Minimum Error-Rate Linear Dispersion Codes for Cooperative Relays

Kuo-ching Liang, Xiaodong Wang, *Senior Member, IEEE*, and Inaki Berenguer

Abstract—Cooperative diversity systems have recently been proposed as a solution to provide spatial diversity for terminals where multiple antennas are not feasible to be implemented. As in multiple-input–multiple-output systems, space–time codes can be used to efficiently exploit the increase in capacity provided in cooperative diversity systems. In this paper, we propose a two-layer linear dispersion (LD) code for cooperative diversity systems and derive a simulation-based optimization algorithm to optimize the LD code and power allocation in terms of block error rate. The proposed code design paradigm can obtain optimal codes under arbitrary fading statistics. Performance comparisons are made to other cooperative diversity schemes. The effect that distances between source, relays, and destination terminals have on the energy allocation between the broadcast and cooperative intervals is also studied.

Index Terms—Cooperative diversity, gradient estimation, linear dispersion (LD) codes, multiple-input–multiple-output (MIMO), stochastic approximation.

I. INTRODUCTION

IN RECENT years, multiple-input–multiple-output (MIMO) systems have been extensively studied because of their promise of enormous capacity gain [1], [2]. Cooperative diversity has been proposed for systems with multiple single-antenna devices as an attempt to realize the spatial diversity gain similar to that of a MIMO system [3]. The mobile terminals share their antennas with other users in the network to create a virtual antenna array and provide spatial diversity for transmission. Fig. 1 shows a typical cooperative diversity system setup. The capacity of wireless relay channels has been analyzed in [4], and relays have also been shown to extend the coverage area and system performance. In [5] and [6], it is also shown that for a two-user cooperative diversity system, the achievable capacity region of the system is increased for a flat block-fading channel and that the users' achievable rates are also less susceptible to channel variations.

Due to the similarity between cooperative diversity and MIMO systems, space–time codes have been proposed as a possible solution. These codes utilize both the spatial and time domains to introduce correlation between signals transmitted

from the different antennas at different time slots [7]. An important class of space–time code is the space–time block codes (STBCs). The Alamouti code [8] is one example of an STBC for a MIMO system with two transmit antennas. In a cooperative diversity system, space–time coding can also be used to take advantage of its MIMO-like properties to obtain spatial diversity, coding gains, and higher spectral efficiency. In [9]–[11], performance analyses are made for various cooperative schemes in a flat fading channel using the Alamouti code, and it is shown that systems using the Alamouti code outperform the systems that employ simple repetition coding.

To retransmit the received signal to the destination, relays can choose from two relay schemes: amplify-and-forward (AF) and decode-and-forward (DF) [12]. Using an AF relay scheme, the relays generate the space–time codewords using the received signal and transmit the space–time codewords to the destination at a predetermined energy level. When using a DF relay scheme, the received signal is first decoded, and then, the decoded symbols are used to generate the space–time codewords to be transmitted. It is easy to see that a cooperative diversity system is different from a MIMO system in one important aspect: All transmitting antennas except those located at the source have imperfect knowledge of the information to be transmitted. The imperfection of the source-to-relay channels is a source of errors for the cooperative diversity system.

In this paper, we propose a new coding scheme for AF cooperative diversity systems based on linear dispersion (LD) codes [13]–[15]. Using LD codes, the energy of the transmitted symbols is spread out along the temporal and spatial dimensions. In [15], an LD coding scheme is proposed for the cooperative diversity system. The LD dispersion matrices are applied to the received signal at the relays, where the source and relays act as a multiantenna system and transmit the resulting LD code matrix. Diversity gain is then analyzed for pairwise error probability. However, no scheme is proposed to construct the code to optimize the error probability.

In our proposed scheme, instead of using the LD code only in the relay-to-destination transmissions, we take a two-layered approach to design an LD code for cooperative diversity systems. The first layer of the LD code is generated at the source by Q r-quadratic-amplitude modulation (QAM) symbols with the first set of dispersion vectors. The second layer of LD code is then generated by the cooperative nodes with the second set of dispersion matrices using the received signals. The power allocated to each node and communication link is also an important design parameter in the performance of the cooperative diversity system [16]. For cooperative diversity systems using LD code, an explicit analytical expression for the block

K.-c. Liang and X. Wang are with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA.

I. Berenguer was with the Laboratory for Communication Engineering, University of Cambridge, CB3 0FD Cambridge, U.K. He is now with McKinsey & Company, 28004 Madrid, Spain.
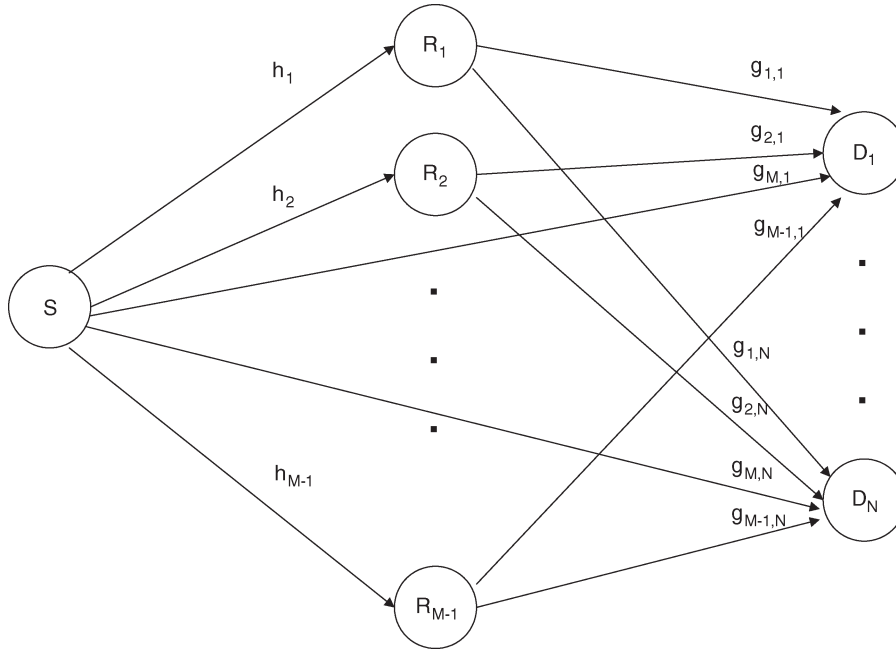
Fig. 1. *(M,N)* cooperative diversity system.

error rate (BLER) does not exist, and therefore, deterministic optimization techniques cannot be used. Here, we employ a simulation-based stochastic approximation technique together with gradient estimation [17] to jointly optimize the LD code and power allocation with respect to the BLER. This method is useful when the objective function and the gradient cannot be evaluated analytically but can be estimated.

The remainder of this paper is organized as follows. In Section II, we describe the system model for cooperative diversity systems using a two-layered LD code. In Section III, we employ simulation-based stochastic approximation and gradient estimation techniques to optimize the LD code and power allocation with respect to the BLER. In Section IV, we present simulation results to compare the performance of the optimized LD code and power allocation with other space–time code designs. Finally, Section IV contains the conclusions.

II. SYSTEM MODEL

Unlike MIMO systems, where the transmit antennas all are part of a single antenna array and each transmit antenna has perfect knowledge of the information to be transmitted, transmit antennas in a cooperative diversity system can belong to several independent terminals, where only the source has perfect knowledge of the symbols to be transmitted. To be able to use the cooperative nodes as relays to transmit information jointly to the destination, the information first needs to be distributed to the cooperative nodes.

Consider the cochannel transmission, cooperative network illustrated in Fig. 1 consisting of one source terminal, $(M-1)$ relay terminals, and one destination terminal, which are denoted as S, R_m, $m = 1, \ldots, M-1$, and D, respectively. The source has one transmit antenna, each relay has one antenna, and the destination has N receive antennas, with the nth receive antenna denoted as $D_n, n = 1, \ldots, N$. The system thus consists

of M transmit antennas and N receive antennas, which we will denote as an (M, N) system. The source-to-relay, relay-to-destination, and source-to-destination channels are assumed to be mutually independent block-fading channels with arbitrary fading statistics and are denoted as h_m, $g_{m,n}$, and $g_{M,n}$, respectively, where, in the transmissions to the destination, we designate the source as the Mth node. In our model, the source does not need knowledge of the channels, and we assume that the destination has knowledge of all the channels h_m, $g_{m,n}$, and $g_{M,n}$, $m = 1, \ldots, M-1$, $n = 1, \ldots, N$, while relay m has knowledge of h_m. The destination's knowledge of the source-to-relay channels can be obtained from the corresponding relays through some control channel, which we assume to be perfect.

In our proposed scheme, each transmission frame consists of two intervals:

1) broadcast interval: source broadcasts information using first-layer LD code to cooperating relays and destination;
2) cooperation interval: source and relays transmit with second-layer LD code to the destination.

Unlike for MIMO systems, a code design for cooperative diversity system needs to take into consideration the broadcast channels between the source and relays. Moreover, the energy allocated in the broadcast interval to transmit the first-layer LD code to the relays has significant impact on system performance. The total energy consumption E_0 is fixed for each transmission frame, and then, energy allocation to each interval has an important effect on the BLER performance of the system. The optimal distribution of energy between the two intervals depends on the space–time code, the statistics of the channels between source, relays, and destination, and the physical distances between the terminals. If insufficient energy is allocated to the broadcast interval, the first-layer LD code received by the relays during the broadcast interval can become so corrupted that the performance of the overall system

is degraded, despite having more energy allocated to the source-to-destination and relay-to-destination links in the cooperation interval. On the other hand, if too much energy is assigned to the broadcast interval for reliable transmission of the first-layer LD code to the relays, the lack of energy during the cooperation interval can still degrade the overall system performance. It is then clear that energy allocation between the two intervals should also be an optimization parameter.

A. Transmission Scheme

Let us define $s_1, s_2, \ldots, s_Q$ as the Q different r-QAM symbols that the source terminal wishes to transmit to the destination terminal, where $s_q = \alpha_q + j\beta_q$, $q = 1, \ldots, Q$, and $\mathbb{E}\{|s_q^2|\} = 1$. For a cooperative diversity system with one source antenna, $M - 1$ relays, and N destination antennas, we can construct the following cooperative transmission scheme.

1) Source forms τ linearly combined symbols $\boldsymbol{k} = [k_1, \ldots, k_\tau]^T$ using Q r-QAM symbols, $s_1, \ldots, s_Q$, and the first-layer dispersion $\tau \times 1$ vectors $\boldsymbol{c}_q$, $\boldsymbol{d}_q$, $q = 1, \ldots, Q$.
2) Source transmits $k_1, \ldots, k_\tau$ to the relays and destination during τ consecutive symbol intervals.
3) Source and relays form the second-layer LD codeword using their received signals corresponding to $k_1, \ldots, k_\tau$ and with dispersion matrices $\boldsymbol{A}_t$, $\boldsymbol{B}_t$, $t = 1, \ldots, \tau$, where $\boldsymbol{A}_t$, $\boldsymbol{B}_t$ have dimensions $(T - \tau) \times M$, and each relay uses one column of the dispersion matrices.

In the broadcast phase, the vectors $\boldsymbol{c}_q$ and $\boldsymbol{d}_q$ disperse the qth symbol along the time dimension. In the cooperation phase, the signal received by the relays at time t is dispersed by $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ along both the time and spatial dimensions. Notice that τ, which is the number of linearly combined symbols transmitted during the broadcast interval, is also the length of the broadcast interval. Since the total length of the frame is fixed at T, the choice of τ will determine the size of both broadcast and cooperation intervals. By choosing $\tau > T/2$, we are devoting more resources to ensure that information received by the cooperative terminals is less error prone. Intuitively, this is done when the source sees poorer channels to the relays compared to the relay-to-destination channels. Conversely, we can make $\tau < T/2$ when the relays see poorer channels to the destination compared to the source-to-relay channels. Thus, it is clear that τ is also a design variable that needs to be optimized.

B. Energy Constraints

The energy allocated to the broadcast interval E_1 and co-operation interval E_2 is constrained by the fixed total energy constraint for one transmission frame E_0, where

$$E_0 = E_1 + E_2, \qquad E_1 > 0, \quad E_2 > 0 \qquad (1)$$

and we can write (1) as a function of an angular coordinate

$$E_1 = E_0 \cos^2 \alpha, \quad E_2 = E_0 \sin^2 \alpha, \qquad \alpha \in (0, \pi). \quad (2)$$

For an (M, N) cooperative diversity system, let us denote the normalized distance between source and relay R_m as d_{SR_m},

$m = 1, \ldots, M - 1$, the normalized distance between source and destination as d_{SD}, and the normalized distance between relay R_m and destination as $d_{R_m D}$. Denote $\rho_{\mathrm{SD},1}, \rho_{\mathrm{SD},2}, \rho_{\mathrm{SR}_m}$, and $\rho_{R_m D}$ as SNR between the source and destination during the broadcast and cooperation phase, SNR between source and relay R_m, and SNR between relay R_m and destination, respectively. By assuming that the received noise has unit variance and incorporating path loss into our model, it follows that

$$\rho_{\mathrm{SD},1} = \frac{E_1}{\tau} \left(\frac{1}{d_{\mathrm{SD}}} \right)^\nu$$

$$\rho_{\mathrm{SR}_m} = \frac{E_1}{\tau} \left(\frac{1}{d_{\mathrm{SR}_m}} \right)^\nu$$

$$\rho_{\mathrm{SD},2} = \frac{1}{M} \frac{E_2}{T - \tau} \left(\frac{1}{d_{\mathrm{SD}}} \right)^\nu$$

$$\rho_{R_m D} = \frac{1}{M} \frac{E_2}{T - \tau} \left(\frac{1}{d_{R_m D}} \right)^\nu \qquad (3)$$

where ν is the path loss exponent. The factor of $1/M$ divides the energy allocated in the cooperation interval evenly among the M transmitting terminals. In the following analysis, unless stated otherwise, we assume that $\nu = 4$ for an urban environment [18].

C. Broadcast Interval

In this and the following section, in the interest of clarity in presenting the system model, we will present the details of the matrices which involves the channel coefficients and the dispersion matrices to Appendix A. This allows us to give a clearer presentation of the transmission scheme proposed in this paper.

For a given τ, let us denote the $\tau \times 1$ complex-valued LD vectors for the source-to-relay transmission as $\boldsymbol{c}_q = [c_{1q}, \ldots, c_{\tau q}]^T$ and $\boldsymbol{d}_q = [d_{1q}, \ldots, d_{\tau q}]^T, q = 1, \ldots, Q$. Recall that $s_q = \alpha_q + j\beta_q$; the $\tau \times 1$ linearly combined symbol vector to be transmitted is then

$$\boldsymbol{k} = \sum_{q=1}^Q (\alpha_q \boldsymbol{c}_q + j\beta_q \boldsymbol{d}_q), \qquad q = 1, \ldots, Q. \qquad (4)$$

The energy constraint is $\mathbb{E}\{\boldsymbol{k}^H \boldsymbol{k}\} \leq \tau$. Using the fact that α_q, β_q are independent identically distributed (i.i.d.) with zero mean and variance 1/2, we get the following constraint

$$\sum_{q=1}^Q \left(\boldsymbol{c}_q^H \boldsymbol{c}_q + \boldsymbol{d}_q^H \boldsymbol{d}_q \right) \leq 2\tau. \qquad (5)$$

Denote h_m, $m = 1, \ldots, M - 1$ as the fading channel coefficients for the source-to-relay channels. Let $\boldsymbol{r}_{R_m}$ be the $\tau \times 1$ received signal vector at relay R_m after passing through a matched filter and normalizing by $|h_m|$. The received signal at relay R_m is then given by

$$\boldsymbol{r}_{R_m} = |h_m| \sqrt{\rho_{\mathrm{SR}_m}} \boldsymbol{k} + \boldsymbol{n}_{R_m}$$

$$\text{with} \quad \boldsymbol{n}_{R_m} \sim \mathcal{N}_\mathcal{C}(\boldsymbol{0}, \boldsymbol{I}), \qquad m = 1, \ldots, M - 1. \quad (6)$$

In the broadcast interval, the destination antenna D_n also receives the transmission from the source, and the $\tau \times 1$ received signal vector at antenna D_n is given as

$$\boldsymbol{r}_{D_n,B} = g_{M,n}\sqrt{\rho_{\mathrm{SD},1}}\boldsymbol{k} + \boldsymbol{n}_{D_n}$$

$$\text{with} \quad \boldsymbol{n}_{D_n} \sim \mathcal{N}_{\mathcal{C}}(\boldsymbol{0}, \boldsymbol{I}), \qquad n = 1, \dots, N. \quad (7)$$

Denoting $k_t = \tilde{\alpha}_t + j\tilde{\beta}_t, t = 1, \dots, \tau$, then from (4), we have the following relationship:

$$\begin{bmatrix} \tilde{\alpha}_1 \\ \vdots \\ \tilde{\alpha}_\tau \\ \tilde{\beta}_1 \\ \vdots \\ \tilde{\beta}_\tau \end{bmatrix}_{2\tau \times 1} = \boldsymbol{\Phi} \underbrace{\begin{bmatrix} \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_Q \\ \beta_Q \end{bmatrix}_{2Q \times 1}}_{\boldsymbol{x}} \quad (8)$$

where $\boldsymbol{\Phi}$ is a $2\tau \times 2Q$ matrix of elements of the first-layer dispersion matrices. Details for the construction of $\boldsymbol{\Phi}$ are given in Appendix A. We further denote $\boldsymbol{y}_1$ as the real-valued received signal vector during the broadcast interval at the relays, i.e.,

$$\boldsymbol{y}_1 \triangleq \left[\Re\{\boldsymbol{r}_{R_1}^T\} \; \Im\{\boldsymbol{r}_{R_1}^T\} \cdots \Re\{\boldsymbol{r}_{R_{M-1}}^T\} \; \Im\{\boldsymbol{r}_{R_{M-1}}^T\} \right]_{2(M-1)\tau \times 1}^T. \quad (9)$$

Using (6) and (8), the received signal $\boldsymbol{y}_1$ at the relays during the broadcast interval can now be written as

$$\boldsymbol{y}_1 = \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n}, \qquad \boldsymbol{n} \sim \mathcal{N}\left(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}\right) \quad (10)$$

with the real-valued $2(M-1)\tau \times 2\tau$ equivalent channel matrix $\boldsymbol{H}$, composed of the source-to-relay channel coefficients, given in Appendix A.

D. Cooperation Interval

In the cooperation interval, source and relays construct a new LD codeword using the received first-layer LD codeword generated during the broadcast interval. We first need to normalize the energy of the received signal during the broadcast interval. From (6), we have

$$\mathbb{E}\left\{\boldsymbol{r}_{R_m}^H \boldsymbol{r}_{R_m}\right\} = |h_m|^2 \rho_{\mathrm{SR}_m}\tau + \tau. \quad (11)$$

Thus, before performing the linear combination on the received signals at the relays using the second-layer dispersion matrices, we multiply by the normalization constant

$$\gamma_{R_m} \triangleq \sqrt{\frac{\tau}{|h_m|^2 \rho_{\mathrm{SR}_m}\tau + \tau}}, \qquad m = 1, \dots, M-1 \quad (12)$$

such that

$$\gamma_{R_m}^2 \, \mathbb{E}\left\{\boldsymbol{r}_{R_m}^H \boldsymbol{r}_{R_m}\right\} = \tau. \quad (13)$$

For the cooperation interval, we use a set of dispersion matrices $\{\boldsymbol{A}_t, \boldsymbol{B}_t\}_{t=1}^{\tau}$, with dimension $(T - \tau) \times M$. From (13), the τ received symbols to be linearly combined have total energy of τ; thus, we normalize the second-layer dispersion matrices as

$$\sum_{t=1}^{\tau} \mathrm{tr}\left(\boldsymbol{A}_t^H \boldsymbol{A}_t + \boldsymbol{B}_t^H \boldsymbol{B}_t\right) \leq 2M(T - \tau). \quad (14)$$

For transmission in the cooperation interval, the source employs the first column of the LD matrices and transmit the following $(T - \tau) \times 1$ signal vector:

$$\boldsymbol{x}_S = \sum_{t=1}^{\tau} \Re\{k_t\}\boldsymbol{a}_{M,t} + j\Im\{k_t\}\boldsymbol{b}_{M,t}. \quad (15)$$

The relay R_m will use the (m)th column of the LD matrices and transmit the following $(T - \tau) \times 1$ signal vector:

$$\boldsymbol{x}_{R_m} = \gamma_{R_m} \sum_{t=1}^{\tau} \left(\Re\{r_{R_m,t}\}\boldsymbol{a}_{m,t} + j\Im\{r_{R_m,t}\}\boldsymbol{b}_{m,t}\right)$$

$$m = 1, \dots, M-1 \quad (16)$$

where $\boldsymbol{a}_{m,t}$ and $\boldsymbol{b}_{m,t}$ are the mth column of the dispersion matrices $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$, respectively. Denote $g_{m,n}$, $m = 1, \dots, M-1$, $n = 1, \dots, N$, and $g_{M,n}$ as the relay-to-destination and source-to-destination channel coefficients. Then, the $(T - \tau) \times 1$ received signal vector at the destination antenna D_n during the cooperation interval is

$$\boldsymbol{r}_{D_n,C} = g_{M,n}\sqrt{\rho_{\mathrm{SD},2}}\boldsymbol{x}_S + \sum_{m=1}^{M-1} g_{m,n}\sqrt{\rho_{R_m D_n}}\boldsymbol{x}_{R_m} + \boldsymbol{v}_{D_n}$$

$$\boldsymbol{v}_{D_n} \sim \mathcal{N}_{\mathcal{C}}(\boldsymbol{0}, \boldsymbol{I}). \quad (17)$$

Define $\boldsymbol{y}_2$ as the real-valued received signal vector at the destination in both broadcast and cooperation intervals as (18), shown at the bottom of the page. From (7), (9), and (17), we can write $\boldsymbol{y}_2$ as

$$\boldsymbol{y}_2 = \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} + \widetilde{\boldsymbol{G}}\boldsymbol{y}_1 + \boldsymbol{u}, \qquad \boldsymbol{u} \sim \mathcal{N}\left(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}\right) \quad (19)$$

where $\boldsymbol{G}$ is the $2NT \times 2\tau$ real-valued equivalent channel matrix for the received signal component at the destination from

$$\boldsymbol{y}_2 \triangleq \left[\Re\{\boldsymbol{r}_{D_1,B}^T\} \; \Im\{\boldsymbol{r}_{D_1,B}^T\} \; \cdots \; \Re\{\boldsymbol{r}_{D_N,B}^T\} \; \Im\{\boldsymbol{r}_{D_N,B}^T\} \; \Re\{\boldsymbol{r}_{D_1,C}^T\} \; \Im\{\boldsymbol{r}_{D_1,C}^T\} \; \cdots \; \Re\{\boldsymbol{r}_{D_N,C}^T\} \; \Im\{\boldsymbol{r}_{D_N,C}^T\} \right]_{2NT \times 1}^T \quad (18)$$

the source during the broadcast and cooperation intervals, and $\widetilde{\boldsymbol{G}}$ is the $2NT \times 2(M-1)\tau$ real-valued equivalent channel matrix for the received signal component at the destination from the relays during the cooperation interval. The channel matrix $\boldsymbol{G}$ consists of the second-layer dispersion matrices and the source-to-destination channel coefficients, whereas $\widetilde{\boldsymbol{G}}$ consists of the second-layer dispersion matrices and the relay-to-destination channel coefficients. The details of the equivalent channel matrices are given in Appendix A.

In order to perform data detection at the destination, we need to write $\boldsymbol{y}_2$ in terms of $\boldsymbol{x}$, which is the real-valued transmitted symbol vector. Substituting (10) into (19), we have

$$\boldsymbol{y}_2 = \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} + \widetilde{\boldsymbol{G}}(\boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n}) + \boldsymbol{u}$$
$$= (\boldsymbol{G} + \widetilde{\boldsymbol{G}}\boldsymbol{H})\boldsymbol{\Phi}\boldsymbol{x} + (\widetilde{\boldsymbol{G}}\boldsymbol{n} + \boldsymbol{u}). \tag{20}$$

In (20), the effective total noise is colored, owing to the noise amplification and recombination at the relays, with covariance $\boldsymbol{\Sigma} = (1/2)(\widetilde{\boldsymbol{G}}\widetilde{\boldsymbol{G}}^T + \boldsymbol{I})$. In order to perform detection, we need to whiten the noise first. That is, define

$$\boldsymbol{z} \triangleq \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{y}_2 = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{G} + \widetilde{\boldsymbol{G}}\boldsymbol{H})\boldsymbol{\Phi}\boldsymbol{x} + \bar{\boldsymbol{u}}, \qquad \bar{\boldsymbol{u}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}). \tag{21}$$

We can then employ the sphere decoder [19]–[21] to perform maximum likelihood (ML) detection on $\boldsymbol{z}$ to obtain $\hat{\boldsymbol{x}}$, which is the ML estimate of $\boldsymbol{x}$.

III. OPTIMIZATION OF TWO-LAYER COOPERATIVE LD CODE

A. Stochastic Approximation and Gradient Estimation

In this section, we develop an algorithm to find the two-layered dispersion matrices and the energy allocation to minimize the BLER for a cooperative diversity system in an arbitrary fading scenario. Since the exact analytical expression of the average BLER for an arbitrary set of dispersion matrices and arbitrary fading statistics does not exist, we have to resort to stochastic gradient algorithms to optimize the average BLER performance with respect to the dispersion matrices and the energy allocation. Here, we employ an optimization scheme based on the Robbins–Monro (R–M) [22] algorithm and use the score-function algorithm for gradient estimation. The R–M algorithm takes the recursive form

$$\varpi_{i+1} = \varpi_i - \sigma_i \widehat{\nabla \Upsilon}(\varpi_i) \tag{22}$$

where ϖ_i is the estimated parameter value at iteration i, $\widehat{\nabla \Upsilon}(\varpi_i)$ is the gradient estimate of the objective function at ϖ_k, and $\{\sigma_i\}$ is a decreasing step size sequence of positive numbers such that

$$\sum_{i=1}^{\infty} \sigma_i = \infty, \quad \sum_{i=1}^{\infty} \sigma_i^2 < \infty. \tag{23}$$

By choosing $\sigma_i = \sigma/i$, where σ is a positive scalar, the above stochastic gradient algorithm will converge to global optimum in a weak sense.

Consider an (M, N) cooperative diversity system with Q symbols to be transmitted, broadcast interval length τ, and total transmission interval of T. Let us define the real-valued channel vectors corresponding, respectively, to channels from source to relay, and channels from relay to destination and source to destination

$$\boldsymbol{\phi} \triangleq [\, \Re\{h_1\} \quad \Im\{h_1\} \quad \cdots \quad \Re\{h_{M-1}\} \quad \Im\{h_{M-1}\} \,]^T$$
$$\boldsymbol{\psi} \triangleq [\, \Re\{g_{1,1}\} \quad \Im\{g_{1,1}\} \quad \cdots \quad \Re\{g_{M,N}\} \quad \Im\{g_{M,N}\} \,]^T. \tag{24}$$

For the cooperative LD code design problem, the optimization parameter set $\boldsymbol{\theta}$ consists of the broadcast dispersion vectors, the cooperative dispersion matrices, and the angular coordinate defining the energy allocation, i.e.,

$$\boldsymbol{\theta} \triangleq \{\{\boldsymbol{c}_q, \boldsymbol{d}_q, q = 1, \dots, Q\}, \quad \{\boldsymbol{A}_t, \boldsymbol{B}_t, t = 1, \dots, \tau\}, \quad \alpha\} \tag{25}$$

with constraints (5) and (14). Define the empirical BLER as $\gamma(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta})$ for the given sets of noise-whitened receive signal vector $\boldsymbol{z}$, information symbol vector $\boldsymbol{x}$, channel realizations $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$, and the given parameter $\boldsymbol{\theta}$. The empirical BLER is then given by an indicator function

$$\gamma(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) \triangleq \mathbb{I}(\hat{\boldsymbol{x}} \neq \boldsymbol{x} | \boldsymbol{z}, \boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) \tag{26}$$

where $\hat{\boldsymbol{x}}$ denotes the decoded symbol vector. Recalling from (21) that $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}_2$, we can thus write the empirical BLER as a function of $\boldsymbol{y}_2$. For given $\boldsymbol{\theta}$, the average BLER is then

$$\Upsilon(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\boldsymbol{x}}\mathbb{E}_{\boldsymbol{\phi}}\mathbb{E}_{\boldsymbol{\psi}}\mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{y}_2 | \boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}}\{\gamma(\boldsymbol{y}_2, \boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta})\}. \tag{27}$$

We want to solve the following optimization problem:

$$\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \Upsilon(\boldsymbol{\theta}) \tag{28}$$

with the constraint set given by

$$\boldsymbol{\Theta} \triangleq \left\{ \sum_{q=1}^{Q} \mathrm{tr}\left(\boldsymbol{c}_q^H \boldsymbol{c}_q + \boldsymbol{d}_q^H \boldsymbol{d}_q\right) \leq 2\tau, \right.$$
$$\left. \sum_{t=1}^{\tau} \mathrm{tr}\left(\boldsymbol{A}_t^H \boldsymbol{A}_t + \boldsymbol{B}_t^H \boldsymbol{B}_t\right) \leq 2M(T-\tau) \right\}. \tag{29}$$

In (27), we have

$$\mathbb{E}_{\boldsymbol{y}_1, \boldsymbol{y}_2 | \boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}}\{\gamma(\boldsymbol{y}_2, \boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta})\}$$
$$= \iint \gamma(\boldsymbol{y}_2, \boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) p(\boldsymbol{y}_1, \boldsymbol{y}_2 | \boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) d\boldsymbol{y}_1 d\boldsymbol{y}_2$$
$$= \iint \gamma(\boldsymbol{y}_2, \boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) p(\boldsymbol{y}_1 | \boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\theta})$$
$$\times p(\boldsymbol{y}_2 | \boldsymbol{y}_1, \boldsymbol{x}, \boldsymbol{\psi}, \boldsymbol{\theta}) d\boldsymbol{y}_1 d\boldsymbol{y}_2. \tag{30}$$

From (10) and (19), it follows that $p(\boldsymbol{y}_1|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\theta})$ and $p(\boldsymbol{y}_2|\boldsymbol{y}_1\boldsymbol{x},\boldsymbol{\psi},\boldsymbol{\theta})$ are both white Gaussian probability density function. Let us denote

$$\mathcal{P}_1 \overset{\triangle}{=} p(\boldsymbol{y}_1|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\theta})$$

$$= \frac{1}{\pi^{(M-1)\tau}} \exp\left[-(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})^T(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})\right] \qquad (31)$$

$$\mathcal{P}_2 \overset{\triangle}{=} p(\boldsymbol{y}_2|\boldsymbol{y}_1,\boldsymbol{x},\boldsymbol{\psi},\boldsymbol{\theta})$$

$$= \frac{1}{\pi^{\mathrm{NT}}} \exp\left[-(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)\right].$$
$$(32)$$

Using (27), the gradient of the average BLER with respect to $\boldsymbol{\theta}$ is then given by

$$\nabla_{\boldsymbol{\theta}}\Upsilon(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x}}\mathbb{E}_{\boldsymbol{\phi}}\mathbb{E}_{\boldsymbol{\psi}}\iint[\nabla_{\boldsymbol{\theta}}\gamma(\boldsymbol{y}_2,\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})]\mathcal{P}_1\mathcal{P}_2$$

$$+ \gamma(\boldsymbol{y}_2,\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})\left[(\nabla_{\boldsymbol{\theta}}\mathcal{P}_1)\mathcal{P}_2 + \mathcal{P}_1(\nabla_{\boldsymbol{\theta}}\mathcal{P}_2)\right]d\boldsymbol{y}_1 d\boldsymbol{y}_2. \quad (33)$$

For ML detection, it is shown in Appendix C that

$$\mathbb{E}_{\boldsymbol{x}}\iint \nabla_{\boldsymbol{\theta}}\gamma(\boldsymbol{y}_2,\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})\mathcal{P}_1\mathcal{P}_2 d\boldsymbol{y}_1 d\boldsymbol{y}_2 = 0. \qquad (34)$$

Then, (33) can be written as

$$\nabla_{\boldsymbol{\theta}}\Upsilon(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x}}\mathbb{E}_{\boldsymbol{\phi}}\mathbb{E}_{\boldsymbol{\psi}}\iint \gamma(\boldsymbol{y}_2,\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})$$

$$\times \left[(\nabla_{\boldsymbol{\theta}}\mathcal{P}_1)\mathcal{P}_2 + \mathcal{P}_1(\nabla_{\boldsymbol{\theta}}\mathcal{P}_2)\right]d\boldsymbol{y}_1 d\boldsymbol{y}_2$$

$$= \mathbb{E}_{\boldsymbol{x}}\mathbb{E}_{\boldsymbol{\phi}}\mathbb{E}_{\boldsymbol{\psi}}\iint \gamma(\boldsymbol{y}_2,\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})$$

$$\times \nabla_{\boldsymbol{\theta}}p(\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})d\boldsymbol{y}_1 d\boldsymbol{y}_2$$

$$= \mathbb{E}_{\boldsymbol{x}}\mathbb{E}_{\boldsymbol{\phi}}\mathbb{E}_{\boldsymbol{\psi}}\iint \gamma(\boldsymbol{y}_2,\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})$$

$$\times \frac{\nabla_{\boldsymbol{\theta}}p(\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})}{p(\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})}$$

$$\times p(\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})d\boldsymbol{y}_1 d\boldsymbol{y}_2$$

$$= \mathbb{E}_{\boldsymbol{x}}\mathbb{E}_{\boldsymbol{\phi}}\mathbb{E}_{\boldsymbol{\psi}}\mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi}}\{\gamma(\boldsymbol{y}_2,\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})\nabla_{\boldsymbol{\theta}}$$

$$\times \log p(\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})\}$$

$$= \mathbb{E}_{\boldsymbol{x}}\mathbb{E}_{\boldsymbol{\phi}}\mathbb{E}_{\boldsymbol{\psi}}\mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi}}\{\gamma(\boldsymbol{y}_2,\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})\nabla_{\boldsymbol{\theta}}$$

$$\times [\log\mathcal{P}_1 + \log\mathcal{P}_2]\}. \quad (35)$$

B. Simulation-Based LD Code Optimization Algorithm

We now present the iterative simulation-based algorithm to optimize the dispersion matrices and the energy allocation. The optimal value for the design variable τ is chosen by evaluating the following algorithm at different values of τ and selecting the one which gives the lowest BLER.

For a given τ in the kth iteration, let $\boldsymbol{\theta}_k$ be the set

$$\boldsymbol{\theta}_k = \left\{\left\{\boldsymbol{c}_q^{(k)},\boldsymbol{d}_q^{(k)},q=1,\ldots,Q\right\}\right.$$

$$\left.\left\{\boldsymbol{A}_t^{(k)},\boldsymbol{B}_t^{(k)},t=1,\ldots,\tau\right\}, \quad \alpha^{(k)}\right\}. \quad (36)$$

Perform the following steps to update the parameter $\boldsymbol{\theta}_{k+1}$ for the next iteration.

1) Generate symbol and signal samples.
   a) Draw L symbol vectors $\boldsymbol{x}(1),\boldsymbol{x}(2),\ldots,\boldsymbol{x}(L)$ uniformly from the constellation set.
   b) Simulate L observations $\boldsymbol{y}_1(1),\boldsymbol{y}_1(2),\ldots,\boldsymbol{y}_1(L)$, where each $\boldsymbol{y}_1(\ell)$ is generated by [cf. (8), (10), and (44)]

$$\boldsymbol{y}_1(\ell) = \boldsymbol{H}(\ell)\boldsymbol{\Phi}(\ell)\boldsymbol{x}(\ell) + \boldsymbol{n}(\ell), \qquad \ell = 1,2,\ldots,L. \quad (37)$$

   c) Simulate L observations $\boldsymbol{y}_2(1),\boldsymbol{y}_2(2),\ldots,\boldsymbol{y}_2(L)$, where each $\boldsymbol{y}_2(\ell)$ is generated by [cf. (46)–(48)]

$$\boldsymbol{y}_2(\ell) = \boldsymbol{G}(\ell)\boldsymbol{\Phi}(\ell)\boldsymbol{x}(\ell) + \widetilde{\boldsymbol{G}}(\ell)\boldsymbol{y}_1(\ell) + \boldsymbol{u}(\ell)$$

$$\ell = 1,2,\ldots,L. \quad (38)$$

   d) Decode $\boldsymbol{x}(\ell)$ based on (21) and compute the empirical BLER $\gamma(\boldsymbol{z}(\ell),\boldsymbol{x}(\ell),\boldsymbol{\phi}(\ell),\boldsymbol{\psi}(\ell),\boldsymbol{\theta}_k)$.
2) Score function method for gradient estimation: Generate the estimate of (35)

$$\widehat{\nabla\Upsilon}(\boldsymbol{\theta}_k) = \frac{1}{L}\sum_{\ell=1}^L \gamma(\boldsymbol{y}_2(\ell),\boldsymbol{x}(\ell),\boldsymbol{\phi}(\ell),\boldsymbol{\psi}(\ell),\boldsymbol{\theta}_k)$$

$$\times \left\{\nabla_{\boldsymbol{\theta}}\left[\log p(\boldsymbol{y}_1(\ell)|\boldsymbol{x}(\ell),\boldsymbol{\phi}(\ell),\boldsymbol{\theta})\right.\right.$$

$$\left.\left. + \log p(\boldsymbol{y}_2(\ell)|\boldsymbol{y}_1(\ell),\boldsymbol{x}(\ell),\boldsymbol{\psi}(\ell),\boldsymbol{\theta})\right]|_{\boldsymbol{\theta}=\boldsymbol{\theta}_k}\right\}. \quad (39)$$

   The expressions of the gradients required in (39) are given in Appendix B.
3) Update parameters: The parameters are updated as

$$\boldsymbol{\theta}_{k+1} = \Pi_{\boldsymbol{\Theta}}\left[\boldsymbol{\theta}_k - \sigma_k\widehat{\nabla\Upsilon}(\boldsymbol{\theta}_k)\right] \quad (40)$$

where $\Pi_{\boldsymbol{\Theta}}(\cdot)$ is a projection operator onto the set $\boldsymbol{\Theta}$. That is, $\Pi_{\boldsymbol{\Theta}}(\cdot)$ normalizes the first-layer dispersion vectors and the second-layer dispersion matrices such that the equalities in (5) and (14) are satisfied, respectively.

C. Complexity Issues and Convergence

1) Implementation Complexity: The code design methodology proposed in this paper is an offline algorithm. Therefore, once the dispersion matrices have been designed, the actual use of the code is similar in complexity to the scheme proposed in [15], with additional memory requirement for storage of the dispersion matrices for different physical system configurations and fading environments. The need for symbol-level

synchronization does incur additional overheads. However, this assumption has been adopted in much of the cooperative diversity literature. Moreover, symbol-level synchronization is also easier to implement in a cellular-network architecture, which is assumed for this paper.

2) Convergence: The time it takes for the stochastic gradient algorithm to converge depends on the size of the system and the fading environment. They both affect the computational cost of the sphere decoder. For offline implementation, the changes in the fading conditions do not affect the performance of the system once the dispersion matrices have been designed.

IV. SIMULATION RESULTS

In this section, we present simulations to demonstrate the performance of cooperative diversity systems using the BLER-optimized LD code obtained by the algorithm given in Section III-B. It is well known that the orthogonal codes are only capacity achieving in a (2,1) system [23]. At data rate that exceeds the rate limit of an orthogonal code for a given system configuration, the orthogonal code no longer yields full diversity. When operating at the same high data rate, the proposed two-layer LD code can be optimized to perform at a lower BLER compared to the orthogonal codes. Because explicit analytical equations do not exist for BLER of the two-layer LD code, the diversity of the system can be obtained by observing the slope of the BLER curves.

Several two-layer LD codes for a cooperative diversity system are optimized for minimum BLER for different physical system configurations and fading environments. Their BLER performances are compared against other coding schemes, including orthogonal codes such as the Alamouti's code, for the range of total energy available for the system. For all examples, the two-layer LD codes are designed at an energy level where the BLER is approximately 10^{-2}. It will be seen that codes designed at a particular SNR also work well for a wide range of SNR.

Example 1—Cooperative LD Code for a (2,2) System: We first examine the performance of the cooperative LD code in a system with one source terminal, one relay terminal, and a destination terminal with two receive antennas. Fig. 2 compares the cooperative LD code with that of the Alamouti code for cooperative relays [9]–[11] in different fading channels for increasing total energy E_0. We can see that the optimized cooperative LD code outperforms the Alamouti code in a wide range of SNR values. At a BLER of 10^{-2}, the cooperative LD code has a 1-dB gain in both the Rician $K = 2$ and Rayleigh fading and a 2-dB gain in Nakagami $m = 0.5$ fading. This shows that the benefit of using the two-layered cooperative LD code over the Alamouti code increases as the channel conditions worsen. In this example, the terminals are all equal in distance and have i.i.d. fading to all other terminals. The BLER performances of cooperative LD code and Alamouti code are compared in Rayleigh, Rician $K = 2$, and Nakagami $m = 0.5$ fading channels. The Rician $K = 2$ channel represents better-than-Rayleigh channels, while Nakagami $m = 0.5$ channel represents worse-than-Rayleigh fading conditions [24]. For the cooperative LD code, we have chosen $T = 4$, $Q = 2$,
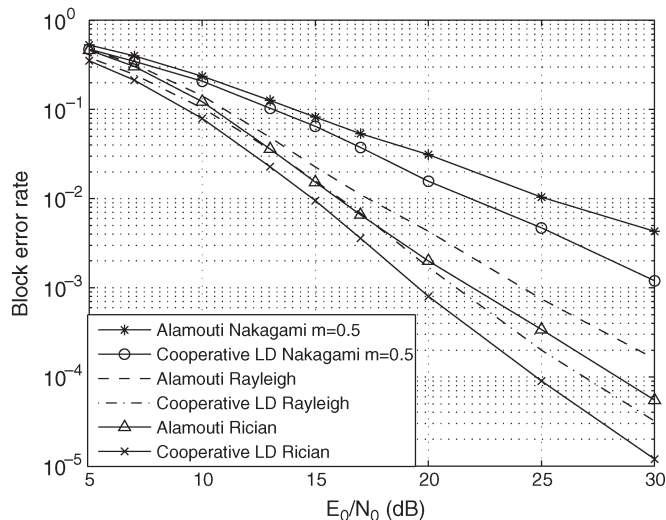

Fig. 2. Cooperative LD code versus Alamouti code for a (2,2) system under different fading environments.

and 16-QAM constellation for rate $R = 2$, and the dispersion matrices are optimized for the given channel statistics. We have found that the optimal length of the broadcast interval is at $\tau = 3$. This means that for this particular physical setup, more resources are needed in the broadcast interval for optimal BLER performance. For fair comparison, we consider the same rate, using 16-QAM constellation for the Alamouti code, and energy is equally divided between the broadcast interval and cooperation interval by choosing $\alpha = \pi/2$.

In [13], it was determined that LD codes with good performance typically have

$$Q = \min(M, N) \times T. \tag{41}$$

In this example, we can see that this constraint is met for one layer of LD code, since for the broadcast interval, the transmitted symbol size is 3, and $\min(M, N) \times T = 3$. In the cooperation interval, we fall short of this limit since we are transmitting three linearly combined symbols while $\min(M, N) \times (T - \tau) = 2$. However, since the first-layer LD code is simply the linear combination of the two 16-QAM symbols, we can see that from the overall system point of view, we are transmitting two 16-QAM symbols over one time period in a (2,2) system, thus satisfying the constraint. For the cooperative LD code, the constraint in (41) becomes

$$Q = \min(M, N) \times (T - \tau). \tag{42}$$

Example 2—Cooperative LD Code for Different System Configurations: In this example, we compare systems with different physical configurations. Fig. 3 shows the performance of an Alamouti code and four cooperative LD codes optimized at different physical system setups. The Alamouti code is simulated under a (2,2) system with $d_{SR_1} = 1$ and $d_{R_1 D_n} = 1$. Cooperative LD code 1 is optimized for a (2,2) system with $d_{SR_1} = 1$ and $d_{R_1 D_n} = 1$. Cooperative LD code 2 is optimized for a (4,2) system with $d_{SR_m} = 1$, $d_{R_m D_n} = 1$. Cooperative
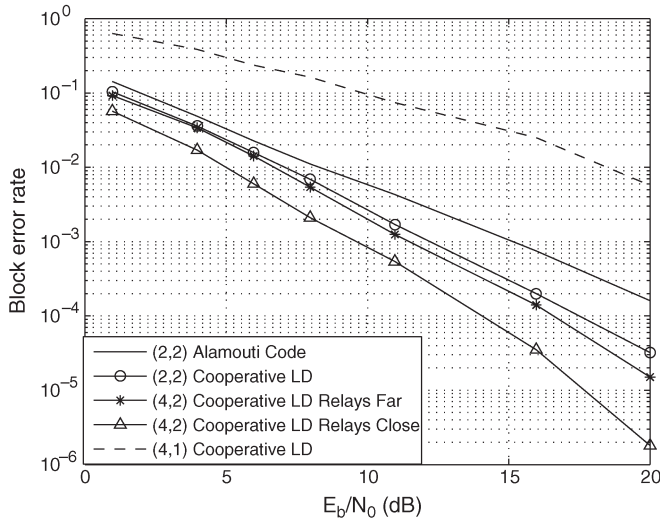
Fig. 3.   Cooperative LD code optimized in different physical configurations versus Alamouti code for a (2,2) system under Rayleigh fading.
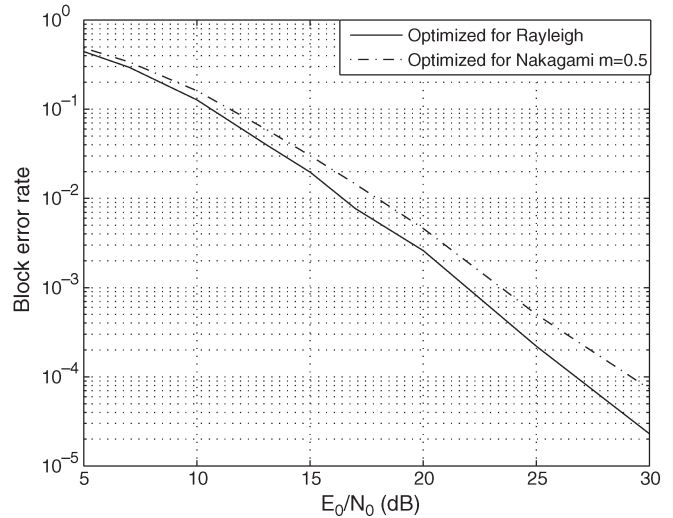

Fig. 5.   Performance in Rayleigh fading channel of codes optimized for Rayleigh and Nakagami $m = 0.5$ fading channels.
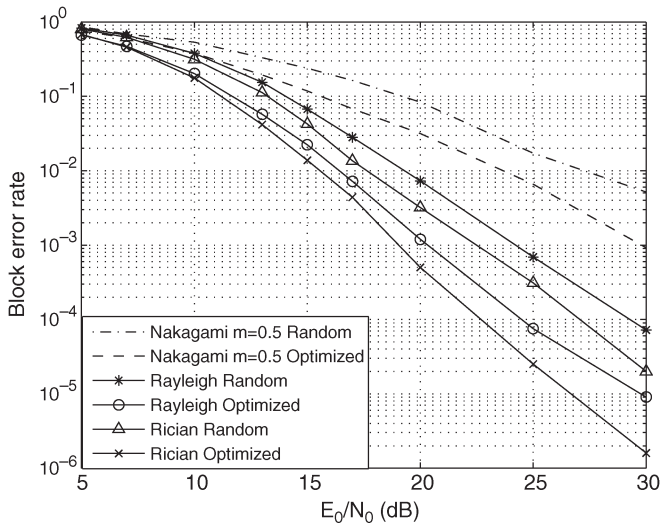

Fig. 4.   Optimized cooperative LD code versus randomly chosen cooperative LD code for a (4,2) system under different fading environments.

LD code 3 is optimized for a (4,2) system with $d_{SR_m} = 0.1$, $d_{R_m D_n} = 1$. Cooperative LD code 4 is optimized for a (4,1) system with $d_{SR_m} = 1$, $d_{R_m D_n} = 1$. All codes are chosen to transmit with $R = 2$ and are compared under Rayleigh fading. We can see that the LD code for a (4,1) system has the worst performance. The LD code for the (2,2) system and both LD codes for the (4,2) system outperforms the Alamouti code for a (2,2) system. It is interesting to see that the LD code for a (4,2) system using relays that are farther away from the source does not offer much improvement over the LD code for a (2,2) system. On the other hand, the LD code for a (4,2) system using relays that are much closer to the source offers a much bigger performance gain, indicating that the choice of relays with which the source chooses to cooperate is critical in improving system performance.

Example 3—Cooperative LD Code for a (4,2) System: We present a cooperative LD code for a system of one source,

three relays, and one two-receive-antenna destination terminal. The terminals are again all equal in distance and have i.i.d. fading to all other terminals. In Fig. 4, we compare the optimized cooperative LD code with randomly generated LD code in different fading environments. For both the optimized cooperative LD code and randomly generated code, we have code length $T = 12$, broadcast interval length $\tau = 8$, and number of substreams transmitted $Q = 8$. The optimal broadcast interval length is chosen by comparing the BLER performance of a different interval length. When QPSK constellation is used, we have rate $R = 16/12$. The energy is evenly divided between the broadcast and cooperation intervals for the randomly chosen code, while for the optimized cooperative LD code, it is determined by the proposed algorithm. We can see that, in general, a randomly chosen code does not provide good performance, and the performance of the cooperative LD code is shown to be significantly improved for all fading environments.

Example 4—Cooperative LD Code for Systems of Different Channel Statistics: In this example, we show that a code designed for one type of fading condition no longer provides the optimal performance when used under other types of fading conditions. Fig. 5 compares two sets of dispersion matrices designed under a different fading environment. Both sets of dispersion matrices are optimized for a (2,2) system: one for the Rayleigh fading channel and one for the Nakagami $m = 0.5$ channel. Both codes are then simulated in a Rayleigh fading environment. Although it is also optimized for BLER, it is clearly shown in Fig. 5 that the Nakagami $m = 0.5$ code does not offer the best performance under Rayleigh fading conditions and that it suffers a 1-dB loss due to the mismatch in channel conditions.

In Fig. 6, we can see how different fading channels and a different system setup affect the resulting optimized code in terms of the energy allocations. In this example, we fixed the distance between source and destination at $d_{SD} = 2$ while varying the ratio of the source-relay and relay-destination distances. In all fading environments, for d_{SR}/d_{RD} between $(1/2, 2)$, the allocation of energy to the broadcast interval
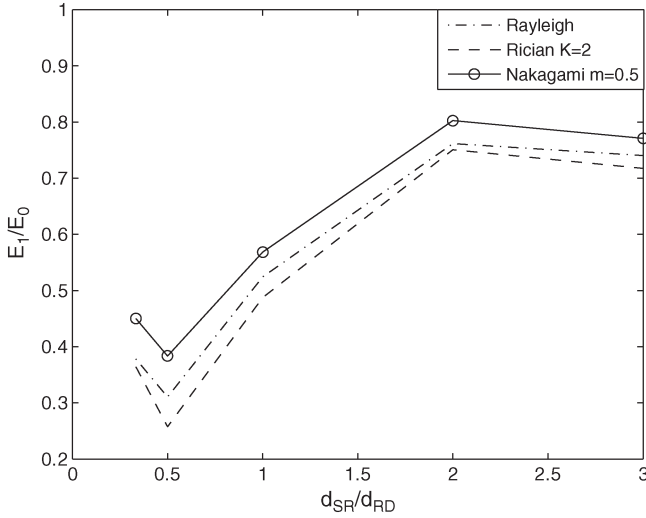
Fig. 6. Ratio of energy allocation in broadcast interval for different $D_{\mathrm{SR}}/D_{\mathrm{RD}}$ ratios.

increases as $d_{\mathrm{SR}}/d_{\mathrm{RD}}$ increases. When the distance between source and relay increases, more energy is needed to improve the SNR of the received signal at the relay during the broadcast interval. We can see that having a clean copy of the symbols to be transmitted at the relays is just as important as having a good channel between the relay and destination. For $d_{\mathrm{SR}}/d_{\mathrm{RD}}$ between (1/3, 1/2), there is a decrease in the allocation of energy to the broadcast interval as the ratio increases. This range of ratio represents the setups of having $d_{\mathrm{RD}} > d_{\mathrm{SD}}$. In this case, there is less path loss between source and destination than relay and destination. Recall that the destination receives a copy of the transmission from the source during the broadcast interval, and the received signal is also used in detection. Thus, BLER performance is improved by increasing the energy allocation to the broadcast interval to better utilize this portion of the transmission. On the other hand, system setups with $d_{\mathrm{SR}}/d_{\mathrm{RD}}$ between (2, 3) represent having $d_{\mathrm{RD}} < d_{\mathrm{SD}}$. The slight decrease in energy allocated to the broadcast interval in this region represents the devoting of more system resources to the second-layer LD code, which is affected by d_{RD}. However, we can see that the decrease in energy is not as dramatic as the decrease in the range of (1/3, 1/2), since the poorly received signal at the relay during the broadcast interval will also adversely affect the BLER. This effect can be seen by comparing the energy allocation in different fading environments. Fig. 6 shows clearly that for all ratios of $d_{\mathrm{SR}}/d_{\mathrm{RD}}$, the worse fading channel will require more energy being allocated to the broadcast interval.

V. Conclusion

In this paper, we have proposed a two-layered LD code-based space–time coding scheme for cooperative diversity systems and a simulation-based optimization algorithm to construct the optimal code based on BLER performance. The proposed code design algorithm can obtain optimal codes under arbitrary fading statistics and arbitrary system configuration in terms of both number of terminals and distances between the terminals by finding the optimal set of code matrices, energy allocation scheme, and broadcast interval length. Since the codes are designed offline, codes for different configurations can be obtained beforehand and chosen by the system to match the detected configuration. In this paper, we have also assumed flat-fading environment. This is a reasonable assumption given that the source terminal, in general, will select relay terminals in close proximity to the source. In the case where the transmitting terminals are far from the destination, both the source and the relay terminals will have relatively equal distance to the destination. Also, since this code design algorithm is simulation-based, appropriate modifications to the transmission model and the corresponding gradient computations can also be designed in a frequency-selective fading environment. Therefore, the same simulation-based code optimization approach can be used for different model assumptions, where other code designs such as delay-diversity codes only work well under the specific assumption of frequency-selective channels.

Appendix A
Definitions

For (8), we define (43), shown at the bottom of the page. For (10), we define

$$
\boldsymbol{H} \triangleq \begin{bmatrix} |h_1|\sqrt{\rho_{\mathrm{SR}_1}}\boldsymbol{I}_{2\tau \times 2\tau} \\ \vdots \\ |h_{M-1}|\sqrt{\rho_{\mathrm{SR}_{M-1}}}\boldsymbol{I}_{2\tau \times 2\tau} \end{bmatrix}_{2(M-1)\tau \times 2\tau}. \tag{44}
$$

For (19), we define (45), shown at the bottom of the next page, where

$$
\mathcal{A}_{m,t} \triangleq \begin{bmatrix} \Re\{\boldsymbol{a}_{m,t}\} & -\Im\{\boldsymbol{a}_{m,t}\} \\ \Im\{\boldsymbol{a}_{m,t}\} & \Re\{\boldsymbol{a}_{m,t}\} \end{bmatrix}_{2(T-\tau)\times 2}
$$

$$
\mathcal{B}_{m,t} \triangleq \begin{bmatrix} -\Im\{\boldsymbol{b}_{m,t}\} & -\Re\{\boldsymbol{b}_{m,t}\} \\ \Re\{\boldsymbol{b}_{m,t}\} & -\Im\{\boldsymbol{b}_{m,t}\} \end{bmatrix}_{2(T-\tau)\times 2}
$$

$$
\boldsymbol{\Phi} \triangleq \begin{bmatrix} \Re\{c_{11}\} & -\Im\{d_{11}\} & \Re\{c_{12}\} & \cdots & \Re\{c_{1Q}\} & -\Im\{d_{1Q}\} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \Re\{c_{\tau 1}\} & -\Im\{d_{\tau 1}\} & \Re\{c_{\tau 2}\} & \cdots & \Re\{c_{\tau Q}\} & -\Im\{d_{\tau Q}\} \\ \Im\{c_{11}\} & \Re\{d_{11}\} & \Im\{c_{12}\} & \cdots & \Im\{c_{1Q}\} & \Re\{d_{1Q}\} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \Im\{c_{\tau 1}\} & \Re\{d_{\tau 1}\} & \Im\{c_{\tau 2}\} & \cdots & \Im\{c_{\tau Q}\} & \Re\{d_{\tau Q}\} \end{bmatrix}_{2\tau \times 2Q} \tag{43}
$$

and

$$\boldsymbol{g}_{M,n} \stackrel{\triangle}{=} \sqrt{\rho_{\mathrm{SD},2}} \begin{bmatrix} \Re\{g_{M,n}\} \\ \Im\{g_{M,n}\} \end{bmatrix}$$

$$\boldsymbol{g}_{m,n} \stackrel{\triangle}{=} \gamma_{R_m}\sqrt{\rho_{\mathrm{RD}}} \begin{bmatrix} \Re\{g_{m,n}\} \\ \Im\{g_{m,n}\} \end{bmatrix} \qquad m=1,\dots,M-1 \quad (46)$$

and

$$\mathbf{P} \stackrel{\triangle}{=} \sqrt{\rho_{\mathrm{SD},1}}\,[\,\mathbf{P}_1 \quad \mathbf{P}_2 \quad \cdots \quad \mathbf{P}_N\,]^T_{2N\tau \times 2\tau} \quad (47)$$

where

$$\mathbf{P}_n \stackrel{\triangle}{=} \begin{bmatrix} \Re\{g_{M,n}\}\boldsymbol{I}_{\tau\times\tau} & \Im\{g_{M,n}\}\boldsymbol{I}_{\tau\times\tau} \\ -\Im\{g_{M,n}\}\boldsymbol{I}_{\tau\times\tau} & \Re\{g_{M,n}\}\boldsymbol{I}_{\tau\times\tau} \end{bmatrix}_{2\tau\times2\tau}. \quad (48)$$

We further define (49), shown at the bottom of the page.

APPENDIX B
GRADIENT DERIVATIONS

A. Gradient Calculation for Energy Allocation

Recall from (35) that

$$\nabla_{\boldsymbol{\theta}} \log p\,(\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}[\log \mathcal{P}_1 + \log \mathcal{P}_2]. \quad (50)$$

From (31), we have $\mathcal{P}_1$ distributed as multivariate Gaussian. Define

$$f_1 \stackrel{\triangle}{=} -(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})^T(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x}). \quad (51)$$

Then, from (2), (3), and (44), we have

$$\begin{aligned} \frac{\partial f_1(\alpha)}{\partial\alpha} &= -\frac{\partial(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})^T}{\partial\alpha}(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x}) \\ &\quad - (\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})^T\frac{\partial(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})}{\partial\alpha} \\ &= -\frac{\sin\alpha}{\cos\alpha}\Big[(\boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})^T(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x}) \\ &\qquad\qquad + (\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})^T(\boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})\Big]. \quad (52) \end{aligned}$$

In the cooperation interval, we have from (32) that $\mathcal{P}_2$ is also multivariate Gaussian distributed. Define

$$f_2 \stackrel{\triangle}{=} -(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1) \quad (53)$$

and

$$\boldsymbol{G} = \boldsymbol{\Lambda} + \bar{\boldsymbol{\Lambda}} \quad (54)$$

where

$$\boldsymbol{\Lambda} \stackrel{\triangle}{=} \begin{bmatrix} \mathbf{P} \\ \mathbf{0}_{2N(T-\tau)\times2\tau} \end{bmatrix}$$

$$\bar{\boldsymbol{\Lambda}} \stackrel{\triangle}{=} \begin{bmatrix} & & \mathbf{0}_{2N\tau\times2\tau} & & & \\ \mathcal{A}_{1,1}\boldsymbol{g}_{M,1} & \cdots & \mathcal{A}_{1,\tau}\boldsymbol{g}_{M,1} & \mathcal{B}_{1,1}\boldsymbol{g}_{M,1} & \cdots & \mathcal{B}_{1,\tau}\boldsymbol{g}_{M,1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}_{1,1}\boldsymbol{g}_{M,N} & \cdots & \mathcal{A}_{1,\tau}\boldsymbol{g}_{M,N} & \mathcal{B}_{1,1}\boldsymbol{g}_{M,N} & \cdots & \mathcal{B}_{1,\tau}\boldsymbol{g}_{M,N} \end{bmatrix}. \quad (55)$$

Thus, $\boldsymbol{G}$ is separated into two matrices: $\boldsymbol{\Lambda}$, which depends only on E_1, and $\bar{\boldsymbol{\Lambda}}$, which depends only on E_2. From (46)–(48), we have

$$\begin{aligned} \frac{\partial f_2(\alpha)}{\partial\alpha} &= -\left[\frac{\sin\alpha}{\cos\alpha}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{x} - \frac{\cos\alpha}{\sin\alpha}(\bar{\boldsymbol{\Lambda}}\boldsymbol{\Phi}\boldsymbol{x} + \widetilde{\boldsymbol{G}}\boldsymbol{y}_1) - \triangledown\widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right]^T \\ &\quad \times (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1) - (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T \\ &\quad \times \left[\frac{\sin\alpha}{\cos\alpha}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{x} - \frac{\cos\alpha}{\sin\alpha}(\bar{\boldsymbol{\Lambda}}\boldsymbol{\Phi}\boldsymbol{x} + \widetilde{\boldsymbol{G}}\boldsymbol{y}_1 - \triangledown\widetilde{\boldsymbol{G}}\boldsymbol{y}_1)\right] \end{aligned} \quad (56)$$

where $\triangledown\widetilde{\boldsymbol{G}}$ is given in (57), shown at the bottom of the next page, with

$$\bar{\gamma}_{R_m} = |h_m|^2\,\frac{\rho_{\mathrm{SR}_m}}{\cos^2\alpha}\cos\alpha\sin\alpha. \quad (58)$$

Combining (52) and (56), the gradient of $\log p(\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})$ with respect to α is given by

$$\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta})|_\alpha = \frac{\partial f_1(\alpha)}{\partial\alpha} + \frac{\partial f_2(\alpha)}{\partial\alpha}. \quad (59)$$

$$\boldsymbol{G} \stackrel{\triangle}{=} \begin{bmatrix} & & \mathbf{P} & & & \\ \mathcal{A}_{M,1}\boldsymbol{g}_{M,1} & \cdots & \mathcal{A}_{M,\tau}\boldsymbol{g}_{M,1} & \mathcal{B}_{M,1}\boldsymbol{g}_{M,1} & \cdots & \mathcal{B}_{M,\tau}\boldsymbol{g}_{M,1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}_{M,1}\boldsymbol{g}_{M,N} & \cdots & \mathcal{A}_{M,\tau}\boldsymbol{g}_{M,N} & \mathcal{B}_{M,1}\boldsymbol{g}_{M,N} & \cdots & \mathcal{B}_{M,\tau}\boldsymbol{g}_{M,N} \end{bmatrix}_{2NT\times2\tau} \quad (45)$$

$$\widetilde{\boldsymbol{G}} \stackrel{\triangle}{=} \begin{bmatrix} & & \mathbf{0}_{2N\tau\times2(M-1)\tau} & & & \\ \mathcal{A}_{1,1}\boldsymbol{g}_{1,1} & \cdots & \mathcal{A}_{1,\tau}\boldsymbol{g}_{1,1} & \mathcal{B}_{1,1}\boldsymbol{g}_{1,1} & \cdots & \mathcal{B}_{M-1,\tau}\boldsymbol{g}_{M-1,1} \\ \mathcal{A}_{1,1}\boldsymbol{g}_{1,2} & \cdots & \mathcal{A}_{1,\tau}\boldsymbol{g}_{1,2} & \mathcal{B}_{1,1}\boldsymbol{g}_{1,2} & \cdots & \mathcal{B}_{M-1,\tau}\boldsymbol{g}_{M-1,2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}_{1,1}\boldsymbol{g}_{1,N} & \cdots & \mathcal{A}_{1,\tau}\boldsymbol{g}_{1,N} & \mathcal{B}_{1,1}\boldsymbol{g}_{1,N} & \cdots & \mathcal{B}_{M-1,\tau}\boldsymbol{g}_{M-1,N} \end{bmatrix}_{2NT\times2(M-1)\tau} \quad (49)$$

B. Gradient Calculation for Dispersion Matrices

First-Layer Dispersion Vectors: Next, we provide the gradient of f_1 and f_2 with respect to $\boldsymbol{c}_{R,q}$. The tth entry of the gradient of $f_1(\boldsymbol{c}_{R,q})$ is

$$\left[\frac{\partial f_1(\boldsymbol{c}_{R,q})}{\partial \boldsymbol{c}_{R,q}}\right]_t = \lim_{\delta\to 0} \frac{f_1(\boldsymbol{c}_{R,q}+\delta\boldsymbol{\epsilon}_t)-f_1(\boldsymbol{c}_{R,q})}{\delta} \quad (60)$$

where $\boldsymbol{\epsilon}_t$ is a $\tau \times 1$ column vector with one at the tth position and zero elsewhere. From (8), we have

$$\boldsymbol{\Phi}_{\boldsymbol{c}_{R,q}+\delta\boldsymbol{\epsilon}_t} = \boldsymbol{\Phi} + \delta\boldsymbol{\Xi}_{t,2q-1}$$
$$\boldsymbol{\Phi}_{\boldsymbol{c}_{I,q}+\delta\boldsymbol{\epsilon}_t} = \boldsymbol{\Phi} + \delta\boldsymbol{\Xi}_{\tau+t,2q-1}$$
$$\boldsymbol{\Phi}_{\boldsymbol{d}_{R,q}+\delta\boldsymbol{\epsilon}_t} = \boldsymbol{\Phi} + \delta\boldsymbol{\Xi}_{\tau+t,2q}$$
$$\boldsymbol{\Phi}_{\boldsymbol{d}_{I,q}+\delta\boldsymbol{\epsilon}_t} = \boldsymbol{\Phi} - \delta\boldsymbol{\Xi}_{t,2q} \quad (61)$$

where $\boldsymbol{\Xi}_{i,j}$ is a $2\tau \times 2Q$ matrix with one at the (i,j) position and zeros everywhere else. Therefore

$$f_1(\boldsymbol{c}_{R,q}+\delta\boldsymbol{\epsilon}_t) = -(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}_{\boldsymbol{c}_{R,q}+\delta\boldsymbol{\epsilon}_t}\boldsymbol{x})^T(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}_{\boldsymbol{c}_{R,q}+\delta\boldsymbol{\epsilon}_t}\boldsymbol{x})$$
$$= -(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x} - \delta\boldsymbol{H}\boldsymbol{\Xi}_{t,2q-1}\boldsymbol{x})^T$$
$$\times (\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x} - \delta\boldsymbol{H}\boldsymbol{\Xi}_{t,2q-1}\boldsymbol{x})$$
$$= f_1 + \delta(\boldsymbol{H}\boldsymbol{\Xi}_{t,2q-1}\boldsymbol{x})^T(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})$$
$$+ \delta(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})^T(\boldsymbol{H}\boldsymbol{\Xi}_{t,2q-1}\boldsymbol{x}) + o(\delta). \quad (62)$$

Therefore, we have

$$\left[\frac{\partial f_1(\boldsymbol{c}_{R,q})}{\partial \boldsymbol{c}_{R,q}}\right]_t = (\boldsymbol{H}\boldsymbol{\Xi}_{t,2q-1}\boldsymbol{x})^T(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})$$
$$+ (\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})^T(\boldsymbol{H}\boldsymbol{\Xi}_{t,2q-1}\boldsymbol{x}). \quad (63)$$

Similarly for the gradients of f_1 with respect to $\boldsymbol{c}_{I,q}$, $\boldsymbol{d}_{R,q}$, and $\boldsymbol{d}_{I,q}$, we have the following expressions:

$$\left[\frac{\partial f_1(\boldsymbol{c}_{I,q})}{\partial \boldsymbol{c}_{I,q}}\right]_t = (\boldsymbol{H}\boldsymbol{\Xi}_{\tau+t,2q-1}\boldsymbol{x})^T(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})$$
$$+ (\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})^T(\boldsymbol{H}\boldsymbol{\Xi}_{\tau+t,2q-1}\boldsymbol{x})$$
$$\left[\frac{\partial f_1(\boldsymbol{d}_{R,q})}{\partial \boldsymbol{d}_{R,q}}\right]_t = (\boldsymbol{H}\boldsymbol{\Xi}_{\tau+t,2q}\boldsymbol{x})^T(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})$$
$$+ (\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})^T(\boldsymbol{H}\boldsymbol{\Xi}_{\tau+t,2q}\boldsymbol{x})$$
$$\left[\frac{\partial f_1(\boldsymbol{d}_{I,q})}{\partial \boldsymbol{d}_{I,q}}\right]_t = -(\boldsymbol{H}\boldsymbol{\Xi}_{t,2q}\boldsymbol{x})^T(\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})$$
$$- (\boldsymbol{y}_1 - \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{x})^T(\boldsymbol{H}\boldsymbol{\Xi}_{t,2q}\boldsymbol{x}). \quad (64)$$

We now compute the gradient of $\log\mathcal{P}_2$ with respect to $\boldsymbol{c}_{R,q}$ for the cooperation intervals. We have

$$f_2(\boldsymbol{c}_{R,q}+\delta\boldsymbol{\epsilon}_t) = -\left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}_{\boldsymbol{c}_{R,q}+\delta\boldsymbol{\epsilon}_t}\boldsymbol{x} + \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right)^T$$
$$\times \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}_{\boldsymbol{c}_{R,q}+\delta\boldsymbol{\epsilon}_t}\boldsymbol{x} + \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right)$$
$$= f_2 + \delta(\boldsymbol{G}\boldsymbol{\Xi}_{t,2q-1}\boldsymbol{x})^T(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)$$
$$+ \delta(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T(\boldsymbol{G}\boldsymbol{\Xi}_{t,2q-1}\boldsymbol{x}) + o(\delta). \quad (65)$$

Therefore, we have

$$\left[\frac{\partial f_2(\boldsymbol{c}_{R,q})}{\partial \boldsymbol{c}_{R,q}}\right]_t = (\boldsymbol{G}\boldsymbol{\Xi}_{t,2q-1}\boldsymbol{x})^T(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)$$
$$+ (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T(\boldsymbol{G}\boldsymbol{\Xi}_{t,2q-1}\boldsymbol{x}). \quad (66)$$

Similarly for the gradients of f_2 with respect to $\boldsymbol{c}_{I,q}$, $\boldsymbol{d}_{R,q}$, and $\boldsymbol{d}_{I,q}$, we have the following expressions:

$$\left[\frac{\partial f_2(\boldsymbol{c}_{I,q})}{\partial \boldsymbol{c}_{I,q}}\right]_t = (\boldsymbol{G}\boldsymbol{\Xi}_{\tau+t,2q-1}\boldsymbol{x})^T(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)$$
$$+ (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T(\boldsymbol{G}\boldsymbol{\Xi}_{\tau+t,2q-1}\boldsymbol{x})$$
$$\left[\frac{\partial f_2(\boldsymbol{d}_{R,q})}{\partial \boldsymbol{d}_{R,q}}\right]_t = (\boldsymbol{G}\boldsymbol{\Xi}_{\tau+t,2q}\boldsymbol{x})^T(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)$$
$$+ (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T(\boldsymbol{G}\boldsymbol{\Xi}_{\tau+t,2q}\boldsymbol{x})$$
$$\left[\frac{\partial f_2(\boldsymbol{d}_{I,q})}{\partial \boldsymbol{d}_{I,q}}\right]_t = -(\boldsymbol{G}\boldsymbol{\Xi}_{t,2q}\boldsymbol{x})^T(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)$$
$$- (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T(\boldsymbol{G}\boldsymbol{\Xi}_{t,2q}\boldsymbol{x}). \quad (67)$$

Second-Layer Dispersion Matrices: We can see that $p(\boldsymbol{y}_1|\boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\theta})$ is independent of the second-layer dispersion matrices; therefore, the gradient with respect to the second-layer dispersion matrices is zero, and we only have to evaluate the gradient $\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_2|\boldsymbol{y}_1, \boldsymbol{x}, \boldsymbol{\psi}, \boldsymbol{\theta})$.

Let us evaluate the gradient of $\log p(\boldsymbol{y}_2|\boldsymbol{y}_1, \boldsymbol{x}, \boldsymbol{\psi}, \boldsymbol{\theta})$ with respect to $\boldsymbol{A}_{R,t}$. Note from (45) and (49) that $\boldsymbol{G}$ depends only upon the Mth columns of the second-layer dispersion matrices, while $\widetilde{\boldsymbol{G}}$ is independent of the Mth columns of the second-layer dispersion matrices.

Let $\boldsymbol{\varsigma}_n$ and $\boldsymbol{\kappa}_m$ be $T-\tau$ and M dimensional vectors with 1 at the nth and mth position, respectively. Define

$$\boldsymbol{G}^{\boldsymbol{A}_{R,t}+\delta\boldsymbol{\varsigma}_n\boldsymbol{\kappa}_M^T} = \boldsymbol{G} + \delta\boldsymbol{\Xi}_{n,M}^{\boldsymbol{A}_{R,t}}$$
$$\widetilde{\boldsymbol{G}}^{\boldsymbol{A}_{R,t}+\delta\boldsymbol{\varsigma}_n\boldsymbol{\kappa}_m^T} = \widetilde{\boldsymbol{G}} + \delta\boldsymbol{\Xi}_{n,m}^{\boldsymbol{A}_{R,t}}, \qquad m = 1,\dots,M-1 \quad (68)$$

$$\nabla\widetilde{\boldsymbol{G}} \triangleq \begin{bmatrix} & & & \boldsymbol{0}_{2N\tau\times 2(M-1)\tau} & & \\ \bar{\gamma}_{R_1}\mathcal{A}_{2,1}\boldsymbol{g}_{1,1} & \cdots & \bar{\gamma}_{R_1}\mathcal{A}_{2,\tau}\boldsymbol{g}_{1,1} & \bar{\gamma}_{R_1}\mathcal{B}_{2,1}\boldsymbol{g}_{1,1} & \cdots & \bar{\gamma}_{R_{M-1}}\mathcal{B}_{M,\tau}\boldsymbol{g}_{M-1,1} \\ \bar{\gamma}_{R_1}\mathcal{A}_{2,1}\boldsymbol{g}_{1,2} & \cdots & \bar{\gamma}_{R_1}\mathcal{A}_{2,\tau}\boldsymbol{g}_{1,2} & \bar{\gamma}_{R_1}\mathcal{B}_{2,1}\boldsymbol{g}_{1,2} & \cdots & \bar{\gamma}_{R_{M-1}}\mathcal{B}_{M,\tau}\boldsymbol{g}_{M-1,2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{\gamma}_{R_1}\mathcal{A}_{2,1}\boldsymbol{g}_{1,N} & \cdots & \bar{\gamma}_{R_1}\mathcal{A}_{2,\tau}\boldsymbol{g}_{1,N} & \bar{\gamma}_{R_1}\mathcal{B}_{2,1}\boldsymbol{g}_{1,N} & \cdots & \bar{\gamma}_{R_{M-1}}\mathcal{B}_{M,\tau}\boldsymbol{g}_{M-1,N} \end{bmatrix} \quad (57)$$

where

$$\boldsymbol{\Xi}_{n,M}^{\boldsymbol{A}_{R,t}} \triangleq \begin{bmatrix} & & \boldsymbol{0}_{2N\tau \times 2\tau} & & & & & \\ 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{A}_{R,t}} \boldsymbol{g}_{M,1} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{A}_{R,t}} \boldsymbol{g}_{M,N} & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{69}$$

and (70), shown at the bottom of the page, with

$$\boldsymbol{\Lambda}_n^{\boldsymbol{A}_{R,t}} \triangleq \begin{bmatrix} \boldsymbol{\varsigma}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\varsigma}_n \end{bmatrix}. \tag{71}$$

Thus, for $m = M$, we have

$$\begin{aligned} f_2\left(\boldsymbol{A}_{R,t} + \delta \boldsymbol{\varsigma}_n \boldsymbol{\kappa}_M^T\right) &= -\left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \delta \boldsymbol{\Xi}_{n,M}^{\boldsymbol{A}_{R,t}} \boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right)^T \\ &\quad \times \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \delta \boldsymbol{\Xi}_{n,M}^{\boldsymbol{A}_{R,t}} \boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right) \\ &= f_2 + \delta \left(\boldsymbol{\Xi}_{n,M}^{\boldsymbol{A}_{R,t}} \boldsymbol{\Phi}\boldsymbol{x}\right)^T (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1) \\ &\quad + \delta(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T \\ &\quad \times \left(\boldsymbol{\Xi}_{n,M}^{\boldsymbol{A}_{R,t}} \boldsymbol{\Phi}\boldsymbol{x}\right) + o(\delta). \end{aligned} \tag{72}$$

Therefore

$$\begin{aligned} \left[\frac{\partial f_2(\boldsymbol{A}_{R,t})}{\partial \boldsymbol{A}_{R,t}}\right]_{n,M} &= \left(\boldsymbol{\Xi}_{n,M}^{\boldsymbol{A}_{R,t}} \boldsymbol{\Phi}\boldsymbol{x}\right)^T \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right) \\ &\quad + \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right)^T \left(\boldsymbol{\Xi}_{n,M}^{\boldsymbol{A}_{R,t}} \boldsymbol{\Phi}\boldsymbol{x}\right). \end{aligned} \tag{73}$$

For $m = 1, \ldots, M - 1$, we have

$$\begin{aligned} f_2\left(\boldsymbol{A}_{R,t} + \delta \boldsymbol{\varsigma}_n \boldsymbol{\kappa}_m^T\right) &= -\left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1 - \delta \boldsymbol{\Xi}_{n,m}^{\boldsymbol{A}_{R,t}} \boldsymbol{y}_1\right)^T \\ &\quad \times \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1 - \delta \boldsymbol{\Xi}_{n,m}^{\boldsymbol{A}_{R,t}} \boldsymbol{y}_1\right) \\ &= f_2 + \delta \left(\boldsymbol{\Xi}_{n,m}^{\boldsymbol{A}_{R,t}} \boldsymbol{y}_1\right)^T (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1) \\ &\quad + \delta(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T \\ &\quad \times \left(\boldsymbol{\Xi}_{n,m}^{\boldsymbol{A}_{R,t}} \boldsymbol{y}_1\right) + o(\delta). \end{aligned} \tag{74}$$

Therefore

$$\begin{aligned} \left[\frac{\partial f_2(\boldsymbol{A}_{R,t})}{\partial \boldsymbol{A}_{R,t}}\right]_{n,m} &= \left(\boldsymbol{\Xi}_{n,m}^{\boldsymbol{A}_{R,t}} \boldsymbol{y}_1\right)^T \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right) \\ &\quad + (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T \left(\boldsymbol{\Xi}_{n,m}^{\boldsymbol{A}_{R,t}} \boldsymbol{y}_1\right). \end{aligned} \tag{75}$$

For the gradients with respect to $\boldsymbol{A}_{I,t}$, $\boldsymbol{B}_{R,t}$, and $\boldsymbol{B}_{I,q}$, similar expressions can be given for $m = M$

$$\begin{aligned} \left[\frac{\partial f_2(\boldsymbol{A}_{I,t})}{\partial \boldsymbol{A}_{I,t}}\right]_{n,M} &= \left(\boldsymbol{\Xi}_{n,M}^{\boldsymbol{A}_{I,t}} \boldsymbol{\Phi}\boldsymbol{x}\right)^T \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right) \\ &\quad + (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T \left(\boldsymbol{\Xi}_{n,M}^{\boldsymbol{A}_{I,t}} \boldsymbol{\Phi}\boldsymbol{x}\right) \end{aligned} \tag{76}$$

$$\begin{aligned} \left[\frac{\partial f_2(\boldsymbol{B}_{R,t})}{\partial \boldsymbol{B}_{R,t}}\right]_{n,M} &= \left(\boldsymbol{\Xi}_{n,M}^{\boldsymbol{B}_{R,t}} \boldsymbol{\Phi}\boldsymbol{x}\right)^T \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right) \\ &\quad + (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T \left(\boldsymbol{\Xi}_{n,M}^{\boldsymbol{B}_{R,t}} \boldsymbol{\Phi}\boldsymbol{x}\right) \end{aligned} \tag{77}$$

$$\begin{aligned} \left[\frac{\partial f_2(\boldsymbol{B}_{I,t})}{\partial \boldsymbol{B}_{I,t}}\right]_{n,M} &= \left(\boldsymbol{\Xi}_{n,M}^{\boldsymbol{B}_{I,t}} \boldsymbol{\Phi}\boldsymbol{x}\right)^T \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right) \\ &\quad + (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T \left(\boldsymbol{\Xi}_{n,M}^{\boldsymbol{B}_{I,t}} \boldsymbol{\Phi}\boldsymbol{x}\right) \end{aligned} \tag{78}$$

and for $m = 1, \ldots, M - 1$

$$\begin{aligned} \left[\frac{\partial f_2(\boldsymbol{A}_{I,t})}{\partial \boldsymbol{A}_{I,t}}\right]_{n,m} &= \left(\boldsymbol{\Xi}_{n,m}^{\boldsymbol{A}_{I,t}} \boldsymbol{y}_1\right)^T \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right) \\ &\quad + (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T \left(\boldsymbol{\Xi}_{n,m}^{\boldsymbol{A}_{I,t}} \boldsymbol{y}_1\right) \end{aligned} \tag{79}$$

$$\begin{aligned} \left[\frac{\partial f_2(\boldsymbol{B}_{R,t})}{\partial \boldsymbol{B}_{R,t}}\right]_{n,m} &= \left(\boldsymbol{\Xi}_{n,m}^{\boldsymbol{B}_{R,t}} \boldsymbol{y}_1\right)^T \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right) \\ &\quad + (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T \left(\boldsymbol{\Xi}_{n,m}^{\boldsymbol{B}_{R,t}} \boldsymbol{y}_1\right) \end{aligned} \tag{80}$$

$$\begin{aligned} \left[\frac{\partial f_2(\boldsymbol{B}_{I,t})}{\partial \boldsymbol{B}_{I,q}}\right]_{n,m} &= \left(\boldsymbol{\Xi}_{n,m}^{\boldsymbol{B}_{I,t}} \boldsymbol{y}_1\right)^T \left(\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1\right) \\ &\quad + (\boldsymbol{y}_2 - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{x} - \widetilde{\boldsymbol{G}}\boldsymbol{y}_1)^T \left(\boldsymbol{\Xi}_{n,m}^{\boldsymbol{B}_{I,t}} \boldsymbol{y}_1\right) \end{aligned} \tag{81}$$

where we have (82) and (83), shown at the bottom of the next page, with (84)–(86), also shown at the bottom of the next page, and with (87)–(89), also shown at the bottom of the next page, with

$$\boldsymbol{\Lambda}_n^{\boldsymbol{B}_{I,t}} \triangleq \begin{bmatrix} -\boldsymbol{\varsigma}_n & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{\varsigma}_n \end{bmatrix}. \tag{90}$$

$$\boldsymbol{\Xi}_{n,m}^{\boldsymbol{A}_{R,t}} \triangleq \begin{bmatrix} & & \boldsymbol{0}_{2N\tau \times 2(M-1)\tau} & & & & & & & & \\ 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{A}_{R,t}} \boldsymbol{g}_{m,1} & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{A}_{R,t}} \boldsymbol{g}_{m,N} & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}, \qquad m = 1, \ldots, M-1 \tag{70}$$

APPENDIX C

Here, we provide the proof for (34). Given symbols $\{s_q\}_{q=1}^Q$ from a finite r-QAM constellation, with $s_q = \alpha_q + j\beta_q$, define the symbol vector $\boldsymbol{x}$ as in (8), $\boldsymbol{x} = [\alpha_1, \beta_1, \ldots, \alpha_Q, \beta_Q]^T$, i.e., each element in $\boldsymbol{x}$ belongs to a $\sqrt{r}$-PAM constellation. Denote the set of all possible symbol vectors $\Omega = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{r^Q}\}$. We now show that for equiprobable symbol vectors and $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}_2$, it is equivalent to showing

$$\sum_{i\in\Omega} \int \left\{\nabla_{\boldsymbol{\theta}}\left[1 - \gamma(\boldsymbol{z}, \boldsymbol{x}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta})\right]\right\} p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta})d\boldsymbol{z} = 0 \tag{91}$$

where

$$p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{NT}} \exp\left[-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{Bx})^T(\boldsymbol{z} - \boldsymbol{Bx})\right] \tag{92}$$

with $\boldsymbol{B} \triangleq \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{G} + \widetilde{\boldsymbol{G}}\boldsymbol{H})\boldsymbol{\Phi}$.

By making the substitution $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}_2$ and change of variable from $\boldsymbol{z}$ to $\boldsymbol{y}$ in (91), we have

$$\sum_{i\in\Omega} \frac{1}{(2\pi)^{NT}} \int \left\{\nabla_{\boldsymbol{\theta}}\left[1 - \gamma(\boldsymbol{y}_2, \boldsymbol{x}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta})\right]\right\}$$
$$\times \exp\left[-\frac{1}{2}\left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{y}_2 - \boldsymbol{Bx}_i\right\|^2\right] |\boldsymbol{\Sigma}^{-\frac{1}{2}}|d\boldsymbol{y}_2 = 0. \tag{93}$$

$$\boldsymbol{\Xi}_{n,M}^{\boldsymbol{A}_{I,t}} \triangleq \begin{bmatrix} & & & \boldsymbol{0}_{2N\tau\times2\tau} & & & \\ 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{A}_{I,t}}\boldsymbol{g}_{M,1} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{A}_{I,t}}\boldsymbol{g}_{M,N} & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{82}$$

$$\boldsymbol{\Xi}_{n,m}^{\boldsymbol{A}_{I,t}} \triangleq \begin{bmatrix} & & & \boldsymbol{0}_{2N\tau\times2(M-1)\tau} & & & & & \\ 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{A}_{I,t}}\boldsymbol{g}_{m,1} & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \ddots & \vdots & \vdots & \ddots & \vdots & & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{A}_{I,t}}\boldsymbol{g}_{m,N} & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}, \qquad m = 1, \ldots, M-1 \tag{83}$$

$$\boldsymbol{\Lambda}_n^{\boldsymbol{A}_{I,t}} \triangleq \begin{bmatrix} \boldsymbol{0} & -\varsigma_n \\ \varsigma_n & \boldsymbol{0} \end{bmatrix} \tag{84}$$

$$\boldsymbol{\Xi}_{n,M}^{\boldsymbol{B}_{R,t}} \triangleq \begin{bmatrix} & & & & & \boldsymbol{0}_{2N\tau\times2\tau} & & \\ 0 & \cdots & 0 & 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{B}_{R,t}}\boldsymbol{g}_{M,1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{B}_{R,t}}\boldsymbol{g}_{M,N} & \cdots & 0 \end{bmatrix} \tag{85}$$

$$\boldsymbol{\Xi}_{n,m}^{\boldsymbol{B}_{R,t}} \triangleq \begin{bmatrix} & & & & & & \boldsymbol{0}_{2N\tau\times2(M-1)\tau} & & & & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{B}_{R,t}}\boldsymbol{g}_{m,1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{B}_{R,t}}\boldsymbol{g}_{m,N} & \cdots & 0 \end{bmatrix} \qquad m = 1, \ldots, M-1 \tag{86}$$

$$\boldsymbol{\Lambda}_n^{\boldsymbol{B}_{R,t}} \triangleq \begin{bmatrix} \boldsymbol{0} & -\varsigma_n \\ \varsigma_n & \boldsymbol{0} \end{bmatrix} \tag{87}$$

$$\boldsymbol{\Xi}_{n,M}^{\boldsymbol{B}_{I,t}} \triangleq \begin{bmatrix} & & & & & \boldsymbol{0}_{2N\tau\times2\tau} & & \\ 0 & \cdots & 0 & 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{B}_{I,t}}\boldsymbol{g}_{M,1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{B}_{I,t}}\boldsymbol{g}_{M,N} & \cdots & 0 \end{bmatrix} \tag{88}$$

$$\boldsymbol{\Xi}_{n,m}^{\boldsymbol{B}_{I,t}} \triangleq \begin{bmatrix} & & & & & & \boldsymbol{0}_{2N\tau\times2(M-1)\tau} & & & & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{B}_{I,t}}\boldsymbol{g}_{m,1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & \boldsymbol{\Lambda}_n^{\boldsymbol{B}_{I,t}}\boldsymbol{g}_{m,N} & \cdots & 0 \end{bmatrix} \qquad m = 1, \ldots, M-1 \tag{89}$$

Rearranging (93), we now have

$$
\sum_{i \in \Omega} \frac{|\boldsymbol{\Sigma}^{-\frac{1}{2}}|}{(2\pi)^{NT}} \int \left\{ \nabla_{\boldsymbol{\theta}} \left[1 - \gamma(\boldsymbol{y}_2, \boldsymbol{x}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) \right] \right\}
$$

$$
\times \exp \left\{ -\frac{1}{2} \left[\boldsymbol{y}_2 - (\boldsymbol{G} + \widetilde{\boldsymbol{G}} \boldsymbol{H}) \boldsymbol{\Phi} \boldsymbol{x}_i \right]^T \right.
$$

$$
\left. \times \boldsymbol{\Sigma}^{-1} \left[\boldsymbol{y}_2 - (\boldsymbol{G} + \widetilde{\boldsymbol{G}} \boldsymbol{H}) \boldsymbol{\Phi} \boldsymbol{x}_i \right] \right\} d\boldsymbol{y}_2
$$

$$
= \left| \boldsymbol{\Sigma}^{-\frac{1}{2}} \right| |\boldsymbol{\Sigma}|^{\frac{1}{2}} \sum_{i \in \Omega} \int \left\{ \nabla_{\boldsymbol{\theta}} \left[1 - \gamma(\boldsymbol{y}_2, \boldsymbol{x}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) \right] \right\}
$$

$$
\times p(\boldsymbol{y}_2 | \boldsymbol{x}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) d\boldsymbol{y}_2
$$

$$
= \sum_{i \in \Omega} \int \int \left\{ \nabla_{\boldsymbol{\theta}} \left[1 - \gamma(\boldsymbol{y}_2, \boldsymbol{x}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) \right] \right\}
$$

$$
\times p(\boldsymbol{y}_1, \boldsymbol{y}_2 | \boldsymbol{x}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) d\boldsymbol{y}_1 d\boldsymbol{y}_2
$$

$$
= \sum_{i \in \Omega} \int \int \left\{ \nabla_{\boldsymbol{\theta}} \left[1 - \gamma(\boldsymbol{y}_2, \boldsymbol{x}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) \right] \right\} \mathcal{P}_1 \mathcal{P}_2 d\boldsymbol{y}_1 d\boldsymbol{y}_2. \tag{94}
$$

Thus, it suffices for us to show (91) to complete the proof.

Recall the system model in (21). With the definition of $\gamma(\cdot)$ in (26), signal models given in (10) and (19), and the ML detection rule, we have

$$
1 - \gamma(\boldsymbol{z}, \boldsymbol{x}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta})
$$

$$
= \mathbb{I}(\hat{\boldsymbol{x}} = \boldsymbol{x}_i | \boldsymbol{z}, \boldsymbol{x} = \boldsymbol{x}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta})
$$

$$
= \prod_{j \in \Omega, j \neq i} \mathbb{I} \left\{ \| \boldsymbol{z} - \boldsymbol{B} \boldsymbol{x}_j \|^2 \geq \| \boldsymbol{z} - \boldsymbol{B} \boldsymbol{x}_i \|^2 \right\}
$$

$$
= \prod_{j \in \Omega, j \neq i} \mathbb{I} \left\{ \underbrace{2(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{B}^T}_{\boldsymbol{a}_{i|j}^T} \boldsymbol{z} - \underbrace{(\| \boldsymbol{B} \boldsymbol{x}_i \|^2 - \| \boldsymbol{B} \boldsymbol{x}_j \|^2)}_{b_{i|j}} \geq 0 \right\}. \tag{95}
$$

Using (31), (32), and (34), we have

$$
\sum_{i \in \Omega} \int \nabla_{\boldsymbol{\theta}} \left[1 - \gamma(\boldsymbol{z}, \boldsymbol{x}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}) \right] \mathcal{P}_z d\boldsymbol{z}
$$

$$
= -\pi^{-NT - (M-1)\tau} \int \int \sum_{i \in \Omega} \sum_{j \in \Omega, j \neq i} \left[\boldsymbol{a}_{i|j}^T (\nabla_{\boldsymbol{\theta}} \boldsymbol{B}^T \boldsymbol{z}) - \nabla_{\boldsymbol{\theta}} b_{i|j} \right]
$$

$$
\times \delta \left(\| \boldsymbol{z} - \boldsymbol{B} \boldsymbol{x}_j \| - \| \boldsymbol{z} - \boldsymbol{B} \boldsymbol{x}_i \| \right)
$$

$$
\times \prod_{l \in \Omega, l \neq j} \mathbb{I} \left(\| \boldsymbol{z} - \boldsymbol{B} \boldsymbol{x}_l \|^2 \geq \| \boldsymbol{z} - \boldsymbol{B} \boldsymbol{x}_j \|^2 \right)
$$

$$
\times \exp \left\{ - \| \boldsymbol{z} - \boldsymbol{B} \boldsymbol{x}_i \|^2 \right\} d\boldsymbol{z} \tag{96}
$$

where $\delta(\cdot)$ denotes the delta function, and we relied on the chain rule for the derivative of the indicator function. Since $\boldsymbol{a}_{i|j} = -\boldsymbol{a}_{j|i}$ and $\nabla_{\boldsymbol{\theta}} b_{i|j} = -\nabla_{\boldsymbol{\theta}} b_{j|i}$, (96) becomes zeros by summing the corresponding (i, j) and (j, i) terms.

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wirel. Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[3] A. Nosratinia and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[4] T. M. Cover and A. A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[5] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity— Part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[6] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity— Part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[7] B. Vucetic and J. Yuan, *Space–Time Coding.* Hoboken, NJ: Wiley, 2003.

[8] S. M. Alamouti, "A simple transmit diversity scheme for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[9] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: Performance limits and space–time signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[10] P. A. Anghel, G. Leus, and M. Kaveh, "Multi-user space–time coding in cooperative networks," in *Proc. ICASSP*, Apr. 2003, vol. 4, pp. 73–76.

[11] P. Herhold, E. Zimmermann, and G. Fettweis, "On the performance of cooperative amplify-and-forward relay networks," in *Proc. 5th Int. ITG Conf. Source Channel Coding*, Jan. 2004, pp. 451–458.

[12] I. Maric and R. Yates, "Forwarding strategies for Gaussian parallel-relay networks," in *Proc. CISS*, Mar. 2004, p. 269.

[13] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1804–1824, Jul. 2002.

[14] R. W. Heath and A. J. Paulraj, "Linear dispersion codes for MIMO systems based on frame theory," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2429–2441, Oct. 2002.

[15] Y. Jing and B. Hassibi, "Wireless networks, diversity and space–time codes," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2004, pp. 463–468.

[16] M. Dohler, A. Gkeilas, and H. Aghvami, "Resource allocation for FDMA-based regenerative multi-hop links," *IEEE Trans. Commun.*, vol. 3, no. 6, pp. 1989–1993, Nov. 2004.

[17] J. Spall, *Introduction to Stochastic Search and Optimization.* Hoboken, NJ: Wiley, 2003.

[18] T. S. Rappaport, *Wireless Communications: Principles and Practice.* Englewood Cliffs, NJ: Prentice-Hall, 1996.

[19] A. M. Chan and I. Lee, "A new reduced-complexity sphere decoder for multiple antenna systems," in *Proc. ICC*, Apr. 2002, vol. 1, pp. 460–464.

[20] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in lattice, including a computational complexity," *Math. Comput.*, vol. 44, pp. 463–471, Apr. 1995.

[21] I. Berenguer and X. Wang, "Space–time coding and signal processing for MIMO communications," *J. Comput. Sci. Technol.*, vol. 18, no. 6, pp. 689–702, Nov. 2003.

[22] H. J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications.* New York: Springer-Verlag, 2003.

[23] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space–time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[24] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.
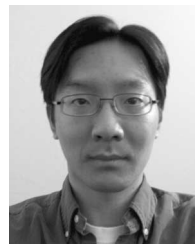
Kuo-ching Liang received the B.S. degree in electrical engineering and computer science from the University of California, Berkeley, in 1999, the M.Eng. degree in electrical engineering from Cornell University, Ithaca, NY, in 2000, and the M.S. degree in electrical engineering from the University of Pennsylvania, Philadelphia, in 2003. He is currently working toward the Ph.D. degree at the Department of Electrical Engineering, Columbia University, New York, NY.

His research interests are in the areas of digital communication and statistical signal processing and its applications in the area of computational biology and bioinformatics.

Xiaodong Wang (S'98–M'98–SM'04) received the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ.

He is currently with the faculty of the Department of Electrical Engineering, Columbia University, New York, NY. His research interests are in the general areas of computing, signal processing, and communications, and he has published extensively in these areas. Among his publications is *Wireless Communication Systems: Advanced Techniques for Signal Reception* (Prentice-Hall, 2003). His current research interests include wireless communications, statistical signal processing, and genomic signal processing.

Dr. Wang received the 1999 NSF CAREER Award and the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award. He served as an Associate Editor for the IEEE Transactions on Communications, the IEEE Transactions on Wireless Communications, the IEEE Transactions on Signal Processing, and the IEEE Transactions on Information Theory.
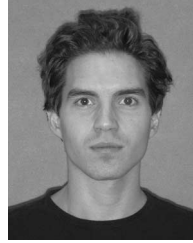
Inaki Berenguer received the Master's and Ph.D. degrees in engineering from Cambridge University, Cambridge, U.K., in 2002 and 2004, respectively, and was a two-year graduate Fulbright Scholar at Columbia University, New York, NY.

From July 1999 to January 2001, he was with Hewlett Packard, Brussels, Belgium. In January 2001, he was with the Wireless Division of STMicroelectronics, San Diego, CA, where he worked until August 2001. In Summer 2004, he worked for Intel Corporation, and from September to December 2004 and in Summer 2005, he was a Research Associate at NEC Laboratories America, Princeton, NJ. Since December 2005, he has been with McKinsey & Company. He is the recipient of the British Council, La Caixa, Cambridge-MIT Institute, and Leonardo Da Vinci Fellowships.