Channel Estimation and Training Design for Two-Way Relay Networks with Power Allocation

Bin Jiang, *Member, IEEE,* Feifei Gao, *Member, IEEE,* Xiqi Gao, *Senior Member, IEEE,* and Arumugam Nallanathan, *Senior Member, IEEE*

***Abstract*—In this paper, we propose a new channel estimation prototype for the *amplify-and-forward* (AF) two-way relay network (TWRN). By allowing the relay to first estimate the channel parameters and then allocate the powers for these parameters, the final data detection at the source terminals could be optimized. Specifically, we consider the classical three-node TWRN where two source terminals exchange their information via a single relay node in between and adopt the maximum likelihood (ML) channel estimation at the relay node. Two different power allocation schemes to the training signals are then proposed to maximize the average effective signal-to-noise ratio (AESNR) of the data detection and minimize the mean-square-error (MSE) of the channel estimation, respectively. The optimal/sub-optimal training designs for both schemes are found as well. Simulation results corroborate the advantages of the proposed technique over the existing ones.**

***Index Terms*—Channel estimation, amplify-and-forward, two-way relay networks, power allocation, training design.**

I. Introduction

EMPLOYING the relay node to assist the bidirectional communications between two source terminals has gained increasing attention recently [1]–[3]. In this scheme, the overall transmission period is divided into two phases. During phase I, both source terminals transmit and the relay receives a superposed signal. In phase II, relay directly broadcasts the overlapped signal to both terminals. From a *network coding*-like manner, the source terminals are still capable of extracting the desired signal. The so derived mechanism is proved effective in terms of enhancing the system throughput compared to its unidirectional counterpart [4], [5], and is then named as two-way relay network (TWRN).

Following this guideline, the capacity analysis and the achievable rate region for *amplify-and-forward* (AF) and *decode-and-forward* (DF) based TWRN were studied in [6]–[13]. In [14] the optimal mapping function at the relay node that minimizes the transmission bit-error rate (BER) was proposed, whose expression is unfortunately hard to implement for practical scenarios. For cases when the channel knowledge is known at the transmitter side, the optimal beamforming at the multi-antenna relay that maximizes the capacity of AF-based TWRN was designed in [15] and [16], while the suboptimal subcarrier sorting and power allocation at the relay were proposed in [17]. When the channel knowledge is only available at the receiver side, the space-time coding (STC) technique that is able to improve the transmission reliability has been developed in [18].

Since most works [6]–[18] assume perfect channel knowledge, the way to obtain the accurate channel information becomes crucial. The earliest attempt was recently made in [19], where two different channel estimation algorithms as well as the corresponding optimal training sequences were derived for an AF based three-node TWRN. However, only the simplest scaling operation at the relay node is assumed in [19] while the relay function is not optimized. A more practical consideration is to equip the relay node with advanced signal processing techniques, which, in fact, has already been used by many existing works, e.g., beamforming [12], [15], [16], resource allocation [17], STC design [18] etc.. Therefore, it is also reasonable to assume that the relay node is capable of performing the channel estimation itself. Letting the relay node estimate the channels is also indispensable as is seen from the following aspects: i) DF based relay network requires the channel knowledge at the relay node; ii) The advanced relay operations, e.g., [12], [15]–[17], need the channel knowledge at the relay node.

In this work, we introduce a new prototype for channel estimation in TWRN that the relay node first estimates the channels from the two terminals and then allocates the power to different channel components during a training-phase, such that the final data detection or channel estimation at the terminal nodes can be optimized. This is a unique property for TWRN and does not exist for unidirectional relay network [4], [5]. As one of the earliest works on the channel estimation for TWRN, we will choose the flat fading assumption to present our main idea, while the extension to the frequency selective channels can be easily made if the orthogonal frequency division multiplex (OFDM) modulation is adopted. We will study the classical three-node TWRN and choose two different optimization objectives, i.e., maximizing the average effective

Manuscript received June 10, 2009; revised September 13, 2009 and February 7, 2010; accepted April 1, 2010. The associate editor coordinating the review of this paper and approving it for publication was G. Abreu.

B. Jiang and X. Q. Gao are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: {bjiang, xqgao@seu.edu.cn).

F. Gao is with the School of Engineering and Science, Jacobs University, Bremen, Germany (e-mail: feifeigao@ieee.org).

A. Nallanathan is with the Department of Electronic Engineering, King's College London, Strand Campus, London WC2R 2LS, United Kingdom (e-mail: nallanathan@ieee.org).

The work of B. Jiang and X. Q. Gao was supported by the National Science Fund for Distinguished Young Scholars (Grant No. 60925004), the National Natural Science Foundation of China (Grant No. 60902009), and the National Science and Technology Major Project of China (Grants No. 2009ZX03003-005 and 2008ZX03003-005). The work of F. Gao was supported in part by the German Research Foundation (DFG) under Grant GA 1654/1-1. Part of this paper has been presented at IEEE ICC, June 14–18 2009, Dresden, Germany.

Digital Object Identifier 10.1109/TWC.2010.06.090870

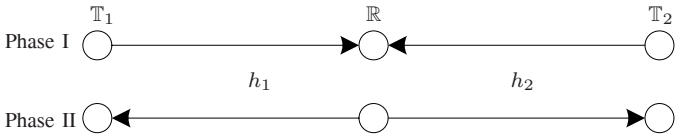

Fig. 1. A typical two-way relay network with two source nodes and one relay node working under the AF relaying scheme.

signal-to-noise ratio (AESNR) and minimizing the channel estimation mean-square errors (MSE) at the source terminals. The same principle can be straightforwardly extended to different objectives, e.g., minimizing the bit error rate (BER), minimizing the outage probability, etc.. Note that power allocation will not be used during the data transmission since the relay node cannot differentiate the signals from two terminals without multiple access technique.

The rest of the paper is organized as follows. Section II presents the system model of the TWRN as well as the channel estimation at the relay node. In Section III and Section IV, we derive the optimal power allocation as well as the training design to maximize the detection AESNR and to minimize the estimation MSE at both source terminals, respectively. Simulation results are then provided in Section V. Finally, conclusions are drawn in Section VI.

Notations: Vectors and matrices are boldface small and capital letters, respectively; the transpose, complex conjugate, Hermitian, and inverse of $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$ and $\mathbf{A}^{-1}$, respectively; $\mathrm{tr}(\mathbf{A})$ represents the trace of $\mathbf{A}$; $\|\mathbf{a}\|$ denotes the two-norm of $\mathbf{a}$, and $\mathrm{diag}\{\mathbf{a}\}$ is a diagonal matrix whose diagonal elements are given by the elements of $\mathbf{a}$; $\mathbf{I}$ is the identity matrix, and $\mathrm{E}\{\cdot\}$ denotes the statistical expectation.

II. Problem Formulation

A. Data Transmission in TWRN

A classical three-node TWRN with two source terminals $\mathbb{T}_1$, $\mathbb{T}_2$, and one relay node $\mathbb{R}$ is shown in Fig. 1. The baseband channel between $\mathbb{T}_i$, $i = 1, 2$, and $\mathbb{R}$ is assumed as circularly symmetric complex Gaussian and is denoted by $h_i \in \mathcal{CN}(0, \sigma_i^2)$. Assuming time division duplex (TDD) mode, the channel is reciprocal such that the channel from $\mathbb{R}$ to $\mathbb{T}_i$ is also described by h_i. Since $\mathbb{T}_i$'s are separated from each other, h_1 and h_2 are considered independent. Moreover, the average transmission powers of $\mathbb{T}_1$, $\mathbb{T}_2$, and $\mathbb{R}$ are denoted by P_1, P_2, P_r, respectively.

One round of data communication between $\mathbb{T}_i$'s is divided into two phases. During phase I, $\mathbb{T}_i$ sends out d_i and $\mathbb{R}$ receives

$$r = h_1 d_1 + h_2 d_2 + n_r, \tag{1}$$

where n_r is the additive white Gaussian noise with variance σ_n^2. A typical scaling factor

$$\kappa_D = \sqrt{\frac{P_r}{P_1\sigma_1^2 + P_2\sigma_2^2 + \sigma_n^2}} \tag{2}$$

is adopted at $\mathbb{R}$ to keep the average power of $\mathbb{R}$ as P_r. During phase II, $\mathbb{R}$ broadcasts the scaled signal, and $\mathbb{T}_1$ and $\mathbb{T}_2$ receive

$$y_1 = \kappa_D h_1^2 d_1 + \kappa_D h_1 h_2 d_2 + \kappa_D h_1 n_r + n_1, \tag{3}$$

$$y_2 = \kappa_D h_2^2 d_2 + \kappa_D h_1 h_2 d_1 + \kappa_D h_2 n_r + n_2, \tag{4}$$

respectively, where n_i, $i = 1, 2$, represents the noise at $\mathbb{T}_i$ and is assumed to have the same distribution as n_r.

Taking $\mathbb{T}_1$ for example, it has been shown in [19] that the maximum likelihood (ML) data detection only relies on the parameters $a \triangleq h_1^2$ and $b \triangleq h_1 h_2$. Moreover, it can be verified that $\mathrm{E}\{|a|^2\} = 2\sigma_1^4$ and $\mathrm{E}\{|b|^2\} = \sigma_1^2\sigma_2^2$, respectively.

B. Channel Estimation in TWRN

In order to be compatible with the data transmission, it is preferred that channel estimation is accomplished from a two-phase training. Denote the training sequence from $\mathbb{T}_1$ and $\mathbb{T}_2$ as $\mathbf{t}_1$ and $\mathbf{t}_2$, respectively, both with length N. The training signal received at $\mathbb{R}$ during phase I is

$$\mathbf{r} = h_1\mathbf{t}_1 + h_2\mathbf{t}_2 + \mathbf{n}_r = \mathbf{T}\mathbf{h} + \mathbf{n}_r, \tag{5}$$

where $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2]$, $\mathbf{h} = [h_1, h_2]^T$, and $\mathbf{n}_r$ is an $N \times 1$ noise vector. We adopt the ML channel estimator at $\mathbb{R}$,[1] and channels are obtained from [20]

$$\hat{\mathbf{h}} = (\mathbf{T}^H\mathbf{T})^{-1}\mathbf{T}^H\mathbf{r} = \mathbf{h} + (\mathbf{T}^H\mathbf{T})^{-1}\mathbf{T}^H\mathbf{n}_r. \tag{6}$$

We propose that $\mathbb{R}$ broadcasts the following training signals in the second phase:

$$\tilde{\mathbf{r}} = \kappa_T(\sqrt{\alpha}\hat{h}_1\mathbf{t}_1 + \sqrt{\beta}\hat{h}_2\mathbf{t}_2) = \kappa_T\mathbf{T}\mathbf{\Lambda}\hat{\mathbf{h}}, \tag{7}$$

where α and β are the weights that control the power allocated to the estimated channels $\hat{h}_1$ and $\hat{h}_2$, $\mathbf{\Lambda} = \mathrm{diag}\{\sqrt{\alpha}, \sqrt{\beta}\}$, and κ_T is the scaling factor to keep the relay power P_r during the training. Substituting (6) into (7) gives

$$\tilde{\mathbf{r}} = \kappa_T(\mathbf{T}\mathbf{\Lambda}\mathbf{h} + \tilde{\mathbf{n}}_r), \tag{8}$$

where $\tilde{\mathbf{n}}_r = \mathbf{T}\mathbf{\Lambda}(\mathbf{T}^H\mathbf{T})^{-1}\mathbf{T}^H\mathbf{n}_r$ is the residual noise.

Remark 1: The signal processing technique in (7) or (8) is named as *denoising*, which is known to reduce the effective noise power. For example, when $\alpha = \beta = 1$, then $\sqrt{\alpha}\hat{h}_1\mathbf{t}_1 + \sqrt{\beta}\hat{h}_2\mathbf{t}_2$ reduces to $h_1\mathbf{t}_1 + h_2\mathbf{t}_2 + \mathbf{P}_T\mathbf{n}_r$, where $\mathbf{P}_T = \mathbf{T}(\mathbf{T}^H\mathbf{T})^{-1}\mathbf{T}^H$ is the projection matrix of the space spanned by $\mathbf{T}$. Clearly, the total noise power reduces by a factor of $N/2$ after the projection, while the signal components do not change.

Denote $Q_i = \mathbf{t}_i^H\mathbf{t}_i$ as the power spent for training at $\mathbb{T}_i$ that is not necessarily the same as NP_i.[2] Define the correlation factor between $\mathbf{t}_1$ and $\mathbf{t}_2$ as $\rho = \mathbf{t}_1^H\mathbf{t}_2/\sqrt{Q_1Q_2}$. From the straightforward computation, the total noise power is obtained as

$$\mathrm{E}\{\tilde{\mathbf{n}}_r^H\tilde{\mathbf{n}}_r\} = \underbrace{\frac{\alpha + \beta - 2|\rho|^2\sqrt{\alpha\beta}}{1 - |\rho|^2}}_{\varphi(\alpha,\beta,|\rho|^2)}\sigma_n^2, \tag{9}$$

[1]Other estimators can also be applied but only changes the following mathematical derivation.

[2]See [21] for optimal power balancing between training and data transmission.

where $\varphi(\alpha,\beta,|\rho|^2)$ is defined as the corresponding term. Since $\alpha+\beta \geq 2\sqrt{\alpha\beta}$, (9) is a monotonically increasing function of $|\rho|$ whose maximum goes to infinity when $|\rho|=1$. Intuitively, the training correlation should be chosen as $|\rho|=0$, i.e., orthogonal training, since it minimizes the noise power sent from $\mathbb{R}$. Nonetheless, we will not make this assumption here, whereas a more general discussion on ρ will be presented later.

From (8) to (9), κ_T can be explicitly written as

$$\kappa_T = \sqrt{\frac{NP_r}{\alpha Q_1\sigma_1^2+\beta Q_2\sigma_2^2+\varphi(\alpha,\beta,|\rho|^2)\sigma_n^2}}. \tag{10}$$

Without loss of generality, we can normalize α and β as

$$\alpha+\beta=1, \quad 0\leq\alpha\leq 1, \tag{11}$$

and hence only the design over α is needed.

In phase II, the received training signal at $\mathbb{T}_1$ is

$$\mathbf{z}_1 = \tilde{\mathbf{r}}h_1+\mathbf{n}_1, \tag{12}$$

where $\mathbf{n}_1$ is an $N\times 1$ noise vector. By substituting (8) into (12), $\mathbf{z}_1$ can be re-expressed as

$$\mathbf{z}_1 = \kappa_T\sqrt{\alpha}a\mathbf{t}_1+\kappa_T\sqrt{\beta}b\mathbf{t}_2+\tilde{\mathbf{n}} = \kappa_T\mathbf{T}\mathbf{\Lambda}\mathbf{h}_e+\tilde{\mathbf{n}}, \tag{13}$$

where $\mathbf{h}_e=[a,b]^T$ is the equivalent channel vector, and $\tilde{\mathbf{n}}=\kappa_T h_1\tilde{\mathbf{n}}_r+\mathbf{n}_1$ is the equivalent noise vector with covariance matrix

$$\mathbf{R}_{\tilde{\mathbf{n}}} = \mathrm{E}\{\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H\} = \sigma_n^2\left(\mathbf{I}_N+\kappa_T^2\sigma_1^2\mathbf{T}\mathbf{\Lambda}(\mathbf{T}^H\mathbf{T})^{-1}\mathbf{\Lambda}\mathbf{T}^H\right). \tag{14}$$

III. Linear Maximum Signal-to-Noise Ratio Channel Estimation

In this section, we consider a linear channel estimator that is able to maximize the AESNR, which is named as *linear maximum signal-to-noise ratio* (LMSNR) channel estimator. The linear estimator is adopted due to its easy implementation.

A. Channel Estimation

Due to symmetry, we only discuss for the terminal $\mathbb{T}_1$. Suppose that a and b are obtained from two linear estimators

$$\hat{a} = \mathbf{u}^H\mathbf{z}_1, \tag{15}$$

$$\hat{b} = \mathbf{v}^H\mathbf{z}_1, \tag{16}$$

respectively. Define $\Delta_a=\hat{a}-a$ and $\Delta_b=\hat{b}-b$ as the channel estimation errors. During the data transmission, the received signal at $\mathbb{T}_1$ is rewritten as

$$y_1 = \kappa_D\hat{a}d_1+\kappa_D\hat{b}d_2\underbrace{-\kappa_D\Delta_b d_2-\kappa_D\Delta_a d_1+\kappa_D h_1 n_r+n_1}_{\tilde{n}} \tag{17}$$

where $\tilde{n}$ is the equivalent noise at $\mathbb{T}_1$. After subtracting the self-signal component $\kappa_D\hat{a}d_1$, the AESNR of the detection can be expressed as (18), shown at the top of the next page, where

$$\begin{aligned}\mathbf{R}_z &= \mathrm{E}\{\mathbf{z}_1\mathbf{z}_1^H\} \\ &= \sigma_n^2\left(\mathbf{I}_N+C_1\mathbf{t}_1\mathbf{t}_1^H+C_2\mathbf{t}_2\mathbf{t}_2^H+C_3\mathbf{t}_1\mathbf{t}_2^H+C_3^*\mathbf{t}_2\mathbf{t}_1^H\right),\end{aligned} \tag{19}$$

with

$$C_1 = \kappa_T^2\alpha\left(\frac{\sigma_a^2}{\sigma_n^2}+\frac{\sigma_1^2}{(1-|\rho|^2)Q_1}\right),$$

$$C_2 = \kappa_T^2\beta\left(\frac{\sigma_b^2}{\sigma_n^2}+\frac{\sigma_1^2}{(1-|\rho|^2)Q_2}\right),$$

$$C_3 = -\kappa_T^2\sqrt{\alpha\beta}\frac{\rho\sigma_1^2}{(1-|\rho|^2)\sqrt{Q_1Q_2}}.$$

Remark 2: Since the effective noise in (17) is non-Gaussian and is correlated with the signal component, maximizing AESNR does not strictly corresponds to maximizing the data throughput but corresponds to maximizing a lower bound of the data throughput.[3] Note that this is a typical way of incorporating the channel estimation errors into the system performance [21].

Following the same approach in [19] where the simple optimization is applied [22], the optimal $\mathbf{u}$ and $\mathbf{v}$ that maximize $\bar{\gamma}$ can be derived as

$$\mathbf{u} = \kappa_T\sqrt{\alpha}\sigma_a^2\mathbf{R}_z^{-1}\mathbf{t}_1, \tag{20}$$

$$\mathbf{v} = \eta\kappa_T\sqrt{\beta}\sigma_b^2\mathbf{R}_z^{-1}\mathbf{t}_2, \tag{21}$$

where

$$\eta = \frac{F_0-\psi_1(\alpha,\beta,|\rho|)}{\psi_2(\alpha,\beta,|\rho|)}, \tag{22}$$

with

$$\begin{aligned}F_0 &= P_1\sigma_a^2+P_2\sigma_b^2+\sigma_1^2\sigma_n^2+\sigma_n^2/\kappa_D^2,\\ \psi_1(\alpha,\beta,|\rho|) &= P_1\sigma_a^4\kappa_T^2\alpha\mathbf{t}_1^H\mathbf{R}_z^{-1}\mathbf{t}_1,\\ \psi_2(\alpha,\beta,|\rho|) &= P_2\sigma_b^4\kappa_T^2\beta\mathbf{t}_2^H\mathbf{R}_z^{-1}\mathbf{t}_2.\end{aligned}$$

The corresponding AESNR can then be reexpressed as

$$\bar{\gamma} = \frac{\eta}{\eta-1}, \tag{23}$$

and channel estimation MSEs are

$$e_a = \mathrm{E}\{|a-\hat{a}|^2\} = \sigma_a^2-\kappa_T^2\alpha\sigma_a^4\mathbf{t}_1^H\mathbf{R}_z^{-1}\mathbf{t}_1, \tag{24}$$

$$e_b = \mathrm{E}\{|b-\hat{b}|^2\} = \sigma_b^2-\left(2\eta-\eta^2\right)\kappa_T^2\beta\sigma_b^4\mathbf{t}_2^H\mathbf{R}_z^{-1}\mathbf{t}_2. \tag{25}$$

B. Optimal Training Design

The optimal training sequences that maximize $\bar{\gamma}$ are characterized by the following proposition.

Proposition 1: If the received SNR at the relay during the training is greater than or equal to -3 dB, i.e., $10\log\left(Q_i\sigma_i^2/\sigma_n^2\right)\geq -3$, $i=1,2$, then the optimal training sequences $\mathbf{t}_1$ and $\mathbf{t}_2$ should be orthogonal, i.e., $\rho=0$.

Proof: See Appendix I. ∎

With the orthogonal training, $\mathbf{R}_z^{-1}\mathbf{t}_i$ reduces to $\mathbf{t}_i/\sigma_n^2(1+C_iQ_i)$, $i=1,2$, and the corresponding channel estimations can be expressed as

$$\hat{a} = \frac{\kappa_T\sqrt{\alpha}\sigma_a^2}{(\kappa_T^2\alpha\sigma_1^2+1)\sigma_n^2+\kappa_T^2\alpha\sigma_a^2Q_1}\mathbf{t}_1^H\mathbf{z}_1, \tag{26}$$

$$\hat{b} = \frac{\eta\kappa_T\sqrt{\beta}\sigma_b^2}{(\kappa_T^2\beta\sigma_1^2+1)\sigma_n^2+\kappa_T^2\beta\sigma_b^2Q_2}\mathbf{t}_2^H\mathbf{z}_1. \tag{27}$$

[3]When we assume that the noise component is Gaussian and is independent from the signal, the data throughput obtained from $\log(1+\text{AESNR})$ is a lower bound of its true value. Then maximizing AESNR could optimize this lower bound.

$$\bar{\gamma} = \frac{\kappa_D^2 \mathrm{E}\{|\hat{b}|^2\} P_2}{\mathrm{E}\{|\tilde{n}|^2\}} = \frac{\kappa_D^2 P_2 \mathbf{v}^H \mathbf{R}_z \mathbf{v}}{\kappa_D^2 \mathrm{E}\{|\mathbf{u}^H \mathbf{z}_1 - a|^2\} P_1 + \kappa_D^2 \mathrm{E}\{|\mathbf{v}^H \mathbf{z}_1 - b|^2\} P_2 + \kappa_D^2 \sigma_1^2 \sigma_n^2 + \sigma_n^2} \tag{18}$$

In practical communication systems, the condition that the received SNR is greater than $-3\mathrm{dB}$ is normally satisfied. Therefore, the optimal training sequences from $\mathbb{T}_i$'s should be orthogonal to each other, which is a consistent result as that in [19].

C. Optimal Power Allocation

The next step is to find the optimal power allocation factor α when $\rho = 0$. After some tedious calculation, we obtain

$$\psi_1(\alpha, \beta, 0) = \frac{F_3 \alpha}{F_1 \sigma_n^2 + F_2 \alpha}, \tag{28}$$

$$\psi_2(\alpha, \beta, 0) = \frac{F_6 \beta}{F_4 \sigma_n^2 + F_5 \beta}, \tag{29}$$

where

$$\begin{aligned} F_1 &= Q_2 \sigma_2^2 + \sigma_n^2, & F_3 &= N P_r P_1 Q_1 \sigma_a^4, \\ F_4 &= Q_1 \sigma_1^2 + \sigma_n^2, & F_6 &= N P_r P_2 Q_2 \sigma_b^4, \\ F_2 &= N P_r Q_1 \sigma_a^2 + \left(N P_r \sigma_1^2 + Q_1 \sigma_1^2 - Q_2 \sigma_2^2 \right) \sigma_n^2, \\ F_5 &= N P_r Q_2 \sigma_b^2 + \left(N P_r \sigma_1^2 + Q_2 \sigma_2^2 - Q_1 \sigma_1^2 \right) \sigma_n^2. \end{aligned}$$

Observing that $\bar{\gamma}$ is a function of a single variable α, the maximum must be obtained at one root of its first order derivative or at $\alpha = 0$ or $\alpha = 1$, regardless of whether $\bar{\gamma}$ is concave or not. One can immediately remove the choices of $\alpha = 0$ and $\alpha = 1$, which fail either the channel estimate of a or b. The derivative of $\bar{\gamma}$ with respect to α is given by (30), which is shown on the top of the next page, where

$$\begin{aligned} \omega_2 &= 1 - \frac{F_4 \left(F_0 F_2^2 - F_2 F_3 \right)}{F_1 F_3 F_5}, \\ \omega_1 &= -2 - \frac{2 F_0 F_2 F_4}{F_3 F_5} \sigma_n^2, \\ \omega_0 &= 1 + \frac{F_4}{F_5} \sigma_n^2 - \frac{F_0 F_1 F_4}{F_3 F_5} \sigma_n^4. \end{aligned}$$

Note that, the existence of two roots indicates that $\bar{\gamma}$ is neither a concave or a convex function if we do not restrict α in the region $[0, 1]$. Therefore, the optimal α is the root of

$$\omega_2 \alpha^2 + \omega_1 \alpha + \omega_0 = 0, \tag{31}$$

which is one of

$$\alpha_{\mathrm{opt}} = \frac{-\omega_1 \pm \sqrt{\omega_1^2 - 4 \omega_0 \omega_2}}{2 \omega_2}, \tag{32}$$

that gives the larger value of $\bar{\gamma}$ and at the same time belongs to $[0, 1]$. Due to the complicated structure of the roots, we cannot theoretically prove whether both roots reside in $[0, 1]$, neither could we know which one gives a larger $\bar{\gamma}$. Nonetheless in practical applications, one can simply check these two candidates and choose the one that yields the highest AESNR. We also find via numerous simulations that $\bar{\gamma}$ is a concave function of $\alpha \in [0, 1]$ and some examples will be presented in the simulation section.

Interestingly, when the transmit SNRs at both the two source terminals are high, there are

$$\lim_{\sigma_n^2 \to 0^+} \omega_2 = 1 - \tau^2, \qquad \lim_{\sigma_n^2 \to 0^+} \omega_1 = -2, \qquad \lim_{\sigma_n^2 \to 0^+} \omega_0 = 1.$$

In this case, we derive the suboptimal power allocation factor

$$\alpha_{\mathrm{sub-opt}} = \lim_{\sigma_n^2 \to 0^+} \alpha_{\mathrm{opt}} = \frac{1}{1 + \tau} \in [0, 1], \tag{33}$$

where $\tau = \frac{Q_1 \sigma_1 \sqrt{P_2}}{Q_2 \sigma_2 \sqrt{P_1}}$ is defined as the effective ratio controller (ERC). This suboptimal power allocation factor is easy to implement, and it demonstrates good performance even at relatively lower SNR region, as will be seen in the later simulation part.

Some insights can be gained from the suboptimal power allocation factor $\alpha_{\mathrm{sub-opt}}$:

- If the two channel links are well balanced in terms of $\frac{Q_1 \sigma_1}{\sqrt{P_1}} = \frac{Q_2 \sigma_2}{\sqrt{P_2}}$, then $\alpha_{\mathrm{sub-opt}} = 0.5$, i.e., equal power allocation. Obviously, whether the link is balanced depends on three different factors, i.e., training power, data power, as well as the channel quality. For example in the recently announced LTE-Advanced specifications [23], the relay is adopted between base stations (BS) and mobile stations (MS). Naturally, it is expected powers of BS and MS are not the same. Depending on the positions of relay and MS (especially when the latter is moving), the channel links are also expected to be different. All these facts corroborates the necessity of the proposed study.
- Through the simulations, we find that the optimal power allocation almost simultaneously maximize the AESNR at $\mathbb{T}_1$ and $\mathbb{T}_2$, especially when SNR is high. This can be explained using sub-optimal power allocation that is near optimal at high SNR region. Let $\alpha'_{\mathrm{sub-opt}}$ and $\beta'_{\mathrm{sub-opt}}$ be the suboptimal power allocation to $\mathbf{t}_1$ and $\mathbf{t}_2$ such that the AESNR at $\mathbb{T}_2$ is maximized. By switching the variables in the derivation of (33), we can easily obtain

$$\beta'_{\mathrm{sub-opt}} = \frac{1}{1 + \frac{1}{\tau}} = \frac{\tau}{1 + \tau}, \tag{34}$$

  which is the same as $\beta_{\mathrm{sub-opt}}$ used to maximize AESNR at $\mathbb{T}_1$.

IV. Linear Minimum Mean-Square-Error Channel Estimation

In this section, we study another popular type of channel estimator, i.e., linear minimum mean-square-error (LMMSE) channel estimator, as well as its corresponding power allocation.

A. Channel Estimation

From (13), the linear MMSE estimation of $\mathbf{h}_e$ and the resultant MSE can be immediately derived from the standard

$$\frac{\partial\bar{\gamma}}{\partial\alpha}=\frac{F_1F_3F_5F_6\sigma_n^2\left(\omega_2\alpha^2+\omega_1\alpha+\omega_0\right)}{\left(\left(F_0-\psi_1(\alpha,\beta,0)-\psi_2(\alpha,\beta,0)\right)\left(F_1\sigma_n^2+F_2\alpha\right)\left(F_4\sigma_n^2+F_5\beta\right)\right)^2} \tag{30}$$

approach [24] as:

$$\hat{\mathbf{h}}_e=\kappa_T\mathbf{R}_{\mathbf{h}_e}\mathbf{\Lambda}\mathbf{T}^H\left(\kappa_T^2\mathbf{T}\mathbf{\Lambda}\mathbf{R}_{\mathbf{h}_e}\mathbf{\Lambda}\mathbf{T}^H+\mathbf{R}_{\tilde{\mathbf{n}}}\right)^{-1}\mathbf{z}_1, \tag{35}$$

and

$$\sigma^2_{\hat{\mathbf{h}}_e}=\mathrm{tr}\left\{\left(\mathbf{R}^{-1}_{\mathbf{h}_e}+\kappa_T^2\mathbf{\Lambda}\mathbf{T}^H\mathbf{R}^{-1}_{\tilde{\mathbf{n}}}\mathbf{T}\mathbf{\Lambda}\right)^{-1}\right\}, \tag{36}$$

where $\mathbf{R}_{\mathbf{h}_e}=\mathrm{E}\{\mathbf{h}_e\mathbf{h}_e^H\}$ is the covariance matrices of $\mathbf{h}_e$. Since $\mathrm{E}\{ab^*\}=0$, it can be computed that $\mathbf{R}_{\mathbf{h}_e}=\mathrm{diag}\{\sigma_a^2,\sigma_b^2\}$.

From (14) and the matrix inversion lemma, $\sigma^2_{\hat{\mathbf{h}}_e}$ can be re-expressed as

$$\sigma^2_{\hat{\mathbf{h}}_e}=\sigma_n^2\mathrm{tr}\left\{\left(\sigma_n^2\mathbf{R}^{-1}_{\mathbf{h}_e}+\mathbf{D}^{-1}\right)^{-1}\right\}, \tag{37}$$

where $\mathbf{D}$ is a 2×2 matrix given by

$$\mathbf{D}=\kappa_T^{-2}\left(\mathbf{\Lambda}\mathbf{T}^H\mathbf{T}\mathbf{\Lambda}\right)^{-1}+\sigma_1^2\left(\mathbf{T}^H\mathbf{T}\right)^{-1}. \tag{38}$$

The (i,j)-th entries of $\mathbf{D}$, i.e., $d_{i,j}$, can be expressed as

$$d_{11}=\frac{\left(A_3+\alpha\sigma_1^2\right)x+A_4}{\alpha Q_1x^2}, \tag{39}$$

$$d_{22}=\frac{\left(A_3+\beta\sigma_1^2\right)x+A_4}{\beta Q_2x^2}, \tag{40}$$

$$d_{12}=d_{21}^*=\frac{\rho\big((A_3+\sqrt{\alpha\beta}\sigma_1^2)x+A_4\big)}{\sqrt{\alpha\beta Q_1Q_2}x^2}, \tag{41}$$

where $x\triangleq 1-|\rho|^2$, and the explicit forms of A_3 and A_4 are defined in Appendix I.

For notation simplicity, we define $\gamma_a=\sigma_a^2/\sigma_n^2$, $\gamma_b=\sigma_b^2/\sigma_n^2$. Then, (37) can be explicitly written as in (42), shown at the top of the next page.

B. Training Design

Define $\mu_i=Q_i/NP_r$, and define the following two sets:

$$\mathcal{A}_1=\begin{cases}[\alpha_0,1], & \text{if } \mu_1\geq\frac{1}{8};\\ [\alpha_0,\alpha_1], & \text{otherwise,}\end{cases} \tag{43}$$

$$\mathcal{A}_2=\begin{cases}[0,1-\beta_0], & \text{if } \mu_2\geq\frac{1}{8};\\ [1-\beta_1,1-\beta_0], & \text{otherwise,}\end{cases} \tag{44}$$

where

$$\alpha_0=\frac{1}{\mu_1^2+4\mu_1+5},\quad \alpha_1=\frac{9-4\mu_1-\sqrt{1-8\mu_1}}{2\left(\mu_1^2-4\mu_1+5\right)},$$
$$\beta_0=\frac{1}{\mu_2^2+4\mu_2+5},\quad \beta_1=\frac{9-4\mu_2-\sqrt{1-8\mu_2}}{2\left(\mu_2^2-4\mu_2+5\right)}.$$

The following proposition provides illumination for our next discussion.

Proposition 2: The optimal $\mathbf{t}_1$ and $\mathbf{t}_2$ which minimized $\sigma^2_{\hat{\mathbf{h}}_e}$ should be orthogonal, if

$$\alpha\in\mathcal{A}_1\bigcap\mathcal{A}_2\ . \tag{45}$$

Proof: See Appendix II. ∎

If both μ_1 and μ_2 are greater than $\frac{1}{8}$, then the condition in Proposition 2 reduces to $\alpha\in[\alpha_0,1-\beta_0]$. For example, if $\mu_1=\mu_2=1$ which is the case when relay has the same power as the terminal and when the training power is the same as the data transmission power, then the condition is $\alpha\in[0.1,0.9]$. When $\mu_i, i=1,2$ become larger, the range of α corresponding to the optimal $\rho=0$ also becomes larger.

C. Power Allocation

Unfortunately, Proposition 2 still cannot guarantee that $\rho=0$ is the optimal correlation factor for all the range of α. Therefore in this paper, we will restrict ourselves to a suboptimal power allocation under $\rho=0$. Nonetheless, we will numerically see in the simulations that $\rho=0$ is a good correlation factor for all the range of α.

For $\rho=0$, $\sigma^2_{\hat{\mathbf{h}}_e}$ can be simplified to

$$\sigma^2_{\hat{\mathbf{h}}_e}=\left(\frac{\gamma_ad_{11}}{\gamma_a+d_{11}}+\frac{\gamma_bd_{22}}{\gamma_b+d_{22}}\right)\sigma_n^2. \tag{46}$$

Observing that $\sigma^2_{\hat{\mathbf{h}}_e}$ is a function of a single variable α, the minimum must be obtained at one root of its first order derivative or at $\alpha=0$ or $\alpha=1$, regardless of whether $\sigma^2_{\hat{\mathbf{h}}_e}$ is convex or not. With the same reason as in Section III.C, one can immediately remove the choices of $\alpha=0$ and $\alpha=1$. The derivative of $\sigma^2_{\hat{\mathbf{h}}_e}$ with respect to α is given by (47), which is shown at the top of the next page, where F_i's are the same as those defined in the previous section. Thus, the optimal α is one of (48), shown at the top of the next page, that belongs to $[0,1]$. When the transmit SNRs at both the two source terminals are high, we obtain the suboptimal solution of α as

$$\alpha_{\mathrm{sub-opt}}=\lim_{\sigma_n^2\to 0^+}\alpha_{\mathrm{opt}}=\frac{1}{1+\varrho}, \tag{49}$$

where $\varrho=\frac{Q_1\sigma_1}{Q_2\sigma_2}$.

Similar discussions as those in Section III.C can be made here, which will be omitted for brevity. Once again, the optimal α depends on the balancedness of the two link, which depends on two different factors, i.e., training power and channel quality. Different from the scenario of maximizing the AESNR, the balancedness for minimizing the MSE is not related with the data power P_i, which is quite reasonable because data power is not involved into the LMMSE channel estimation.

V. Simulation Results

In this section, we numerically study the performance of our proposed channel estimation algorithms and the training designs under various scenarios. The channel variances are set as $\sigma_1^2=\sigma_2^2=1$. For all examples, we fix $P_1=P_2=P_r$, and the SNR is defined as P_1/σ_n^2. The parameter N is set to be 4 and the phase of ρ is randomly taken. The training power Q_2 is set to be NP_2, while Q_1/Q_2 is taken as 6.3 (8dB in log-unit) for most examples unless otherwise mentioned. Hence, the

$$\sigma^2_{\hat{\mathbf{h}}_e} = \frac{(\gamma_a+\gamma_b)d_{11}d_{22}+\gamma_a\gamma_b d_{11}+\gamma_a\gamma_b d_{22}-(\gamma_a+\gamma_b)|d_{12}|^2}{\gamma_a\gamma_b+d_{11}d_{22}+\gamma_b d_{11}+\gamma_a d_{22}-|d_{12}|^2}\sigma_n^2 \tag{42}$$

$$\frac{\partial\sigma^2_{\hat{\mathbf{h}}_e}}{\partial\alpha} = \frac{\left(Q_2\sigma_b^2F_4\left(F_1\sigma_n^2+F_2\alpha\right)^2-Q_1\sigma_a^2F_1\left(F_4\sigma_n^2+F_5\beta\right)^2\right)NP_r\sigma_n^2}{\left(F_1\sigma_n^2+F_2\alpha\right)^2\left(F_4\sigma_n^2+F_5\beta\right)^2} \tag{47}$$

$$\alpha_{\text{opt}} = \frac{Q_1\sigma_a^4F_1F_5^2+\left(Q_1\sigma_a^4F_5+Q_2\sigma_b^4F_2\right)F_1F_4\sigma_n^2 \pm\sqrt{Q_1Q_2F_1F_4}\sigma_a^2\sigma_b^2\left(F_2F_5+F_1F_5\sigma_n^2+F_2F_4\sigma_n^2\right)}{Q_1\sigma_a^4F_1F_5^2-Q_2\sigma_b^4F_4F_2^2} \tag{48}$$

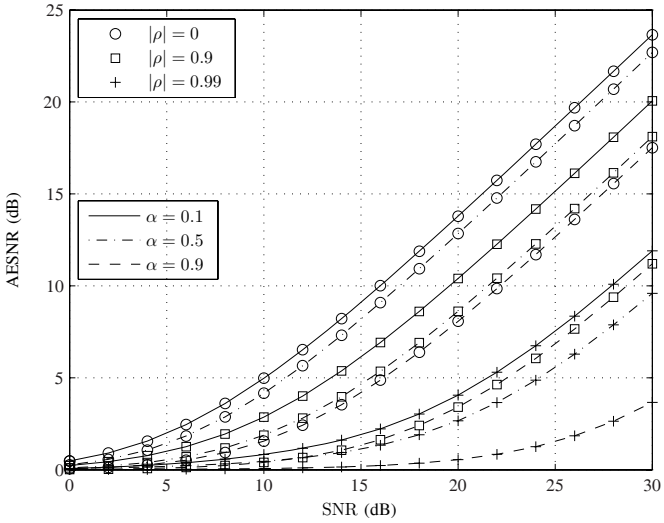

Fig. 2. AESNR versus SNR for $|\rho| = 0, 0.9, 0.99$ and $\alpha = 0.1, 0.5, 0.9$.

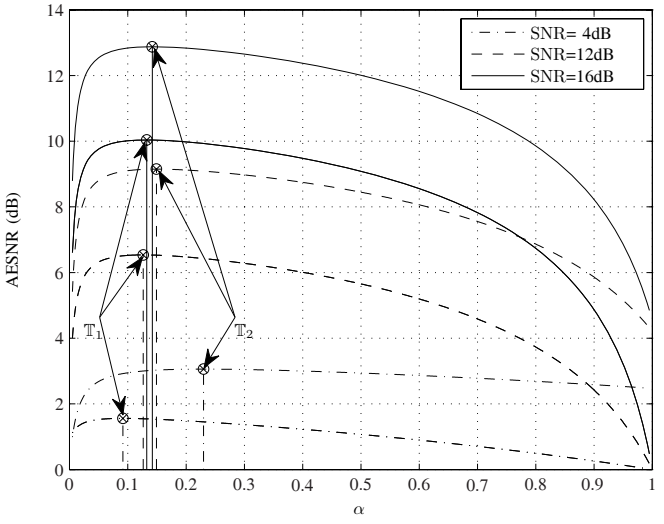

Fig. 3. AESNR versus α for SNR$= 4, 12, 16$ dB at $|\rho| = 0$.

parameters τ (defined as $\frac{Q_1\sigma_1\sqrt{P_2}}{Q_2\sigma_2\sqrt{P_1}}$) and ϱ (defined as $\frac{Q_1\sigma_1}{Q_2\sigma_2}$) are 6.3, and it is expected that $\alpha_{\text{sub-opt}}$ is 0.137 for both LMSNR and LMMSE channel estimators. Unless otherwise mentioned, the simulations results are made for terminal $\mathbb{T}_1$ only.

A. LMSNR Channel Estimation

1) Optimality of $|\rho|$: In the first example, we numerically verify the optimal choice of $|\rho|$ by taking $|\rho| = 0$, $|\rho| = 0.9$, and $|\rho| = 0.99$, while for each $|\rho|$ we take three different power allocation factors $\alpha = 0.1$, $\alpha = 0.5$ and $\alpha = 0.9$, respectively. The performance AESNRs versus the SNR are shown in Fig. 2. Note that we do not demonstrate for $|\rho|$ between 0 and 0.9 because the resultant lines are close to that of $|\rho| = 0$, which make the figure illegible. Performance on different α will be given in the next example. One important observation from Fig. 2 is that, for any value of α, $|\rho| = 0$ always yields the best performance, which numerically verifies our previous claim in Proposition 1.

2) Optimality of α: Fig. 3 shows the AESNR versus the power allocation α for both terminals $\mathbb{T}_1$ and $\mathbb{T}_2$. Three different SNRs, i.e., 4 dB, 12 dB, and 16 dB are involved. The parameter ρ is taken as 0 for all SNRs.

We first discuss over the three AESNR curves for $\mathbb{T}_1$. Numerically, it is observed that the AESNR is a concave function of α within $[0, 1]$, which indicates a unique maximum point corresponding to the optimal power allocation. The intersection of α_{opt} with AESNR is marked by circle, while the numerically obtained optimal α (from one-dimensional search) is marked by cross. Clearly, our theoretical analysis matches the numerical results quite well. It needs to be emphasized that, the value of $\alpha_{\text{sub-opt}}$ is 0.137, which almost overlaps with α_{opt} when the SNR is as high as 16 dB. Although it is away from the α_{opt} when the SNR is low, the corresponding performance does not deteriorate much compared to its optimal value. A more clear demonstration is shown in the next example.

Now we look at the rest three AESNR curves for $\mathbb{T}_2$. The optimal β is first obtained and then the corresponding α is found from $1-\beta$. It is seen that the optimal power allocation for $\mathbb{T}_2$ is not the same as that for $\mathbb{T}_1$ at the low SNR region. However, they are very close at high SNR region as we have analyzed in the main text part, i.e., both approach to $\alpha_{\text{sub-opt}}$. Nonetheless, we also observe that the AESNR curves are quite flat near the $\alpha_{\text{sub-opt}}$. So the performance loss by using $\alpha_{\text{sub-opt}}$ is quite small, and we then claim that using $\alpha_{\text{sub-opt}}$ could almost maximize AESNR for both terminals simultaneously.

3) Comparison with the existing method: We then compare our method with the existing one in [19], where an equal power allocation $\alpha = \beta = 0.5$ is applied. The optimal correlation factor is $\rho = 0$ for both methods. The results from the optimal power allocation α_{opt} and the sub-optimal power

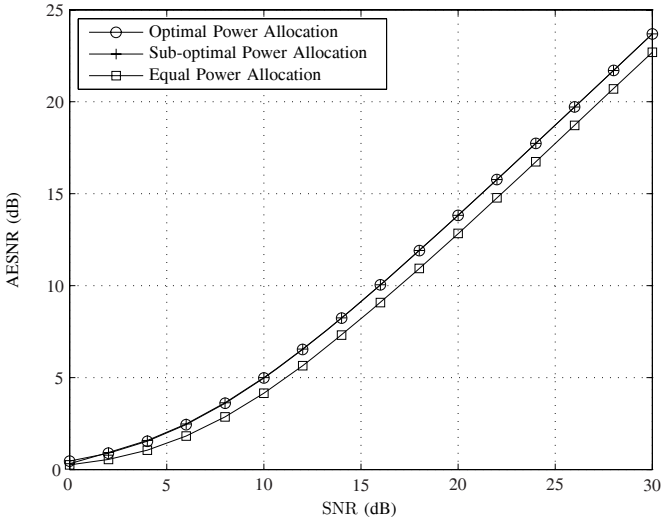

Fig. 4. AESNR versus SNR under optimal, sub-optimal, and equal power allocations.

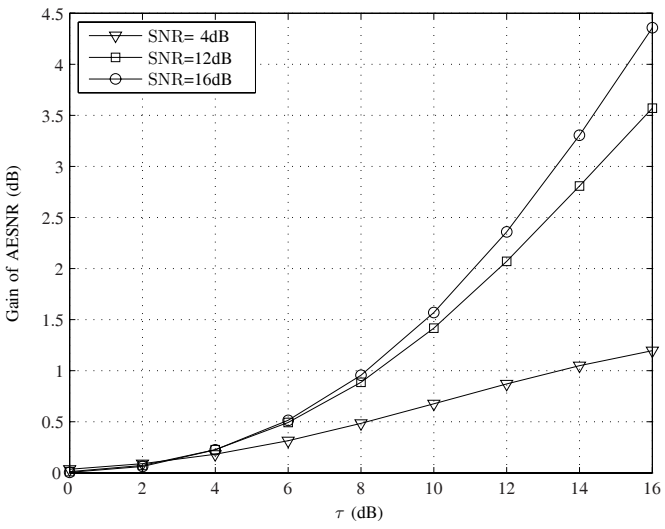

Fig. 5. The AESNR gain of the optimal allocation over the equal power allocation versus τ for SNR$= 4, 12, 16$ dB.

allocation $\alpha_{\mathrm{sub-opt}}$ are both displayed. As shown in Fig. 4, the performance of our proposed method with $\alpha_{\mathrm{sub-opt}}$ attaches that with α_{opt} even at lower SNR region, which implies a very good approximation by using $\alpha_{\mathrm{sub-opt}}$, only. Moreover, the proposed method with optimal power allocation outperforms the method in [19] by 1 dB.

4) Performance for different τ: From the expression of $\alpha_{\mathrm{sub-opt}}$, we know that the gain of the proposed technique over [19] is related to τ, i.e., the balancedness between the two channel links. It is then of interest to check how the performance gain varies with different values of τ. Define the AESNR gain as

$$\text{AESNR gain} = \frac{\text{AESNR with optimal power allocation}}{\text{AESNR with equal power allocation}}.$$

The results are shown in Fig. 5 for three SNRs 4 dB, 12 dB, and 16 dB, respectively. Clearly, as τ becomes larger, i.e., more unbalanced scenario, the AESNR gain from applying the optimal power allocation becomes larger.

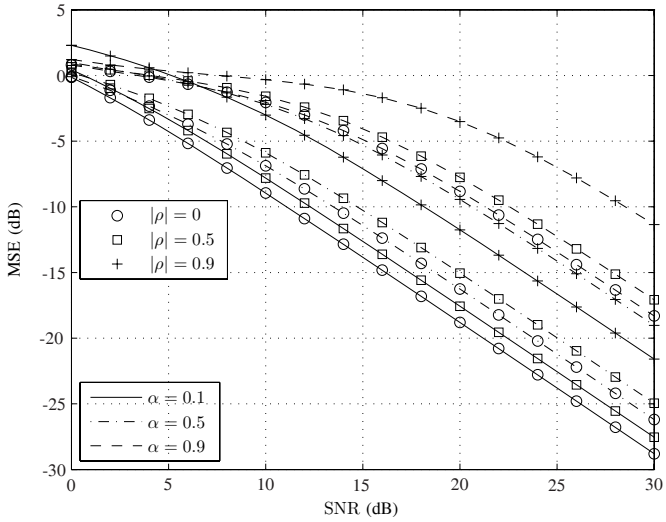

Fig. 6. MSE versus SNR for $|\rho| = 0, 0.9, 0.99$ and $\alpha = 0.1, 0.5, 0.9$.

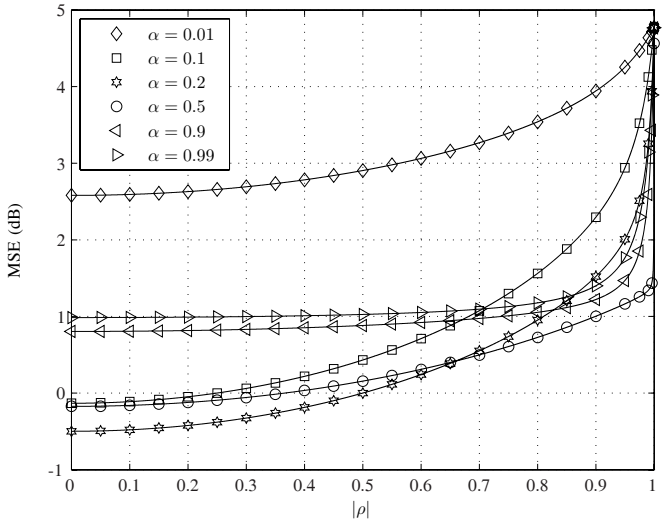

Fig. 7. MSE versus $|\rho|$ for $\alpha = 0.01, 0.1, 0.2, 0.5, 0.9, 0.99$ at SNR$= 0$dB.

B. LMMSE Channel Estimation

1) Optimality of $|\rho|$: The performance of MSEs versus the SNR are shown in Fig. 6 for three different $|\rho|$ as $|\rho| = 0$, $|\rho| = 0.5$, and $|\rho| = 0.9$, while for each $|\rho|$ we consider three different power allocation factor $\alpha = 0.1$, $\alpha = 0.5$ and $\alpha = 0.9$. It is seen that $|\rho| = 0$ provides the optimal values for all three choice of α.

To further verify the optimality of $|\rho| = 0$ in a numerical way, we fix SNR$= 0$ dB and show the performance of MSEs versus $|\rho|$ in Fig. 7, for a wide range of α from 0.01 to 0.99. Obviously, $|\rho| = 0$ still performs the best over different scenarios. We need to emphasize that the optimality of $|\rho| = 0$ is only numerically examined in this paper.

2) Optimality of α: In this example, we check the optimality of α_{opt} in (48) for the choice $|\rho| = 0$. Three different SNRs 4 dB, 12 dB, and 16 dB are adopted. The performance MSEs versus α are shown in Fig. 8 for both terminal $\mathbb{T}_1$ and $\mathbb{T}_2$. It is seen that MSE curves are convex functions of α within $[0, 1]$ for all three SNR values. The vertical lines give the position of α_{opt} whose intersection with MSE curves are

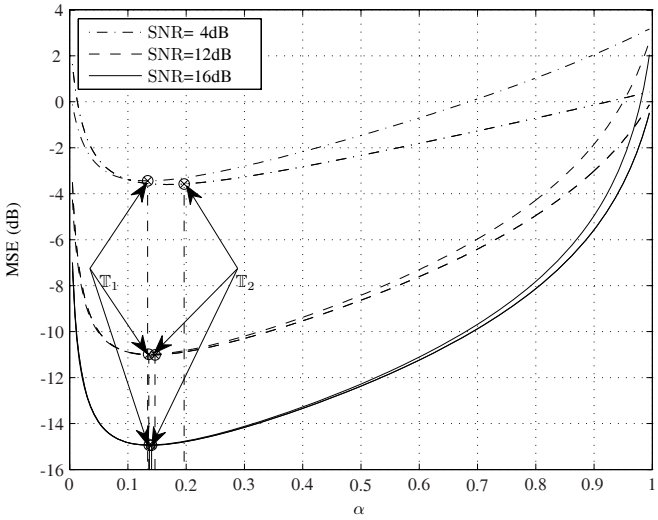

Fig. 8. MSE versus α for SNR$= 4, 12, 16$ dB at $|\rho| = 0$.

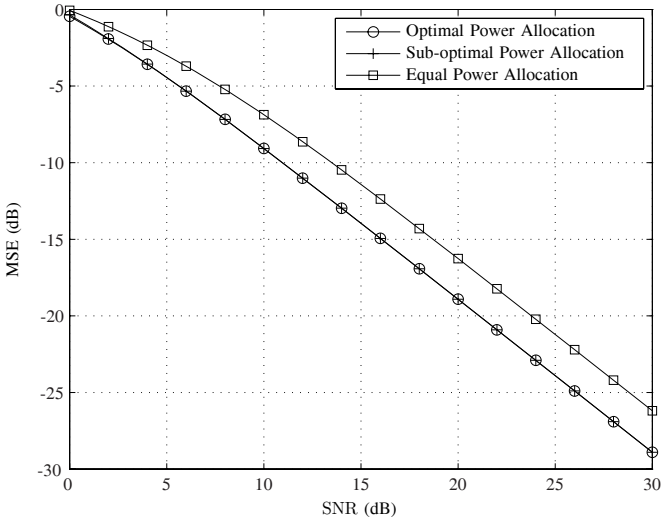

Fig. 9. MSE versus SNR under optimal, sub-optimal, and equal power allocations..

marked by circle. The numerically searched optimal α on MSE lines are marked by cross. Clearly, our theoretical analysis matches the numerical results quite well. Similarly to Fig. 3, the suboptimal $\alpha_{\rm sub-opt}$ is far from $\alpha_{\rm opt}$ when the SNR equals to 4 dB, but the MSE degeneration is not remarkable. Moreover, the optimal power allocations for $\mathbb{T}_1$ and $\mathbb{T}_2$ are not the same at the low SNR region but are very close at high SNR region. Once again, we observe that the MSE curves are quite flat near $\alpha_{\text{sub-opt}}$. So the performance loss by using $\alpha_{\text{sub-opt}}$ is quite small.

3) Comparison with the existing method: We then compare our method with the existing LMMSE channel estimator [19]. Both $\alpha_{\rm opt}$ and $\alpha_{\rm sub-opt}$ are examined and the performance results are shown in Fig. 9. Clearly, the performance of $\alpha_{\rm sub-opt}$ attaches that from $\alpha_{\rm opt}$ even at lower SNR region, which implies a very good approximation of replacing $\alpha_{\rm opt}$ by $\alpha_{\rm sub-opt}$. Moreover, the proposed method with optimal power allocation outperforms the existing LMMSE from one to 3 dB.

4) Performance for different ϱ: As in the LMSNR estimator, the gain of the proposed method over the existing one comes from the unbalanced channel links in the relay networks. We then demonstrate this effect by changing the parameter ϱ in $\alpha_{\rm sub-opt}$. Define the MSE gain as

$$\text{MSE gain} = \frac{\text{MSE with equal power allocation}}{\text{MSE with optimal power allocation}}.$$

The resultant MSEs versus ϱ are displayed in Fig. 10 for SNR $= 4$, 12, and 16 dB, respectively. Clearly, the larger the ϱ is, i.e., the lager the unbalancedness is, the greater the MSE gain is observed.

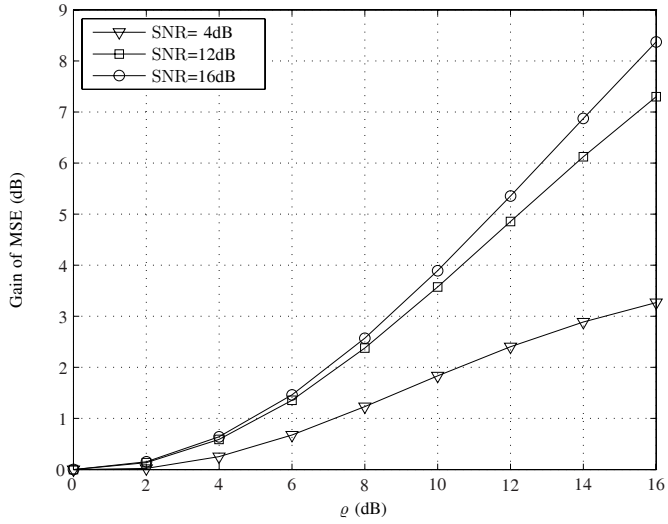

Fig. 10. The MSE gain of the optimal allocation over the equal power allocation versus ϱ for SNR$= 4, 12, 16$ dB.

VI. Conclusions

In this paper, we introduced a new principle to implement channel estimation for AF based TWRN by letting the relay node estimate the channels first and then allocate the power to optimize for certain criterions. We considered two linear estimators, i.e., LMSNR and LMMSE, and studied the corresponding training design and the optimal/suboptimal power allocations. Nonetheless, the proposed principle can be readily extended to different channel estimators with different optimizing criterions. For unbalanced channel links, where the definition of unbalancedness varies for different scenarios, the proposed method outperforms the existing techniques that trivially apply equal power allocation at the relay node. Finally, numerous simulation results were provided to corroborate the proposed studies.

Appendix I
Proof of Proposition 1

In order to maximize the AESNR, we first need to obtain the values $\mathbf{R}_z^{-1}\mathbf{t}_i$ and $\mathbf{t}_i^H\mathbf{R}_z^{-1}\mathbf{t}_i$, $i = 1, 2$.

Lemma 1: If an invertible $N \times N$ matrix $\mathbf{\Theta}$ has the form

$$\mathbf{\Theta} = \mathbf{I} + \sum_{i=1}^{M}\sum_{j=1}^{M}[\mathbf{A}]_{ij}\mathbf{x}_i\mathbf{x}_j^H, \tag{50}$$

where $\mathbf{x}_i$'s are $N \times 1$ column vectors and $\mathbf{A}$ is an $M \times M$ matrix, then the inverse of $\mathbf{\Theta}$ is

$$\mathbf{\Theta}^{-1} = \mathbf{I} - \sum_{i=1}^{M}\sum_{j=1}^{M} [\mathbf{B}]_{ij}\mathbf{x}_i\mathbf{x}_j^H, \tag{51}$$

where $\mathbf{B}$ is an $M \times M$ matrix with the form

$$\mathbf{B} = (\mathbf{I} + \mathbf{A\Omega})^{-1}\mathbf{A}, \tag{52}$$

and $\mathbf{\Omega}$ is an $M \times M$ matrix with $[\mathbf{\Omega}]_{ij} = \mathbf{x}_i^H\mathbf{x}_j$. Furthermore, we have

$$\mathbf{\Theta}^{-1}\mathbf{x}_q = \mathbf{x}_q - \sum_{i=1}^{M} \left[(\mathbf{I} + \mathbf{A\Omega})^{-1}\mathbf{A\Omega}\right]_{iq}\mathbf{x}_i, \tag{53}$$

and

$$\mathbf{x}_p^H\mathbf{\Theta}^{-1}\mathbf{x}_q = \left[\mathbf{\Omega}(\mathbf{I} + \mathbf{A\Omega})^{-1}\right]_{pq}. \tag{54}$$

Proof: The product of the right hand sides (RHSs) of (50) and (51), denoted as $\mathbf{G}$, can be expressed as

$$\mathbf{G} = \mathbf{I} + \sum_{i=1}^{M}\sum_{j=1}^{M}\left([\mathbf{A}]_{ij} - [\mathbf{B}]_{ij} - \sum_{p=1}^{M}\sum_{q=1}^{M}[\mathbf{A}]_{ip}[\mathbf{\Omega}]_{pq}[\mathbf{B}]_{qj}\right)\mathbf{x}_i\mathbf{x}_j^H, \tag{55}$$

$$= \mathbf{I} + \sum_{i=1}^{M}\sum_{j=1}^{M}\left([\mathbf{A} - \mathbf{B} - \mathbf{A\Omega B}]_{ij}\right)\mathbf{x}_i\mathbf{x}_j^H. \tag{56}$$

Substituting (52) into (56) yields $\mathbf{G} = \mathbf{I}$. Thus, the RHS of (51) is the inverse of $\mathbf{\Theta}$.

From (51), $\mathbf{\Theta}^{-1}\mathbf{x}_q$ and $\mathbf{x}_p^H\mathbf{\Theta}^{-1}\mathbf{x}_q$ can be respectively expressed as

$$\mathbf{\Theta}^{-1}\mathbf{x}_q = \mathbf{x}_q - \sum_{i=1}^{M}\sum_{j=1}^{M}[\mathbf{B}]_{ij}[\mathbf{\Omega}]_{jq}\mathbf{x}_i = \mathbf{x}_q - \sum_{i=1}^{M}[\mathbf{B\Omega}]_{iq}\mathbf{x}_i, \tag{57}$$

$$\mathbf{x}_p^H\mathbf{\Theta}^{-1}\mathbf{x}_q = [\mathbf{\Omega}]_{pq} - \sum_{i=1}^{M}[\mathbf{\Omega}]_{pi}[\mathbf{B\Omega}]_{iq} = [\mathbf{\Omega} - \mathbf{\Omega B\Omega}]_{pq}. \tag{58}$$

Substituting (52) into (57) and (58) yields (53) and (54), respectively. This completes the proof of Lemma 1. ∎

For the parameters used in Lemma 1, let $\mathbf{x}_i = \mathbf{t}_i$, $\mathbf{\Theta} = \mathbf{R}_z/\sigma_n^2$, and $\mathbf{A}$ be a 2×2 matrix with $[\mathbf{A}]_{11} = C_1$, $[\mathbf{A}]_{22} = C_2$, and $[\mathbf{A}]_{12} = [\mathbf{A}]_{21}^* = C_3$, then $\mathbf{t}_i\mathbf{R}_z^{-1}\mathbf{t}_i$, $i = 1, 2$, can be computed as (59) and (60), which are shown on the top of the next page.

Define $x = 1 - |\rho|^2$. Then after tedious re-organization, $\psi_1(\alpha, \beta, |\rho|)$ becomes

$$\psi_1(x) \triangleq \psi_1(\alpha, \beta, \sqrt{1-x^2}) = A_0\frac{\varsigma_3 x^3 + \varsigma_2 x^2 + \varsigma_1 x}{\xi_3 x^3 + \xi_2 x^2 + \xi_1 x + \xi_0}, \tag{61}$$

where

$$\varsigma_3 = \beta A_2, \qquad \varsigma_2 = \beta\sigma_1^2 + A_3, \qquad \varsigma_1 = A_4,$$
$$\xi_3 = \alpha\beta A_1 A_2, \qquad \xi_0 = A_4 A_5\sigma_1^2 + A_4^2,$$
$$\xi_2 = (A_1 + A_2 + \sigma_1^2)\alpha\beta\sigma_1^2 + (\alpha A_1 + \beta A_2 + 2\sqrt{\alpha\beta}\sigma_1^2)A_3 + A_3^2,$$
$$\xi_1 = (\alpha A_1 + \beta A_2 + 2\sqrt{\alpha\beta}\sigma_1^2)A_4 + A_3 A_5\sigma_h^2 + 2A_3 A_4,$$

and

$$A_0 = \frac{P_1\sigma_a^4 Q_1\alpha}{\sigma_n^2}, \quad A_1 = \frac{Q_1\sigma_a^2}{\sigma_n^2}, \quad A_2 = \frac{Q_2\sigma_b^2}{\sigma_n^2}, \quad A_4 = \frac{A_5\sigma_n^2}{NP_r},$$
$$A_3 = \frac{\alpha Q_1\sigma_1^2 + \beta Q_2\sigma_2^2 + 2\sqrt{\alpha\beta}\sigma_n^2}{NP_r}, \qquad A_5 = (\sqrt{\alpha} - \sqrt{\beta})^2.$$

From the quotient rule in calculus [25], the first order derivative of $\psi_1(x)$ can be expressed as (62), shown on the top of the next page.

To see the sign of (62), we only need to consider its numerator. Obviously, the coefficients of x and the constant term are nonnegative. Furthermore, we can verify that

$$\varsigma_3\xi_2 - \varsigma_2\xi_3 = \beta A_2\left((A_2 + \sigma_1^2)\alpha\beta\sigma_1^2 + (\beta A_2 + 2\sqrt{\alpha\beta}\sigma_1^2)A_3 + A_3^2\right) \geq 0,$$
$$\varsigma_3\xi_1 - \varsigma_1\xi_3 = \beta A_2\left((\beta A_2 + 2\sqrt{\alpha\beta}\sigma_1^2)A_4 + \sigma_1^2 A_3 A_5 + 2A_3 A_4\right) \geq 0.$$

Therefore, the coefficients of x^4 and x^3 are also nonnegative. After some algebraic operations, the coefficient of x^2 can be expressed as

$$\varsigma_2\xi_1 - \varsigma_1\xi_2 + 3\varsigma_3\xi_0 = B_0 + B_1\left(\tilde{A}_2(2\alpha - 6\sqrt{\alpha\beta} + 5\beta) - (\alpha - 4\sqrt{\alpha\beta})\right), \tag{63}$$

where $\tilde{A}_2 = \frac{Q_2\sigma_2^2}{\sigma_n^2}$ is the received SNR from $\mathbb{T}_2$ at $\mathbb{R}$ during the training, while B_0 and B_1 are positive coefficients whose explicit forms are omitted for brevity. Bearing in mind that $\alpha + \beta = 1$, we obtain

$$\begin{cases} \alpha - 4\sqrt{\alpha\beta} < 0, & \frac{1}{17} < \beta \leq 1; \\ \frac{2\alpha - 6\sqrt{\alpha\beta} + 5\beta}{\alpha - 4\sqrt{\alpha\beta}} \geq 2, & 0 \leq \beta < \frac{1}{17}. \end{cases} \tag{64}$$

From (64), we see that if $\tilde{A}_2 \geq \frac{1}{2}$, then the coefficient $(\varsigma_2\xi_1 - \varsigma_1\xi_2 + 3\varsigma_3\xi_0)$ is no less than zero. Hence $\frac{\partial\psi_1}{\partial x} \geq 0$ and $\psi_1(\alpha, \beta, |\rho|)$ is a monotonically increasing function of x and is, hence, a monotonically decreasing function of $|\rho|$.

Similarly, we know that if $\tilde{A}_1 = \frac{Q_1\sigma_1^2}{\sigma_n^2} \geq \frac{1}{2}$, then $\frac{\partial\psi_2(\alpha,\beta,\sqrt{1-x^2})}{\partial x} \geq 0$ and $\psi_2(\alpha, \beta, |\rho|)$ is a monotonically decreasing function of $|\rho|$.

Combined with (22) and (23), we know that $\bar{\gamma}$ is a monotonically decreasing function of $|\rho|$ whose maximum is attained when $|\rho| = 0$. This completes the proof of Proposition 1.

Appendix II
Proof of Proposition 2

Define

$$g_1 = \frac{d_{11}}{\det\mathbf{D}}, \qquad g_2 = \frac{d_{22}}{\det\mathbf{D}}, \qquad g_3 = \frac{|d_{12}|^2}{\det\mathbf{D}}, \tag{65}$$

where $\det\mathbf{D} = d_{11}d_{22} - |d_{12}|^2$ is the determinant of $\mathbf{D}$. Since $\det\mathbf{D}$ is positive, g_i, $i = 1, 2, 3$ are positive numbers. Then, (37) can be alternatively written as

$$\sigma_{\hat{\mathbf{h}}_e}^2 = \frac{\gamma_a^{-1} + \gamma_b^{-1} + g_1 + g_2}{(\gamma_a^{-1} + g_2)(\gamma_b^{-1} + g_1) - g_3}\sigma_n^2. \tag{66}$$

$$\mathbf{t}_1^H \mathbf{R}_z^{-1} \mathbf{t}_1 = \frac{Q_1(1+C_2Q_2(1-|\rho|^2))}{\left(1+C_1Q_1+C_2Q_2+2\Re(C_3\rho^*)\sqrt{Q_1Q_2}+(C_1C_2-|C_3|^2)(1-|\rho|^2)Q_1Q_2\right)\sigma_n^2}, \tag{59}$$

$$\mathbf{t}_2^H \mathbf{R}_z^{-1} \mathbf{t}_2 = \frac{Q_2(1+C_1Q_1(1-|\rho|^2))}{\left(1+C_1Q_1+C_2Q_2+2\Re(C_3\rho^*)\sqrt{Q_1Q_2}+(C_1C_2-|C_3|^2)(1-|\rho|^2)Q_1Q_2\right)\sigma_n^2} \tag{60}$$

$$\frac{\partial \psi_1}{\partial x} = A_0 \frac{(\varsigma_3\xi_2-\varsigma_2\xi_3)x^4+2(\varsigma_3\xi_1-\varsigma_1\xi_3)x^3+(\varsigma_2\xi_1-\varsigma_1\xi_2+3\varsigma_3\xi_0)x^2+2\varsigma_2\xi_0 x+\varsigma_1\xi_0}{\left(\xi_3x^3+\xi_2x^2+\xi_1x+\xi_0\right)^2} \tag{62}$$

The derivatives of $\sigma^2_{\hat{\mathbf{h}}_e}$ with respect to g_1, g_2, and g_3 are given by

$$\frac{\partial \sigma^2_{\hat{\mathbf{h}}_e}}{\partial g_1} = -\frac{\left(\gamma_a^{-1}+g_2\right)^2+g_3}{\left(\left(\gamma_a^{-1}+g_2\right)\left(\gamma_b^{-1}+g_1\right)-g_3\right)^2}\sigma_n^2 < 0\,, \tag{67}$$

$$\frac{\partial \sigma^2_{\hat{\mathbf{h}}_e}}{\partial g_2} = -\frac{\left(\gamma_b^{-1}+g_1\right)^2+g_3}{\left(\left(\gamma_a^{-1}+g_2\right)\left(\gamma_b^{-1}+g_1\right)-g_3\right)^2}\sigma_n^2 < 0\,, \tag{68}$$

$$\frac{\partial \sigma^2_{\hat{\mathbf{h}}_e}}{\partial g_3} = \frac{\gamma_a^{-1}+\gamma_b^{-1}+g_1+g_2}{\left(\left(\gamma_a^{-1}+g_2\right)\left(\gamma_b^{-1}+g_1\right)-g_3\right)^2}\sigma_n^2 > 0\,, \tag{69}$$

respectively. From (67)–(69), we know that $\sigma^2_{\hat{\mathbf{h}}_e}$ is monotonically decreasing with g_1 and g_2, whereas is monotonically increasing with g_3.

With some tedious manipulations and re-organizations, the derivatives of g_1 and g_2 with respect to x can be expressed as

$$\frac{\partial g_1}{\partial x} = B_{1,1} + B_{1,2}\left(\left(A_3+\alpha\sigma_1^2\right)^2-\left(\alpha-\sqrt{\alpha\beta}\right)^2\sigma_1^4\right), \tag{70}$$

$$\frac{\partial g_2}{\partial x} = B_{2,1} + B_{2,2}\left(\left(A_3+\beta\sigma_1^2\right)^2-\left(\beta-\sqrt{\alpha\beta}\right)^2\sigma_1^4\right), \tag{71}$$

where $B_{i,j}$'s are all positive coefficients whose explicit forms are omitted for brevity. From the fact that $A_3 > \mu_1\sigma_1^2$, we can verify that (70) and (71) are positive if $\alpha \in \mathcal{A}_1$. Therefore, g_i, $i = 1, 2$ are monotonically increasing function of x, whose maximums are achieved simultaneously when $x = 1$. Meanwhile, g_3 achieves its minimum when $x = 1$. Hence, the channel estimation MSE $\sigma^2_{\hat{\mathbf{h}}_e}$ at $\mathbb{T}_1$ achieves its minimum when $x = 1$; namely, $\rho = 0$.

Similarly, we can prove that if $\alpha \in \mathcal{A}_2$, then the channel estimation MSE at $\mathbb{T}_2$ achieves its minimum when $\rho = 0$. This completes the proof of Proposition 2.

ACKNOWLEDGMENT

The authors would like to thank the editor and the reviewers for their insightful comments and thorough review. Their suggestions have led to significant improvement of the manuscript.

REFERENCES

[1] S. Zhang, S. C. Liew, and P. P. Lam, "Physical-layer network coding," in *Proc. MobiComm*, Sep. 2006, pp. 358-365.

[2] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: analog network coding," Computer Science and Artificial Intelligence Laboratory Technical Report, MIT-CSAIL-TR-2007-012, Feb. 23, 2007.

[3] B. Rankov and A. Wittneben, "Spectral efficient signaling for half-duplex relay channels," in *Proc. Asilomar Conf.*, Pacific Grove, USA, Oct. 2005, pp. 1066-1071.

[4] J. N. Laneman and G. W. Wornell, "Distributed space-time block coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415-2425, Oct. 2003.

[5] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[6] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proc. IEEE ISIT*, Seattle, USA, July 2006, pp. 1668-1672.

[7] S. J. Kim, N. Devroye, P. Mitran, and V. Tarokh, "Achievable rate regions for bi-directional relaying," submitted to *IEEE Trans. Inf. Theory.* [Online]. Available: arXiv:0808.0954v1.

[8] C.-H. Liu, F. Xue, and J. G. Andrews, "Network coding with two-way relaying: achievable rate regions and diversity-multiplexing tradeoffs," submitted to *IEEE Trans. Wireless Commun.* [Online]. Available: arXiv:0902.2260v1.

[9] A. Avestimehr, A. Sezgin, and D. Tse, "Approximate capacity of the two way relay channel: a deterministic approach," in *Allerton Conf. Commun., Control, Comput.*, Monticello, IL, 2008. [Online]. Available: arXiv:0808.3145v1.

[10] R. Vaze and R. Heath Jr, "On the capacity and diversity-multiplexing tradeoff of the two-way relay channel," submitted to *IEEE Trans. Inf. Theorey*. [Online]. Available: arXiv:0810.3900v2

[11] M. P. Wilson, K. Narayanan, H. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bi-directional relaying," in *Allerton Conf. Commun., Control, Comput.*, Monticello, IL, 2007.

[12] T. Oechtering and H. Boche, "Optimal transmit strategies in multi antenna bidirectional relaying," in *Proc. IEEE ICASSP'07*, Honolulu, HI, USA, Apr. 2007, pp. 145-148.

[13] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498-3516, Oct. 2007.

[14] T. Cui, T. Ho, and J. Kliewer, "Memoryless relay strategies for two-way relay channels: performance analysis and optimization," in *Proc. IEEE ICC*, Beijing, China, May 2008, pp. 1139-1143.

[15] Y.-C. Liang and R. Zhang, "Optimal analogue relaying with multi-antenna for physical layer network coding," in *Proc. IEEE ICC*, Beijing China, May 2008, pp. 3893-3897.

[16] R. Zhang, Y.-C. Liang, C. C. Chai, and S. G. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no.5, pp. 699-712, June 2009.

[17] C. K. Ho, R. Zhang and Y.-C. Liang, "Two-way relaying over OFDM: optimized tone permutation and power allocation," in *Proc. IEEE ICC*, Beijing, China, May 2008, pp. 3908-3912.

[18] T. Cui, F. Gao, T. Ho, and A. Nallanathan, "Distributed space-time coding for two-way wireless relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 658-671, Feb. 2009.

[19] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3024-3033, Oct. 2009.

[20] M. Biguesh and A. B. Gershman, "Training based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884-893, Mar. 2006.

[21] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951-963 Apr. 2003.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2006.

[23] Further Advancements for E-UTRA Physical Layer Aspects (Release 9), 3GPP, Technical Specification Group Radio Access Network.

[24] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs NJ: Prentice Hall, 1987.
[25] G. Strang, *Calculus*. Wellesley MA: Cambridge Press, 1991.

Bin Jiang (S'06, M'10) received the B.S. degree in electrical engineering from Southeast University, Nanjing, China, in 2002. He is currently working toward the Ph.D. degree in communication and signal processing with the School of Information Science and Engineering, Southeast University, Nanjing, China. His research interests lie in the area of signal processing and wireless communication for MIMO systems. Mr. Jiang received the Science and Technology Progress Awards of the State Education Ministry of China in 2009.

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 jointed 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 has co-authored more than 60 refereed IEEE journal and conference papers and has served as a TPC member for IEEE ICC (2008, 2009), IEEE GLOBECOM (2008, 2009), IEEE VTC (2008, 2009) and IEEE PIMRC (2009).

Xiqi Gao (SM'07) received the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 1997. He joined the Department of Radio Engineering, Southeast University, in April 1992. Now he is a professor of information systems and communications. From September 1999 to August 2000, he was a visiting scholar at Massachusetts Institute of Technology, Cambridge, and Boston University, Boston, MA. From August 2007 to July 2008, he visited the Darmstadt University of Technology, Darmstadt, Germany, as a Humboldt scholar.

His current research interests include broadband multicarrier communications, MIMO wireless communications, channel estimation and turbo equalization, and multirate signal processing for wireless communications. He serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Dr. Gao received the Science and Technology Progress Awards of the State Education Ministry of China in 1998, 2006 and 2009.

Arumugam Nallanathan (S'97−M'00−SM'05) received the B.Sc. with honors from the University of Peradeniya, Sri-Lanka, in 1991, the CPGS from the Cambridge University, United Kingdom, in 1994 and the Ph.D. from the University of Hong Kong, Hong Kong, in 2000, all in Electrical Engineering. He was an Assistant Professor in the Department of Electrical and Computer Engineering, National University of Singapore, Singapore from August 2000 to December 2007. Currently, he is a Senior Lecturer in the Department of Electronic Engineering at King's College London, United Kingdom. His research interests include cognitive radio, relay networks, MIMO-OFDM systems, ultra-wide bandwidth (UWB) communication and localization. In these areas, he has published over 150 journal and conference papers. He is a co-recipient of the Best Paper Award presented at 2007 IEEE International Conference on Ultra-Wideband (ICUWB'2007).

He currently serves on the Editorial Board of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and IEEE SIGNAL PROCESSING LETTERS as an Associate Editor. He served as a Guest Editor for EURASIP JOURNAL OF WIRELESS COMMUNICATIONS AND NETWORKING: Special issue on UWB Communication Systems- Technology and Applications. He currently serves as the Secretary for the Signal Processing and Communications Electronics Technical Committee of IEEE Communications Society. He served as the General Track Chair for the IEEE VTC'2008-Spring, Co-Chair for the IEEE GLOBECOM'2008 Signal Processing for Communications Symposium and IEEE ICC'2009 Wireless Communications Symposium.