Title
Maximum a Posteriori Based Channel Estimation Strategy for Two-Way Relaying Channels

Journal
The IEEE Transactions on Wireless Communications, 13(1)

Authors
Xie, Xinqian
Peng, Mugen
Zhao, Bin
et al.

Publication Date
2013-12-16

Peer reviewed

Maximum a Posteriori Based Channel Estimation Strategy for Two-Way Relaying Channels

Xinqian Xie, Mugen Peng, *Senior Member, IEEE*, Bin Zhao, Wenbo Wang, *Member, IEEE*, and Yingbo Hua, *Fellow, IEEE*

***Abstract*—Wireless network coding can significantly improve the spectrum efficiency for relaying transmission when receivers can acquire accurate channel state information (CSI). In this paper, the channel estimation problem for two-way relaying channels is considered where two sources exchange information through an amplify-and-forward relay employing analog network coding protocol. By taking advantage of the apriori information of wireless channels to further improve channel estimation accuracy, the maximum a posteriori (MAP) based estimation schemes are developed to estimate the composite source-source channel coefficients and the amplitude of individual source-relay channels with apriori knowledge of channel distribution information (CDI). Variations of MAP estimation algorithms are also developed for systems under practical constraints where channel CDI needs to be estimated. In particular, scale MAP estimator as well as a long term estimation algorithm is developed to effectively control the negative impact of CDI estimation error on MAP estimation performance. The simulation results show that the MAP based estimation strategies consistently outperform maximum likelihood estimation methods in the measure of mean square error, thus establishes the advantage of presented MAP based schemes.**

***Index Terms*—Channel estimation, two-way relay, analog network coding, maximum a posteriori.**

I. INTRODUCTION

RECENTLY, wireless relaying has attracted a lot of research interests due to its capability in enhancing long-range communications and enlarging the coverage of cellular networks. Half-duplex relay operating in time division duplex mode would reduce system spectrum efficiency. Cooperative network coding protocol allows signals from multiple sources to be mixed at the relay node before being forwarded to their destination simultaneously, thus has the potential to bring in significant improvement to system spectrum efficiency [1]. [2] demonstrates a 2-fold increase in system throughput when network coding is deployed in two-way relaying channels (TWRC). Accurate channel state information (CSI) is required at each source node in order to recover signal of interest by subtracting self-interference signal. Most research work on wireless network coding assume perfect CSI or ideal channel estimation. Since network coding is very sensitive to channel estimation error [3] and practical training based estimation techniques consume both bandwidth and energy overhead [4], optimal training sequence design has been developed in [5] to minimize estimation error and to reduce training overhead for both maximum likelihood estimator and maximum signal-to-noise ratio estimator. Likewise, with the assistance of superimposed training sequence [6], individual CSI could be estimated [7] to support various diversity techniques including sub-carrier pairing [8], to achieve further improvement of system performance. In [9], the optimal relay selection scheme was developed to achieve full diversity for multiple relays scenarios, and multiple-input multiple output (MIMO) technology has shown its advantage to further improve the throughput for network coded relay networks [10]. It should be noticed that most channel estimators are developed for systems without the a priori knowledge of channel fading statistics. When channel distribution information are available, Bayesian estimation method could be applied to reduce estimation residue [11]. For example, the maximum a posteriori (MAP) estimator can be employed to minimize the Bayes risk for a hit-or-miss cost (HMC) function [16].

In this paper, we use the Bayesian approach to solve channel estimation problem in TWRC. Particularly, the main contributions of this work are listed as follows:

- The MAP based channel estimation algorithms are developed under the assumption of perfect CDI at each source to estimate both the composite source-source channels and individual source-relay channel amplitudes. We also design the MAP based estimation schemes for systems under practical constraint where noise variance needs to be estimated.
- Variations of MAP estimation algorithms have been developed for systems with practical constraints. In particular, iterative least square-MAP algorithm has been developed for system with unknown noise variance. An improved channel estimation strategy is provided where instantaneous channel estimates are utilized to calculate the CDI and it back sustains the MAP based instantaneous channel estimation. A scale MAP mechanism has been incorporated into the estimation strategy to control the negative impact of estimation error.

The rest of this paper is organized as follows. Section II

Manuscript received May 13, 2013; revised July 31, 2013; accepted November 3, 2013. The associate editor coordinating the review of this paper and approving it for publication was Y. Sanada.

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61222103, No.61072058), the National Basic Research Program of China (973 Program) (Grant No. 2013CB336600), the State Major Science and Technology Special Projects (No. 2012ZX03001031, No. 2012ZX03001037), the Beijing Natural Science Foundation (Grant No. 4131003), and the Specialized Research Fund (Grant No. 20120005140002).

X. Xie, M. Peng, and W. Wang are with the Key Laboratory of Universal Wireless Communications for Ministry of Education, Beijing University of Posts and Telecommunications, China (e-mail: xxmbupt@gmail.com, {pmg, wbwang}@bupt.edu.cn). The corresponding author is M. Peng.

B. Zhao is with Cambridge Silicon Radio, UK (e-mail: zb34111@yahoo.com).

Y. Hua is with the Department of Electrical Engineering, College of Engineering University of California, Riverside, CA, USA (e-mail: yhua@ee.ucr.edu).

Digital Object Identifier 10.1109/TWC.2013.120313.130864

describes the transmission scheme of two way analog network coding protocol. In section III, the MAP based estimation schemes are presented including both composite and individual channel estimations. Then in section IV, the scale MAP as well as the improved long term estimation strategy is presented. The simulation results are shown in section V, and followed by the conclusions in section VI.

Notations: The transpose, Hermitian and inverse of matrices are denoted by $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$, respectively. $\|\cdot\|$ denotes the two-norm of vectors. $|\cdot|$, $\angle(\cdot)$ and $\Re\{\cdot\}$ denote the magnitude, phase and the real part of the complex arguments, respectively. $\mathfrak{E}\{\cdot\}$ denotes the expectation of random variables.

II. System Model

Consider a two-way relaying channel where two sources $\mathbb{S}_i\,(i=1,2)$ exchange information via a relay $\mathbb{R}$ operating in half-duplex mode as shown in Fig.1. All the three nodes are equipped with a single antenna. The channel coefficient between $\mathbb{S}_i$ and $\mathbb{R}$ is denoted by h_i. P_{Si} denotes transmitting power of $\mathbb{S}_i$ and P_R denotes transmitting power of $\mathbb{R}$. The data symbol and training sequence sent from $\mathbb{S}_i$ are denoted by s_i and $\mathbf{t}_i$, respectively. Analog network coding (ANC) protocol for TWRC has two phases. In phase 1, $\mathbb{S}_1$ and $\mathbb{S}_2$ simultaneously transmit data s_1 and s_2 to $\mathbb{R}$, respectively. In phase 2, $\mathbb{R}$ forwards the received signal to both sources after amplifying it by a factor α.

Assumption:

- The channels are assumed to be quasi-static flat fading, i.e., they stay identical within one transmission block but vary from block to block.
- The fading coefficient h_i is assumed to be circularly symmetric complex Gaussian random variable with zero mean and υ_i variance, and h_1 and h_2 are assumed to be independent with each other.
- Both sources and relay has full knowledge of training sequences as well as the amplifying factor α. When channel variance υ_i and noise variance at relay denoted by σ_R^2 are available, α is given by $\alpha = \sqrt{\frac{P_R}{\upsilon_1 P_{S1}+\upsilon_2 P_{S2}+\sigma_R^2}}$. Otherwise, α is set to be a const which is independent of υ_i and σ_R^2.

Without loss of generality, we only consider the signal detection at $\mathbb{S}_1$. The received signal at $\mathbb{S}_1$ can be given as

$$y_S = \alpha h_1^2 s_1 + \alpha h_1 h_2 s_2 + \alpha h_1 n_R + n_S,$$

where $\mathbf{n}_R$ represents the additive white Gaussian noise (AWGN) with zero mean and σ_R^2 variance at $\mathbb{R}$, and n_S stands for the AWGN at $\mathbb{S}_1$. With full knowledge of s_1 and perfect channel side information (CSI) at $\mathbb{S}_1$, the self-interference term can be completed removed and symbol s_2 can be effectively detected. Therefore, the problem lies in how to estimate the instantaneous composite channel coefficients $h_a \triangleq h_1^2$ and $h_b \triangleq h_1 h_2$.

To tackle the problem, the famous training based estimation technique is utilized where $\mathbb{S}_i$ transmits the N length training sequence $\mathbf{t}_i$ to the relay simultaneously. Thus $\mathbb{R}$ can observe

$$\mathbf{x}_R = h_1\mathbf{t}_1 + h_2\mathbf{t}_2 + \mathbf{w}_R, \quad (1)$$

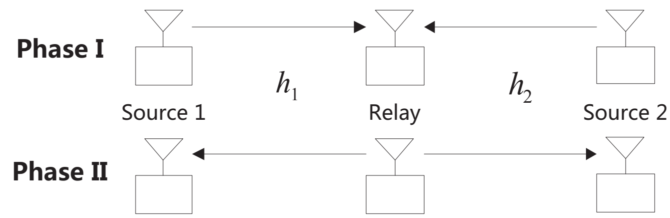

Fig. 1. The bidirectional communication topology

where $\mathbf{w}_R$ represents the AWGN vector of $N\times 1$ dimension at $\mathbb{R}$ with each entry of zero mean and σ_R^2 variance. For the j-th entry of $\mathbf{t}_i$, the peak power constraint is holden as $\|t_{i,j}\|^2 = P_{Si}, \forall j \in [1,N]$. Then $\mathbb{R}$ forwards the superimposed sequence to both sources after scaling $\mathbf{x}_R$ by α. Due to the symmetry of the two sources, we only focus on the channel estimation issue at $\mathbb{S}_1$ and that of $\mathbb{S}_2$ can be settled similarly. After relay forwarding $\alpha\mathbf{x}_R$, $\mathbb{S}_1$ receives

$$\mathbf{x}_S = \alpha h_1^2\mathbf{t}_1 + \alpha h_1 h_2\mathbf{t}_2 + \alpha h_1\mathbf{w}_R + \mathbf{w}_S, \quad (2)$$

where $\mathbf{w}_S$ is the AWGN vector at $\mathbb{S}_1$ with each entry of zero mean and σ_S^2 variance. Hence, total $2N$ symbol periods are required to accomplish the training process. Since each source has priori knowledge of $\mathbf{t}_i$, effective estimation methods can be performed to obtain the required channel coefficients in terms of the observation $\mathbf{x}_S$.

III. Channel Estimation With MAP Based Estimator

In this section, we present the MAP based estimation method to obtain the required channel coefficients for TWRC. The channel variance υ_i and noise variance σ_R^2 and σ_S^2 are assumed to be known at both sources and relay.

A. MAP-based Composite Channel Estimation

The task of this part is to estimate the composite channel coefficients h_a and h_b. Since the under estimated parameters are complex value, both amplitude and phase are needed to be obtained. Let $a \triangleq |h_a|$, $b \triangleq |h_b|$ and the phase of h_a and h_b are denoted by $\theta_a \in [0,2\pi)$ and $\theta_b \in [0,2\pi)$, respectively. The under estimated parameters can be written in a vector form by $\mathbf{\Theta} \triangleq [a,b,\theta_a,\theta_b]^T$.

By definition, the MAP estimation is known to be given by

$$\hat{\mathbf{\Theta}} = \arg\max_{\Theta}\{p(\mathbf{x}_S|\mathbf{\Theta})\,p(\mathbf{\Theta})\}, \quad (3)$$

where $p(\mathbf{x}_S|\mathbf{\Theta})$ represents the conditional probability density function (p.d.f) and $p(\mathbf{\Theta})$ is the corresponding joint apriori p.d.f. In this case, $p(\mathbf{x}_S|\mathbf{\Theta})$ can be easily given by

$$p(\mathbf{x}_S|\mathbf{\Theta}) = \frac{1}{\pi^N(\alpha^2 a\sigma_R^2+\sigma_S^2)^N} \times \exp\left\{-\frac{\|\mathbf{x}_S - \alpha a e^{j\theta_a}\mathbf{t}_1 - \alpha b e^{j\theta_b}\mathbf{t}_2\|^2}{(\alpha^2 a\sigma_R^2+\sigma_S^2)}\right\}. \quad (4)$$

For simplicity, σ_R^2 and σ_S^2 are supposed to be equal in following parts of this work as $\sigma_R^2=\sigma_S^2=\sigma_n^2$, and the two sources are with equal transmitting power as $P_{S1}=P_{S2}=P_S$. After straightforward derivation shown in the Appendix A, the joint p.d.f of $\boldsymbol{\Theta} = [a, b, \theta_a, \theta_b]^T$ is obtained as

$$p(\boldsymbol{\Theta}) = \frac{b}{2\pi^2 a \upsilon_1 \upsilon_2} \exp\left\{ -\left(\frac{a}{\upsilon_1} + \frac{b^2}{\upsilon_2 a}\right)\right\}, \quad a > 0. \quad (5)$$

After some manipulation, (3) can be transformed as

$$\hat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} \left\{ -\frac{\|\mathbf{x}_S - \alpha a e^{j\theta_a}\mathbf{t}_1 - \alpha b e^{j\theta_b}\mathbf{t}_2\|^2}{\sigma_n^2(\alpha^2 a + 1)} - N\log\left(\alpha^2 a + 1\right) - \frac{a}{\upsilon_1} - \frac{b^2}{\upsilon_2 a} + \log\left(\frac{b}{a}\right)\right\}, \quad (6)$$

where the terms inside the max operator is denoted by $\mathcal{L}(\mathbf{x}_S|\boldsymbol{\Theta})$ which is named as the updated log-likelihood function (l.l.f). Hence, it is effective to obtain the estimates utilizing proper 4-dimensional searching methods which would endure high computational complexity. In order to reduce the computational complexity, we adopt following process to solve the MAP estimation problem.

Note that θ_b is only related to the first term of $\mathcal{L}(\mathbf{x}_S|\boldsymbol{\Theta})$, thus $\hat{\theta}_b$ can be obtained by

$$\hat{\theta}_b = \arg\min_{\theta_b} \|\mathbf{x}_S - \alpha a e^{j\theta_a}\mathbf{t}_1 - \alpha b e^{j\theta_b}\mathbf{t}_2\|^2 \quad (7)$$

with given a, b and θ_a. Then we have

$$\hat{\theta}_b = \angle\left\{\frac{\mathbf{t}_2^H\left(\mathbf{x}_S - \alpha a e^{j\theta_a}\mathbf{t}_1\right)}{\alpha\|\mathbf{t}_2\|^2}\right\}. \quad (8)$$

The equation above, when injected into $\mathcal{L}(\mathbf{x}_S|\boldsymbol{\Theta})$, gives

$$\mathcal{L}_1(\mathbf{x}_S|a, b, \theta_a) = -\frac{\|\mathbf{z}_1\|^2 + \alpha^2 b^2\|\mathbf{t}_2\|^2 - 2\alpha\|\mathbf{z}_1^H\mathbf{t}_2\|b}{\sigma_n^2(\alpha^2 a + 1)} - N\log\left(\alpha^2 a + 1\right) - \frac{a}{\upsilon_1} - \frac{b^2}{\upsilon_2 a} + \log\left(\frac{b}{a}\right), \quad (9)$$

with $\mathbf{z}_1 \triangleq \mathbf{x}_S - \alpha a e^{j\theta_a}\mathbf{t}_1$. With given a and θ_a, $\hat{b}$ is obtained by

$$\hat{b} = \arg\max_b \mathcal{L}_1(\mathbf{x}_S|a, b, \theta_a) \quad (10)$$

After straightforward calculation, the first order derivative of (9) is derived as

$$\dot{\mathcal{L}}_1(b) = \frac{\partial\mathcal{L}_1(\mathbf{x}_S|a, b, \theta_a)}{\partial b} = -B_1 b + B_2 + \frac{1}{b}, \quad (11)$$

with

$$B_1 = \frac{2\alpha^2\|\mathbf{t}_2\|^2}{\sigma_n^2(\alpha^2 a + 1)} + \frac{2}{a\upsilon_2}, \quad B_2 = \frac{2\alpha\|\mathbf{z}_1^H\mathbf{t}_2\|}{\sigma_n^2(\alpha^2 a + 1)}.$$

By forcing $\dot{\mathcal{L}}_1(b) = 0$, we have $\hat{b} = \frac{\pm\sqrt{B_2^2 + 4B_1} + B_2}{2B_1}$. Since $\lim_{b\to-\infty}\dot{\mathcal{L}}_1(b) \to +\infty$ and $\lim_{b\to\infty}\dot{\mathcal{L}}_1(b) \to -\infty$, it can be known that $\frac{\sqrt{B_2^2+4B_1}+B_2}{2B_1} > 0$ is the local maximal point and $\frac{-\sqrt{B_2^2+4B_1}+B_2}{2B_1} < 0$ is the local minimal one which is not effective in this estimation. Therefore, $\hat{b}$ is derived by

$$\hat{b} = \frac{\sqrt{B_2^2 + 4B_1} + B_2}{2B_1}, \quad (12)$$

with given a and θ_a. Then substituting (12) into (9), $\mathcal{L}(\mathbf{x}_S|a, \theta_a, b)$ transforms to $\mathcal{L}_2(\mathbf{x}_S|a, \theta_a)$ which has only two remaining parameters a and θ_a. Here, we omit to derive the expression of $\mathcal{L}_2(\mathbf{x}_S|a, \theta_a)$since it is rather complicated. It can be testified that $\mathcal{L}_2(\mathbf{x}_S|a, \theta_a)$ is non-concave with respect to a, so that the numerical methods such as 2-dimension searching can be utilized to achieve feasible estimation and the computational complexity would be extremely high. In order to reduce the calculating consumption, we provide the iterative algorithm as Table I shows to obtain the MAP estimates.

TABLE I
ITERATIVE ALGORITHM OF COMPOSITE CHANNEL ESTIMATION

- **Initialize** $\{\hat{a}, \hat{\theta}_a\}$ with a certain estimator of lower complexity and $\hat{\boldsymbol{\Theta}} = \emptyset$.
- **Repeat**
  - Calculate current $\hat{\theta}_b$ in terms of (8) with $\{\hat{a}, \hat{\theta}_a\}$ injected into it.
  - Calculate current $\hat{b}$ by plugging $\{\hat{a}, \hat{\theta}_a\}$ into (12).
  - Update $\hat{\theta}_a$ by $\hat{\theta}_a = \angle\left\{\frac{\mathbf{t}_1^H\left(\mathbf{x}_S - \alpha\hat{b}e^{j\hat{\theta}_b}\mathbf{t}_2\right)}{\alpha\|\mathbf{t}_1\|^2}\right\}$.
  - Update $\hat{a}$ by
  $$\hat{a} = \arg\max_a \left\{ -N\log\left(\alpha^2 a + 1\right) - \log(a) - \frac{a}{\upsilon_1} - \frac{\hat{b}^2}{\upsilon_2 a} - \frac{\|\mathbf{x}_S - \alpha a e^{j\hat{\theta}_a}\mathbf{t}_1 - \alpha\hat{b}e^{j\hat{\theta}_b}\mathbf{t}_2\|^2}{\sigma_n^2(\alpha^2 a + 1)}\right\}. \quad (13)$$
  - Renew $\hat{\boldsymbol{\Theta}} = \{\hat{a}, \hat{b}, \hat{\theta}_a, \hat{\theta}_b\}$.
- **Until** termination criterion is satisfied.
- **Return** $\hat{\boldsymbol{\Theta}}$.

Lemma *1:* The proposed iterative algorithm is convergent and the limit point of iteration is a stationary point.

Proof: Note that, one round of iteration consists of four steps where each step aims at renewing one entry of $\boldsymbol{\Theta}$. For notation simplification, we further denote θ_i to be the i-th entry of $\boldsymbol{\Theta}$. Consider the step of renewing θ_i, the renewed estimate of θ_i denoted by $\hat{\theta}_i^{new}$ would satisfy $\mathcal{L}\left(\mathbf{x}_S|\hat{\theta}_i^{new}\right) > \mathcal{L}\left(\mathbf{x}_S|\hat{\theta}_i\right)$ with given other parameters in $\boldsymbol{\Theta}$. This indicates that the value of $\mathcal{L}(\mathbf{x}_S|\boldsymbol{\Theta})$ would strictly increase after each step, thus it would also strictly increase after one round of iteration. Besides, it can be testified that $\mathcal{L}(\mathbf{x}_S|\boldsymbol{\Theta})$ is continuous with respect to θ_i and $\mathcal{L}(\mathbf{x}_S|\boldsymbol{\Theta}) < +\infty$ for $a > 0$. Therefore, we can conclude that the iterative algorithm is convergent.

Denote the limited point θ_i to be $\bar{\theta}_i$. At the limit point, the solution will not change if we continue the iteration. Otherwise, the value of $\mathcal{L}(\mathbf{x}_S|\boldsymbol{\Theta})$ can be further increased which contradicts its convergence behavior. Since the resultant estimate in each step is the global maximum as well as the local maximum, thus we can have

$$\frac{\partial\mathcal{L}(\mathbf{x}_S|\boldsymbol{\Theta})}{\partial\theta_i}\left(\theta_i - \bar{\theta}_i\right) \le 0 \quad (14)$$

which implies the stationarity of the iteration result. ■

Remark:

- The initialization of the algorithm needs to guess the value of $\{a, \theta_a\}$, and it is very important since it affects the convergence point. So the ML estimator presented in [5] is recommended to perform the initial guess.
- Focusing on the updating of $\hat{a}$, the first order derivation of the under maximized function can be organized to a

cubic equation with respect to a so that its peak points expression can be given. Hence, $\hat{a}$ can be renewed by the optimal value among its local maximal points and the end point $a = 0$.

B. MAP-Based Individual Channel Estimation

The individual channel coefficients are useful to support diversity techniques such as subcarrier pairing [8] and relay selection [9], where only channel amplitude information is needed. Therefore, we only focus on estimating the individual channel amplitude denoted by $u_i = |h_i|, i = 1, 2$. Note that, u_i can be calculated by straightforward calculation in terms of composite channel estimates, but the transformation from composite channel to individual channel would increase the estimated errors. Intuitively, let $\hat{a} = a + \Delta a$ and $\hat{b} = b + \Delta b$, the normalized MSE of b yields $\mathfrak{E}\{\frac{\|\Delta b\|^2}{b^2}\}$. When we utilize the composite channel estimates $\hat{a}$ and $\hat{b}$ to calculate individual estimate $\hat{u}_2$, the normalized MSE of u_2 can be approximately computed as $\mathfrak{E}\{\frac{\|\Delta b\|^2}{b^2} + \frac{\|\Delta a\|^2}{2a^2}\}$ which is larger than that of b. This implies that the transformation from composite channel to individual channel would enlarge estimated errors. Thus we focus on directly obtaining the estimate of individual channel with MAP based estimation. The under estimated parameters is defined by $\mathbf{\Theta}' = [u_1, u_2, \theta_a, \theta_b]^T$. Our aim is to estimate the magnitude u_i, while θ_a and θ_b are used for assistance.

Since h_1 and h_2 are independent with each other, $p\left(\mathbf{\Theta}'\right)$ is given by

$$p\left(\mathbf{\Theta}'\right) = \frac{u_1 u_2}{\pi^2 \upsilon_1 \upsilon_2} \exp\left\{-\left(\frac{u_1^2}{\upsilon_1} + \frac{u_2^2}{\upsilon_2}\right)\right\}. \tag{15}$$

Recalling the conditional p.d.f in (4), $\mathbf{\Theta}'$ can be obtained by

$$\mathbf{\Theta}' = \arg\max_{\mathbf{\Theta}'}\left\{-\frac{\|\mathbf{x}_S - \alpha u_1^2 e^{j\theta_a}\mathbf{t}_1 - \alpha u_1 u_2 e^{j\theta_b}\mathbf{t}_2\|^2}{\sigma_n^2\left(\alpha^2 u_1^2 + 1\right)} - N\log\left(\alpha^2 u_1^2 + 1\right) - \left(\frac{u_1^2}{\upsilon_1} + \frac{u_2^2}{\upsilon_2}\right) + \log\left(u_1 u_2\right)\right\} \tag{16}$$

where the terms inside the max operator is denoted by $\mathcal{L}\left(\mathbf{x}_S|\mathbf{\Theta}'\right)$.

Recalling (8), $\hat{\theta}_b$ can be obtained similarly as

$$\hat{\theta}_b = \angle\left(\frac{\mathbf{t}_2^H}{\alpha\|\mathbf{t}_2\|^2}\mathbf{z}_1'\right) \tag{17}$$

with given u_1, u_2 and θ_a, where $\mathbf{z}_1' \triangleq \mathbf{x}_S - \alpha u_1^2 e^{j\theta_a}\mathbf{t}_1$. Substituting above equation into $\mathcal{L}\left(\mathbf{x}_S|\mathbf{\Theta}'\right)$, and utilizing derivative approach, $\hat{u}_2$ is derived as

$$\hat{u}_2 = \frac{\sqrt{C_2^2 + 4C_1} + C_2}{2C_1} \tag{18}$$

with given u_1 and θ_a, where

$$C_1 = \frac{2\alpha^2 u_1^2\|\mathbf{t}_2\|^2}{\sigma_n^2\left(\alpha^2 u_1^2 + 1\right)} + \frac{2}{\upsilon_2}, \quad C_2 = \frac{2\alpha u_1\|\mathbf{z}_1^H\mathbf{t}_2\|}{\sigma_n^2\left(\alpha^2 u_1^2 + 1\right)}.$$

With (17) and (18) substituted into $\mathcal{L}\left(\mathbf{x}_S|\mathbf{\Theta}'\right)$, the resultant $\mathcal{L}\left(\mathbf{x}_S|u_1, \theta_a\right)$ has only two remaining parameters u_1 and θ_a. Similar to the composite estimation, the iterative algorithm is also effective to obtain $\hat{u}_1$ and $\hat{u}_2$ which is shown in Table II.

TABLE II
ITERATIVE ALGORITHM OF INDIVIDUAL CHANNEL ESTIMATION

- **Initialize** $\{\hat{u}_1, \hat{\theta}_a\}$ with a certain estimator of lower complexity, and $\hat{\mathbf{\Theta}}' = \phi$.
- **Repeat**
  - Calculate current $\hat{\theta}_b$ in terms of (17) with $\{\hat{u}_1, \hat{\theta}_a\}$ injected in it.
  - Calculate current $\hat{u}_2$ by plugging $\{\hat{u}_1, \hat{\theta}_a\}$ into (18).
  - Update $\hat{\theta}_a$ by $\hat{\theta}_a = \angle\left\{\frac{\mathbf{t}_1^H\left(\mathbf{x}_S - \alpha\hat{u}_1\hat{u}_2 e^{j\hat{\theta}_b}\mathbf{t}_2\right)}{\alpha\|\mathbf{t}_1\|^2}\right\}$.
  - Update $\hat{u}_1$ as
  $$\hat{u}_1 = \arg\max_{u_1}\left\{-N\log\left(\alpha^2 u_1^2 + 1\right) + \log\left(u_1\right) - \frac{u_1^2}{\upsilon_1} - \frac{\|\mathbf{x}_S - \alpha u_1^2 e^{j\hat{\theta}_a}\mathbf{t}_1 - \alpha u_1\hat{u}_2 e^{j\hat{\theta}_b}\mathbf{t}_2\|^2}{\sigma_n^2\left(\alpha^2 a + 1\right)}\right\}. \tag{19}$$
  - Renew $\hat{\mathbf{\Theta}}' = \{\hat{u}_1, \hat{u}_2, \hat{\theta}_a, \hat{\theta}_b\}$.
- **Until** termination criterion is satisfied.
- **Return** $\hat{\mathbf{\Theta}}'$.

Remark:

- Above algorithm is convergent and stationary which can be proved in similar way as the proof of Lemma 1.
- It is necessary to mention that the composite and individual channel estimation can achieve same estimates for $|h_1|$ and $|h_2|$ utilizing ML estimator. However, the two estimations would be different utilizing MAP estimator due to the difference of the chosen apriori p.d.fs. We use simulations to compare the performance of the two MAP estimation methods in the simulation part.

C. Asymptotic Behavior of MAP Estimation

The following proposition is given to show the asymptotic behavior of MAP estimator.

Proposition *1:* The MAP estimate converges to ML solution when accumulative training SNR is sufficiently large, i.e., $\frac{NP_S}{\sigma_n^2} \to \infty$.

Proof: Define $Q \triangleq \frac{NP_S}{\sigma_n^2}$, and the limitation of the normalized l.l.f with respect to Q is

$$\lim_{Q\to\infty}\frac{\mathcal{L}_{MAP}\left(\mathbf{x}_S;\mathbf{\Theta}\right)}{Q} = -\frac{\|\mathbf{x}_S - \alpha h_a\mathbf{t}_1 - \alpha h_b\mathbf{t}_2\|^2}{NP_S\left(\alpha^2 a + 1\right)} - \frac{\sigma_n^2}{P_S}\log\left(\alpha^2 a + 1\right) - \lim_{Q\to\infty}\left(\frac{a}{Q\upsilon_1} + \frac{b^2}{Qa\upsilon_2} - \frac{1}{Q}\log\left(\frac{b}{a}\right)\right) = \frac{\mathcal{L}_{ML}\left(\mathbf{x}_S;\mathbf{\Theta}\right)}{Q} \neq 0. \tag{20}$$

Then it yields

$$\lim_{Q\to\infty}\hat{\mathbf{\Theta}}_{MAP} = \lim_{Q\to\infty}\{\arg\max_{\mathbf{\Theta}}\mathcal{L}_{MAP}\left(\mathbf{x}_S;\mathbf{\Theta}\right)\} = \arg\max_{\mathbf{\Theta}}\lim_{Q\to\infty}\frac{\mathcal{L}_{MAP}\left(\mathbf{x}_S;\mathbf{\Theta}\right)}{Q} = \arg\max_{\mathbf{\Theta}}\frac{\mathcal{L}_{ML}\left(\mathbf{x}_S;\mathbf{\Theta}\right)}{Q} = \hat{\mathbf{\Theta}}_{ML}, \tag{21}$$

thus the proposition is achieved. ■

Considering a much stronger constraint, i.e., $\frac{P_S}{\sigma_n^2} \to \infty$, we can have

$$\lim_{\frac{P_S}{\sigma_n^2}\to\infty} \frac{\mathcal{L}_{MAP}\left(\mathbf{x}_S;\mathbf{\Theta}\right)}{P_S/\sigma_{n^2}} = -\frac{\|\mathbf{x}_S-\alpha h_a\mathbf{t}_1-\alpha h_b\mathbf{t}_2\|^2}{P_S\left(\alpha^2 a+1\right)} \quad (22)$$
$$-\lim_{\frac{P_S}{\sigma_n^2}\to\infty}\left(\frac{\sigma_n^2}{P_S}\left(\frac{a}{\upsilon_1}+\frac{b^2}{a\upsilon_2}\right)-\frac{\sigma_n^2}{P_S}\log\left(\frac{b}{a}\right)\right)$$
$$-\lim_{\frac{P_S}{\sigma_n^2}\to\infty}\frac{\sigma_n^2}{P_S}N\log\left(\alpha^2 a+1\right) = -\frac{\|\mathbf{x}_S-\alpha h_a\mathbf{t}_1-\alpha h_b\mathbf{t}_2\|^2}{P_S\left(\alpha^2 a+1\right)}.$$

Then the MAP estimate attains

$$\begin{aligned}\lim_{\frac{P_S}{\sigma_n^2}\to\infty}\hat{\mathbf{\Theta}}_{MAP} &= \arg\max_{\mathbf{\Theta}}\{-\frac{\|\mathbf{x}_S-\alpha h_a\mathbf{t}_1-\alpha h_b\mathbf{t}_2\|^2}{P_S\left(\alpha^2 a+1\right)}\}\\ &= \arg\max_{\mathbf{\Theta}}\{-\|\mathbf{x}_S-\alpha h_a\mathbf{t}_1-\alpha h_b\mathbf{t}_2\|^2\}\\ &= \hat{\mathbf{\Theta}}_{LS}. \end{aligned} \quad (23)$$

Hence, it can be concluded that the MAP estimation degrades to the LS solution when $\frac{P_S}{\sigma_n^2}$ is sufficiently large.

Proposition *2:* In memoryless quasi-static flat fading channel with independent fading coefficients, both MAP and ML estimation problem are related to minimum Kullback-Leibler divergence problem when training sequence length N is sufficiently large.

Proof: See in the appendix B. ■

D. MSE Comparison Between MAP and ML Estimations

Owing to the nonlinearity of the MAP estimation, it is hard to derive the closed expression of the MSE or as such. In this part, the MSE of h_b with given h_a is analyzed. In [5], it has been proved that the orthogonal training is the optimal design for TWRC, thus $\mathbf{t}_1^H\mathbf{t}_2 = 0$ is assumed here.

In [5], the MSE of h_b denoted by δ_1 for ML estimation is given as

$$\delta_1 = \frac{1}{NP_S}\{\underbrace{\sigma_n^2\left(a+1/\alpha^2\right)}_{\delta_E}\}. \quad (24)$$

For MAP estimation, recalling (12), the MAP estimate $\hat{b}$ with given a is rewritten as

$$\hat{b}_{MAP} = \sqrt{\left(\frac{B_2}{2B_1}\right)^2+\frac{1}{B_1}}+\frac{B_2}{2B_1}. \quad (25)$$

Since $\|\mathbf{t}_2\|^2 = NP_S$, it is known that $B_1, B_2 \sim O\left(N\right)$, $\frac{B_2}{2B_1}\sim O\left(1\right)$ and $\frac{1}{B_1}\sim O\left(\frac{1}{N}\right)$, where $O\left(\cdot\right)$ represents the equivalent decreasing or increasing order of the argument. When N is sufficiently large, we can have $\frac{1}{B_1} \ll \frac{B_2}{2B_1}$. Utilizing the Taylor series expansion, $\hat{b}_{MAP}$ is approximately derived as

$$\hat{b}_{MAP} \approx \frac{B_2}{B_1}+\frac{1}{B_2} = g\left(N\right)\frac{\|\mathbf{t}_2^H\mathbf{z}_1\|}{\alpha\|\mathbf{t}_2\|^2}+\frac{1}{B_2},$$

with $g\left(N\right) = \frac{1}{1+\frac{\sigma_n^2\left(\alpha^2 a+1\right)}{\alpha^2 a\upsilon_2 P_S N}}$. With $g\left(N\right)$ expanded, $\hat{b}_{MAP}$ can be further written as

$$\hat{b}_{MAP} \approx \left(1-B_3\right)\frac{\|\mathbf{t}_2^H\mathbf{z}_1\|}{\alpha\|\mathbf{t}_2\|^2}+\frac{1}{B_2}, \quad (26)$$

with $B_3 = \frac{\sigma_n^2\left(\alpha^2 a+1\right)}{\alpha^2 a\upsilon_2 P_S N}\sim O\left(\frac{1}{N}\right)$. After some derivation, the MSE of h_b with MAP estimation denoted by δ_2 can be written by

$$\begin{aligned}\delta_2 &= \mathfrak{E}\Big\{\Big\|\left(1-B_3\right)\frac{\mathbf{t}_2^H\mathbf{z}_1}{\|\mathbf{t}_2\|^2}+\frac{1}{B_2}e^{j\hat{\theta}_b}-h_b\Big\|^2\Big\}\\ &= \left(1-B_3\right)^2\frac{\delta_E}{NP_S}+B_3^2 b^2+\mathfrak{E}\{\frac{1}{B_2^2}\}-\mathfrak{E}\{\frac{B_3 h_b\delta_E}{2\mathbf{t}_2^H\mathbf{z}_1/\alpha}\}\\ &\quad -\mathfrak{E}\{\frac{B_3 h_b\delta_E}{2\mathbf{z}_1^H\mathbf{t}_2/\alpha}\}+\mathfrak{E}\{\frac{\left(1-B_3\right)^2\mathbf{t}_2^H\mathbf{n}_E}{\|\mathbf{t}_2\|^2}\cdot\frac{\delta_E}{2\mathbf{t}_2^H\mathbf{z}_1/\alpha}\}\\ &\quad +\mathfrak{E}\{\frac{\left(1-B_3\right)^2\mathbf{n}_E^H\mathbf{t}_2}{\|\mathbf{t}_2\|^2}\cdot\frac{\delta_E}{2\mathbf{z}_1^H\mathbf{t}_2/\alpha}\}, \end{aligned} \quad (27)$$

since $\mathfrak{E}\{\mathbf{t}_2^H\mathbf{n}_E\} = 0$ with $\mathbf{n}_E \triangleq a\mathbf{n}_R+\mathbf{n}_S/\alpha$. Note that the last four terms of above equation contains the term $\mathbf{t}_2^H\mathbf{z}_1 = \alpha h_b\|\mathbf{t}_2\|^2+\mathbf{t}_2^H\mathbf{n}_E$ in the denominator. Over high SNR region, we can attain $\mathbf{t}_2^H\mathbf{z}_1\approx\alpha h_b\|\mathbf{t}_2\|^2$. After tedious calculation, δ_2 is approximately derived as

$$\begin{aligned}\delta_2 &\approx \frac{\left(1-B_3\right)^2\delta_E}{NP_S}+B_3^2 b^2+\left[\frac{\delta_E}{bNP_S+NP_S\delta_E}\right]^2-\frac{B_3\delta_E}{NP_S}\\ &< \frac{\delta_E}{NP_S}+\left(\frac{\delta_E}{NP_S}\right)^2\left(\frac{b^2}{a^2\upsilon_2^2}+\frac{1}{b^2}-\frac{3}{a\upsilon_2}\right)\\ &\quad +\left(\frac{\delta_E}{NP_S}\right)^3\left(\frac{1}{a\upsilon_2}\right). \end{aligned} \quad (28)$$

It can be observed that all the three terms in above equation are asymptotically decreasing over N, the last term can be removed since it has the highest order, i.e., $O\left(\frac{1}{N^3}\right)$. After some simplification, (28) is rewritten as

$$\begin{aligned}\delta_2 &\lessapprox \frac{\delta_E}{NP_S}+\left(\frac{\delta_E}{NP_S\upsilon_2}\right)^2\left(\frac{b^2}{a^2}+\frac{\upsilon_2^2}{b^2}-\frac{3\upsilon_2}{a}\right)\\ &= \frac{\delta_E}{NP_S}+\frac{1}{a}\left(\frac{\delta_E}{NP_S\upsilon_2}\right)^2\left(c+\frac{\upsilon_2^2}{c}-3\upsilon_2\right) \end{aligned} \quad (29)$$

with $c\triangleq b^2$. Focusing on the function $f\left(c\right) = c+\frac{\upsilon_2^2}{c}-3\upsilon_2$, it can be testified that $f\left(c\right)\le 0$ when $\frac{3-\sqrt{5}}{2}\upsilon_2\le c\le\frac{3+\sqrt{5}}{2}\upsilon_2$. Since c satisfies the exponential distribution with parameter $\frac{1}{\upsilon_2}$, we can calculate that $P\{f\left(c\right)\le 0\}\ge\int_{\frac{3-\sqrt{5}}{2}\upsilon_2}^{\frac{3+\sqrt{5}}{2}\upsilon_2}\frac{1}{\upsilon_2}e^{-\frac{x}{\upsilon_2}}\mathrm{d}x > 0.6$ which indicates that

$$P\{\delta_2<\delta_1\}\ge 0.6. \quad (30)$$

Namely, the MAP estimation performs better than ML estimation with high probability. The average performance of the MAP and ML estimations are evaluated and compared in the simulation part.

E. Complexity Comparison

Since the source nodes acquire the knowledge of $\mathbf{t}_i$, the terms such as $\|\mathbf{t}_i\|^2$ and $\mathbf{t}_1^H\mathbf{t}_2$ can be obtained utilizing offline computation. Hence, only the computation of the terms that contains $\mathbf{x}_S$ consumes online operations including multiplication and addition. Consider the iterative step in MAP estimation algorithm, the complexity is mainly determined by the calculation of $\|\mathbf{x}_S\|^2$, $\mathbf{t}_1^H\mathbf{x}_S$ and $\mathbf{t}_2^H\mathbf{x}_S$ due to the fact that their complexity are determined by N. Each of them consists of N multiplication operations and $(N-1)$ addition operations. In particular, it is necessary to stress that once the corresponding terms are calculated, the results can be used in each iteration. Hence, the computation complexity of each iteration is independent with N. Then the total computational complexity of the iteration is approximately written as

$$3\left[N \cdot o\left(m\right) + \left(N-1\right) \cdot o\left(a\right)\right] + 4i \cdot o\left(b\right), \tag{31}$$

where $o\left(m\right)$ and $o\left(a\right)$ represents the computation complexity of a complex multiplication and addition, respectively. $o\left(b\right)$ denotes the average complexity in calculating a closed form expression that is unrelated to N, and i denotes the iterative time. For the ML solution in [5], its complexity are also mainly determined by the terms $\|\mathbf{x}_S\|^2$, $\mathbf{t}_1^H\mathbf{x}_S$ and $\mathbf{t}_2^H\mathbf{x}_S$. Thus its complexity is calculated as $3\left[N \cdot o\left(m\right) + \left(N-1\right) \cdot o\left(a\right)\right]$, and the ML initialization would not increase the complexity for the iterative MAP estimation algorithm. Then the total complexity for MAP estimation is as (31) shows. Obviously, the complexity of iterative MAP algorithm is always higher than that of ML solution, and it is affected by the iterative time. However, their computation complexity are of the same level for bounded iterative time.

IV. Improved Channel Estimation Strategy

In above estimation schemes, both noise variance and CDI are assumed to be known at the receivers. Here, we first focus on the occasion of unknown noise variance, then consider the case of unknown CDI.

A. With Unknown and Partial Information of Noise Variance

In this part, we focus on the two occasions, i.e., the noise variance σ_n^2 is unknown at the source, and the statistical information of σ_n^2 is acquired.

1) With unknown σ_n^2: Obviously, when σ_n^2 is unknown, the MAP or ML estimation can not be implemented because the corresponding l.l.f can hardly be constructed. Intuitively, the least square (LS) estimator can be effectively utilized since it only needs the knowledge of $\mathbf{t}_i$. Thus the estimates of composite channels of LS solution are obtained by

$$\left[\hat{h}_a \, \hat{h}_b\right] = \frac{1}{\alpha}\left(\mathbf{T}^H\mathbf{T}\right)^{-1}\mathbf{T}^H \cdot \mathbf{x}_S \tag{32}$$

with $\mathbf{T} = \left[\mathbf{t}_1 \, \mathbf{t}_2\right]$. Then the estimate of $\sigma_E^2 = \left(\alpha^2|h_a|\sigma_n^2+1\right)\sigma_n^2$ is computed by

$$\hat{\sigma}_E^2 = \frac{1}{N}\|\mathbf{x}_S - \alpha\hat{h}_a\mathbf{t}_1 - \alpha\hat{h}_b\mathbf{t}_2\|^2. \tag{33}$$

Treating $\hat{\sigma}_E^2$ as a deterministic value, the estimate of $\mathbf{\Theta}$ can be renewed by

$$\hat{\mathbf{\Theta}} = \arg\max_{\mathbf{\Theta}}\Bigg\{-\frac{\|\mathbf{x}_S - \alpha h_a\mathbf{t}_1 - \alpha h_b\mathbf{t}_2\|^2}{\hat{\sigma}_E^2} - \frac{a}{\upsilon_1} - \frac{b^2}{a\upsilon_2} + \log\left(\frac{b}{a}\right)\Bigg\}, \tag{34}$$

where the terms inside the max operator is denoted by $\mathcal{L}\left(\mathbf{x}_S|\mathbf{\Theta},\hat{\sigma}_E^2\right)$. To improve the estimation accuracy, an iterative algorithm is developed which is shown as Table III shows.

We can expect that the estimating accuracy would be improved compared with conventional LS estimation due to the usage of apriori information.

TABLE III
Iterative Algorithm with Unknown Noise Variance

- **Initialize** $\{\hat{h}_a, \hat{h}_b\}$ with LS estimation in terms of (32).
- **Repeat**
  - Calculate $\hat{\sigma}_E^2$ in terms of (33).
  - Renew $\hat{\mathbf{\Theta}}$ by $\hat{\mathbf{\Theta}} = \arg\max_{\mathbf{\Theta}} \mathcal{L}\left(\mathbf{x}_S|\mathbf{\Theta},\hat{\sigma}_E^2\right)$.
  - Renew $\hat{h}_a = \hat{a}e^{j\hat{\theta}_a}$ and $\hat{h}_b = \hat{b}e^{j\hat{\theta}_b}$.
- **Until** termination criterion is satisfied.
- **Return** $\hat{\mathbf{\Theta}}$.

2) With Partial Information of σ_n^2: Here, we concern the occasion that the distribution information of σ_n^2 is acquired at the receiver. Considering σ_n^2 as one of the remaining parameters, the MAP estimation is executed as

$$\{\hat{\mathbf{\Theta}}, \hat{\sigma}_n^2\} = \arg\max_{\theta,\sigma_n^2} p\left(\mathbf{x}_S|\mathbf{\Theta},\sigma_n^2\right)p\left(\mathbf{\Theta},\sigma_n^2\right), \tag{35}$$

where $p\left(\mathbf{\Theta},\sigma_n^2\right)$ is the joint p.d.f. Obviously, the optimization problem has total five parameters so that its computational complexity is extremely high utilizing five dimensional searching method, and the estimation accuracy can not be guaranteed. In order to reduce the computational complexity, we treat σ_n^2 as a redundant factor and the estimation problem becomes

$$\hat{\mathbf{\Theta}} = \arg\max_{\mathbf{\Theta}} p\left(\mathbf{x}_S|\mathbf{\Theta},\sigma_n^2\right)p\left(\mathbf{\Theta}\right). \tag{36}$$

Obviously, the estimation problem can be hardly solved with unknown σ_n^2. To tackle the problem, the expectation of the posteriori probability with respect to σ_n^2 is chosen to be the objective function, which is given as

$$\hat{\mathbf{\Theta}} = \arg\max \mathfrak{E}\{p\left(\mathbf{x}_S|\mathbf{\Theta},\sigma_n^2\right)p\left(\mathbf{\Theta}\right)\}. \tag{37}$$

Here, it is assumed that $\mathbf{\Theta}$ and σ_n^2 are independent of each other and σ_n^2 satisfies the inverse-gamma distribution whose p.d.f is given by

$$p\left(\sigma_n^2\right) = \frac{\lambda}{\sigma_n^4}\exp\left(-\frac{\lambda}{\sigma_n^2}\right), \tag{38}$$

where λ is a positive factor. With the change of $\xi \triangleq \frac{1}{\sigma_n^2}$ and the definition of $\mathbf{z}_S \triangleq \mathbf{x}_S - \alpha h_a\mathbf{t}_1 - \alpha h_b\mathbf{t}_2$, we have

$$\begin{aligned}
&\mathfrak{E}\{p\left(\mathbf{x}_S|\mathbf{\Theta},\sigma_n^2\right)\} \qquad (39)\\
&= \int_0^\infty \frac{\xi^N}{\left[\pi\left(\alpha^2 a+1\right)\right]^N}\exp\Big\{-\frac{\|\mathbf{z}_S\|^2\xi}{\alpha^2 a+1}\Big\}\cdot\lambda\exp\left(-\lambda\xi\right)\mathrm{d}\xi\\
&= \frac{\lambda}{\left[\pi\left(\alpha^2 a+1\right)\right]^N}\int_0^\infty \xi^N\exp\left\{-\left(\frac{\|\mathbf{z}_S\|^2\xi}{\alpha^2 a+1}+\lambda\right)\xi\right\}\mathrm{d}\xi.
\end{aligned}$$

After straightforward calculation, above expectation is derived as

$$\mathfrak{E}\{p\left(\mathbf{x}_S|\mathbf{\Theta},\sigma_n^2\right)\} = \frac{\lambda\Gamma\left(N+1\right)}{\pi^N\left(\alpha^2 a+1\right)^N\eta^{N+1}}, \tag{40}$$

where $\eta = \frac{\|\mathbf{z}_S\|^2\xi}{\alpha^2 a+1} + \lambda$, and $\Gamma\left(z\right)$ is the Gamma function given by $\Gamma\left(z\right) = \int_0^\infty x^{z-1}e^{-x}dx$. Hence, the estimation can be successfully executed since σ_n^2 has been removed from the problem.

B. With Unknown CDI

In this part, we deal with the occasion that CDI is unknown at the receivers. With the assumption that channels during each block have independent and identical distributions, the desired CDI can be obtained in terms of multiple observations of instantaneous channel coefficients so as to perform MAP method to enhance instantaneous estimation accuracy.

1) Obtaining Channel Statistic: Consider the long term transmission which consists of multiple blocks. One may calculate the CDI by fitting all the instantaneous channel estimates during current block and previous blocks. But using fitting methods may result in complicated p.d.fs whose mathematical expression may be hardly obtained. For simplification, we constrain the channel to satisfy the empirical channel models such as Rayleigh fading or Nakagami-m channels, so that the p.d.f expression can be obtained in terms of its statistic, e.g., mean and variance. In particular, only the variance υ_1 and υ_2 are required to be estimated for Rayleigh fading channels.

Denote the channel variance estimate of υ_i during the k-th block by $\hat{\upsilon}_i(k)$. For composite channel estimation, the instantaneous estimates of $a(k)$ and $b(k)$ are denoted by $\hat{a}(k)$ and $\hat{b}(k)$, respectively. Then $\hat{\upsilon}_i(k)$ can be estimated by

$$\hat{\upsilon}_1(k) = \frac{1}{k}\sum_{l=1}^{k}\hat{a}(l), \hat{\upsilon}_2(k) = \frac{\sum_{l=1}^{k}\left[\hat{b}(l)\right]^2}{k\cdot\hat{\upsilon}_1(k)} \quad (41)$$

Proposition *3:* The estimation for υ_i is unbiased when orthogonal training is utilized, i.e., $\mathbf{t}_1^H\mathbf{t}_2 = 0$, and the symbol signal-to-noise ratio $\frac{P_S}{\sigma_n^2}$ is sufficiently large.

Proof: The proposition is equivalent to prove $\lim_{k\to\infty}\mathfrak{E}\{\hat{\upsilon}_i(k)\} = \upsilon_i$, where the expectation is taken with respect to the noise vectors. It is known that $\hat{\mathbf{\Theta}}_{MAP} \to \hat{\mathbf{\Theta}}_{LS}$ when $\frac{P_S}{\sigma_n^2}$ is sufficiently large, thus $\hat{a}(l)$ and $\hat{b}(l)$ can be approximately obtained in terms of the LS estimation as

$$\hat{a}(l) = \left|\frac{\mathbf{t}_1^H}{\|\mathbf{t}_1\|^2}\mathbf{x}_S(l)\right|, \hat{b}(l) = \left|\frac{\mathbf{t}_2^H}{\|\mathbf{t}_2\|^2}\mathbf{x}_S(l)\right|. \quad (42)$$

With $\mathbf{t}_1^H\mathbf{t}_2 = 0$, $\hat{a}(l)$ can be specifically written as

$$\hat{a}(l) = \left|a(l)e^{j\theta_a(l)} + \frac{\mathbf{t}_1^H}{\|\mathbf{t}_1\|^2}(\alpha h_1\mathbf{n}_R + \mathbf{n}_S)\right|. \quad (43)$$

It can be observed that $\hat{a}(l)$ satisfies Rician distribution with $a(l)$ mean and $\frac{\left[\alpha^2 a(l)+1\right]\sigma_n^2}{NP_S}$ variance. Thus we have

$$\mathfrak{E}\{\hat{a}(l)\} = \sqrt{\frac{\pi(\alpha^2 a(l)+1)}{2NP_S}}\sigma_n L_{\frac{1}{2}}\left(-\frac{a(l)^2 NP_S}{2(\alpha^2 a(l)+1)\sigma_n^2}\right), \quad (44)$$

where $L_{\frac{1}{2}}(x) = e^{\frac{x}{2}}\left[(1-x)I_0(x) - xI_1(x)\right]$ and $I_\beta(x)$ represents the modified Bessel function of the first kind with order β. For $L_{\frac{1}{2}}(x)$, the approximation $L_{\frac{1}{2}}(x) = 2\sqrt{\frac{|x|}{\pi}}$ is holden on the condition of $x \to -\infty$, which leads to $\mathfrak{E}\{\hat{a}(l)\} \approx a(l)$. In accordance with the Law of Large Numbers, we can get $\lim_{k\to\infty}\mathfrak{E}\{\hat{\upsilon}_1(k)\} = \upsilon_1$.

While for $\hat{\upsilon}_2(k)$, we can approximately get

$$\hat{\upsilon}_2(k) \approx \frac{\sum_{l=1}^{k}\left[b(l)+\Delta b_l\right]^2}{\sum_{l=1}^{k}a(l)}\left[1 - \frac{1}{\left(1+\frac{\sum_{l=1}^{k}\Delta a_l}{\sum_{l=1}^{k}a(l)}\right)^2}\frac{\sum_{l=1}^{k}\Delta a_l}{\sum_{l=1}^{k}a(l)}\right] \quad (45)$$

with $\Delta a_l = \hat{a}(l) - a(l)$ and $\Delta b_l = \hat{b}(l) - b(l)$. When $\frac{P_S}{\sigma_n^2}$ is sufficiently large, it yields $\mathfrak{E}\{\hat{\upsilon}_2(k)\} \approx \frac{\sum_{l=1}^{k}[b(l)]^2}{\sum_{l=1}^{k}a(l)}$ which results in $\lim_{k\to\infty}\mathfrak{E}\{\hat{\upsilon}_2(k)\} = \frac{k\upsilon_1\upsilon_2}{k\upsilon_1} = \upsilon_2$. The proof is finished. ■

More specifically, if the two individual channels are of identical distribution with υ variance, then the channel variance estimate during the k-th block denoted by $\hat{\upsilon}(k)$ can be estimated by

$$\hat{\upsilon}(k) = \frac{1}{2k}\sum_{l=1}^{k}\left[\hat{a}(l) + \frac{2}{\pi}\hat{b}(l)\right]. \quad (46)$$

Proposition *4:* The estimation for υ in (46) is unbiased when orthogonal training is utilized, i.e., $\mathbf{t}_1^H\mathbf{t}_2 = 0$, and the symbol signal-to-noise ratio $\frac{P_S}{\sigma_n^2}$ is sufficiently large.

Proof: Similarly to $\hat{a}(l)$, $\hat{b}(l)$ also satisfies the Rician distribution with $b(l)$ mean and the same variance as $\hat{a}(l)$. $\mathfrak{E}\{\hat{b}(l)\} \approx b(l)$ can be obtained in similar way with $\frac{P_S}{\sigma_n^2}$ to be sufficiently large. Then the expectation of $\hat{\upsilon}(k)$ yields

$$\mathfrak{E}\{\hat{\upsilon}(k)\} = \frac{1}{2k}\sum_{l=1}^{k}\left[a(l) + \frac{2}{\pi}b(l)\right]. \quad (47)$$

Utilizing the Law of Large Numbers, $\lim_{k\to\infty}\mathfrak{E}\{\hat{\upsilon}(k)\} = \upsilon$ is achieved since $\mathfrak{E}\{a(l)\} = \upsilon$ and $\mathfrak{E}\{b(l)\} = \frac{\pi}{2}\upsilon$. This indicates the estimation is unbiased. ■

Although $\hat{\upsilon}(k)$ is unbiased, the estimation accuracy is strongly determined by k. To evaluate its performance, the MSE of $\hat{\upsilon}(k)$ is analyzed as the following corollary shows.

Lemma *2:* The MSE of $\hat{\upsilon}(k)$ is upper bounded by

$$\text{MSE}\{\hat{\upsilon}(k)\} \le \left(\frac{1}{4} + \frac{1}{\pi^2}\right)\frac{\gamma}{k} + \frac{\upsilon^2}{\pi^2 k} \quad (48)$$

where $\gamma = \frac{(\alpha^2\upsilon+1)\sigma_n^2}{NP_S}$.

Proof: By definition, the MSE of $\hat{\upsilon}(k)$ is given by

$$\mathfrak{E}\{|\hat{\upsilon}(k) - \upsilon|^2\} = \mathfrak{E}\{\hat{\upsilon}(k)^2\} - 2\upsilon\mathfrak{E}\{\hat{\upsilon}(k)\} + \upsilon^2, \quad (49)$$

where the expectation is taken with respect to all the $a(l)$, $b(l)$ and noise vectors. The first term of above equation yields

$$\mathfrak{E}\{\hat{\upsilon}(k)^2\} = \frac{1}{4k^2}\mathfrak{E}\left\{\left(\sum_{l=1}^{k}\hat{a}(l)\right)^2 + \frac{4}{\pi^2}\left(\sum_{l=1}^{k}\hat{b}(l)\right)^2 + \frac{4}{\pi}\left(\sum_{l=1}^{k}\hat{a}(l)\right)\left(\sum_{l=1}^{k}\hat{b}(l)\right)\right\} \quad (50)$$

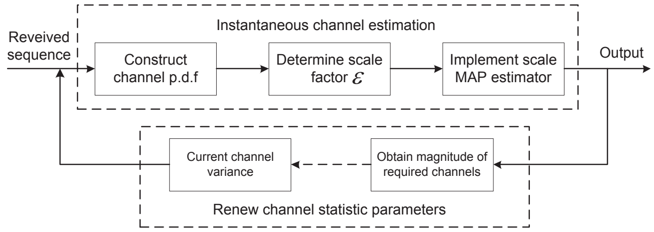

Fig. 2. The long term estimation strategy

TABLE IV
LONG TERM CHANNEL ESTIMATION STRATEGY

Begin(In the k-th block)

- **Step 1:** Construct the apriori p.d.f in terms of $\hat{\upsilon}(k-1)$.
- **Step 2:** Determine the value of ε by $\varepsilon = \varepsilon(k)$.
- **Step 3:** Estimate composite channel coefficients utilizing SMAP estimation.
- **Step 4:** Renew $\hat{\upsilon}(k)$ in accordance with (58).

End

with

$$\mathfrak{E}\left\{\left(\sum_{l=1}^{k}\hat{a}(l)\right)^2\right\} \approx k\left(2\upsilon^2 + \underbrace{\frac{(\alpha^2\upsilon+1)\sigma_n^2}{NP_S}}_{\gamma}\right) + \left(k^2-k\right)\upsilon^2, \quad (51)$$

$$\mathfrak{E}\left\{\left(\sum_{l=1}^{k}\hat{b}(l)\right)^2\right\} \approx k\left(\upsilon^2+\gamma\right) + \left(k^2-k\right)\frac{\pi^2}{4}\upsilon^2, \quad (52)$$

$$\mathfrak{E}\left\{\left(\sum_{l=1}^{k}\hat{a}(l)\right)\left(\sum_{l=1}^{k}\hat{b}(l)\right)\right\} \lessapprox \frac{\pi k^2}{2}\upsilon^2. \quad (53)$$

Recalling $\mathfrak{E}\{\hat{\upsilon}(k)\} = \upsilon$, we can finally obtain

$$\mathfrak{E}\{|\hat{\upsilon}(k)-\upsilon|^2\} \leq \left(\frac{1}{4}+\frac{1}{\pi^2}\right)\frac{\gamma}{k} + \frac{\upsilon^2}{\pi^2 k} \quad (54)$$

as the lemma shows. ■

Remark:

- As (54) shows, the MSE of $\hat{\upsilon}(k)$ is related to its true value υ and the noise term γ. When $\gamma \to 0$, the MSE converges to $\frac{\upsilon^2}{\pi^2 k}$ which is treated as the innate error of the estimation. However, the MSE converges to zero when k is sufficiently large which indicates the estimation is adequate.
- While for individual estimation, the estimate of υ can be similarly obtained by

$$\hat{\upsilon}'(k) = \frac{1}{2k}\sum_{l=1}^{k}\left[\hat{u}_1^2(l) + \hat{u}_2^2(l)\right],$$

where $\hat{u}_1(k)$ and $\hat{u}_2(k)$ denote the estimates of $|h_1|$ and $|h_2|$, respectively. It is necessary to mention that $\hat{\upsilon}'(k)$ would endure more severe errors than $\hat{\upsilon}(k)$. As a consequence, it is recommended to utilize the solution as (46) shows.

Since the MSE of $\hat{\upsilon}_k$ is a decreasing function with respect to k, for small value of k, the estimate $\hat{\upsilon}(k)$ may endure severe estimation error which would lead to imprecise CDI. As a result, the instantaneous estimation accuracy would be negatively affected when MAP estimation is adopted.

2) Scale MAP Estimator: Motivated to maintain the improvement of MAP estimation and prevent the negative affection caused by estimated errors of CDI, we develop the scale MAP (SMAP) estimator as

$$\hat{\mathbf{\Theta}} = \arg\max_{\mathbf{\Theta}} p(\mathbf{x}|\mathbf{\Theta})\, p^{\varepsilon}(\mathbf{\Theta}), \quad (55)$$

where ε is a given factor. SMAP estimation coincides with ML solution at $\varepsilon = 0$ whereas it attains the conventional MAP estimation at $\varepsilon = 1$. Intuitively, when the estimated $p(\theta)$ is accuracy, ε is recommended to be a large value in order to take the advantage of apriori information. When the obtained apriori p.d.f endures estimated errors, the value of ε should be decreased, and to the extreme, the ML estimation is implemented instead of MAP estimation. Thus it is possible to achieve the optimal estimates by selecting an optimal ε. In general, ε is recommended to be zero or small value in the beginning transmission period with small k since the estimate of υ endures severe error. With transmission going on, ε should be increased due to the declination of estimated error. When k is large enough, the estimate error of ε can be ignored so that ε is set by unit value to implement the conventional MAP estimation. As a result, the scaling factor function is apparently a non-decrease function with respect to k. Due to the difficulty in obtaining the optimal design, we present a simple linear approach to evaluate the effectiveness of SMAP estimation. Two thresholds k_{min} and k_{max} are introduced where $k < k_{min}$ indicates that the estimate of $\hat{\upsilon}(k)$ is not adequate, and $k \geq k_{max}$ means the estimate of variance converges to its true value. Then we have

$$\varepsilon(k) = \begin{cases} 0, k < k_{min}, \\ 1, k \geq k_{max}, \end{cases} \quad (56)$$

and when $k_{min} \leq k < k_{max}$, $\varepsilon(k)$ is given by

$$\varepsilon(k) = \frac{k-k_{min}}{k_{max}-k_{min}}. \quad (57)$$

3) Proposed Channel Estimation Strategy: The long term channel estimation strategy with implementing SMAP estimation is shown in Fig.2. Focusing on the k-th block, after estimating $\hat{a}(k)$ and $\hat{b}(k)$, $\hat{\upsilon}(k)$ is renewed by

$$\hat{\upsilon}(k) = \frac{2(k-1)\cdot\hat{\upsilon}(k-1) + \hat{a}(k) + \frac{2}{\pi}\hat{b}(k)}{2k}, \quad (58)$$

where $\hat{\upsilon}(0)$ is set to be zero. Then $\hat{\upsilon}(k)$ can be utilized to construct the apriori p.d.f for instantaneous channel estimation in the $(k+1)$-th block. The estimation strategy is shown in Table IV. The estimation process of SMAP estimation is similar to that of conventional MAP estimation, thus we omit its presentation.

C. Extending to A General Scenario

In this part, we consider a more general scenario where M transmitters transmit training sequence to one receiver simultaneously. Let θ_m denote the composite channel coefficient from the m-th transmitter to the receiver, and $\mathbf{t}_m$ denote the

N length training sequence sent by the m-th transmitter. The sequence observation vector $\mathbf{x}$ is

$$\mathbf{x} = \sum_{m=1}^{M} \theta_m \mathbf{t}_m + \mathbf{w}_E, \tag{59}$$

$\mathbf{w}_E$ is the equivalent noise vector. Since each transmitter-receiver link is not restricted to a single hop, $\mathbf{w}_E$ would endure a complex expression which is relevant to instantaneous channel coefficients. In that case, the noise variance σ_E^2 can not be previously known, thus the iterative algorithm presented in Table I can be applied.

Actually, if the joint p.d.f of $\mathbf{\Theta} = \{\theta_1, \ldots, \theta_M\}$ is known apriori to the receiver, the MAP estimation algorithm can be directly applied. However, in a highly complex network, the channel p.d.f may not be available to the receiver but the channel mean and variance can be estimated. To take advantage of the channel statistical information on the condition of unknown channel distribution, we utilize the Gaussian distribution to attain a effective solution. Let μ_m and υ_m denote the mean and variance of θ_m, and all the M parameters are treated as independent with each other. In accordance with the definition of MAP estimator, and after some manipulation, the estimation yields

$$\hat{\mathbf{\Theta}} = \arg\min_{\mathbf{\Theta}} \left\{ \frac{\|\mathbf{x} - \sum_{m=1}^{M} \theta_m \mathbf{t}_m\|^2}{\sigma_E^2} + \sum_{m=1}^{M} \frac{\|\theta_m - \mu_m\|^2}{\upsilon_m} \right\}. \tag{60}$$

Thus the estimate of θ_m is obtained by

$$\hat{\theta}_m = \frac{\frac{\mathbf{t}_m^H \mathbf{w}_E}{NP_S} \upsilon_m + \frac{\sigma_E^2}{NP_S} \mu_m}{\frac{\sigma_E^2}{NP_S} + \upsilon_m}. \tag{61}$$

By straightforward calculation, the average MSE of the MAP estimation is computed by

$$\delta_{MAP} = \frac{\sigma_E^2}{NP_S} \frac{\upsilon_m}{\upsilon_m + \frac{\sigma_E^2}{NP_S}}, \tag{62}$$

while that of ML estimation is $\delta_{ML} = \frac{\sigma_E^2}{NP_S}$. The inequality $\delta_{MAP} < \delta_{ML}$ is always holden which implies that the estimation can benefit from the acquired channel statistical information though the specific distribution is unknown.

V. Performance Evaluation

In this section, the Monte Carlo simulations are implemented to numerically examine the performance of the presented MAP based channel estimation schemes. Since the MAP estimator belongs to the class of extremum estimators, the famous ML estimator is chosen for comparison. In particular, the instantaneous channel h_1 and h_2 are generated by two independent complex Gaussian random variables with zero mean and $\upsilon_1 = \upsilon_2 = \upsilon$ variance. The typical value of υ is set to be $\upsilon = 1$. The transmitting power of source nodes and relay are set as equal, and the Gaussian noise variance at relay and source are assumed to be unit value. Hence, the signal-to-noise ratio (SNR) is defined as $\frac{P_S}{\sigma_n^2} = P_S$. The orthogonal training, i.e., $\mathbf{t}_1^H \mathbf{t}_2 = 0$ is utilized in the simulation. To evaluate the channel estimation accuracy, the average mean square error (MSE) are computed. The iterative time in the MAP estimation algorithm is set to be 5 in all the simulations. Totally 10^4 Monte-Carlo runs are adopted for average.

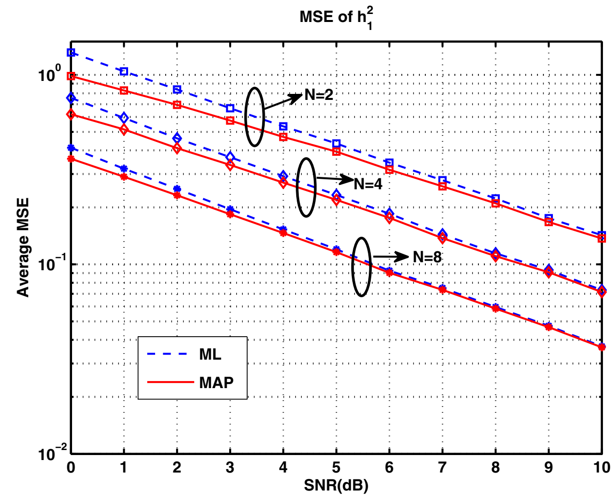

Fig. 3. Average MSEs of h_1^2 versus SNR for ML and MAP estimations.

A. On Composite Channel Estimation

First, we focus on the composite channel estimation where the average MSE of h_1^2 and $h_1 h_2$ are evaluated. The average MSE results for h_1^2 and $h_1 h_2$ are shown in Fig.3 and Fig.4, respectively. It is observed that the MAP estimation attains lower MSE than that of ML estimation during all SNR region. Compared the two figures, the MSE gain of $h_1 h_2$ is much significant than that of h_1^2, e.g., the performance gain of h_1^2 is less than 1dB while that of $h_1 h_2$ which is more than 2dB on the condition of $SNR \leq 5$dB for $N = 2$. However, the corresponding gain attained by MAP method is reduced with increasing SNR. Comparing the MSE curves of different N, it is seen that the performance gain becomes less with larger N, e.g, MAP estimation obtains about 2dB gain of $h_1 h_2$ for $N = 2$ and the gain decreases to less than 1dB for $N = 4$. Hence, it indicates that the MAP estimation achieve significant performance gain over low $\frac{NP_S}{\sigma_n^2}$ region compared with ML method, and the two methods perform almost the same over high $\frac{NP_S}{\sigma_n^2}$ region.

Then we compare the proposal with the variance squared maximum likelihood (VSML) estimator in [12], and the sample-average estimator (SAE) presented in [13] in terms of the average MSE of h_1^2 which is shown in Fig.5. For VSML and SAE methods, 8-PSK modulation is utilized for each training symbol. Focusing on the three curves of the same N, it can be observed that the proposed MAP algorithm can achieve the lowest MSE of h_1^2 among the three methods which demonstrates the advantage of our proposal.

Moreover, to show how the variance affects the performance of MAP algorithm, we simulate the MSEs of h_1^2 with different channel variance υ which is shown in Fig.6. The training sequence length is set to be $N = 2$. In accordance with the curves, it can be seen that the performance improvement of MAP algorithm differs with different channel variances, and

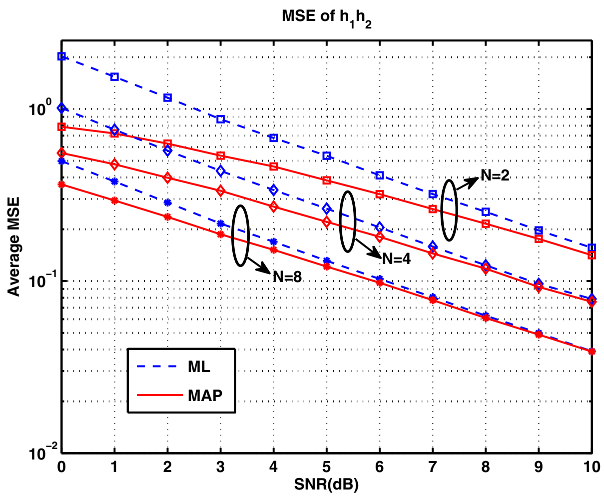

Fig. 4. Average MSEs of h_1h_2 versus SNR for ML and MAP estimations.

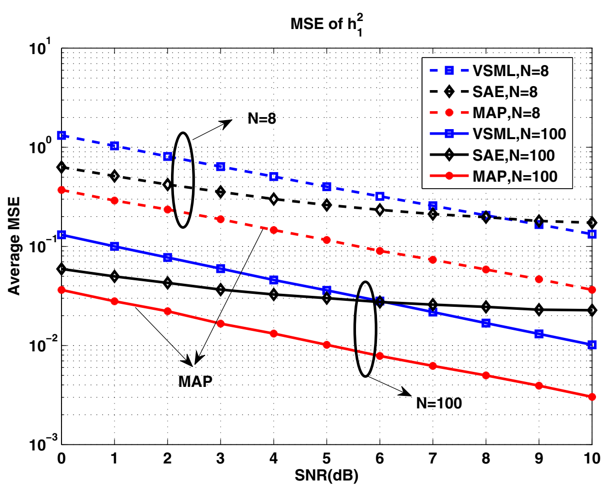

Fig. 5. MSE versus SNR of different estimation methods.

the improvement increases a lot with decreasing υ. This implies that the improvement of MAP algorithm highly depends on the statistical behavior of under estimated parameters.

B. On Individual Channel Estimation

The individual channel estimation is concerned where the average MSEs for $|h_1|$ and $|h_2|$ are evaluated. Both the ML estimation and the composite MAP solution are taken for comparison. For composite MAP approach, the composite channels are estimated first, then the individual amplitudes are calculated by $|\hat{h_1}| = \sqrt{\hat{a}}$ and $|\hat{h_2}| = \frac{\hat{b}}{\sqrt{\hat{a}}}$. As the curves in Fig.7 and Fig.8 shows, the individual MAP estimation attains lower MSE than the two compared methods. Comparing the MSE between $|h_1|$ and $|h_2|$, it is seen that the MSE of $|h_2|$ is much higher than that of $|h_1|$. This is because the individual estimation of $|h_2|$ would have $|h_1|$ in the denominator. Minute error of $|h_1|$ may cause negative impact on the estimation accuracy of $|h_2|$ in a large scale. Besides, the MSEs of individual channels are lower than that calculated by composite channels because extra calculation between estimates would enlarge the estimated errors.

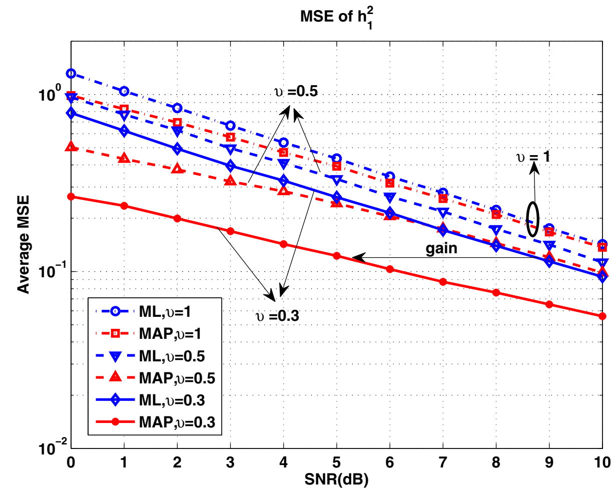

Fig. 6. MSE versus SNR of different channel variance.

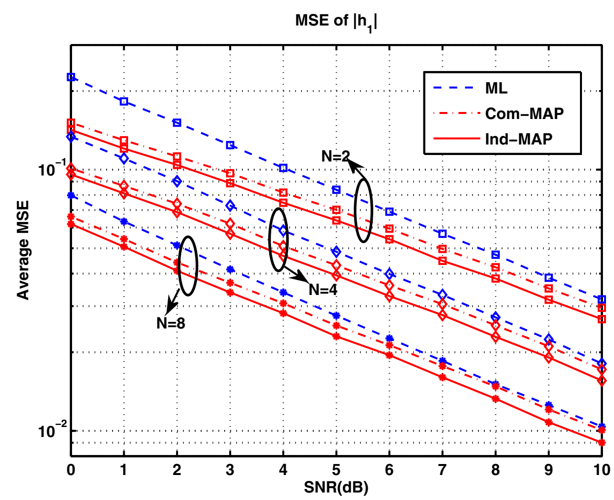

Fig. 7. Average MSEs of $|h_1|$ versus SNR for ML and MAP estimations.

We also investigate the individual estimation at $\mathbb{S}_2$ which is shown in Fig.9 where the MSEs of $|h_1|$ and $|h_2|$ versus SNR are plotted of different sequence length. It can be observed that the MSE of $|h_2|$ is much less than that of $|h_1|$ which is different from the estimation results in $\mathbb{S}_1$. Combining the results in Fig.7, Fig.8 and Fig.9, it can be concluded that the estimation for $|h_i|$ is more accuracy than $|h_{-i}|$ at $\mathbb{S}_i$ where h_{-i} denotes the opposite channel of h_i. This is because the observation received at $\mathbb{S}_i$ contains more information about h_i which results in the performance difference.

C. With Unknown Noise Variance

Next, we evaluate the performance of the iterative MAP estimation algorithm presented in Section IV. The iterative time is set to be 5. Restricted to the paper length, the average MSEs of composite channel coefficients h_1^2 is shown in Fig.10. As the curves show, the MAP estimation achieve lower average MSE than conventional LS estimation. Thus it indicates that the apriori CDI can improve estimation accuracy even though the utilized apriori p.d.f is constructed with calculated noise variance. Focusing on the curve of $N = 2$, we can see that the

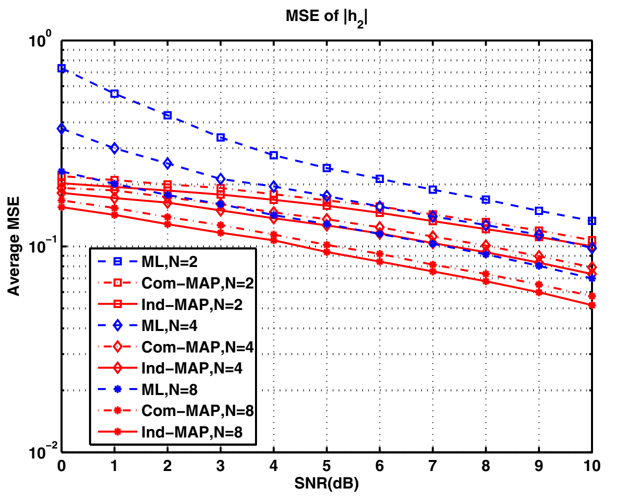

Fig. 8. Average MSEs of $|h_2|$ versus SNR for ML and MAP estimations.

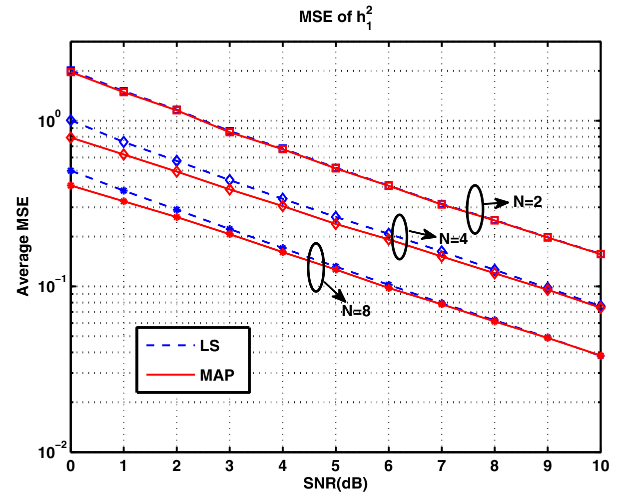

Fig. 10. Average MSEs of h_1^2 versus SNR for LS and MAP estimations with unknown noise variance.

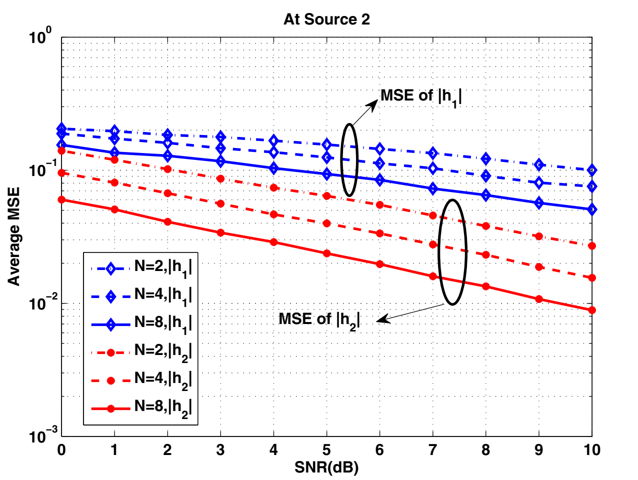

Fig. 9. MSE versus SNR of different sequence length at source 2

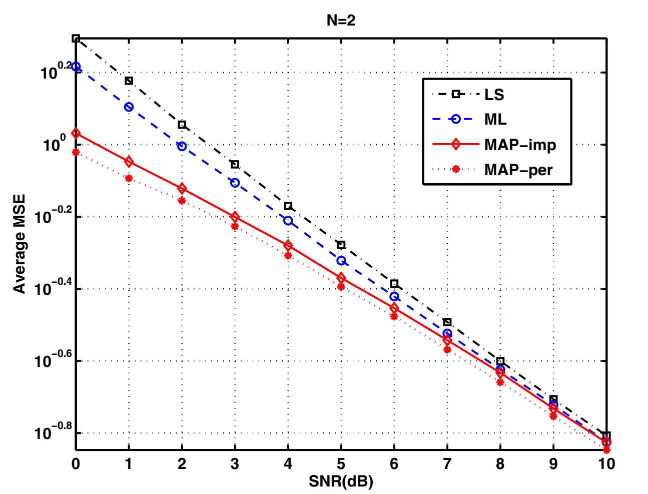

Fig. 11. Average MSE summation of h_1^2 and $h_1 h_2$ versus SNR.

MSE performance of the two schemes are almost the same. This because the calculated noise variance may endure severe error with small N which would degenerate the performance gain of MAP estimation. Meanwhile, it can be observed that the MSE gain of $N=4$ is more significant than that of $N=8$ for the reason that the MAP estimation always performs better with low training consumption

D. With Imperfect CDI

After that, we evaluate the performance of the presented long term estimation strategy in terms of the normalized summation of average MSE of h_1^2 and $h_1 h_2$, i.e, $\frac{1}{2}\left[\mathrm{MSE}\left(h_1^2\right)+\mathrm{MSE}\left(h_1 h_2\right)\right]$. We simulated 10000 blocks of transmission, and the factor ε in the k-th block is set to be $\varepsilon\left(k\right)=\frac{k}{10000}$. The training sequence length is set as $N=2$. The curve in Fig.11 marked *MAP-per* is assumed to obtain perfect channel statistic information whereas *MAP-imp* is the proposed estimation strategy with linear scale function. It is seen that the MAP estimation with estimated channel variance performs worse than that with perfect channel knowledge. However, the proposed estimation strategy can achieve lower MSE than the ML and LS estimation which indicates that the linear function of ε is effective in this case. In addition, it can be expected that the performance of proposed estimation strategy can be further improved by optimizing the scale function.

E. Evaluation under More Practical Channel Models

In above parts, all the evaluations are under the quasi-static flat Rayleigh fading channel model. Here, we evaluate its performance under the time varying channel following auto regressive (AR) fading model. The block size is defined to have 128 symbols, but the equivalent channel coherence interval spans only 16 symbols. Therefore, each block is divided into 8 sub-blocks where channel coefficient of each subblock satisfies the first order AR model [14], and each sub-block has 2 symbols for training. The channel coefficient in the k-th subblock follows

$$h_i\left(k\right)=\varrho h_i\left(k-1\right)+\mu_i\left(k\right)$$

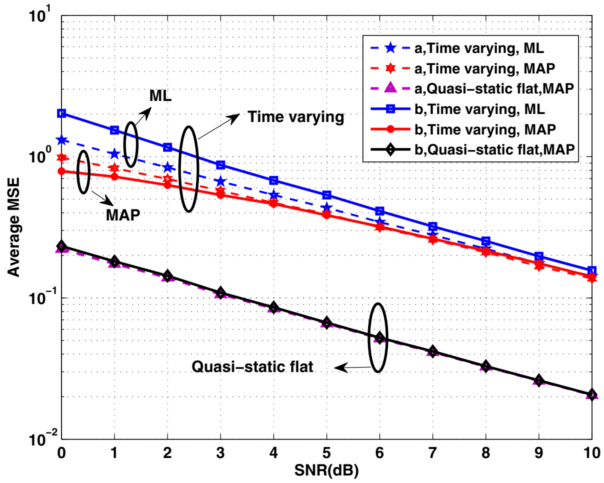

Fig. 12. MSE versus SNR of different channel models.

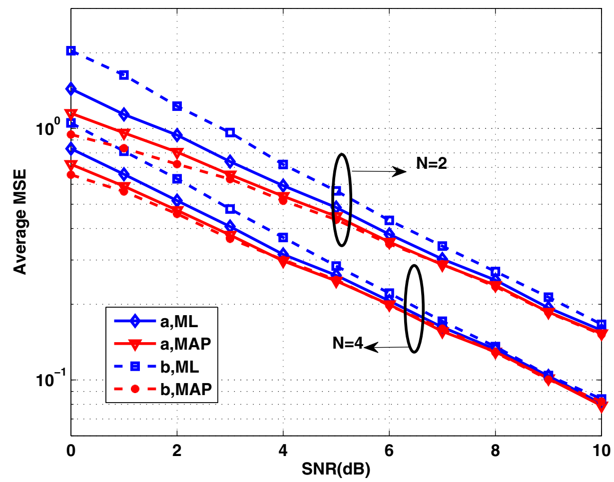

Fig. 13. Average MSE of h_1^2 and $h_1 h_2$ versus SNR under SCM scenario.

with $\mu_i(k)$ satisfies complex Gaussian distribution with zero mean and $\left(1-\varrho^2\right)\upsilon_i$ variance, and ϱ is set to be 0.9 in the simulation. For comparison, the first 16 symbols of every 128 symbols are used for training in block fading channel. In accordance with the curves in Fig.12, we can see that the corresponding MSE of time varying channel is much higher than that of flat fading channel. This can be expected since the accumulative training SNR for each channel estimation has been reduced which results in the performance degradation. However, the proposed MAP method can also achieve significant gain compared with the ML solution which indicates that the proposed MAP algorithm is effective in time varying channels.

To investigate the effectiveness of the proposed MAP estimation algorithm to practical channel model, we evaluate its performance under a more practical channel model, i.e., the spatial channel model (SCM) presented in the 3rd Generation Partnership Project (3GPP) [15]. We choose the third scenario with the following main parameters.

- Power delay profile: Pedestrian B.
- Number of paths: 6.
- Relative path power of each path: $\{0.0, -0.9, -4.9, -8.0, -7.8, -23.9\}$(dB).
- Delay of each path: $\{0, 200, 800, 1200, 2300, 3700\}$(ns).
- Mobility speed of source: $30km/h$.
- Direction of travel: -22.5 degree.
- Angle of arrival: 22.5 degree (odd numbered paths), -67.5 degree (even numbered paths).

In Fig.13, the MSE of h_1^2 and h_1h_2 with ML estimation and MAP solution are plotted with different sequence length. It can be observed that the MAP method performs better than ML one which implies the proposal is also effective under practical channel scenario.

VI. Conclusions

The training based channel estimation for TWRC is studied in this paper. To make use of the apriori CDI of channel coefficients, the MAP based estimation schemes are presented to perform effective estimation including the estimation of composite source-source channel coefficients and that of individual source-relay channels which needs the knowledge of corresponding channel p.d.f. Considering that CDI is also needed to be estimated and is not error free, an improved channel estimation strategy along with the scale MAP estimator is developed where CDI is calculated by instantaneous channel coefficients block by block, and it can be put back to implement scale MAP estimator. Simulation results show the proposed MAP estimators can achieve lower MSE than conventional ML estimator. Meanwhile, it is also testified that the proposed MAP based estimation strategies also outperforms the ML estimation since it can achieve lower MSE.

Appendix A

The amplitude and phase of a complex variance can be treated as independent, then the joint p.d.f is given by

$$p\left(a,b,\theta_a,\theta_b\right) = p_1\left(a,b\right)p_2\left(\theta_a,\theta_b\right).$$

Since h_i satisfies complex Gaussian distribution with zero mean and variance υ_i, the p.d.f of $|h_i|$ is given as [17]

$$f_{|h_i|}\left(x\right) = \frac{2x}{\upsilon_i}e^{\frac{x^2}{\upsilon_i}}.$$

With $|h_1| = \sqrt{a}$ and $|h_2| = \frac{b}{\sqrt{a}}$, the determinant of Jacobian matrix is calculated as

$$\mathbf{J} = \frac{\partial\left(|h_1|,|h_2|\right)}{\partial\left(a,b\right)} = \frac{1}{2a}.$$

Then the joint p.d.f of a and b is obtained by

$$p_1\left(a,b\right) = \frac{1}{2a}\cdot\frac{2\sqrt{a}}{\upsilon_1}e^{-\frac{a}{\upsilon_1}}\cdot\frac{2b}{\upsilon_2\sqrt{a}}e^{-\frac{b^2}{\upsilon_2 a}} = \frac{2b}{a\upsilon_1\upsilon_2}e^{-\left(\frac{a}{\upsilon_1}+\frac{b^2}{a\upsilon_2}\right)}.$$

As for $\theta_a = \mod\left(2\theta_1, 2\pi\right)$ and $\theta_b = \mod\left(\left(\theta_1+\theta_2\right), 2\pi\right)$ where $\mod\left(y,x\right)$ denotes the modulo operation of y over x, the joint p.d.f can be calculated by $p_2\left(\theta_a,\theta_b\right) = \frac{1}{4\pi^2}$, where $\theta_1,\theta_2 \in [0, 2\pi)$. Thus $p\left(a,b,\theta_a,\theta_b\right)$ is finally derived as

$$p\left(a,b,\theta_a,\theta_b\right) = \frac{b}{2\pi^2 a\upsilon_1\upsilon_2}e^{-\left(\frac{a}{\upsilon_1}+\frac{b^2}{a\upsilon_2}\right)}.$$

APPENDIX B

Consider the average MAP estimation log likelihood metric over N training symbols of a single memoryless fading block, the estimate yields

$$\begin{aligned}\hat{\mathbf{\Theta}}_{MAP} &= \arg\max_{\mathcal{C}} \left\{ \frac{1}{N} \log \left(p(\mathcal{C}) \prod_{i=1}^{N} p(\mathbf{x}(i)|\mathcal{C}) \right) \right\} \\ &= \arg\max_{\mathcal{C}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \log p(\mathbf{x}(i)|\mathcal{C}) + \frac{1}{N} \log p(\mathcal{C}) \right\},\end{aligned}$$

where $\mathbf{x}(i)$ denotes the i-th channel measurement of observation vector. $p(\mathbf{x}(i)|\mathcal{C})$ is the likelihood function of channel observation for the i-th training symbol assuming channel coefficient vector $\mathcal{C}$.

Choose $N = K\lceil|\log p(\mathcal{C})|\rceil$, for $p(\mathcal{C}) > 0, \forall \mathcal{C}$, where $K \to \infty$ and $\lceil\cdot\rceil$ is the ceiling operation, then

$$\begin{aligned}&\lim_{N\to\infty} \hat{\mathbf{\Theta}}_{MAP} \\ &= \lim_{N\to\infty} \arg\max_{\mathcal{C}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \log p(\mathbf{x}(i)|\mathcal{C}) + \frac{1}{N} \log p(\mathcal{C}) \right\} \\ &= \arg\max_{\mathcal{C}} \left\{ \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \log p(\mathbf{x}(i)|\mathcal{C}) + \lim_{N\to\infty} \frac{1}{N} \log p(\mathcal{C}) \right\} \\ &= \arg\max_{\mathcal{C}} \int_{\mathbf{x}} \log p(\mathbf{x}|\mathcal{C})\, p(\mathbf{x}|\mathbf{\Theta})\, d\mathbf{x} = \lim_{N\to\infty} \hat{\mathbf{\Theta}}_{ML},\end{aligned}$$

where $p(\mathbf{x}|\mathbf{\Theta})$ is the empirical distribution of channel observation which has converged to its theoretical counterpart when $N \to \infty$, namely the distribution of channel observation conditioned on channel coefficient $\mathbf{\Theta}$.

It is straightforward to show that it is an optimization problem with $\int_{\mathbf{x}} p(\mathbf{x}|\mathcal{C})\, d\mathbf{x} = 1$. The Kullback-Leibler divergence formula

$$-D(p(\mathbf{x}|\mathbf{\Theta})\,\|p(\mathbf{x}|\mathcal{C})) = \int_{\mathbf{x}} p(\mathbf{x}|\mathbf{\Theta}) \log \frac{p(\mathbf{x}|\mathcal{C})}{p(\mathbf{x}|\mathbf{\Theta})} \le 0$$

leads to

$$\int_{\mathbf{x}} (\mathbf{x}|\mathbf{\Theta}) \log p(\mathbf{x}|\mathcal{C})\, d\mathbf{x} \le \int_{\mathbf{x}} (\mathbf{x}|\mathbf{\Theta}) \log p(\mathbf{x}|\mathbf{\Theta})\, d\mathbf{x}.$$

Thus both MAP and ML estimation are related to minimum Kullback-Leibler divergence problem, where optimal solution is achieved when $D(p(\mathbf{x}|\mathbf{\Theta})\,\|p(\mathbf{x}|\mathcal{C})) = 0$, e.g., $p(\mathbf{x}|\mathcal{C}) = \kappa p(\mathbf{x}|\mathbf{\Theta}), \forall \mathbf{x}$, where κ is a constant. Since $p(\mathbf{x}|\mathcal{C})$ and $p(\mathbf{x}|\mathbf{\Theta})$ are both p.d.f, $\kappa = 1$, $\mathcal{C} = \mathbf{\Theta}$ is the unique optimal solution.

REFERENCES

[1] M. Peng, C. Yang, Z. Zhao, W. Wang, and H. Chen, "Cooperative network coding in relay-based IMT-advanced systems," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 76–84, Apr. 2012.

[2] S. Zhang, S. Liew, and P. Lam, "Physical layer network coding," in *Proc. 2006 ACM MobiCom*, pp. 358–365.

[3] C. Zhang, J. Ge, J. Li, and X. Shi, "Performance analysis and power allocation for a two-way amplify-and-forward relay with channel estimation errors," *IET Commun.*, vol. 6, no. 12, pp. 1846–1855, 2012.

[4] M. Biguesh and A. 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.

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

[6] F. Gao, B. Jiang, X. Gao, and X. Zhang, "Superimposed training based channel estimation for OFDM modulated amplify-and-forward relay networks," *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 2029–2039, July 2011.

[7] S. Zhang, F. Gao, and C. Pei, "Optimal training design for individual channel estimation in two-way relay networks," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4987–4991, Sept. 2012.

[8] H. Zhang, Y. Liu, and M. Tao, "Resource allocation with subcarrier pairing in OFDMA two-way relay networks," *IEEE Wireless Commun. Lett.*, vol. 1, no. 2, pp. 61–64, Apr. 2012.

[9] L. Song, "Relay selection for two-way relaying with amplify-and-forward protocols," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1954–1959, May 2011.

[10] S. Xu and Y. Hua, "Optimal design of spatial source-and-relay matrices for a non-regenerative two-way MIMO relay system," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1645–1655, May 2011.

[11] P. Lioliou, M. Viberg, and M. Matthaiou, "Bayesian approach to channel estimation for AF MIMO relaying systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1440–1451, Sept. 2012.

[12] S. Abdallah and I. Psaromiligkos, "Blind channel estimation for MPSK-based amplify-and-forward two-way relaying," in *Proc. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2828–2831.

[13] Q. Zhao, Z. Zhou, and B. Vucetic, "Low complexity semi-blind channel estimation algorithm in two-way relay networks," in *Proc. 2011 IEEE Personal Indoor and Mobile Radio Communications*, pp. 1443–1447.

[14] S. Savazzi and U. Spagnolini, "Optimizing training lengths and training intervals in time-varying fading channels," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1098–1112, 2009.

[15] "Spatial channel model for multiple input multiple output simulations," 3GPP TR 25.996, release 6, Sept. 2003.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory.*

[17] A. Goldsmith. *Wirelss Communications*, Stanford University Press.

Xinqian Xie received the B.S. degree in telecommunication engineering from Beijing University of Posts and Communications (BUPT), China, in 2010. He is currently pursuing the Ph.D. degree in Wireless Signal Processing and Networks Lab at BUPT. His research interests include cooperative communications, estimation and detection theory.

Mugen Peng (M'05-SM'11) is currently a Professor with the Key Laboratory of Universal Wireless Communication (Ministry of Education) at Beijing University of Posts and Telecommunications (BUPT), China. He received his Ph.D. degree in Communication and Information Engineering from BUPT in 2005. His research interests are in the areas of cooperative communication, green communications, wireless network coding, self-organization networking, and interference modeling in wireless communications. He has published over 100 papers in refereed journals and conference proceedings, edited eight books and translated four books, contributed more than 30 proposals to 3GPP and CCSA standards, and has been granted more than 20 patents in China. He is leading several projects funded by NSFC, China MOST, and industries.

Bin Zhao received the B.S.E.E. and M.S.E.E. degrees from Shanghai Jiaotong University, Shanghai, China, in 1995 and 1998 respectively. He received a Ph.D. degree in electrical engineering from West Virginia University (Morgantown, WVa, U.S.A.) in 2004, where he worked as a research assistant in the Wireless Communications Research Laboratory. He is currently a senior engineer with Cambridge Silicon Radio. His research interests are in the areas of communications theory, error-correction coding, sensor networking, and information theory. He was a DSP Engineer with Huawei Technologies Co. Ltd. from Sep. 1997 to July 1999, a Communications Engineer with Efficient Channel Coding Inc. (now ViaSat Inc.) from July 2004 to Dec. 2005, and a Posdoc Researcher with University of California, Riverside, Department of Electrical Engineering from Dec 2005 to May 2007.

Wenbo Wang is currently the dean of Telecommunication Engineering at Beijing University of Posts and Telecommunications (BUPT). He received the BS degree, the MS and Ph.D. Degrees from BUPT in 1986, 1989 and 1992 respectively. Now he is the Assistant Director of academic committee of Key Laboratory of Universal Wireless Communication (Ministry of Education) in BUPT. His research interests include radio transmission technology, Wireless network theory, Broadband wireless access and Software radio technology. Prof. Wenbo Wang has published more than 200 journal and international conference papers and holds 12 patents and has published six books.

Yingbo Hua (S'86-M'88-SM'92-F'02) received a B.S. degree (1982) from Southeast University, Nanjing, China, a M.S. degree (1983) and a Ph.D. degree (1988) from Syracuse University, Syracuse, NY. He was a Lecturer (1990-1992), a Senior Lecturer (1993-1995), and a Reader and Associate Professor (1996-2000) with the University of Melbourne, Australia. He was a Visiting Faculty Member with Hong Kong University of Science and Technology (1999-2000), and a Consultant with Microsoft Research, WA (summer 2000). Since 2001, he has been with the University of California at Riverside, where he is a Senior Full Professor.

He has served as Editor, Guest Editor, Member of Editorial Board and/or Member of Steering Committee for IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, *EURASIP Signal Processing*, *IEEE Signal Processing Magazine*, IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, and IEEE WIRELESS COMMUNICATION LETTERS. He has been a Member of IEEE Signal Processing Society's Technical Committees for Underwater Acoustic Signal Processing, Sensor Array and Multichannel Signal Processing, and Signal Processing for Communication and Networking. He has served as member of Technical and/or Advisory Committees for over 40 international conferences and workshops. He has authored near three hundreds of articles and coedited three volumes of books, with more than six thousands of citations, in the fields of Sensing, Signal Processing and Communications. He is a Fellow of IEEE and AAAS.