

Timing Estimation and Resynchronization for Amplify-and-Forward Communication Systems

Xiao Li, Chengwen Xing, Yik-Chung Wu, and S. C. Chan, *Member, IEEE*

Abstract—This paper proposes a general framework to effectively estimate the unknown timing and channel parameters, as well as design efficient timing resynchronization algorithms for asynchronous amplify-and-forward (AF) cooperative communication systems. In order to obtain reliable timing and channel parameters, a least squares (LS) estimator is proposed for initial estimation and an iterative maximum-likelihood (ML) estimator is derived to refine the LS estimates. Furthermore, a timing and channel uncertainty analysis based on the Cramér–Rao bounds (CRB) is presented to provide insights into the system uncertainties resulted from estimation. Using the parameter estimates and uncertainty information in our analysis, timing resynchronization algorithms that are robust to estimation errors are designed jointly at the relays and the destination. The proposed framework is developed for different AF systems with varying degrees of timing misalignment and channel uncertainties and is numerically shown to provide excellent performances that approach the synchronized case with perfect channel information.

Index Terms—Amplify-and-forward (AF), asynchronous, channel estimation, Cramér–Rao bound (CRB), relay, resynchronization.

I. INTRODUCTION

COOPERATIVE distributed MIMO systems, which suggest the sharing of antennas among several single-antenna terminals to cooperatively transmit data [1], have been advocated by many researchers because of their great potential in achieving comparable link reliability and system capacity to traditional multiple antenna systems [2]–[5].

Nevertheless, this type of system also presents many practical difficulties in system design and implementations due to its distributed nature, the most important of which is timing synchronization. It has been analytically and numerically shown in [6]–[8] that timing asynchronism brings considerable performance degradation to such systems. Furthermore, diversity gain [9] and system capacity [10] may also be lost as a result of the intersymbol interference (ISI) caused by imperfect synchronization. Therefore, appropriate mechanisms that deal with the

asynchronous reception in distributed systems become especially essential.

As countermeasures, some delay-robust transmission and coding schemes at the transmitter side have been proposed [11]–[16] to bypass the issue. However, these schemes impose restrictions on transmission and reception protocols and hence reduce the flexibility of cooperation strategies. On the other hand, in order for many existing cooperation schemes to work efficiently, it is necessary to achieve perfect timing synchronization, including (but not limited to) cooperative relays [7], [8], distributed space–time coding [17], [18], cooperative eigen-coding scheme and distributed unitary space–time modulation (USTM) [21]. Thus, instead of employing a special scheme at the transmitter side to combat asynchronism, a general algorithm that can effectively resynchronize the timing misalignment is of tremendous value to mitigate the ISI at the destination receiver side.

As a first investigation on designing effective algorithms to resynchronize the received signal at the destination, [22] proposed a general framework to design estimation and timing resynchronization algorithms in decode-and-forward (DF) relay systems. However, for amplify-and-forward (AF) systems, most of the existing works assume perfect timing synchronization [17]–[19], [23]–[28] and the timing resynchronization problem remains widely untouched because unlike DF systems, the timing resynchronization algorithm design for AF systems is closely related to the processing at relays, which varies with transmission schemes.

This paper develops a general framework for the estimation of timing and channels, as well as the design of joint timing resynchronization algorithms at the relays and the destination for dual-hop cooperative AF relay systems. The contributions of this paper are summarized as follows. First, as a computationally efficient method, a least squares (LS) solution is derived to perform initial timing and channel estimation. Then, an iterative maximum-likelihood (ML) estimator is proposed to refine the LS estimates. Secondly, we present a Cramér–Rao bound (CRB) analysis, which provides insights into the system uncertainties obtained from estimation. Third, based on the uncertainty analysis, robust timing resynchronization algorithms are designed jointly at relays and the destination to minimize the recovered data mean-squared error (MSE) averaged over timing and channel uncertainties. Simulation results show that the symbol error rate (SER) performance of systems using the proposed algorithms approaches that of the ideal case when the uncertainties and timing misalignment are relatively mild.

The rest of this paper is organized as follows. In Section II, the system model for the considered relay system is presented. The joint timing and channel estimation problem is investigated in Section III, followed by a thorough CRB analysis. The design of

Manuscript received July 20, 2009; accepted December 10, 2009. First published December 31, 2009; current version published March 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Roberto Lopez-Valcarce. This work was supported by Grant GRF HKU 7181/07E.

X. Li is with the Department of Electrical and Computer Engineering, University of California, Davis CA 95616-5294 USA (e-mail: eceli@ucdavis.edu).

C. Xing, Y.-C. Wu, and S.C. Chan are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail : cwxing@eee.hku.hk; ycwu@eee.hku.hk; scchan@eee.hku.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2009.2039837

joint relay and destination timing resynchronization algorithms is discussed in Section IV. Section V provides numerical results to validate the proposed estimation and resynchronization algorithms. Finally, the paper is concluded in Section VI.

Notation: $\Re(\cdot)$ and $\Im(\cdot)$ take the real and imaginary parts of a complex quantity. The operation $\text{tr}\{\mathbf{A}\}$ takes the trace of matrix \mathbf{A} and the notation $\mathbf{A}^{1/2}$ is the square root of matrix \mathbf{A} by Cholesky decomposition. The operator $\text{diag}(\mathbf{x})$ denotes a diagonal matrix with the elements of \mathbf{x} located along the main diagonal, while $\text{diag}[\mathbf{A}, \mathbf{B}, \dots]$ represents a block diagonal matrix with $[\mathbf{A}, \mathbf{B}, \dots]$ being the diagonal matrix elements. Superscripts $(\cdot)^*$, $(\cdot)^H$ and $(\cdot)^T$ denote the conjugate, the conjugate transpose and the transpose operators respectively while \mathbf{I}_K indicates a $K \times K$ identity matrix. Notation $\|\mathbf{x}\|$ ($\|\mathbf{x}\|_{\mathbf{W}}$) represent the L_2 norm (the weighted L_2 norm with \mathbf{W} being the weighting matrix) of vector \mathbf{x} and $\mathbb{E}_{\theta}\{\cdot\}$ assumes the expectation with respect to variable θ . Finally, \odot and \otimes stand for the Hadamard and Kronecker product, respectively.

II. SYSTEM MODEL

In this paper, an AF cooperative communication system is considered, as shown in Fig. 1. The system consists of a source \mathcal{S} , a destination \mathcal{D} and K relays \mathcal{R}_k scattered in the middle, which are all equipped with single antennas. The propagation channels are assumed to be quasi-static and flat fading [2]. Practically speaking, the receiver at the destination \mathcal{D} has no timing and channel information before transmission. Thus, a training is used to obtain these parameter estimates [19] prior to the transmission of data, and the transmission contains the following two periods.

- **Training period:** The source \mathcal{S} transmits a training sequence to the K relays \mathcal{R}_k . At each relay, a superimposed training is added onto the original training after some relay processing, such as distributed space-time coding [17] or a linear precoding [19]. At the destination \mathcal{D} , after sampling, a joint estimation of the multiple timing offsets and channels is performed. With the timing and channel estimates, the design of joint relay and destination timing resynchronization algorithms is carried out at the destination \mathcal{D} .

- **Data transmission period:** After the training period, the source \mathcal{S} transmits a data sequence to the K relays \mathcal{R}_k . After encoding the incoming message using a specific coding scheme, each relay \mathcal{R}_k preprocesses the signal using the designed relay resynchronization algorithm and forwards the incoming message to the destination \mathcal{D} . At \mathcal{D} , the incoming signal is further resynchronized with the proposed algorithm to mitigate the distortions brought by the asynchronism. Finally, the signal is demodulated and decoded.

A. Received Signals and Processing at Relays

In the first hop, each relay \mathcal{R}_k receives signals from \mathcal{S} , which is a point-to-point communication system and the synchronization is straightforward. All the conventional synchronization techniques can be used [31]–[33]. Thus, without loss of generality, it is assumed that the signal from \mathcal{S} is received without timing offsets after matched-filtering. The received signal vector \mathbf{r}_k at the k th relay \mathcal{R}_k can then be written as

$$\mathbf{r}_k = f_k \mathbf{s} + \mathbf{n}_k \quad (1)$$

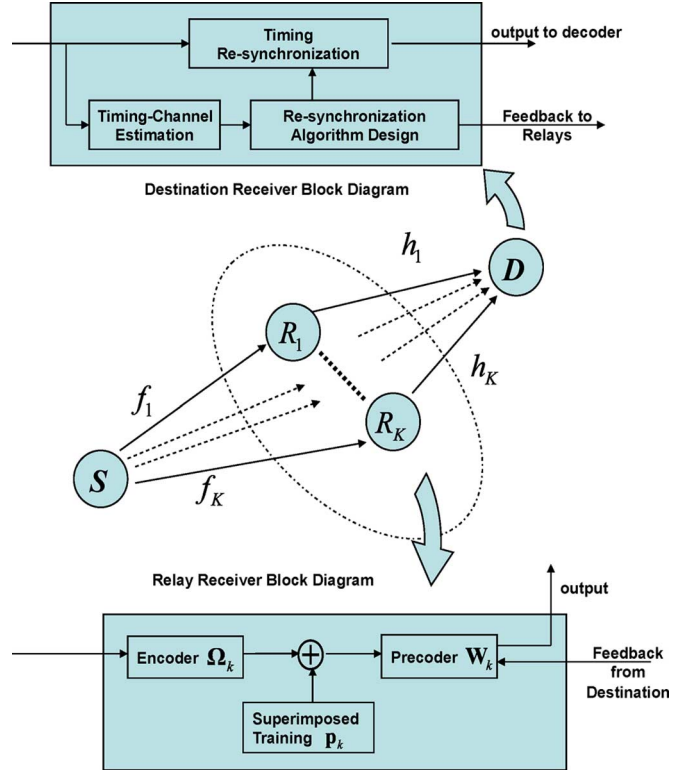


Fig. 1. A typical amplify-and-forward (AF) relay system.

where \mathbf{s} is a length- L sequence transmitted from the source \mathcal{S} . The scalar f_k is the complex channel coefficient from \mathcal{S} to \mathcal{R}_k and is assumed to be a zero mean, circular complex Gaussian random variable with unknown variance. The term \mathbf{n}_k is a vector containing circular complex Gaussian noise elements with zero mean and covariance $\mathbf{R}_{\mathbf{n}_k}$.

Upon reception, \mathcal{R}_k processes the incoming signal \mathbf{r}_k according to a specific transmission scheme. This process can be generally written as

$$\mathbf{x}_k = \mathbf{W}_k(\Omega_k \mathbf{r}_k + \mathbf{p}_k) \quad (2)$$

where the notations \mathbf{p}_k , Ω_k and \mathbf{W}_k are explained in the following.

- **Superimposed Training \mathbf{p}_k :** The sequence \mathbf{p}_k is a separate training with covariance $\mathbf{R}_{\mathbf{p}_k}$ imposed by \mathcal{R}_k , which is only used during the training period for estimation purpose. During the data transmission period, \mathbf{p}_k is set to zero and all the power is used for transmitting data \mathbf{s} from \mathcal{S} .

- **General Encoding Matrix Ω_k :** The operator Ω_k is a predefined $R \times L$ matrix that generates *distributed space-time codes* [17], [18]. Meanwhile, if distributed space-time coding is not employed during transmission, the encoding matrix is simply chosen as $\Omega_k = \mathbf{I}_L$.

- **General Precoding Matrix \mathbf{W}_k :** The matrix \mathbf{W}_k is a general $R \times R$ precoding matrix after encoding the message using Ω_k . Traditionally, for single antenna relay systems, \mathbf{W}_k is a scaling operation with $\mathbf{W}_k = w_k \mathbf{I}_L$ and it is referred to as *distributed beamforming or power allocation*. The scalar w_k is designed by normalizing the transmitted power at each relay [17]–[19] or optimizing certain system performance criteria, including (but not limited to) recovered data MSE and received

signal-to-noise ratio (SNR) [23]–[26]. In order to make our discussions more general, the notation \mathbf{W}_k is considered as non-diagonal in Section IV.

Note that different from traditional multiple antenna systems which perform precoding and encoding across different antennas, the received vector \mathbf{r}_k here contains a block of symbol received in time, and all the encoding $\mathbf{\Omega}_k$ and precoding \mathbf{W}_k are applied across time at different relays.

Then substituting (1) into (2), we can expand the signal model as

$$\mathbf{x}_k = f_k \cdot \mathbf{W}_k \mathbf{\Omega}_k \mathbf{s} + \mathbf{W}_k \mathbf{p}_k + \mathbf{W}_k \mathbf{e}_k, \quad (3)$$

where $\mathbf{e}_k \triangleq \mathbf{\Omega}_k \mathbf{n}_k$ is the equivalent noise vector with its covariance matrix calculated as

$$\mathbf{R}_{\mathbf{e}_k} = \mathbf{\Omega}_k \mathbf{R}_{\mathbf{n}_k} \mathbf{\Omega}_k^H. \quad (4)$$

B. Received Signal at Destination

Due to hardware imperfections and diverse relay locations, the signals arriving at \mathbb{D} from each relay \mathbb{R}_k are not synchronized to each other. Hence, before matched filtering, the received signal (within $0 \leq t \leq L_o T$) at the destination \mathbb{D} can be expressed as

$$y(t) = \sum_{k=1}^K h_k \sum_{i=-L_g}^{L_o+L_g+1} \mathbf{x}_k(i) g(t-iT-\epsilon_k T) + v(t), \quad (5)$$

where $\mathbf{x}_k(i)$ is the i th element in vector \mathbf{x}_k in (3) and $g(t)$ stands for the pulse shaping filter. Also, h_k is the complex channel coefficient between \mathbb{R}_k and \mathbb{D} and has the same statistics as f_k . The symbol $\epsilon_k \in [0, 1)$ is the normalized timing offset while T is the symbol duration. The last term $v(t)$ is the zero mean, circular complex Gaussian noise at \mathbb{D} . Here, the output code length R is taken to be $R = L_o + 2L_g$, with L_o representing the observation interval while L_g being the effective duration of the tail of $g(t)$ on one side.

Upon reception, the waveform $y(t)$ is sampled at \mathbb{D} by an oversampling ratio $Q \geq 2$ and thus the sample interval is $T_s = T/Q$. After putting the received samples into a vector $\mathbf{y} \triangleq [y(0), y(T_s), \dots, y((L_o Q - 1)T_s)]^T$, we have [32], [33]

$$\mathbf{y} = \sum_{k=1}^K h_k f_k \mathbf{A}_{\epsilon_k} \mathbf{W}_k \mathbf{\Omega}_k \mathbf{s} + \sum_{k=1}^K h_k \mathbf{A}_{\epsilon_k} \mathbf{W}_k \mathbf{p}_k + \sum_{k=1}^K h_k \mathbf{A}_{\epsilon_k} \mathbf{W}_k \mathbf{e}_k + \mathbf{v} \quad (6)$$

where

$$\begin{aligned} \mathbf{A}_{\epsilon_k} &\triangleq [\mathbf{a}_{-L_g}(\epsilon_k), \dots, \mathbf{a}_0(\epsilon_k), \dots, \mathbf{a}_{L_o+L_g-1}(\epsilon_k)] \\ \mathbf{a}_i(\epsilon_k) &\triangleq [g(-iT - \epsilon_k T), g(-iT + T_s - \epsilon_k T), \dots, \\ &\quad g(-iT + (L_o Q - 1)T_s - \epsilon_k T)]^T \end{aligned}$$

$$\begin{aligned} \mathbf{s} &\triangleq [s(0), \dots, s(L-1)]^T \\ \mathbf{p}_k &\triangleq [p_k(-L_g), \dots, p_k(0), \dots, p_k(L_o + L_g - 1)]^T \\ \mathbf{e}_k &\triangleq [e_k(-L_g), \dots, e_k(0), \dots, e_k(L_o + L_g - 1)]^T \\ \mathbf{v} &\triangleq [v(0), v(T_s), \dots, v((L_o Q - 1)T_s)]^T. \end{aligned}$$

After defining $\mathbf{A}_{\epsilon} \triangleq [\mathbf{A}_{\epsilon_1}, \dots, \mathbf{A}_{\epsilon_K}]$, $\mathbf{H} \triangleq \text{diag}[h_1, \dots, h_K] \otimes \mathbf{I}_{L_o+2L_g}$, $\mathbf{F} \triangleq \text{diag}[f_1, \dots, f_K] \otimes \mathbf{I}_{L_o+2L_g}$, $\mathbf{W} \triangleq \text{diag}[\mathbf{W}_1, \dots, \mathbf{W}_K]$ and $\mathbf{\Omega} \triangleq [\mathbf{\Omega}_1^H, \dots, \mathbf{\Omega}_K^H]^H$, the general model in (6) can be rewritten compactly as

$$\mathbf{y} = \mathbf{A}_{\epsilon} \mathbf{H} \mathbf{W} \mathbf{F} \mathbf{\Omega} \mathbf{s} + \mathbf{A}_{\epsilon} \mathbf{H} \mathbf{W} \mathbf{p} + \mathbf{A}_{\epsilon} \mathbf{H} \mathbf{W} \mathbf{e} + \mathbf{v} \quad (7)$$

where $\mathbf{p} \triangleq [\mathbf{p}_1^H, \dots, \mathbf{p}_K^H]^H$ and $\mathbf{e} \triangleq [\mathbf{e}_1^H, \dots, \mathbf{e}_K^H]^H$.

In the next section, estimation algorithms are presented to estimate the channels ($\mathbf{f} \triangleq [f_1, \dots, f_K]^T$, $\mathbf{h} \triangleq [h_1, \dots, h_K]^T$) and timing ($\epsilon \triangleq [\epsilon_1, \dots, \epsilon_K]^T$) parameters. Then in Section IV, the resynchronization algorithms at the relays and the destination are discussed.

III. JOINT TIMING AND CHANNEL ESTIMATION IN THE TRAINING PERIOD

During the training period, there is no timing and channel information. Hence, without loss of generality, the precoder is chosen as $\mathbf{W} = \mathbf{I}_{K(L_o+2L_g)}$ since in general \mathbf{W} depends on timing and channel information obtained from estimation in this period. Thus, the system model (7) in the training period is simplified as

$$\mathbf{y}_t = \mathbf{A}_{\epsilon} \mathbf{H} \mathbf{F} \mathbf{\Omega} \mathbf{s}_t + \mathbf{A}_{\epsilon} \mathbf{H} \mathbf{p} + \mathbf{A}_{\epsilon} \mathbf{H} \mathbf{e} + \mathbf{v} \quad (8)$$

where \mathbf{y}_t is the received signal at \mathbb{D} during the training period.

In order to derive the joint timing and channel estimator, we start with an equivalent model to (8) using the diagonal structure of \mathbf{H} and \mathbf{F}

$$\begin{aligned} \mathbf{y}_t &= \underbrace{[\mathbf{A}_{\epsilon_1} \mathbf{\Omega}_1 \mathbf{s}_t, \dots, \mathbf{A}_{\epsilon_K} \mathbf{\Omega}_K \mathbf{s}_t]}_{\triangleq \mathbf{M}(\epsilon)} \underbrace{\begin{bmatrix} f_1 h_1 \\ \vdots \\ f_K h_K \end{bmatrix}}_{\triangleq \boldsymbol{\xi}} \\ &+ \underbrace{[\mathbf{A}_{\epsilon_1} \mathbf{p}_1, \dots, \mathbf{A}_{\epsilon_K} \mathbf{p}_K]}_{\triangleq \mathbf{N}(\epsilon)} \underbrace{\begin{bmatrix} h_1 \\ \vdots \\ h_K \end{bmatrix}}_{\triangleq \mathbf{h}} + \mathbf{z} \\ &= \mathbf{M}(\epsilon) \boldsymbol{\xi} + \mathbf{N}(\epsilon) \mathbf{h} + \mathbf{z} \end{aligned} \quad (9)$$

where $\mathbf{z} = \mathbf{A}_{\epsilon} \mathbf{H} \mathbf{e} + \mathbf{v}$ is the equivalent compound noise at \mathbb{D} with covariance

$$\mathbf{R}_{\mathbf{z}} = \mathbb{E}\{\mathbf{z}\mathbf{z}^H\} = \mathbf{A}_{\epsilon} \mathbf{H} \mathbf{R}_{\mathbf{e}} \mathbf{H}^H \mathbf{A}_{\epsilon}^H + \mathbf{R}_{\mathbf{v}}. \quad (10)$$

In (10), $\mathbf{R}_{\mathbf{v}}$ is the covariance of \mathbf{v} and $\mathbf{R}_{\mathbf{e}} = \text{diag}[\mathbf{R}_{\mathbf{e}_1}, \dots, \mathbf{R}_{\mathbf{e}_K}]$ with $\mathbf{R}_{\mathbf{e}_k}$ defined in (4). Now that the estimation of separate channels \mathbf{f} and \mathbf{h} in each hop becomes the estimation of the composite channel $\boldsymbol{\xi} \triangleq \mathbf{f} \odot \mathbf{h}$ and the second hop channel \mathbf{h} .

Remark: For estimation purpose, \mathbf{s}_t , $\mathbf{\Omega}_k$ and \mathbf{p}_k should be chosen to minimize the estimation errors (e.g., estimation MSE

[19], Cramér–Rao bound [35]). The design of $\mathbf{\Omega}_k$ and training \mathbf{s}_t has been discussed in [17], [19]. Since the focus of this paper is to design the timing resynchronization algorithms, here without loss of generality, \mathbf{p}_k is chosen as white sequences whereas \mathbf{s}_t and $\mathbf{\Omega}_k$ are chosen according to [19] and [17] respectively.

A. Least Squares Estimation

As a straightforward approach, we can estimate the parameters ϵ , ξ and \mathbf{h} by minimizing the norm of error \mathbf{z}

$$\Lambda_{\text{LS}}(\epsilon, \xi, \mathbf{h}) = \left\| \mathbf{y}_t - \underbrace{[\mathbf{M}(\epsilon) \quad \mathbf{N}(\epsilon)]}_{\triangleq \mathbf{Q}(\epsilon)} \begin{bmatrix} \xi \\ \mathbf{h} \end{bmatrix} \right\|^2. \quad (11)$$

When ϵ are fixed, it can be readily shown that the Least Squares (LS) estimates of ξ and \mathbf{h} are

$$\begin{bmatrix} \hat{\xi} \\ \hat{\mathbf{h}} \end{bmatrix} = [\mathbf{Q}(\epsilon)^H \mathbf{Q}(\epsilon)]^{-1} \mathbf{Q}(\epsilon)^H \mathbf{y}_t. \quad (12)$$

After substituting (12) into (11) and ignoring some scaling constants and irrelevant terms, a cost function that only depends on ϵ is obtained as

$$\Lambda_{\text{LS}}(\epsilon, \hat{\xi}, \hat{\mathbf{h}}) = \|(\mathbf{I} - \mathbf{Q}(\epsilon)[\mathbf{Q}(\epsilon)^H \mathbf{Q}(\epsilon)]^{-1} \mathbf{Q}(\epsilon)^H) \mathbf{y}_t\|^2. \quad (13)$$

Then the timing offsets ϵ are estimated as

$$\hat{\epsilon}_{\text{LS}} = \arg \min_{\epsilon} \Lambda_{\text{LS}}(\epsilon, \hat{\xi}, \hat{\mathbf{h}}). \quad (14)$$

The above minimization (14) is a multidimensional problem, which imposes high computational complexity at \mathbb{D} . To cope with this issue, alternating projection [34] can be used to reduce the K -dimensional minimization into a series of one-dimensional searches. Finally, the channel estimates are obtained by putting $\hat{\epsilon}_{\text{LS}}$ back to (12).

B. Iterative Maximum-Likelihood Estimation

The LS solution derived above is simple and straightforward but it does not consider the effect of the channel-dependent noise \mathbf{z} . For higher accuracy, we here derive a ML estimator. From (9), the likelihood function of timing ϵ and channel ξ, \mathbf{h} is obtained as

$$P(\mathbf{y}_t | \epsilon, \xi, \mathbf{h}) = \frac{1}{\pi^{L_o Q} \det(\mathbf{R}_z)} \times \exp\left(-\|\mathbf{y}_t - \mathbf{M}(\epsilon)\xi - \mathbf{N}(\epsilon)\mathbf{h}\|_{\mathbf{R}_z^{-1}}^2\right) \quad (15)$$

where \mathbf{R}_z was defined in (10) and is related to ϵ and \mathbf{h} . Therefore, the exact ML estimator is highly complex. In the sequel, we propose an iterative ML solution to refine the estimates from the LS solution.

Maximizing (15) is equivalent to maximizing the log-likelihood function

$$\Lambda_{\text{ML}}(\epsilon, \xi, \mathbf{h}) = C - \log \det[\mathbf{R}_z(\epsilon, \mathbf{h})] - \|\mathbf{y}_t - \mathbf{M}(\epsilon)\xi - \mathbf{N}(\epsilon)\mathbf{h}\|_{\mathbf{R}_z^{-1}(\epsilon, \mathbf{h})}^2 \quad (16)$$

where C is an irrelevant constant and the dependency of \mathbf{R}_z on timing ϵ and channel \mathbf{h} is explicitly stated as $\mathbf{R}_z(\epsilon, \mathbf{h})$. Then the iterative algorithm is initialized as $(\hat{\epsilon}_{\text{ML}}^{(0)}, \hat{\xi}_{\text{ML}}^{(0)}, \hat{\mathbf{h}}_{\text{ML}}^{(0)}) = (\hat{\epsilon}_{\text{LS}}, \hat{\xi}_{\text{LS}}, \hat{\mathbf{h}}_{\text{LS}})$ and the parameter

estimates $(\hat{\epsilon}_{\text{ML}}^{(i+1)}, \hat{\xi}_{\text{ML}}^{(i+1)}, \hat{\mathbf{h}}_{\text{ML}}^{(i+1)})$ in the $(i+1)$ th iteration are updated in turn, starting from the composite channel ξ as follows.

• Updating $\hat{\xi}_{\text{ML}}^{(i+1)}$:

The composite channel is updated as $\hat{\xi}_{\text{ML}}^{(i+1)} = \arg \max_{\xi} \Lambda_{\text{ML}}(\hat{\epsilon}_{\text{ML}}^{(i)}, \xi, \hat{\mathbf{h}}_{\text{ML}}^{(i)})$. Since $\mathbf{R}_z(\epsilon, \mathbf{h})$ is not a function of ξ , the above problem is actually a weighted LS problem [35]

$$\hat{\xi}_{\text{ML}}^{(i+1)} = \arg \min_{\xi} \left\| \mathbf{y}_t - \mathbf{M}(\hat{\epsilon}_{\text{ML}}^{(i)}) \xi - \mathbf{N}(\hat{\epsilon}_{\text{ML}}^{(i)}) \hat{\mathbf{h}}_{\text{ML}}^{(i)} \right\|_{\mathbf{R}_z^{-1}(\hat{\epsilon}_{\text{ML}}^{(i)}, \hat{\mathbf{h}}_{\text{ML}}^{(i)})}^2$$

and the solution can be computed analytically as

$$\begin{aligned} \hat{\xi}_{\text{ML}}^{(i+1)} &= \left[\mathbf{M}(\hat{\epsilon}_{\text{ML}}^{(i)})^H \mathbf{M}(\hat{\epsilon}_{\text{ML}}^{(i)}) + \mathbf{R}_z^{-1}(\hat{\epsilon}_{\text{ML}}^{(i)}, \hat{\mathbf{h}}_{\text{ML}}^{(i)}) \right]^{-1} \\ &\quad \times \mathbf{M}(\hat{\epsilon}_{\text{ML}}^{(i)})^H \left[\mathbf{y}_t - \mathbf{N}(\hat{\epsilon}_{\text{ML}}^{(i)}) \hat{\mathbf{h}}_{\text{ML}}^{(i)} \right]. \end{aligned} \quad (17)$$

• Updating $\hat{\mathbf{h}}_{\text{ML}}^{(i+1)}$:

The second hop channel \mathbf{h} is updated as $\hat{\mathbf{h}}_{\text{ML}}^{(i+1)} = \arg \max_{\mathbf{h}} \Lambda_{\text{ML}}(\hat{\epsilon}_{\text{ML}}^{(i)}, \hat{\xi}_{\text{ML}}^{(i+1)}, \mathbf{h})$ and evaluated as

$$\begin{aligned} \hat{\mathbf{h}}_{\text{ML}}^{(i+1)} &= \arg \min_{\mathbf{h}} \left\{ \log \det \left[\mathbf{R}_z(\hat{\epsilon}_{\text{ML}}^{(i)}, \mathbf{h}) \right] \right. \\ &\quad \left. + \left\| \mathbf{y}_t - \mathbf{M}(\hat{\epsilon}_{\text{ML}}^{(i)}) \hat{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\hat{\epsilon}_{\text{ML}}^{(i)}) \mathbf{h} \right\|_{\mathbf{R}_z^{-1}(\hat{\epsilon}_{\text{ML}}^{(i)}, \mathbf{h})}^2 \right\}. \end{aligned} \quad (18)$$

The above expression is highly nonlinear but can be solved via the gradient descent and related algorithms (i.e., conjugate gradient and quasi-Newton methods [20]), which requires the partial derivative of $\partial \Lambda_{\text{ML}}(\hat{\epsilon}_{\text{ML}}^{(i)}, \hat{\xi}_{\text{ML}}^{(i+1)}, \mathbf{h}) / \partial \mathbf{h}^*$. The partial derivatives are calculated elementwise as shown in (41) in Appendix A.

• Updating $\hat{\epsilon}_{\text{ML}}^{(i+1)}$:

Finally, the timing offsets ϵ are updated as $\hat{\epsilon}_{\text{ML}}^{(i+1)} = \arg \max_{\epsilon} \Lambda_{\text{ML}}(\epsilon, \hat{\xi}_{\text{ML}}^{(i+1)}, \hat{\mathbf{h}}_{\text{ML}}^{(i+1)})$ and is equivalent to

$$\begin{aligned} \hat{\epsilon}_{\text{ML}}^{(i+1)} &= \arg \min_{\epsilon} \left\{ \log \det \left[\mathbf{R}_z(\epsilon, \hat{\mathbf{h}}_{\text{ML}}^{(i+1)}) \right] \right. \\ &\quad \left. + \left\| \mathbf{y}_t - \mathbf{M}(\epsilon) \hat{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\epsilon) \hat{\mathbf{h}}_{\text{ML}}^{(i+1)} \right\|_{\mathbf{R}_z^{-1}(\epsilon, \hat{\mathbf{h}}_{\text{ML}}^{(i+1)})}^2 \right\} \end{aligned} \quad (19)$$

where an exhaustive search is performed using alternating projection.

With the results above, the joint timing and channel estimation algorithm is now summarized in Algorithm 1.

Algorithm 1: Joint Timing and Channel Estimation Algorithm

- (1) Perform coarse estimation using the Least Squares solution
 - Obtain $\hat{\epsilon}_{\text{LS}} = \arg \min_{\epsilon} \Lambda_{\text{LS}}(\epsilon, \hat{\xi}, \hat{\mathbf{h}})$ in (13);
 - Obtain $\begin{bmatrix} \hat{\xi}_{\text{LS}} \\ \hat{\mathbf{h}}_{\text{LS}} \end{bmatrix} = [\mathbf{Q}(\hat{\epsilon}_{\text{LS}})^H \mathbf{Q}(\hat{\epsilon}_{\text{LS}})]^{-1} \mathbf{Q}(\hat{\epsilon}_{\text{LS}})^H \mathbf{y}_t$.
- (2) If higher accuracy is desired, perform fine estimation using the iterative Maximum Likelihood solution
 - **Initialize** $(\hat{\epsilon}_{\text{ML}}^{(0)}, \hat{\xi}_{\text{ML}}^{(0)}, \hat{\mathbf{h}}_{\text{ML}}^{(0)}) = (\hat{\epsilon}_{\text{LS}}, \hat{\xi}_{\text{LS}}, \hat{\mathbf{h}}_{\text{LS}})$;
 - **Repeat** $i = 0, 1, \dots$

$$\hat{\xi}_{\text{ML}}^{(i+1)} = \arg \max_{\xi} \Lambda_{\text{ML}}(\hat{\epsilon}_{\text{ML}}^{(i)}, \xi, \hat{\mathbf{h}}_{\text{ML}}^{(i)}) \text{ in (17);}$$

$$\hat{\mathbf{h}}_{\text{ML}}^{(i+1)} = \arg \max_{\mathbf{h}} \Lambda_{\text{ML}}(\hat{\epsilon}_{\text{ML}}^{(i)}, \hat{\xi}_{\text{ML}}^{(i+1)}, \mathbf{h}) \text{ in (18);}$$

$$\hat{\epsilon}_{\text{ML}}^{(i+1)} = \arg \max_{\epsilon} \Lambda_{\text{ML}}(\epsilon, \hat{\xi}_{\text{ML}}^{(i+1)}, \hat{\mathbf{h}}_{\text{ML}}^{(i+1)}) \text{ in (19);}$$

- **until** $\|\hat{\xi}_{\text{ML}}^{(i+1)} - \hat{\xi}_{\text{ML}}^{(i)}\|^2$ and $\|\hat{\mathbf{h}}_{\text{ML}}^{(i+1)} - \hat{\mathbf{h}}_{\text{ML}}^{(i)}\|^2$ and $\|\hat{\epsilon}_{\text{ML}}^{(i+1)} - \hat{\epsilon}_{\text{ML}}^{(i)}\|^2$ are smaller than corresponding thresholds ζ_{ξ} , ζ_h and ζ_{ϵ} .

C. Timing and Channel Estimation Uncertainty Analysis

Notice that all the parameters obtained during the training period suffer from random estimation errors. That is $\epsilon = \hat{\epsilon} + \delta_{\hat{\epsilon}}$, $\xi = \hat{\xi} + \delta_{\hat{\xi}}$ and $\mathbf{h} = \hat{\mathbf{h}} + \delta_{\hat{\mathbf{h}}}$, where $\delta_{\hat{\epsilon}} \triangleq [\delta_{\hat{\epsilon}_1}, \dots, \delta_{\hat{\epsilon}_K}]^T$, $\delta_{\hat{\xi}} \triangleq [\delta_{\hat{\xi}_1}, \dots, \delta_{\hat{\xi}_K}]^T$ and $\delta_{\hat{\mathbf{h}}} \triangleq [\delta_{\hat{h}_1}, \dots, \delta_{\hat{h}_K}]^T$ are the estimation errors of the corresponding parameters. If the statistical information of timing and channel uncertainties ($\delta_{\hat{\epsilon}}$, $\delta_{\hat{\xi}}$, $\delta_{\hat{\mathbf{h}}}$) is known to the receiver, then it is possible to design timing resynchronization algorithms that are robust to the estimation errors. Here we propose to use the CRB as a measure to provide insights into the uncertainties of parameters because the ML estimator asymptotically approaches the error performance predicted by the CRB [35].

Since the parameters of interest $(\epsilon, \xi, \mathbf{h})$ contain both real and complex-valued elements, we define $\boldsymbol{\theta} \triangleq [\epsilon, \Re\{\xi\}^T, \Im\{\xi\}^T, \Re\{\mathbf{h}\}^T, \Im\{\mathbf{h}\}^T]^T$. Denote $\boldsymbol{\mu} \triangleq \mathbf{M}(\epsilon)\xi + \mathbf{N}(\epsilon)\mathbf{h}$ and based on the probability density function in (15), the (i, j) th entry of the Fisher information matrix (FIM) \mathbf{J} is calculated as [35]

$$[\mathbf{J}]_{i,j} = 2\Re \left\{ \frac{\partial \boldsymbol{\mu}^H}{\partial \theta_i} \mathbf{R}_z^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \right\} + \text{tr} \left(\mathbf{R}_z^{-1} \frac{\partial \mathbf{R}_z}{\partial \theta_i} \mathbf{R}_z^{-1} \frac{\partial \mathbf{R}_z}{\partial \theta_j} \right) \quad (20)$$

where θ_i is the i th element in vector $\boldsymbol{\theta}$.

For the calculation of FIM \mathbf{J} , the corresponding components are computed as

$$\begin{aligned} \frac{\partial \boldsymbol{\mu}^H}{\partial \epsilon_i} &= \zeta_i^* (\mathbf{D}_{\epsilon_i} \boldsymbol{\Omega}_i \mathbf{s}_t)^H + h_i^* (\mathbf{D}_{\epsilon_i} \mathbf{p}_i)^H \\ \frac{\partial \mathbf{R}_z}{\partial \epsilon_i} &= \mathbf{D}_{\epsilon_i} \mathbf{H}_i \mathbf{R}_{e_i} \mathbf{H}_i^H \mathbf{A}_{\epsilon_i}^H + \mathbf{A}_{\epsilon_i} \mathbf{H}_i \mathbf{R}_{e_i} \mathbf{H}_i^H \mathbf{D}_{\epsilon_i}^H \\ \frac{\partial \boldsymbol{\mu}^H}{\partial \Re\{\xi_i\}} &= j \cdot \frac{\partial \boldsymbol{\mu}^H}{\partial \Im\{\xi_i\}} = (\mathbf{A}_{\epsilon_i} \boldsymbol{\Omega}_i \mathbf{s}_t)^H \\ \frac{\partial \boldsymbol{\mu}^H}{\partial \Re\{h_i\}} &= j \cdot \frac{\partial \boldsymbol{\mu}^H}{\partial \Im\{h_i\}} = (\mathbf{A}_{\epsilon_i} \mathbf{p}_i)^H \\ \frac{\partial \mathbf{R}_z}{\partial \Re\{h_i\}} &= \mathbf{A}_{\epsilon_i} \mathbf{R}_{e_i} \mathbf{H}_i^H \mathbf{A}_{\epsilon_i}^H + \mathbf{A}_{\epsilon_i} \mathbf{H}_i \mathbf{R}_{e_i} \mathbf{A}_{\epsilon_i}^H \\ \frac{\partial \mathbf{R}_z}{\partial \Im\{h_i\}} &= j \cdot \mathbf{A}_{\epsilon_i} \mathbf{R}_{e_i} \mathbf{H}_i^H \mathbf{A}_{\epsilon_i}^H - j \cdot \mathbf{A}_{\epsilon_i} \mathbf{H}_i \mathbf{R}_{e_i} \mathbf{A}_{\epsilon_i}^H \end{aligned}$$

where j is the imaginary unit $j = \sqrt{-1}$ and $\mathbf{D}_{\epsilon_i} \triangleq \partial \mathbf{A}_{\epsilon_i} / \partial \epsilon_i$. Substituting the above results back to (20), we can obtain the CRB by inverting the FIM \mathbf{J} (i.e., $\text{CRB}(\boldsymbol{\theta}) = \mathbf{J}^{-1}$). Note that

$$\begin{bmatrix} \epsilon \\ \xi \\ \mathbf{h} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{I}_K & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K & j \cdot \mathbf{I}_K & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_K & j \cdot \mathbf{I}_K \end{bmatrix}}_{\triangleq \boldsymbol{\Pi}} \cdot \boldsymbol{\theta}, \quad (21)$$

then the CRB matrix of the original set of complex-valued parameters $(\epsilon, \xi, \mathbf{h})$ can be evaluated as [35]

$$\text{CRB}(\epsilon, \xi, \mathbf{h}) = \boldsymbol{\Pi} \cdot \mathbf{J}^{-1} \cdot \boldsymbol{\Pi}^H \quad (22)$$

and can be written in a block matrix form as

$$\text{CRB}(\epsilon, \xi, \mathbf{h}) = \begin{bmatrix} \mathbf{C}_{\epsilon, \epsilon} & \mathbf{C}_{\epsilon, \xi} & \mathbf{C}_{\epsilon, \mathbf{h}} \\ \mathbf{C}_{\xi, \epsilon} & \mathbf{C}_{\xi, \xi} & \mathbf{C}_{\xi, \mathbf{h}} \\ \mathbf{C}_{\mathbf{h}, \epsilon} & \mathbf{C}_{\mathbf{h}, \xi} & \mathbf{C}_{\mathbf{h}, \mathbf{h}} \end{bmatrix}, \quad (23)$$

where $\mathbf{C}_{\epsilon, \epsilon}$, $\mathbf{C}_{\xi, \xi}$, and $\mathbf{C}_{\mathbf{h}, \mathbf{h}}$ are the $K \times K$ CRB matrices for ϵ , ξ and \mathbf{h} respectively. According to [35] and [36], given the estimates $(\hat{\epsilon}, \hat{\xi}, \hat{\mathbf{h}})$, the true values of timing and channel parameters asymptotically follow a multivariate Gaussian distribution as

$$\begin{aligned} \epsilon &\sim \mathcal{N}(\hat{\epsilon}, \mathbf{C}_{\epsilon, \epsilon}) \\ \xi &\sim \mathcal{CN}(\hat{\xi}, \mathbf{C}_{\xi, \xi}) \\ \mathbf{h} &\sim \mathcal{CN}(\hat{\mathbf{h}}, \mathbf{C}_{\mathbf{h}, \mathbf{h}}). \end{aligned} \quad (24)$$

IV. JOINT RELAY AND DESTINATION TIMING RESYNCHRONIZATION ALGORITHMS

In this section, the joint design of timing resynchronization algorithms at both \mathbb{R}_k and \mathbb{D} is addressed under timing and channel uncertainties. Recall the general model in (7). During the data transmission period, the superimposed training \mathbf{p} in (7) is set to zero and all the powers at relays are used to transmit data from \mathbb{S} . Thus, the model becomes

$$\mathbf{y}_d = \mathbf{A}_{\epsilon} \mathbf{H} \mathbf{W} \mathbf{F} \boldsymbol{\Omega} \mathbf{s}_d + \mathbf{A}_{\epsilon} \mathbf{H} \mathbf{W} \mathbf{e} + \mathbf{v} \quad (25)$$

where \mathbf{y}_d is the received signal at \mathbb{D} during the data transmission period. Using the diagonal structure of \mathbf{W} , \mathbf{H} , and \mathbf{F} , we could rearrange the model in the data transmission period as

$$\mathbf{y}_d = \mathbf{A}_{\epsilon} \mathbf{W} \boldsymbol{\Xi} \boldsymbol{\Omega} \mathbf{s}_d + \mathbf{A}_{\epsilon} \mathbf{W} \mathbf{H} \mathbf{e} + \mathbf{v} \quad (26)$$

where $\boldsymbol{\Xi} = \mathbf{H} \mathbf{F}$ is the composite channel matrix.

For asynchronous AF systems, the signals from \mathbb{R}_k 's are no longer perfectly aligned with each other at the destination \mathbb{D} and hence there is ISI from adjacent symbols within the data block. Also, the inaccurate channel obtained from estimation will result in further performance degradation. Ultimately, the diversity gain achieved by distributed space-time coding [17]–[19] or optimal beamforming [24]–[27] will be lost due to the asynchronism and channel uncertainties. In the following, we jointly design robust timing resynchronization strategies for both relays and the destination in distributed AF systems.

Traditionally, for beamforming-based AF systems, the design of relay processing is focused on choosing the optimal precoding matrix \mathbf{W}_k to optimize certain performance criteria, such as minimum transmit power [23], maximum system capacity [24], [25], minimum recovered data MSE [29], [30] and maximum received SNR [26]. On the other hand, for distributed space-time coding-based AF systems, the design of relay processing is to normalize the transmitted power at each relay [17]–[19]. Unfortunately, the above mentioned works are based on the assumption of perfect synchronization, and \mathbf{W}_k is a scaled diagonal matrix $\mathbf{W}_k = w_k \mathbf{I}_{L_o + 2L_g}$. However, a diagonal scaling operation at relays may not be optimal under

asynchronism due to the ISI. Therefore, in the following, \mathbf{W}_k is considered as a general matrix and designed jointly with the timing equalizer \mathbf{G} at the destination \mathbb{D} .

A. Problem Formulation

Here, we propose to jointly design timing resynchronization strategies at relay \mathbf{W} and destination \mathbf{G} to minimize the recovered data MSE under a power constraint

$$\min_{\mathbf{G}, \mathbf{W}} \text{MSE}(\mathbf{G}, \mathbf{W}) \triangleq \mathbb{E} \{ \|\mathbf{G}\mathbf{y}_d - \mathbf{T}(\boldsymbol{\eta})\mathbf{s}_d\|^2 \} \quad (27)$$

$$\text{s.t. } \mathcal{E}(\mathbf{W}) \leq P_D \quad (28)$$

where the expectation in (27) is taken with respect to the statistics of $\boldsymbol{\epsilon}$, $\boldsymbol{\xi}$, \mathbf{h} , \mathbf{s}_d , \mathbf{e} and \mathbf{v} . The matrix $\mathbf{T}(\boldsymbol{\eta})$ is a $L_o \times (L_o + 2L_g)$ circulant matrix with the first row $\boldsymbol{\eta}$ being

$$\boldsymbol{\eta} \triangleq [R_{gg}(-L_g), \dots, R_{gg}(0), \dots, R_{gg}(L_g) \quad \mathbf{0}_{1 \times (L_o - 1)}] \quad (29)$$

where $[R_{gg}(-L_g), \dots, R_{gg}(0), \dots, R_{gg}(L_g)]$ stands for the ideal zero-ISI sampled waveform after matched filtering, with $R_{gg}(\epsilon)$ being the autocorrelation function of $g(t)$ at $t = \epsilon T$. The matrix $\mathbf{T}(\boldsymbol{\eta})$ can be interpreted as a windowing operation selecting the length- L_o block of data of interest for detection.

Besides, $\mathcal{E}(\mathbf{W})$ is the average signal power received at the destination \mathbb{D} from the K relays

$$\mathcal{E}(\mathbf{W}) \triangleq \mathbb{E} \{ \|\mathbf{A}_\epsilon \mathbf{W} \boldsymbol{\Xi} \Omega \mathbf{s}_d + \mathbf{A}_\epsilon \mathbf{W} \mathbf{H} \mathbf{e}\|^2 \} \quad (30)$$

where the expectation is taken with respect to $\boldsymbol{\epsilon}$, $\boldsymbol{\xi}$, \mathbf{h} , \mathbf{s}_d and \mathbf{e} . On the other hand, P_D is used to specify a threshold at the destination \mathbb{D} to limit the power from the K relays [28], [37], [38]. This constraint is especially important for cooperative communication systems (e.g., "ad hoc" wireless networks, sensor networks, etc.) where relay nodes cooperate together to transmit a single data stream, which may result in a much higher transmit power than previously allowed.

Using the fact that the timing and channel uncertainties follow an asymptotic Gaussian distribution (24), the average MSE of the recovered data and the power constraint are evaluated in Appendix B to be

$$\begin{aligned} \text{MSE}(\mathbf{G}, \mathbf{W}) &= \text{tr} \{ \mathbf{G} \mathbf{A}_\epsilon \mathbf{W} \mathbf{R} \mathbf{W}^H \mathbf{A}_\epsilon^H \mathbf{G}^H \} + \text{tr} \{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \} \\ &\quad + \text{tr} \{ \mathbf{G} \mathbf{R}_v \mathbf{G}^H \} - \text{tr} \{ \mathbf{G} \mathbf{A}_\epsilon \mathbf{W} \hat{\boldsymbol{\Xi}} \Omega \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \} \\ &\quad - \text{tr} \left\{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s^H \Omega^H \hat{\boldsymbol{\Xi}}^H \mathbf{W}^H \mathbf{A}_\epsilon^H \mathbf{G}^H \right\} \end{aligned} \quad (31)$$

and

$$\mathcal{E}(\mathbf{W}) = \text{tr} \{ \mathbf{A}_\epsilon \mathbf{W} \mathbf{R} \mathbf{W}^H \mathbf{A}_\epsilon^H \} \quad (32)$$

where $\mathbf{R} \triangleq \hat{\boldsymbol{\Xi}} \Omega \mathbf{R}_s \Omega^H \hat{\boldsymbol{\Xi}}^H + \mathbf{R}_{\delta_\xi} \odot (\Omega \mathbf{R}_s \Omega^H) + \hat{\mathbf{H}} \mathbf{R}_e \hat{\mathbf{H}}^H + \mathbf{R}_{\delta_h} \odot \mathbf{R}_e$ and furthermore, $\mathbf{R}_{\delta_\xi} = \mathbf{C}_{\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\xi}}} \otimes \mathbf{I}_{L_o + 2L_g}$, $\mathbf{R}_{\delta_h} = \mathbf{C}_{\hat{\mathbf{h}}, \hat{\mathbf{h}}} \otimes \mathbf{I}_{L_o + 2L_g}$, and \mathbf{R}_s is the covariance of the data sequence $\mathbf{R}_s \triangleq \mathbb{E}_{\mathbf{s}_d} \{ \mathbf{s}_d \mathbf{s}_d^H \}$.

B. Joint Design of Timing Equalizer \mathbf{G} and Precoder \mathbf{W}

In order to derive the optimal pair (\mathbf{G}, \mathbf{W}) , we first differentiate (31) with respect to \mathbf{G}^H and set the derivative to zero. It follows that the optimal \mathbf{G} in terms of \mathbf{W} is

$$\mathbf{G} = \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s^H \Omega^H \hat{\boldsymbol{\Xi}}^H \mathbf{W}^H \mathbf{A}_\epsilon^H (\mathbf{A}_\epsilon \mathbf{W} \mathbf{R} \mathbf{W}^H \mathbf{A}_\epsilon^H + \mathbf{R}_v)^{-1}. \quad (33)$$

After substituting (33) into (31), it is proved in Appendix-C that the optimization problem (27) can be reformulated as

$$\min_{\mathbf{W}} \text{tr} \left\{ \Phi \left(\mathbf{R}^{H/2} \mathbf{W}^H \mathbf{A}_\epsilon^H \mathbf{R}_v^{-1} \mathbf{A}_\epsilon \mathbf{W} \mathbf{R}^{1/2} + \mathbf{I} \right)^{-1} \right\} \quad (34)$$

$$\text{s.t. } \text{tr} \{ \mathbf{A}_\epsilon \mathbf{W} \mathbf{R} \mathbf{W}^H \mathbf{A}_\epsilon^H \} \leq P_D \quad (35)$$

where $\Phi \triangleq \mathbf{R}^{-1/2} \hat{\boldsymbol{\Xi}} \Omega \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s^H \Omega^H \hat{\boldsymbol{\Xi}}^H \mathbf{R}^{-H/2}$.

Since \mathbf{R}_v and Φ are both Hermitian, according to eigendecomposition we can write $\mathbf{R}_v = \mathbf{U}_v \boldsymbol{\Lambda}_v \mathbf{U}_v^H$ and $\Phi = \mathbf{U}_\Phi \boldsymbol{\Lambda}_\Phi \mathbf{U}_\Phi^H$, where \mathbf{U}_v and \mathbf{U}_Φ are the unitary matrices containing eigenvectors that correspond to the eigenvalues in $\boldsymbol{\Lambda}_v$ and $\boldsymbol{\Lambda}_\Phi$. After transforming the variable \mathbf{W} by $\bar{\mathbf{W}} = \mathbf{U}_v^H \mathbf{A}_\epsilon \mathbf{W} \mathbf{R}^{1/2} \mathbf{U}_\Phi$, the optimization problem is further simplified as

$$\begin{aligned} \min_{\bar{\mathbf{W}}} \text{tr} \left\{ \boldsymbol{\Lambda}_\Phi (\bar{\mathbf{W}}^H \boldsymbol{\Lambda}_v^{-1} \bar{\mathbf{W}} + \mathbf{I})^{-1} \right\} \\ \text{s.t. } \text{tr} \{ \bar{\mathbf{W}}^H \bar{\mathbf{W}} \} \leq P_D. \end{aligned}$$

It is known that the optimal $\bar{\mathbf{W}}$ for the above problem has the following structure [39]:

$$\bar{\mathbf{W}}^o = \begin{bmatrix} \text{diag}(\beta_1, \dots, \beta_N) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (36)$$

where $N = \min\{\text{rank}(\boldsymbol{\Lambda}_v), K(L_o + 2L_g)\}$. Denote λ_v^i and λ_Φ^i as the i th diagonal element in matrices $\boldsymbol{\Lambda}_v$ and $\boldsymbol{\Lambda}_\Phi$, then the above problem becomes

$$\begin{aligned} \min_{\beta_i} \sum_{i=1}^N \frac{\lambda_\Phi^i}{\beta_i^2 / \lambda_v^i + 1} + \sum_{i=N+1}^{K(L_o + 2L_g)} \lambda_\Phi^i \\ \text{s.t. } \sum_{i=1}^N \beta_i^2 \leq P_D. \end{aligned}$$

Applying the method of Lagrange multiplier onto the above optimization problem, the optimal solution is readily obtained as

$$\beta_i^2 = \left(\sqrt{\frac{\lambda_v^i \lambda_\Phi^i}{\nu}} - \lambda_v^i \right)^+ \quad (37)$$

where $(x)^+ = x$ if $x \geq 0$ and $(x)^+ = 0$ if $x < 0$. The constant ν is chosen to satisfy the constraint $\sum_{i=1}^N \beta_i^2 \leq P_D$.

With the elements β_i calculated in (37), the optimal $\bar{\mathbf{W}}^o$ can be obtained as in (36). Then according to the transformation

$\bar{\mathbf{W}} = \mathbf{U}_v^H \mathbf{A}_{\hat{\epsilon}} \mathbf{W} \mathbf{R}^{1/2} \mathbf{U}_{\Phi}$, the optimal choice \mathbf{W}^o should satisfy the following equation:

$$\mathbf{A}_{\hat{\epsilon}} \mathbf{W}^o = \underbrace{\mathbf{U}_v \bar{\mathbf{W}}^o \mathbf{U}_{\Phi}^H \mathbf{R}^{-1/2}}_{\triangleq \boldsymbol{\Upsilon}_o}. \quad (38)$$

Since $\mathbf{A}_{\hat{\epsilon}} = [\mathbf{A}_{\hat{\epsilon}_1}, \dots, \mathbf{A}_{\hat{\epsilon}_K}]$ and $\mathbf{W}^o = \text{diag}[\mathbf{W}_1^o, \dots, \mathbf{W}_K^o]$, the above equation can also be written as

$$[\mathbf{A}_{\hat{\epsilon}_1} \mathbf{W}_1^o, \dots, \mathbf{A}_{\hat{\epsilon}_K} \mathbf{W}_K^o] = \boldsymbol{\Upsilon}_o \quad (39)$$

and the corresponding block of \mathbf{W}_k^o can be calculated as

$$\mathbf{W}_k^o = (\mathbf{A}_{\hat{\epsilon}_k}^H \mathbf{A}_{\hat{\epsilon}_k})^{-1} \mathbf{A}_{\hat{\epsilon}_k}^H \boldsymbol{\Upsilon}_o [(k-1)R+1 : kR] \quad (40)$$

with $R = L_o + 2L_g$ and $\boldsymbol{\Upsilon}_o[(k-1)R+1 : kR]$ denotes the submatrix containing the $[(k-1)R+1]$ th to the (kR) th columns of $\boldsymbol{\Upsilon}_o$.

V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, the performances of the proposed algorithms for estimation and timing resynchronization in different AF systems is demonstrated by Monte Carlo simulations, where each point is obtained by averaging over 10^4 runs. In all simulations, QPSK modulation is used and the pulse shaping filter $g(t)$ is assumed to be root-raised cosine waveform with effective tail length $L_g = 4$, roll-off factor 0.22 and normalized energy $\int_{-\infty}^{\infty} g^2(t) dt = 1$.

The length of the sequence is $L_o = 64$ and the oversampling ratio is set to $Q = 2$. The channel coefficients f_k 's and h_k 's are modeled as independent identically distributed (i.i.d.) complex Gaussian random variables with zero mean and unit variance. Finally, the noise covariance is taken as $\mathbf{R}_{n_k} = \sigma_{n_k}^2 \mathbf{I}_L$ at \mathbb{R}_k and $\mathbf{R}_v = \sigma_v^2 \mathbf{I}_{L_o Q}$ at \mathbb{D} . For simplicity, the noise powers at relays and destination are set to be the same $\sigma^2 \triangleq \sigma_{n_1}^2 = \dots = \sigma_{n_K}^2 = \sigma_v^2$, and the SNR is defined as $\text{SNR} = E_s / \sigma^2$, where E_s is the average transmitted signal power $\mathbb{E}\{|s(i)|^2\} = E_s$.

A. Estimation Performance

In the training period, the training sequence \mathbf{s}_t from the source and the encoding matrix $\boldsymbol{\Omega}_k$ are chosen according to [19] as $\mathbf{s}_t = \sqrt{E_s} \mathbf{1}_{(L_o+2L_g) \times 1}$ and $\boldsymbol{\Omega}_k$ specified in [17]. As introduced later, the same $\boldsymbol{\Omega}_k$ is also used for distributed space-time coding-based AF systems during the data transmission period as well. On the other hand, the superimposed training sequences \mathbf{p}_k 's at \mathbb{R}_k 's are all generated as $[\exp(-j\phi_{-L_g}), \dots, \exp(-j\phi_{L_o+L_g})]$, where ϕ_i is uniformly distributed between $[-\pi, \pi]$.

Here, the performances of the proposed LS estimator and iterative ML estimator are presented. For the iterative ML estimator, the corresponding thresholds for convergence are set as $\zeta_{\xi} = \zeta_h = \zeta_{\epsilon} = K \cdot 10^{-5}$ (with K being the number of relays), and approximately our simulations after three iterations can arrive at stable convergent results. In Fig. 2, the timing estimation MSE at \mathbb{D} and the corresponding CRB are plotted as a function of SNR ($K = 2$ and $K = 4$). It can be seen that for both cases, the CRB and the MSE for the LS and iterative ML estimators coincide in high SNR region, indicating that the LS solution pro-

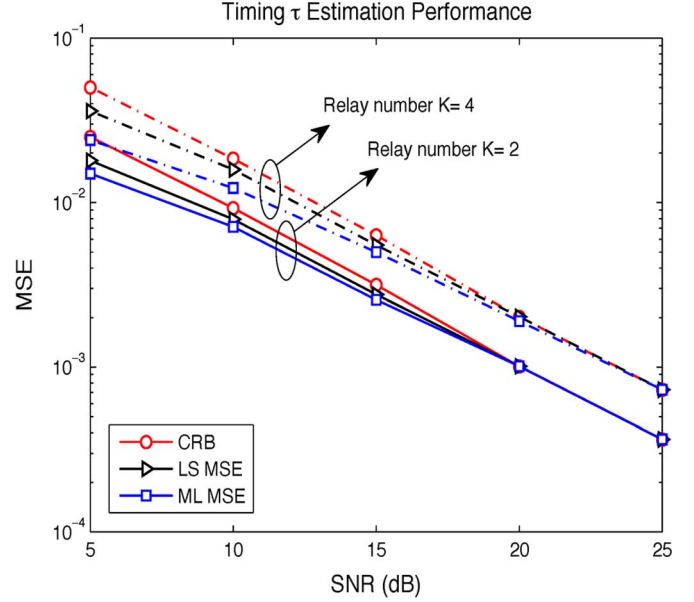


Fig. 2. Estimation performance of coarse LS and iterative ML estimators for timing offsets ϵ for both $K = 2$ and $K = 4$.

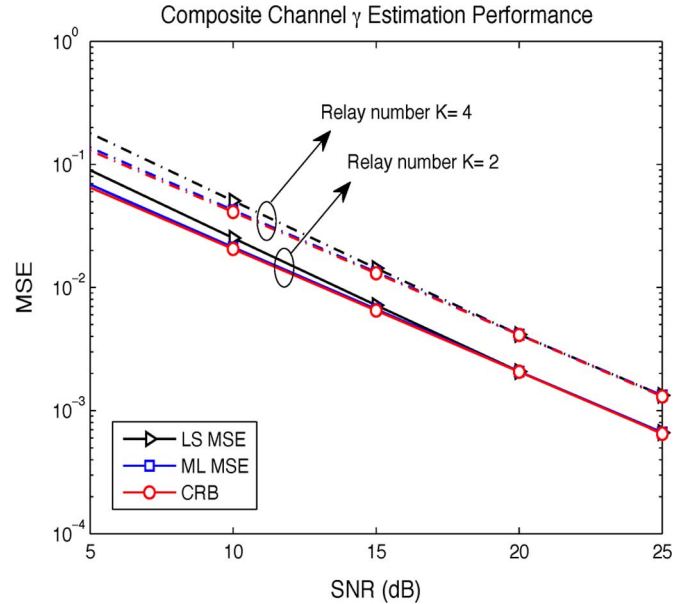


Fig. 3. Estimation performance of coarse LS and iterative ML estimators for composite channel ξ for both $K = 2$ and $K = 4$.

vides very close accuracy to the ML estimator and the CRB. At low SNR, the MSE of the LS estimate is slightly higher than those obtained from the ML estimator because the LS estimate does not consider the effects of the timing-channel-dependent noise, which are significant in low SNR region. It is also observed that for both cases at low SNR, the LS and ML MSEs fall below the CRB due to the fading environment. This phenomenon has been discussed in great details in [22] and hence it is not repeated here due to the space limitation.

In Fig. 3, the MSE of the composite channel estimate and the corresponding CRB are plotted. It is obvious from Fig. 3 that the performances of both the LS and ML channel estimators touch the CRB when the SNR is high. When the SNR is low, the

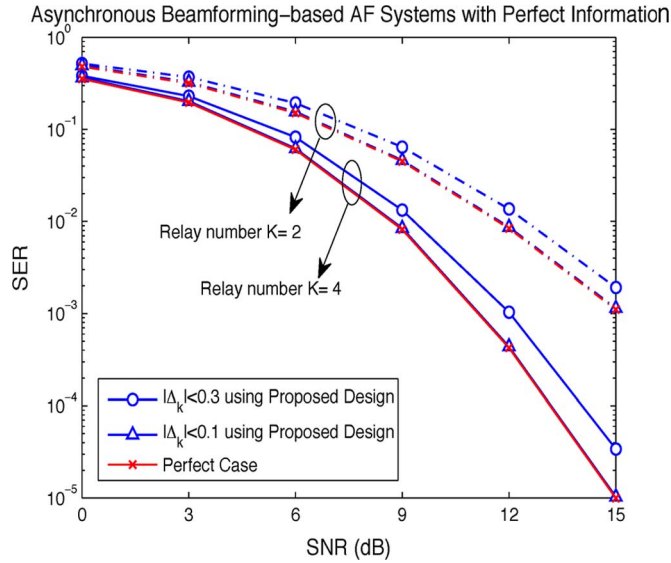


Fig. 4. SER performance for asynchronous beamforming-based AF systems with Perfect Information using QPSK modulation, $|\Delta_k| < 0.1$ and $|\Delta_k| < 0.3$ for $K = 2$ and $K = 4$.

MSE performance of the LS solution starts to deviate from the CRB while the ML solution remains close to the CRB. Similar performance can be observed for the estimation of the second hop channel \mathbf{h} and thus it is not presented here.

B. Timing Resynchronization Performance

The performance of the proposed joint relay and destination timing resynchronization algorithms is thoroughly illustrated in this section. In order to better illustrate the effects of timing offsets, we represent the offsets as $\epsilon_k = \epsilon_o + \Delta_k$, where ϵ_o is the common offset with respect to a certain time frame at \mathbb{D} and Δ_k is the residual offset. Furthermore, we separately consider the results for distributed beamforming-based and distributed space-time coding-based AF systems.

1) *Distributed Beamforming-Based AF Systems*: In the beamforming-based AF system, $\Omega_k = \mathbf{I}_R$. The proposed design is first examined in Fig. 4 for $K = 2$ and 4 relay systems when perfect timing and channel information are available under $|\Delta_k| < 0.1$ and $|\Delta_k| < 0.3$. When the timing asynchronism is relatively mild $|\Delta_k| < 0.1$, the performance of the proposed joint design (\mathbf{G}, \mathbf{W}) overlaps with the performance of the perfect case. Even under severe timing misalignment $|\Delta_k| < 0.3$, the performance of the proposed design still remains at a satisfactory level and the degradation brought by the increase of delay difference is acceptable.

Meanwhile in Fig. 5, the performance of the joint relay and destination timing resynchronization algorithms (\mathbf{G}, \mathbf{W}) is further examined in $K = 2$ and 4 relay systems under $|\Delta_k| < 0.1$ and with timing and channel uncertainties. Especially, the proposed design is compared with a nonrobust design where the parameter estimates are directly used without considering their estimation errors. It is obvious that the proposed design is effective in dealing with system uncertainties and outperforms the nonrobust design by almost an order of magnitude. As can be seen, under mild asynchronism $|\Delta_k| < 0.1$, the proposed joint timing resynchronization algorithm provides very close

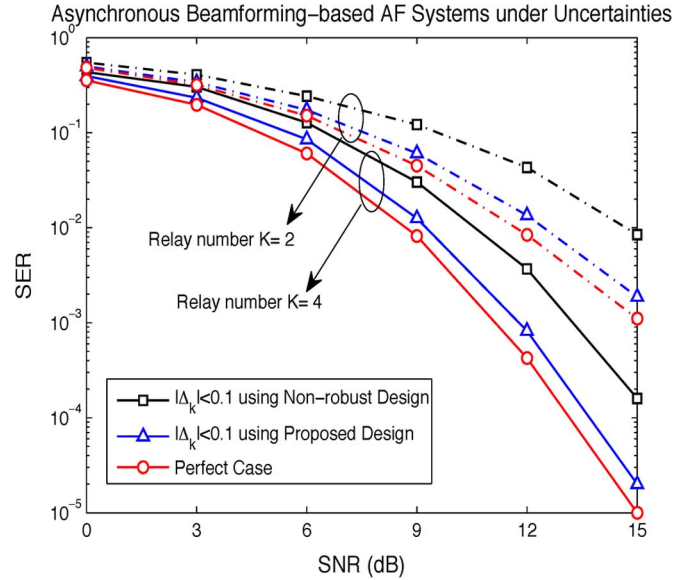


Fig. 5. SER performance for asynchronous beamforming-based AF systems with uncertainties using QPSK modulation, $|\Delta_k| < 0.1$ for $K = 2$ and $K = 4$.

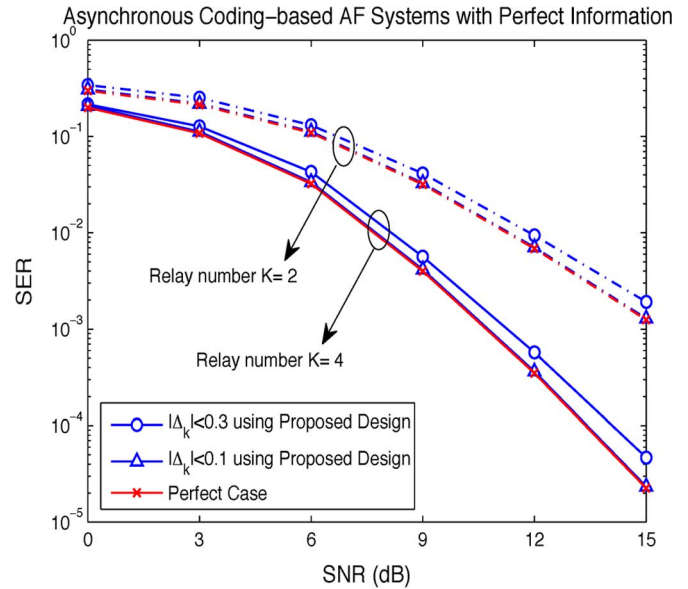


Fig. 6. SER performance for asynchronous coding-based AF systems with Perfect Information using QPSK modulation, $|\Delta_k| < 0.1$ and $|\Delta_k| < 0.3$ for $K = 2$ and $K = 4$.

performance to the perfect case, in which there is no asynchronism and system uncertainty. This shows that the proposed algorithm can effectively mitigate the ISI in beamforming-based AF systems in terms of SER performance, especially in salvaging the diversity gain that may be lost due to asynchronism [9].

2) *Distributed Coding-Based AF Systems*: For the distributed coding-based AF system, during the data transmission period, the coding matrices Ω_k are generated randomly based on isotropic distribution on the space of $(L_o + 2L_g) \times (L_o + 2L_g)$ unitary matrices as in [17] and [18]. The performance of the proposed robust joint relay and destination timing resynchronization algorithm is examined for coding-based AF systems in Figs. 6 and 7, respectively. It can be seen that similar phenomena to that of beamforming-based systems can be observed for asynchronous coding-based AF systems as well and it can

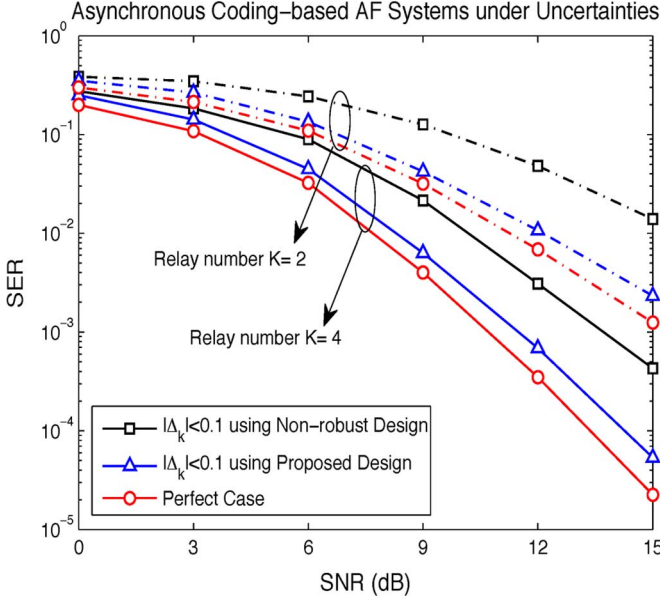


Fig. 7. SER performance for asynchronous coding-based AF systems with uncertainties using QPSK modulation, $|\Delta_k| < 0.1$ for $K = 2$ and $K = 4$.

be easily noted that asynchronous coding-based AF systems generally preserve its diversity order using the proposed joint relay and destination resynchronization algorithms.

VI. CONCLUSION

In this paper, the problem of joint timing offset and channel estimation, and furthermore the joint design of relay and destination timing resynchronization algorithms was considered for amplify-and-forward cooperative relay systems. The LS estimator, iterative ML estimator and CRB were derived for the timing and channel parameters. Then, aiming at minimizing the recovered data MSE at the destination and taking into account the uncertainties in the estimated timing and channel parameters, the robust joint timing resynchronization algorithm at relay and destination was derived. The proposed framework is a general methodology that can be applied to asynchronous AF systems with beamforming or distributed space-time coding. Furthermore, this framework integrates the estimation process into the design problem that deals with asynchronism and system

uncertainties. Simulation results have well supported our presented analysis and also verified the efficiency of the estimation and timing compensation algorithms.

APPENDIX A

CALCULATION OF PARTIAL DERIVATIVE

$$\partial \Lambda_{\text{ML}}(\boldsymbol{\epsilon}_{\text{ML}}^{(i)}, \boldsymbol{\xi}_{\text{ML}}^{(i+1)}, \mathbf{h}) / \partial \mathbf{h}^*$$

The partial derivative with respect to \mathbf{h}^* can be obtained elementwise as

$$\frac{\partial \Lambda_{\text{ML}}(\boldsymbol{\epsilon}_{\text{ML}}^{(i)}, \boldsymbol{\xi}_{\text{ML}}^{(i+1)}, \mathbf{h})}{\partial \mathbf{h}^*} = \begin{bmatrix} \frac{\partial \Lambda_{\text{ML}}(\boldsymbol{\epsilon}_{\text{ML}}^{(i)}, \boldsymbol{\xi}_{\text{ML}}^{(i+1)}, \mathbf{h})}{\partial h_1^*} \\ \vdots \\ \frac{\partial \Lambda_{\text{ML}}(\boldsymbol{\epsilon}_{\text{ML}}^{(i)}, \boldsymbol{\xi}_{\text{ML}}^{(i+1)}, \mathbf{h})}{\partial h_K^*} \end{bmatrix} \quad (41)$$

With the expression of $\Lambda_{\text{ML}}(\boldsymbol{\epsilon}_{\text{ML}}^{(i)}, \boldsymbol{\xi}_{\text{ML}}^{(i+1)}, \mathbf{h})$ in (16), the partial derivative with respect to h_j^* is written as

$$\begin{aligned} & \frac{\partial \Lambda_{\text{ML}}(\boldsymbol{\epsilon}_{\text{ML}}^{(i)}, \boldsymbol{\xi}_{\text{ML}}^{(i+1)}, \mathbf{h})}{\partial h_j^*} \\ &= - \frac{\partial \log \det [\mathbf{R}_{\mathbf{z}}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h})]}{\partial h_j^*} \\ &= - \frac{\partial \left\| \mathbf{y}_t - \mathbf{M}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \boldsymbol{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \mathbf{h} \right\|_{\mathbf{R}_{\mathbf{z}}^{-1}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h})}^2}{\partial h_j^*} \end{aligned}$$

with each individual component in the above equation is computed as follows.

• Calculation of the first term:

With complex variable differentiation [35], we have

$$\begin{aligned} \frac{\partial \log \det \mathbf{R}_{\mathbf{z}}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h})}{\partial h_j^*} &= \text{tr} \left\{ \mathbf{R}_{\mathbf{z}}^{-1}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h}) \cdot \frac{\partial \mathbf{R}_{\mathbf{z}}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h})}{\partial h_j^*} \right\} \\ &= \text{tr} \left\{ \mathbf{R}_{\mathbf{z}}^{-1}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h}) \cdot \mathbf{A}_{e_j^{(i)}} \mathbf{H}_j \mathbf{R}_{e_j} \mathbf{A}_{e_j^{(i)}}^H \right\}. \end{aligned}$$

• **Calculation of the second term:** Similarly, the second term can be obtained using the chain rule as [see the equation shown at the bottom of the page]. By complex variable differentiation,

$$\begin{aligned} & \frac{\partial \left\| \mathbf{y}_t - \mathbf{M}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \boldsymbol{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \mathbf{h} \right\|_{\mathbf{R}_{\mathbf{z}}^{-1}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h})}^2}{\partial h_j^*} \\ &= \frac{\partial \left(\mathbf{y}_t - \mathbf{M}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \boldsymbol{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \mathbf{h} \right)^H}{\partial h_j^*} \mathbf{R}_{\mathbf{z}}^{-1}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h}) \left(\mathbf{y}_t - \mathbf{M}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \boldsymbol{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \mathbf{h} \right) \\ &+ \left(\mathbf{y}_t - \mathbf{M}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \boldsymbol{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \mathbf{h} \right)^H \mathbf{R}_{\mathbf{z}}^{-1}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h}) \frac{\partial \left(\mathbf{y}_t - \mathbf{M}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \boldsymbol{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \mathbf{h} \right)}{\partial h_j^*} \\ &+ \left(\mathbf{y}_t - \mathbf{M}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \boldsymbol{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \mathbf{h} \right)^H \frac{\partial \mathbf{R}_{\mathbf{z}}^{-1}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h})}{\partial h_j^*} \left(\mathbf{y}_t - \mathbf{M}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \boldsymbol{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \mathbf{h} \right). \end{aligned}$$

we can further compute the corresponding components in the above expression as

$$\begin{aligned} \frac{\partial \mathbf{R}_z^{-1}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h})}{\partial h_j^*} &= \mathbf{R}_z^{-1}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h}) \frac{\partial \mathbf{R}_z(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h})}{\partial h_j^*} \mathbf{R}_z^{-1}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}, \mathbf{h}) \\ \frac{\partial \left(\mathbf{y}_t - \mathbf{M}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \boldsymbol{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \mathbf{h} \right)^H}{\partial h_j^*} &= \left(\mathbf{A}_{\epsilon_j^{(i)}} \mathbf{P}_j \right)^H \\ \frac{\partial \left(\mathbf{y}_t - \mathbf{M}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \boldsymbol{\xi}_{\text{ML}}^{(i+1)} - \mathbf{N}(\hat{\boldsymbol{\epsilon}}_{\text{ML}}^{(i)}) \mathbf{h} \right)}{\partial h_j^*} &= \mathbf{0}. \end{aligned}$$

APPENDIX B

PROOF OF MSE(\mathbf{G}, \mathbf{W}) IN (31) AND $\mathcal{E}(\mathbf{W})$ IN (32)

A. MSE(\mathbf{G}, \mathbf{W}) Expression

Substituting (26) into (27), we have

$$\text{MSE}(\mathbf{G}, \mathbf{W}) = \mathbb{E}_{\boldsymbol{\epsilon}, \boldsymbol{\xi}, \mathbf{h}, \mathbf{s}_d, \mathbf{e}, \mathbf{v}} \left\{ \left\| \mathbf{G}(\mathbf{A}_{\boldsymbol{\epsilon}} \mathbf{W} \boldsymbol{\Xi} \boldsymbol{\Omega} \mathbf{s}_d + \mathbf{A}_{\boldsymbol{\epsilon}} \mathbf{W} \mathbf{H} \mathbf{e} + \mathbf{v}) - \mathbf{T}(\boldsymbol{\eta}) \mathbf{s}_d \right\|^2 \right\}. \quad (42)$$

Using the fact that \mathbf{s}_d , \mathbf{e} and \mathbf{v} are uncorrelated, we have

$$\begin{aligned} \text{MSE}(\mathbf{G}, \mathbf{W}) &= \\ &\text{tr} \left\{ \mathbf{G} \mathbb{E}_{\boldsymbol{\epsilon}} \left\{ \mathbf{A}_{\boldsymbol{\epsilon}} \mathbf{W} \mathbb{E}_{\boldsymbol{\Xi}} \left\{ \boldsymbol{\Xi} \boldsymbol{\Omega} \mathbb{E}_{\mathbf{s}_d} \left\{ \mathbf{s}_d \mathbf{s}_d^H \right\} \boldsymbol{\Omega}^H \boldsymbol{\Xi}^H \right\} \mathbf{W}^H \mathbf{A}_{\boldsymbol{\epsilon}}^H \right\} \mathbf{G}^H \right\} \\ &+ \text{tr} \left\{ \mathbf{G} \mathbb{E}_{\boldsymbol{\epsilon}} \left\{ \mathbf{A}_{\boldsymbol{\epsilon}} \mathbf{W} \mathbb{E}_{\mathbf{H}} \left\{ \mathbf{H} \mathbb{E}_{\mathbf{e}} \left\{ \mathbf{e} \mathbf{e}^H \right\} \mathbf{H}^H \right\} \mathbf{W}^H \mathbf{A}_{\boldsymbol{\epsilon}}^H \right\} \mathbf{G}^H \right\} \\ &- \text{tr} \left\{ \mathbf{G} \mathbb{E}_{\boldsymbol{\epsilon}} \left\{ \mathbf{A}_{\boldsymbol{\epsilon}} \mathbf{W} \mathbb{E}_{\boldsymbol{\Xi}} \left\{ \boldsymbol{\Xi} \boldsymbol{\Omega} \mathbb{E}_{\mathbf{s}_d} \left\{ \mathbf{s}_d \mathbf{s}_d^H \right\} \right\} \right\} \mathbf{T}^H(\boldsymbol{\eta}) \right\} \\ &- \text{tr} \left\{ \mathbf{T}(\boldsymbol{\eta}) \mathbb{E}_{\boldsymbol{\epsilon}} \left\{ \mathbb{E}_{\boldsymbol{\Xi}} \left\{ \mathbb{E}_{\mathbf{s}_d} \left\{ \mathbf{s}_d \mathbf{s}_d^H \right\} \boldsymbol{\Omega}^H \boldsymbol{\Xi}^H \right\} \mathbf{W}^H \mathbf{A}_{\boldsymbol{\epsilon}}^H \right\} \mathbf{G}^H \right\} \\ &+ \text{tr} \left\{ \mathbf{T}(\boldsymbol{\eta}) \mathbb{E}_{\mathbf{s}_d} \left\{ \mathbf{s}_d \mathbf{s}_d^H \right\} \mathbf{T}^H(\boldsymbol{\eta}) \right\} + \text{tr} \left\{ \mathbf{G} \mathbb{E}_{\mathbf{v}} \left\{ \mathbf{v} \mathbf{v}^H \right\} \mathbf{G}^H \right\}. \end{aligned} \quad (43)$$

Since timing offsets $\boldsymbol{\epsilon}$ are entangled through $\mathbf{A}_{\boldsymbol{\epsilon}}$ in a highly nonlinear manner, the expectations over channel uncertainties are tackled first and then the timing uncertainties are tackled later using Taylor series expansion.

• **Channel Uncertainties** : With the channel uncertainties $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}} + \boldsymbol{\delta}_{\hat{\boldsymbol{\xi}}}$ and $\mathbf{h} = \hat{\mathbf{h}} + \boldsymbol{\delta}_{\hat{\mathbf{h}}}$, we can write

$$\begin{aligned} \boldsymbol{\Xi} &= \hat{\boldsymbol{\Xi}} + \boldsymbol{\delta}_{\hat{\boldsymbol{\Xi}}} \\ \mathbf{H} &= \hat{\mathbf{H}} + \boldsymbol{\delta}_{\hat{\mathbf{H}}} \end{aligned}$$

where $\hat{\boldsymbol{\Xi}} = \text{diag}[\hat{\boldsymbol{\xi}}] \otimes \mathbf{I}_{L_o+2L_g}$, $\boldsymbol{\delta}_{\hat{\boldsymbol{\Xi}}} = \text{diag}[\boldsymbol{\delta}_{\hat{\boldsymbol{\xi}}}] \otimes \mathbf{I}_{L_o+2L_g}$ and $\boldsymbol{\delta}_{\hat{\mathbf{H}}} = \text{diag}[\boldsymbol{\delta}_{\hat{\mathbf{h}}}] \otimes \mathbf{I}_{L_o+2L_g}$. Using the asymptotic Gaussian distribution in (24), the expectations over $\boldsymbol{\Xi}$ and \mathbf{H} can be evaluated as

$$\mathbb{E}_{\boldsymbol{\Xi}} \{ \boldsymbol{\Xi} \boldsymbol{\Omega} \mathbf{R}_s \boldsymbol{\Omega}^H \boldsymbol{\Xi}^H \} = \hat{\boldsymbol{\Xi}} \boldsymbol{\Omega} \mathbf{R}_s \boldsymbol{\Omega}^H \hat{\boldsymbol{\Xi}}^H + \mathbf{R}_{\boldsymbol{\delta}_{\hat{\boldsymbol{\Xi}}}} \odot (\boldsymbol{\Omega} \mathbf{R}_s \boldsymbol{\Omega}^H) \quad (44)$$

$$\mathbb{E}_{\mathbf{H}} \{ \mathbf{H} \mathbf{R}_e \mathbf{H}^H \} = \hat{\mathbf{H}} \mathbf{R}_e \hat{\mathbf{H}}^H + \mathbf{R}_{\boldsymbol{\delta}_{\hat{\mathbf{H}}}} \odot \mathbf{R}_e \quad (45)$$

$$\mathbb{E}_{\boldsymbol{\Xi}} \{ \boldsymbol{\Xi} \boldsymbol{\Omega} \mathbf{R}_s \} = \hat{\boldsymbol{\Xi}} \boldsymbol{\Omega} \mathbf{R}_s \quad (46)$$

where

$$\begin{aligned} \mathbf{R}_{\boldsymbol{\delta}_{\hat{\boldsymbol{\Xi}}}} &= \mathbf{C}_{\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\xi}}} \otimes \mathbf{I}_{L_o+2L_g} \\ \mathbf{R}_{\boldsymbol{\delta}_{\hat{\mathbf{H}}}} &= \mathbf{C}_{\hat{\mathbf{h}}, \hat{\mathbf{h}}} \otimes \mathbf{I}_{L_o+2L_g} \\ \mathbf{R}_e &= \text{diag}[\mathbf{R}_{e_1}, \dots, \mathbf{R}_{e_K}] \end{aligned}$$

with $\mathbf{C}_{\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\xi}}}$ and $\mathbf{C}_{\hat{\mathbf{h}}, \hat{\mathbf{h}}}$ being the CRB matrices for $\boldsymbol{\xi}$ and \mathbf{h} defined in (23). Note that the estimates are used when evaluating the CRB matrices because the CRB depends on the true values of the parameters.

• **Timing Uncertainty** : On the other hand, due to the nonlinearity presented in the system model, we expand the matrix \mathbf{A}_{ϵ_k} using Taylor series around the estimates $\hat{\epsilon}_k$ as follows:

$$\mathbf{A}_{\epsilon_k} = \mathbf{A}_{\hat{\epsilon}_k} + \mathbf{D}_{\hat{\epsilon}_k} \delta_{\hat{\epsilon}_k} + O(\boldsymbol{\rho}_k \delta_{\hat{\epsilon}_k}^2),$$

where $\mathbf{D}_{\epsilon_i} = \partial \mathbf{A}_{\epsilon_i} / \partial \epsilon_i$ and the symbol $O(\boldsymbol{\rho}_k \delta_{\hat{\epsilon}_k}^2)$ represents those matrix terms with order higher than $\delta_{\hat{\epsilon}_k}^2$. With the above Taylor expansion and the distribution in (24), then for an arbitrary square matrix \mathbf{Z} containing square submatrices $\mathbf{Z}_{i,j}$ ($i, j = 1, \dots, K$), we have

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\epsilon}} \{ \mathbf{A}_{\boldsymbol{\epsilon}} \mathbf{Z} \mathbf{A}_{\boldsymbol{\epsilon}}^H \} &= \mathbb{E}_{\boldsymbol{\epsilon}} \left\{ \underbrace{\sum_{i,j=1}^K \delta_{\hat{\epsilon}_j} \mathbf{A}_{\hat{\epsilon}_i} \mathbf{Z}_{i,j} \mathbf{D}_{\hat{\epsilon}_j}^H}_{=0} \right\} + \mathbb{E}_{\boldsymbol{\epsilon}} \left\{ \underbrace{\sum_{i,j=1}^K \delta_{\hat{\epsilon}_i} \mathbf{D}_{\hat{\epsilon}_i} \mathbf{Z}_{i,j} \mathbf{A}_{\hat{\epsilon}_j}^H}_{=0} \right\} \\ &+ \mathbb{E}_{\boldsymbol{\epsilon}} \left\{ \underbrace{\sum_{i,j=1}^K \delta_{\hat{\epsilon}_i} \delta_{\hat{\epsilon}_j} \mathbf{D}_{\hat{\epsilon}_i} \mathbf{Z}_{i,j} \mathbf{D}_{\hat{\epsilon}_j}^H + O(\boldsymbol{\rho}_k \delta_{\hat{\epsilon}_k}^2)}_{\text{high order terms, neglected}} \right\} \\ &+ \sum_{i,j=1}^K \mathbf{A}_{\hat{\epsilon}_i} \mathbf{Z}_{i,j} \mathbf{A}_{\hat{\epsilon}_j}^H \\ &\simeq \mathbf{A}_{\hat{\boldsymbol{\epsilon}}} \mathbf{Z} \mathbf{A}_{\hat{\boldsymbol{\epsilon}}}^H \end{aligned} \quad (47)$$

and similarly $\mathbb{E}_{\boldsymbol{\epsilon}} \{ \mathbf{A}_{\boldsymbol{\epsilon}} \} \simeq \mathbf{A}_{\hat{\boldsymbol{\epsilon}}}$. Substituting the above results back to (43), we arrive at

$$\begin{aligned} \text{MSE}(\mathbf{G}, \mathbf{W}) &= \text{tr} \{ \mathbf{G} \mathbf{A}_{\hat{\boldsymbol{\epsilon}}} \mathbf{W} \mathbf{R} \mathbf{W}^H \mathbf{A}_{\hat{\boldsymbol{\epsilon}}}^H \mathbf{G}^H \} + \text{tr} \{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \} \\ &+ \text{tr} \{ \mathbf{G} \mathbf{R}_v \mathbf{G}^H \} - \text{tr} \{ \mathbf{G} \mathbf{A}_{\hat{\boldsymbol{\epsilon}}} \mathbf{W} \hat{\boldsymbol{\Xi}} \boldsymbol{\Omega} \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \} \\ &- \text{tr} \{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s^H \boldsymbol{\Omega}^H \hat{\boldsymbol{\Xi}}^H \mathbf{W}^H \mathbf{A}_{\hat{\boldsymbol{\epsilon}}}^H \mathbf{G}^H \} \end{aligned}$$

where $\mathbf{R} \triangleq \hat{\boldsymbol{\Xi}} \boldsymbol{\Omega} \mathbf{R}_s \boldsymbol{\Omega}^H \hat{\boldsymbol{\Xi}}^H + \mathbf{R}_{\boldsymbol{\delta}_{\hat{\boldsymbol{\Xi}}}} \odot (\boldsymbol{\Omega} \mathbf{R}_s \boldsymbol{\Omega}^H) + \hat{\mathbf{H}} \mathbf{R}_e \hat{\mathbf{H}}^H + \mathbf{R}_{\boldsymbol{\delta}_{\hat{\mathbf{H}}}} \odot \mathbf{R}_e$.

B. $\mathcal{E}(\mathbf{W})$ Expression

Similar to the procedures shown above, we can readily write the expression of $\mathcal{E}(\mathbf{W})$ as

$$\begin{aligned} \mathcal{E}(\mathbf{W}) &\triangleq \mathbb{E}_{\epsilon, \xi, \mathbf{h}, \mathbf{s}_d, \mathbf{e}} \{ \|\mathbf{A}_\epsilon \mathbf{W} \Xi \Omega \mathbf{s}_d + \mathbf{A}_\epsilon \mathbf{W} \mathbf{H} \mathbf{e}\|^2 \} \\ &= \text{tr} \left\{ \mathbb{E}_\epsilon \left\{ \mathbf{A}_\epsilon \mathbf{W} \mathbb{E}_\Xi \left\{ \underbrace{\Xi \Omega \mathbb{E}_{\mathbf{s}_d} \{ \mathbf{s}_d \mathbf{s}_d^H \}}_{=\mathbf{R}_s} \Omega^H \Xi^H \right\} \mathbf{W}^H \mathbf{A}_\epsilon^H \right\} \right\} \\ &\quad + \text{tr} \left\{ \mathbb{E}_\epsilon \left\{ \mathbf{A}_\epsilon \mathbf{W} \mathbb{E}_\mathbf{H} \left\{ \underbrace{\mathbf{H} \mathbb{E}_e \{ \mathbf{e} \mathbf{e}^H \}}_{=\mathbf{R}_e} \mathbf{H}^H \right\} \mathbf{W}^H \mathbf{A}_\epsilon^H \right\} \right\} \\ &= \text{tr} \{ \mathbf{A}_\epsilon \mathbf{W} \mathbf{R} \mathbf{W}^H \mathbf{A}_\epsilon^H \}. \end{aligned} \quad (48)$$

APPENDIX C PROOF OF (34)

Substituting (33) into (31), the MSE expression becomes

$$\begin{aligned} \text{MSE}(\mathbf{W}) &= \text{tr} \{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \} - \text{tr} \left\{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s^H \Omega^H \hat{\Xi}^H \mathbf{W}^H \mathbf{A}_\epsilon^H \right. \\ &\quad \left. \times (\mathbf{A}_\epsilon \mathbf{W} \mathbf{R} \mathbf{W}^H \mathbf{A}_\epsilon^H + \mathbf{R}_v)^{-1} \mathbf{A}_\epsilon \mathbf{W} \Xi \Omega \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \right\}. \end{aligned}$$

Using the transformation of variable $\tilde{\mathbf{W}} = \mathbf{R}_v^{-1/2} \mathbf{A}_\epsilon \mathbf{W} \mathbf{R}^{1/2}$, we have

$$\begin{aligned} \text{MSE}(\tilde{\mathbf{W}}) &= \text{tr} \{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \} - \text{tr} \left\{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s^H \Omega^H \hat{\Xi}^H \mathbf{R}^{-H/2} \right. \\ &\quad \left. \times \tilde{\mathbf{W}}^H (\tilde{\mathbf{W}} \tilde{\mathbf{W}}^H + \mathbf{I})^{-1} \tilde{\mathbf{W}} \mathbf{R}^{-1/2} \hat{\Xi} \Omega \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \right\}. \end{aligned}$$

Furthermore, using the matrix inversion lemma $\mathbf{A}^H (\mathbf{A} \mathbf{A}^H + \mathbf{I})^{-1} \mathbf{A} = \mathbf{I} - (\mathbf{A}^H \mathbf{A} + \mathbf{I})^{-1}$, the above expression can be simplified as

$$\begin{aligned} \text{MSE}(\tilde{\mathbf{W}}) &= \text{tr} \{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \} - \text{tr} \left\{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s^H \Omega^H \hat{\Xi}^H \mathbf{R}^{-H/2} \right. \\ &\quad \left. \times [\mathbf{I} - (\tilde{\mathbf{W}}^H \tilde{\mathbf{W}} + \mathbf{I})^{-1}] \mathbf{R}^{-1/2} \hat{\Xi} \Omega \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \right\} \\ &= \text{MSE}_0 + \text{tr} \left\{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s^H \Omega^H \hat{\Xi}^H \mathbf{R}^{-H/2} \right. \\ &\quad \left. \times (\tilde{\mathbf{W}}^H \tilde{\mathbf{W}} + \mathbf{I})^{-1} \mathbf{R}^{-1/2} \hat{\Xi} \Omega \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \right\} \end{aligned} \quad (49)$$

where

$$\begin{aligned} \text{MSE}_0 &= \text{tr} \{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \} \\ &\quad - \text{tr} \left\{ \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s^H \Omega^H \hat{\Xi}^H \mathbf{R}^{-1} \hat{\Xi} \Omega \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \right\} \end{aligned} \quad (50)$$

is a constant that is irrelevant to $\tilde{\mathbf{W}}$.

Denote $\Phi \triangleq \mathbf{R}^{-1/2} \hat{\Xi} \Omega \mathbf{R}_s \mathbf{T}^H(\boldsymbol{\eta}) \mathbf{T}(\boldsymbol{\eta}) \mathbf{R}_s^H \Omega^H \hat{\Xi}^H \mathbf{R}^{-H/2}$ and substitute $\tilde{\mathbf{W}} = \mathbf{R}_v^{-1/2} \mathbf{A}_\epsilon \mathbf{W} \mathbf{R}^{1/2}$ back to (49), the cost function (31) is now equivalent to

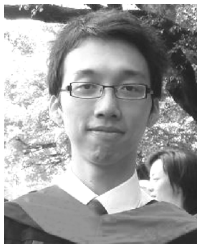
$$\begin{aligned} \overline{\text{MSE}}(\mathbf{W}) &= \text{tr} \left\{ \Phi \left(\mathbf{R}^{H/2} \mathbf{W}^H \mathbf{A}_\epsilon^H \mathbf{R}_v^{-1} \mathbf{A}_\epsilon \mathbf{W} \mathbf{R}^{1/2} + \mathbf{I} \right)^{-1} \right\} \end{aligned} \quad (51)$$

where the constant MSE_0 is dropped.

REFERENCES

- [1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, Part I, II," *IEEE Trans. Commun.*, vol. 51, pp. 1927–1948, Nov. 2003.
- [2] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.
- [3] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, pp. 1089–1098, Aug. 2004.
- [4] P. A. Anghel and M. Kaveh, "Exact symbol error probability of a cooperative network in a Rayleigh-fading environment," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1416–1421, Sep. 2004.
- [5] A. Ribeiro, X. Cai, and G. B. Giannakis, "Symbol error probabilities for general cooperative links," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1264–1273, May 2005.
- [6] J. Mietzner and P. A. Hoher, "Distributed space-time codes for cooperative wireless networks in the presence of different propagation delays and path losses," in *Proc. 3rd IEEE Sensor Array Multichannel Signal Processing Workshop*, Barcelona, Spain, Jul. 2004, pp. 264–268.
- [7] R. C. Palat, A. Annamalai, and J. H. Reed, "Upper bound on bit error rate for time synchronization errors in bandwidth limited distributed MIMO networks," in *Proc. IEEE WCNC*, Apr. 2006, vol. 4, pp. 2058–2063.
- [8] Y. Mei, Y. Hua, A. Swami, and B. Daneshmand, "Combating synchronization errors in cooperative relays," in *Proc. IEEE ICASSP*, Mar. 2005, vol. 3, pp. 369–372.
- [9] S. Jagannathan, H. Aghajan, and A. Goldsmith, "The effect of time synchronization errors on the performance of cooperative MISO systems," in *Proc. IEEE GLOBECOM*, Nov. 2004, pp. 102–107.
- [10] R. A. Iltis and R. Cagley, "Channel estimation and carrier offset control for cooperative MIMO sensor networks," in *Proc. IEEE 39th Asilomar Conf. Signals, Systems, Computing*, Pacific Grove, CA, Nov. 2005, vol. 1, pp. 210–214.
- [11] B. Azimi-Sadjadi and A. Mercado, "Diversity gain for cooperating nodes in multi-hop wireless networks," in *Proc. IEEE VTC—Fall*, Sep. 2004, vol. 2, pp. 1483–1487.
- [12] R. Djapic, A.-J. Van der Veen, and L. Tong, "Synchronization and packet separation in wireless ad hoc networks by known modulus algorithm," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 51–64, Jan. 2005.
- [13] P. Stoica and E. Lindskog, "Space-time block coding for channels with intersymbol interference," in *Proc. IEEE 35th Asilomar Conf. Signals, Systems, Computing*, Pacific Grove, CA, Nov. 2001, vol. 1, pp. 252–256.
- [14] X. Li, "Space-time coded multi-transmission among distributed transmitters without perfect synchronization," *IEEE Signal Process. Lett.*, vol. 11, no. 12, pp. 948–951, Jan. 2005.
- [15] Y. Shang and X.-G. Xia, "Space-time trellis codes with asynchronous full diversity up to fractional symbol delays," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2473–2479, Jul. 2008.
- [16] Z. Li and X.-G. Xia, "A simple Alamouti space-time transmission scheme for asynchronous cooperative systems," *IEEE Signal Process. Lett.*, vol. 14, no. 11, pp. 804–807, 2008.
- [17] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3524–3536, Dec. 2006.
- [18] Y. Jing and H. Jafarkhani, "Using orthogonal and quasi-orthogonal designs in wireless relay networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4106–4118, Nov. 2007.
- [19] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Wireless Commun.*, vol. 7, no. 5, pp. 1907–1916, May 2008.
- [20] T. Cui and C. Tellambura, "Semiblind channel estimation and data detection for OFDM systems with optimal pilot design," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 1053–1062, May 2007.

- [21] T. Wang, Y. Yao, and G. B. Giannakis, "Non-coherent distributed space-time processing for multiuser cooperative transmission," in *Proc. IEEE GLOBECOM*, Nov. 2005, pp. 3738–3742.
- [22] X. Li, Y.-C. Wu, and E. Serpedin, "Timing synchronization in decode-and-forward cooperative communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1444–1455, Apr. 2009.
- [23] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z.-Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4306–4316, Sep. 2008.
- [24] N. Khajehnouri and A. H. Sayed, "Distributed MMSE relay strategies for wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3336–3348, Jul. 2007.
- [25] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.
- [26] Y. Jing and H. Jafarkhani, "Network beamforming using relays with perfect channel information," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2499–2517, Jun. 2009.
- [27] W. Guan and H. W. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, no. 7, pp. 517–519, Jul. 2008.
- [28] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5062–5073, Oct. 2008.
- [29] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.
- [30] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communication systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.
- [31] M. Oerder and H. Meyr, "Digital filter and square timing recovery," *IEEE Trans. Commun.*, vol. 36, pp. 605–612, May 1988.
- [32] J. Riba, J. Sala, and G. Vazquez, "Conditional maximum likelihood timing recovery: Estimators and bounds," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 835–850, Apr. 2001.
- [33] Y.-C. Wu and E. Serpedin, "Design and analysis of feedforward symbol timing estimators based on the conditional maximum likelihood principle," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1908–1918, May 2005.
- [34] M.-O. Pun, M. Morelli, and C.-C. J. Kuo, "Maximum-likelihood synchronization and channel estimation for OFDMA uplink transmissions," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 726–736, Apr. 2006.
- [35] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 1998.
- [36] E. G. Larsson, Y. Selen, and P. Stoica, "Adaptive equalization for frequency-selective channels of unknown length," *IEEE Trans. Veh. Technol.*, vol. 54, pp. 568–579, Mar. 2005.
- [37] M. Gastpar, "On capacity under receive and spatial spectrum-sharing constraints," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 471–487, Feb. 2007.
- [38] M. Gastpar, "On capacity under received-signal constraints," in *Proc. Allerton Conf. Communications, Control, Computing*, Monticello, IL, Oct. 2004, pp. 1322–1331.
- [39] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2399, Sep. 2003.



Xiao Li received the B.Eng. degree (*summa cum laude*) from Sun Yat-sen (Zhongshan) University in 2007 and the M.Phil. degree from the Department of Electrical and Electronic Engineering at the University of Hong Kong in 2009. He is currently working towards the Ph.D. degree at the University of California, Davis.

His current research interests include statistical signal processing, optimization, communications and information theory, with applications in wireless networks and communication systems.



Chengwen Xing received the B.Eng. degree from Xidian University, Xian, China, in 2005. He is currently working towards the Ph.D. degree in the Department of Electrical and Electronic Engineering, the University of Hong Kong, Hong Kong.

His current research interests include statistical signal processing, convex optimization, multivariate statistics, combinatorial optimization, and cooperative communication systems.



Yik-Chung Wu received the B.Eng. (E.E.E.) degree and the M.Phil. degree from the University of Hong Kong (HKU) in 1998 and 2001, respectively, and the Ph.D. degree from the Texas A&M University in 2005. During his study at Texas A&M University, he was fully supported by the prestigious Croucher Foundation scholarship.

After receiving the Master's degree, he was a Research Assistant in the University of Hong Kong. From August 2005 to August 2006, he was a Member of Technical Staff with the Thomson

Corporate Research, Princeton, NJ. Since September 2006, he has been an Assistant Professor with the University of Hong Kong. His research interests are in the general area of signal processing and communication systems, and in particular receiver algorithm design, synchronization techniques, channel estimation, and equalization.

Dr. Wu was a TPC member for the IEEE Vehicular Technology Conference Fall 2005, the IEEE GLOBECOM 2006 and 2008, and the International Conference on Communications (ICC) 2007 and 2008. He is currently an Associate Editor for the IEEE COMMUNICATIONS LETTERS.



S. C. Chan (S'87–M'92) received the B.Sc. (Eng.) and Ph.D. degrees from the University of Hong Kong in 1986 and 1992, respectively.

He joined the City Polytechnic of Hong Kong in 1990 as an Assistant Lecturer and later as a University Lecturer. Since 1994, he has been with the Department of Electrical and Electronic Engineering, University of Hong Kong, where he is currently a Professor. He held visiting positions in Microsoft Corporation, Redmond, CA, Microsoft Research Asia, the University of Texas at Arlington, and the

Nanyang Technological University. His main research area is signal processing and applications. His previous research covers fast transform algorithms, filter design and realization, multirate and array signal processing, adaptive and communications signal processing, data compression, and image-based rendering.

Dr. Chan has served in a number of professional and administrative committees. He is currently a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society, an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I and the *Journal of Signal Processing Systems*, and was the Chairman of the IEEE HK Chapter of Signal Processing. He was the Special Session Chairman of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2003 and is an organizing committee member of the IEEE 2010 ICIP.