# Novel blind full multipath Two Way Relay Network (TWRN), OFDM channel estimation using Machine Learning

Parthapratim De [1]

[1]Inst for Infocomm Research

October 30, 2023

**Abstract**

Most two way relay network (TWRN) half-duplex channel estimation algorithms have been developed for single path channels, except for those in frequency domain OFDM systems. We derive a novel time-domain, blind Maximum A posteriori Probability (MAP) estimation method for multipath estimation in TWRN OFDM systems. Since a TWRN half-duplex system is a cascade of two/more bidirectional transmission systems, there are multiple forward and reverse, individual, as well as composite/cascade (of two individual) channels (unlike *only* one channel in traditional transmission). The situation is further complicated in the case of multipath channels. Additionally, TWRN systems suffer from having noise components at different nodes, including colored (non-white) at the receiver terminal node. Thus most recent research works concentrate on the easier task of estimating single-path (flat frequency) channels, that too by pilot-based, and sometimes, even by suboptimal least squares (LS) methods. However, in this paper, forward, composite/individual mulipath channel estimators developed are semiblind (for enhanced spectral efficiency), and which turn out to be nonlinear. Moreover, an unique "Factor Analysis Alternating Maximization" method (used in psychometrics and some Machine Learning (MLe) applications, but not in signal processing, communication/TWRN systems), is used, in a novel manner, to overcome the colored noise problem, allowing one to derive novel, closed-form, analytical expressions for reverse individual channel $\mathbf{h}$ estimation, (with its convergence provided), which is unavailable in existing literature. Non-trivial Cramer Rao bounds have also been derived for these novel multipath channel estimators. Comprehensive simulation results show the novel forward, reverse, composite and individual channel estimation methods perform much better than the existing ones.

# Novel blind full multipath Two Way Relay Network (TWRN), OFDM channel estimation using Machine Learning

Parthapratim De
Institute for Infocomm Research, Singapore pd4267@yahoo.com

***Abstract*— Most two way relay network (TWRN) half-duplex channel estimation algorithms have been developed for single path channels, except for those in frequency domain OFDM systems. We derive a novel time-domain, blind Maximum A posteriori Probability (MAP) estimation method for multi-path estimation in TWRN OFDM systems. Since a TWRN half-duplex system is a cascade of two/more bidirectional transmission systems, there are multiple forward and reverse, individual, as well as composite/cascade (of two individual) channels (unlike *only* one channel in traditional transmission). The situation is further complicated in the case of multipath channels. Additionally, TWRN systems suffer from having noise components at different nodes, including colored (non-white) at the receiver terminal node. Thus most recent research works concentrate on the easier task of estimating single-path (flat frequency) channels, that too by pilot-based, and sometimes, even by suboptimal least squares (LS) methods. However, in this paper, forward, composite/individual mulipath channel estimators developed are semiblind (for enhanced spectral efficiency), and which turn out to be nonlinear. Moreover, an unique "Factor Analysis Alternating Maximization" method (used in psychometrics and some Machine Learning (MLe) applications, but not in signal processing, communication/TWRN systems), is used, in a novel manner, to overcome the colored noise problem, allowing one to derive novel, closed-form, analytical expressions for reverse individual channel $\mathbf{h}$ estimation, (with its convergence provided), which is unavailable in existing literature. Non-trivial Cramer Rao bounds have also been derived for these novel multipath channel estimators. Comprehensive simulation results show the novel forward, reverse, composite and individual channel estimation methods perform much better than the existing ones.**

## I. Introduction

Advanced 5G and 6G wireless systems are required to deliver very high data rate communications. Two-way relay networks (TWRN) play an important part in building *spectrally efficient* communications. In TWRN amplify-forward (AF) mode, the relay node (in between two/many terminals) amplifies the faded signal it receives from the two terminals. TWRN may thus be considered as a distributed MIMO system, by reducing the number of transmit/receive antennas on the source/destination nodes. Additionally, the relay facilitates a bidirectional system, as it sends data to both the terminals in the same timeslot, thereby improving the spectrally efficiency and information rate. The spectral efficiency can be further enhanced by transmitting only few *pilot* subcarriers and resorting to semiblind data/channel estimation.

This interest in TWRNs has sparked significant research in its various aspects. Starting from joint relay and antenna selection [1], multiple-relay network, it has been extended to asynchronous TWRNs [2]. Advanced research has been performed in timing-offset estimation [3], [4], carrier offset and phase noise estimation in TWRNs [5], [6], [7]. The research has then continued into full-duplex ones (for even more spectrally efficient systems) [8], [9] and expanded to relaying with Massive MIMO (MMIMO) [10] and cloud based multi-way MIMO TWRNS [11]. Distributed beamforming [12], energy efficiency of millimeter-wave TWRNs [13] and time-varying MIMO dual hop relay (though one-directional) channel estimation [14] have also been investigated. One advantage of TWRN AF protocol is that no signal processing is required at the relay ($R$) node, in that $R$ just amplifies its received signal signal and transmits it to both the terminal nodes $T_1$, $T_2$. This results in cheaper, reduced complexity relay components, without sacrificing on transmission latency. However, this comes at the cost of advanced signal processing/hardware at the receiver terminal, as it has to semi-blindly estimate the signal (from the other terminal), for interference cancellation (IC)), as well as estimate the reverse channel (comprising its own channel $\mathbf{h}$ to the relay $R$, assuming channel reciprocity) to enable self-interference cancellation (SIC) caused by its own (*self*) signal. Thus, developing low-complexity and fast TWRN AF algorithms are still an area of significant research, to enable superior, spectrally efficient, reception of high data rate communications in emerging 5G/6G wireless modems.

The concatenation of two/more bidirectional transmission systems, (in TWRN half-duplex systems), involves two/more forward/reverse, individual/composite (cascade of two individual) channels, which is unlike only one individual channel in traditional transmission. This situation is particularly very complicated in long multipath channels, which require careful modelling and estimation of multiple (and even coupled) parameters. A single-input, single-output (SISO) system, with a single antenna (at the relay, as well as the 2 terminals) is considered here. This is a difficult scenario, as multiple antennas in MIMO relays may *facilitate* TWRN channel estimation [15].

An innovative *time-domain* model of a two-way, half-duplex relay TWRN over a *block* based orthogonal division multiple access transmission system (OFDM), in a full multipath channel, is developed here, (as only *few* channel paths have to be estimated in a time-domain method, instead of estimating a large number of subcarriers in a frequency domain approach [16]). Most of existing works concentrate on *single* channel path/tapTWRN systems. Only few works address multipath channel estimation in OFDM TWRN system The alternative TWRN estimators in multipath OFDM channels [16], [9], by neglecting all noise components, resort to least squares (LS) methods, which will shown to be suboptimal (see (10) below), and suffer in performance, in comparison to the optimal ML (let alone MAP) methods, in estimating the reverse channel. Additionally, *individual* channels in [16] are estimated in the frequency domain (see expressions in Sec. III. A. 3, and equation (38) in [16]). This calculation may involve some scalar divisions, which are prone to noise enhancement at channel spectral nulls.

The contributions of the paper are:

i) Development of a novel *time-domain* model of a SISO, two-way half-duplex relay network (TWRN) over a *block* based orthogonal division multiple access transmission system (OFDM), in *full multipath* channels, in contrast to most existing works on (pilot based) single-path/tap (frequency flat) channel estimation (Section II). The time-domain model of multitude of TWRN multipath (i. e. frequency selective) channels require appropriate signal models and estimation techniques.

ii) Formulation of the forward, composite multipath TWRN channels $\mathbf{b}$ in Section III, along with new derivation of the prior probability density function (pdf) of multipath *composite* channels, required for subsequent MAP estimation.

iii) Novel blind, *non-linear*, (full mulipath) TWRN forward composite channel ($\mathbf{b}$) estimator in Section IV, along with estimation of *individual* multipath channel $\mathbf{h}$, $\mathbf{g}$ (from composite channel $\mathbf{b}$) (Section IV. A). Semiblind estimation is achieved by using an EM algorithm (which is more complicated than the one for single-path scalar channel [3]).

iv) Reverse channel estimation is complicated by a colored (channel dependent) noise at the receiver terminal, as shown in equation (10) below and [17], thereby rendering the Least Squares (LS) estimators suboptimal. In spite of that, *existing* TWRN multipath channel estimators [16], [9] still resort to LS methods (to simplify matters), resulting in inferior performance.

v) Though expressions for the ML based single-path, reverse TWRN channel estimator can be derived [17], [3], it is difficult to derive closed form expressions for MAP based reverse channel estimation (*even* for single-path channels), as seen in [5], and for full-duplex single-path TWRN networks [8]. Our paper overcomes this problem by deriving an unique, innovative MAP based reverse multipath channel ($\mathbf{a}$ and $\mathbf{h}$) estimators, using *Factor Analysis Alternating Maximization* (a concept used sometimes in statistics and machine learning, but not used in signal processing and wireless communications), whose performance is superior *over LS methods* (Section V). A closed form expression for the reverse channel estimator is developed, and iterated using a *second* EM algorithm in a novel way, and this estimator's convergence is proved.

v) *Comprehensive* and clear comparisons with existing TWRN channel estimators (Sec. VI),

vi) Derivation of Cramer-Rao bounds for forward composite channel $\mathbf{b}$, and reverse channel $\mathbf{h}$ (Sec. VII),

vii) Simulation Results, illustrating the superior performance of the novel estimators, over existing methods (Sec. VIII), for varying number of subcarriers, channel length, and varying number of OFDM blocks for all forward and reverse independent and composite channels, along with estimating time offsets and including CP-OFDM TWRN channels

*Notations:* Bold upper-case symbols $\mathbf{A}$ denote matrices. Bold lower-case symbols $\mathbf{b}$ denote vectors. $\mathbf{I}_i$ is an identity matrix of size $i \times i$, $\mathbf{0}_{j,k}$ is a $j \times k$-sized zero matrix. Also, $\mathbf{A}(i:j, k:l)$ denotes the $i$th to $j$th rows and $k$th to $l$th columns of the matrix $\mathbf{A}$. For any vector $\mathbf{s}$ and matrix $\mathbf{A}$, $\|\mathbf{s}\|^2_{\mathbf{A}}$ denotes $\mathbf{s}^H\mathbf{A}\mathbf{s}$. $\otimes$ denotes the Kronecker product.

## II. System Model

The TWRN considered in this paper consists of two terminals $T1$, $T2$ and one relay node $R$, though extension to multiple relays and terminals is simpleforward. Consider a OFDM system, with $\mathbf{s}_i^{(k)} = [s_i^{(k)}(N-1), \ldots, s_i^{(k)}(0)]^T$ be the $i$th transmitted OFDM block's data from $k$th terminal; $k = 1, 2$ are the two terminals. ($N$ is the discrete Fourier transform (DFT) size). The time-domain signals $\mathbf{u}_i^{(k)} := [u_i^{(k)}(N-1), \ldots, u_i^{(k)}(0)]^T$ are obtained by taking the inverse DFT (IDFT) of $\mathbf{s}_i^{(k)}$. The $i$th transmitted OFDM block symbol consists of $\mathbf{u}_i^{(k)}$, padded with a guard interval of $Z$ ($\geq 0$) zero samples (or preceded with a cyclic prefix). The guard interval enables simple subcarrier-by-subcarrier equalization in the frequency domain at the receiver, provided that the guard interval length is greater than or equal to the channel delay spread $L$; the $i$th transmitted OFDM block $\mathbf{d}_i$ is of size $P = N + Z = N + L$. Thus $\mathbf{d}_i^{(k)} = [\mathbf{u}^{(k)T}_i, \mathbf{0}_{1,L}]^T \in \mathcal{C}^P$. A few (pilot) subcarriers in $\mathbf{d}_i^{(2)}$, say 7 (out of 64 in one OFDM block/symbol) subcarriers, are assumed to be known at $T_1$ )pilots), but others $64 - 7 = 57$ are unknown (blind subcarriers).

**Note** Extension to OFDM with cyclic-prefix(CP) can be done, as in [18], where a CP system, after a specific signal remodulation (using 2 consecutive OFDM blocks in equation (12), [18]), can be converted into a zero-padded (ZP) system (also see equations (11) to (14), Section III.A., [18]). Thus we consider ZP systems only here.

Define the $i$th OFDM block's, $P \times (L+1)$-sized, transmitted data matrix from Terminal $k$, ($k = 1, 2$ terminals),

$$\mathbf{D}_i^{(k)}(P-1) = \begin{bmatrix} d_i^{(k)}(P-1) & 0 & \cdots & 0 \\ d_i^{(k)}(P-2) & d_i^{(k)}(P-1) & 0 & 0 \\ & & \ddots & \\ 0 & \cdots & d_i^{(k)}(0) & d_i^{(k)}(1) \\ 0 & \cdots & & d_i^{(k)}(0) \end{bmatrix}$$
$$\in \mathcal{C}^{P \times (L+1)}, \quad (1)$$

where $d_i^{(k)}(n) = 0, N \leq n \leq P-1$. The received signal (at relay $R$), $P \times 1$-sized $\bar{\mathbf{r}}_i$, (with sufficient guard interval), is

$$\begin{aligned} \bar{\mathbf{r}}_i &\stackrel{\Delta}{=} [r_i(P-1)\, r_i(P-2) \cdots,\ r_i(0)]^T \\ &= \mathbf{D}_i^{(1)}(P-1)\mathbf{h} + \mathbf{D}_i^{(2)}(P-1)\mathbf{g} + \mathbf{n}_r, \end{aligned} \quad (2)$$

where $\mathbf{h} = [h_L\, h_{L-1} \cdots h_0]^T$ is $(L+1)$-path multipath channel from $T1$ to $R$; $\mathbf{g} = [g_L\, g_{L-1} \cdots g_0]^T$ is $(L+1)$-path multipath channel from $T2$ to $R$, $\mathbf{n}_r$ is the AWGN at $R$. Relay $R$ then *adds* another zero padding of $L$ null subcarriers to $\bar{\mathbf{r}}_i$, and sends out a signal $\mathbf{r}_i = [\bar{\mathbf{r}}_i^T, \mathbf{0}_{1,L}]^T \in \mathcal{C}^{\tilde{P} \times 1}$, $\tilde{P} = P + L$, to both the terminals $T1$, $T2$. Note that $r_i(n) = 0$, $P \leq n \leq \tilde{P} - 1$. Defining the data matrix (corresponding to $\mathbf{r}_i(n)$) as

$$\mathbf{D}_i^{(r)}(\tilde{P}-1) = \begin{bmatrix} r_i(\tilde{P}-1) & 0 & \cdots & 0 \\ r_i(\tilde{P}-2) & r_i(\tilde{P}-1) & 0 & \cdots \\ & & \ddots & \\ 0 & 0 & \cdots & r_i(0) \end{bmatrix}$$
$$\begin{aligned} &\in \mathcal{C}^{(\tilde{P}\times(L+1)} = Toeplitz([r_i(\tilde{P}-1)\, r_i(\tilde{P}-2)\, \cdots\, 0]), \\ &= [\tilde{\mathbf{D}}_i^{(1)}(\tilde{P}-1)\mathbf{h}|\,\tilde{\mathbf{D}}_i^{(1)}(\tilde{P})\mathbf{h}|\, \cdots, |\tilde{\mathbf{D}}_i^{(1)}(\tilde{P}+L-1)\mathbf{h}] \\ &+ [\tilde{\mathbf{D}}_i^{(2)}(\tilde{P}-1)\mathbf{g}|\,\tilde{\mathbf{D}}_i^{(2)}(\tilde{P})\mathbf{g}|\, \cdots\, \tilde{\mathbf{D}}_i^{(2)}(\tilde{P}+L-1)\mathbf{g}] \\ &+ [\mathbf{n}_r(\tilde{P}-1)|\,\mathbf{n}_r(\tilde{P}), |\, \cdots |\mathbf{n}_r(\tilde{P}+L-1)], \end{aligned} \tag{3}$$

where $Toeplitz(\mathbf{x})$ is a Toeplitz matrix with $\mathbf{x}$ as its first column, and (defining $\tilde{\mathbf{D}}_i^{(k)}(n)$ as a $\tilde{P}\times(L+1)$-sized extension of $\mathbf{D}_i^{(k)}(n)$). Then the received signal, at terminal $T1$, (from the relay $R$), is $\mathbf{z}_i(n)$, given by

$$\begin{aligned} \mathbf{z}_i &\stackrel{\Delta}{=} [z_i(\tilde{P}-1)\, z_i(\tilde{P}-2)\, \cdots,\, z_i(0)]^T \\ &= \mathbf{D}_i^{(r)}(\tilde{P}-1)[h_L\, h_{L-1}\, \cdots\, h_0]^T + \mathbf{n}_1 \in \mathcal{C}^{\tilde{P}\times 1}, \end{aligned} \tag{4}$$

where $\mathbf{n}_1$ is the AWGN at $T1$.

Three general assumptions are made as follows:

($\mathcal{A}1$) The symbol sequence of each terminal $k$, $s^{(k)}(n)$ is temporally white with zero mean and unit variance, and is statistically uncorrelated with $s^{(k)}(n-m)$ for $m \neq 0$.

($\mathcal{A}2$) The two noise sequences $w(n)$ are both stationary, and temporally and spatially white with zero mean and variance $\sigma^2$.

($\mathcal{A}3$) The symbol sequences $s^{(1)}(n)$ and $s^{(2)}(n)$ are statistically uncorrelated with each other, and also with the noise sequences $w(n)$.

Note that because the IDFT matrix is unitary, the assumptions ($\mathcal{A}1$) and ($\mathcal{A}3$) also hold for the time-domain sequences $u_i^{(k)}(n)$.

## III. Formulation of the forward, composite, mulipath TWRN channel $\mathbf{b}$

The system model, above, is further extended. Equation (4) gives

$$\begin{aligned} \mathbf{z}_i &= [\tilde{\mathbf{D}}_i^{(1)}(\tilde{P}-1)\mathbf{h}|\cdots|\tilde{\mathbf{D}}_i^{(1)}(\tilde{P}+L-1)\mathbf{h}]\mathbf{h} \\ &+ [\tilde{\mathbf{D}}_i^{(2)}(\tilde{P}-1)\mathbf{g}|\cdots|\mathbf{D}_i^{(2)}(\tilde{P}+L-1)\mathbf{g}]\mathbf{h} \\ &+ [\tilde{\mathbf{n}}_r(\tilde{P}-1)|\cdots|\tilde{\mathbf{n}}_r(\tilde{P}+L-1)]\mathbf{h} + \mathbf{n}_1 \\ &= [\tilde{\mathbf{D}}_i^{(1)}(\tilde{P}-1)|\cdots|\tilde{\mathbf{D}}_i^{(1)}(\tilde{P}+L-1)] \begin{bmatrix} \mathbf{h} & \mathbf{0} & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{h} \end{bmatrix} \mathbf{h} \\ &+ [\tilde{\mathbf{D}}_i^{(2)}(\tilde{P}-1)|\,\cdots|\,\tilde{\mathbf{D}}_i^{(2)}(\tilde{P}+L-1)] \begin{bmatrix} \mathbf{g} & \mathbf{0} & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{g} \end{bmatrix} \mathbf{h} \\ &+ [\tilde{\mathbf{n}}_r(\tilde{P}-1)|\,|\cdots|\tilde{\mathbf{n}}_r(\tilde{P}+L-1)])\,\mathbf{h} + \mathbf{n}_1. \end{aligned} \tag{5}$$

Then defining $\bar{\bar{\mathbf{D}}}_i^{(k)} \stackrel{\Delta}{=} [\tilde{\mathbf{D}}_i^{(1)}(\tilde{P}-1)|\,\cdots|\tilde{\mathbf{D}}_i^{(k)}(\tilde{P}-1+L)] \in \mathcal{C}^{\tilde{P}\times(L+1)^2}$, $k=1,2$, equation (5) becomes

$$\mathbf{z}_i = \bar{\bar{\mathbf{D}}}_i^{(1)} \begin{bmatrix} \mathbf{h} & \mathbf{0} & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{h} \end{bmatrix} \mathbf{h} + \bar{\bar{\mathbf{D}}}_i^{(2)} \begin{bmatrix} \mathbf{g} & \mathbf{0} & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{g} \end{bmatrix} \mathbf{h} + \tilde{\mathbf{n}}, \tag{6}$$

We finally have

$$\begin{aligned} \mathbf{z}_i &= \bar{\bar{\mathbf{D}}}_i^{(1)} \begin{bmatrix} h_L\mathbf{h} \\ h_{L-1}\mathbf{h} \\ \vdots \\ h_0\mathbf{h} \end{bmatrix} + \bar{\bar{\mathbf{D}}}_i^{(2)} \begin{bmatrix} g_L\mathbf{h} \\ g_{L-1}\mathbf{h} \\ \vdots \\ g_0\mathbf{h} \end{bmatrix} \\ &+ ([\mathbf{n}_r(\tilde{P}-1)\, \cdots\, \mathbf{n}_r(\tilde{P}+L-1)]\mathbf{h} + \mathbf{n}_1), \\ &= \bar{\bar{\mathbf{D}}}_i^{(1)}(\mathbf{h}\otimes\mathbf{h}) + \bar{\bar{\mathbf{D}}}_i^{(2)}(\mathbf{g}\otimes\mathbf{h}) + \bar{\mathbf{n}}, \end{aligned} \tag{7}$$

where $\otimes$ is the Kronecker product. The associated composite channel vectors, defined as $\tilde{\mathbf{a}} = (\mathbf{h}\otimes\mathbf{h})$, $\tilde{\mathbf{b}} = (\mathbf{g}\otimes\mathbf{h})$, are both $(L+1)^2\times 1$-sized vectors. $\mathbf{n}\stackrel{\Delta}{=}[\bar{\mathbf{n}}_r(n)\,\cdots\,\bar{\mathbf{n}}_r(n-L)])\mathbf{h}+\mathbf{n}_1$ is the $\tilde{P}\times 1$ (overall) noise vector. However, in equation (7), *only* $\bar{L} = (2L+1)$ (incidentally, the length of the convolution of two $(L+1)$ sequences) columns of $\bar{\bar{\mathbf{D}}}_i^{(k)}$, $k=1,2$ are linearly independent. The linearly independent columns of $\bar{\bar{\mathbf{D}}}_i^{(k)}$ are its $1, 2, \cdots, (L+1), 2(L+1), 3(L+1), \cdots, (L+1)^2$-th columns. Thus we define modified data matrices $\bar{\mathbf{D}}_i^{(k)} \stackrel{\Delta}{=} \bar{\bar{\mathbf{D}}}_i^{(k)}(:, 1, 2, \cdots, (L+1), 2(L+1), 3(L+1), \cdots, (L+1)^2)$. The corresponding parameters $\bar{\mathbf{a}}/\bar{\mathbf{b}}$ include a subset of $\bar{L}$ components of $\tilde{\mathbf{a}}/\tilde{\mathbf{b}}$. Moreover, $\bar{\mathbf{a}}$, is given by

$$\begin{aligned} \bar{\mathbf{a}} &\stackrel{\Delta}{=} \begin{bmatrix} 1 & \mathbf{0}_{L^2+2L} & & & & & & & & & \\ 0 & 1 & 0_3 & 1\mathbf{0}_{L^2+2L-5} & & & & & & & \\ 0_L & 1 & 0_{L-1} & 1 & 0_{L-1} & 1 & 0_{L-1} & 1 & 0_{L-1} & 1 \\ \vdots & & & & & & & & & \end{bmatrix} \tilde{\mathbf{a}} \\ &= \mathbf{T}(\mathbf{h}\otimes\mathbf{h}) = (\mathbf{h}*\mathbf{h}), \end{aligned} \tag{8}$$

where $\mathbf{h}*\mathbf{h} \in \mathcal{C}^{\bar{L}\times 1}$ is the *linear* convolution of $\mathbf{h}$ with $\mathbf{h}$.

**Note**: The Kronecker product vectors $\tilde{\mathbf{a}} = (\mathbf{h}\otimes\mathbf{h})$ and $\tilde{\mathbf{b}}$ consist of $(L+1)^2$ elements, each of which is a single term, consisting of a pairwise product of elements of $\mathbf{h}/\mathbf{g}$ with $\mathbf{h}$ (see equation (7)). $\mathbf{T}$ is the transformation matrix which sums up some elements of $\tilde{\mathbf{a}}(\tilde{\mathbf{b}})$ to form a *reduced* subset of $\bar{L}$ elements, which is actually the convolution $\bar{\mathbf{a}} = \mathbf{h}*\mathbf{h}$. Equation (8) thus provides a relation between $\otimes$ and $*$ operators. The *complex valued* $\bar{L}\times 1$ vector $\bar{\mathbf{a}}$ can be expressed in terms of its magnitude and phase components by

$$\bar{\mathbf{a}} = diag(\mathbf{a})[e^{j\psi_{\mathbf{a}}}],\ \bar{\mathbf{b}} = diag(\mathbf{b})[e^{j\psi_{\mathbf{b}}}], \tag{9}$$

where $\mathbf{a}$, $\mathbf{b}$ and $\psi_{\mathbf{a}}, \psi_{\mathbf{b}}$ are the magnitudes and phases of the composite channels $\mathbf{a}, \mathbf{b}$ respectively. Then the magnitude of equation (7) becomes

$$\begin{aligned} \mathbf{z}_i &= \bar{\mathbf{D}}_i^{(1)}(\mathbf{h}*\mathbf{h}) + \bar{\mathbf{D}}_i^{(2)}(\mathbf{g}*\mathbf{h}) + (\mathbf{H}\mathbf{n}_r + \mathbf{n}_1) \\ &= \bar{\mathbf{D}}_i^{(1)}\mathbf{a} + \bar{\mathbf{D}}_i^{(2)}\mathbf{b} + (\mathbf{H}\mathbf{n}_r + \mathbf{n}_1), \end{aligned} \tag{10}$$

where $\mathbf{H}$ is the Toeplitz matrix with $\mathbf{h}$ as its first column. (The phases of all variables in (10), including composite channels $\bar{\mathbf{a}}, \bar{\mathbf{b}}$, will be dealt with separately). The overall noise $\mathbf{n} = (\mathbf{H}\mathbf{n}_r + \mathbf{n}_1)$ then depends on the multipath channel $\mathbf{h}$, (as also in single-path channels [17]), has a non-diagonal correlation matrix, $\mathbf{C} = (\mathbf{H}\mathbf{H}^H + \mathbf{I})\sigma^2$, making it a colored noise. The least-squares (LS) solutions, by neglecting all noise terms, is then *not* equivalent to the optimal ML estimate, and thus inadequate to solve the relay composite channel estimation problem [17].

### A. Derivation of prior pdfs of composite channels

Next, to obtain the *maximum aposteriori probability (MAP)* estimate of the associated channels, the joint prior probability density function (pdf) of composite channels (in (7)), has to be determined [5], [19]. Recall that in [5], only the prior pdf of scalar (single-path) channel parameters $a = h_0^2$ and $b = g_0 h_0$ needed to be determined, while here the composite channels are convolutions $\mathbf{b} = \mathbf{g} * \mathbf{h}$, $\mathbf{a} = \mathbf{h} * \mathbf{h}$, determining whose apriori pdfs, is considerably more complicated. For *simplicity and ease of presentation*, consider two $3\,(L+1=3)$-path individual channels whose magnitudes are $\mathbf{h} = [h_3,\, h_2,\, h_1]$ and $\mathbf{g} = [g_3,\, g_2,\, g_1]$ respectively (their phases will be determined separately). Then magnitudes of composite channels $\mathbf{a} = \mathbf{h} * \mathbf{h} = [a_1, a_2, \cdots, a_{\bar{L}}]^T$, $\mathbf{b} = \mathbf{g} * \mathbf{h} = [b_1, b_2, \cdots, b_{\bar{L}}]^T$, given by

$$
\begin{aligned}
&a_1 = h_1^2,\, a_2 = 2h_1h_2,\, a_3 = 2h_1h_3 + h_2^2,\, a_4 = 2h_2h_3, \\
&a_5 = h_3^2 . b_1 = h_1g_1,\, b_2 = h_2g_1 + h_1g_2,\, b_3 = g_1h_3 + g_2h_2 \\
&\quad + g_3h_1, b_4 = g_2h_3 + g_3h_2,\, b_5 = g_3h_3.
\end{aligned}
$$

*Lemma 1*: The logarithm of the joint aprior pdf, $f(\mathbf{a}, \mathbf{b})$, of the magnitudes of the composite channels $\mathbf{a}, \mathbf{b}$, is given by

$$
\begin{aligned}
&log f(\mathbf{a},\mathbf{b}) = log f(\mathbf{a}) log f(\mathbf{b}|\mathbf{a}) = log f(\mathbf{a}) + log(\frac{2b_1}{a_1 v}) \\
&- \frac{b_1^2}{a_1 v} + log(\frac{2(b_2\sqrt{a_1} - \alpha b_1)}{a_1 v}) - \frac{(b_2\sqrt{a_1} - \alpha b_1)^2}{a_1 v} \\
&+ log(\frac{2(b_3 a_1 - \bar{\beta})}{a_1 v}) - \frac{(b_3 a_1 - \bar{\beta})^2}{a_1^2 v} \\
&+ log([2(b_4\sqrt{a_1} - \sqrt{a_5}\beta)/(\alpha^2\sqrt{a_1}v)]) - (b_4\sqrt{a_1} \\
&- \sqrt{a_5}\beta)^2/(\alpha^4 a_1 v) + log(2b_5/(a_5 v)) - b_5^2/(a_5 v). \qquad (11)
\end{aligned}
$$

Proof: See Appendix B. The results can be *easily extended* to arbitrary values of $L+1$, (channel length), as illustrated in Appendix B. The method (in Appendix B) can be programmed on a computer for any generic value of $L$

Then the gradient of the log prior-pdf, $\bigtriangledown_{\mathbf{b}}[log f(\mathbf{b}|\mathbf{a})]$, is

$$
\bigtriangledown_{\mathbf{b}}\,[log f(\mathbf{b}|\mathbf{a})] = \begin{bmatrix} 1/b_1 \\ \sqrt{a_1}/(b_2\sqrt{a_1} - \alpha b_1) \\ a_1/(b_3 a_1 - \bar{\beta}) \\ \sqrt{a_1}/(b_4\sqrt{a_1} - \sqrt{a_5}\beta) \\ 1/b_5 \end{bmatrix}
+ \begin{bmatrix} -(2b_1/(a_1 v))[1 - \alpha\sqrt{a}_1 + (\sqrt{a}_5 - \alpha^2)/(a_1) - \sqrt{a}_5/\alpha^3] \\ -(2b_2/v)[1 - (\alpha/a_1^2) + (a_5/(\alpha^4 a_1))] \\ -2b_3/(a_1 v) \\ -2b_4/(a_1 v) \\ -2b_5/(a_1 v) \end{bmatrix}, \qquad (12)
$$

## IV. Derivation of novel blind, full mulipath TWRN forward, composite channel ($\mathbf{b}$) *Non-Linear* estimator

Defining $\bar{\mathbf{z}}_i \triangleq [\mathbf{z}_i - \bar{\mathbf{D}}_i^{(1)}\mathbf{a}] = \bar{\mathbf{D}}_i^{(2)}\mathbf{b} + \mathbf{n}$. First, we assume that the composite channel parameter $\mathbf{a} = (\mathbf{h} * \mathbf{h})$ is known from the last iteration, from which $\mathbf{h}$ can be determined (by deconvolution), or as in Appendix A. Then $\mathbf{H}$ is formed and used to compute the noise correlation matrix $\mathbf{C}$. Also, terminal $T1$ knows its own transmitted data $\bar{\mathbf{D}}_i^{(1)}$. Thus $\bar{\mathbf{z}}_i$ can be computed. MAP estimation of the composite channel $\mathbf{b}$ requires the *aposteriori* pdf of $\mathbf{b}$, (given data $\bar{\mathbf{z}}_i$, and $\mathbf{a}$ (from last iteration)). This is given by

$$
\begin{aligned}
f(\mathbf{b}|\bar{\mathbf{z}}_i, \mathbf{a}) &= \frac{f(\bar{\mathbf{z}}_i|\mathbf{b},\mathbf{a}) f(\mathbf{b}|\mathbf{a})}{f(\mathbf{z}_i|\mathbf{a})} \\
&= \frac{e^{-\frac{1}{2}[\bar{\mathbf{z}}_i - \bar{\mathbf{D}}_i^{(2)}\mathbf{b}]^H \mathbf{C}^{-1}[\bar{\mathbf{z}}_i - \bar{\mathbf{D}}_i^{(2)}\mathbf{b}]} \times f(\mathbf{b}|\mathbf{a})}{f(\mathbf{z}_i|\mathbf{a})} \qquad (13)
\end{aligned}
$$

The *aposteriori* pdf, $f(\mathbf{b}|\bar{\mathbf{z}}_i, \mathbf{a})$, is then maximized with respect to (w.r.t) $\mathbf{b}$, to find MAP estimate [20]- [26]. This is equivalent to maximizing $\mathcal{L} \triangleq \log_e(f(\bar{\mathbf{z}}_i|\mathbf{b},\mathbf{a}) f(\mathbf{b}|\mathbf{a}))$ w.r.t $\mathbf{b}$. Then

$$
\mathcal{L} = -\frac{1}{2}[\bar{\mathbf{z}}_i - \bar{\mathbf{D}}_i^{(2)}\mathbf{b}]^H \mathbf{C}^{-1}[\bar{\mathbf{z}}_i - \bar{\mathbf{D}}_i^{(2)}\mathbf{b}] + log(f(\mathbf{b}|\mathbf{a})), \qquad (14)
$$

with $log(f(\mathbf{b}|\mathbf{a}))$ given in (11). However, maximizing (14), at terminal $T1$, with respect to the composite channel parameter $diag(\mathbf{b})$, also requires knowledge of $\bar{\mathbf{D}}_i^{(2)}$, $i$th OFDM data block transmitted from *other terminal* $T2$ to $R$. Since data $\bar{\mathbf{D}}_i^{(2)}$ is unknown at $T1$ [27]- [29], an Expectation-Maximization (EM) algorithm is applied to the likelihood function $\mathcal{L}$ (in (14). An EM algorithm is used for TWRN relay channel estimation in [3], but *only single path/tap* channels are considered, which simplifies the situation considerably. Moreover, [3] does *not* exploit the MAP optimality criterion. Since $\bar{\mathbf{D}}_i^{(2)}$ is unknown at $T1$, the conditional expectation of the likelihood function (14) (over the unknown data $\bar{\mathbf{D}}_i^{(2)}$, given $\bar{\mathbf{z}}_i$), i. e. the E-step in EM algorithm, is computed,

$$
\begin{aligned}
&E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}\{\mathcal{L}\} = -\frac{1}{2} E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}\{[\bar{\mathbf{z}}_i - \bar{\mathbf{D}}_i^{(2)}\mathbf{b}]^H \mathbf{C}^{-1}[\bar{\mathbf{z}}_i - \bar{\mathbf{D}}_i^{(2)}\mathbf{b}]\} \\
&- \tilde{P}\log_e(\det(\mathbf{C}) - E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}\{log(f(\mathbf{b}|\mathbf{a})\}. \qquad (15)
\end{aligned}
$$

However,*if a ML* criterion based estimator is employed, then there is no priori pdf $\{log(f(\mathbf{b}|\mathbf{a})\}$ term, so that its likelihood function becomes

$$
\begin{aligned}
&E^{ML}_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}\{\mathcal{L}\} = -\frac{1}{2} E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}\{[\bar{\mathbf{z}}_i - \bar{\mathbf{D}}_i^{(2)}\mathbf{b}]^H \mathbf{C}^{-1}[\bar{\mathbf{z}}_i - \bar{\mathbf{D}}_i^{(2)}\mathbf{b}]\} \\
&- \tilde{P}\log_e \det(\mathbf{C}) \qquad (16)
\end{aligned}
$$

The 1st term of (15) is

$$
\begin{aligned}
&-\frac{1}{2}([\bar{\mathbf{z}}_i^H \mathbf{C}^{-1}\bar{\mathbf{z}}_i] + \mathbf{b}^H E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}[\bar{\mathbf{D}}_i^{(2)H}\mathbf{C}^{-1}\bar{\mathbf{D}}_i^{(2)}]\mathbf{b} \\
&- 2\Re\{\bar{\mathbf{z}}_i^H \mathbf{C}^{-1} E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}[\bar{\mathbf{D}}_i^{(2)}]\mathbf{b}\}), \qquad (17)
\end{aligned}
$$

where two Expectation terms, namely, $E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}[\bar{\mathbf{D}}_i^{(2)H}\mathbf{C}^{-1}\bar{\mathbf{D}}_i^{(2)}]$ and $E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}[\bar{\mathbf{D}}_i^{(2)}]$ have to be derived, for which the required pdfs have to be determined. We assume that the transmitted discrete data $d_i^{(2)}$ is independent, uniformly distributed, with pdf $f(d_i^{(2)} = A_l) = \frac{1}{M}$, $A_l = 1, 2, \cdots, M$. Then each column $c$ of $\bar{\mathbf{D}}_i^{(2)}$, denoted by $[\bar{\mathbf{D}}^{(2)}]_c$, is a $\tilde{P}$-sized column vector, with only $N$ non-zero random entries $\mathbf{A}_l = [A_{l,0}\, A_{l,1}\, \cdots, A_{l,N-1}]^H$ (for the $l$th random experiment).

*Lemma 2*: Defining for the $i$th OFDM block's, $\mathbf{X}_i \triangleq \sum_{l=1}^{M^N} [\mathbf{A}_{l,i}]_{\bar{L}columns} \frac{A}{B}$, (15) becomes

$$
\begin{aligned}
&E_{\bar{\mathbf{D}}_i^{(2)}}\{\mathcal{L}\} = \bar{\mathbf{z}}_i^H \mathbf{C}^{-1}\bar{\mathbf{z}}_i - 2\Re\{\bar{\mathbf{z}}_i^H \mathbf{C}^{-1}\mathbf{X}\mathbf{b}\} \\
&+ (\frac{B}{A})\mathbf{b}^H(\sum_{l=1}^{M^N}[\mathbf{X}\mathbf{C}^{-1}\mathbf{X}^H])\mathbf{b} + other\, terms. \qquad (18)
\end{aligned}
$$

Proof: See Appendix C.

Obviously, minimizing $E_{\bar{\mathbf{D}}_i^{(2)}}\{\mathcal{L}\}$ involves maximizing $2\Re\{\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}\mathbf{b}\}$, the 2nd term in (18). $\{\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}\mathbf{b}\}$ has the maximum value, when it is real (i. e. does *not* have any imaginary component), which then gives $2\Re\{\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}\}\mathbf{b} = 2\|\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}diag(\mathbf{b})]$. Then the term $2\Re\{\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}\mathbf{b}\}$ in (18) can be re-written as

$$2\Re\{\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}\mathbf{b}\} = 2\Re\{\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}diag(\mathbf{b})[e^{j\psi_{\mathbf{b}}}]\}, \quad (19)$$

where $[e^{j\psi_{\mathbf{b}}}]$ is a $\tilde{L}\times 1$ vector. By estimating $\psi_{\mathbf{b}}$ as

$$\psi_{\mathbf{b}} = -\angle\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}, \quad (20)$$

we have

$$\begin{aligned}2\Re\{\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}\mathbf{b}\} &= 2\Re\{\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}diag(\mathbf{b})[e^{j\psi_{\mathbf{b}}}]\}\\ &= 2\|\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}diag(\mathbf{b})\|. \quad (21)\end{aligned}$$

With the choice of $\psi_{\mathbf{b}}$ in (20), $A$ becomes

$$\begin{aligned}A &= exp^{-\frac{1}{2}\sum_{l=1}^{\bar{L}}[\bar{\mathbf{z}}_i-\mathbf{A}_l\mathbf{b}]^H\mathbf{C}^{-1}[\bar{\mathbf{z}}_i-\mathbf{A}_l\mathbf{b}]}\\ &= exp^{-\frac{1}{2}\sum_{l=1}^{\bar{L}}[\|\bar{\mathbf{z}}_i^H\|^2_{\mathbf{C}^{-1}}+2\Re\{\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}\mathbf{b}\}+\|\mathbf{b}^H\mathbf{X}\|^2_{\mathbf{C}^{-1}}]}\\ &= exp^{-\frac{1}{2}\sum_{l=1}^{\bar{L}}[\|\bar{\mathbf{z}}_i^H\|^2_{\mathbf{C}^{-1}}+2\|\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}diag(\mathbf{b})\|+\|\mathbf{b}^H\mathbf{X}]\|^2_{\mathbf{C}^{-1}}]}\\ B &= \sum_{t=1}^{M^N}[exp^{-\frac{1}{2}[\|\bar{\mathbf{z}}_i^H\|^2_{\mathbf{C}^{-1}}+2\|\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}diag(\mathbf{b})\|+\|\mathbf{b}^H\mathbf{X}]\|^2_{\mathbf{C}^{-1}}]}]^{\bar{L}},\\ \mathbf{X} &= \sum_{l=1}^{M^N}\mathbf{A}_l\frac{A}{B} \quad (22)\end{aligned}$$

Then both $A$ and $B$ are real and positive.

It is to be noted that Hermitian, non-negative definite noise correlation matrix $\mathbf{C}$ has non-negative (possibly positive) eigen-values $\{\lambda_i\}$'s; $\det(\mathbf{C}) = \prod_k\lambda_k$, $\lambda_k - \mathbf{C}'s$ kth eigen-value, thus its determinant is also non-negative ( possibly positive). Thus, $\mathbf{C}$ permits a Cholesky decomposition $\mathbf{C} = \mathbf{C}_T\mathbf{C}_T^H$, $\mathbf{C}_T$being the Cholesky factor. Now, $\psi_{\mathbf{b}}$ (in equation (20) can be re-written as

$$\psi_{\mathbf{b}} = -\angle\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}. \quad (23)$$

Using trace operator's properties, $a = \text{trace}(a), a : scalar$, $\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA})$, and applying (23), (22), the 3rd term in (18) is

$$\begin{aligned}&\sum_{l=1}^{M^N}\text{trace}\{\mathbf{b}^H[\mathbf{A}_l\mathbf{C}^{-1}\mathbf{A}_l^H]\mathbf{b}\}\\ &= \frac{\bar{A}}{B}\sum_{l=1}^{M^N}\text{trace}\{\|diag(\mathbf{b})^2\mathbf{A}_l(\mathbf{C}_T^H)^{-1}\|^2\}, \quad (24)\end{aligned}$$

where $\bar{A} = A(e^{\|\mathbf{z}_i^H\mathbf{C}^{-1}\mathbf{A}_l\|^2})$.

The gradient of the log of prior-pdf, $\bigtriangledown_{\mathbf{b}}[logf(\mathbf{b}|\mathbf{a})]$, in (12), has reciprocal coefficients of $\{1/b_i\}$ in the first term on its RHS, and coefficients of its linear term $b_i$ in the 2nd term on its RH Using matrix identities $\frac{\partial\text{trace}(\mathbf{AX})}{\partial\mathbf{X}} = \mathbf{A}$, $\frac{\partial|\mathbf{X}|}{\partial\mathbf{X}} = |\mathbf{X}|\mathbf{X}^{-1}$, $\frac{(\partial\mathbf{X}^H\mathbf{AX}}{\partial\mathbf{X})} = 2\mathbf{AX}$ (for Hermitian $\mathbf{A}$), [30].

Differentiation of (18), with respect to the parameter $\mathbf{b}$ gives

$$\begin{aligned}\frac{\partial E_{\bar{\mathbf{D}}_i^{(2)}}\{\mathcal{L}\}}{\partial\mathbf{b}} &= 2\frac{A}{B}\sum_{l=1}^{M^N}[\mathbf{A}_l^H\mathbf{C}^{-1}\mathbf{A}_l\mathbf{b} - (|\bar{\mathbf{z}}_i^H\mathbf{A}_l\mathbf{C}^{-1}|)]\\ &+ \begin{bmatrix}1/b_1\\ \sqrt{a_1}/(b_2\sqrt{a_1}-\alpha b_1)\\ a_1/(b_3a_1-\bar{\beta})\\ \sqrt{a_1}/(b_4\sqrt{a_1}-\sqrt{a_5}\beta)\\ 1/b_5\end{bmatrix}\\ &+ \begin{bmatrix}-(2b_1/(a_1v))[1-\alpha\sqrt{a}_1+(\sqrt{a}_5-\alpha^2)/(a_1)-\sqrt{a}_5/\alpha^3]\\ -(2b_2/v)[1-(\alpha/a_1^2)+(a_5/(\alpha^4a_1))]\\ -2b_3/(a_1v)\\ -2b_4/(a_1v)\\ -2b_5/(a_1v)\end{bmatrix},\end{aligned} \quad (25)$$

which reduces to

$$\begin{aligned}\frac{\partial E\{\mathcal{L}\}}{\partial\mathbf{b}} &= 2\frac{A}{B}\sum_{l=1}^{M^N}[\mathbf{A}_l^H\mathbf{C}^{-1}\mathbf{A}_l\mathbf{b} - (|\bar{\mathbf{z}}_i^H\mathbf{A}_l\mathbf{C}^{-1}|)] + \mathbf{F}\frac{1}{\mathbf{b}} + \mathbf{G}\mathbf{b},\\ &\mathbf{F} = diag(1, \sqrt{a_1}/[\sqrt{a_1}-\alpha(b_1^{(l-1)}/b_2^{(l-1)})],\\ &a_1/[a_1-(\bar{\beta}/b_3^{(l-1)})], \sqrt{a_1}/[\sqrt{a_1}-(\sqrt{a_5}\beta/b_4^{(l-1)})], 1],\\ &\mathbf{G} = diag(-(2/(a_1v))[1-\alpha\sqrt{a}_1+(\sqrt{a}_5-\alpha^2)/(a_1)-\\ &\sqrt{a}_5/\alpha^3], -(2/v)[1-(\alpha/a_1^2)+(a_5/(\alpha^4a_1))], -2/(a_1v),\\ &-2/(a_1v)-2/(a_1v)), \quad (26)\end{aligned}$$

where the superscript $^{(l-1)}$ refers to the earlier $(l-1)$ EM iteration.

Putting $\frac{\partial E_{\tilde{\mathbf{D}}_i^{(2)}}\{\mathcal{L}\}}{\partial|\bar{\mathbf{b}}|} = \mathbf{0}$, we need to solve the equation (below) for estimating the channel magnitude $\mathbf{b}$,

$$\begin{aligned}&\mathbf{W}(\mathbf{b}^{(l)}.)^2 + \mathbf{B}\mathbf{b}^{(l)} + \mathbf{D}^{(l)} = \mathbf{0},\\ &\mathbf{W} = 2\frac{A}{B}\sum_{l=1}^{M^N}[\mathbf{A}_l^H\mathbf{C}^{-1}\mathbf{A}_l] + \mathbf{G},\\ &\mathbf{B} = -\frac{A}{B}\sum_{l=1}^{M^N}(|\bar{\mathbf{z}}_i^H\mathbf{A}_l\mathbf{C}^{-1}|)\\ &\mathbf{D} = \mathbf{F}. \quad (27)\end{aligned}$$

Equation (27) is a matrix quadratic, non-linear equation, *unlike* linear MMSE or LS algorithms, typically used in one-way single channel estimation. The reciprocal coefficients of $\{1/b_i\}$ in $\bigtriangledown_{\mathbf{b}}[logf(\mathbf{b}|\mathbf{a})]$, in its 1st term in RHS of (12), gives rise to the quadratic coefficients of $\mathbf{b}$ in (27) and coefficients of its linear term $b_i$ in the 2nd term on its RHS. The estimate of composite channel magnitude $\mathbf{b}$ is given by that solution of (27) that gives non-negative entries of $\mathbf{b}$ (since $\mathbf{b}$ is the magnitude of the forward composite channel vector)

$$\begin{aligned}&\mathbf{b} = (2\mathbf{W})^{-1}[-\mathbf{B} + (\mathbf{B}^2 - 4\mathbf{WD})^{1/2}],\\ &\psi_{\mathbf{b}} = -\angle\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X} \quad (28)\end{aligned}$$

$(.)^{1/2}$ is the matrix square root operator.

**Note**: Multipath channel estimation (equation (16) [16], equation (40) [9]) uses LS methods, by neglecting all noise terms. It is different from MAP or even ML method, (which uses the noise correlation matrix $\mathbf{C}$).

### A. Estimation of Individual Channels from Composite Channel Estimation

In addition to estimation of composite channels $\mathbf{a}, \mathbf{b}$ at terminals $T1, T2$, *individual channels* $\mathbf{h}$ *and* $\mathbf{g}$ estimation may be needed for beamforming, power allocation at the relay node and the 2 terminals [17], [16], [5], to obtain more efficient directional transmission between relay node $R$ and $T1, T2$. Let us start with $3(L+1=3)$ channels whose magnitudes are $\mathbf{h} = [h_3,\, h_2,\, h_1]$ and $\mathbf{g} = [g_3,\, g_2,\, g_1]$ respectively (their phases will be determined separately). Then the individual channel $\mathbf{h}$ can be reconstructed from $\mathbf{a}$ by time-domain deconvolution as

$$h_1 = \sqrt{a_1},\, h_2 = \frac{a_2}{2h_1} = \frac{a_2}{2\sqrt{a_1}},\, h_3 = \sqrt{a_5} \tag{29}$$

Similarly, $L = 4$, then $\mathbf{h}, \mathbf{a}$ are $5 \times 1$ and $(\bar{L} = 9) \times 1$-sized vectors. Then $\mathbf{h}$ is obtained from $\mathbf{a}$ by, $\mathbf{h} = [\sqrt{a_1},\, a_2/(2\sqrt{a_1}),\, (a_4 - a_8/(2\sqrt{a_{\bar{L}}}))/a_2,\, a_8/(2\sqrt{a_{\bar{L}}}),\, \sqrt{a_{\bar{L}}}]^T$. From the other composite channel $\mathbf{b} = \mathbf{g} * \mathbf{h} = [b_1 b_2 \cdots b_{\bar{L}}]$, individual channel $\mathbf{g}$ is reconstructed by

$$\begin{aligned} &g_1 = b_1/h_1 = b_1/\sqrt{a_1},\, g_3 = b_5/\sqrt{a_5} \\ &g_2 = (b_2\sqrt{a_1} - \alpha b_1)/a_1, \alpha \overset{\Delta}{=} (2a_4 - a_2)/(2\sqrt{a_5}), \end{aligned} \tag{30}$$

Extension to generic $L$ multipath channels is given at the end of Appendix A.

## V. Estimation of Reverse Channel Parameter $\mathbf{a}$

Reverse channel estimation is complicated by a colored (channel dependent) noise at the receiver terminal, as shown in equation (10) and [17], thereby rendering the LS estimator ineffective. Though expressions for the ML based single-path, reverse channel estimator can be derived [17], [3], it is difficult to derive closed form expressions for MAP based reverse channel estimation (*even* for single-path channels), as noted in [5], and for full-duplex single-path TWRN networks [8]. Our paper overcomes this problem by deriving an unique, innovative MAP based multipath, reverse channel ($\mathbf{a}$ and $\mathbf{h}$) estimators, using Factor Analysis (a concept used sometimes in statistics and machine learning, but not used in signal processing and wireless communications). Factor Analysis allows us to convert the associated difficult maximum likelihood (ML) problem, into a tractable likelihood function, which can be maximized analytically.

From (7), we have, for the $l$th (received) OFDM block,

$$\begin{aligned} \tilde{\mathbf{z}}_l &\overset{\Delta}{=} [\mathbf{z}_l - \tilde{\mathbf{D}}_l^{(2)}\mathbf{b}] = \tilde{\mathbf{D}}_l^{(1)}\mathbf{a} \\ &+ ([\mathbf{n}_r(n) \cdots \mathbf{n}_r(n-L)]\mathbf{h} + \mathbf{n}_{1,l}, \end{aligned} \tag{31}$$

which can be re-written as

$$\tilde{\mathbf{z}}_l = \mathbf{A}_T\mathbf{d}_l^{(1)} + \mathbf{H}\mathbf{n}_{r,l} + \mathbf{n}_{1,l}, \tag{32}$$

$\mathbf{H}$ and $\mathbf{A}_T$ are the Toeplitz matrices constructed from $\mathbf{h}$ and $\mathbf{a}$ respectively. Extending (equation (14) in [33]), we have, for $S = 3$ OFDM blocks in a time-invariant channel,

$$\begin{aligned} \begin{bmatrix} \tilde{\mathbf{z}}_1 \\ \tilde{\mathbf{z}}_2 \\ \tilde{\mathbf{z}}_3 \end{bmatrix} &= diag(\mathbf{A}_T,\, \mathbf{A}_T,\, \mathbf{A}_T) \begin{bmatrix} \mathbf{d}_1^{(1)} \\ \mathbf{d}_2^{(1)} \\ \mathbf{d}_3^{(1)} \end{bmatrix} \\ &+ diag(\mathbf{H}, \mathbf{H}, \mathbf{H}) \begin{bmatrix} \mathbf{n}_{r1} \\ \mathbf{n}_{r2} \\ \mathbf{n}_{r3} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \mathbf{n}_3 \end{bmatrix}, \\ &= diag(\mathbf{A}_T,\, \mathbf{A}_T,\, \mathbf{A}_T) \begin{bmatrix} \mathbf{d}_1^{(1)} \\ \mathbf{d}_2^{(1)} \\ \mathbf{d}_3^{(1)} \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{n}}_1 \\ \bar{\mathbf{n}}_2 \\ \bar{\mathbf{n}}_3. \end{bmatrix} \end{aligned} \tag{33}$$

$\bar{\mathbf{n}}_l = \mathbf{H}\mathbf{n}_{r,l} + \mathbf{n}_l$ is the total noise, where the first noise term $\mathbf{H}\mathbf{n}_{r,l}$ is still Gaussian (as a filtered Gaussian noise remains Gaussian [19]), but colored (non-white). Thus, $\bar{\mathbf{n}}_l$ is a *Gaussian, colored* noise. Since $E\{\mathbf{d}_l^{(1)}\} = \mathbf{0}$, $E\{\mathbf{d}_l^{(1)}(\mathbf{d}_m^{(1)})^H\} = \delta(l-m)$, and $E\{(\mathbf{n}_r)_l\} = \mathbf{0}$, $E\{(\mathbf{n}_r)_l((\mathbf{n}_r)_m)^H\} = \sigma^2\delta(l-m)$, (and similarly for noise sequence $\mathbf{n}_l$). The mean and covariance matrix of $\mathbf{z}_l$ are

$$\begin{aligned} &\tilde{\mathbf{z}}_l = \mathbf{A}_T\mathbf{d}_l^{(1)} + \mathbf{H}\mathbf{n}_{r,l} + \mathbf{n}_{1,l}, \\ &\mu_{\tilde{\mathbf{z}}} = \mathbf{0} \\ &Cov(\tilde{\mathbf{z}}_l) = \mathbf{H}E\{\mathbf{n}_r\mathbf{n}_r^H\}\mathbf{H}^H + \mathbf{A}_T E\{\mathbf{d}_l^{(1)}{\mathbf{d}_l^{(1)}}^H\}\mathbf{A}_T^H \\ &\quad + E\{(\mathbf{n}_{1l}{\mathbf{n}_{1l}}^H)\} = \sigma^2(\mathbf{H}\mathbf{H}^H + \mathbf{I}) + \mathbf{A}_T\mathbf{A}_T^H = \tilde{\mathbf{C}}. \end{aligned} \tag{34}$$

$\tilde{\mathbf{C}}$ is the covariance matrix here (*different* from $\mathbf{C}$ in equation (14), Sec. IV). Suppose at this stage, $\mathbf{a}$ and its Toeplitz extension $\mathbf{A}_T$ are known from the last $(l-1)$th iteration. Then,

$$\begin{aligned} \hat{\mathbf{H}} &= max_{\mathbf{H}}\mathcal{L}(\mathbf{H}, \hat{\mathbf{A}}) \overset{\Delta}{=} \sum_{l=1}^{S} log(p(\tilde{\mathbf{z}}_l|\mathbf{H}, \mathbf{A})) \\ &= \sum_{l=1}^{S} (\tilde{\mathbf{z}}_l - \mu_{\tilde{\mathbf{z}}})^H(\tilde{\mathbf{C}})^{-1}(\tilde{\mathbf{z}}_l - \mu_{\tilde{\mathbf{z}}})(2\pi)^{-\bar{P}/2}|det(\tilde{\mathbf{C}})|^{-\frac{1}{2}}. \end{aligned} \tag{35}$$

Using (34), we have

$$\begin{aligned} &max_{\mathbf{H}}\mathcal{L}(\mathbf{H}, \mathbf{A}) = \sum_{l=1}^{S}([\tilde{\mathbf{z}}_l^H](\tilde{\mathbf{C}})^{-1}[\tilde{\mathbf{z}}_l]^H) \\ &(2\pi)^{-\bar{P}/2}|det(\tilde{\mathbf{C}})|^{-\frac{1}{2}}, \tilde{\mathbf{C}} = \sigma^2(\mathbf{I} + \mathbf{H}\mathbf{H}^H) + \mathbf{E}, \mathbf{E} = \mathbf{A}\mathbf{A}^H. \end{aligned} \tag{36}$$

However, maximizing $\mathcal{L}(\mathbf{H}, \hat{\mathbf{A}})$ is difficult, for which there is no known method, see [20].

It is here, that a technique called “Factor Analysis” (used in machine learning for separating mixture of distributions [31], [32], and very recently in speech and signal processing [33]), comes to the rescue. The received signal (observation vector) $\tilde{\mathbf{z}}_l$ can be viewed as filtering white $\mathbf{n}_r$ through the unknown $\mathbf{H}$ factor, (along with other factors) in the factor analysis model (32). Thus, $\mathbf{n}_r$ can be considered as a *latent/hidden variable*, which generates filtered $\mathbf{H}\mathbf{n}_r$ (the 2nd term in equation (32). Since, $\mathbf{H}\mathbf{n}_r$, has non-diagonal covariance matrix, it is *correlated, almost signal-like*. Rewriting (36),

$$\mathcal{L}(\mathbf{H},\hat{\mathbf{E}}) = \sum_{l=1}^{S} log(p(\tilde{\mathbf{z}}_l|\mathbf{H},\mathbf{E})) = \sum_{l=1}^{S} log(\sum_{\mathbf{n}_r} p(\tilde{\mathbf{z}}_l,\mathbf{n}_r;\mathbf{H},\mathbf{E}))$$
$$= \sum_{l=1}^{S} log(\sum_{\mathbf{n}_r} f(\mathbf{n}_r|\tilde{\mathbf{z}}_l) \times \frac{p(\tilde{\mathbf{z}}_l,\mathbf{n}_r;\mathbf{H},\mathbf{E})}{f(\mathbf{n}_r|\tilde{\mathbf{z}}_l)})$$
$$= \sum_{l=1}^{S} log(E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}\{\frac{p(\tilde{\mathbf{z}}_l,\mathbf{n}_r;\mathbf{H},\mathbf{E})}{f(\mathbf{n}_r|\tilde{\mathbf{z}}_l)}\}), \quad (37)$$

by multiplying both numerator and denominator by $f(\mathbf{n}_r|\tilde{\mathbf{z}}_l)$. Using Jensen's in-equality for concave $log$ function, i. e. $log(E\{X\}) \geq E\{log(X)\}$ , we have

$$\mathcal{L}(\mathbf{H},\mathbf{A}) \geq \sum_{l=1}^{S} E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}\{log(\frac{p(\tilde{\mathbf{z}}_l|\mathbf{n}_r;\mathbf{H},\mathbf{E})f(\mathbf{n}_r|\tilde{\mathbf{z}}_l)}{f(\mathbf{n}_r|\tilde{\mathbf{z}}_l)})\}$$
$$= \sum_{l=1}^{S}\sum_{\mathbf{n}_r} f(\mathbf{n}_r|\tilde{\mathbf{z}}_l) \times log(p(\tilde{\mathbf{z}}_l|\mathbf{n}_r;\mathbf{H},\mathbf{E}). \quad (38)$$

This reduces to

$$\geq \sum_{l=1}^{S} E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}\{log(p(\tilde{\mathbf{z}}_l|\mathbf{n}_r;\mathbf{H},\hat{\mathbf{E}})\} \implies max_{\mathbf{H}}\mathcal{L}(\mathbf{H},\mathbf{E})$$
$$= max_{\mathbf{H}}\,[max_{f(\mathbf{n}_r|\tilde{\mathbf{z}}_l)}\,[E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}\sum_{l=1}^{S}\{log(p(\tilde{\mathbf{z}}_l|\mathbf{n}_r;\mathbf{H},\mathbf{E})\}]], \quad (39)$$

which falls within the category of EM algorithms, with the last equation in (39), encompassing its E (Expectation) and M (Maximization) steps. Maximization of $\mathcal{L}$, over the pdf $f(\mathbf{n}_r|\tilde{\mathbf{z}}_l)$, followed by maximization over $\mathbf{H}$, are carried out in an *alternating fashion*, as in [33]. This *requires* that the pdf $p(\tilde{\mathbf{z}}_l|\mathbf{n}_r)$ be determined first, after which its *Expectation (over the hidden variable* $\mathbf{n}_r|\tilde{\mathbf{z}}_l$*)* is computed in (39). To compute (39), the conditional pdfs, $f(\mathbf{n}_r|\tilde{\mathbf{z}}_l)$ and $p(\tilde{\mathbf{z}}_l|\mathbf{n}_r;\mathbf{H},\mathbf{E})$, have to be determined. The conditional random variables $(\mathbf{n}_r|\tilde{\mathbf{z}}_l)$ and $(\tilde{\mathbf{z}}_l|\mathbf{n}_)$ are both Gaussian distributed, with pdfs $f(\mathbf{n}_r|\tilde{\mathbf{z}}_l) \sim N(\mu_{\mathbf{n}_r|\tilde{\mathbf{z}}_l},[Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l)])$ and $p(\tilde{\mathbf{z}}_l|\mathbf{n}_r) \sim N(\mu_{\tilde{\mathbf{z}}_l|\mathbf{n}_r}, Cov(\tilde{\mathbf{z}}_l|\mathbf{n}_r)$ respectively. Using $\tilde{\mathbf{z}}_l = \mathbf{H}(\mathbf{n}_r)_l + \mathbf{A}_T^{(k-1)}\mathbf{d}_l^{(1)} + (\mathbf{n})_l$ (from equation (34)), we have

*Lemma 3*

The conditional means and covariances of the pdfs $p(\tilde{\mathbf{z}}_l|\mathbf{n}_r;\mathbf{H},\mathbf{A})$ and $f(\mathbf{n}_r|\tilde{\mathbf{z}}_l)$ are given by

$$\mu_{\tilde{\mathbf{z}}_l|\mathbf{n}_r} = \mathbf{H}\mathbf{n}_r, Cov(\tilde{\mathbf{z}}_l|\mathbf{n}_r) = \sigma^2\mathbf{I} \quad (40)$$
$$\mu_{\mathbf{n}_r|\tilde{\mathbf{z}}_l} = \sigma^2\mathbf{H}^H(\tilde{\mathbf{C}})^{-1}(\tilde{\mathbf{z}}_l),$$
$$Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l) = \sigma^2[\mathbf{I} - \sigma^2\mathbf{H}^H[\tilde{\mathbf{C}}]^{-1}\mathbf{H}]. \quad (41)$$

Proof: See Appendix D.

Continuing from (39) and employing Lemma 3, it follows

$$\sum_{l=1}^{S} E\{log(p(\tilde{\mathbf{z}}_l|\mathbf{n}_r;\mathbf{H},\mathbf{A})\}$$
$$= \sum_{l=1}^{S} E_{\mathbf{n}|\tilde{\mathbf{z}}_l}[-log(2\pi)^{-\tilde{P}/2} - log(det(Cov(\tilde{\mathbf{z}}_l|\mathbf{n}_r))^{-\frac{1}{2}})$$
$$- \frac{1}{2}(\tilde{\mathbf{z}}_l - \mu_{\tilde{\mathbf{z}}_l|\mathbf{n}_r})^H(Cov(\tilde{\mathbf{z}}_l|\mathbf{n}_r))^{-1}(\tilde{\mathbf{z}}_l - \mu_{\tilde{\mathbf{z}}_l|\mathbf{n}_r})],$$
$$= \sum_{l=1}^{S} E[-log(2\pi)^{-\tilde{P}/2} - log(det(\sigma^2\mathbf{I})^{-\frac{1}{2}}) -$$
$$\frac{1}{2}(\tilde{\mathbf{z}}_l - \mathbf{H}\mathbf{n}_r)^H\frac{1}{\sigma^2}(\tilde{\mathbf{z}}_l - \mathbf{H}\mathbf{n}_r)], \quad (42)$$
$$= \frac{1}{2}\sum_{l=1}^{S}[E(\frac{\tilde{\mathbf{z}}_l^H\tilde{\mathbf{z}}_l}{\sigma^2}) - \frac{(E[\mathbf{n}_r^H]\mathbf{H}^H\tilde{\mathbf{z}}_l)}{\sigma^2} - E(\frac{\tilde{\mathbf{z}}_l^H\mathbf{H}\mathbf{n}_r}{\sigma^2})$$
$$+ E(\frac{(\mathbf{n}_r^H\mathbf{H}^H\mathbf{H}\mathbf{n}_r}{\sigma^2})] + other\,terms. \quad (43)$$

Then using $a = \mathrm{trace}(a), a : scalar, \mathrm{trace}(\mathbf{AB}) = \mathrm{trace}(\mathbf{BA})$, equation (43) gives the log-likelihood (for $k$th iteration) as

$$\mathcal{L} = \frac{1}{2}\sum_{l=1}^{S}\mathrm{trace}[E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l},(\frac{\tilde{\mathbf{z}}_l^H\tilde{\mathbf{z}}_l}{\sigma^2}) - \frac{E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}(\mathbf{n}_r^H)\mathbf{H}^H\tilde{\mathbf{z}}_l}{\sigma^2}$$
$$- \frac{\tilde{\mathbf{z}}_l^H\mathbf{H}E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}(\mathbf{n}_r)}{\sigma^2} + \frac{(\mathbf{H}^H\mathbf{H})E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}(\mathbf{n}_r^H\mathbf{n}_r)}{\sigma^2}]$$
$$+ other\,terms, \quad (44)$$

Now, one has to maximize the likelihood function over $\mathbf{H}$, for which we take the differentiation of (44) with respect to $\mathbf{H}$, and equate it to $\mathbf{0}$. This provides an estimate of the channel parameter $\mathbf{H}$ matrix, (rather than the composite channel $\tilde{\mathbf{a}}$ vector). Using the identity , $\frac{\partial(\mathrm{trace}(\mathbf{ABA}^T\mathbf{C}))}{\partial\mathbf{A}} = \mathbf{CAB} + \mathbf{C}^H\mathbf{AB}^H$, (for any generic we have ,

$$\frac{\partial E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}(log(\mathcal{L})}{\partial\mathbf{H}} = \frac{1}{2}\sum_{l=1}^{S} 2\frac{\mathbf{H}}{\sigma^2}E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}(\mathbf{n}_r^H\mathbf{n}_r)$$
$$- 2[\frac{\tilde{\mathbf{z}}_l}{\sigma^2}E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}(\mathbf{n}_r^H)] = \mathbf{0}. \quad (45)$$

Finally, the estimate $\hat{\mathbf{H}}$ is

$$\hat{\mathbf{H}} = (\sum_{l=1}^{S}[\tilde{\mathbf{z}}_l\mu^H_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}]) \times (\sum_{l=1}^{S}[Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l) + \mu_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}\mu^H_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}])^{-1}, \quad (46)$$

by using $E(\mathbf{n}_r^H\mathbf{n}_r) = Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l) + \mu_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}\mu^H_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}$. If apriori information about the pdf $\mathbf{H}$ is utilized, then the MAP estimate of $\mathbf{H}$ is obtained by defining $\mathbf{W} = \frac{\partial f(\mathbf{h})}{\partial\mathbf{H}} = (\frac{(\prod_{i=0}^{L}h_i)}{(2\pi\sigma^2)^{L+1}}e^{-\sum_{j=0}^{L}\frac{h_i^2}{2\sigma^2}}) \times \mathbf{T}$, where $\mathbf{T}$ is a Toeplitz matrix with its first row given by $[(\sigma^2 - h_L^2), (\sigma^2 - h_{L-1}^2), \cdots, (\sigma^2 - h_0^2), 0, \cdots, 0]$. Then equation (46) is modified by

$$\hat{\mathbf{H}} = (\sum_{l=1}^{S}[\tilde{\mathbf{z}}_l\mu^H_{\mathbf{n}_r|\tilde{\mathbf{z}}_l} + \mathbf{W}] \times$$
$$(\sum_{l=1}^{S}[Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l) + \mu_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}\mu^H_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}])^{-1} \quad (47)$$

$\mathbf{H}$ is initialized by its LS estimate,

$$\hat{\mathbf{H}}^{(0)} = (\sum_{l=1}^{S}[\mathbf{D}^{(1)^H}\tilde{\mathbf{z}}_l]) \times (\sum_{l=1}^{S}\mathbf{D}^{(1)^H}\mathbf{D}^{(1)})^{-1}, \quad (48)$$

In (46), one needs to compute substitute the values $\mu_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}$ and $Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l)$, using equation (41) (Lemma 3). Then the $l$th iteration of $\hat{\mathbf{H}}^{(l)}$, is computed (from its $(l-1)$th iteration estimate), by

$$\begin{aligned}\hat{\mathbf{H}}^{(l)} = &(\sum_{l=1}^{S}\sigma^2[\tilde{\mathbf{z}}_l\tilde{\mathbf{z}}_l^H\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}]) \\ &\times(\sum_{l=1}^{S}\sigma^2[[\mathbf{I}-\sigma^2\mathbf{H}^{(l-1)H}\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}] \\ &+\sigma^2[\mathbf{H}^{(l-1)H}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{z}}_l\tilde{\mathbf{z}}_l^H\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}]])^{-1} \quad (49)\end{aligned}$$

where $\tilde{\mathbf{C}}^{(l-1)} = \sigma^2(\mathbf{I}+\mathbf{H}^{(l-1)}\mathbf{H}^{(l-1)H}) + \mathbf{A}_T^{(l-1)}\mathbf{A}_T^{(l-1)^H}$. The estimated channel matrix $\mathbf{H}$ being a Toeplitz matrix, the individual channel $\mathbf{h}$ is estimated by

$$\hat{\mathbf{h}}^{(l)}(k) = mean(diag(\mathbf{H}^{(l)},k)), k = 0,1,\cdots,L, \quad (50)$$

where $diag(\mathbf{H},k)$ is $k$th sub-diagonal of $\mathbf{H}$ matrix ($diag(\mathbf{H},0)$ is the main diagonal of $\mathbf{H}$). The NRMSE of $\hat{\mathbf{h}}$, in (50) is less than than of LS estimates, Fig. 7. Then the reverse composite channel vector $\mathbf{a}$ can be estimated by

$$\hat{\mathbf{a}}^{(l)} = \hat{\mathbf{h}}^{(l)} * \hat{\mathbf{h}}^{(l)}. \quad (51)$$

So this algorithm proceeds, as follows: In every $l$th EM iteration, determine $\hat{\mathbf{h}}^{(l)}$ is obtained by equation (50). Then we estimate $\hat{\mathbf{a}}^{(l)}$ by (51), from which the matrix $\mathbf{A}_T^{(l)}$ is constructed from a Toeplitz extension of $\hat{\mathbf{a}}^{(l)}$. These are all used in the $(l+1)$th iteration to compute $\mathbf{H}^{(l+1)}$, $\mathbf{h}^{(l+1)}$, by using (49) and (50). Since fixed values of $\hat{\mathbf{a}}^{(l)}$ and $\hat{\mathbf{A}}_T^{(l)}$ are used to estimate $\hat{\mathbf{H}}^{(l+1)}$ at $(l+1)$th iteration. $\hat{\mathbf{H}}^{(l+1)}$ is then fixed at its present value, and then it is used to estimate $\hat{\mathbf{a}}^{(l+1)}$ and $\hat{\mathbf{A}}_T^{(l+1)}$, and so on. Thus the above parameters are coupled together, which has, till date, been solved by numerical means only, as in existing literature [5], [8]. The complete algorithm is tabulated in Table II.

### A. Convergence

It can be shown that (49) converges to the optimal solution. After processing an adequate number ($S$) of OFDM blocks,

$$\begin{aligned}&\hat{\mathbf{H}}^{(l)} \to \sigma^2[E(\tilde{\mathbf{z}}_l\tilde{\mathbf{z}}_l^H)\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}]) \times (\sigma^2[[\mathbf{I}-\sigma^2\mathbf{H}^{(l-1)H} \\ &\cdot\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}] + \sigma^2[\mathbf{H}^{(l-1)H}\tilde{\mathbf{C}}^{-1}E(\tilde{\mathbf{z}}_l\tilde{\mathbf{z}}_l^H)\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}]])^{-1} \\ &= (\sigma^2[\tilde{\mathbf{C}}\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}]) \times (\sigma^2[[\mathbf{I}-\sigma^2\mathbf{H}^{(l-1)H}\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}] \\ &+\sigma^2[\mathbf{H}^{(l-1)H}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{C}}\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}]])^{-1} \\ &= (\sigma^2[\mathbf{H}^{(l-1)}])]) \times [\sigma^2\mathbf{I}]^{-1} = \hat{\mathbf{H}}^{(l-1)}, \quad (52)\end{aligned}$$

by recalling that $E(\tilde{\mathbf{z}}_l\tilde{\mathbf{z}}_l^H) = \tilde{\mathbf{C}}$ (equation (34)). The above convergence *confirms the validity* of the reverse channel estimator in (49). Estimating $\mathbf{h}$ in this way may be *less* erroneous, than estimating ($\mathbf{a} = (\mathbf{h}*\mathbf{h})$ first, and then obtaining $\hat{\mathbf{h}}$ by de-convolving a noisy estimate of $\mathbf{a}$). Equation (49) is a non-linear equation, which is *unlike* linear MMSE or LS algorithms, typically used in one-way individual channel estimation. Recall, estimation is being done at terminal $T1$, and thus knows its own transmitted data $\tilde{\mathbf{D}}_i^{(1)}$.

TABLE I

**Two Way Relay Networks (TWRN) Composite Channel Estimation**

| | |
|---|---|
| **Step A.** | **Estimation of Composite Channel** $\mathbf{b}$ |
| **Step A. 1** | 2 $L$-tap individual channels $\mathbf{h} = [h_L\, h_{L-1} \cdots h_0]$, $\mathbf{g} = [g_L\, g_{L-1} \cdots g_0]$. Define the magnitude of composite channels $\mathbf{a} = (\mathbf{h}*\mathbf{h})$, $\mathbf{b} = (\mathbf{g}*\mathbf{h})$, $*$ : linear convolution. $\tilde{P}\times\bar{L}$ matrices $\tilde{\mathbf{D}}_i^{(k)}$, $k=1,2$ are Toeplitz matrices constructed from the transmitted data from 2 terminals $T_k$, $\bar{L} = 2L+1$. |
| **Step A. 2** | Let $\mathbf{a}$ be *known (or, from previous iteration)*, From $\mathbf{a}$, determine $\mathbf{h}$ by deconvolution. *For example*, for $(L+1) = 3$ path multipath channels $\mathbf{h},\mathbf{g}$, the composite channel $\mathbf{a}$ if of length $\bar{L} = 2L+1 = 5$. Then $\mathbf{h} = [\sqrt{a_1}, a_2/(2\sqrt{a_1}), \sqrt{a_{2L+1)}}]$ can be easily determined. See Appendix A for details. |
| **Step A. 3** | Construct channel matrix $\mathbf{H}$ as the Toeplitz matrix from $\mathbf{h}$. Then compute $\mathbf{C} = (\mathbf{H}\mathbf{H}^H + \mathbf{I})\sigma^2$. |
| **Step A. 4** | Moreover, Terminal $T_1$ knows its own transmitted data $\tilde{\mathbf{D}}_i^{(1)}$. |
| **Step A. 5** | Define: $\bar{\mathbf{z}}_i \triangleq [\mathbf{z}_i - \tilde{\mathbf{D}}_i^{(1)}\mathbf{a}] = \tilde{\mathbf{D}}_i^{(2)}\mathbf{b} + \mathbf{n}(n)$ Let $\mathbf{A}_l$ be the $\bar{P}\times\bar{L}$-sized alphabet matrix at $l$th computer experiment, corresponding to $\tilde{\mathbf{D}}_i^{(2)}$ ($i$th OFDM block transmitted block from terminal $T2$). Define $A = \sum_{l=1}^{\bar{L}} exp^{-\frac{1}{2}[\lVert\bar{\mathbf{z}}_i^H\rVert^2_{\mathbf{C}^{-1}} + 2\lVert\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}diag(\mathbf{b})\rVert + \lVert\mathbf{b}^H\mathbf{X}]\rVert^2_{\mathbf{C}^{-1}}]}$, $B = \sum_{t=1}^{M^N}[exp^{-\frac{1}{2}[\lVert\bar{\mathbf{z}}_i^H\rVert^2_{\mathbf{C}^{-1}} + 2\lVert\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}diag(\mathbf{b})\rVert + \lVert\mathbf{b}^H\mathbf{X}]\rVert^2_{\mathbf{C}^{-1}}]}]^{\bar{L}}$, $\mathbf{X} = \sum_{l=1}^{M^N}\mathbf{A}_l\frac{A}{B}$ |
| **Step A.6** | Phase of composite (complex) $\bar{b}b$ is given by $\psi_{\mathbf{b}}^{(l)} = -\angle\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X} = -\angle\bar{\mathbf{b}}^{(l)H}\mathbf{A}_l^H\mathbf{C}^{-1}\mathbf{A}_l$, with $\bar{\mathbf{b}}$ on RHS of (equation above) having been computed at the last $(l-1)$ EM iteration and $\psi_{\mathbf{b})}$ on LHS, is the phase estimate at the current $l$th EM iteration. |
| **Step A.7** | *Expectation of Maximum A-posteriori (MAP)* criterion $\{\mathcal{L}\}$ for estimating composite channel $\mathbf{b}$: $E_{\bar{\mathbf{D}}_i^{(2)}}\{\mathcal{L}\} = -\text{trace}\{[\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\bar{\mathbf{z}}_i]\} - \text{trace}\{\lVert[\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}[\bar{\mathbf{b}}]\rVert]\} + \sum_{l=1}^{M^{2L+1}}\frac{\bar{A}}{B}\mathbf{A}_l^H\mathbf{C}^{-1}\text{trace}\{\lVert diag(\mathbf{b})\mathbf{A}_l^H\mathbf{C}_T)^H\rVert^2\} + other\,terms$, $\bar{A} = A(e^{\lVert\bar{\mathbf{b}}^H\mathbf{A}_l\rVert^2})$. |
| **Step A.8** | Differentiation of $E[\mathcal{L}]$ with respect to $\mathbf{b}$, $\frac{\partial E\{\mathcal{L}\}}{\partial\mathbf{b}} = 2\frac{A}{B}\sum_{l=1}^{M^N}[\mathbf{A}_l^H\mathbf{C}^{-1}\mathbf{A}_l\mathbf{b} - (\lvert\bar{\mathbf{z}}_i^H\mathbf{A}_l\mathbf{C}^{-1}\rvert)] + \mathbf{F}\frac{1}{\mathbf{b}} + \mathbf{G}\mathbf{b}.$, $\mathbf{F}$ and $\mathbf{G}$ given in equation (26). where the superscript $^{(l-1)}$ refers to the earlier $(l-1)$ EM iteration. $\frac{\partial E_{\tilde{\mathbf{D}}_i^{(2)}}\{\mathcal{L}\}}{\partial\mathbf{b}} = 2\frac{\bar{A}}{B}\sum_{l=1}^{M^{2L+1}}\mathbf{A}_l^H\mathbf{C}^{-1}(\mathbf{C}_T\mathbf{A}_l)\mathbf{b} - 2\sum_{l=1}^{M^L}\sqrt{\frac{A}{B}}(\lvert\bar{\mathbf{z}}_i^H\mathbf{A}_l\mathbf{C}^{-1}\rvert) + \mathbf{F}\frac{1}{\mathbf{b}} + \mathbf{G}\mathbf{b}$. $\mathbf{F}$ and $\mathbf{G}$ given in equation (26). |
| **Step A.9** | Putting $\frac{\partial E_{\tilde{\mathbf{D}}_i^{(2)}}\{\mathcal{L}\}}{\partial\mathbf{b}}$ to $\mathbf{0}$, equation to estimate the magnitude of the composite channel $\mathbf{b}$: $\mathbf{W}(\mathbf{b}.)^2 + \mathbf{B}\mathbf{b} + \mathbf{D} = \mathbf{0}, \mathbf{W} = \frac{\bar{A}}{B}\sum_{l=1}^{M^{2L+1}}\mathbf{A}_l^H\mathbf{C}^{-1}(\mathbf{C}_T\mathbf{A}_l) + \mathbf{B} = -\sum_{l=1}^{M^{2L+1}}\sqrt{\frac{A}{B}}(\lvert\bar{\mathbf{z}}_i^H\mathbf{A}_l\mathbf{C}^{-1}\rvert)\mathbf{D} = \mathbf{F}\ (.)^{1/2}$ : matrix square root. Finally using Step **Step A.8** , estimate $\tilde{\mathbf{b}} = diag(\hat{\bar{\mathbf{b}}})e^{j\hat{\psi}_{\mathbf{b}}}$. |

## VI. Comparison with Existing Methods

Our novel semiblind multipath relay estimation algorithm is specifically compared with 1. [17], 2. [16], 3. [3], 4. [5], 5. [9].

### A. Comparison with [17], [16]

[17] is a relay estimation method in single-carrier, continuous transmission (CT not block-based system with a *single* path, which eases the problem considerably. In case of multipath channels, the forward channel $\mathbf{b}$ is estimated by a LS method [16], by neglecting all noise terms. The LS method (*not equivalent to ML*) in this case, as the overall noise ($\tilde{\mathbf{n}} = \mathbf{H}\mathbf{n}_r + \mathbf{n}_1$) in (10) *depends* on the multipath channel $\mathbf{h}$, (as for single-path channels in [17]), making it a colored noise. This fact makes LS solutions inadequate to estimate the reverse relay channel $\mathbf{a}$. A ML method is

TABLE II
**Two Way Relay Networks (TWRN) Cascaded Channel Estimation**

**Step B** **Estimation of Composite Channel a**

**Step B.1** For $l$th iteration, assume that $\mathbf{a}^{(l-1)}$ known from $(l-1)$th EM iteration.

**Step B.1** From received signal at Terminal 1, define $\tilde{\mathbf{z}}_l \triangleq [\mathbf{z}_l - \tilde{\mathbf{D}}_l^{(2)}\bar{\mathbf{b}}] = \tilde{\mathbf{z}}_l = \mathbf{A}_T\mathbf{d}_l^{(1)} + \mathbf{H}(\mathbf{n}_r)_l + (\mathbf{n}_1)_l,$. $\mathbf{H}$ and $\mathbf{A}_T$ are the Toeplitz matrices constructed from $\mathbf{h}$ and $\mathbf{a}$.

**Step B.2** Mean and covariance matrix of $\tilde{\mathbf{z}}_l$ are $\mu_{\tilde{\mathbf{z}}} = \mathbf{0}$, $Cov(\tilde{\mathbf{z}}_l) = \sigma^2(\mathbf{H}\mathbf{H}^H + \mathbf{I}) + \mathbf{A}_T\mathbf{A}_T^H = \tilde{\mathbf{C}}$

**Step B.3** $max_{\mathbf{H}}\mathcal{L}(\mathbf{H}, \hat{\mathbf{A}})$ solved using *Method of alternating Maximization.*

$$max_{\mathbf{H}}\,[max_{f(\mathbf{n}_r|\tilde{\mathbf{z}}_l)}\,[E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}[\sum_{l=1}^{S}\{log(p(\tilde{\mathbf{z}}_l|\mathbf{n}_r, \mathbf{H})\}]]$$

Maximization of $\mathcal{L}$, over the pdf $f(\mathbf{n}_r|\tilde{\mathbf{z}}_l)$, followed by maximization over $\mathbf{H}$, are carried out in an *alternating fashion*.
Use EM algorithm with Expectation (E) and Maximization (M) steps.

**Step B.4** Then the likelihood function is

$$\mathcal{L} = \frac{1}{2}\sum_{l=1}^{S}\text{trace}[E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l},(\frac{\tilde{\mathbf{z}}_l^H\tilde{\mathbf{z}}_l}{\sigma^2}) - \frac{E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}(\mathbf{n}_r^H)\mathbf{H}^H\tilde{\mathbf{z}}_l}{\sigma^2}$$
$$- \frac{\tilde{\mathbf{z}}_l^H\mathbf{H}E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}(\mathbf{n}_r)}{\sigma^2} + \frac{(\mathbf{H}^H\mathbf{H})E_{\mathbf{n}_r|\tilde{\mathbf{z}}_l}(\mathbf{n}_r^H\mathbf{n}_r)}{\sigma^2}] + \cdot$$

**Step B.5** Using *EM* algorithm and *Factor Analysis*, $l$th iteration $\hat{\mathbf{H}}^{(l)\,H}$, from its $(l-1)$th iteration estimate, given by

$$\hat{\mathbf{H}}^{(l)} = (\sum_{l=1}^{S}\sigma^2[\tilde{\mathbf{z}}_l\tilde{\mathbf{z}}_l^H\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}] + \mathbf{W})$$
$$\times(\sum_{l=1}^{S}\sigma^2[[\mathbf{I} - \sigma^2\mathbf{H}^{(l-1)\,H}\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}]$$
$$+\sigma^2[\mathbf{H}^{(l-1)\,H}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{z}}_l\tilde{\mathbf{z}}_l^H\tilde{\mathbf{C}}^{-1}\mathbf{H}^{(l-1)}]])^{-1},$$

where $\mathbf{W} = (\frac{(\prod_{i=0}^{L}h_i)}{(2\pi\sigma^2)^{L+1}}e^{-\sum_{j=0}^{L}\frac{h_i^2}{2\sigma^2}}) \times \mathbf{T}$, where $\mathbf{T}$ is a Toeplitz matrix with its first row given by $[(\sigma^2 - h_L^2)$, $(\sigma^2 - h_{L-1}^2)$, $\cdots$, $(\sigma^2 - h_0^2)$, $0$, $\cdots$, $0]$.

**Step B.6** $\mathbf{H}$ is initialized by its LS estimate,

$$\hat{\mathbf{H}}^{(0)} = (\sum_{l=1}^{S}[\mathbf{D}^{(1)\,H}\tilde{\mathbf{z}}_l]) \times (\sum_{l=1}^{S}\mathbf{D}^{(1)\,H}\mathbf{D}^{(1)})^{-1}.$$

**Step B.7** From Toeplitz $\hat{\mathbf{H}}$, $\hat{\mathbf{h}}$ is obtained by averaging over its diagonal and sub-diagonal entries.

**Step B.8** *For next iteration*, calculate $\mathbf{a} = \mathbf{h} * \mathbf{h}$ (linear convolution), and $\mathbf{A}_T$ is a Toeplitz extension of $\mathbf{a}$.

**Step B.9** Estimating $\mathbf{h}$ in this way is *less* erroneous, than estimating $\mathbf{a} = (\mathbf{h}*\mathbf{h})$ first, and then obtaining $\hat{\mathbf{h}}$ by de-convolving a noisy estimate of $\mathbf{a}$.

thus used in [17] to estimate single-tap $\mathbf{a}$, with enhanced performance.

On the other hand, [16] estimates multipath channels in an OFDM system via a time-domain method. However, composite channels $\mathbf{a}$ and $\mathbf{b}$ are jointly estimated by a LS method *only, neglecting all noise terms*, see equation (16), [16], which is clearly *sub-optimal* for estimating $\mathbf{a}$ in noisy situations, as overall noise is colored, depending on individual channel $\mathbf{h}$, as explained in [17], and above paragraph. Simulations show inferior performance in noisy multipath cascaded channels, compared to our novel method here, but extraction of individual channel estimation is done in equivalent frequency domain (see expressions in Sec. III. A. 3, and equation (38) in [16])-which may lead to noise enhancement, due to scalar divisions, at channel spectral null frequencies); also requires some restrictions on each channel length (along with sign ambiguity). Additionally, [16] is an entirely pilot-data based estimation method, not exploiting prior channel pdf as in MAP estimators, while our novel method is *blind/semi-blind*, thus making it spectrally efficient, by allowing more OFDM blocks/symbols to be transmitted in a certain amount of time.

**Note**: Multipath channel estimation (equation (16) [16], equation (40) [9]) use LS methods (not MAP or even ML method, which uses the noise correlation matrix $\mathbf{C}$ in the associated likelihood function in equation (16)).

### B. Comparison with [4], [3], [2]

[4], [3] estimate a single-path/tap $L + 1 = 1$ TWRN channel; the estimation is semi-blind, as much of the transmitted symbols from the *other terminal* (except for few pilot symbols is unknown but the method is designed for a *single path* channels. The semi-blindness is achieved by using an EM algorithm (hidden variables being the unknown transmitted symbols from the other terminal, similar to our approach).

Also, [3] (and [2]) considers a half-duplex *single-path TWRN* $\mathbf{h}, \mathbf{g}$ channels, with timing offsets and pulse-shaping filters, using a ML method, along with optimization of pulse-shaping filters and training sequences. In *asynchronous* [3], it is assumed that the data from terminal $T_1$ arrives at relay $R$ arrives earlier than from other terminal $T_2$ by a timing offset of $\tau$. The value of $\tau$ may be summation of integral multiple of symbol period and a fraction (of symbol period) part. Equation (12) in [3] shows the received signal (at $T_1$) consists a term, which is the summation of few (instead of a single, in synchronous case) data symbols transmitted from Terminal $T_2$, which leads to a multipath channel $\mathbf{g}$. Specifically Sec. III. B. [3] considers a rectangular pulse truncated to 2 symbol periods, so the received signal (at $T_1$) consists of interference, corresponding to 2 transmitted data symbols from $T_2$ (denoted by $\mathbf{s}_2^k$, $k = 1, 2$) in equation (31) [3]. This results in a $L + 1 = 2$ tap multipath channel, which is just a particular case of our generic $L + 1$ (for any $L$) $\mathbf{g}$ channel. Our method can thus be *easily* applied to asynchronous system, with timing offsets (though oversampling may be required to handle the fractional part of timing offset). *Not* only that, [3], [4], do *not* exploit the channel pdf information (if available), as it is not based on MAP optimality. Our novel method, on the other hand, employs a *MAP* method for a multipath channels.

### C. Comparison with [5] and [6]

: TWRN channel estimators are derived using the MAP criterion in [5], and show improved over ML estimators in certain cases. However, [5] deals with *only* single-tap/path channels, while ours deal with multipath channels, which is a more complicated situation in two-way cascaded channels. These methods also suffer from another disadvantage in that *even for single-path channels, MAP based [5] does not provide close-form expressions* for the ML reverse channel ($\mathbf{a}$) estimation problem; instead the likelihood function is maximized via (exhaustive search, in both magnitude and phase dimensions) or suboptimal search and numerical techniques [5]. While, *close-form expressions* can be derived for our novel MAP reverse multipath channel estimator (by employing "Factor Analysis" technique).

### D. Comparison with [8]

[8] estimates single-path channels, but for full duplex, AF TWRN systems. Due to the feedback in the self interference link, at the relay, the overall system cannot avoid being a multipath (ISI) channel. However, the multipath weights are all related to each other (which makes it easier), whereas ours is a random multipath channel, with multipath tap values uncorrelated to each other. Another problem with [8] is that it does not provide *close-form solutions* for the ML reverse channel ($\mathbf{a}$) estimation problem, unlike our novel MAP reverse channel estimator. By incorporating the self-interference in full duplex networks, our novel method may be extended to full duplex TWRN multipath channels. However, this is not provided in this paper, due to brevity (space limitations).

### E. Comparison with [9]

[9] employs a EM-based method for semi-blinding estimating a full-duplex TWRN OFDM multipath channel, along with frequency offsets. However, composite channels $\mathbf{a}$ and $\mathbf{b}$ are each estimated by a suboptimal LS method *only*, (see equation (40) in [9]). The noise terms $\mathbf{w}_1(n)$ (in equation (9), and described above equation (11), pp. 5, [9]) is taken as white noise (with a diagonal correlation matrix), which is clearly *sub-optimal* for estimating $\mathbf{a}$ in noisy situations, (because the colored noise term in (10) depends on individual channel $\mathbf{h}$, as seen in equation (10) above and in [17])). It provides for spectrally efficient, semiblind estimation (via EM algorithm), but, being semiblind, performs *worse* than *even* all training-pilots-based [16].Not only that, another disadvantage of [9] is that it does *not* provide methods for *individual channels* $\mathbf{h}$ *and* $\mathbf{g}$ estimation, which may be needed for beamforming, power allocation at the relay node and the 2 terminals [17], [16], [5], nor does it use MAP optimality criterion.

Ours is a time-domain, MAP based TWRN multipath channel estimator, which works well in reduced guard interval. The MAP likelihood functions, for both $\mathbf{a}$ and $\mathbf{b}$ channels, are in the time-domain. The colored noise situation, in eqn (10) leads to a complicated likelihood function, with noise correlation matrix $\mathbf{C} = \sigma^2(\mathbf{H}\mathbf{H}^H + \mathbf{I})$, which is difficult to maximize, in estimating reverse channel $\mathbf{a}$. We overcome this problem by employing a *novel "Factor Analysis"* method. The "Factor Analysis" method allows us to transform the associated likelihood function to a simpler form, which can be maximized *analytically*, see equation (49). We also provide individual channels $\mathbf{h}$ and $\mathbf{g}$ estimation methods.

To the best of our knowledge, Factor Analysis method has *not* been employed for TWRN channel estimation *before*. Even for single-path channels, MAP based [5] and [8] do not provide *close-form solutions* for the ML reverse channel ($\mathbf{a}$) estimation problem; instead the likelihood function is maximized via (exhaustive search, in both channel magnitude and phase dimensions) and numerical techniques [5].

## VII. Derivation of Cramer-Rao Bounds (CRB) for Cascaded Channels in TWRN

Derivation of Cramer Rao(CR) lower bounds (CRLB) for forward and reverse channels is quite complicated, especially for multipath TWRN channels.

*Lemma 4*: The CRLBs for forward ($\mathbf{b}$) and reverse channels ($\mathbf{H}/\mathbf{h}$) are given by

$$CRLB_{\mathbf{b}} = \mathbf{J_{b,b}}^{-1} = (1/4)\sum_l ((||\mathbf{H}||^2 + \mathbf{I})\sigma^2)$$
$$([||\mathbf{A}_l^H||^2]^{-1}) \tag{53}$$
$$\mathbf{J_{H,H}} = \frac{4}{\sigma^4}[E\{\|\mu^H\tilde{\mathbf{z}}_l\|^2\} + E\{\|Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l)\mathbf{H}]\|^2\} - E\{\mu^H\tilde{\mathbf{z}}_l Cov^*(\mathbf{n}_r|\tilde{\mathbf{z}}_l)\mathbf{H}^*\} - E\{Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l)\mathbf{H}\mu^*\tilde{\mathbf{z}}_l^*\}, \tag{54}$$

from which $\mathbf{J_{h,h}}$ is easily obtained.

*Proof*: See Appendix E.

## VIII. Simulation Results

The channels from both terminals $T_j$, $j = 1, 2$ to the relay $R$, are assumed to have five or three taps, each represented by a symmetric complex Gaussian random variable with zero mean and unit variance. The novel algorithms for MAP multi-path TWRN channel estimators are compared existing methods. Their performances have been compared in terms of i) varying number of subcarriers (size) of each OFDM block, ii) varying SNRs, iii) varying length of multipath channels ($L + 1$), iv) number of OFDM blocks used in estimation. The CRBs have been *derived* for both forward (and) reverse composite/individual channels, used as a lower bound benchmark of performance for all estimators.

The data signals $\{s_i^{(k)}(n)\}$ are binary phase shift keying (BPSK)/quadrature phase shift keying (QPSK) modulated, with a single transmit/receive antenna at relay $R$ and at each of 2 terminals $T1$, $T2$. Few ($\bar{T}$) of the transmitted subcarriers $\{d_i^{(k)}(n)\}$, in each OFDM block, are known at the receiver (pilots), the rest are *unknown* data subcarriers. Simulation results are obtained by averaging over 100 trials; for each computer trial, independent and identically distributed complex Gaussian channel coefficients with zero mean and unit variance (Rayleigh fading channel) are generated. The receiver signal-to-noise ratio (SNR) is defined as $\mathrm{SNR} = \frac{\mathrm{E}(||\mathbf{y}(n)-\mathbf{w}(n)||^2)}{\mathrm{E}(||\mathbf{w}(n)||^2)}$, ($\mathbf{w}(n)$ : AWGN noise); performance of different estimators measured by normalized MSE (NRMSE)

$$\mathrm{NRMSE} = \frac{1}{100}\sum_{p=1}^{100}\left\{\frac{\sum_{\ell=0}^{L}||\mathbf{Ch}^{(p)}(\ell) - \hat{\mathbf{Ch}}^{(p)}(\ell)||_F^2}{\sum_{\ell=0}^{L}||\mathbf{Ch}^{(p)}(\ell)||_F^2}\right\}, \tag{55}$$

for any generic channel $\mathbf{Ch}$. $\mathbf{Ch}$ could be composite channels $\mathbf{a}$, $\mathbf{b}$ or individual channels $\mathbf{g}$, $\mathbf{h}$.

As mentioned earlier, multiple forward and reverse, individual as well as composite channels estimators (for novel semiblind, and existing training, semiblind algorithms) have to be *simulated*, for performance comparisons:

i) Novel Semiblind MAP based Composite $\mathbf{b}, \mathbf{a}$ (Forward and Reverse) Multipath Channel Estimation, using EM algorithm, in ZP-OFDM multipath transmission system, (denoted by "Semiblind-EM-ChbEst" in Figs/plots)

ii) Blind MAP channel estimation of Reverse individual channel $\mathbf{h}$ and individual channel $\mathbf{g}$. Factor Analysis and a 2nd EM algorithm, (denoted by "Est-ChH-Novel" and "Est-ChG-Novel")

iii) Novel training pilots based all composite and individual channel estimators, (denoted by "Training-Chb-Est")

iv) Existing OFDM multipath (composite and individual) channel estimator [16], (denoted by "Gao-Chb-Est", "Est-ChH-Gao" and "Est-ChG-Gao")

v) Another existing semiblind OFDM multipath (composite) channel estimator [9], (denoted by "Chak-Chb-Est", "Est-ChH-Chak" and "Est-ChG-Chak")

vi) Another existing ML based semiblind single-path/tap composite channel, with timing offsets in asynchronous systems, [3], and compared with our novel method, (denoted by "[4]-Timing Offset")

vii) Simulation of cyclic-prefix (CP) based OFDM system,

viii) Simulation of CRB lower bounds of both composite and individual channels, (denoted by "CRB-Chb" and "CRB-HH").

It is to be noted that *only few* OFDM blocks/symbols are used here, as compared to that in [3] etc.

Fig. 1 a) considers a SISO OFDM system with 64 subcarriers and plots NRMSE of forward composite $\mathbf{b}$ estimated by our novel method, [16], [9] in a 5 multipath channel. Of this, *only* $\bar{T} = 7$ training subcarriers are employed as pilot subcarriers; and 6 OFDM blocks are used. It is to be recalled that even forward, composite (multipath) channel estimation (equation (16) [16], equation (40) [9]) uses suboptimal LS methods (not MAP or even ML estimator, which uses the *noise correlation matrix* $\mathbf{C}$ in its associated likelihood function in equation (16)). The associated CRB lower bounds are plotted in Fig. 1 b). Note that our novel methods, (using the entire received OFDM block of $\tilde{P}$ subcarriers, instead of removing the ZP/CP subcarriers from the received OFDM block, as in most existing estimators) exploit the additional information in received signal block (even those corresponding to the transmitted zero subcarriers). Fig. 2 considers the same system with more training subcarriers ($\bar{T} = 14$), where the performance of "Training" and "Semiblind-EM" estimators are very close to each other. Semiblind estimators perform well, as the channel information is *still embedded* in its received OFDM blocks (which can be extracted by advanced signal processing techniques, as we do in Section IV). The NRMSE of the novel estimator is pretty close to CR lower bound (for most of the SNR range). This result holds, even for long multipath TWRN channel, which has mostly been avoided in the existing literature.

Fig. 3 compares the performance of different forward composite channel ($\mathbf{b}$) estimators for a 32 subcarrier (for each OFDM block) system. Figs 3 a) and b) compares the performances, for $2, 4, 6$ OFDM blocks; the performance improves with more OFDM blocks, as expected.

And in the more difficult reverse channel ($\mathbf{a}$, $\mathbf{h}$) estimation case in Figs. 4-6, the noise is correlated (see equation (10) above). The existing multipath reverse channel ($\mathbf{a}$, $\mathbf{h}$) estimators (equation (16) [16], equation (40) [9]) use suboptimal LS methods, and thus their performance degrades. The novel reverse channel estimator takes care of all correlated noise terms and uses a MAP (*enhanced* from ML) criterion/cost function. Due to the difficulty in reverse channel ($\mathbf{a}$, $\mathbf{h}$) estimation, its NRMSE is higher than that of $\mathbf{b}$ estimator. Simulation results are provided for varying channel length, number of subcarriers and number of OFDM blocks used in estimation. Fig. 5 b) shows that the NRMSE of reverse channel estimator, using 6 blocks, is close to CRB lower bounds. Also, individual channel $\mathbf{g}$ is estimated from composite channel $\mathbf{b}$, and individual reverse channel $\mathbf{h}$ (after they have been estimated), as in Sec. V. A., and its NRMSE also plotted in Figs. 4-6. The NRMSE of $\hat{\mathbf{g}}$ is higher than that of the *other* individual channel $\hat{\mathbf{h}}$, as $\hat{\mathbf{h}}$ is estimated directly from the received signal, whereas estimation of $\hat{\mathbf{g}}$, in equation (60), involves some divisions (by noisy estimates of other channels). Such a situation has also been witnessed for MAP estimated single-tap channel - (see Figs. 7, 8 in [5]).

Fig. 7 plots the NRMSE of reverse channel $\mathbf{h}$ vs. EM iteration no. for SNRs of $2$, $13$ dB. The Factor Analysis-EM algorithm convergences quickly. However, there is a sharp drop in NRMSE of $\mathbf{h}$ vs iteration no., at low SNR of $2dB$, while there is slight drop in NRMSE (from its initial estimate), vs iteration no., at higher SNR of 13 dB, indicating superiority of our novel algorithms, especially in noisy situations.

Fig. 8 compares our novel semiblind-EM channel estimator with that of [3]. We consider a rectangular pulse truncated to 2 symbol periods, so the received signal (at $T_1$) consists of interference, corresponding to 2 transmitted data symbols from $T_2$ (denoted by $\mathbf{s}_2^{(l)}$, $l = 1, 2$) in equation (31) [3]. This results in a $L + 1 = 2$ tap multipath channel, which is just a particular case of our generic $L + 1$-sized $\mathbf{g}$ channel, which again affects composite channel $\mathbf{b}$. Our estimators performs better than ML-based estimator in [3], as we use apriori channel pdf in our novel MAP estimator.

Fig. 9 simulates the composite and individual channel estimators in a CP-OFDM multipath transmission system, with similar results. Following the discussion in **Note** in Sec. II, and also [18], CP-OFDM is converted into a ZP-OFDM system, then our novel methods are applied to it. Even for CP-OFDM system, our novel estimators perform appreciably better than the existing ones.

## IX. Conclusions

Semiblind estimation of multiple forward, reverse, individual and composite channels in bi-directional AF TWRN systems, continues to be a very active area of research to deliver spectrally efficient, high data rate 6G systems. Superior reception and demodulation require non-trivial sophisticated, fast IC/SIC receiver architectures. at the terminals. The noise (including colored noise), at different relay and terminal nodes, make LS methods [16], [9], inadequate for demodulation. As a result, most existing works have concentrated on the easier task of *single*-path channel estimation. These disadvantages are overcome by developing optimal MAP estimators in this paper. Of particular importance is the derivation of an closed-form analytical expression for multipath *reverse* iterative channel estimator, via *the innovative Factor Analysis approach* and using an Alternating Maximization method, and whose performance is superior *to LS methods*. The convergence of the reverse channel $\mathbf{h}$ estimator, proved in (52) , *confirms the validity* of the reverse channel estimator in (49). Estimating $\mathbf{h}$ in this way may be *less* erroneous, than estimating ($\mathbf{a} = (\mathbf{h} * \mathbf{h})$ first, and then obtaining $\hat{\mathbf{h}}$ by de-convolving a noisy estimate of $\mathbf{a}$). *Even for single-path channels, MAP based [5] and [8] do not provide close-form expressions* for the ML reverse channel $(\mathbf{a})$ estimation problem. Instead the likelihood function is maximized via (exhaustive search, in both magnitude and phase dimensions) or suboptimal numerical techniques [5], without providing any insight into expression for the reverse channel estimators. Asynchronization, timing offsets etc. particularly affect TWRN performance, as there may be two/multiple terminal and relay nodes involved. Also bidirectional communications complicate the situation, as the timing offset in a timeslot affect the transmission/reception in the next timeslot. Thus effects of asynchronization, timing offsets have received substantial interest in existing works [12], [2], [4], [3]; however, as discussed in Sec. VI. B., this asynchronization can be accommodated within the *generic* framework of our novel TWRN *full* multipath channel estimators. Simulation results demonstrate the supremacy of the novel estimators over existing methods. Future work include developing time-domain method of full mulipath, along with alleviation of phase noise and carrier offset effects, as well as time-varying TWRN channel estimation in advanced *full-duplex* (FD) TWRN and for other emerging communication systems (with *cascaded channels*), e. g., unmanned aero vehicle (UAV)s [35], Intelligent Reflecting Systems (IRS) [36].

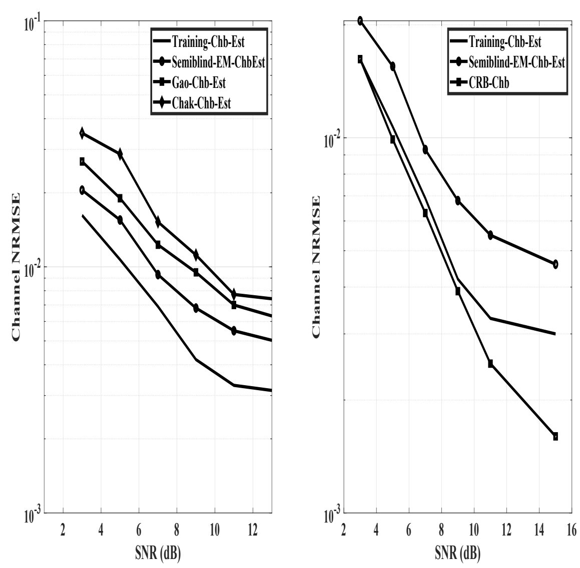



Fig. 1

A) RELAY 5-PATH COMPOSITE (FORWARD) CHANNEL $\mathbf{b}$ NRMSE, 64 SUBCARRIERS IN OFDM BLOCK, 6 BLOCKS, 7 TRAINING SUBCARRIERS ONLY, B) CHANNEL $\mathbf{b}$ NRMSE VS CRAMER-RAO BOUNDS

## REFERENCES


[1] J. Li, L. J. Cimini, J. Ge, C. Zhang, and H. Feng, "Optimal and suboptimal joint relay and antenna selection for two-way amplify and forward relaying," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 980–993, Feb. 2017.

[2] R. Rahimi and S. Shahbazpanahi, "Asynchronous two-way MIMO relaying: A multi-relay scenario," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4270–4287, Aug. 2018.

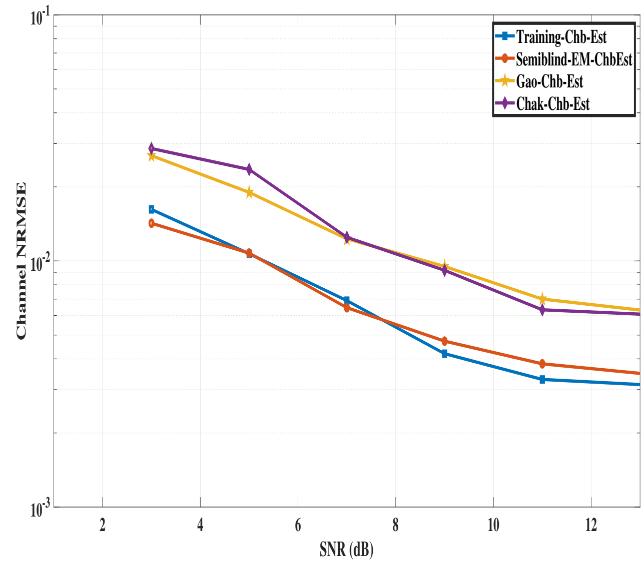




Fig. 2

MORE (14) PILOT SUBCARRIERS, RELAY COMPOSITE (FORWARD) 5-PATH CHANNEL $\mathbf{b}$ NRMSE, 64 SUBCARRIERS IN OFDM BLOCK, 6 BLOCKS


[3] S. Abdallah et. al., "Semi-Blind Joint Timing-Offset and Channel Estimation for AF Two-Way Relaying, " *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 450-463, Jan. 2020.

[4] S. Abdallah, A. I. Salameh, and M. Saad, "Spectrum efficient joint frequency offset and channel estimation for time-asynchronous amplify and forward two-way relay networks," *IEEE Access*, vol. 7, pp. 71 972– 71 985, May 2019.

[5] X. Xie et. al., "Maximum a Posteriori Based Channel Estimation Strategy for Two-Way Relaying Channels, " *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2613-2627, Apr. 2015.

[6] X. Xie et. al., "Maximum a Posteriori Based Phase Noise and Carrier Offset Estimation for Two-Way Relaying Channels, " *IEEE Trans. Wireless Commun.*, 2017.

[7] R. Wang et. al., "Channel Estimation, Carrier Recovery, and Data Detection in the Presence of Phase Noise in OFDM Relay Systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1186-1205, Feb. 2016.

[8] X. Li et. al., "Channel Estimation for Residual Self-Interference in Full Duplex Amplify-and-Forward Two-way Relays," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 4970-4983, Aug. 2018.

[9] S. Chakraborty et. al., "Iterative SAGE-Based Joint MCFOs and Channel Estimation for Full-Duplex Two-Way Multi-Relay Systems in Highly Mobile Environment," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7379-7394, Nov. 2018.

[10] Z. Zhang, Z. Chen, M. Shen, B. Xia, W. Xie, and Y. Zhao, "Performance analysis for training-based multipair two-way full-duplex relaying with massive antennas," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6130– 6145, Jul. 2018.

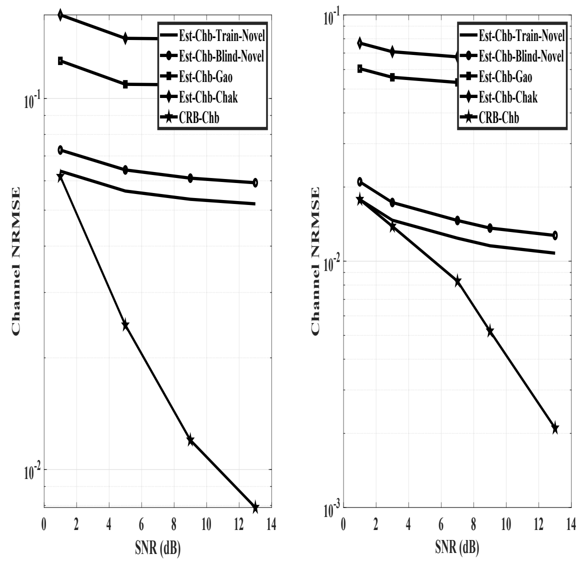
[11] F. L. Duarte et. al., " Cloud-Driven Multi-Way Multiple-Antenna Relay Systems: Joint Detection, Best-User-Link Selection and Analysis," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3342-3354, Jun. 2020.

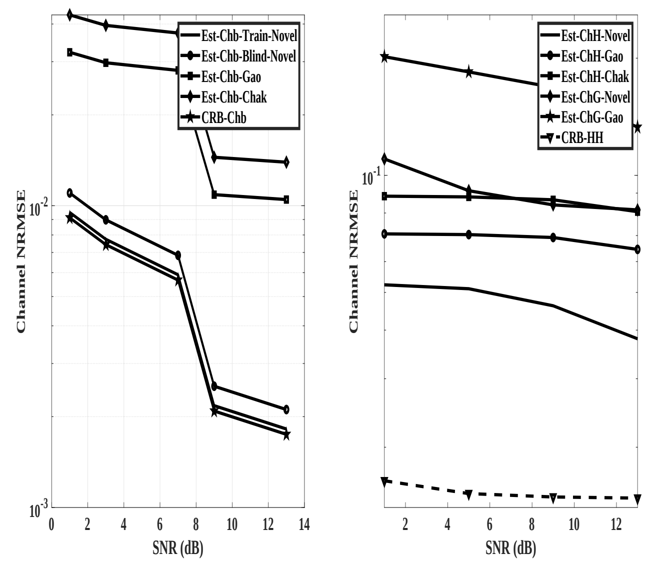
[12] R. Vahidnia, S. Shahbazpanahi, and A. Minasian, "Pre-channel equalization and distributed beamforming in asynchronous single-carrier bidirectional relay networks," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3968–3983, Jan. 2016.

[13] Z. Wei et. al., "Energy Efficiency of Millimeter-Wave Full-Duplex Relaying Systems: Challenges and Solutions," *IEEE Access*, vol. 4, 2017.

[14] C. W. R. Chiong, Y. Rong and Y. Xiang, "Channel Estimation for Time-Varying MIMO Relay Systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 6752-6762, Dec. 2016.

[15] F. Roemer et. al, "Tensor-Based Channel Estimation and Iterative Refinements for Two-Way Relaying With Multiple Antennas and Spatial Reuse," *IEEE Trans. Sig. Proc.*, vol. 58, no. 11, pp. 5720-5735, Nov. 2011.




;

Fig. 3

A) RELAY COMPOSITE (FORWARD) 5-PATH CHANNEL $\mathbf{b}$ NRMSE, 32 SUBCARRIERS OFDM, 2 BLOCKS, B) 4 OFDM BLOCKS

Fig. 4

A) RELAY COMPOSITE (FORWARD) 5-PATH CHANNEL $\mathbf{b}$ NRMSE, 32 SUBCARRIERS OFDM, 6 BLOCKS, B) REVERSE 5-PATH CHANNEL ESTIMATION (FACTOR ANALYSIS, 2 OFDM BLOCKS


[16] F. Gao et. al., "Channel estimation for OFDM modulated two-way relay networks," *IEEE Trans. Sig. Proc.*, vol. 57, no. 11, pp. 4443-4455, Nov. 2009.

[17] F. Gao et. al., "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans Comm*, vo.57, no. 10, pp. 3024-3033, Oct. 2009.

[18] F. Gao et. al., "Robust Subspace Blind Channel Estimation for Cyclic Prefixed MIMO OFDM Systems: Algorithm, Identifiability and Performance Analysis, " *IEEE Journal on Selected Areas in Communications*, vol.26, no. 2, pp.378 - 388, Feb. 2008.

[19] A. Papoulis, "Probability, random variables and stochastic processes," Prengtice Hall, 4th ed., Jul. 2017.

[20] A. Ng, "Machine Learning", CS229, Stanford Univ.

[21] Coursera lectures (Stanford Univ.) on Bayesian Machine Learnnig.

[22] J. Tay, Friedman and Tibshirani, "Principal component guided sparse lasso regression", *arxiv 2019, Stanford Univ*.

[23] G. Strang, *Linear Algebra and Learning from Data*, Cambridge Press, 2019.

[24] J. Ma, S. Zhang, H. Li, F. Gao, S. Jin, "Sparse Bayesian Learning for the Time-varying Massive MIMO Channels: Acquisition and Tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925-1938, March 2019.

[25] P. De et. al. , "Multi-Stage Kalman Filtering (MSKF) based timevarying Sparse Channel Estimation with Fast Convergence,", IEEE Open Journal of Signal Proc, vol. 3, Jan 2022, pp. 21-35. https://doi.org/10.1109/OJSP.2021.3132583

[26] L. Gan, N. Narisetty et, al. "Bayesian Regularization for graphical models with unequal shrinkage", *Journal of American Statistical Assoc*, vol. 114, 2019 - issue 527. https://www.tandfonline.com/doi/full/10.1080/01621459.2018.1482755.

[27] P. De et. al., "Linear Prediction Based Semiblind Channel Estimation for Multiuser OFDM with insufficient guard interval," *IEEE Trans. Wireless Commun.,* vol. 8, no. 12, pp. 5728-5737, Dec. 2009.

[28] P. De, "Semiblind sparse channel estimation using reduced rank filtering," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1418-1433, March 2018.

[29] P. De, "Fast, Reduced Rank Filtering Based Semiblind MIMO OFDM Sparse Channel Estimation", *IEEE Systems Journal*, vol. 15, no. 1, pp. 1036-1048, March 2021.

[30] T. Minka, "Old and new matrix algebra useful for statistics," *MIT notes*, Dec. 28, 2000.

[31] A. Klami et. al., "Group Factor Analysis," *IEEE Trans Neural Netw. & Learn. Syst.*, vol. 9, Sep. 2016.

[32] S. A. Mulaik, *Foundations of Factor Analysis*, CRC Press, 2nd. ed., 2009.

[33] D. Ramirez et. al.,"Multichannel Factor Analysis with Common and Unique Factors," *IEEE Trans. Sig. Proc.*, pp. 1-14, 2019.

[34] Tareq Y. Al-Naffouri, "An EM-Based Forward-Backward Kalman Filter for the Estimation of Time-Variant Channels in OFDM," *IEEE Trans. Commun.*, vol. 55, no. 7, pp. , Jul. 2008.

[35] J. Du et. al. "Robust Tensor-based algorithm for UAV-assisted IoT communication system via nested PARFAC analyis," *IEEE Trans. Sig. Proc.*, vol. 70, pp. 5117-5132, Oct 2022.

[36] S. Zhang and R. Zhang, "Capacity Characterization for Intelligent Reflecting Surface Aided MIMO Communication," *IEEE Journal on Selected Areas in Commun*, Jul 2020.


.

## Appendix A: Reconstruction of individual channels $\mathbf{h}, \mathbf{g}$ from composite channels $\mathbf{a}, \mathbf{b}$

Let us start with $3(L+1=3)$ channels whose magnitudes are $\mathbf{h} = [h_3,\, h_2,\, h_1]$ and $\mathbf{g} = [g_3,\, g_2,\, g_1]$ respectively (their phases will be determined separately). Then $\mathbf{a} = \mathbf{h} * \mathbf{h} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \end{bmatrix}$, given by

$$a_1 = h_1^2,\, a_2 = 2h_1h_2,\, a_3 = 2h_1h_3 + h_2^2, \\ a_4 = 2h_2h_3, a_5 = h_3^2. \tag{56}$$

Then the magnitude of the reverse composite channel is given by $\mathbf{a} = [\mathbf{h} * \mathbf{h}] = [a_1, a_2, \cdots, a_{\bar{L}}]$, where $h_l$ , is the $l$th channel tap *magnitude* of $\mathbf{h}$. (the phase of $\mathbf{h}$ will be dealt with later on). Then the individual channel $\mathbf{h}$ can be reconstructed from $\mathbf{a}$ by time-domain deconvolution as

$$h_1 = \sqrt{a_1},\, h_2 = \frac{a_2}{2h_1} = \frac{a_2}{2\sqrt{a_1}},\, h_3 = \sqrt{a_5} \tag{57}$$

Also, the following relationships exist,

$$h_2 = (a_4)/(2\sqrt{a_5}) = \sqrt{a_3 - 2\gamma},\, \gamma \stackrel{\Delta}{=} \sqrt{a_1 a_5} \tag{58}$$

The 2 equations above will be used in Appendix B to derive apriori pdfs of composite channels $\mathbf{a}, \mathbf{b}$. Similarly, $L = 4$, then $\mathbf{h}, \mathbf{a}$ are $5 \times 1$ and $(\bar{L} = 9) \times 1$-sized vectors. Then $\mathbf{h}$ is obtained from $\mathbf{a}$ by, $\mathbf{h} = [\sqrt{a_1},\, a_2/(2\sqrt{a_1}),\, (a_4 - a_8/(2\sqrt{a_{\bar{L}}}))/a_2,\, a_8/(2\sqrt{a_{\bar{L}}}),\, \sqrt{a_{\bar{L}}}]^T$ Extension to generic $L$ multipath channels is given at the end of this section.

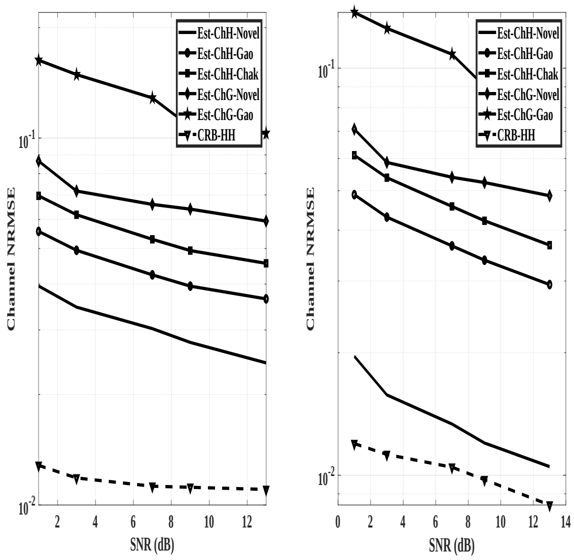



Fig. 5

REVERSE RELAY 5-PATH CHANNEL $\mathbf{h}$ NRMSE, 32 SUBCARRIERS OFDM, USING FACTOR ANALYSIS, A) 4 OFDM BLOCKS, B) 6 OFDM BLOCKS.

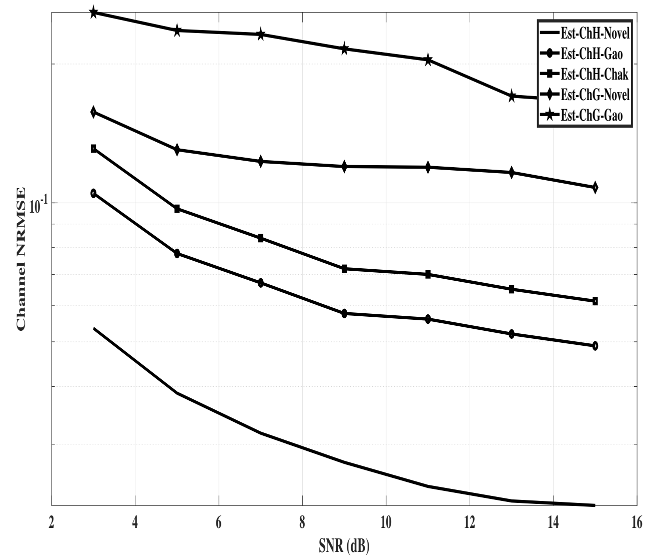



Fig. 6

REVERSE RELAY 3-PATH CHANNEL $\mathbf{h}$ NRMSE, 32 SUBCARRIERS OFDM, 3 OFDM BLOCKS.

The other composite channel $\mathbf{b} = \mathbf{g} * \mathbf{h}$ is given by

$$\begin{aligned}&\mathbf{b} = [b_1 b_2 \cdots b_{\bar{L}}],\\ &b_1 = h_1 g_1,\, b_2 = h_2 g_1 + h_1 g_2,\, b_3 = g_1 h_3 + g_2 h_2 + g_3 h_1,\\ &b_4 = g_2 h_3 + g_3 h_2,\, b_5 = g_3 h_3. \end{aligned} \tag{59}$$

Then from $\mathbf{b}$, individual channel $\mathbf{g}$ is reconstructed by

$$\begin{aligned}&g_1 = b_1/h_1 = b_1/\sqrt{a_1},\, g_3 = b_5/\sqrt{a_5},\\ &g_2 = (b_2\sqrt{a_1} - \alpha b_1)/a_1,\, \alpha \stackrel{\Delta}{=} (2a_4 - a_2)/(2\sqrt{a_5}). \end{aligned} \tag{60}$$

The above analysis can be *easily* extended to *any* value of $L+1, \bar{L} = (2L+1)$.The first and the last (i. e. $\bar{L}$)th points of the convolution have only one term, involving one $h_i$ coefficient (for $\mathbf{a}_i$) or a product $h_i g_i$ (for $\mathbf{b}_i$), This allows us to find the conditional pdfs of $a_1$, $a_{\bar{L}}$, $b_1$, $b_{\bar{L}}$ directly. The second and second-last terms of the convolution are a summation of one/two terms invloving 2 coefficients of $h_i, g_i$ (see $a_2, a_4, b_2, b_4$). Their pdfs are determined next. Continuing in this way, the third and third-last of the convolution are a summation of two or third terms, their pdfs are next determined in the same way as $L+1 = 3, \bar{L} = 5$ example above. This is then continued for the other points of the convolution. This analysis *holds* for any arbitrary value of $L$. If value of $L$ is known apriori, the equations for reconstructing individual channels $\mathbf{h}, \mathbf{g}$ from composite channels, can be easily determined beforehand. This method can be programmed on a computer for a generic value of $L$. *Moreover*, equations (56) to (60) will be used in Appendix B to evaluate the apriori pdfs of the composite channels $\mathbf{a}$, $\mathbf{b}$.

**Appendix B: Proof of Lemma 1: Derivation of Apriori PDFs of composite channels $\mathbf{a}, \mathbf{b}$**

Since complex $\bar{\mathbf{h}}, \bar{\mathbf{g}}$ are Gaussian distributed with variance $v$, their magnitudes $\mathbf{h}, \mathbf{g}$ are Rayleigh distributed. Then the pdf of each component of $\mathbf{h}$ is $f_{h_i}(x) = \frac{2x}{v_1} e^{-\frac{x^2}{v}}$. Then using (56) to (58), and using the formula for pdf of a function of random variable, the pdf of $\mathbf{a}$ is

$$\begin{aligned}&f_{a_1}(a_1) = f_{h_1}(h_1 = \sqrt{a_1})/(|\partial a_1/\partial h_1|)\\ &= (2\sqrt{a_1}/v)e^{-a_1/v}/2\sqrt{a_1} = e^{-a_1/v}/v,\\ &f_{a_5}(a_5) = f_{h_3}(h_3) = \sqrt{a_5})/(|\partial a_5/\partial h_3|)\\ &= (2\sqrt{a_5}/v)e^{-a_5/v}/2\sqrt{a_5} = e^{-a_5/v}/v\\ &h_2 = \frac{a_2}{2\sqrt{a_1}} \implies f_{a_2}(a_2|a_1) = f_{h_2}(h_2 =\\ &(a_2/2\sqrt{a_1}))/(2\sqrt{a_1}) = (a_2/(a_1 v))e^{-a_2^2/(4a_1 v)},\\ &h_2 = (a_4)/2\sqrt{a_5} \implies f(a_4|a_5) = (2a_4)/(va_5)\times\\ &e^{-a_4{}^2/(4a_5 v)},\\ &h_2 = \sqrt{a_3 - 2\sqrt{a_1 a_5}} \implies f_{a_3}(a_3|a_1, a_5)\\ &= ((2\sqrt{a_3 - 2\gamma})/v) \times e^{-(a_3-2\gamma)/v},\, \gamma = \sqrt{a_1 a_5},\\ &f_{\mathbf{a}}(\mathbf{a}) = f_{a_1}(a_1) f_{a_2}(a_2|a_1) f_{a_3}(a_3|a_1, a_5) f(a_4|a_5) f_{a_5}(a_5) \end{aligned} \tag{61}$$

The phase of $\mathbf{a}$ is =$[\theta_1, \theta_2, \cdots, \theta_{\bar{L}}]$ are distributed jointly by unform distribution as

$$f([\theta_1, \theta_2, \cdots, \theta_5]) = \frac{1}{(2\pi)^{\bar{L}}}, 0 < \theta_i < 2\pi.$$

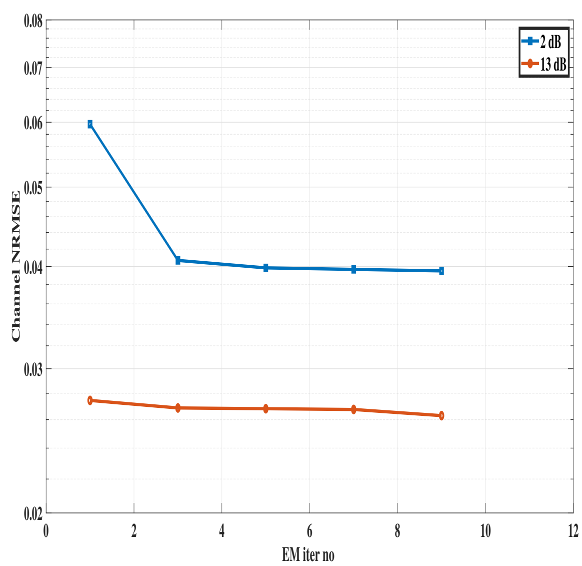



Fig. 7
REVERSE RELAY CHANNEL $\mathbf{h}$ ESTIMATION VS EM ITERATION NO., 4 OFDM BLOCKS, FOR DIFFERENT SNRS.

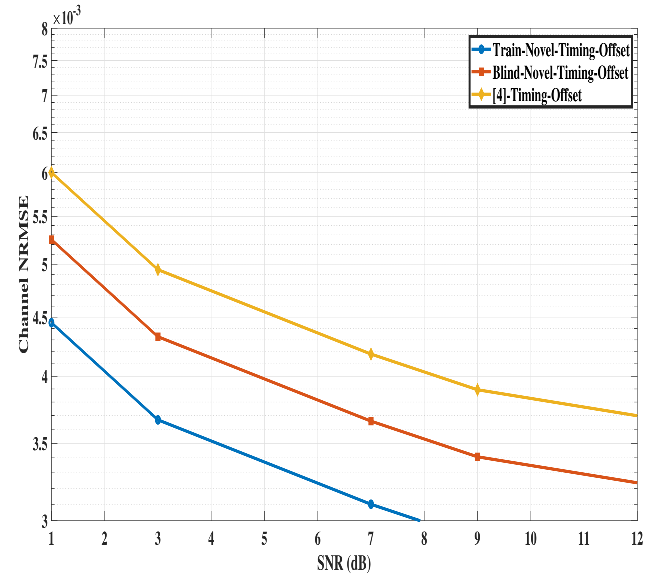



Fig. 8
NRMSE OF ESTIMATORS, WITH TIMING OFFSETS

Similarly, employing (60), the pdf of $\mathbf{b} = \mathbf{g} * \mathbf{h}$ is

$$\begin{aligned}
&f_{b_1}(b_1|a_1) = f_{g_1}(g_1 = b_1/\sqrt{a_1})/(|\partial b_1/\partial g_1|) \\
&= (2b_1/(a_1 v))e^{-b_1^2/(a_1 v)}, \\
&f_{b_5}(b_5) = f_{g_3}(g_3 = b_5/\sqrt{a_5})/(|\partial b_5/\partial g_3|) \\
&= (2b_5/(a_5 v))e^{-b_5^2/(a_5 v)} \\
&f(b_2|b_1, a_2, a_4, a_5) = 2(b_2\sqrt{a_1} - \alpha b_1)/(a_1 v) \\
&e^{-(b_2\sqrt{a_1}-\alpha b_1)^2/a_1 v}, \alpha = (2a_4 - a_2)/(2\sqrt{a_5}), \\
&f(b_4) = [2(b_4\sqrt{a_1} - \sqrt{a_5}\beta)/(\alpha^2\sqrt{a_1}v)] \\
&e^{-(b_4\sqrt{a_1}-\sqrt{a_5}\beta)^2/(\alpha^4 a_1 v)},\ \beta = b_2\sqrt{a_1} - \alpha b_1. \qquad (62)
\end{aligned}$$

Now it remains for us to determine

$$\begin{aligned}
&b_3 = h_3 g_1 + h_2 g_2 + h_1 g_3 \implies \\
&g_3 = (b_3 a_1 - \bar{\beta})/a_1,\ \bar{\beta} = [b_1\sqrt{a_5} + \alpha(b_2 - \alpha b_1)], \\
&f_{b_3}(b_3|a_i, b_i) = 2(b_3 a_1 - \bar{\beta})/(a_1 v)e^{-(b_3 a_1-\bar{\beta})^2/(a_1^2 v)}. \qquad (63)
\end{aligned}$$

Then the joint pdf of $\{\ \mathbf{a}, \mathbf{b}\}$ is

$$\begin{aligned}
&f(\mathbf{a}, \mathbf{b}) = e^{-a_1/v}/v \times a_2/(a_1 v)e^{-a_2^2/(4a_1 v)} \\
&\times ((2\sqrt{a_3 - 2\gamma})/v)e^{-(a_3-2\gamma)/v}\times \\
&2a_4/(va_5)e^{-{a_4}^2/(4a_5 v)} \times e^{-a_5/v}/v\times \\
&2b_1/(a_1 v)e^{-b_1^2/(a_1 v)} \times 2(b_2\sqrt{a_1} - \alpha b_1)/(a_1 v) \\
&e^{-(b_2\sqrt{a_1}-\alpha b_1)^2/a_1 v} \times 2(b_3 a_1 - \bar{\beta})/(a_1 v) \\
&e^{-(b_3 a_1-\bar{\beta})^2/(a_1^2 v)} \times [2(b_4\sqrt{a_1} - \sqrt{a_5}\beta)/(\alpha^2\sqrt{a_1}v)] \\
&e^{-(b_4\sqrt{a_1}-\sqrt{a_5}\beta)^2/(\alpha^4 a_1 v)} \times 2b_5/(a_5 v)e^{-b_5^2/(a_5 v)} \qquad (64)
\end{aligned}$$

The logarithm of $f(\mathbf{a}, \mathbf{b})$, is then given in (11).

As in Appendix A, $f(\mathbf{a}, \mathbf{b})$ can be derived for any arbitrary value of $L$. This method can be programmed on a computer for a generic value of $L$. Since individual channels $\mathbf{h}, \mathbf{g}$ are Rayleigh distributed, expressing the composite channels $\mathbf{a}, \mathbf{b}$ as functions of $\mathbf{h}, \mathbf{g}$ (see Appendix A), allow their pdfs to be determined, by using the formula for the pdf of functions of random variables.

**C: Proof of Lemma 2**

The 1st term of (15) is

$$\begin{aligned}
&-\frac{1}{2}([\bar{\mathbf{z}}_i^H \mathbf{C}^{-1}\bar{\mathbf{z}}_i] + \mathbf{b}^H E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}[\bar{\mathbf{D}}_i^{(2)H}\mathbf{C}^{-1}\bar{\mathbf{D}}_i^{(2)}]\mathbf{b} \\
&-2\Re\{\bar{\mathbf{z}}_i^H \mathbf{C}^{-1} E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}[\bar{\mathbf{D}}_i^{(2)}]\mathbf{b}\}), \qquad (65)
\end{aligned}$$

In (65), it is seen that one needs to compute two Expectation terms, namely, $E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}[\bar{\mathbf{D}}_i^{(2)H}\mathbf{C}^{-1}\bar{\mathbf{D}}_i^{(2)}]$ and $E_{\bar{\mathbf{D}}_i^{(2)}|\bar{\mathbf{z}}_i}[\bar{\mathbf{D}}_i^{(2)}]$, for which the required pdfs have to be derived. We assume that the transmitted discrete data $d_i^{(2)}$ is independent, uniformly distributed, with pdf $f(d_i^{(2)} = A_l) = \frac{1}{M}$, $A_l = 1, 2, \cdots, M$. Then each column $c$ of $\bar{\mathbf{D}}_i^{(2)}$, denoted by $[\bar{\mathbf{D}}^{(2)}]_c$ has the pdf, (skipping the subscript $\{i\}$, for ease of notation), $f([\bar{\mathbf{D}}^{(2)}]_c = \mathbf{A}_l]) = \frac{1}{M}^N$, since each column has *only* $N$ random $d_i^{(2)}$'s ($N$: number of subcarriers in each OFDM block), the rest are zero entries in $[\bar{\mathbf{D}}^{(2)}]_c$. Now, the conditional pdf $f(\bar{\mathbf{D}}_c^{(2)} = \mathbf{A}_l|\bar{\mathbf{z}}_i, \mathbf{b})$ is given by

$$\begin{aligned}
&f(\bar{\mathbf{D}}_c^{(2)} = \mathbf{A}_l|\bar{\mathbf{z}}_i, \mathbf{b}) = \frac{f(\bar{\mathbf{z}}_i|\mathbf{b}, \bar{\mathbf{D}}_c^{(2)} = \mathbf{A}_l)f(\bar{\mathbf{D}}_c^{(2)}) = \mathbf{A}_l)}{\sum_{l=1}^{MN} f(\bar{\mathbf{z}}_i|\mathbf{b}, \bar{\mathbf{D}}_c^{(2)} = \mathbf{A}_l)f(\bar{\mathbf{D}}_c^{(2)}) = \mathbf{A}_l)} \\
&= \frac{exp^{-\frac{1}{2}[\bar{\mathbf{z}}_i-\mathbf{A}_l\mathbf{b}]^H\mathbf{C}^{-1}[\bar{\mathbf{z}}_i-\mathbf{A}_l\mathbf{b}]}}{\sum_{t=1}^{MN} exp^{-\frac{1}{2}[\bar{\mathbf{z}}_i-\mathbf{A}_t\mathbf{b}]^H\mathbf{C}^{-1}[\bar{\mathbf{z}}_i-\mathbf{A}_l\mathbf{b}]}} \qquad (66)
\end{aligned}$$

Finally over the independent $\bar{L}$ columns of $\bar{\mathbf{D}}_c^{(2)}$, the data matrix $\bar{\mathbf{D}}^{(2)}$ has the pdf

$$\begin{aligned}
&f(\bar{\mathbf{D}}^{(2)} = [\mathbf{A}_l]_{\bar{L}\,columns}|\bar{\mathbf{z}}_i, \mathbf{b}) \\
&= \frac{(exp^{-\frac{1}{2}\sum_{l=1}^{\bar{L}}[\bar{\mathbf{z}}_i-\mathbf{A}_l\mathbf{b}]^H\mathbf{C}^{-1}[\bar{\mathbf{z}}_i-\mathbf{A}_l\mathbf{b}]})}{[\sum_{t=1}^{MN} exp^{-\frac{1}{2}[\bar{\mathbf{z}}_i-\mathbf{A}_t\mathbf{b}]^H\mathbf{C}^{-1}[\bar{\mathbf{z}}_i-\mathbf{A}_l\mathbf{b}]}]^{\bar{L}}} \overset{\Delta}{=} \frac{A}{B}, \qquad (67)
\end{aligned}$$

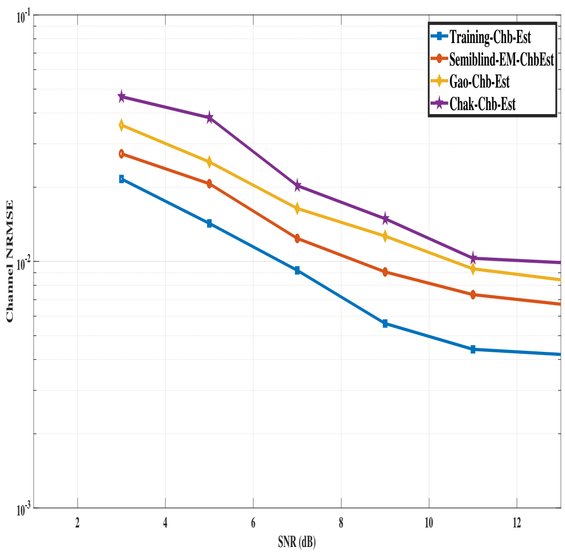



Fig. 9
CP-OFDM COMPOSITE (FORWARD) 5-PATH CHANNEL $\mathbf{b}$ NRMSE, 64 SUBCARRIERS IN OFDM BLOCK, 6 BLOCKS

where $A = exp^{-\frac{1}{2}\sum_{l=1}^{\bar{L}}[\bar{\mathbf{z}}_i-\mathbf{A}_l\mathbf{b}]^H\mathbf{C}^{-1}[\bar{\mathbf{z}}_i-\mathbf{A}_l\mathbf{b}]}$ and $B = [\sum_{t=1}^{M^N} exp^{-\frac{1}{2}[\bar{\mathbf{z}}_i-\mathbf{A}_t\mathbf{b}]^H\mathbf{C}^{-1}[\bar{\mathbf{z}}_i-\mathbf{A}_l\mathbf{b}]}]^{\bar{L}}$. Then

$$E[\bar{\mathbf{D}}_i^{(2)}=\mathbf{A}_l|\mathbf{b},\mathbf{z}_i] = \sum_{l=1}^{M^N}[\mathbf{A}_l f(\bar{\mathbf{D}}^{(2)}|\bar{\mathbf{z}}_i,\mathbf{b})] = \sum_{l=1}^{M^N}\mathbf{A}_l\frac{A}{B}, \quad (68)$$

$$E[\bar{\mathbf{D}}_i^{(2)}\mathbf{C}^{-1}\bar{\mathbf{D}}_i^{(2)H}|\bar{\mathbf{b}},\mathbf{z}_i] = \sum_{l=1}^{M^N}\mathbf{A}_l\mathbf{C}^{-1}\mathbf{A}_l{}^H\frac{A}{B}, \quad (69)$$

where $\mathbf{A}_l$ is the alphabet for $\bar{\mathbf{D}}^{(2)}$, at the $l$th experiment, and and $E[\bar{\mathbf{D}}_i^{(2)}\mathbf{C}^{-1}\bar{\mathbf{D}}_j^{(2)H}|\bar{\mathbf{b}},\mathbf{z}_i] \approx \mathbf{0}$, $i \neq j$ (see note [6] in [34]). We use the notation $\mathbf{s}^H\mathbf{C}^{-1}\mathbf{s}\triangleq\|\mathbf{s}\|^2_{\mathbf{C}^{-1}}$ (for any vector $\mathbf{s}$)), Defining for the $i$th OFDM block's, $\mathbf{X}_i\triangleq\sum_{l=1}^{M^N}[\mathbf{A}_{l,i}]_{\bar{L}\,columns}\frac{A}{B}$, we have

$$E_{\bar{\mathbf{D}}_i^{(2)}}\{\mathcal{L}\} = \bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\bar{\mathbf{z}}_i - 2\Re\{\bar{\mathbf{z}}_i^H\mathbf{C}^{-1}\mathbf{X}\mathbf{b}\}$$
$$+(\frac{B}{A})\mathbf{b}^H(\sum_{l=1}^{M^N}[\mathbf{X}\mathbf{C}^{-1}\mathbf{X}^H])\mathbf{b} + other\ terms. \quad (70)$$

**D: Proof of Lemma 3**

For 2 Gaussian random variables $x_1$ and $x_2$, the conditional pdf of $x_1|x_2$ is *still* Gaussian [20] and its conditional mean and covariance matrix are $\mu_{x_1|x_2} = \mu_{x_1} + Cov(x_1,x_2)[Cov(x_2)]^{-1}(x_2-\mu_2)$, $Cov(x_1|x_2) = Cov(x_1) - Cov(x_1,x_2)[Cov(x_2)]^{-1}Cov(x_2,x_1)$. Consider the 2 dimensional Gaussian random vector $\mathbf{x} = \{[\tilde{\mathbf{z}}_i^T,\mathbf{n}_r^T]^T\}$. Then, Using (34), we have

$$\mu_{\mathbf{x}} = \mu_{[\tilde{\mathbf{z}}_l,\mathbf{n}_{rl}]^H} = [\mathbf{0},\mathbf{0}]^H,$$
$$Cov(\mathbf{n}_r,\tilde{\mathbf{z}}_l) = E\{\mathbf{n}_r(\mathbf{A}\mathbf{d}_l^{(1)}+\mathbf{H}\mathbf{n}_{rl}+\mathbf{n}_{1l})^H\}$$
$$= E\{\mathbf{n}_r\mathbf{d}_l^{(1)H}\} + E\{\{\mathbf{n}_r(\mathbf{n}_r^H\}\mathbf{H}^H + E\{\mathbf{n}_r\mathbf{n}_1^H)\} = \sigma^2\mathbf{H}^H,$$
$$Cov(\tilde{\mathbf{z}}_l) = \tilde{\mathbf{C}} \quad (71)$$

Then

$$Cov(\mathbf{x}) = \begin{bmatrix} Cov(\tilde{\mathbf{z}}_l) & E(\tilde{\mathbf{z}}_l\mathbf{n}_r^H) \\ E(\mathbf{n}_r\tilde{\mathbf{z}}_l^H) & E(\mathbf{n}_r\mathbf{n}_r^H)\end{bmatrix} = \begin{bmatrix}\tilde{\mathbf{C}} & \sigma^2\mathbf{H} \\ \sigma^2\mathbf{H}^H & \sigma^2\mathbf{I}\end{bmatrix}. \quad (72)$$

Then

$$\mu_{\tilde{\mathbf{z}}_l|\mathbf{n}_r} = \mu_{\mathbf{z}_l} + Cov(\tilde{\mathbf{z}},\mathbf{n}_r)[Cov(\mathbf{n}_r)]^{-1}(\mathbf{n}_r)$$
$$= (\sigma^2)\mathbf{H})[\sigma^2\mathbf{I}]^{-1}\mathbf{n}_r = \mathbf{H}\mathbf{n}_r,$$
$$Cov(\tilde{\mathbf{z}}_l|\mathbf{n}_r) = Cov(\tilde{\mathbf{z}}_l) - Cov(\tilde{\mathbf{z}}_l,\mathbf{n}_r)[Cov(\mathbf{n}_r)]^{-1}$$
$$\cdot Cov(\mathbf{n}_r,\tilde{\mathbf{z}}_l) = \tilde{\mathbf{C}} - (\sigma^2\mathbf{H})\frac{\mathbf{I}}{\sigma^2}\sigma^2\mathbf{H}^H = \sigma^2\mathbf{I}.$$

Defining differently, $\mathbf{x} = [\mathbf{n}_r^T\,\tilde{\mathbf{z}}_l^T]^T$, and proceeding in the same way as (72), we have

$$\mu_{\mathbf{n}_r|\tilde{\mathbf{z}}_l} = \mu_{\mathbf{n}_r} + Cov(\mathbf{n}_r,\tilde{\mathbf{z}}_l)[Cov(\tilde{\mathbf{z}}_l)]^{-1}(\tilde{\mathbf{z}}_l)$$
$$= \mathbf{0} + \sigma^2\mathbf{H}^H(\tilde{\mathbf{C}})^{-1}\tilde{\mathbf{z}}_l,$$
$$Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l) = Cov(\mathbf{n}_r) - Cov(\mathbf{n}_r,\tilde{\mathbf{z}}_l)[Cov(\tilde{\mathbf{z}}_l)]^{-1}$$
$$\cdot Cov(\tilde{\mathbf{z}}_l,\mathbf{n}_r) = \sigma^2\mathbf{I} - [\sigma^2\mathbf{H}^H[\tilde{\mathbf{C}}]^{-1}\mathbf{H}\sigma^2]. \quad (73)$$

**E: Proof of Lemma 4,**

**Derivation of Cramer-Rao Lower Bounds (CRLB) for estimating composite channels $\mathbf{a}, \mathbf{b}$**

**E.1 CRLB for $\mathbf{b}$ channel**

The CR bounds for ML channel estimates (without using any *prior* channel information), are derived in this Subsection.

Re-defining $\{\mathcal{L}\} = E_{\bar{\mathbf{D}}_i^{(2)}}\{\mathcal{L}\}$ (with slight abuse of notation), the complex Fisher information matrix (FIM) is defined, [17],

$$\mathbf{J}_{\mathbf{b},\mathbf{b}} = E\{\frac{\partial\{\mathcal{L}\}}{\partial\bar{\mathbf{b}}^*}(\frac{\partial\{\mathcal{L}\}}{\partial\bar{\mathbf{b}}^*})^H\} = E\{\frac{\partial\{\mathcal{L}\}}{\partial\bar{\mathbf{b}}^*}\frac{\partial\{\mathcal{L}\}}{\partial\bar{\mathbf{b}}^T}\}, \quad (74)$$

Since complex $\mathbf{b} = \mathbf{b}_a + j\mathbf{b}_b$ and using the formula for derivatives of (product of functions), we have

$$\frac{\partial E\{\mathcal{L}\}}{\partial\mathbf{b}^*} = \frac{\partial E\{\mathcal{L}\}}{\partial\mathbf{b}_a} + j\frac{\partial E\{\mathcal{L}\}}{\partial\mathbf{b}_b}$$
$$\frac{\partial E\{\mathcal{L}\}}{\partial\mathbf{b}_a} = \frac{\partial[\bar{\mathbf{z}}_i - \sum_l\mathbf{A}_l\mathbf{b}]^H\mathbf{C}^{-1}[\bar{\mathbf{z}}_i - \sum_{l'}\mathbf{A}_{l'}\mathbf{b}]}{\partial\mathbf{b}_a}$$
$$= -\sum_l\mathbf{A}_l^H\mathbf{C}^{-1}[\bar{\mathbf{z}}_i - \sum_{l'}\mathbf{A}_{l'}\mathbf{b}] + [\bar{\mathbf{z}}_i - \sum_l\mathbf{A}_l\mathbf{b}]^H$$
$$\cdot\mathbf{C}^{-1}(-\sum_{l'}\mathbf{A}_{l'})$$
$$\frac{\partial E\{\mathcal{L}\}}{\partial\mathbf{b}_b} = \sum_l j\mathbf{A}_l^H\mathbf{C}^{-1}[\bar{\mathbf{z}}_i - \sum_{l'}\mathbf{A}_{l'}\mathbf{b}]$$
$$+[\bar{\mathbf{z}}_i - \sum_l\mathbf{A}_l\mathbf{b}]^H\mathbf{C}^{-1}(-j\sum_{l'}\mathbf{A}_{l'}). \quad (75)$$

Then, we have

$$\frac{\partial E\{\mathcal{L}\}}{\partial\mathbf{b}^*} = -2\sum_l\mathbf{A}_l^H\mathbf{C}^{-1}[\bar{\mathbf{z}}_i - \sum_{l'}\mathbf{A}_{l'}\mathbf{b}]$$
$$\frac{\partial\{\mathcal{L}\}}{\partial\bar{\mathbf{b}}^T} = (\frac{\partial\{\mathcal{L}\}}{\partial\bar{\mathbf{b}}})^T = ([\bar{\mathbf{z}}_i - \sum_l\mathbf{A}_l\mathbf{b}]^H\mathbf{C}^{-1}(-\sum_{l'}\mathbf{A}_{l'}))^T \quad (76)$$

Then the FIM is given by

$$\begin{aligned}\mathbf{J}_{\mathbf{b},\mathbf{b}} &= E\{\frac{\partial\{\mathcal{L}\}}{\partial\bar{\mathbf{b}}^*}\frac{\partial\{\mathcal{L}\}}{\partial\bar{\mathbf{b}}^T}\\ &= 4[(-\sum_l \mathbf{A}_l^H)\mathbf{C}^{-1}[\bar{\mathbf{z}}_i - \sum_{l'}\mathbf{A}_{l'}\mathbf{b}]][([\bar{\mathbf{z}}_i - \sum_l \mathbf{A}_l\mathbf{b}]^H\\ &\cdot \mathbf{C}^{-1}(-\sum_{l'}\mathbf{A}_{l'}))^T]\\ &= 4[(\sum_l \mathbf{A}_l^H)\mathbf{C}^{-1}[\bar{\mathbf{z}}_i - \sum_{l'}\mathbf{A}_{l'}\mathbf{b}]][(\sum_{l'}\mathbf{A}_{l'}^H)\\ &\cdot \mathbf{C}^{-1}[\bar{\mathbf{z}}_i - \sum_l \mathbf{A}_l\mathbf{b}]^*]\\ &= 4[\sum_l ||\mathbf{A}_l^H\mathbf{C}^{-1}[\mathbf{Hn}+\mathbf{n}]||^2]\\ &= 4[\sum_l ||\mathbf{A}_l^H||^2][(||\mathbf{H}||^2+\mathbf{I})\sigma^2)^{-2}][||(\mathbf{H}||^2+\mathbf{I})\sigma^2]\\ &= 4[\sum_l ||\mathbf{A}_l^H||^2][(||\mathbf{H}||^2+\mathbf{I})\sigma^2)^{-2}(||\mathbf{H}||^2+\mathbf{I})\sigma^2)]\\ &= 4[\sum_l ||\mathbf{A}_l^H||^2][(||\mathbf{H}||^2+\mathbf{I})\sigma^2]^{-1} \end{aligned} \tag{77}$$

since

$$\bar{\mathbf{z}}_i = \sum_l \mathbf{A}_l\mathbf{b} + \mathbf{Hn} + \mathbf{n} \Rightarrow [\bar{\mathbf{z}}_i - \sum_l \mathbf{A}_l\mathbf{b}] = \mathbf{Hn} + \mathbf{n}, \tag{78}$$

and

$$|\mathbf{C}| = |\mathbf{HH}^H + \mathbf{I}|\sigma^2 = (|\mathbf{H}|^2 + \mathbf{I})\sigma^2$$

Finally, (77) becomes

$$\begin{aligned}&\mathbf{J}_{\mathbf{b},\mathbf{b}} = 4[\sum_l \mathbf{A}_l^H[\mathbf{Hn}+\mathbf{n}_1]\mathbf{C}^{-1}][-\mathbf{C}^{-*}[\mathbf{H}^*\mathbf{n}^*+\mathbf{n}_1^*]\\ &\sum_{l'}(\mathbf{A}_{l'}^T) = 4E\{[\sum_l \mathbf{A}_l^H||[\mathbf{Hn}+\mathbf{n}_1]\mathbf{C}^{-1}]||^2[\sum_{l'}\mathbf{A}_{l'}^T)]\}\}\\ &= 4tr\{[\sum_l \mathbf{A}_l^H[||\mathbf{H}||^2+\mathbf{I}]\sigma^2(||\mathbf{HH}^H||+\mathbf{I})^{-2}\sigma^{-4}(\mathbf{A}_l^*)^H\\ &= 4tr\{\sum_l((||\mathbf{H}||^2+\mathbf{I})\sigma^2)^{-1}(||\mathbf{A}_l||^{2H}) \end{aligned} \tag{79}$$

Then the Cramer Rao bound for parameter $\mathbf{b}(CRLB_{\mathbf{b}}$ is

$$CRLB_{\mathbf{b}} = \mathbf{J}_{\mathbf{b},\mathbf{b}}{}^{-1} = (1/4)\sum_l((||\mathbf{H}||^2+\mathbf{I})\sigma^2)[||\mathbf{A}_l^H||^2]^{-1} \tag{80}$$

since data $\mathbf{A}$ (generated in $l$th and $l^{'}$th experiments) are zero mean and uncorrelated with each other. Since in training-based estimation of $\mathbf{b}$, transmitted data $\tilde{\mathbf{D}}_l^{(2)}$ (from terminal T2) is available, then

$$CRLB_{\mathbf{b}} = (1/4)\sum_l((||\mathbf{H}||^2+\mathbf{I})\sigma^2)[||\tilde{\mathbf{D}}_l^{(2)}||^2]^{-1}, \tag{81}$$

for $l$th experiment.

**C.2 CRLB for $\mathbf{H}$ (and $\mathbf{h}$) channel**

Let complex channel matrix $\mathbf{H} = \mathbf{H}_R + j\mathbf{H}_I$. From (42),

$$\begin{aligned}&\mathcal{L} =\\ &\sum_{l=1}^{S} E_{\mathbf{n}|\tilde{\mathbf{z}}_l}[\frac{1}{2}(\tilde{\mathbf{z}}_l - [\mathbf{H}_R + j\mathbf{H}_I]\mathbf{n}_r)^H\frac{1}{\sigma^2}\\ &(\tilde{\mathbf{z}}_l - [\mathbf{H}_R + j\mathbf{H}_I]\mathbf{n}_r)] + \cdots, \end{aligned} \tag{82}$$

$$\frac{\partial\mathcal{L}}{\partial\mathbf{H}} = \frac{1}{2}[\frac{\partial\mathcal{L}}{\partial\mathbf{H}_R} - j\frac{\partial\mathcal{L}}{\partial\mathbf{H}_I}]$$

Using product rule of differentiation,

$$\begin{aligned}&\partial\mathcal{L}/\partial\mathbf{H}_R =\\ &\frac{1}{\sigma^2}E\{[\partial(\tilde{\mathbf{z}}_l - [\mathbf{H}_R + j\mathbf{H}_I]\mathbf{n}_r)^H/\partial H_R](\tilde{\mathbf{z}}_l - [\mathbf{H}_R + j\mathbf{H}_I]\mathbf{n}_r)]\\ &+ (\tilde{\mathbf{z}}_l - [\mathbf{H}_R + j\mathbf{H}_I]\mathbf{n}_r)^H[\partial(\tilde{\mathbf{z}}_l - [\mathbf{H}_R + j\mathbf{H}_I]\mathbf{n}_r)/\partial H_R]\}\\ &= \frac{1}{\sigma^2}E\{-\mathbf{n}_r{}^H[\tilde{\mathbf{z}}_l - [\mathbf{H}_R + j\mathbf{H}_I]\mathbf{n}_r)] + (\tilde{\mathbf{z}}_l^H - \mathbf{n}_r^H\\ &\cdot [\mathbf{H}_R - j\mathbf{H}_I])(-\mathbf{n}_r)\} \end{aligned} \tag{83}$$

Similarly,

$$\begin{aligned}&\partial\mathcal{L}/\partial\mathbf{H}_I =\\ &\frac{1}{\sigma^2}E\{(\partial(\tilde{\mathbf{z}}_l^H - \mathbf{n}_r^H[\mathbf{H}_R - j\mathbf{H}_I])/(\partial H_I)(\tilde{\mathbf{z}}_l - [\mathbf{H}_R + j\mathbf{H}_I]\mathbf{n}_r)\\ &+ (\tilde{\mathbf{z}}_l^H - \mathbf{n}_r^H[\mathbf{H}_R - j\mathbf{H}_I])\partial(\tilde{\mathbf{z}}_l - [\mathbf{H}_R + j\mathbf{H}_I]\mathbf{n}_r)/\partial H_I\}\\ &= \frac{1}{\sigma^2}E\{[(-j\mathbf{n}_r^H)(\tilde{\mathbf{z}}_l - [\mathbf{H}_R + j\mathbf{H}_I]\mathbf{n}_r)] + [(\tilde{\mathbf{z}}_l^H - \mathbf{n}_r^H\\ &\cdot [\mathbf{H}_R - j\mathbf{H}_I](-\mathbf{n}_r))(j)]\} \end{aligned}$$

Finally, we have

$$\begin{aligned}&\partial\mathcal{L}/\partial\mathbf{H} = \partial\mathcal{L}/\partial\mathbf{H}_R - j\partial\mathcal{L}/\partial\mathbf{H}_I\\ &= -\frac{2}{\sigma^2}E\{[\tilde{\mathbf{z}}_l^H - \mathbf{n}_r^H\mathbf{H}^H]\mathbf{n}_r\}\\ &\partial\mathcal{L}/\partial\mathbf{H}^* = \partial\mathcal{L}/\partial\mathbf{H}_R + j\partial\mathcal{L}/\partial\mathbf{H}_I\\ &= -\frac{2}{\sigma^2}E\{\mathbf{n}_r{}^H[\tilde{\mathbf{z}}_l - \mathbf{Hn}_r]\} \end{aligned} \tag{84}$$

$$\mathbf{J}_{\mathbf{H},\mathbf{H}} = E\{\frac{\partial\{\mathcal{L}\}}{\partial\bar{\mathbf{H}}^*}(\frac{\partial\{\mathcal{L}\}}{\partial\bar{\mathbf{H}}})^T, \tag{85}$$

$$\begin{aligned}&\mathbf{J}_{\mathbf{H},\mathbf{H}} = (\frac{2}{\sigma^2})^2[E\{(\mu^H\tilde{\mathbf{z}}_l - Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l)\mathbf{H}][E\{(\mu^*\tilde{\mathbf{z}}_l^* -\\ &Cov^*(\mathbf{n}_r|\tilde{\mathbf{z}}_l)\mathbf{H}^*] = (\frac{2}{\sigma^2})^2[E\{\|(\mu^H\tilde{\mathbf{z}}_l - Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l)\mathbf{H})\|^2\} \end{aligned} \tag{86}$$

Then (86) becomes

$$\begin{aligned}&\mathbf{J}_{\mathbf{H},\mathbf{H}} = \frac{4}{\sigma^4}[E\{\|\mu^H\tilde{\mathbf{z}}_l\|^2\} + E\{\|Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l)\mathbf{H}]\|^2\} - E\{\mu^H\\ &\tilde{\mathbf{z}}_l Cov^*(\mathbf{n}_r|\tilde{\mathbf{z}}_l)\mathbf{H}^*\} - E\{Cov(\mathbf{n}_r|\tilde{\mathbf{z}}_l)\mathbf{H}\mu^*\tilde{\mathbf{z}}_l^*\}, \end{aligned} \tag{87}$$

from which $\mathbf{J}_{\mathbf{h},\mathbf{h}}$ is easily obtained.