Training Sequence Design for Discriminatory Channel Estimation in Wireless MIMO Systems

Tsung-Hui Chang, *Member, IEEE*, Wei-Cheng Chiang, Y.-W. Peter Hong, *Member, IEEE*, and Chong-Yung Chi, *Senior Member, IEEE*

***Abstract*—This paper proposes a training-based channel estimation scheme for achieving quality-of-service discrimination between legitimate and unauthorized receivers in wireless multiple-input multiple-output (MIMO) channels. The proposed method has applications ranging from user discrimination in wireless TV broadcast systems to the prevention of eavesdropping in secret communications. By considering a wireless MIMO system that consists of a multiple-antenna transmitter, a legitimate receiver (LR) and an unauthorized receiver (UR), we propose a multi-stage training-based discriminatory channel estimation (DCE) scheme that aims to optimize the channel estimation performance of the LR while limiting the channel estimation performance of the UR. The key idea is to exploit the channel estimate fed back from the LR at the beginning of each stage to enable the judicious use of artificial noise (AN) in the training signal. Specifically, with knowledge of the LR's channel, AN can be properly superimposed with the training data to degrade the UR's channel without causing strong interference on the LR. The channel estimation performance of the LR in earlier stages may not be satisfactory due to the inaccuracy of the channel estimate and constraints on the UR's estimation performance, but can improve rapidly in later stages as the quality of channel estimate improves. The training data power and AN power are optimally allocated by minimizing the normalized mean-square error (NMSE) of the LR subject to a lower limit constraint on the NMSE of the UR. The proposed DCE scheme is then extended to the case with multiple LRs and multiple URs. Simulation results are presented to demonstrate the effectiveness of the proposed DCE scheme.**

***Index Terms*—Artificial noise, MIMO channel estimation, quality-of-service (QoS) discrimination, secret communications, training signal design.**

I. Introduction

THE need for discrimination between the quality-of-service (QoS) of different receivers in wireless systems appear in many applications, such as the QoS discrimination between paid and unpaid subscribers in TV broadcast systems, and the prevention of eavesdropping by unauthorized receivers (i.e., eavesdroppers) in secret communications. Conventionally, these issues have been addressed through the use of application-level techniques such as user authentication [1] or cryptography [2]. However, these methods may be subject to vulnerabilities and difficulties in key distribution and management, especially in highly dynamic wireless networks [3], [4]. Recent developments in physical-layer secrecy [5]–[9] have shown that this information security problem can also be handled in the physical layer by directly exploiting the different fading characteristics between the legitimate and unauthorized receivers' channels. It has been shown that the transmitter can reliably broadcast message signals to the legitimate receiver without having the unauthorized receiver infer any information from the message if QoS discrimination between the legitimate and the unauthorized receivers is guaranteed. These physical-layer techniques, therefore, can provide information security without involving any complex key exchange and management. Besides, they can be used as a complement to higher-layer security techniques. For example, one can use these physical-layer secrecy techniques to strengthen the security of the key exchanging process [3], [4]. Most of the existing works on physical-layer secrecy focus on the study of the maximum achievable perfect secrecy rate, i.e., the so called perfect secrecy capacity [5]–[9], from an information theoretic point of view, or on the design of channel coding or beamforming strategies [10]–[15] under the assumption that both the legitimate and unauthorized receivers perfectly know their channel state information (CSI). We instead investigate the problem of QoS discrimination from the channel estimation perspective through the design of training signals for discriminating between channel estimation performances at receivers. The proposed training and channel estimation schemes, therefore, identify a completely new problem that has not been addressed in the literature.

In this paper, we consider a wireless multiple-input multiple-output (MIMO) system that consists of a multiple-antenna transmitter and two (multiple-antenna) receivers, namely, the legitimate receiver (LR) and the unauthorized receiver (UR). The UR is assumed to be passive and it would not emit signals to prevent communication between the transmitter and the LR. To discriminate between the LR and the UR's reception performances, we propose a multistage training-based channel estimation scheme that aims to optimize the channel estimation performance at the LR while constraining the estimation performance attainable by the UR. Inspired by the work of Goel and Negi [11], discrimination of channel estimation performance is achieved by utilizing artificial noise (AN) in the left null space of the LR's

Manuscript received November 11, 2009; accepted August 09, 2010. Date of publication August 19, 2010; date of current version November 17, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alper Tunga Erdogan. The work of Y.-W. Peter Hong is supported by the National Science Council (NSC), R.O.C., under Grants NSC 98-2219-E-007-004 and NSC 98-2218-E-009-008-MY3. The work of C.-Y. Chi is supported by the NSC, R.O.C., under Grants NSC-98-2219-E-007-005 and NSC-98-2219-E-007-003. This work was presented in part at the 2009 IEEE GLOBECOM, Honolulu, HI, November 30–December 4, 2009.

T.-H. Chang and W.-C. Chiang are with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan 30013, R.O.C (e-mail: tsunghui.chang@gmail.com; g9764532@oz.nthu.edu.tw).

Y.-W. P. Hong and C.-Y. Chi are with the Institute of Communications Engineering and the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan 30013, R.O.C (e-mail: ywhong@ee.nthu.edu.tw; cychi@ee.nthu.edu.tw).

Digital Object Identifier 10.1109/TSP.2010.2068292

channel to degrade the estimation performance at the UR. This requires the transmitter to obtain knowledge of the CSI of the LR, which can be achieved through feedback at the end of each stage. The quality of the channel estimate obtained by the UR is constrained due to the use of AN while the channel estimate at the LR can be refined through multi-stage training. The more accurate the knowledge of the LR's CSI at the transmitter is, the more effective the use of AN can be. Discrimination between the channel estimation qualities leads to differences in the effective channel experienced by each receiver. From a signal processing point of view, this leads to performance discrimination in data detection and channel decoding achieved by the LR and the UR, and will make the UR have a low probability to intercept the information message. [3], [16].

Specifically, in the first stage of the proposed discriminatory channel estimation (DCE) scheme, the transmitter broadcasts a training signal, as in conventional MIMO training schemes [17], [18], to perform preliminary training on the LR's channel. Assume that the signal is received by both the LR and the UR, and that linear minimum mean square error (LMMSE) channel estimation is performed at both receivers. The preliminary training signal power must be restricted in order to constrain the UR's channel estimation performance, but doing this also limits the quality of the LR's channel estimate. By having the LR feedback its preliminary channel estimate, the transmitter can send in the next stage another training signal that contains AN in the left null space of the LR's channel. The AN-aided training signal can degrade the UR's estimation performance while allowing the LR to refine its channel estimate. Multiple stages of this feedback-and-retraining process can be performed to further refine the channel estimate at the LR. We show that the transmitter must be more conservative when utilizing AN in early stages, since the lack of precise channel knowledge will cause noise leakage into the LR's channel [19] and thus corrupt its channel estimate. Therefore, we propose to judiciously design the power allocation between the training data and the AN by minimizing the normalized mean square error (NMSE) performance of the LR subject to a performance constraint on the UR. In the case with only one feedback-and-retraining stage, we show that the power allocation problem, though not convex, can be solved by a simple line search. However, in the case with multiple stages of feedback and retraining, the power allocation problem becomes intractable. We instead propose to obtain an approximate solution by using the monomial approximation and condensation method [20] in the context of geometric program (GP). The condensation method involves solving a sequence of convex GPs, and hence a suboptimal but effective power allocation scheme can be efficiently obtained by interior point methods [21]. We will also extend the proposed training design and DCE scheme to the scenario with multiple LRs and multiple URs. Simulation results will show that the proposed design can render LRs to acquire accurate channel estimates while holding the NMSE of URs at a high value.

The rest of this paper is organized as follows. The wireless MIMO system model and the proposed DCE scheme are presented in Section II. We first focus on the scenario with one LR and one UR. In Section III, we analyze the channel estimation performance with a single feedback-and-retraining stage, and present a design criterion for optimal allocation of the training data and AN powers. In Section IV, we study the optimum training design and power allocation policy for the case with multiple stages of feedback and retraining. The proposed DCE scheme is extended to systems with multiple LRs and multiple URs in Section V. In Section VI, the efficacy of the proposed method is demonstrated by computer simulations. Finally, the conclusion is drawn in Section VII.

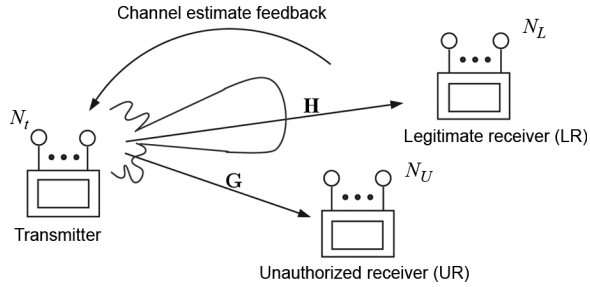


Fig. 1. Diagram of a wireless MIMO system consisting of a multi-antenna transmitter, a multi-antenna legitimate receiver (LR) and a multi-antenna unauthorized receiver (UR).

II. Problem Statement and Signal Model

As shown in Fig. 1, we consider a wireless MIMO system that consists of a transmitter, a legitimate receiver (LR) and an (passive) unauthorized receiver (UR) (e.g., the unpaid user in TV broadcast systems or the eavesdropper in secret communications). The system will be extended to that with multiple LRs and URs in Section V. We assume that the transmitter, the LR and the UR have N_t, N_L, and N_U antennas, respectively. To enable channel estimation at the LR, the transmitter must broadcast a sequence of training signals, but this may also allow the UR to perform channel estimation. To prevent the UR from benefiting from the broadcasted training signal, we propose a multi-stage discriminatory channel estimation (DCE) scheme described as follows.

- **Preliminary training**: In the initial stage, the transmitter first emits a sequence of regular training signals (that consists of only training data) for preliminary channel estimation at the LR.
- **(Multiple) feedback and retraining**: In the next stage, the LR sends back the channel estimate obtained in the previous stage to the transmitter (while the UR is allowed to intercept the fedback channel estimate[1]). The fedback channel estimate enables the transmitter to use artificial noise (AN) to degrade the channel estimation performance of the UR. Specifically, the transmitter broadcasts another sequence of training signals which superimposes training data with AN that is placed in the left null space of the LR's channel estimate. Both the LR and the UR will make use of

[1]As will be seen in Section III-A, the UR can perform the optimal LMMSE channel estimation without using the intercepted channel estimate. Moreover, if the transmitter employs the so called secrecy channel coding [6], [8] in the data transmission phase, this intercepted channel estimate is useless and does not enable the UR to intercept the information message sent by the transmitter.

the training signals in both stages to refine their channel estimates, but the performance of the UR will be constrained due to AN. The feedback-and-retraining process can be repeated multiple times, if necessary.

Assume that the feedback-and-retraining process is performed K times such that there is a total of $K+1$ stages in the training process, and that the channels from the transmitter to the LR and to the UR remain static over the entire training process.[2] Let $\mathbf{X}_k \in \mathbb{C}^{T_k \times N_t}$ denote the transmitted training signal matrix in stage k, with the signal length equal to T_k. The signals received by the LR and the UR are respectively given by

$$\text{LR}: \mathbf{Y}_k = \mathbf{X}_k\mathbf{H} + \mathbf{W}_k, \tag{1}$$

$$\text{UR}: \mathbf{Z}_k = \mathbf{X}_k\mathbf{G} + \mathbf{V}_k, \tag{2}$$

for $k = 0, 1, \ldots, K$, where the parameters are defined as follows:

$\mathbf{H} \in \mathbb{C}^{N_t \times N_L}$	The MIMO channel matrix of the LR. The elements of $\mathbf{H}$ are assumed to be independent and identically distributed (i.i.d.) random variables with zero mean and variance equal to σ_H^2.
$\mathbf{G} \in \mathbb{C}^{N_t \times N_U}$	The MIMO channel matrix of the UR. The elements of $\mathbf{G}$ are assumed to be i.i.d. random variables with zero mean and variance equal to σ_G^2.
$\mathbf{W}_k \in \mathbb{C}^{T_k \times N_L}$	Additive (spatially and temporally) white Gaussian noise (AWGN) matrix at the LR, with the power of each entry equal to σ_w^2.
$\mathbf{V}_k \in \mathbb{C}^{T_k \times N_U}$	AWGN matrix at the UR, with the power of each entry equal to σ_v^2.

Let P_k be the training data power at stage k, for $k = 0, \ldots, K$. The training signal matrices $\mathbf{X}_k$, $k = 0, \ldots, K$, are designed as follows. In the initial stage (i.e., $k = 0$), the transmitter broadcasts a pure training signal

$$\mathbf{X}_0 = \sqrt{\frac{P_0 T_0}{N_t}}\mathbf{C}_0 \tag{3}$$

for preliminary channel estimation at the LR, where $\mathbf{C}_0 \in \mathbb{C}^{T_0 \times N_t}$ is the training data matrix satisfying $\text{Tr}(\mathbf{C}_0^H\mathbf{C}_0) = N_t$ in which $\text{Tr}(\cdot)$ represents the trace of a matrix. The LR uses the received signal $\mathbf{Y}_0$ and the training signal $\mathbf{X}_0$ to obtain a preliminary estimate of $\mathbf{H}$, denoted by $\hat{\mathbf{H}}_0$, and sends $\hat{\mathbf{H}}_0$ back to the transmitter.

With knowledge of $\hat{\mathbf{H}}_0$ at the transmitter, we can then design the training signals that contain AN in the left null space of $\hat{\mathbf{H}}_0$ [11] in an attempt to corrupt the UR's channel estimate. Specifically, by assuming that $N_t > N_L$, the training signal in stage 1 is given by

$$\mathbf{X}_1 = \sqrt{\frac{P_1 T_1}{N_t}}\mathbf{C}_1 + \mathbf{A}_1 \cdot \mathbf{N}_{\hat{H}_0}^H \tag{4}$$

where $\mathbf{C}_1 \in \mathbb{C}^{T_1 \times N_t}$ represents the training data matrix satisfying $\text{Tr}(\mathbf{C}_1^H\mathbf{C}_1) = N_t$, $\mathbf{N}_{\hat{H}_0} \in \mathbb{C}^{N_t \times (N_t - N_L)}$ is a matrix that spans the left null space of $\hat{\mathbf{H}}_0$ and satisfies $\mathbf{N}_{\hat{H}_0}^H\mathbf{N}_{\hat{H}_0} = \mathbf{I}_{N_t - N_L}$ (the $(N_t - N_L)$ by $(N_t - N_L)$ identity matrix) and $\mathbf{N}_{\hat{H}_0}^H\hat{\mathbf{H}}_0 = \mathbf{0}$, and $\mathbf{A}_1 \in \mathbb{C}^{T_1 \times (N_t - N_L)}$ is an AN matrix with each entry being i.i.d. zero-mean, complex Gaussian random variables with variance equal to $\sigma_{a,1}^2$. It is worthwhile to notice that the LR may also suffer from the AN added in (4) since the estimate $\hat{\mathbf{H}}_0$ is in general not perfect [19]. Therefore, the allocation between the training data power P_1 and the AN power $\sigma_{a,1}^2$ should be designed carefully.

The training design rule used in (4) also applies to $\mathbf{X}_k$ for $k = 2, \ldots, K$, if multiple feedback-and-retraining stages are performed. This will be discussed in detail in Section IV. In the next section, we focus on the case of $K = 1$ (one feedback-and-retraining stage) and present a design criterion for the discrimination between the estimation performance of the LR and that of the UR.

III. Discriminatory Channel Estimation for $K = 1$

In this section, we first analyze the channel estimation performances of the LR and the UR, considering that the training signal design proposed in the previous section is used with $K = 1$. For simplicity of formulation, we assume that orthogonal training data are used, that is, $\mathbf{C}_0^H\mathbf{C}_0 = \mathbf{I}_{N_t}$ and $\mathbf{C}_1^H\mathbf{C}_1 = \mathbf{I}_{N_t}$. Then, we find the optimal set of training data powers P_0, P_1, and the AN power $\sigma_{a,1}^2$ that minimizes the NMSE of the LR subject to a lower limit constraint on the NMSE of the UR.

A. NMSE Analysis and Design Criterion

Let us assume that both the LR and the UR employ the LMMSE criterion[3] [23] for channel estimation. In that case, the preliminary channel estimate of $\mathbf{H}$ at the LR in the initial stage (i.e., stage 0) is given by

$$\hat{\mathbf{H}}_0 = \sigma_H^2\mathbf{X}_0^H\left(\sigma_H^2\mathbf{X}_0\mathbf{X}_0^H + \sigma_w^2\mathbf{I}_{T_0}\right)^{-1}\mathbf{Y}_0 \tag{5}$$

$$\triangleq \mathbf{H} + \Delta\mathbf{H}_0 \tag{6}$$

where $\Delta\mathbf{H}_0 \in \mathbb{C}^{N_t \times N_L}$ denotes the estimation error matrix. One can show that $\Delta\mathbf{H}_0$ has the correlation matrix given by [23]

$$\begin{aligned}\text{E}\left\{\Delta\mathbf{H}_0(\Delta\mathbf{H}_0)^H\right\} &= N_L\left(\frac{1}{\sigma_H^2}\mathbf{I}_{N_t} + \frac{P_0T_0}{N_t\sigma_w^2}\mathbf{C}_0^H\mathbf{C}_0\right)^{-1} \\ &= N_L\left(\frac{1}{\sigma_H^2} + \frac{P_0T_0}{N_t\sigma_w^2}\right)^{-1}\mathbf{I}_{N_t}.\end{aligned} \tag{7}$$

The NMSE of $\Delta\mathbf{H}_0$ is then defined as

$$\text{NMSE}_L^{(0)} \triangleq \frac{\text{Tr}\left(\text{E}\{\Delta\mathbf{H}_0(\Delta\mathbf{H}_0)^H\}\right)}{N_tN_L} = \left(\frac{1}{\sigma_H^2} + \frac{P_0T_0}{N_t\sigma_w^2}\right)^{-1}. \tag{8}$$

[2]The impact of time-varying channels on the proposed DCE scheme will be investigated by computer simulations in Section VI.

[3]Readers who are interested in the proposed DCE scheme using best linear unbiased estimation (BLUE) are referred to [22].

In the next stage (i.e., stage 1), the LR can make use of both $\mathbf{Y}_0$ and $\mathbf{Y}_1$ as well as knowledge of $\hat{\mathbf{H}}_0$ to refine its channel estimate. Specifically, by (1), (3), and (4), we have

$$\begin{aligned}\mathbf{Y} &\triangleq \begin{bmatrix}\mathbf{Y}_0\\ \mathbf{Y}_1\end{bmatrix}\\ &= \begin{bmatrix}\sqrt{\frac{P_0T_0}{N_t}}\mathbf{C}_0\\ \sqrt{\frac{P_1T_1}{N_t}}\mathbf{C}_1\end{bmatrix}\mathbf{H}+\begin{bmatrix}\mathbf{W}_0\\ \mathbf{A}_1\mathbf{N}_{\hat{H}_0}^H\mathbf{H}+\mathbf{W}_1\end{bmatrix}\\ &\overset{(a)}{=} \begin{bmatrix}\sqrt{\frac{P_0T_0}{N_t}}\mathbf{C}_0\\ \sqrt{\frac{P_1T_1}{N_t}}\mathbf{C}_1\end{bmatrix}\mathbf{H}+\begin{bmatrix}\mathbf{W}_0\\ -\mathbf{A}_1\mathbf{N}_{\hat{H}_0}^H\Delta\mathbf{H}_0+\mathbf{W}_1\end{bmatrix}\\ &\triangleq \bar{\mathbf{C}}\mathbf{H}+\bar{\mathbf{W}}\end{aligned} \tag{9}$$

where (a) follows from (6) and the fact that $\mathbf{N}_{\hat{H}_0}^H\hat{\mathbf{H}}_0=\mathbf{0}$. In (9), we have defined

$$\bar{\mathbf{C}}=\begin{bmatrix}\sqrt{\frac{P_0T_0}{N_t}}\mathbf{C}_0\\ \sqrt{\frac{P_1T_1}{N_t}}\mathbf{C}_1\end{bmatrix}\text{ and }\bar{\mathbf{W}}=\begin{bmatrix}\mathbf{W}_0\\ -\mathbf{A}_1\mathbf{N}_{\hat{H}_0}^H\Delta\mathbf{H}_0+\mathbf{W}_1\end{bmatrix}. \tag{10}$$

Applying the LMMSE criterion to (9), the LMMSE estimate of $\mathbf{H}$ in stage 1 is given by

$$\hat{\mathbf{H}}_1=\sigma_H^2\bar{\mathbf{C}}^H\left(\sigma_H^2\bar{\mathbf{C}}\bar{\mathbf{C}}^H+\mathbf{R}_{\bar{W}}\right)^{-1}\mathbf{Y} \tag{11}$$

where $\mathbf{R}_{\bar{W}}=\mathrm{E}\{\bar{\mathbf{W}}\bar{\mathbf{W}}^H\}$ is the correlation matrix of $\bar{\mathbf{W}}$. The associated NMSE of $\hat{\mathbf{H}}_1$ can be computed as [23]

$$\mathrm{NMSE}_L^{(1)}=\frac{\mathrm{Tr}\left(\left(\frac{1}{N_L\sigma_H^2}\mathbf{I}_{N_t}+\bar{\mathbf{C}}^H\mathbf{R}_{\bar{W}}^{-1}\bar{\mathbf{C}}\right)^{-1}\right)}{N_tN_L}. \tag{12}$$

By the statistical independence among $\mathbf{W}_0$, $\mathbf{W}_1$ and $\mathbf{A}_1$, one can show that

$$\mathbf{R}_{\bar{W}}=\begin{bmatrix}N_L\sigma_w^2\mathbf{I}_{T_0} & \mathbf{0}\\ \mathbf{0} & \left(\mathrm{E}\{||\mathbf{N}_{\hat{H}_0}^H\Delta\mathbf{H}_0||_F^2\}\sigma_{a,1}^2+N_L\sigma_w^2\right)\mathbf{I}_{T_1}\end{bmatrix}. \tag{13}$$

Recall that $\hat{\mathbf{H}}_0$ and $\Delta\mathbf{H}_0$ are statistically uncorrelated due to the orthogonality principle [23]. By (7), (8) and the fact that $\mathbf{N}_{\hat{H}_0}^H\mathbf{N}_{\hat{H}_0}=\mathbf{I}_{N_t-N_L}$, one can show that

$$\mathrm{E}\{||\mathbf{N}_{\hat{H}_0}^H\Delta\mathbf{H}_0||_F^2\}=N_L(N_t-N_L)\cdot\mathrm{NMSE}_L^{(0)}. \tag{14}$$

Substituting (13) and (14) into (12) yields (15) and (16), shown at the bottom of the page, where (a) follows from (8) and the premises $\mathbf{C}_0^H\mathbf{C}_0=\mathbf{I}_{N_t}$ and $\mathbf{C}_1^H\mathbf{C}_1=\mathbf{I}_{N_t}$.

The NMSE performance of the UR can be analyzed as follows. By (2), (3) and (4), the received signal at the UR is given by

$$\begin{aligned}\mathbf{Z}&\triangleq\begin{bmatrix}\mathbf{Z}_0\\ \mathbf{Z}_1\end{bmatrix}=\begin{bmatrix}\sqrt{\frac{P_0T_0}{N_t}}\mathbf{C}_0\\ \sqrt{\frac{P_1T_1}{N_t}}\mathbf{C}_1\end{bmatrix}\mathbf{G}+\begin{bmatrix}\mathbf{V}_0\\ \mathbf{A}_1\mathbf{N}_{\hat{H}_0}^H\mathbf{G}+\mathbf{V}_1\end{bmatrix}\\ &\triangleq\bar{\mathbf{C}}\mathbf{G}+\bar{\mathbf{V}}\end{aligned} \tag{17}$$

where $\bar{\mathbf{V}}$ is the noise matrix involving $\mathbf{V}_0$ and $\mathbf{V}_1$ in the first line of (17). By applying the LMMSE criterion, the UR can obtain the channel estimate as

$$\hat{\mathbf{G}}=\sigma_G^2\bar{\mathbf{C}}^H\left(\sigma_G^2\bar{\mathbf{C}}\bar{\mathbf{C}}^H+\mathrm{E}\{\bar{\mathbf{V}}\bar{\mathbf{V}}^H\}\right)^{-1}\mathbf{Z}\triangleq\mathbf{G}+\Delta\mathbf{G} \tag{18}$$

where the covariance matrix of $\bar{\mathbf{V}}$ can be computed as

$$\begin{aligned}\mathbf{R}_{\bar{V}}&\triangleq\mathrm{E}\{\bar{\mathbf{V}}\bar{\mathbf{V}}^H\}\\ &=\begin{bmatrix}N_U\sigma_v^2\mathbf{I}_{T_0} & \mathbf{0}\\ \mathbf{0} & N_U\left((N_t-N_L)\sigma_{a,1}^2\sigma_G^2+\sigma_v^2\right)\mathbf{I}_{T_1}\end{bmatrix}.\end{aligned} \tag{19}$$

Note from (18) and (19) that the UR can perform the optimal LMMSE channel estimation with no need of $\hat{\mathbf{H}}_0$, even though the UR is assumed able to intercept it. By (18) and (19), the NMSE of the UR in stage 1 is given by

$$\begin{aligned}\mathrm{NMSE}_U^{(1)}&\triangleq\frac{\mathrm{Tr}\left(\mathrm{E}\{\Delta\mathbf{G}(\Delta\mathbf{G})^H\}\right)}{N_tN_U}\\ &=\frac{\mathrm{Tr}\left(\left(\frac{1}{N_U\sigma_G^2}\mathbf{I}_{N_t}+\bar{\mathbf{C}}^H\mathbf{R}_{\bar{V}}^{-1}\bar{\mathbf{C}}\right)^{-1}\right)}{N_tN_U} &(20)\\ &=\left(\frac{1}{\mathrm{NMSE}_U^{(0)}}+\frac{\frac{P_1T_1}{N_t}}{(N_t-N_L)\sigma_{a,1}^2\sigma_G^2+\sigma_v^2}\right)^{-1} &(21)\end{aligned}$$

where

$$\mathrm{NMSE}_U^{(0)}=\left(\frac{1}{\sigma_G^2}+\frac{P_0T_0}{N_t\sigma_v^2}\right)^{-1} \tag{22}$$

is the NMSE of the UR when using only the training signal in stage 0 for channel estimation. Note from (21) and (22) that

$$\mathrm{NMSE}_U^{(1)}<\mathrm{NMSE}_U^{(0)} \tag{23}$$

$$\mathrm{NMSE}_L^{(1)}=\frac{N_L\mathrm{Tr}\left(\left(\frac{1}{\sigma_H^2}\mathbf{I}_{N_t}+\frac{P_0T_0}{\sigma_w^2N_t}\mathbf{C}_0^H\mathbf{C}_0+\frac{\frac{P_1T_1}{N_t}}{\mathrm{NMSE}_L^{(0)}(N_t-N_L)\sigma_{a,1}^2+\sigma_w^2}\mathbf{C}_1^H\mathbf{C}_1\right)^{-1}\right)}{N_tN_L} \tag{15}$$

$$\overset{(a)}{=}\left(\frac{1}{\mathrm{NMSE}_L^{(0)}}+\frac{\frac{P_1T_1}{N_t}}{\mathrm{NMSE}_L^{(0)}\cdot(N_t-N_L)\sigma_{a,1}^2+\sigma_w^2}\right)^{-1} \tag{16}$$

whenever $P_1T_1 > 0$. This implies that, compared to using $\mathbf{Z}_0$ only, the UR can obtain better channel estimation performance by using both $\mathbf{Z}_0$ and $\mathbf{Z}_1$, even though $\mathbf{Z}_1$ is poisoned by AN.

With (16) and (21), we propose to find the optimal set of training data powers P_0, P_1 and AN power $\sigma_{a,1}^2$ by solving the following optimization problem:

$$\min_{P_0,P_1,\sigma_{a,1}^2\geq 0} \quad \mathrm{NMSE}_{\mathrm{L}}^{(1)} \tag{24a}$$

$$\text{s.t.} \quad \mathrm{NMSE}_{\mathrm{U}}^{(1)} \geq \gamma, \tag{24b}$$

$$\frac{\mathrm{E}\{\|\mathbf{X}_0\|_F^2\}+\mathrm{E}\{\|\mathbf{X}_1\|_F^2\}}{T_0+T_1} \leq P_{\mathrm{ave}}, \tag{24c}$$

where $\gamma > 0$ is the preassigned lower limit on the UR's achievable NMSE, and $P_{\mathrm{ave}} > 0$ is the maximum average transmission power. By (3) and (4), the constraint in (24c) becomes

$$\begin{aligned}\mathrm{E}\{\|\mathbf{X}_0\|_F^2\}+\mathrm{E}\{\|\mathbf{X}_1\|_F^2\} &= P_0T_0+\left(P_1+(N_t-N_L)\sigma_{a,1}^2\right)T_1 \\ &\leq P_{\mathrm{ave}}(T_0+T_1)\end{aligned} \tag{25}$$

which represents an average energy constraint on the training signals over both stages. We can see that the design criterion in (24) aims to minimize the NMSE of the LR while enforcing the NMSE of the UR to be no less than γ.

It is interesting to remark that, when

$$\gamma < \left(\frac{1}{\sigma_G^2}+\frac{P_{\mathrm{ave}}(T_0+T_1)}{N_t\sigma_v^2}\right)^{-1} \tag{26}$$

where the right-hand side is the minimum NMSE achievable at the UR without any interference from the AN, problem (24) yields the trivial solution where $\sigma_{a,1}^2 = 0$ and P_0, P_1 (≥ 0) are any values that satisfy $P_0T_0 + P_1T_1 = P_{\mathrm{ave}}(T_0 + T_1)$. This implies that no AN is needed and thus there is no performance discrimination between receivers. Moreover, by (22), (23), and (24b), it must hold that

$$\gamma \leq \mathrm{NMSE}_{\mathrm{U}}^{(1)} \leq \mathrm{NMSE}_{\mathrm{U}}^{(0)} = \left(\frac{1}{\sigma_G^2}+\frac{P_0T_0}{N_t\sigma_v^2}\right)^{-1} \overset{\mathrm{(a)}}{\leq} \sigma_G^2 \tag{27}$$

where (a) follows from the fact that $P_0T_0 \geq 0$. Hence, in the following, we shall consider only the interesting case of

$$\left(\frac{1}{\sigma_G^2}+\frac{P_{\mathrm{ave}}(T_0+T_1)}{N_t\sigma_v^2}\right)^{-1} \leq \gamma \leq \sigma_G^2. \tag{28}$$

B. Optimal Power Allocation

To illustrate how problem (24) can be solved, let us define $a = P_0T_0$, $b = P_1T_1$, $c = (N_t - N_L)\sigma_{a,1}^2$ and

$$\tilde{\gamma} \triangleq \left(\frac{1}{\gamma}-\frac{1}{\sigma_G^2}\right)N_t\sigma_v^2 \geq 0. \tag{29}$$

In this case, the condition in (28) reduces to

$$0 \leq \tilde{\gamma} \leq P_{\mathrm{ave}}(T_0+T_1). \tag{30}$$

Then we can reformulate problem (24) into the following maximization problem:

$$\max_{a,b,c\geq 0} \quad a+\frac{(N_t\sigma_w^2+\sigma_H^2\cdot a)\cdot b}{N_t\sigma_w^2+\sigma_H^2\cdot a+N_t\sigma_H^2\cdot c} \tag{31a}$$

$$\text{s.t.} \quad a+\frac{\sigma_v^2\cdot b}{\sigma_G^2\cdot c+\sigma_v^2} \leq \tilde{\gamma}, \tag{31b}$$

$$a+T_1\cdot c+b \leq P_{\mathrm{ave}}(T_0+T_1). \tag{31c}$$

By close inspection of the problem structure of (31), we show in the following proposition that the three-dimensional optimization problem can be solved by a simple one-dimensional line search. The proof is given in Appendix A.

Proposition 1: *Let $\{a^\star, b^\star, c^\star\}$ be the optimum solution to the nonconvex optimization problem in (31) with $0 \leq \tilde{\gamma} \leq P_{\mathrm{ave}}(T_0+T_1)$. For*

$$\eta \triangleq N_t\left(\frac{\sigma_v^2}{\sigma_G^2}-\frac{\sigma_w^2}{\sigma_H^2}\right) > \tilde{\gamma}, \tag{32}$$

the optimal solution to problem (31) is given by $c^\star = 0$ (i.e., no AN is needed) and any $a^\star$, $b^\star \geq 0$ such that $a^\star + b^\star = \tilde{\gamma}$ (e.g., $a^\star = b^\star = \tilde{\gamma}/2$). On the other hand, for $\eta \leq \tilde{\gamma}$, the optimum value of a (i.e., $a^\star$) can be obtained by solving the following one-variable optimization problem:

$$\max_{a} \quad a+\frac{(N_t\sigma_w^2+\sigma_H^2 a)\cdot b(a)}{N_t\sigma_w^2+N_t\sigma_H^2\cdot c(a)+\sigma_H^2 a} \tag{33a}$$

$$\text{s.t.} \quad \max\{\eta,0\} \leq a \leq \tilde{\gamma} \tag{33b}$$

where

$$c(a) = \frac{P_{\mathrm{ave}}(T_0+T_1)-\tilde{\gamma}}{T_1+\sigma_G^2\left(\frac{\tilde{\gamma}-a}{\sigma_v^2}\right)} \tag{34}$$

$$b(a) = \sigma_G^2\left(\frac{\tilde{\gamma}-a}{\sigma_v^2}\right)c(a)+\tilde{\gamma}-a. \tag{35}$$

The associated values of $b^\star$ and $c^\star$ are given by $b(a^\star)$ and $c(a^\star)$, respectively.

Proposition 1 implies that a globally optimal solution of problem (31) and, thus, problem (24) can be conveniently obtained via a simple line search with respect to a over the interval $[\max\{\eta,0\},\tilde{\gamma}]$, provided that $\eta \leq \tilde{\gamma}$; otherwise a simple closed-form solution can be readily obtained with $c^\star = 0$ and, e.g., $a^\star = b^\star = \tilde{\gamma}/2$. To gain more insights from Proposition 1, one may interpret η in (32) as a measure of channel quality difference between the UR and the LR. When $\eta > \tilde{\gamma}$, it implies that the UR is under a much worse channel condition (either with a larger noise power or a smaller channel variance) than the LR, so there is no need to use AN to interfere with the UR. However, when $\eta \leq \tilde{\gamma}$ or even $\eta \leq 0$, which implies that the UR has a comparable or even better channel condition than the LR, additional AN has to be transmitted in order to limit the UR's channel estimation performance.

IV. Discriminatory Channel Estimation for $K > 1$

We notice that, even with the optimal power allocation, the NMSE performance of the LR may be limited if we are only allowed to perform one stage of feedback and retraining. This is true especially when γ is set to a high value or when the UR has a much higher signal-to-noise ratio (SNR) than the LR (e.g., when $\sigma_v^2/\sigma_G^2 \ll \sigma_w^2/\sigma_H^2$). In either of these cases, the training energy P_0T_0 in stage 0 must be small in order to meet the constraint in (24b), which then degrades the quality of the LR's preliminary channel estimate. The lack of accuracy in the preliminary channel estimate (i.e., the large value of $\mathrm{NSME}_{\mathrm{L}}^{(0)}$) will restrict the use of AN in stage 1 and thereby limit the ability of the training scheme to discriminate the LR's and the UR's performances [see (16)]. Fortunately, this problem can be resolved by performing multiple stages of feedback and retraining as described in this section.

We consider the case where the channel estimate feedback and retraining process is repeated K times $(K > 1)$ so that there is a total of $K+1$ stages of training when including the initial stage. For $k > 0$, let $\hat{\mathbf{H}}_{k-1}$ be the channel estimate obtained at the LR from the observations $\mathbf{Y}_0, \ldots, \mathbf{Y}_{k-1}$ and training data $\mathbf{C}_0, \ldots, \mathbf{C}_{k-1}$. In the kth stage of training, the discriminatory training signal $\mathbf{X}_k$ is given by

$$\mathbf{X}_k = \sqrt{\frac{P_kT_k}{N_t}}\mathbf{C}_k + \mathbf{A}_k\mathbf{N}_{\hat{H}_{k-1}}^H \tag{36}$$

where $\mathbf{C}_k \in \mathbb{C}^{T_k \times N_t}$ is the deterministic training data matrix satisfying $\mathbf{C}_k^H\mathbf{C}_k = \mathbf{I}_{N_t}$, $\mathbf{N}_{\hat{H}_{k-1}} \in \mathbb{C}^{N_t \times (N_t - N_L)}$ is a matrix which spans the left null space of $\hat{\mathbf{H}}_{k-1}$, and the AN matrix $\mathbf{A}_k \in \mathbb{C}^{T_k \times (N_t - N_L)}$ consists of i.i.d. complex Gaussian random variables with zero mean and variance equal to $\sigma_{a,k}^2$. Denote by

$$\Delta\mathbf{H}_{k-1} = \hat{\mathbf{H}}_{k-1} - \mathbf{H} \tag{37}$$

the estimation error matrix at stage $k-1$. For the case with multiple times of feedback and retraining, the signal received at the LR over the $K+1$ stages is given by

$$\mathbf{Y}_0 = \sqrt{\frac{P_0T_0}{N_t}}\mathbf{C}_0\mathbf{H} + \mathbf{W}_0 \tag{38}$$

$$\begin{aligned}\mathbf{Y}_k &= \sqrt{\frac{P_kT_k}{N_t}}\mathbf{C}_k\mathbf{H} + \mathbf{A}_k\mathbf{N}_{\hat{H}_{k-1}}^H\mathbf{H} + \mathbf{W}_k \\ &= \sqrt{\frac{P_kT_k}{N_t}}\mathbf{C}_k\mathbf{H} - \mathbf{A}_k\mathbf{N}_{\hat{H}_{k-1}}^H\Delta\mathbf{H}_{k-1} + \mathbf{W}_k\end{aligned} \tag{39}$$

for $k = 1, \ldots, K$. Following the analysis as presented in Section III-A, one can show by induction that the NMSE obtained by the LR through K feedback-and-retraining stages can be expressed in the following recursive form

$$\mathrm{NMSE}_L^{(K)} = \left(\frac{1}{\mathrm{NMSE}_L^{(K-1)}} + \frac{\frac{P_KT_K}{N_t}}{\mathrm{NMSE}_L^{(K-1)} \cdot (N_t - N_L)\sigma_{a,K}^2 + \sigma_w^2}\right)^{-1}. \tag{40}$$

with the initial $\mathrm{NMSE}_{\mathrm{L}}^{(0)}$ given in (8). Similarly, the NMSE of the UR at stage K can be obtained as

$$\begin{aligned}&\mathrm{NMSE}_U^{(K)} \\ &\quad= \left(\frac{1}{\mathrm{NMSE}_U^{(K-1)}} + \frac{\frac{P_KT_K}{N_t}}{(N_t - N_L)\sigma_{a,K}^2\sigma_G^2 + \sigma_v^2}\right)^{-1} \\ &\quad= \left(\frac{1}{\sigma_G^2} + \frac{P_0T_0}{N_t\sigma_v^2} + \sum_{k=1}^{K}\frac{\frac{P_kT_k}{N_t}}{(N_t - N_L)\sigma_{a,k}^2\sigma_G^2 + \sigma_v^2}\right)^{-1}.\end{aligned} \tag{41}$$

With (40) and (41), we can jointly optimize the power values of $\{P_0, P_1, \ldots, P_K, \sigma_{a,1}^2, \ldots, \sigma_{a,K}^2\}$, by using a design criterion as in (24). Specifically, we consider the following optimization problem:

$$\min_{P_0, P_k, \sigma_{a,k}^2 \geq 0,\ k=1,\ldots,K} \quad \mathrm{NMSE}_L^{(K)} \tag{42a}$$

$$\text{s.t.} \quad \mathrm{NMSE}_U^{(K)} \geq \gamma, \tag{42b}$$

$$\sum_{k=0}^{K}\mathrm{E}\{\|\mathbf{X}_k\|_F^2\} \leq P_{\mathrm{ave}}\bar{T}, \tag{42c}$$

where $\bar{T} = \sum_{k=0}^{K} T_k$, and (42c) is the average energy constraint. By (36), the constraint in (42c) can be expressed as

$$P_0T_0 + \sum_{k=1}^{K}\left(P_kT_k + \sigma_{a,k}^2(N_t - N_L)T_k\right) \leq P_{\mathrm{ave}}\bar{T}. \tag{43}$$

In comparison with (24), the optimization problem (42) is much more involved due to the recursive structure in (40). In fact, (42) is a nonconvex optimization problem, and the globally optimal solution becomes intractable as $K > 1$. However, we present in Appendix B that the problem in (42) can be handled efficiently by using the monomial approximation and the condensation method (a successive convex approximation method) in the context of geometric programming (GP) [20]. The condensation method basically involves solving a sequence of convex GPs, provided that a feasible initial power allocation is given [24]. Since GP can be efficiently solved by interior point methods in a polynomial-time complexity [21], an approximate solution to problem (42) can be efficiently and reliably obtained. For example, one can use the MATLAB optimization toolbox CVX [25] to solve the GPs. The condensation method for solving problem (42) is summarized in Table I.

An initial set of feasible power values of problem (42) can be conveniently obtained as follows. Suppose that one has obtained the optimum solutions $P_0^\star$ and $P_1^\star$ and $(\sigma_{a,1}^\star)^2$ of problem (24) (with $K = 1$) through the simple line search method (see Proposition 1). We can tentatively obtain a set of power values as

$$P_k^\star = \frac{P_{\mathrm{ave}}P_1^\star}{P_1^\star + (N_t - N_L)(\sigma_{a,1}^\star)^2} \tag{44}$$

$$(\sigma_{a,k}^\star)^2 = \frac{P_{\mathrm{ave}}(\sigma_{a,1}^\star)^2}{P_1^\star + (N_t - N_L)(\sigma_{a,1}^\star)^2} \tag{45}$$

for $k = 2, \ldots, K$. It can be verified by (43) that the set of power values $\{P_0^\star, P_1^\star, \ldots, P_k^\star, (\sigma_{a,1}^\star)^2, \ldots, (\sigma_{a,k}^\star)^2\}$ is feasible to

TABLE I
CONDENSATION METHOD FOR PROBLEM (42)

Given an initial set of feasible power values $\{P_0, P_1, \ldots, P_K, \sigma_{a,1}^2, \ldots, \sigma_{a,K}^2\}$, and a solution accuracy $\epsilon > 0$.

Step 1. Set

$$\bar{a}_0 = \left(\frac{1}{\sigma_H^2} + \frac{P_0 T_0}{N_t \sigma_w^2}\right)^{-1},\ \bar{t}_0 = \bar{a}_0^{-1}$$
$$\bar{a}_k = \frac{P_k T_k}{N_t},\ \bar{b}_k = \sigma_G^2 \sigma_{a,k}^2 (N_t - N_L) + \sigma_v^2,$$
$$\bar{t}_k = \bar{t}_{k-1} + \frac{\bar{t}_{k-1} P_k T_k / N_t}{(N_t - N_L)\sigma_{a,k}^2 + \bar{t}_{k-1}\sigma_w^2},\ k = 1, \ldots, K,$$

Step 2. Compute $\{\xi_{k,i}\}_{i=1}^4$, $k = 1, \ldots, K$, by (A.17) and (A.18).

Step 3. Solve (A.20) (e.g., by CVX [25]), and obtain the optimal solution as $\{a_0^\star, a_k^\star, b_k^\star, t_0^\star, t_k^\star,\ k = 1, \ldots, K\}$.

Step 4. If $(t_K^\star - \bar{t}_K)/\bar{t}_K > \epsilon$, then let

$$\bar{a}_0 = a_0^\star,\ \bar{t}_0 = t_0^\star,$$
$$\bar{a}_k = a_k^\star,\ \bar{b}_k = b_k^\star,\ \bar{t}_k = t_k^\star,\ k = 1, \ldots, K,$$

and go to Step 2; otherwise, output

$$P_0^\star = \left(\frac{1}{a_0^\star} - \frac{1}{\sigma_H^2}\right)\frac{N_t \sigma_w^2}{T_0},\ P_k^\star = \frac{a_k^\star N_t}{T_k},$$
$$(\sigma_{a,k}^\star)^2 = \frac{b_k^\star - \sigma_v^2}{\sigma_G^2 (N_t - N_L)},\ k = 1, \ldots, K,$$

as an approximate solution to problem (42).

(42c) [(43)] but may not satisfy the constraint in (42b). To solve this, one can simultaneously scale down $\{P_0^\star, P_1^\star, \ldots, P_K^\star\}$ such that (42b) holds with equality to obtain an initial set of feasible power values for problem (42). By our experience in simulations, this simple initialization approach works well for the condensation method in Table I.

V. DISCRIMINATORY CHANNEL ESTIMATION WITH MULTIPLE LRS AND URS

In the previous sections, we have focused on the scenario where there is only one LR and one UR. In this section, we extend the proposed DCE scheme to the case with multiple LRs and multiple URs.

Assume that there are M_L LRs, each equipped with N_L antennas, and M_U URs, each equipped with N_U antennas, and that $N_t > M_L N_L$. Let $\mathbf{H}_i \in \mathbb{C}^{N_t \times N_L}$ denote the MIMO channel matrix between the transmitter and the ith LR, and let $\mathbf{G}_j \in \mathbb{C}^{N_t \times N_U}$ denote the MIMO channel matrix between the transmitter and the jth UR. The elements of $\mathbf{H}_i$ and $\mathbf{G}_j$ are i.i.d. random variables with zero mean and variances equal to $\sigma_{H,i}^2$ and $\sigma_{G,j}^2$, respectively. The received signals at LR i and UR j are respectively given by

$$\text{LR } i: \mathbf{Y}_{i,k} = \mathbf{X}_k \mathbf{H}_i + \mathbf{W}_{i,k}, \quad i = 1, \ldots, M_L \tag{46}$$
$$\text{UR } j: \mathbf{Z}_{j,k} = \mathbf{X}_k \mathbf{G}_j + \mathbf{V}_{j,k}, \quad j = 1, \ldots, M_U \tag{47}$$

for $k = 0, \ldots, K$, where $\mathbf{W}_{i,k} \in \mathbb{C}^{T_k \times N_U}$ and $\mathbf{V}_{j,k} \in \mathbb{C}^{T_k \times N_U}$ are the AWGN with the power of each entry equal to $\sigma_{w,i}^2$ and $\sigma_{v,j}^2$, respectively. Let $\hat{\mathbf{H}}_{i,k} \in \mathbb{C}^{N_t \times N_L}$ be the channel estimate of the ith LR obtained at stage k, and assume that all the LRs send back their channel estimates to the transmitter. Then similar to (3) and (36), the training signals are given by

$$\mathbf{X}_0 = \sqrt{\frac{P_0 T_0}{N_t}}\mathbf{C}_0 \tag{48}$$
$$\mathbf{X}_k = \sqrt{\frac{P_k T_k}{N_t}}\mathbf{C}_k + \mathbf{A}_k \mathbf{N}_{\hat{\mathcal{H}}_{k-1}}^H \tag{49}$$

where $\mathbf{N}_{\hat{\mathcal{H}}_{k-1}} \in \mathbb{C}^{N_t \times (N_t - M_L N_L)}$ is a matrix that spans the left null space of the channel estimate

$$\hat{\mathcal{H}}_{k-1} = [\hat{\mathbf{H}}_{1,k-1}, \hat{\mathbf{H}}_{2,k-1}, \ldots, \hat{\mathbf{H}}_{M_L,k-1}] \in \mathbb{C}^{N_t \times M_L N_L}. \tag{50}$$

Since the training signals are broadcast to all the receivers, both LRs and URs can use the same training signals $\mathbf{X}_0, \ldots, \mathbf{X}_K$ to perform channel estimation. Hence, the NMSE of the LRs and that of the URs have the same structures as those in (40) and (41), respectively. The training design formulated in (24) and (42) can then be extended to the case with multiple LRs and multiple URs by minimizing the worst NMSE performance among LRs subject to NMSE constraints on all the URs. The design formulation is given as follows

$$\min_{\substack{P_0, P_k, \sigma_{a,k}^2 \geq 0, \\ k=1,\ldots,K}} \left\{\max_{i=1,\ldots,M_L} \text{NMSE}_L^{(i,K)}\right\} \tag{51a}$$
$$\text{s.t.}\ \ \text{NMSE}_U^{(j,K)} \geq \gamma,\ \forall\, j = 1, \ldots, M_U, \tag{51b}$$
$$\sum_{k=0}^{K} \mathrm{E}\{\|\mathbf{X}_k\|_F^2\} \leq P_{\text{ave}} \bar{T}. \tag{51c}$$

Here, $\text{NMSE}_L^{(i,K)}$ and $\text{NMSE}_U^{(i,K)}$ denote the NMSEs of the ith LR and the jth UR at stage K, respectively. Problem (51) can be efficiently approximated by the monomial approximation and condensation method in a way similar to that presented in Section IV and Appendix B.

The design problem in (51) can be simplified for the special case where $\sigma_{H,1}^2 = \sigma_{H,2}^2 = \cdots = \sigma_{H,M_L}^2$ and $\sigma_{G,1}^2 = \sigma_{G,2}^2 = \cdots = \sigma_{G,M_U}^2$. In particular, by (40) and (41) one can see that the LR with the largest $\sigma_{w,i}^2$ would have the largest $\text{NMSE}_L^{(i,K)}$, and the UR with the smallest $\sigma_{v,j}^2$ would have the smallest $\text{NMSE}_U^{(i,K)}$. In this case, one needs only to consider the performances of these two users in the training signal design. For example, if $\sigma_{w,1}^2 \leq \sigma_{w,2}^2 \leq \cdots \leq \sigma_{w,M_L}^2$ and $\sigma_{v,1}^2 \leq \sigma_{v,2}^2 \leq \cdots \leq \sigma_{v,M_U}^2$, then the design problem in (51) can be reduced to

$$\min_{P_0, P_k, \sigma_{a,k}^2 \geq 0,\ k=1,\ldots,K} \quad \text{NMSE}_L^{(M_L,K)} \tag{52a}$$
$$\text{s.t.}\ \ \text{NMSE}_U^{(1,K)} \geq \gamma, \tag{52b}$$
$$\sum_{k=0}^{K} \mathrm{E}\{\|\mathbf{X}_k\|_F^2\} \leq P_{\text{ave}} \bar{T} \tag{52c}$$

which is similar to problem (42) in this special case.

VI. SIMULATION RESULTS AND DISCUSSIONS

In this section, simulation results are presented to demonstrate the efficacy of the proposed training signal design and DCE scheme. We first consider the wireless system shown in

Fig. 1 with one LR and one UR. We assumed that the transmitter has four antennas ($N_t = 4$), and both the LR and the UR have two antennas ($N_L = N_U = 2$). The elements of the channel matrices $\mathbf{H}$ and $\mathbf{G}$ were i.i.d. complex Gaussian random variables with zero mean and unit variance ($\sigma_H^2 = \sigma_G^2 = 1$). The training data matrices $\mathbf{C}_k, k = 0, \ldots, K$, were randomly drawn from semi-unitary $T_k \times N_t$ matrices. The average power P_{ave} was set to 30 dBm ($P_{\text{ave}} = 1$) and the training signal length per stage was set to

$$T_0 = T_0 = \cdots = T_K = \left\lfloor \frac{300}{K+1} \right\rfloor. \tag{53}$$

Note from (53) that we have partitioned the training sequence into $(K+1)$ equal-length segments, and fixed the total training length (which was 300)[4] for any choice of K. Except for the 4th simulation example below, we assumed that the channel matrices $\mathbf{H}$ and $\mathbf{G}$ were static during the whole training process. In the simulation, we consider an "NMSE lower bound"

$$\text{NMSE lower bound} = \left(\frac{1}{\sigma_H^2} + \frac{P_{\text{ave}} \sum_{k=0}^{K} T_k}{N_t \sigma_w^2} \right)^{-1} \tag{54}$$

which represents the best NMSE performance achievable at the LR without using any AN (i.e., $\sigma_{a,k}^2 = 0$ for all k). Note that this bound also corresponds to the NMSE performance achieved by the conventional (non-discriminatory) MIMO channel estimation scheme [17]. With $\sigma_H^2 = \sigma_G^2 = 1$ and $P_{\text{ave}} = 1$, the SNRs at the LR and the UR were defined as

$$\text{SNR}_{\text{L}} = \frac{\sum_{k=0}^{K} \text{E}\{\|\mathbf{X}_k \mathbf{H}\|_F^2\}}{\sum_{k=0}^{K} \text{E}\{\|\mathbf{W}_k\|_F^2\}} = \frac{1}{\sigma_w^2} \tag{55}$$

$$\text{SNR}_{\text{U}} = \frac{\sum_{k=0}^{K} \text{E}\{\|\mathbf{X}_k \mathbf{G}\|_F^2\}}{\sum_{k=0}^{K} \text{E}\{\|\mathbf{V}_k\|_F^2\}} = \frac{1}{\sigma_v^2} \tag{56}$$

respectively. Each simulation result was obtained by averaging over 1000 channel realizations.

1) Example 1: In this example, we examine the performance of the proposed DCE scheme by assuming that $\text{SNR}_{\text{L}} = \text{SNR}_{\text{U}}$. In Fig. 2(a), we present the NMSE performances of the LR and the UR for $\gamma = 0.1$ and $\gamma = 0.03$ with only one stage of feedback and retraining (i.e., with $K = 1$). The optimal power values P_0, P_1 and $\sigma_{a,1}^2$ are obtained via (24) and Proposition 1. First, one can see from Fig. 2(a) that NMSEs of the UR are successfully constrained above 0.1 and 0.03, respectively. Second, we see that the NMSE performance discrimination between the LR and the UR is limited for $\gamma = 0.1$; while it is more appreciable for $\gamma = 0.03$, demonstrating a tradeoff between the value of γ and the achievable NMSE of the LR. We can also see from the figure that a large gap exists between the NMSE attainable by the LR and the NMSE lower bound. This result implies that the LR's channel estimation performance must be sacrificed compared to the conventional training scheme in order to constrain the UR's channel estimation performance. The NMSE performance of the LR and the UR for $K = 11$ are displayed in Fig. 2(b). The optimal power values $\{P_k\}_{k=0}^K$ and $\{\sigma_{a,k}^2\}_{k=1}^K$ were calculated via (42) using the condensation method in Table I. The solution accuracy ϵ in Table I was set to 10^{-3} ($\epsilon = 10^{-3}$). By comparing Fig. 2(b) with Fig. 2(a), it can be seen that the LR's NMSE performance is greatly improved for both $\gamma = 0.1$ and for $\gamma = 0.03$, whereas the NMSEs of the UR are still above the specified level. Moreover, we can see that the gap between the LR's NMSE and the NMSE lower bound is significantly reduced for the case of $\gamma = 0.03$, showing the efficacy of the proposed DCE scheme.

Fig. 2(c) illustrates how the NMSEs of the LR decreases with K (i.e., the number of feedback-and-retraining stages) for $\text{SNR}_{\text{L}} = 20$ dB. As can be seen from this figure, a significant NMSE performance discrimination can be achieved with $K = 3$, showing that the proposed DCE scheme actually works well even with a small number of K.

To illustrate how multiple stages of feedback and retraining can improve the DCE performance, we plot the optimized power values of $\{P_k\}_{k=0}^K$ and $\{\sigma_{a,k}^2\}_{k=1}^K$ in Fig. 2(d) for $\gamma = 0.03$ and $\text{SNR}_{\text{L}} = \text{SNR}_{\text{U}} = 25$ dB. We can see that the optimized P_0 is relatively small in order to limit the UR's best NMSE performance to 0.03. After that, the optimized P_k as well as AN powers $\sigma_{a,k}^2$ monotonically increases since the $\text{NMSE}_L^{(k)}$ gradually decreases from one training stage to another [see (40)].

2) Example 2: In this example, we examine the detection performances of receivers using the channel estimates obtained with the proposed DCE scheme. We considered the scenario where the transmitter sends to the LR a 4-by-4 complex orthogonal space-time block code (OSTBC) which has $N_t = 4$ and $T = 4$ (the code length), and contains three QAM source symbols per code block [27]. Both the LR and the UR will use their channel estimates obtained with DCE to decode the unknown symbols.[5] The simulation was conducted with $N_L = N_U = 2$ and $\text{SNR}_{\text{L}} = \text{SNR}_{\text{U}}$. The average symbol error rates (SERs) of the LR and the UR were obtained by averaging over 50 000 channel realizations and OSTBCs. Fig. 3(a) presents the associated average SERs for 64-QAM OSTBC and $\gamma = 0.1$. Note that we only plot the SER curve of the UR for $K = 1$ since the UR has almost the same average symbol error performances for all the values of K. One can see from this figure that, with increased K, the SERs of the LR gradually improve while that of the UR remains around 0.5. We should mention that the transmission rates in the simulation are kept the same for all K since the total training length has been fixed to 300 [see (53)]. In Fig. 3(b), the associated SERs for 64-QAM OSTBC and $\gamma = 0.03$ are presented. In comparison with Fig. 3(a), one can see that the performance discrimination becomes much more evident for $\gamma = 0.03$, even with one feedback-and-retraining stage only ($K = 1$). For $K \geq 3$, we see from this figure that the LR can achieve SER performance close to that with perfect CSI; whereas the UR's SER is still limited to around 0.1. In Fig. 3(c), we further display the SERs for 256-QAM OSTBC and $\gamma = 0.03$. Similar performance trends can be observed. It is worthwhile to note that the SER results in Fig. 3 are consistent with the channel estimation performance in Fig. 2(c).

3) Example 3: In this example, we consider a wireless system with 2 LRs (say LR 1 and LR 2) and 3 URs (say UR 1, UR 2,

[4] In IEEE 802.11a wireless LAN systems [26], the training sequence length for channel acquisition is around 284 samples.

[5] It is worthwhile to note that square OSTBCs (i.e., with $N_t = T$) in general cannot be properly decoded without CSI at the receiver. See [28] and [16] for the details. Therefore, the UR has to use its channel estimate for symbol decoding.

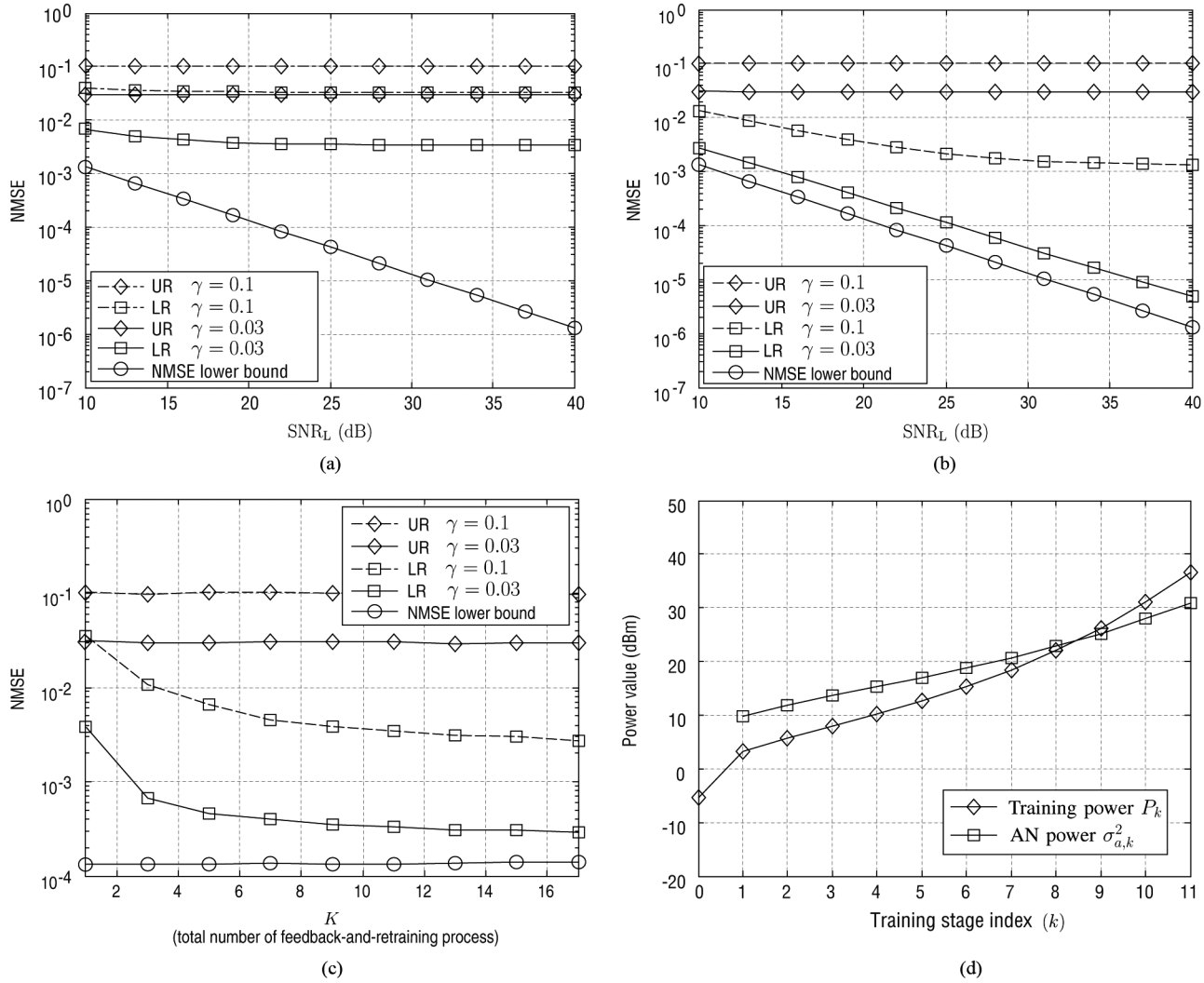


Fig. 2. Simulation results of NMSE performances of the proposed DCE scheme for $N_t = 4$, $N_L = 2$, $N_U = 2$ and $\mathrm{SNR_L} = \mathrm{SNR_U}$: (a) $K = 1$, (b) $K = 11$, (c) $\mathrm{SNR_L} = 20$ dB, and (d) $\gamma = 0.03$, $K = 11$, $\mathrm{SNR_L} = 25$ dB.

and UR 3) (see Section V). We set $N_t = 6$ and $N_L = N_U = 2$, and let all the channels of LRs and URs have unit variance (i.e., $\sigma^2_{H,1} = \sigma^2_{H,2} = 1$ and $\sigma^2_{G,1} = \sigma^2_{G,2} = \sigma^2_{G,3} = 1$). The SNRs of LR 1 and LR 2 are set to 30 and 20 dB, respectively. The SNRs of UR 1, UR 2 and UR 3 were set to 40, 30, and 20 dB. According to (52), we only consider to minimize the NMSE of LR 2 subject to an NMSE constraint on UR 1. Fig. 4 displays the NMSE performances of LRs and URs for $\gamma = 0.03$. We can see from this figure that the NMSEs of all URs are successfully constrained above 0.03, while the NMSEs of the two LRs rapidly decrease as K increases. Moreover, LR 1 has better NMSE performance than LR 2 since the former has a higher SNR value.

4) Example 4: As the final simulation example, we investigate the performance of the proposed DCE scheme in time-varying channels. The motivation of conducting such a simulation is that the channel from the transmitter to the LR and that from the transmitter to the UR may have changed from one stage to another during the DCE process. For simplicity, we assumed one LR and one UR, and set $N_t = 4$, $N_L = N_U = 2$, and $\mathrm{SNR_L} = \mathrm{SNR_U}$ in this simulation example. Let us define $\mathbf{H}[t]$ and $\mathbf{G}[t]$ as the transmitter-to-LR and transmitter-to-UR channel matrices at time t for $t = 1, \ldots, 300$. We assumed that each of the entries of $\mathbf{H}[t]$ and $\mathbf{G}[t]$ vary from one time sample to another following the rule of Jakes' channel model [29]. Let $H_{i,j}[t]$ and $G_{i,j}[t]$ be the (i,j)th entry of $\mathbf{H}[t]$ and $\mathbf{G}[t]$, respectively. Suppose that the LR and the UR have the same maximum Doppler frequency f_D. The autocorrelation of $H_{i,j}[t]$ and that of $G_{i,j}[t]$ by Jakes' model are given by $\mathrm{E}\{H^*_{i,j}[t]H_{i,j}[t+1]\} = \mathrm{E}\{G^*_{i,j}[t]G_{i,j}[t+1]\} = J_0(2\pi f_D T_s)$ where the superscript "*" denotes the complex conjugate, $J_0(\cdot)$ is the zero-order Bessel function of the first kind and $1/T_s$ is the system sampling rate. Since both $\mathbf{H}[t]$ and $\mathbf{G}[t]$ are time-varying from $t = 1$ to $t = 300$, we take the NMSEs of the channel estimates of $\mathbf{H}[300]$ and $\mathbf{G}[300]$ as the performance measures. In Fig. 5(a) and (b), we show the NMSE performance for the Doppler rate $f_D T_s = 10^{-4}$, and for $K = 3$ and $K = 11$, respectively. By comparing Fig. 5(b) with Fig. 2(b), it can be observed that the LR's NMSE performance degrades and exhibits an error floor since the added AN may leak into LR's channel due to the time-varying channels. To further see how the time-varying channel can affect the proposed DCE scheme, we show in Fig. 5(c) the NMSE performance with respect to the

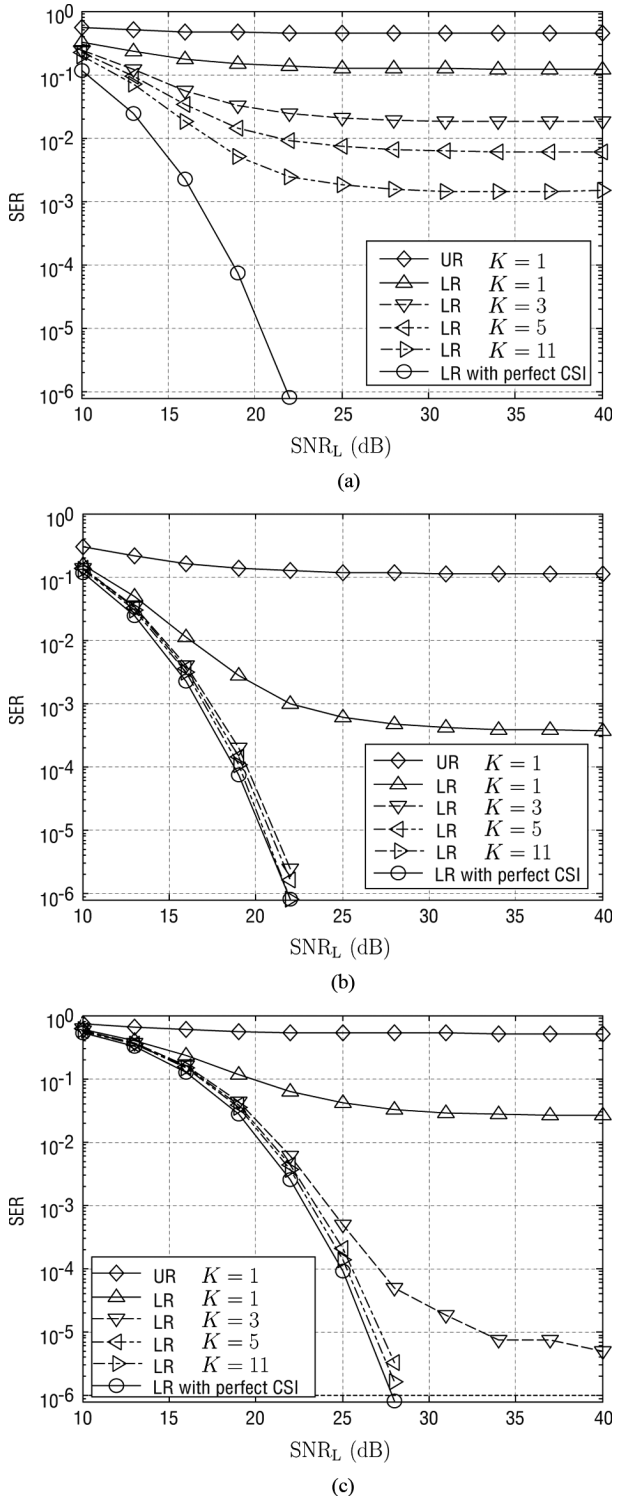


Fig. 3. Symbol error rates of the LR and the UR in an OSTBC system with $N_t = 4$, $N_L = 2$, $N_U = 2$ and $\mathrm{SNR_L} = \mathrm{SNR_U}$. The channel estimates obtained by the proposed DCE scheme were used for OSTBC decoding: (a) 64-QAM OSTBC, $\gamma = 0.1$, (b) 64-QAM OSTBC, $\gamma = 0.03$, and (c) 256-QAM OSTBC, $\gamma = 0.03$.

Doppler rate $f_D T_s$ for $\gamma = 0.03$ and $\mathrm{SNR_L} = \mathrm{SNR_U} = 25$ dB. It can be observed from this figure that not only the NMSE performance but also the discrimination between the LR and the UR can deteriorate with increased Doppler rate $f_D T_s$. Nevertheless, one can see from Fig. 5(c) that successful performance

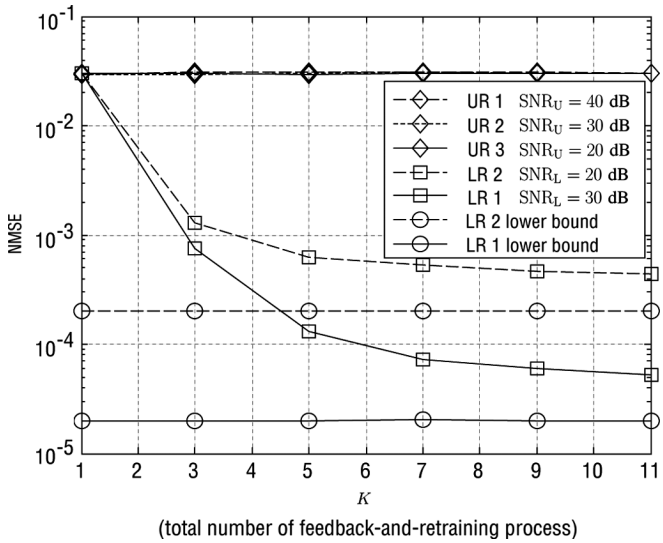


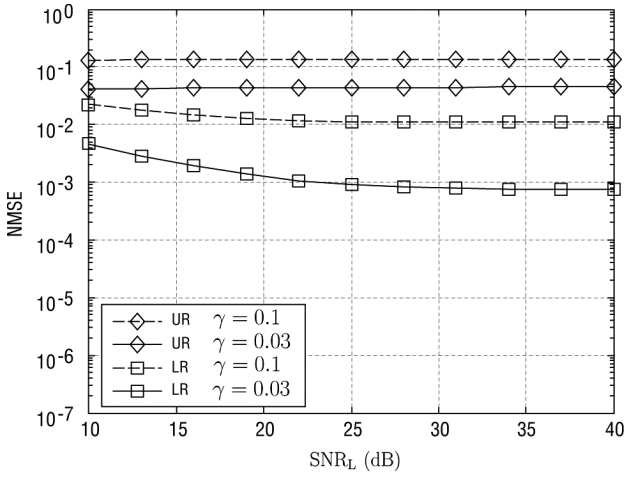
Fig. 4. Simulation results of NMSE performances of the proposed DCE scheme for $N_t = 6$, $N_L = 2$, $N_U = 2$ and $\gamma = 0.03$, with 2 LRs and 3 URs.

discrimination between users can be achieved for Doppler rate $f_D T_s \leq 10^{-4}$, which is a typical Doppler rate value for existing wideband wireless applications. For example, in WiMAX systems [30], the Doppler rate is around 8.33×10^{-5} when the mobile user is moving with a speed of 100 km/h.

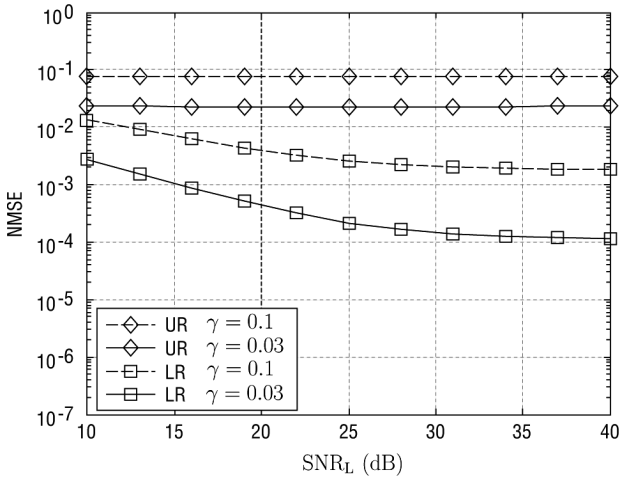
VII. Conclusions and Future Works

In this paper, we have presented a training signal design and a DCE scheme for discriminating between the LMMSE channel estimation performances of an LR and an UR in wireless MIMO communications. We have shown that, with the channel estimates fed back from LR, the performance discrimination between the LR and the UR can be effectively achieved by utilizing the AN together with multiple stages of feedback and retraining. To optimally allocate the training signal powers and AN powers over multiple training stages, we have proposed a design criterion [see (24) and (42)] which minimizes the LR's NMSE subject to a lower limit constraint on the UR's NMSE and an average power constraint at the transmitter. We have shown how the design problem in (24) can be solved by a simple line search method for the case with only one feedback-and-retraining stage (i.e., $K = 1$). When multiple feedback-and-retraining stages $(K > 1)$ are performed, we have shown how an approximate solution to problem (42) can be efficiently obtained by the condensation method in GP. The presented simulation results have shown that the proposed DCE scheme is effective even with three stages of feedback and retraining $(K = 3)$.

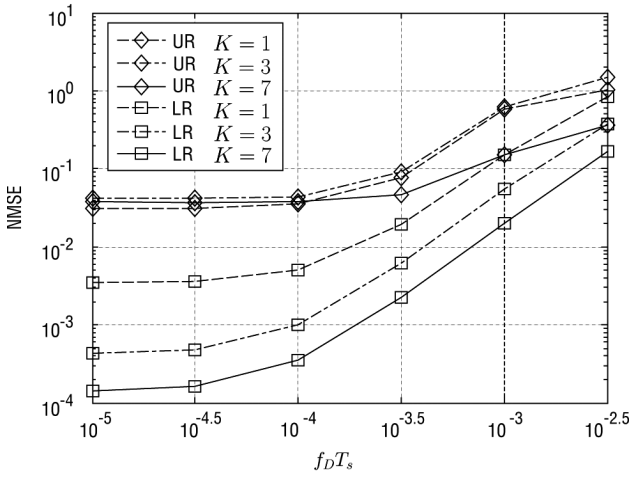
While we have seen the effectiveness of the proposed DCE scheme, there exist several practical issues that must be considered before the proposed scheme can be actually deployed. Specifically, since the proposed DCE scheme involves channel estimate feedback, which in general is not perfect due to limited feedback channel bandwidth [31], the impact of imperfect channel estimate on the proposed scheme must be taken into account. Second, since the feedback-and-retraining process requires the LR to perform channel estimation multiple times and longer training time due to channel estimate feedback, it is



(a)



(b)



(c)

Fig. 5. Simulation results of NMSE performances of the proposed DCE scheme under time-varying channels, for $N_t = 4$, $N_L = N_U = 2$ and $\mathrm{SNR_L} = \mathrm{SNR_U}$: (a) $K = 3$, $f_D T_s = 10^{-4}$, (b) $K = 11$, $f_D T_s = 10^{-4}$, and (c) $\gamma = 0.03$, $\mathrm{SNR_L} = \mathrm{SNR_U} = 25$ dB.

important to devise improved schemes to reduce the complexity burden at the LR and the required number of feedback-and-retraining stages. For example, the low-complexity recursive least squares algorithm [32] may be applied to (36) to implement LR's LMMSE channel estimation in a recursive, stage-by-stage manner. Further research efforts to improve the discrimination performance of the proposed DCE scheme in time-varying channels are also needed. These practical issues certainly bring some interesting and challenging research directions in the future.

The presented DCE scheme may also suggest relevant future research directions in information theory. For example, one may consider analyzing the achievable perfect secrecy rate under imperfect CSI at the LR and the UR. This is in sharp contrast to most existing works where both receivers are assumed to have perfect CSI [6]–[9]. It is anticipated that a higher perfect secrecy rate can be achieved if the UR does not have an accurate channel estimate.

APPENDIX A
PROOF OF PROPOSITION 1

Here we prove Proposition 1. We see from (31b) that a feasible a must satisfy $a \leq \tilde{\gamma}$. Suppose that a feasible $a \leq \tilde{\gamma}$ is given to problem (31). We can first find the optimal values of b and c as a function of a by solving the following optimization problem:

$$\max_{b,c\geq 0} \quad a + \frac{(N_t\sigma_w^2 + \sigma_H^2 a)b}{N_t\sigma_w^2 + \sigma_H^2 a + N_t\sigma_H^2 c} \tag{A1a}$$

$$\text{s.t.} \quad \frac{\sigma_v^2 b}{\sigma_G^2 c + \sigma_v^2} \leq \tilde{\gamma} - a, \tag{A1b}$$

$$T_1 c + b \leq P_{\rm ave}(T_0 + T_1) - a. \tag{A1c}$$

Let $\{b^\star(a), c^\star(a)\}$ be the optimal solution of problem (A1). One can inspect that constraint(A1b) must be active when the optimal objective value is achieved; otherwise, one can always obtain a larger objective value by decreasing $c^\star(a)$. If constraint (A1b) is not active even when $c^\star(a) = 0$, then one can instead increase $b^\star(a)$ to obtain a larger objective value. Since we only consider the interesting case where $\tilde{\gamma} \leq P_{\rm ave}(T_0 + T_1)$ [see (30)], constraint (A1b) must be active with $b^\star(a) = \tilde{\gamma} - a$ if $c^\star(a) = 0$. Hence, we have that

$$b^\star(a) = \left(\frac{\sigma_G^2}{\sigma_v^2}\, c^\star(a) + 1\right)(\tilde{\gamma} - a). \tag{A2}$$

By substituting (A2) into(A1), problem (A1a) reduces to

$$\max_{c>0}\ a + \left(\frac{\left(\frac{\sigma_G^2}{\sigma_v^2}\right)c + 1}{N_t\sigma_H^2 c + N_t\sigma_w^2 + \sigma_H^2 a}\right)(\tilde{\gamma} - a)(N_t\sigma_w^2 + \sigma_H^2 a) \tag{A3a}$$

$$\text{s.t.} \ \left(T_1 + \sigma_G^2\frac{\tilde{\gamma} - a}{\sigma_v^2}\right)c \leq P_{\rm ave}(T_0 + T_1) - \tilde{\gamma}. \tag{A3b}$$

One can show that the objective function in (A3a) is monotonically decreasing with respect to c for

$$a < \eta \triangleq N_t\left(\frac{\sigma_v^2}{\sigma_G^2} - \frac{\sigma_w^2}{\sigma_H^2}\right). \tag{A4}$$

In this case, the optimum value of $c^\star(a)$ is equal to 0, and the associated objective value is given by $\tilde{\gamma}$.

On the other hand, if $a \geq \eta$ the objective function in (A3a) is monotonically non-decreasing with respect to c, and thus the optimum value of $c^\star(a)$ of problem (A3) will activate the constraint in(A3b) leading to

$$c^\star(a) = \frac{P_{\text{ave}}(T_0+T_1) - \tilde{\gamma}}{T_1 + \sigma_G^2\left(\frac{\tilde{\gamma}-a}{\sigma_v^2}\right)}. \tag{A5}$$

Substituting (A5) into(A3), we have that for $a \geq \eta$ the optimum objective value of(A3) is given by

$$a+(\tilde{\gamma}-a)\left(\frac{\left(\frac{\sigma_G^2}{\sigma_v^2}\right)c^\star+1}{\left(\frac{N_t\sigma_H^2}{N_t\sigma_w^2+\sigma_H^2 a}\right)c^\star+1}\right) \geq a+(\tilde{\gamma}-a) = \tilde{\gamma} \tag{A6}$$

which is no less than $\tilde{\gamma}$. From the above analysis, we can conclude with the following two results:

Case of $\eta \leq \tilde{\gamma}$: As seen by (A6), for $\max\{\eta, 0\} \leq a \leq \tilde{\gamma}$ the associated objective value is no less than that for $a < \eta$. Therefore, we can without of generality have the value of $a^\star$ of problem (31) lie in the interval $[\max\{\eta, 0\}, \tilde{\gamma}]$. Since the corresponding values of b and c are respectively given by (A2) and (A5), we can obtain the value of $a^\star$ by solving the one-dimensional problem (33).

Case of $\eta > \tilde{\gamma}$: Since $a \leq \tilde{\gamma}$ we can only have $a < \eta$ and thus $c^\star(a) = 0$ and $b^\star(a) = \tilde{\gamma} - a$. Therefore, the optimum solution of problem (31) is given by $c^\star = 0$ and any $a^\star, b^\star \geq 0$ satisfying $a^\star + b^\star = \tilde{\gamma}$. Proposition 1 is thus proved. ∎

Appendix B
Condensation Method for Problem (42)

We show here how problem (42) can be handled efficiently by the monomial approximation and condensation method. First, one can explicitly write problem (42) as

$$\min_{\substack{P_0,P_k,\sigma_{a,k}^2\geq 0,\\ k=1,\ldots,K}} \quad \text{NMSE}_L^{(K)} \tag{A7a}$$

$$\text{s.t.} \quad \frac{1}{\text{NMSE}_L^{(k)}} = \frac{1}{\text{NMSE}_L^{(k-1)}} + \frac{\frac{P_kT_k}{N_t}}{\text{NMSE}_L^{(k-1)}\cdot(N_t-N_L)\sigma_{a,k}^2+\sigma_w^2}, \quad k=1,\ldots,K, \tag{A7b}$$

$$\frac{1}{\text{NMSE}_L^{(0)}} = \frac{1}{\sigma_H^2} + \frac{P_0T_0}{N_t\sigma_w^2}, \tag{A7c}$$

$$\frac{P_0T_0}{N_t\sigma_v^2} + \sum_{k=1}^{K}\frac{\frac{P_kT_k}{N_t}}{(N_t-N_L)\sigma_{a,k}^2\sigma_G^2+\sigma_v^2} \leq \frac{1}{\gamma} - \frac{1}{\sigma_G^2}, \tag{A7d}$$

$$P_0T_0 + \sum_{k=1}^{K}\left(P_kT_k + \sigma_{a,k}^2(N_t-N_L)T_k\right) \leq P_{\text{ave}}\bar{T}, \tag{A7e}$$

where constraints (A7b) and (A7c) are due to (40) and (8), and (A7d) is due to (41) and (42b), respectively. The key idea of the reformulation is to introduce the auxiliary variables

$$t_k = \frac{1}{\text{NMSE}_L^{(k)}} \geq 0, \quad k=0,\ldots,K \tag{A8}$$

and re-express(A7) as

$$\min_{\substack{P_0,P_k,\sigma_{a,k}^2,t_0,t_k\geq 0,\\ k=1,\ldots,K}} \quad t_K^{-1} \tag{A9a}$$

$$\text{s.t.} \quad t_k \leq t_{k-1} + \frac{\frac{t_{k-1}P_kT_k}{N_t}}{(N_t-N_L)\sigma_{a,k}^2 + t_{k-1}\sigma_w^2}, \quad k=1,\ldots,K, \tag{A9b}$$

$$t_0 \leq \frac{1}{\sigma_H^2} + \frac{P_0T_0}{N_t\sigma_w^2}, \tag{A9c}$$

$$\frac{P_0T_0}{N_t\sigma_v^2} + \sum_{k=1}^{K}\frac{\frac{P_kT_k}{N_t}}{(N_t-N_L)\sigma_{a,k}^2\sigma_G^2+\sigma_v^2} \leq \frac{1}{\gamma} - \frac{1}{\sigma_G^2}, \tag{A9d}$$

$$P_0T_0 + \sum_{k=1}^{K}\left(P_kT_k + \sigma_{a,k}^2(N_t-N_L)T_k\right) \leq P_{\text{ave}}\bar{T}. \tag{A9e}$$

Note that in (A9b) and (A9c) we have replaced the equalities with inequalities since one can show that the two inequalities must be active at the global optimum; otherwise one can always achieve a smaller objective value. To further simplify the expression of (A9), we define

$$a_0 = \left(\frac{1}{\sigma_H^2} + \frac{P_0T_0}{N_t\sigma_w^2}\right)^{-1} \geq 0, \tag{A10a}$$

$$a_k = \frac{P_kT_k}{N_t} \geq 0, \tag{A10b}$$

$$b_k = \sigma_G^2\sigma_{a,k}^2(N_t-N_L) + \sigma_v^2 \geq 0, \tag{A10c}$$

for $k = 1,\ldots,K$, for change of variables. Substituting(A10) into (A9) gives rise to

$$\min_{\substack{a_0,a_k,b_k,t_0,t_k\geq 0,\\ k=1,\ldots,K}} \quad t_K^{-1} \tag{A11a}$$

$$\text{s.t.} \quad t_k \leq t_{k-1} + \frac{t_{k-1}a_k\sigma_G^2}{b_k - \sigma_v^2 + t_{k-1}\sigma_w^2\sigma_G^2}, \quad k=1,\ldots,K, \tag{A11b}$$

$$t_0a_0 \leq 1, \tag{A11c}$$

$$\alpha\left(\left(\frac{\sigma_w^2}{\sigma_v^2}\right)a_0^{-1} + \sum_{k=1}^{K}a_kb_k^{-1}\right) \leq 1, \tag{A11d}$$

$$\beta\left(N_t\sigma_w^2a_0^{-1} + \sum_{k=1}^{K}\left(N_ta_k + \left(\frac{T_k}{\sigma_G^2}\right)b_k\right)\right) \leq 1 \tag{A11e}$$

where $\alpha = \left(1/\gamma - 1/\sigma_G^2 + \sigma_w^2/(\sigma_H^2\sigma_v^2)\right)^{-1}$ and $\beta = \left(P_{\text{ave}}\bar{T} + (N_t\sigma_w^2/\sigma_H^2)\sum_{k=1}^{K}\sigma_v^2T_k/\sigma_G^2\right)^{-1}$. Note from problem (A11) that we have reformulated the constraints in

$$\min_{\substack{a_0,a_k,b_k,t_0,t_k\geq 0,\\ k=1,\ldots,K}} \quad t_K^{-1} \tag{A20a}$$

$$\text{s.t.} \quad \frac{b_k t_k + t_k t_{k-1}\sigma_w^2\sigma_G^2 + t_{k-1}\sigma_v^2}{\left(\frac{t_k\sigma_v^2}{\xi_{k,1}}\right)^{\xi_{k,1}}\left(\frac{b_k t_{k-1}}{\xi_{k,2}}\right)^{\xi_{k,2}}\left(\frac{\sigma_w^2\sigma_G^2 t_{k-1}^2}{\xi_{k,3}}\right)^{\xi_{k,3}}\left(\frac{\sigma_G^2 a_k t_{k-1}}{\xi_{k,4}}\right)^{\xi_{k,4}}} \leq 1,\ k=1,\ldots,K, \tag{A20b}$$

$$t_0 a_0 \leq 1, \tag{A20c}$$

$$\alpha\left(\left(\frac{\sigma_w^2}{\sigma_v^2}\right)a_0^{-1} + \sum_{k=1}^{K} a_k b_k^{-1}\right) \leq 1, \tag{A20d}$$

$$\beta\left(N_t\sigma_w^2 a_0^{-1} + \sum_{k=1}^{K}\left(N_t a_k + \left(\frac{T_k}{\sigma_G^2}\right)b_k\right)\right) \leq 1. \tag{A20e}$$

(A11c)–(A11e) such that the optimization variables are all on the left-hand side of the inequalities. Since problem (A11) involves only non-negative optimization variables, we seek to handle it by the GP [20]. A standard GP can be expressed as follows:

$$\min_{x_1,\ldots,x_n\geq 0} \quad f_0(x_1,\ldots,x_n) \tag{A12a}$$

$$\text{s.t.} \quad f_i(x_1,\ldots,x_n) \leq 1,\ i=1,\ldots,N \tag{A12b}$$

$$h_i(x_1,\ldots,x_n) = 1, i=1,\ldots,M \tag{A12c}$$

where $h_i(x_1,\ldots,x_n)$ are monomials taking the form

$$h_i(x_1,\ldots,x_n) = c_i x_1^{\alpha_{i,1}} x_2^{\alpha_{i,2}}\ldots x_n^{\alpha_{i,n}} \tag{A13}$$

in which $c_i \geq 0$ and $\{\alpha_{i,j}\}_{j=1}^n$ are real numbers for all $i = 1,\ldots,N$, and $f_i(x_1,\ldots,x_n)$ are posynomials taking the form

$$f_i(x_1,\ldots,x_n) = \sum_{\ell=1}^{L_i} d_{i,\ell}\, x_1^{\beta_{i,\ell,1}} x_2^{\beta_{i,\ell,2}}\ldots x_n^{\beta_{i,\ell,n}} \tag{A14}$$

in which $d_{i,\ell} \geq 0$ and $\beta_{i,\ell,j}$ are real numbers for all i,ℓ and j. As seen, a posynomials is simply non-negative weighted sum of monomials. Comparing problem (A11) with the standard GP in (A12), one can see that the objective function and all the inequality constraints of problem (A11) are posynomials, except for the constraints in (A11b). In fact, the inequality constraints in (A11b) can only be expressed as ratios of posynomials as

$$\frac{b_k t_k + t_k t_{k-1}\sigma_w^2\sigma_G^2 + t_{k-1}\sigma_v^2}{t_k\sigma_v^2 + b_k t_{k-1} + \sigma_w^2\sigma_G^2 t_{k-1}^2 + \sigma_G^2 a_k t_{k-1}} \leq 1, \quad k=1,\ldots,K \tag{A15}$$

which are known to be difficult to deal with in general. To resolve these hard constraints and to obtain an effective approximate solution of problem (A11), we apply the condensation method and the monomial approximation technique [24]. The idea of the monomial approximation technique is to locally approximate the posynomial in the denominator of the left-hand side of (A15) by a monomial function, and the condensation method is to find an approximate solution of problem (A11) by successively solving the monomial-approximated problem. Specifically, given a feasible point $\{\bar{a}_0,\bar{a}_k,\bar{b}_k,\bar{t}_0,\bar{t}_k,k=1,\ldots,K\}$ of problem (A11), one can show [24] that

$$t_k\sigma_v^2 + b_k t_{k-1} + \sigma_w^2\sigma_G^2 t_{k-1}^2 + \sigma_G^2 a_k t_{k-1} \geq \left(\frac{t_k\sigma_v^2}{\xi_{k,1}}\right)^{\xi_{k,1}}\left(\frac{b_k t_{k-1}}{\xi_{k,2}}\right)^{\xi_{k,2}}\left(\frac{\sigma_w^2\sigma_G^2 t_{k-1}^2}{\xi_{k,3}}\right)^{\xi_{k,3}}\left(\frac{\sigma_G^2 a_k t_{k-1}}{\xi_{k,4}}\right)^{\xi_{k,4}} \tag{A16}$$

by the inequality of arithmetic and geometric means, where

$$\xi_{k,1} = \frac{\bar{t}_k\sigma_v^2}{f(\bar{a}_k,\bar{b}_k,\bar{t}_k,\bar{t}_{k-1})} \tag{A17a}$$

$$\xi_{k,2} = \frac{\bar{b}_k\bar{t}_{k-1}}{f(\bar{a}_k,\bar{b}_k,\bar{t}_k,\bar{t}_{k-1})} \tag{A17b}$$

$$\xi_{k,3} = \frac{\sigma_w^2\sigma_G^2\bar{t}_{k-1}^2}{f(\bar{a}_k,\bar{b}_k,\bar{t}_k,\bar{t}_{k-1})} \tag{A17c}$$

$$\xi_{k,4} = \frac{\sigma_G^2\bar{a}_k\bar{t}_{k-1}}{f(\bar{a}_k,\bar{b}_k,\bar{t}_k,\bar{t}_{k-1})} \tag{A17d}$$

and

$$f(\bar{a}_k,\bar{b}_k,\bar{t}_k,\bar{t}_{k-1}) = \bar{t}_k\sigma_v^2 + \bar{b}_k\bar{t}_{k-1} + \sigma_w^2\sigma_G^2\bar{t}_{k-1}^2 + \sigma_G^2\bar{a}_k\bar{t}_{k-1}. \tag{A18}$$

The right-hand side of (A16) is a monomial function. The monomial approximation technique is to replace (A15) by the following constraint:

$$\frac{b_k t_k + t_k t_{k-1}\sigma_w^2\sigma_G^2 + t_{k-1}\sigma_v^2}{\left(\frac{t_k\sigma_v^2}{\xi_{k,1}}\right)^{\xi_{k,1}}\left(\frac{b_k t_{k-1}}{\xi_{k,2}}\right)^{\xi_{k,2}}\left(\frac{\sigma_w^2\sigma_G^2 t_{k-1}^2}{\xi_{k,3}}\right)^{\xi_{k,3}}\left(\frac{\sigma_G^2 a_k t_{k-1}}{\xi_{k,4}}\right)^{\xi_{k,4}}} \leq 1 \tag{A19}$$

which are posynomials, and guarantee the satisfaction of (A15). Therefore, given a feasible point of $\{\bar{a}_0,\bar{a}_k,\bar{b}_k,\bar{t}_0,\bar{t}_k,k=1,\ldots,K\}$, problem (A11) is locally approximated by the problem [see (A20a)–(A20e) at the top of the page]. Since (A20) now is a standard GP [see (A12)], it can be efficiently solved by general purpose interior point solvers such as CVX [25] in a polynomial-time complexity. Then the condensation method [24] finds an approximate solution of problem (A11) by iteratively solving a sequence of problems (A20) with each initial feasible point $\{\bar{a}_0,\bar{a}_k,\bar{b}_k,\bar{t}_0,\bar{t}_k,k=1,\ldots,K\}$ given by the optimum solution of (A20) in the previous iteration. In Table I, we summarize the condensation method for problem (42).

REFERENCES

[1] R. M. Needham and M. D. Schroeder, "Using encryption for authentication in large networks of computers," *Commun. ACM*, vol. 21, no. 12, pp. 993–999, Dec. 1978.

[2] C. E. Landwehr and D. M. Goldschlag, "Secure issues in networks with internet access," *Proc. IEEE*, vol. 85, pp. 2034–2051, Dec. 1997.

[3] L. Mucchi, L. S. Ronga, and E. D. Re, "A novel approach for physical layer cryptography in wireless networks," *Springer Netherlands Wireless Personal Commun.*, vol. 53, pp. 329–347, Mar. 2010.

[4] L. Mucchi, L. S. Ronga, and E. D. Re, "Design and implementation of physical layer private key setting for wireless networks," in *Proc. IEEE ICC*, Dresden, Germany, Jun. 14–18, 2009, pp. 1–5.

[5] A. D. Wyner, "The wiretap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, 1975.

[6] P. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inf. Theory*, vol. 54, pp. 4687–4698, Oct. 2008.

[7] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas I: The MISOME wiretap channel," *IEEE. Trans. Inf. Theory*, vol. 55, pp. 3088–3104, Jul. 2010.

[8] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas II: The MIMOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, pp. 5515–5532, Nov. 2010.

[9] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," in *Proc. IEEE ISIT*, Toronto, ON, Canada, Jul. 6–11, 2008, pp. 524–528.

[10] X. Li and J. Hwu, "Using antenna array redundancy and channel diversity for secure wireless transmissions," *J. Commun.*, vol. 2, no. 3, pp. 24–32, May 2007.

[11] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," *IEEE. Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2180–2189, Jun. 2008.

[12] L. Dong, Z. Han, A. P. Petropulu, and H. V. Poor, "Amplify-and-forward based cooperation for secure wireless communications," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, pp. 2613–2616, Apr. 2009.

[13] L. Dong, Z. Han, A. P. Petropulu, and H. V. Poor, "Secure wireless communications via cooperation," in *Proc. Allerton Conf. Communication, Control, Computing*, Monticello, IL, Sep. 23–26, 2008, pp. 1132–1138.

[14] A. Mukherjee and A. L. Swindlehurst, "Fixed-rate power allocation strategies for enhanced secrecy in MIMO wiretap channels," in *Proc. IEEE SPAWC*, Perugia, Italy, Jun. 21–24, 2009, pp. 344–348.

[15] A. L. Swindlehurst, "Fixed SINR solutions for the MIMO wiretap channel," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 19–24, 2009, pp. 2437–2440.

[16] A. O. Hero, "Secure space-time communication," *IEEE. Trans. Inf. Theory*, vol. 49, pp. 3235–3249, Dec. 2003.

[17] T. F. Wong and B. Park, "Training sequence optimization in MIMO systems with colored interference," *IEEE Trans. Commun.*, vol. 52, pp. 1939–1947, Nov. 2004.

[18] J. H. Kotecha and A. M. Sayeed, "Training signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 546–557, Feb. 2004.

[19] Y.-L. Liang, Y.-S. Wang, T.-H. Chang, Y.-W. Hong, and C.-Y. Chi, "On the impact of quantized channel feedback in guaranteeing secrecy using artificial noise," in *Proc. IEEE ISIT*, Seoul, Korea, Jul. 3, 2009, pp. 2351–2355.

[20] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optim. Eng.*, vol. 8, pp. 67–127, Apr. 2007.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[22] T.-H Chang, W.-Y Hong, and C.-Y Chi, "Training signal design for discriminatory channel estimation," in *Proc. IEEE GLOBECOM*, Honolulu, HI, Dec. 4, 2009, pp. 1–6.

[23] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall Int., 1993.

[24] M. Chiang, C.-W. Tamd, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," in *Resource Allocation in Next Generation Wireless Networks*, W. Li and Y. Pan, Eds. Commack, NY: Nova, 2005, ch. 13.

[25] M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming Jun. 2009 [Online]. Available: http://stanford.edu/~boyd/cvx

[26] *Standard for Local and Metropolitan Area Networks- Part 11: Wireless LAN Media Access Control and Physical Layer Specifications: High-Speed Physical Layer in the 5 GHz Band*, IEEE Std 802.11a-1999, Sep. 1999.

[27] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[28] J. Via and I. Santamaria, "On the blind identifiability of orthogonal space-time block codes from second order statistics," *IEEE Trans. Inf. Theory*, vol. 54, pp. 709–722, Feb. 2008.

[29] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

[30] *IEEE Standard for Local and Metropolitan Area Networks- Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std 802.16-2004, Jun. 2004.

[31] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE. Trans. Inf. Theory*, vol. 52, pp. 5045–5060, Nov. 2006.

[32] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall Int., 2001.

Tsung-Hui Chang (S'07–M'08) received the B.S. degree in electrical engineering and the Ph.D. degree in communications engineering from the National Tsing Hua University (NTHU), Hsinchu, Taiwan, R.O.C., in 2003 and 2008, respectively.

During September 2006 and February 2008, he was an exchange Ph.D. student of the University of Minnesota, Minneapolis. Currently, he is a Postdoctoral Research Fellow with the Institute of Communications Engineering, NTHU. His research interests are widely in wireless communications, digital signal processing and convex optimization and its applications.

Wei-Cheng Chiang received the B.S. degree from the Department of Communications Engineering from the National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 2008 and the M.S. degree in communications engineering from the National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 2010.

His research interests are in wireless communications and signal processing.

Y.-W. Peter Hong (S'01–M'05)) received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, R.O.C., in 1999 and the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY, in 2005.

He joined the Institute of Communications Engineering and the Department of Electrical Engineering at National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in fall 2005, where he is now an Associate Professor. He was also a visiting scholar at the University of Southern California from June to August 2008. His research interests include cooperative communications, distributed signal processing for sensor networks, and PHY-MAC cross-layer designs for wireless networks.

Dr. Hong received the Best Paper Award for Young Authors from the IEEE IT/COM Society Taipei/Tainan Chapter in 2005 and the Best Paper Award among unclassified papers in MILCOM 2005. He also received the Junior Faculty Research Award and the Outstanding Teaching Award from the College of EECS at the National Tsing Hua University in 2009 and 2010, respectively. He is a Co-Editor (along with A. Swami, Q. Zhao, and L. Tong) of the book entitled *Wireless Sensor Networks: Signal Processing and Communications Perspectives* (Wiley, 2007) and is a coauthor (along with W.-J. Huang and C.-C. Jay Kuo) of the book entitled *Cooperative Communications and Networking: Technologies and System Design*. He has served as the Publication Chair and the TPC Track Co-Chair of the Vehicular Technology Conference (VTC) 2010-Spring for the track on Cognitive Radio and Cooperative Communications and has also served as the TPC Co-Chair of WASN 2009. He is currently serving as the Publicity Co-Chair of ISITA/ISSSTA 2010. He is also a Guest Editor of the EURASIP Special Issue on Cooperative MIMO Multicell Networks and IJSNET Special Issue on Advances in Theory and Applications of Wireless, Ad Hoc, and Sensor Networks.

Chong-Yung Chi (S'83–M'83–SM'89) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1983.

From 1983 to 1988, he was with the Jet Propulsion Laboratory, Pasadena, CA. He has been a Professor with the Department of Electrical Engineering since 1989 and the Institute of Communications Engineering (ICE) since 1999 (also the Chairman of ICE from 2002 to 2005), National Tsing Hua University, Hsinchu, Taiwan, R.O.C. He has published more than 160 technical papers, including more than 50 journal papers (mostly in IEEE TRANSACTIONS ON SIGNAL PROCESSING), two book chapters, and more than 100 peer-reviewed conference papers, as well as a graduate-level textbook *Blind Equalization and System Identification* (Springer-Verlag, 2006). His current research interests include signal processing for wireless communications, convex analysis and optimization for blind source separation, and biomedical and hyperspectral image analysis.

Dr. Chi has been a Technical Program Committee member for many IEEE sponsored and co-sponsored workshops, symposiums, and conferences on signal processing and wireless communications, including Co-Organizer and General Co-Chairman of the 2001 IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC) and Co-Chair of the Signal Processing for Communications (SPC) Symposium, ChinaCOM 2008, and Lead Co-Chair of the SPC Symposium, ChinaCOM 2009. He is currently serving as Track Chair for MIMO, Signal Processing, and Smart in Antennas, 2011 IEEE Radio and Wireless Symposium in Radio and Wireless Week (RWW) 2011. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2001 to 2006, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II from 2006 to 2007, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I from 2008 to 2009, and the IEEE SIGNAL PROCESSING LETTERS from June 2006 to May 2010. He was also a Member of the Editorial Board of the *EURASIP Signal Processing Journal* from 2005 to 2008 and an Editor (2003–2005) as well as a Guest Editor (2006) of the *EURASIP Journal on Applied Signal Processing*. Currently, he is a member of the IEEE Signal Processing Committee on *Signal Processing Theory and Methods*.