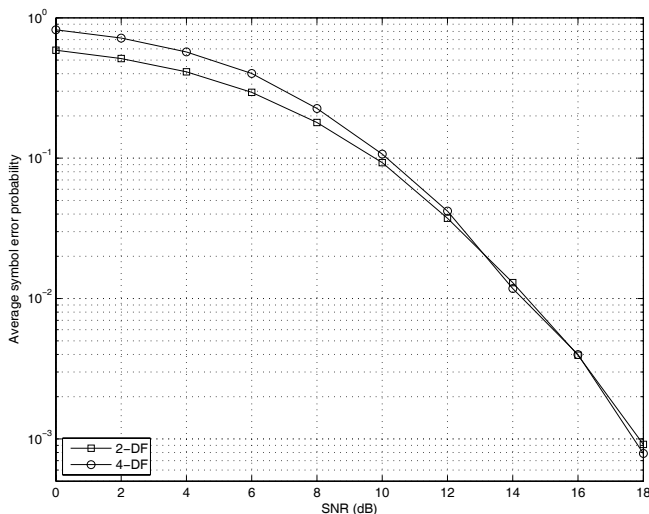

Fig. 6. Average symbol error rate comparison between 2-DF strategy and 4-DF protocol in a multiple antenna system with $N=M=2$.

evaluation of the decoding metric is given in Table I. Considering the performance, we find that Genie Suboptimal Decision Feedback and DF give a good tradeoff between performance and complexity.

Fig. 4 shows the average SER comparison of the proposed strategies when there is noise variance mismatch. 10 noise samples are used to estimate the noise variance. "Mismatch" denotes the performance by using the estimated noise variance. We find that 2-AF Approximate ML and 2-AF Genie Suboptimal Decision Feedback are robust to noise variance

mismatch, while 2-AF Suboptimal ML and 2-DF ML incur about 1.5-dB loss at high SNR. We note that noise variance could be estimated accurately in practice, either data aided or non-data aided. There are lots of references on noise variance estimation, see e.g., [21].

B. Multiple Antenna Systems

In this subsection, we consider multiple antenna systems. In all simulations, we choose $M=N=2$ and $R=1$. One terminal chooses the diagonal constellation diag$\{e^{j\frac{\pi}{2}l}, e^{j\frac{\pi}{2}l}\}$, $l=0,...,3$ and the other constellation uses the diagonal constellation diag$\{e^{j(\frac{\pi}{2}l+\frac{\pi}{4})}, e^{j(\frac{\pi}{2}l+\frac{3\pi}{4})}\}$, $l=0,...,3$.

Fig. 5 compares the decision feedback detectors with the conventional detector having perfect knowledge of $\mathbf{H}_i\mathbf{H}_i^T$ and 4-AF with $L=16$. The observations are similar to those in Fig. 1. The decision feedback detectors perform better than 4-AF in low SNR while 4-AF has a better performance in high SNR. The detector with perfect $\mathbf{H}_i\mathbf{H}_i^T$ has a large gain over both 2-AF with decision feedback detectors and 4-AF. This finding suggests that in multiple antenna systems, it may be preferable to estimate $\mathbf{H}_i\mathbf{H}_i^T$ directly by using pilots. Similar observations are obtained in Fig. 6, which compares 2-DF with 4-DF. 2-DF performs better than 4-DF in low SNR. Thus, the proposed protocols are useful in low SNR when multiple antennas are used.

V. CONCLUSION

This paper considered non-coherent transmission in TWRC. Differential AF and DF strategies were proposed for both single-antenna and multiple-antenna systems. These new schemes are the counterparts of the traditional non-coherent detection or the differential detection in point-to-point communications. We derived ML detectors for both AF and DF protocols. The DF protocol could be considered as performing differential network coding at the physical layer. To reduce the complexity of the ML detector, several suboptimal alternatives were proposed including decision feedback detectors and prediction based detectors. Moreover, the protocols were extended to the multiple-antenna case, and a DUSTM design criterion was derived. We found that the proposed protocols are especially useful when the required average data rate is high. Future work could investigate several interesting topics, including the detailed performance analysis of the proposed strategies, especially of the AF protocols and the synchronization errors.

REFERENCES

[1] B. Rankov and A. Wittneben, "Spectral efficient signaling for half-duplex relay channels," in *Proc. Asilomar Conf. Signals, Systems Computers*, Oct. 2005, pp. 1066-1071.

[2] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: physical-layer network coding," in *Proc. ACM Mobicom*, 2006, pp. 358-365.

[3] P. Popovski and H. Yomo, "The anti-packets can increase the achievable throughput of a wireless multi-hop network," in *Proc. IEEE ICC*, June 2006, pp. 3885-3890.

[4] T. Cui, F. Gao, T. Ho, and A. Nallanathan, "Distributed space-time coding for two-way wireless relay networks," in *Proc. IEEE ICC*, May 2008.

[5] T. Cui, T. Ho, and J. Kliewer, "Some results on relay strategies for memoryless two-way relay channels," in *Proc. Inform. Theory Applications Workshop*, Jan. 2008.

[6] ——, "Relay strategies for memoryless two-way relay channels: performance analysis and optimization," in *Proc. IEEE ICC*, May 2008.

[7] W. Nam, S.-Y. Chung, and Y. H. Lee, "Capacity bounds for two-way relay channels," in *Proc. Int. Zurich Seminar Commun.*, Mar. 2008.

[8] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204-1216, July 2000.

[9] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," to appear in *IEEE Trans. Commun.*

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

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

[12] T. Cui and C. Tellambura, "Bound intersection detection for multiple-symbol differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 53, no. 12, pp. 2114-2123, Dec. 2005.

[13] Y. Jing and H. Jafarkhani, "Distributed differential space-time coding for wireless relay networks," in *Proc. Inform. Theory Applications Workshop*, Jan. 2007.

[14] T. Cui, F. Gao, and C. Tellambura, "Differential modulation for two-way wireless communications: differential network coding at physical layer," Caltech, Tech. Rep., Mar. 2008.

[15] I. Gradshteyn, I. Ryzhik, and A. Jeffrey, *Table of Integrals, Series, and Products*, 5th ed. Academic Press, 1994.

[16] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, 2001.

[17] T. Cui and C. Tellambura, "Joint data detection and channel estimation for OFDM systems," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 670-679, Apr. 2006.

[18] W.-K. Ma, P.-C. Ching, and Z. Ding, "Semidefinite relaxation based multiuser detection for M-ary PSK multiuser systems," *IEEE Trans. Signal Processing*, vol. 52, no. 10, pp. 2862-2872, Oct. 2004.

[19] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[20] H. L. V. Trees, *Detection, Estimation, and Modulation Theory-Part I*. Wiley, 1968.

[21] T. Cui and C. Tellambura, "Power delay profile and noise variance estimation for OFDM," *IEEE Commun. Lett.*, vol. 10, no. 1, pp. 25-27, Jan. 2006.
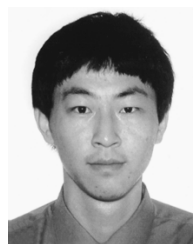
Tao Cui (S'04) received the M.Sc. degree in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, in 2005, and the M.S. degree from the Department of Electrical Engineering, California Institute of Technology, Pasadena, USA, in 2006. He is currently working 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.

Feifei Gao (S'05-M'09) 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 the Ph.D degree from National University of Singapore in 2007. He was a Research Fellow at Institute for Infocomm Research, A*STAR, Singapore in 2008. He joined the School of Engineering and Science at Jacobs University, Bremen, Germany in 2009, where he is currently an Assistant Professor. 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. He has co-authored more than 50 refereed IEEE journal and conference papers and has served as a TPC member for IEEE ICC, IEEE GLOBECOM, IEEE VTC, and IEEE PIMRC.

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 Diversity and Fading Countermeasures, Multiple-Input Multiple-Output (MIMO) Systems and Space-Time Coding, and Orthogonal Frequency Division Multiplexing (OFDM).

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