# A Conditional ℓ1 Regularized MMSE Channel Estimation Technique for IBI Channels

Takano, Yasuhiro
Su, Hsuan-Jung
Juntti, Markku
Matsumoto, Tad

(Citation)
IEEE Transactions on Wireless Communications,17(10):6720-6734

(Issue Date)
2018-10

(Resource Type)
journal article

(Version)
Accepted Manuscript



(URL)
https://hdl.handle.net/20.500.14094/90005442

# A Conditional $\ell 1$ Regularized MMSE Channel Estimation Technique for IBI Channels

Yasuhiro Takano, *Member, IEEE,* Hsuan-Jung Su, *Senior Member, IEEE,* Markku Juntti, *Senior Member, IEEE,* and Tad Matsumoto, *Fellow, IEEE*

***Abstract*—Inter-block-interference (IBI) caused as a result of pursuing the spectral efficiency can deteriorate channel estimation performance. For this problem, previously-proposed chained turbo estimation (CHATES) performs IBI cancelation by using the soft replica of the transmitted signal. The IBI cancelation technique can, however, suffer from a mean squared error (MSE) floor problem, since the soft replica is unavailable at the first turbo iteration. The IBI problem can be avoided by using channel impulse response (CIR) length constraint. Nevertheless, as shown in this paper, the IBI avoidance approach is difficult to be performed independently since it requires unbiased second-order statistics. This paper proposes, therefore, a new conditional $\ell 1$ regularized minimum mean square error (MMSE) channel estimation algorithm by jointly utilizing the IBI avoidance/cancelation and subspace techniques. Simulation results verify that the proposed algorithm solves the MSE floor problem, and, hence, improves the bit error rate (BER) convergence performance in realistic IBI channels including the effect of pulse shaping filters.**

***Index Terms*—Inter-block-interference (IBI), subspace-based channel estimation, $\ell 1$ norm regularization, compressive sensing, turbo equalization.**

## I. Introduction

TERMINALS in a mobile Internet-of-Things (IoT) network are expected to work with limited batteries for long duration while being required to achieve high spectral efficiency demanded for most wireless communication systems (e.g., [1]). Uplink multiple-input multiple-output (MIMO) transmission (TX) assuming single-carrier frequency division multiple access (SC-FDMA) is, hence, an attractive option to improve energy- and spectral-efficiencies [2], [3] when we consider massive machine type communication (MTC) systems (e.g., [4]) and/or the physical layer security (e.g., [5]) for future IoT systems. There has been, however, criticism that it centralizes the complexity required for the whole system into the receiver side, due to the inter-symbol-interference (ISI) problem.

Nevertheless, overheads for data transmission in a TX frame format, such as cyclic prefix (CP) and training sequence (TS) sections, provide a reasonable trade-off between the receiver complexity and the spectral efficiency. For example, by assuming CP-transmission, frequency domain turbo equalization algorithms (e.g., [6]) can solve the ISI problem with the same complexity order as the orthogonal frequency division multiple access (OFDMA) receivers. Moreover, an inter-block-interference (IBI)-free received TS of at least $WN_T$ symbols enables the receiver to accurately estimate a channel impulse response (CIR) of at most $W$ symbols for $N_T$ TX streams under the least squares (LS) criterion [7]. Hence, a TX format is, as depicted in Fig. 1, necessary to have either (a) a TS whose length is greater than $W(N_T + 1)$ symbols, or (b) a TS of $WN_T$ symbols and a guard interval (GI) section to avoid the IBI. The TX format (b) is preferable for battery life longevity of IoT terminals. However, it decreases the spectral efficiency. Hence, chained turbo equalization (CHATUE) [8], [9] and chained turbo estimation (CHATES) [10] algorithms were proposed to improve the spectral efficiency. They enable us to utilize a TX format without CP nor GI, as illustrated in Fig. 1(c).

The CHATES algorithm inherits the CHATUE concept and performs the IBI cancelation to improve the channel estimation accuracy in a turbo receiver. The subspace-based CHATES algorithm can asymptotically achieve the Cramér-Rao bound (CRB) after performing sufficient turbo iterations. However, the CHATES technique can suffer from a mean squared error (MSE) floor problem in the first iteration because the soft replica of the transmitted data is unavailable.

The MSE floor problem can be improved by applying an IBI avoidance strategy. For example, the *a priori* aided compressive sampling matching pursuit (PA-CoSaMP) technique [11] performs $\ell 1$ regularized channel estimation based on the IBI avoidance. However, a receiver has to estimate the *effective* CIR length $w$ depicted in Fig. 1 as well. We can determine the CIR length by using past channel estimates under the wide-sense stationary uncorrelated scattering (WSSUS) assumption (e.g., [12]). Nevertheless, as the data processing inequality implies, it is difficult to track an increasing CIR length correctly by using the past $\ell 1$ regularized estimates. The IBI avoidance approach requires, hence, ordinary $\ell 2$ channel estimates.

This paper proposes, given such a background, a new MIMO channel estimation algorithm by combining subspace-based IBI cancelation and $\ell 1$ regularized IBI avoidance algo-

Y. Takano is with Kobe University, 1-1 Rokkodai, Nada, Kobe, 657-8501, Japan (e-mail: takano@eedept.kobe-u.ac.jp).

H.-J. Su is with National Taiwan University, Taipei 10617, Taiwan (e-mail: hjs@ntu.edu.tw).

M. Juntti and T. Matsumoto are with the Centre for Wireless Communications, University of Oulu, FI-90014, Oulu, Finland (e-mail: markku.juntti@oulu.fi).

T. Matsumoto is with Japan Advanced Institute of Science and Technology (JAIST), 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan (e-mail: matumoto@jaist.ac.jp).

The work of Y. Takano was supported in part by JSPS Kakenhi (C) No. 17K06423 and in part by the Telecommunications Advancement Foundation. The work of H.-J. Su was supported in part by the Ministry of Science and Technology (MOST), Taiwan, under grants 106-2923-E-002-015-MY3 and 106-2221-E-002-030. The work of M. Juntti has been financially supported by Academy of Finland 6Genesis Flagship (grant 318927). The work of T. Matsumoto was supported in part by JSPS Kakenhi (B) No.15H04007.

rithms. Subsequent results below verify that the proposed algorithm formulated with a conditional $\ell 1$ regularized minimum mean square error (MMSE) criterion ameliorates the MSE floor problem in the first turbo iteration. A turbo receiver using the new channel estimation algorithm can, hence, improve the bit error rate (BER) convergence performance over that of the conventional receiver in *realistic* propagation scenarios.

Note that, in practice, the CIRs observed at a receiver include the leakage effect [13]–[15] of pulse shaping required to perform digital signal processing. For example, the Pedestrian-B (PB) [16] channel model is composed of 6 path components. However, as detailed in Section II, the receiver observes the filtered waveform and cannot detect the original 6 paths directly. The CIRs assumed in this paper do not exhibit exactly sparse nature in the observed-domain. We notice that, however, the CIRs follow *structured sparsity* [17] in a signal subspace of the eigen domain.

Contributions of this paper are summarized as follows:

- We assume CIRs including the impact of the pulse shaping filters. In the realistic CIRs, the new conditional $\ell 1$ MMSE estimator, referred to as the adaptive IBI-managed MMSE (AIM-MMSE), asymptotically achieves the CRB while the conventional $\ell 1$ LS estimators [11], [18] do not in general.
- Comparative study between the $\ell 2$-based IBI cancelation and $\ell 1$-based IBI avoidance techniques is provided. This paper shows that the $\ell 1$-based IBI avoidance by itself does not always outperform the $\ell 2$-based technique. However, we can improve the receiver performance by jointly utilizing them.
- BER performance using the proposed algorithm is also evaluated. The simulation results verify that the spectrally efficient TX format (c) in Fig. 1 improves the BER performance over that of the TX format (b) even when channel estimation is actually performed.
- The PA-CoSaMP-based algorithms have a problem in determining an *effective* CIR length to avoid the IBI. The proposed algorithm solves the problem by utilizing its analytical MSE performance and, unlike [11], it requires no empirical adjusting factors.

As a future work of this study, an OFDMA system jointly utilizing the proposed algorithm with [19] has a potential to improve the spectral efficiency further than that in [18], since we do not assume known pilot signals in the data section. Moreover, the proposed estimation framework can be extended into massive MIMO and/or mmWave MIMO systems, according to the analytical MSE performance, which is left as future study.

After Introduction, Section II shows our MIMO system. Section III summarizes basic strategies to combat IBI problems. Section IV details the estimation techniques and proposes the new IBI countermeasure. Section VI verifies the effectiveness of the new algorithm via computer simulations. Section VII shows concluding remarks.

**Notations**: The bold lower-case $\mathbf{x}$ and upper-case $\mathbf{X}$ denote a vector and a matrix, respectively. For matrix $\mathbf{X}$, its transpose and transposed conjugate are denoted as $\mathbf{X}^{\mathsf{T}}$ and $\mathbf{X}^{\mathsf{H}}$, respectively. $\mathbf{X}^{-1}$ and $\mathbf{X}^{\dagger}$ denote the matrix inverse

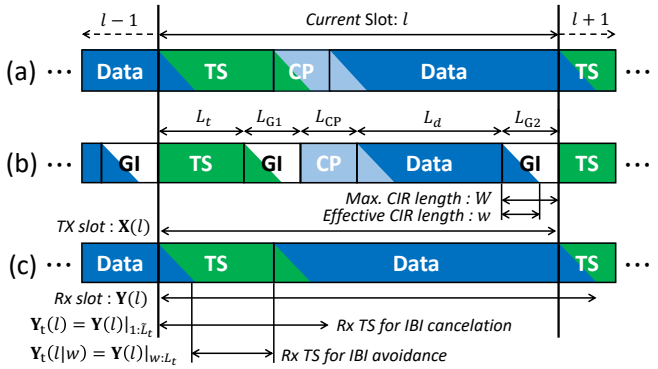



Fig. 1. TX formats having **(a)** a long TS with length $L_t \geq W(N_T+1)$, **(b)** a short TS $L_t \geq WN_T$ with GIs, and **(c)** a short TS without GI/CP sections, where $N_T$, $L_t$, and $W$ denote, respectively, the numbers of TX streams, TS length, and CIR length in symbol. The triangular parts illustrate the IBIs due to multipath channels.

TABLE I
OPERATORS

| Operator | Definition |
|---|---|
| $\mathbf{X}\|_{\mathcal{A}}$ | Submatrix composed of the column vectors in $\mathbf{X}$ corresponding to the set $\mathcal{A}$, where $\mathcal{CAL}$-font is used for index sets. A set of consecutive numbers $\{i, \cdots, j\}$ is denoted by $\mathcal{A} = \{i:j\}$, where $\{i:j\} = \emptyset$ if $i > j$. |
| $\mathbf{J}_{\mathcal{A}}$ | Factor matrix $\mathbf{J}_{\mathcal{A}} \stackrel{\text{def}}{=} \mathbf{I}\|_{\mathcal{A}}$ to denote a compressed/sparse matrix, where $\mathbf{I}$ is an identify matrix. e.g., $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5]$, $\mathcal{A} = \{1, 3\} \Rightarrow \mathbf{X}\mathbf{J}_{\mathcal{A}} \cdot \mathbf{J}_{\mathcal{A}}^{\mathsf{T}} = [\mathbf{x}_1, \mathbf{x}_3] \cdot \mathbf{J}_{\mathcal{A}}^{\mathsf{T}} = [\mathbf{x}_1, \mathbf{0}, \mathbf{x}_3, \mathbf{0}, \mathbf{0}]$. |
| $\mathtt{vec}(\mathbf{X})$ | $MN \times 1$ vector composed by stacking the columns of $\mathbf{X} \in \mathbb{C}^{M \times N}$. |
| $\mathtt{mat}_N(\mathbf{x})$ | Inversion of the vectorization: $\mathtt{mat}_N\{\mathtt{vec}(\mathbf{X})\} = \mathbf{X}$. |
| $\otimes$ | Kronecker product: $\mathfrak{A} = \mathbf{B} \otimes \mathbf{C}$, where $\mathfrak{frak}$-font is used for Kronecker product matrices. |
| $\mathtt{diag}(\mathbf{X})$ | Vector composed of the diagonal elements of $\mathbf{X}$. |
| $\mathtt{D}_{\mathtt{IAG}}(\mathbf{x})$ | Diagonal matrix formed with the vector $\mathbf{x}$. |
| $\mathtt{tplz}_M\{\mathbf{r}\}$ | $M \times N$ Toeplitz matrix whose first row is length $N$ vector $\mathbf{r}$. |
| $\|\mathbf{X}\|^2_{\mathbf{A}\times\mathbf{B}}$ | Weighted matrix Frobenius norm: $\mathtt{tr}\{\mathbf{BXAX}^{\mathsf{H}}\}$ for $\mathbf{X} \in \mathbb{C}^{M \times N}$ with positive definite matrices $\mathbf{A}$ and $\mathbf{B}$. Moreover, $\|\mathbf{X}\|^2_{\mathbf{A}} = \|\mathbf{X}\|^2_{\mathbf{A}\times\mathbf{I_M}}$ and $\|\mathbf{X}\|^2 = \|\mathbf{X}\|^2_{\mathbf{I_N}\times\mathbf{I_M}}$, where $\mathbf{I_M}$ is the $M \times M$ identity matrix. |
| $\|\mathbf{X}\|_1$ | Matrix $\ell 1$ norm: $\sum_{i=1}^{M}\sum_{j=1}^{N}\|x_{ij}\|$ where $x_{ij}$ is the $(i,j)$-th entry of $\mathbf{X}$. |
| $\|\mathbf{X}\|_0$ | Matrix $\ell 0$ norm: the number of non-zero entries of $\mathbf{X}$. |
| $\mathbb{E}^{L}_{j=l}[\mathbf{X}(j)]$ | Average of $\mathbf{X}(j) \in \mathbb{C}^{M \times N}$: $\frac{1}{L}\sum_{j=l-L+1}^{l}\mathbf{X}(j)$ for the past $L$ samples from the timing $l$.<br>Moreover, $\mathbb{E}[\mathbf{X}(j)] = \mathbb{E}^{\infty}_{j=l}[\mathbf{X}(j)]$. |
| $\mathbb{K}[\mathbf{X}(j)]$ | Column-wise covariance matrix: $\mathbb{E}[\mathbf{X}^{\mathsf{H}}(j)\mathbf{X}(j)]$. |
| $\mathbb{R}[\mathbf{X}(j)]$ | Row-wise covariance matrix: $\mathbb{E}[\mathbf{X}(j)\mathbf{X}^{\mathsf{H}}(j)]$. |
| $\mathbb{P}(\mathbf{X})$ | Projection matrix $\mathbf{X}\mathbf{X}^{\dagger}$ of $\mathbf{X}$. |
| $\mathbb{I}(\mathbf{X})$ | The indicator function of $\mathbf{X}$. |
| $\mathtt{AAD}(\mathcal{D}, \sigma_z^2)$ | Adaptive active-set detection [20] that recursively determines significant CIR taps from domain $\mathcal{D}$, according to analytical MSE performance given noise variance $\sigma_z^2$. Formal definition is shown in Section IV-A2. |

and the Moore-Penrose pseudo-inverse of $\mathbf{X}$, respectively. The Cholesky decomposition of $\mathbf{X}$ is denoted by $\mathbf{X}^{\mathrm{H}/2}\mathbf{X}^{1/2}$. $\lceil\cdot\rceil$ denotes the ceiling function. Moreover, this paper uses operators summarized in Table I.

## II. System Model

### A. Transmitter

As depicted in Fig. 2, an $L_{\mathrm{info}}$-bit binary data information sequence $b(i)$ is channel-encoded into a coded frame $c(i_c)$ by a rate $R_c$ convolutional code (CC) with generator polynomials $(g_1, \cdots, g_{\lceil 1/R_c \rceil})$ and is interleaved by an interleaver ($\Pi$). The interleaved coded frame $c\left(\Pi(i_c)\right)$ is serial-to-parallel (S/P)-converted into $N_T$ data segments for MIMO transmission using $N_T$ TX antennas. A data segment is further divided into $N_{\mathrm{S}}$ slots such that the WSSUS assumption holds for a slot duration. A data block is modulated into binary phase shift keyed (BPSK) symbols $x_{d,k}(j)$ with variance $\sigma_x^2$ and the modulation multiplicity $M_b = 1$.

The $k$-th TX antenna transmits data symbols $\mathbf{x}_{d,k}(l) = [x_{d,k}(1), \cdots, x_{d,k}(L_d)]^{\mathsf{T}}$ together with a length-$L_t$ symbol TS $\mathbf{x}_{t,k}(l)$ using single carrier signaling, where $l$ denotes the slot timing index and the data symbol length $L_d$ in a slot is defined as $L_d = L_{\mathrm{info}}/(R_c N_T N_{\mathrm{S}} M_b)$. As depicted in Fig. 1(b), the TX format may generally have a length-$L_{\mathrm{CP}}$ symbol CP and two GIs whose lengths are $L_{\mathrm{G1}}$ and $L_{\mathrm{G2}}$ symbols, respectively. We use the general TX format as a benchmark to show channel estimation performance in the system without IBI problems. However, TX formats without CP and/or GI sections such as that shown in Figs. 1(a) and (c) are also used to verify the effectiveness of new IBI countermeasure proposed in this paper.

### B. Received signal

*1) Signal model:* The received signal $\mathbf{y}_n(l)$ at the $n$-th receive (Rx) antenna suffers from ISI due to fading frequency selectivity, and from complex additive white Gaussian noise (AWGN) as well. The CIR length is assumed at most $W$ symbols. The received signals with $N_R$ Rx antennas can be described in a matrix form as,

$$\mathbf{Y}(l) = \mathbf{H}(l)\mathbf{X}(l) + \mathbf{Z} + \mathbf{Z}_{\mathrm{IBI}} \tag{1}$$

with $\mathbf{Z}_{\mathrm{IBI}} = \mathbf{H}(l-1)\mathbf{X}_{\triangle}(l-1) + \mathbf{H}(l+1)\mathbf{X}_{\triangledown}(l+1)$, where the first two terms correspond to the transmitted block of the *current* slot timing $l$ whereas the last term represents the IBI from the *past* $(l-1)$-th and *future* $(l+1)$-th slots, respectively. Specifically,

$$\begin{array}{rcl} \mathbf{X}(l) & = & [\mathbf{X}_1^{\mathsf{T}}(l), \cdots, \mathbf{X}_{N_T}^{\mathsf{T}}(l)]^{\mathsf{T}} \quad \in \quad \mathbb{C}^{WN_T \times \tilde{L}_{\mathrm{S}}}, \\ \mathbf{H}(l) & = & [\mathbf{H}_1(l), \cdots, \mathbf{H}_{N_T}(l)] \quad \in \quad \mathbb{C}^{N_R \times WN_T}, \end{array} \tag{2}$$

where we denote $\tilde{L}_{\mathrm{S}} = L_{\mathrm{S}} + W - 1$ with the slot length $L_{\mathrm{S}} = L_t + L_{\mathrm{G1}} + L_{\mathrm{CP}} + L_d + L_{\mathrm{G2}}$. We first describe the transmitted data $\mathbf{X}(l)$ and noise $\mathbf{Z}$ matrices. The channel matrix $\mathbf{H}(l)$ is, then, shown in the next subsection.

The data matrix $\mathbf{X}_k(l)$ of the $k$-th TX stream is given by

$$\mathtt{tplz}_W\left\{[\mathbf{x}_{t,k}^{\mathsf{T}}(l), \mathbf{0}_{L_{\mathrm{G1}}}^{\mathsf{T}}, \mathbf{x}_{\mathrm{CP},k}^{\mathsf{T}}(l), \mathbf{x}_{d,k}^{\mathsf{T}}(l), \mathbf{0}_{L_{\mathrm{G2}}+W-1}^{\mathsf{T}}]\right\}$$

with $\mathbf{x}_{\mathrm{CP},k} = \mathbf{x}_{d,k}(l)|_{(L_d - L_{\mathrm{CP}}+1):L_d}$, where the operation $\mathtt{tplz}_W\{\mathbf{r}\}$ constructs a $W \times N_R$ Toeplitz matrix whose first row vector is $\mathbf{r} \in \mathbb{C}^{1\times N_R}$.

The noise row-vector $\mathbf{z}_n$ in $\mathbf{Z}$ follows the Complex normal distribution $\mathcal{CN}(\mathbf{0}, \sigma_z^2 \mathbf{I}_{L_{\mathrm{S}}+W-1})$ and has the uncorrelated property: $\mathbb{E}[\mathbf{z}_i^{\mathsf{H}}\mathbf{z}_j] = 0$. The IBI matrix $\mathbf{Z}_{\mathrm{IBI}}$ is defined by using

$$\mathbf{X}_{\triangle}(l-1) = \left[\mathbf{X}(l-1)|_{(L_{\mathrm{S}}-W+2):L_{\mathrm{S}}},\ \mathbf{O}_{WN_T \times L_{\mathrm{S}}}\right], \tag{3}$$

$$\mathbf{X}_{\triangledown}(l+1) = \left[\mathbf{O}_{WN_T \times L_{\mathrm{S}}},\ \mathbf{X}(l+1)|_{1:(W-1)}\right], \tag{4}$$

where (3) and (4) become the zero matrices when $L_{\mathrm{G2}} \geq W - 1$.

*2) CIR parameters:* Under the WSSUS assumption, channel parameters are fixed in a slot duration $L_{\mathrm{S}}T_s$, where $T_s$ is the symbol interval. Hence, the $(n,j)$-th entry $h_l^{(k,n)}(j)$ of the CIR matrix $\mathbf{H}_k(l)$ can be written as $h_l^{(k,n)}(j) = h^{(k,n)}[(j + lL_{\mathrm{S}})T_s]$. The channel parameter $h^{(k,n)}[t]$, $t \in \mathbb{R}$, is composed of $r_k$ resolvable paths, each having complex gain $g_r^{(k,n)}[t]$ and path delay $\tau_r^{(k,n)}$, where the path gain and delay parameters follow channel profiles defined in a standard specification (e.g., [16]).

The received signal is observed via pulse shaping filters. The receiver may, hence, describe the parameter $h^{(k,n)}[t]$ as that including the effect of the pulse shaping filters [13]–[15]:

$$h^{(k,n)}[t] = p[t] \star \left\{\textstyle\sum_{r=1}^{r_k} g_r^{(k,n)}[t]\, \delta[t - \tau_r^{(k,n)}]\right\}, \tag{5}$$

where $\star$ denotes the convolution operator and $\delta[t]$ is the Dirac delta function. The coefficient $p[t]$ represents a pulse shaping filter logically combining those in the transmitter and the receiver. In most systems, the coefficient $p[t]$ represents an impulse response of the raised cosine filter[1] given by $p[t] = \mathrm{sinc}(t/T_s)\frac{\cos(\pi\alpha t/T_s)}{1-(2\alpha t/T_s)^2}$, where $\alpha$ is the roll-off factor (e.g., [21]).

The expected variance of the CIR matrix $\mathbf{H}_k(l)$ for the $k$-th TX stream is $\mathbb{E}[\|\mathbf{H}_k(l)\|^2] = \sigma_{\mathbf{H}}^2$ with a constant $\sigma_{\mathbf{H}}^2$. The CIR matrix satisfies a property that the spatial covariance matrix $\mathbb{R}[\mathbf{H}_k(l)] \stackrel{\mathrm{def}}{=} \mathbb{E}[\mathbf{H}_k(l)\mathbf{H}_k^{\mathsf{H}}(l)]$ is of full-rank by assuming no unknown interferences [22].

### C. Receiver

As depicted in Fig. 2, the receiver performs channel estimation (EST) jointly over the Rx antennas while also obtaining the extrinsic LLR $\lambda_{\mathrm{EQU},k}^e$ for the $k$-th TX stream by means of the CHATUE [9] (EQU) technique. The $N_T$ log-likelihood ratio (LLR) $\lambda_{\mathrm{EQU},k}^e$ sequences are parallel-to-serial (P/S)-converted to form an extrinsic LLR sequence $\lambda_{\mathrm{EQU}}^e$ corresponding to the interleaved coded frame $c(\Pi(i_c))$ at the transmitter. An *a priori* LLR $\lambda_{\mathrm{DEC}}^a$ for the channel decoder ($\mathrm{CC}^{-1}$) is obtained by deinterleaving $\lambda_{\mathrm{EQU}}^e$. The channel decoder performs decoding for $\lambda_{\mathrm{DEC}}^a$ by using the Bahl, Cocke, Jelinek and Raviv (BCJR) algorithm [23], and outputs the *a posteriori* LLR $\lambda_{\mathrm{DEC}}^p$. After several iterations, $\mathrm{CC}^{-1}$ outputs the estimates of the transmitted sequence $\hat{\mathbf{b}}$ by making a hard decision on $\lambda_{\mathrm{DEC}}^p$.

[1]The root raised cosine filter is often used as the pulse shaping filter in both the transmitter and the receiver.

EQU utilizes the soft replica[2] of the transmitted symbols $\hat{\mathbf{x}}_{d,k}$ which is generated from the equalizer's *a priori* LLR $\lambda^a_{\rm EQU}$. Note that LLR $\lambda^a_{\rm EQU}$ is the interleaved version of the extrinsic LLR $\lambda^e_{\rm DEC}$ which is obtained as $\lambda^e_{\rm DEC} = \lambda^p_{\rm DEC} - \lambda^a_{\rm DEC}$ according to the turbo principle. The soft replica for EST may, however, be generated from the *a posteriori* LLR $\lambda^p_{\rm DEC}$ to improve the convergence performance in a moderate to high signal-to-noise ratio (SNR) regime.

## III. Preliminaries

### A. Signal model

Assuming a TX format shown in Fig. 1(b), the receiver performs channel estimation by using both the TS $\mathbf{X}_t(l)$ and soft replica $\hat{\mathbf{X}}_d(l)$ matrices, where we denote

$$\mathbf{X}_t(l) = [\mathbf{X}^{\mathsf{T}}_{t,1}(l), \cdots, \mathbf{X}^{\mathsf{T}}_{t,N_T}(l)]^{\mathsf{T}} \in \mathbb{C}^{WN_T \times \tilde{L}_t}, \quad (6)$$

$$\hat{\mathbf{X}}_d(l) = [\hat{\mathbf{X}}^{\mathsf{T}}_{d,1}(l), \cdots, \hat{\mathbf{X}}^{\mathsf{T}}_{d,N_T}(l)]^{\mathsf{T}} \in \mathbb{C}^{WN_T \times \tilde{L}_d}. \quad (7)$$

with $\tilde{L}_t = L_t + W - 1$ and $\tilde{L}_d = L_d - W + 1$. Moreover,

$$\mathbf{X}_{t,k}(l) = \mathtt{tplz}_W\left\{[\mathbf{x}^{\mathsf{T}}_{t,k}(l), \mathbf{0}^{\mathsf{T}}_{W-1}]\right\},$$
$$\hat{\mathbf{X}}_{d,k}(l) = \mathtt{tplz}_W\left\{[\hat{\mathbf{x}}^{\mathsf{T}}_{d,k}(l)|_{W:L_d}]\right\}.$$

Correspondingly, we define the received data matrix as $\mathbf{Y}_d(l) = \mathbf{Y}(l)|_{\mathfrak{d}+W:\mathfrak{d}+L_d}$ with the offset $\mathfrak{d} = L_t + L_{\rm G1} + L_{\rm CP}$ to the data section. The received data matrix $\mathbf{Y}_d(l)$ can always avoid IBI, since $L_d \gg W(N_T+1)$. However, $L_t \geq W(N_T+1)$ does not always hold for a short TS length $L_t$. Hence, we need to fully utilize the received TS of length $\tilde{L}_t$ symbols. However, as depicted in Fig. 1(c), the received TS matrix $\mathbf{Y}_t(l) = \mathbf{Y}(l)|_{1:\tilde{L}_t}$ can suffer from the IBI seriously when $N_{{\rm G}i} < W-1$ for $i = 1, 2$. Concretely, $\mathbf{Y}_t(l)$ is written as

$$\mathbf{Y}_t(l) = \mathbf{H}(l)\mathbf{X}_t(l) + \mathbf{Z}_t + \mathbf{Z}_{t,\rm IBI}, \quad (8)$$

where the matrix $\mathbf{Z}_{t,\rm IBI}$ describes the IBI from the *past* and *future* blocks: $\mathbf{Z}_{t,\rm IBI} = \mathbf{H}(l-1)\mathbf{X}_p(l-1) + \mathbf{H}(l)\mathbf{X}_f(l)$. The $WN_T \times \tilde{L}_t$ matrices $\mathbf{X}_p(l-1)$ and $\mathbf{X}_f(l)$ are defined similarly to (3) and (4), by using the data sequences transmitted in the $(l-1)$-th and $l$-th slot timings, respectively. The $N_R \times \tilde{L}_t$ noise matrix $\mathbf{Z}_t$ is $\mathbf{Z}|_{1:\tilde{L}_t}$.

### B. IBI avoidance and compressive estimation

As illustrated in Fig. 1(c), we can avoid the IBI problem easily by choosing the range of the received TS as

$$\mathbf{Y}_{t,[w]}(l) \stackrel{\rm def}{=} \mathbf{Y}_t(l)|_{w:L_t}, \quad (9)$$

given *effective* CIR length $w$. For $\forall w$, an $N_R \times WN_T$ CIR estimate $\hat{\mathbf{H}}(l)$ is obtained by

$$\hat{\mathbf{H}}(l) = \mathbf{Y}_{t,[w]}(l)\,(\mathbf{X}_t(l)|_{w:L_t})^{\dagger}. \quad (10)$$

However, the LS estimate (10) is inaccurate for a long $w$ such that $\mathtt{rank}\{\mathbf{X}_t(l)|_{w:L_t}\} < WN_T = \mathtt{col}\{\mathbf{H}(l)\}$, which means that the LS estimation problem is ill-conditioned. The operator $\mathtt{col}\{\mathbf{M}\}$ denotes the column size of matrix $\mathbf{M}$.

[2]In the case of BPSK, as shown in [24], the $j$-th entry in $\hat{\mathbf{x}}_{d,k}$ is generated as $\hat{x}_{d,k}(j) = \sigma_x \tanh(\lambda^a_{{\rm EQU},k}(j)/2)$, where $\lambda^a_{{\rm EQU},k}(j)$ denotes the $j$-th S/P-converted equalizer's *a priori* LLR for the $k$-th TX stream.

The accuracy problem can be improved by using the compressive estimation approach, if the CIR $\mathbf{H}(l)$ is a sparse matrix supported with a column index set $\mathcal{A}$. We refer to $\mathcal{A}$ as *active-set*, hereafter. By using a matrix $\mathbf{J}_{\mathcal{A}} = \mathbf{I}_{WN_T}|_{\mathcal{A}}$, we denote the sparse CIR $\mathbf{H}_{\mathcal{A}}(l)$ by $\mathbf{H}_{\mathcal{A}}(l) = \mathbf{G}_{\mathcal{A}}(l)\mathbf{J}^{\mathsf{T}}_{\mathcal{A}}$, where $\mathbf{G}_{\mathcal{A}}(l) \stackrel{\rm def}{=} \mathbf{H}(l)\mathbf{J}_{\mathcal{A}}$ is a column-*shrunk* $N_R \times |\mathcal{A}|$ CIR matrix. The estimate matrix can be obtained, as $\hat{\mathbf{H}}_{\mathcal{A}} = \hat{\mathbf{G}}_{\mathcal{A}}\mathbf{J}^{\mathsf{T}}_{\mathcal{A}}$, where $\hat{\mathbf{G}}_{\mathcal{A}}(l) = \mathbf{Y}_{t,[w]}(l)\,\mathbf{\Phi}^{\dagger}_{t,\mathcal{A},[w]}$ with

$$\mathbf{\Phi}_{t,\mathcal{A},[w]}(l) \stackrel{\rm def}{=} \mathbf{J}^{\mathsf{T}}_{\mathcal{A}}\mathbf{X}_t(l)|_{w:L_t}. \quad (11)$$

The MSE performance of $\hat{\mathbf{G}}_{\mathcal{A}}(l)$ is proportional to the CRB if $\mathbf{\Phi}_{t,\mathcal{A},[w]}(l)$ is a fat matrix:

$$\mathtt{rank}\{\mathbf{\Phi}_{t,\mathcal{A},[w]}(l)\} = |\mathcal{A}| = \mathtt{col}\{\mathbf{G}_{\mathcal{A}}\}. \quad (12)$$

Nevertheless, (12) does not always hold for a significantly long $w$ and a dense $\mathcal{A}$. Note that, moreover, the CIR length $w$ and the active-set $\mathcal{A}$ are the parameters to be estimated.

### C. IBI cancelation

The IBI cancelation strategy can be taken for any CIR length $w$ as long as replica signals of the IBI components are available. The CHATES algorithm [10] cancels the IBI by

$$\tilde{\mathbf{Y}}^{[i]}_t(l) = \mathbf{Y}_t(l) - \hat{\mathbf{Z}}^{[i-1]}_{t,\rm IBI}. \quad (13)$$

The cancelation term $\hat{\mathbf{Z}}^{[i-1]}_{t,\rm IBI}$ is written, as

$$\hat{\mathbf{Z}}^{[i-1]}_{t,\rm IBI} = \hat{\mathbf{H}}^{[I_{\rm TB}]}(l-1)\hat{\mathbf{X}}^{[I_{\rm TB}]}_p(l-1) + \hat{\mathbf{H}}^{[i-1]}(l)\hat{\mathbf{X}}^{[i-1]}_f(l)$$

for the current slot timing $l$ at the $i$-th turbo iteration, when a coded frame is transmitted in $N_{\rm S} = 1$ slot duration. The estimate of the first IBI component $\hat{\mathbf{H}}^{[I_{\rm TB}]}(l-1)\hat{\mathbf{X}}^{[I_{\rm TB}]}_p(l-1)$ can be obtained in the last $I_{\rm TB}$-th turbo iteration at the previous slot timing. The second IBI component $\hat{\mathbf{H}}^{[i-1]}(l)\hat{\mathbf{X}}^{[i-1]}_f(l)$ can be estimated as a result of the previous $[i-1]$-th turbo iteration at the current slot timing. Nevertheless, $\hat{\mathbf{X}}^{[i-1]}_f(l)$ is unavailable in the first iteration.

### D. Subspace projection

As widely recognized (e.g., [25]), the covariance matrix of MIMO channel parameters $\mathbf{H}_k(l) \in \mathbb{C}^{N_R \times W}$ for the $k$-th TX stream can be written by using the singular value decomposition (SVD), as

$$\mathbb{K}[\mathbf{H}_k(l)] \stackrel{\rm def}{=} \mathbb{E}\left[\mathbf{H}^{\mathsf{H}}_k(l)\,\mathbf{H}_k(l)\right] = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{U}^{\mathsf{H}}_k, \quad (14)$$

with $W \times W$ unitary matrix $\mathbf{U}_k$, where the rank of the singular matrix $\mathbf{\Lambda}_k \in \mathbb{C}^{W \times W}$ is the number $r_k$ of multipaths.

The subspace channel model assumption [22] can, therefore, be utilized to develop a channel estimation technique. Specifically, a receiver observes the channel matrix,

$$\mathbf{H}_k(l) = \mathbf{B}_k(l)\mathbf{V}^{\mathsf{H}}_k, \quad (15)$$

such that the SVD for the covariance matrix of (15) is consistent with (14). The $N_R \times r_k$ matrix $\mathbf{B}_k(l)$ is slot-dependent. However, the $W \times r_k$ matrix $\mathbf{V}_k \stackrel{\rm def}{=} \mathbf{U}_k|_{1:r_k}$ is independent of the slot timing, since it can be seen as a

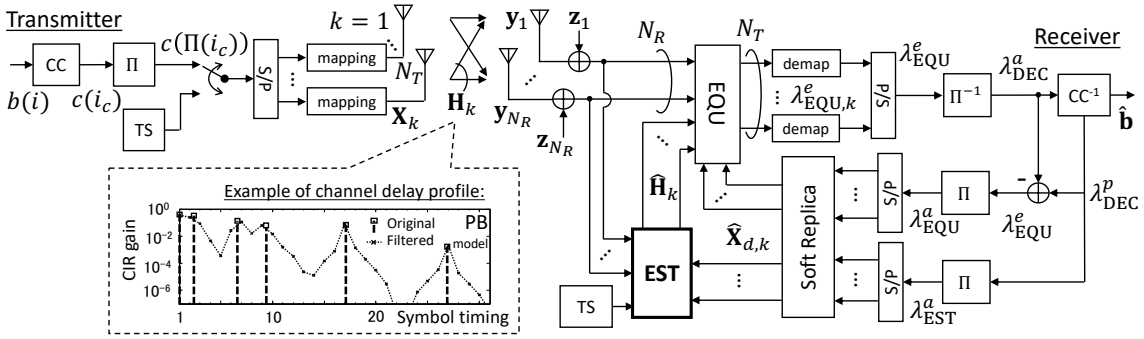



Fig. 2. The system model. The channel delay profile depicted in the left bottom is based on the PB model in a TX bandwidth of 7 MHz with a carrier frequency 2 GHz. The raised cosine filter with roll-off $\alpha = 0.3$ is used for pulse shaping.

time-invariant FIR filter representing the response of pulse shaping filters and multipath channels. As detailed later, we can suppress estimation errors by multiplying a channel estimate $\hat{\mathbf{H}}_k(l)$ with a projection matrix $\mathbb{P}(\mathbf{V}_k) \stackrel{\text{def}}{=} \mathbf{V}_k\mathbf{V}_k^\dagger$, since $\mathtt{rank}\{\mathbb{P}(\mathbf{V}_k)\} = r_k \ll W$ and $\mathbf{H}_k(l)\mathbb{P}(\mathbf{V}_k) = \mathbf{H}_k(l)$.

Let us rewrite the channel model (15) as to describe for $N_T$ TX streams, since the next section discusses channel estimation techniques jointly performed for all the TX streams. Specifically, (15) is rewritten as

$$\mathbf{H}(l) = \mathbf{B}(l)\mathbf{V}^{\mathsf{H}} \tag{16}$$

with the $N_R \times \bar{r}N_T$ gain matrix $\mathbf{B}(l) = [\mathbf{B}_1(l), \cdots, \mathbf{B}_{N_T}(l)]$. The $WN_T \times \bar{r}N_T$ response matrix $\mathbf{V}$ is given by $\bigoplus_{k=1}^{N_T}\mathbf{V}_k$. The operator $\bigoplus$ indicates the matrix direct sum.

## IV. Channel Estimation

### A. $\ell 1$ LS with IBI avoidance ($\ell 1$ LS-IA) channel estimation

*1) Problem formulation:* We apply the IBI avoidance approach shown in Section III-B to the turbo receiver. Given effective CIR length $w$ and the domain $\mathcal{D}$ of CIR taps, the column-shrunk $N_R \times |\mathcal{A}|$ LS channel estimate $\hat{\mathbf{G}}_{\mathcal{A}}^{\text{LS}}$ can be obtained by

$$\hat{\mathbf{G}}_{\mathcal{A}}^{\text{LS}}(l) = \underset{\mathbf{G}_{\mathcal{A}},\ \mathcal{A}\subseteq\mathcal{D}}{\mathtt{arg\ min}}\ \mathcal{L}_{\ell 1}^{td}\left(l, \mathbf{G}_{\mathcal{A}}\mathbf{J}_{\mathcal{A}}^{\mathsf{T}} \mid w, \mathcal{D}\right) + \lambda\|\mathbf{G}_{\mathcal{A}}\|_1 \tag{17}$$

with a Lagrange multiplier $\lambda$ [26]. We perform the IBI avoidance by letting the domain

$$\mathcal{D} \subseteq \mathcal{D}_w \stackrel{\text{def}}{=} \textstyle\bigcup_{k'=0}^{N_T-1}\{(1+k'W) : (w+k'W)\}.$$

The joint log-likelihood function $\mathcal{L}_{\ell 1}^{td}(l, \mathbf{G}_{\mathcal{A}}\mathbf{J}_{\mathcal{A}}^{\mathsf{T}} \mid w, \mathcal{D})$ is defined by $\mathcal{L}_t(l, \mathbf{G}_{\mathcal{A}}\mathbf{J}_{\mathcal{A}}^{\mathsf{T}} \mid w) + \mathcal{L}_d(l, \mathbf{G}_{\mathcal{A}}\mathbf{J}_{\mathcal{A}}^{\mathsf{T}})$ under $\mathcal{A} \subset \mathcal{D}$, where

$$\begin{aligned}\mathcal{L}_t(l, \mathbf{G}_{\mathcal{A}}\mathbf{J}_{\mathcal{A}}^{\mathsf{T}} \,|\, w) &= \tfrac{1}{\sigma_z^2}\|\mathbf{Y}_{t,[w]}(l) - \mathbf{G}_{\mathcal{A}}\mathbf{\Phi}_{t,\mathcal{A},[w]}(l)\|^2_{\mathbf{K}_{\mathbf{Z}_t}^{-1}},\\ \mathcal{L}_d(l, \mathbf{G}_{\mathcal{A}}\mathbf{J}_{\mathcal{A}}^{\mathsf{T}}) &= \tfrac{1}{\sigma_z^2}\|\mathbf{Y}_d(l) - \mathbf{G}_{\mathcal{A}}\hat{\mathbf{\Phi}}_{d,\mathcal{A}}(l)\|^2_{\mathbf{I}_{\tilde{L}_d}\times\mathbf{\Gamma}}.\end{aligned}$$

The weighted matrix Frobenius norm is denoted by $\|\mathbf{M}\|^2_{\mathbf{A}\times\mathbf{B}} = \mathtt{tr}\{\mathbf{BMAM}^{\mathsf{H}}\}$ for conformable matrices $\mathbf{M}$, $\mathbf{A}$, and $\mathbf{B}$. Moreover, $\|\mathbf{M}\|^2_{\mathbf{A}} = \|\mathbf{M}\|^2_{\mathbf{A}\times\mathbf{I}}$.

The transmitted TS matrix $\mathbf{\Phi}_{t,\mathcal{A},[w]}(l)$ is given by (11). Accordingly, the data matrix is defined by $\hat{\mathbf{\Phi}}_{d,\mathcal{A}}(l) = \mathbf{J}_{\mathcal{A}}^{\mathsf{T}}\hat{\mathbf{X}}_d(l)$. Since the IBI problem is avoided, the covariance matrices[3] is given by $\mathbf{K}_{\mathbf{Z}_t} = \mathbf{I}_{\tilde{L}_t}$ and

$$\mathbf{\Gamma} = \left[\mathbf{I}_{N_R} + (\Delta\hat{\sigma}_d^2/\sigma_z^2)\mathbf{R}_{\mathbf{H}}\right]^{-1}, \tag{18}$$

where we denote $\mathbf{R}_{\mathbf{H}} = \mathbb{R}[\mathbf{H}(l)]$ and

$$\Delta\hat{\sigma}_d^2 = \sigma_x^2 - \textstyle\sum_{k=1}^{N_T}\|\hat{\mathbf{x}}_{d,k}(l)\|^2/(L_dN_T). \tag{19}$$

*2) Solution to* (17)*:* For conciseness, this subsection omits the parameter $l$ relevantly. Problem (17) can be solved by the expectation-maximization (EM) algorithm [27] which alternately performs the following two subproblems:

$$\begin{cases}\hat{\mathbf{G}}_{\mathcal{A}^{[n]}} &= \underset{\mathbf{G}_{\mathcal{A}^{[n]}}}{\mathtt{arg\ min}}\ \mathcal{L}_{\ell 1}^{td}\left(\mathbf{G}_{\mathcal{A}^{[n]}}\mathbf{J}_{\mathcal{A}^{[n]}}^{\mathsf{T}} \mid w, \mathcal{A}^{[n]}\right) & (20)\\ \mathcal{A}^{[n+1]} &= \underset{\mathcal{A}\subseteq\mathcal{A}^{[n]}}{\mathtt{arg\ min}}\ \mathcal{L}_{\ell 1}^{td}\left(\mathbf{G}_{\mathcal{A}}\mathbf{J}_{\mathcal{A}}^{\mathsf{T}} \mid w, \mathcal{A}^{[n]}, \hat{\mathbf{d}}_{\mathbf{H}}^{[n]}\right). & (21)\end{cases}$$

The active-set $\mathcal{A}^{[n+1]}$ updated for the next iteration is recursively reduced from $\mathcal{A}^{[n]}$, where the initial active-set $\mathcal{A}^{[0]}$ is set at the given domain $\mathcal{D}$. The delay profile vector $\hat{\mathbf{d}}_{\mathbf{H}}^{[n]}$ is estimated as $\mathbf{J}_{\mathcal{A}^{[n]}}\mathtt{diag}\{\hat{\mathbf{G}}_{\mathcal{A}^{[n]}}^{\mathsf{H}} \cdot \hat{\mathbf{G}}_{\mathcal{A}^{[n]}}\}$, where the solution to (20) can be obtained via

$$\mathtt{vec}\{\hat{\mathbf{G}}_{\mathcal{A}}^{\text{LS}}\} = \mathfrak{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}^{\dagger} \cdot \mathtt{vec}\{\mathbf{R}_{\mathbf{Y\Phi}_{\mathcal{A}}}\} \tag{22}$$

with $\mathfrak{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}} = \mathfrak{J}_{\mathcal{A}}\mathfrak{R}_{\mathbf{\Phi\Phi}}\mathfrak{J}_{\mathcal{A}}^{\mathsf{T}}$ and $\mathbf{R}_{\mathbf{Y\Phi}_{\mathcal{A}}} = \mathbf{R}_{\mathbf{Y\Phi}}\mathbf{J}_{\mathcal{A}}$. We denote $\mathfrak{J}_{\mathcal{A}} = \mathbf{J}_{\mathcal{A}} \otimes \mathbf{I}_{N_R}$ and

$$\begin{aligned}\mathfrak{R}_{\mathbf{\Phi\Phi}} &= \mathbf{R}^{\mathsf{T}}_{\mathbf{\Phi\Phi}_{t,[w]}} \otimes \mathbf{I}_{N_R} + \hat{\mathbf{R}}^{\mathsf{T}}_{\mathbf{XX}_d} \otimes \hat{\mathbf{\Gamma}}, & (23)\\ \mathbf{R}_{\mathbf{Y\Phi}} &= \mathbf{R}_{\mathbf{Y\Phi}_{t,[w]}} + \hat{\mathbf{\Gamma}}\mathbf{R}_{\mathbf{YX}_d}, & (24)\end{aligned}$$

where

$$\begin{aligned}\mathbf{R}_{\mathbf{\Phi\Phi}_{t,[w]}} &= (\mathbf{X}_t|_{w:L_t})(\mathbf{X}_t|_{w:L_t})^{\mathsf{H}}, \quad \hat{\mathbf{R}}_{\mathbf{XX}_d} = \hat{\mathbf{X}}_d\hat{\mathbf{X}}_d^{\mathsf{H}},\\ \mathbf{R}_{\mathbf{Y\Phi}_{t,[w]}} &= (\mathbf{Y}_t|_{w:L_t})(\mathbf{X}_t|_{w:L_t})^{\mathsf{H}}, \quad \mathbf{R}_{\mathbf{YX}_d} = \mathbf{Y}_d\hat{\mathbf{X}}_d^{\mathsf{H}}.\end{aligned}$$

The matrix $\hat{\mathbf{\Gamma}}$ can be obtained by using $\mathbf{R}_{\mathbf{H}} \approx \hat{\mathbf{H}}^{[i-1]}(\hat{\mathbf{H}}^{[i-1]})^{\mathsf{H}}$ in (18), where $\hat{\mathbf{H}}^{[i-1]}$ is the channel estimate obtained by the previous $[i-1]$-th[4] turbo iteration.

[3]Technically, the covariance matrix $\mathbf{K}_{\mathbf{Z}_t}$ is defined, as $\mathbf{K}_{\mathbf{Z}_t} = \mathbb{K}[\Delta\mathcal{Y}_t]/(N_R\sigma_z^2)$, where $\Delta\mathcal{Y}_t$ denotes the residual matrix in the likelihood function $\mathcal{L}_t(\cdot)$.

[4]For the first turbo iteration, $i = 1$, $\mathbf{R}_{\mathbf{H}}$ is discarded in (18) since the channel estimation is performed with the TS only.

Sub-problem (21) can be solved by the adaptive active-set detection (AAD) algorithm [20], However, the LS solution (22) using the pseudo-inverse matrix is not always computed accurately for a long effective CIR length $w$. Hence, we revise the AAD algorithm by imposing a cardinality regularization for IBI channels. Specifically,

$$\begin{aligned}\mathcal{A}^{[n+1]} &= \mathtt{AAD}\left(\mathcal{A}^{[n]}, \hat{\mathbf{d}}_{\mathbf{H}}^{[n]}, |\Delta\hat{\mathbf{d}}_{\mathbf{H}}^{[n]}|, f_{\mathbf{\Phi}}(\sigma_z^2, \mathcal{A}^{[n]})\right) \\ &\stackrel{\text{def}}{=} \left\{ j \;\middle|\; \begin{array}{l} \hat{d}_{\mathbf{H},j}^{[n]} > \frac{|\Delta\hat{\mathbf{d}}_{\mathbf{H}}^{[n]}|}{\|\hat{\mathbf{d}}_{\mathbf{H}}^{[n]}\|_0} + \frac{f_{\mathbf{\Phi}}(\sigma_z^2, \mathcal{A}^{[n]})}{|\mathcal{A}^{[n]}|} \\ j \in \mathcal{A}^{[n]} \cap \mathcal{T}(\hat{\mathbf{d}}_{\mathbf{H}}^{[n]}, E_w) \end{array} \right\} \end{aligned} \quad (25)$$

where the parameter $\hat{d}_{\mathbf{H},j}^{[n]}$ is the $j$-th entry of the vector $\hat{\mathbf{d}}_{\mathbf{H}}^{[n]}$ and the superscript $[n]$ represents the $n$-th AAD iteration. The operation $\mathcal{T}(\mathbf{x}, E)$ forms an index subset of the *top* $E$ entries in vector $\mathbf{x}$. The AAD algorithm iteratively performs (22) and (25) at most $N_{\text{AAD}}$ times until the Bayesian information criterion (BIC) [28] of (22) converges.[5] We define the noise threshold function by an approximated MSE of $\hat{\mathbf{G}}_{\mathcal{A}^{[n]}}$, as

$$f_{\mathbf{\Phi}}(\sigma_z^2, \mathcal{A}^{[n]}) = \sigma_z^2 \mathtt{tr}\{\mathfrak{R}^{\dagger}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}^{[n]}}}\}, \quad (26)$$

under an assumption that CIRs unsupported with $\mathcal{A}^{[n]}$ are minor. The absolute error is given by $|\Delta\hat{\mathbf{d}}_{\mathbf{H}}^{[n]}| \approx f_{\mathbf{\Phi}}(\sigma_z^2, \mathcal{A}^{[n]})$ since the delay profile $\hat{\mathbf{d}}_{\mathbf{H}}^{[n]}$ is obtained from the LS estimate (22).

The operation $\mathcal{T}(\hat{\mathbf{d}}_{\mathbf{H}}^{[n]}, E_w)$ is used to guarantee that the matrix pseudo-inversions in (22) and (26) can be replaced with the ordinary matrix inversion after the initialization at $n = 0$, where the parameter $E_w$ is given by $E_w = \mathtt{min}\{|\mathcal{A}^{[0]}|, \alpha\bar{L}_{td}^{\ell 1}(w)\}$ with the reference signal length[6] $\bar{L}_{td}^{\ell 1}(w) \stackrel{\text{def}}{=} \mathtt{tr}\{\mathfrak{R}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{D}_w}}\}/(wN_TN_R)$. We define the factor $\alpha$ so that the covariance matrix $\mathfrak{R}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}}$ is composed of fat matrices. Since it is sufficient if the TS submatrix $\mathbf{\Phi}_{t,\mathcal{D}_w,[w]}$ is fat, we design the factor such that $\alpha < \alpha_0$, where

$$\alpha_0 \stackrel{\text{def}}{=} \left[\mathtt{max}\{w \mid wN_T < L_t - w + 1, \forall w \le W\}\right]/W. \quad (27)$$

### *B. $\ell 1$ MMSE with IBI avoidance ($\ell 1$ MMSE-IA) channel estimation*

*1) Problem formulation:* A conditional $\ell 1$ MMSE-IA channel estimation is formulated as the MMSE problem of (20). Under the subspace channel assumption (16), the estimate can be written in the form of $\hat{\mathbf{H}}_{\mathcal{A}}^{\text{IA}}(l) = \hat{\mathbf{B}}(l)\hat{\mathbf{V}}_{\mathcal{A}}^{\mathsf{H}} \cdot \mathbf{J}_{\mathcal{A}}^{\mathsf{T}}$, where

$$\begin{aligned}&(\hat{\mathbf{B}}(l), \hat{\mathbf{V}}_{\mathcal{A}}) = \\ &\quad \underset{\mathbf{B}(l), \mathbf{V}_{\mathcal{A}}, \mathcal{A}\subseteq\mathcal{D}}{\mathtt{arg\,min}} \; \mathbb{E}_{j=l}^{L_M}\left[\mathcal{L}_{td}\left(j, \mathbf{B}(j)\mathbf{V}_{\mathcal{A}}^{\mathsf{H}}\mathbf{J}_{\mathcal{A}}^{\mathsf{T}} \mid w, \mathcal{D}\right)\right]. \end{aligned} \quad (28)$$

*2) Solution to* (28)*:* The solution of the gain matrix $\hat{\mathbf{B}}(l)$ is obtained from a conditional LS problem given the subspace vectors $\mathbf{V}_{\mathcal{A}}$, as

$$\begin{aligned}\hat{\mathbf{B}}(l) &= \underset{\mathbf{B}(l)}{\mathtt{arg\,min}}\, \mathcal{L}_{td}\left(l, \mathbf{B}(l)\mathbf{V}_{\mathcal{A}}^{\mathsf{H}}\mathbf{J}_{\mathcal{A}}^{\mathsf{T}} \mid w, \mathcal{A}, \mathbf{V}_{\mathcal{A}}\right) \\ &= \mathtt{mat}_{N_R}\{\hat{\tilde{\mathbf{g}}}_{\mathcal{A}}^{\text{LS}}(l)\}(\tilde{\mathbf{V}}_{\mathcal{A}}^{\mathsf{H}})^{\dagger}, \end{aligned} \quad (29)$$

with $\tilde{\mathbf{V}}_{\mathcal{A}} = \bigoplus_{k=1}^{N_T} \mathfrak{Q}_{\mathcal{A},kk}\mathbf{V}_{\mathcal{A}_k}$, where the $|\mathcal{A}_k|N_R \times |\mathcal{A}_k|N_R$ matrix $\mathfrak{Q}_{\mathcal{A},kk}$ is the $k$-th block diagonal matrix of $\mathfrak{Q}_{\mathcal{A}}$ calculated by the Cholesky decomposition of $\mathbb{E}[\mathfrak{R}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}}(l)]$, and active-subset $\mathcal{A}_k$ corresponds to the $k$-th TX stream in $\mathcal{A}$. The operation $\mathtt{mat}_N(\mathbf{x})$ forms an $N \times M$ matrix from the argument vector $\mathbf{x} \in \mathbb{C}^{NM\times 1}$, i.e., $\mathbf{x} = \mathtt{vec}\{\mathtt{mat}_N\{\mathbf{x}\}\}$. The length $N_R|\mathcal{A}|$ vector $\hat{\tilde{\mathbf{g}}}_{\mathcal{A}}^{\text{LS}}(l)$ is obtained by the noise whitening transformation,[7] $\mathfrak{w}(\cdot)$:

$$\begin{aligned}\hat{\tilde{\mathbf{g}}}_{\mathcal{A}}^{\text{LS}}(l) &= \mathfrak{w}(\mathfrak{Q}_{\mathcal{A}}, \hat{\mathbf{g}}_{\mathcal{A}}^{\text{LS}}(l)) \\ &\stackrel{\text{def}}{=} \mathfrak{Q}_{\mathcal{A}}\hat{\mathbf{g}}_{\mathcal{A}}^{\text{LS}}(l) - \left\{\mathfrak{Q}_{\mathcal{A}} - \bigoplus_{k=1}^{N_T}\mathfrak{Q}_{\mathcal{A},kk}\right\}\mathbf{g}_{\mathcal{A}}(l) \end{aligned} \quad (30)$$

for the $\ell 1$ LS channel estimate vector $\hat{\mathbf{g}}_{\mathcal{A}}^{\text{LS}}(l)$ obtained by the $\ell 1$ LS-IA.

The projection onto the subspace $\mathbf{V}_{\mathcal{A}}$ can be obtained by performing the principal component analysis (PCA) for the covariance matrix $\mathbb{K}_{j=l}^{L_M}[\mathtt{mat}_{N_R}\{\hat{\tilde{\mathbf{g}}}_{\mathcal{A}}^{\text{LS}}(j)\}]$, where $\mathcal{A}$ has to be consistent over the past $L_M$ slots to compute the PCA accurately. However, the active-set $\mathcal{A}$ can be changed in the middle of communication, which makes it difficult to straightforwardly apply the EM algorithm composed of (20) and (21) in order to obtain the subspace $\mathbf{V}_{\mathcal{A}}$. Note that, however,

$$\mathtt{tr}\left\{\mathbf{V}_{\mathcal{A}}^{\mathsf{H}}\mathbf{J}_{\mathcal{A}}^{\mathsf{T}} \cdot \mathbf{V} - \mathbf{I}_{\bar{r}N_T}\right\}/(\bar{r}N_T) \approx 0 \quad (31)$$

holds as long as the active-set $\mathcal{A}$ supports the CIRs approximately. We can, therefore, approximate (29) by re-using a subspace projector of $\mathbf{V}$ obtained by the $\ell 2$ MMSE with IBI cancelation ($\ell 2$ MMSE-IC) algorithm in the previous frame timing, where the algorithm is shown later in Section IV-C. Specifically,

$$\hat{\mathbf{B}}(l) \approx \hat{\tilde{\mathbf{H}}}_{\text{LS}}(l)(\tilde{\mathbf{V}}^{\mathsf{H}})^{\dagger} \quad (32)$$

by using the basis vectors $\tilde{\mathbf{V}}$ instead of $\tilde{\mathbf{V}}_{\mathcal{A}}$, where $\hat{\tilde{\mathbf{H}}}_{\text{LS}}(l) = \mathtt{mat}_{N_R}\{\hat{\tilde{\mathbf{h}}}_{\text{LS}}(l)\}$ with

$$\hat{\tilde{\mathbf{h}}}_{\text{LS}}(l) = \mathfrak{w}(\bar{\mathfrak{Q}}, (\mathbf{J}_{\mathcal{A}} \otimes \mathbf{I}_{N_R})\hat{\mathbf{g}}_{\mathcal{A}}^{\text{LS}}(l)). \quad (33)$$

The $WN_TN_R \times WN_TN_R$ matrix[8] $\bar{\mathfrak{Q}}$ is the Cholesky decomposition of

$$\bar{\mathfrak{R}}_{\mathbf{XX}} \approx (\mathbb{R}[\mathbf{X}_t(j)] + \mathbb{R}[\mathbf{X}_d(j)]) \otimes \mathbf{I}_{N_R}. \quad (34)$$

By (16) and (32), we have

$$\hat{\tilde{\mathbf{H}}}_{\mathcal{A}}^{\text{IA}}(l) = \hat{\tilde{\mathbf{H}}}_{\text{LS}}(l)\,\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}}, \quad (35)$$

where the projection matrix is defined as $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}} = \bigoplus_{k=1}^{N_T}\mathbb{P}(\hat{\tilde{\mathbf{V}}}_k)$ with the noise-whitened subspace estimate vectors $\hat{\tilde{\mathbf{V}}}_k$. The $\ell 1$ MMSE-IA estimate can, therefore, be obtained after the noise coloring transformation, as

$$\hat{\mathbf{H}}_{\mathcal{A}}^{\text{IA}}(l) = \mathtt{mat}_{N_R}\left\{\left(\bigoplus_{k=1}^{N_T}\bar{\mathfrak{S}}_{kk}\right) \cdot \mathtt{vec}\left\{\hat{\tilde{\mathbf{H}}}_{\mathcal{A}}^{\text{IA}}(l)\right\}\right\}, \quad (36)$$

[5]The BIC of $\hat{\mathbf{G}}_{\mathcal{A}^{[0]}}^{\text{LS}}$ is discarded if the initial active-set $\mathcal{A}^{[0]} = \mathcal{D}$ does not meet the cardinality regularization.

[6]$\bar{L}_{td}^{\ell 1}(w) \approx (L_t - w + 1) + \gamma\hat{\sigma}_d^2(L_d - W + 1)$ with $\gamma = \sigma_z^2/(\sigma_z^2 + \Delta\hat{\sigma}_d^2 N_T\sigma_{\mathbf{H}}^2)$.

[7]For the sake of conciseness, (30) is described so that it performs the noise whitening for all the $N_T$ TX streams jointly. However, (30) requires the channel vector $\mathbf{g}_{\mathcal{A}}(l)$ to be estimated. Nevertheless, (30) is solvable by using the back substitution per TX stream [22], where we approximate the $k$-th stream of $\mathbf{g}_{\mathcal{A}}(l)$ with that of the vectorized solution (36).

[8]We can pre-compute $\bar{\mathfrak{Q}}$ since $\mathbb{R}[\mathbf{X}_d(j)]$ may be calculated off-line by assuming the interleaved coded frame $c(\Pi(i_c))$ at the transmitter is an independent and identically distributed (i.i.d.) random sequence.

where $\bar{\mathfrak{S}}_{kk} \in \mathbb{C}^{WN_R \times WN_R}$ is the $k$-th block diagonal matrix of $\bar{\mathfrak{Q}}^{-1}$.

*3) CIR length estimation and active-set detection:* The effective CIR length $w$ in the $\ell 1$ MMSE-IA algorithm is estimated by

$$\hat{w} \;=\; \underset{1\le w\le W}{\mathtt{arg\,min}}\, \mathrm{MSE}(\hat{\mathbf{H}}^{\mathrm{IA}}_{\bar{\mathcal{A}}_w} \mid w), \tag{37}$$

where $\mathrm{MSE}(\hat{\mathbf{H}}^{\mathrm{IA}}_{\bar{\mathcal{A}}_w} \mid w) \overset{\mathrm{def}}{=} \mathbb{E}[\|\hat{\mathbf{H}}^{\mathrm{IA}}_{\bar{\mathcal{A}}_w}(l) - \mathbf{H}(l)\|^2]$. The expected active-set $\bar{\mathcal{A}}_w$ used in (37) can be obtained by performing the AAD algorithm:

$$\bar{\mathcal{A}}_w = \mathtt{AAD}\left(\mathcal{D}_w, \bar{\mathbf{d}}^{\mathrm{IA}}_{\ell 1}, \mathrm{CRB}_{\bar{L}_{td}}(\sigma_z^2), \mathrm{CRB}_{\bar{L}^{\ell 1}_{td}(w)}(\sigma_z^2)\right), \tag{38}$$

where $\bar{L}_{td} \overset{\mathrm{def}}{=} \mathtt{tr}\{\bar{\mathfrak{R}}_{\mathbf{XX}}\}/(WN_TN_R) = L_t + L_d - W + 1$. When $\bar{\mathcal{A}}_w$ supports the CIR correctly, the MSE performance of the $\ell 1$ MMSE-IA follows

$$\mathrm{CRB}_{\hat{L}^{\ell 1}_{td}(w)}(\sigma_z^2) \;=\; \sigma_z^2 \bar{r}_w N_T N_R / \bar{L}^{\ell 1}_{td}(w), \tag{39}$$

where we denote $\bar{r}_w = \mathtt{min}(\bar{r}, w)$. The delay profile vector $\bar{\mathbf{d}}^{\mathrm{IA}}_{\ell 1}$ is computed from the past $\ell 1$ channel estimates:

$$\bar{\mathbf{d}}^{\mathrm{IA}}_{\ell 1} \;=\; \mathtt{diag}\{\mathbb{K}^{L_M}_{j=l}[\hat{\mathbf{H}}^{\mathrm{IA}}_{\ell 1}(j-1)]\}. \tag{40}$$

The active-set used in the LS estimate (33) can be obtained by using the EM algorithm composed of (20) and (21) given domain $\mathcal{D} = \bar{\mathcal{A}}_{\hat{w}}$ (38), where we modify (21), as

$$\mathcal{A}^{[n+1]} = \mathtt{AAD}\left(\mathcal{A}^{[n]}, \hat{\mathbf{d}}^{[n]}_{\ell 1}, |\Delta\hat{\mathbf{d}}^{[n]}_{\ell 1}|, \frac{\bar{r}_w N_T}{|\mathcal{A}^{[n]}|} f_{\mathbf{\Phi}}(\sigma_z^2, \mathcal{A}^{[n]})\right). \tag{41}$$

The delay profile estimate[9] $\hat{\mathbf{d}}^{[n]}_{\ell 1}$ is given by

$$\beta \mathbf{J}_{\mathcal{A}^{[n]}} \mathtt{diag}\{\hat{\mathbf{G}}^{\mathsf{H}}_{\mathcal{A}^{[n]}} \hat{\mathbf{G}}_{\mathcal{A}^{[n]}}\} + (1-\beta)\bar{\mathbf{d}}^{\mathrm{IA}}_{\ell 1}$$

with $\beta = 1/\mathtt{min}\{l, (L_M+1)\}$. The absolute error $|\Delta\hat{\mathbf{d}}^{[n]}_{\ell 1}|$ is defined accordingly.

**Remark:** However, the $\ell 1$ regularized estimates derived from (40) can be inaccurate. This is because $\bar{\mathbf{d}}^{\mathrm{IA}}_{\ell 1}$ has to contain the entries unsupported by $\bar{\mathcal{A}}_w \subseteq \mathcal{D}_w$, in oder to evaluate the MSE performance bounded by the following proposition. The proof is shown in Section V.

**Proposition 1** (MSE bound of the $\ell 1$ MMSE-IA)**.** *Let us denote* $\mathbf{H}^{\perp}_{\mathcal{A}} \overset{\mathrm{def}}{=} \mathbf{H} - \mathbf{H}_{\mathcal{A}}$.

$$\mathrm{MSE}(\hat{\mathbf{H}}^{\mathrm{IA}}_{\mathcal{A}} \mid w) = \frac{\bar{r}_w N_T}{|\mathcal{A}|} f_{\mathbf{\Phi}}(\sigma_z^2, \mathcal{A}) + \mathtt{tr}\{\mathbb{K}[\hat{\mathbf{H}}^{\perp}_{\mathcal{A}}]\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}}\} \tag{42}$$

$$\geq \mathrm{CRB}_{\bar{L}^{\ell 1}_{td}(w)}(\sigma_z^2) + \mathtt{tr}\{\mathbb{K}[\hat{\mathbf{H}}^{\perp}_{\bar{\mathcal{A}}_w}]\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}}\}, \tag{43}$$

*where* $\mathtt{tr}\{\mathbb{K}[\mathbf{H}^{\perp}_{\mathcal{A}}]\}$ *describes the biased error due to CIRs unsupported with* $\mathcal{A}$.

[9]For samples $s(j)$, $\bar{s}(l) \overset{\mathrm{def}}{=} \mathbb{E}^{L+1}_{j=l}[s(j)] = \{s(l) + L\mathbb{E}^{L+1}_{j=l-1}[s(j)]\}/(L+1) = \beta s(l) + (1-\beta)\bar{s}(l-1)$ with $\beta = 1/(L+1)$.

### C. $\ell 2$ MMSE-IC

*1) Problem formulation:* The $\ell 2$ MMSE-IC channel estimation is formulated as

$$\hat{\mathbf{H}}^{\mathrm{IC}}_{\ell 2}(l) = \underset{\mathbf{H}(l)}{\mathtt{arg\,min}}\, \mathbb{E}^{L_M}_{j=l}\left[\mathcal{L}^{td}_{\ell 2}(j, \mathbf{H}(j))\right] \tag{44}$$

with the joint log-likelihood function $\mathcal{L}^{td}_{\ell 2}(l, \mathbf{H}(j)) = \tilde{\mathcal{L}}^{[i]}_t(j, \mathbf{H}) + \mathcal{L}_d(j, \mathbf{H})$, where

$$\tilde{\mathcal{L}}^{[i]}_t(j, \mathbf{H}) \;=\; \tfrac{1}{\sigma_z^2}\|\tilde{\mathbf{Y}}^{[i]}_t(j) - \mathbf{H}\mathbf{X}_t(j)\|^2_{\mathbf{K}^{-1}_{\mathbf{Z}_t,\ell 2}}.$$

The negative log-likelihood function $\tilde{\mathcal{L}}^{[i]}_t(j, \mathbf{H})$ indicates that we perform the IBI cancelation using (13), in order to fully utilize the short but ideally uncorrelated TS signals. The covariance matrix $\mathbf{K}_{\mathbf{Z}_t,\ell 2}$ has, accordingly, to be re-defined, as

$$\begin{aligned}\mathbf{K}_{\mathbf{Z}_t,\ell 2} \;=\; \mathbf{I}_{\tilde{L}_t} + \tfrac{1}{\sigma_z^2 N_R}\mathtt{D}_{\mathtt{IAG}}\big[\, \sigma_x^2 \mathbb{I}(\mathbf{X}_p^{\mathsf{H}} + \mathbf{X}_f^{\mathsf{H}}) \\ \cdot \left\{\sigma_x^2 \cdot \Delta\bar{\mathbf{d}}_{\mathbf{H}} + \Delta\hat{\sigma}_d^2 \cdot \bar{\mathbf{d}}_{\mathbf{H}}\right\}\big],\end{aligned} \tag{45}$$

by taking account of residual errors involved in the IBI cancelation (13), where the operator $\mathtt{D}_{\mathtt{IAG}}(\mathbf{x})$ forms a diagonal matrix from its argument vector $\mathbf{x}$ and $\mathbb{I}(\cdot)$ denotes the indicator function. The expected delay profile $\bar{\mathbf{d}}_{\mathbf{H}}$ and its error $\Delta\bar{\mathbf{d}}_{\mathbf{H}}$ are, respectively, given by

$$\bar{\mathbf{d}}_{\mathbf{H}} = \mathtt{diag}\{\bar{\mathbf{K}}_{\mathbf{H}}\} \tag{46}$$

and $\Delta\bar{\mathbf{d}}_{\mathbf{H}} \approx \{\mathrm{CRB}_{\bar{L}_{td,\ell 2}}(\sigma_z^2)/(WN_T)\}\mathbf{1}_{WN_T}$, where

$$\bar{\mathbf{K}}_{\mathbf{H}} \;=\; \mathbb{K}^{L_M}_{j=l}[\hat{\mathbf{H}}^{\mathrm{IC}}_{\ell 2}(j-1)]. \tag{47}$$

*2) Solution:* The solution can be described similarly as that of the $\ell 1$ MMSE-IA by applying the following three modifications *a), b),* and *c).*

*a) Noise whitening:* We replace the noise whitening transformation (33) with

$$\hat{\tilde{\mathbf{h}}}_{\mathrm{LS}}(l) = \mathfrak{w}(\bar{\mathfrak{Q}}, \hat{\mathbf{h}}^{\mathrm{LS}}_{\ell 2}(l)), \tag{48}$$

where $\hat{\mathbf{h}}^{\mathrm{LS}}_{\ell 2}(l) = \mathfrak{R}^{\dagger}_{\mathbf{XX}}(l) \cdot \mathtt{vec}\{\mathbf{R}_{\mathbf{YX}}(l)\}$ by using

$$\mathfrak{R}_{\mathbf{XX}}(l) \;=\; \mathbf{R}^{\mathsf{T}}_{\mathbf{XX}_t}(l) \otimes \mathbf{I}_{N_R} + \hat{\mathbf{R}}^{\mathsf{T}}_{\mathbf{XX}_d}(l) \otimes \hat{\mathbf{\Gamma}}. \tag{49}$$

with $\mathbf{R}_{\mathbf{XX}_t}(l) = \mathbf{X}_t(l)\mathbf{K}^{-1}_{\mathbf{Z}_t,\ell 2}\mathbf{X}^{\mathsf{H}}_t(l)$. $\mathbf{R}_{\mathbf{YX}}(l)$ is defined similarly to (24). Note that, as the number of turbo iterations increases, $\mathbf{K}_{\mathbf{Z}_t,\ell 2} \to \mathbf{I}_{\tilde{L}_t}$ and $\hat{\mathbf{\Gamma}} \to \mathbf{I}_{N_R}$ are expected since $\Delta\hat{\sigma}_d^2 \to 0$. We hence ensure that $\mathbb{E}[\mathfrak{R}_{\mathbf{XX}}(l)] \to \bar{\mathfrak{R}}_{\mathbf{XX}}$ (34).

*b) Subspace projection:* The $\ell 2$ MMSE-IC derives the subspace vectors $\hat{\tilde{\mathbf{V}}}_k$ by performing the PCA for $\mathbb{K}^{L_M}_{j=l}\left[\hat{\tilde{\mathbf{H}}}^{\mathrm{LS}}_{\ell 2,k}(j)\right]$, for $k = N_T, ..., 1$, where the $N_R \times W$ matrix $\hat{\tilde{\mathbf{H}}}^{\mathrm{LS}}_{\ell 2,k}(j)$ is the $k$-th submatrix of $\mathtt{mat}_{N_R}\{\hat{\tilde{\mathbf{h}}}_{\mathrm{LS}}(j)\}$, corresponding to the $k$-th TX stream.

*c) Noise coloring:* The $\ell 2$ MMSE-IC finally obtains the solution to (44) by performing the subspace projection (35) and the noise coloring transformation (36).

### D. AIM-MMSE channel estimation

*1) Problem description:* The $\ell 2$ MMSE-IC is expected to achieve the CRB asymptotically after performing sufficient turbo iterations. However, the $\ell 2$ MMSE-IC can suffer from the MSE floor problem in the first several turbo iterations. The $\ell 1$ MMSE-IA can improve the MSE floor problem by the IBI avoidance strategy. Nevertheless, the IBI avoidance requires the unbiased delay profile vector to determine the optimal CIR length constraint.

*2) The AIM-MMSE algorithm:* We propose a new ad-hoc MMSE channel estimation algorithm, referred to as AIM-MMSE, to compensate for the drawbacks of the IBI avoidance and cancelation strategies at the same time. The proposed algorithm exploits the $\ell 1$ MMSE-IA in the first turbo iteration, and then, switches to the $\ell 2$ MMSE-IC after performing *relevant* turbo iterations, where we select the algorithm that takes the minimum between their approximated MSE performances:

$$\mathrm{MSE}(\hat{\mathbf{H}}^{\mathrm{IA}}_{\ell 1}) \approx \mathtt{min}_{1\le w\le W}\mathrm{MSE}(\hat{\mathbf{H}}^{\mathrm{IA}}_{\bar{\mathcal{A}}_w} \mid w), \quad (50)$$

$$\mathrm{MSE}(\hat{\mathbf{H}}^{\mathrm{IC}}_{\ell 2}) \approx \sigma_z^2 \frac{\bar{r}}{W} \mathtt{tr}\left\{\mathfrak{R}^{\dagger}_{\mathbf{XX}}\right\}. \quad (51)$$

Algorithm 1 summarizes the AIM-MMSE channel estimation, where the set of input parameters $(i, l)$ denotes the counters of turbo iterations and slots, respectively. It should be noted that Step 2 performs the CIR length estimation (37) by using the expected delay profile (46) computed from the unconstraint $\ell 2$ MMSE-IC estimates. Steps 5 to 11 describe whether the IBI avoidance or the IBI cancelation approach is selected adaptively according to the analytical MSE performances (50) and (51). However, the $\ell 1$ LS-IA technique is used in the first $\lceil W/N_R \rceil$ slots since performance of the PCA is not converged yet.

In all slot timings, we perform the $\ell 2$ MMSE-IC at the final $I_{\mathrm{TB}}$-th iteration in order to update the input matrices $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}}$ and $\bar{\mathbf{K}}_{\mathbf{H}}$ for the next slot timing. Note that in the very first slot timing $l = 1$, we may initialize the input matrices as $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}} = \mathbf{I}_{WN_T}$ and $\bar{\mathbf{K}}_{\mathbf{H}} = \mathtt{D}_{\mathtt{IAG}}\{\mathbf{d}_0\} \otimes \mathbf{I}_{N_T}$ with $\mathbf{d}_0 = \sigma^2_{\mathbf{H}} W\,[W, W-1, \cdots, 1]^{\mathsf{T}}/\{2(W+1)\}$.

### E. Computational complexity order

Table II(a) summarizes the computational complexity orders $\mathcal{O}(\cdot)$ required for the channel estimation algorithms discussed in this section, where the received signal length is at most $\tilde{L}_{td} = L_t + L_d$. It should be emphasized that, due to $3 \ll WN_TN_R$, the complexity order required for the proposed AIM-MMSE is equivalent to that of the $\ell 2$ MMSE-IC, although the number of operations is increased slightly to select the best IBI management strategy according to the IBI length/variance and the LLRs' accuracy.

*1) The $\ell 1$ LS-IA channel estimation:* Table II(b) shows the details of the complexity order needed for the $\ell 1$ LS-IA algorithm. The right-most column describes the number of executions in this algorithm, where we assume $1 \le N_{\mathrm{AAD}} \ll W$. The complexity order needed for (22) is, however, at most $\mathcal{O}(2W^3N_T^3N_R^3)$ in $N_{\mathrm{AAD}}$ iterations. This is because the pseudo-inverse matrix $\mathfrak{R}^{\dagger}_{\mathbf{\Phi\Phi}_{\mathcal{D}_w}}$ in (22) is computed only for the initialization and, after that, it is replaced with the ordinary matrix inverse $\mathfrak{R}^{-1}_{\mathbf{\Phi\Phi}_{\mathcal{A}[n]}}$ due to the maximum cardinality regularization. Moreover, by using the matrix inversion update algorithm,[10] the complexity order for $\mathfrak{R}^{-1}_{\mathbf{\Phi\Phi}_{\mathcal{A}[n]}}$ is $\mathcal{O}(W^3N_T^3N_R^3)$ in $N_{\mathrm{AAD}}$ iterations. Therefore, the $\ell 1$ LS-IA requires $\mathcal{O}(W^2N_T^2\tilde{L}_{td} + 2W^3N_T^3N_R^3)$.

*2) The $\ell 1$ MMSE-IA channel estimation:* Following the $\ell 1$ LS-IA, the $\ell 1$ MMSE-IA algorithm performs the noise whitening (33) and coloring (36) transformations, and the subspace projection (35), with the complexity orders, respectively, $\mathcal{O}(N_T \cdot W^2N_R^2)$, $\mathcal{O}(N_T \cdot W^2N_R^2)$, and $\mathcal{O}(N_T \cdot W^2N_R)$. Note that the $\ell 1$ MMSE-IA does not update the noise whitening factor matrix $\bar{\mathbf{\Omega}}$ and the projection matrix $\mathbf{\Pi}_{\hat{\mathbf{V}}}$, since they are, respectively, pre-computed off-line and obtained in the $\ell 2$ MMSE-IC algorithm at the previous frame timing. Therefore, the complexity order for the $\ell 1$ MMSE-IA is $\mathcal{O}(W^2N_T^2\tilde{L}_{td} + 2W^3N_T^3N_R^3)$ since it is dominated by that of the $\ell 1$ LS-IA.

*3) The $\ell 2$ MMSE-IC channel estimation:* The $\ell 2$ MMSE-IC performs, first of all, the IBI cancelation (13) using sparse matrices $\hat{\mathbf{X}}_p$ and $\hat{\mathbf{X}}_f$, the covariance matrix update (49) and the ordinary $\ell 2$ LS channel estimation (22) whose complexity orders are, respectively, $\mathcal{O}(W^2N_TN_R)$, $\mathcal{O}(W^2N_T^2\tilde{L}_{td})$, and $\mathcal{O}(W^3N_T^3N_R^3)$. As detailed in [20], [22], it performs, then, the noise whitening/coloring transformations, the PCA[11] for the $W \times W$ covariance matrices of the LS channel estimates, and the subspace projection for $N_T$ TX streams. Their complexity orders are, respectively, $\mathcal{O}(N_T \cdot W^2N_R^2)$, $\mathcal{O}(N_T \cdot W^3)$, and $\mathcal{O}(N_T \cdot W^2N_R)$. The complexity order for the $\ell 2$ MMSE-IC is, hence, $\mathcal{O}(W^2N_T^2\tilde{L}_{td} + 1W^3N_T^3N_R^3)$ since it is dominated by that of (49) and (22).

*4) The AIM-MMSE channel estimation:* The AIM-MMSE selects the best IBI management strategy under the minimum

**Algorithm 1** The AIM-MMSE channel estimation.

**Input:** $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}}$, $\bar{\mathbf{K}}_{\mathbf{H}}$, $\hat{\mathbf{X}}_d$, and $(i, l)$.

1: Update $\Delta\hat{\sigma}_d^2$ (19) and $\mathfrak{R}_{\mathbf{XX}}$ (49) by using $\bar{\mathbf{K}}_{\mathbf{H}}$ and $\hat{\mathbf{X}}_d$.
2: Estimate CIR length $\hat{w}$ by (37).
3: Compute $\mathrm{MSE}(\hat{\mathbf{H}}^{\mathrm{IA}}_{\ell 1})$ and $\mathrm{MSE}(\hat{\mathbf{H}}^{\mathrm{IC}}_{\ell 2})$ respectively by (50) and (51).
4: Select ALGO = the $\ell 2$ MMSE-IC (44) as defualt.
5: **if** $\mathrm{MSE}(\hat{\mathbf{H}}^{\mathrm{IA}}_{\ell 1}) < \mathrm{MSE}(\hat{\mathbf{H}}^{\mathrm{IC}}_{\ell 2})$ **and** $i < I_{\mathrm{TB}}$ **then**
6:  **if** $l \le \lceil W/N_R \rceil$ **then**
7:  Select ALGO = the $\ell 1$ LS-IA (17).
8:  **else**
9:  Select ALGO = the $\ell 1$ MMSE-IA (28).
10:  **end if**
11: **end if**
12: Perform the selected ALGO and obtain $\hat{\mathbf{H}}_{\mathrm{AIM}}$.
13: Update $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}}$ and $\bar{\mathbf{K}}_{\mathbf{H}}$ (47) if the $\ell 2$ MMSE-IC was performed.

**Output:** $\hat{\mathbf{H}}_{\mathrm{AIM}}$, $\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}}$, and $\bar{\mathbf{K}}_{\mathbf{H}}$.

[10] Given inverse matrix of an $M \times M$ Hermitian matrix, the complexity order needed to compute the inversion of its arbitrary rank-$N$-downsized submatrix is $\mathcal{O}(M^2N + N^3)$ [20], [29].

[11] The complexity order required for the PCA for an $M \times M$ covariance matrix is dominated by the SVD requiring $\mathcal{O}(M^3)$ [30], where the complexity to update $\mathbb{K}_{j=l}^{L_M}[\hat{\mathbf{H}}^{\mathrm{LS}}_{\ell 2,k}(j)]$ is negligible by using the recursion shown in footnote 9.

TABLE II

(a) COMPLEXITY ORDERS.

| Algorithm | Complexity order |
|---|---|
| $\ell 1$ LS-IA | $\mathcal{O}(W^2N_T^2\tilde{L}_{td} + 2W^3N_T^3N_R^3)$ |
| $\ell 1$ MMSE-IA | $\mathcal{O}(W^2N_T^2\tilde{L}_{td} + 2W^3N_T^3N_R^3)$ |
| $\ell 2$ MMSE-IC | $\mathcal{O}(W^2N_T^2\tilde{L}_{td} + 1W^3N_T^3N_R^3)$ |
| AIM-MMSE | $\mathcal{O}(W^2N_T^2\tilde{L}_{td} + 3W^3N_T^3N_R^3)$ |

(b) DETAILS FOR THAT OF THE $\ell 1$ LS-IA.

| Symbol | Eqn. | Complexity | Exec. counts |
|---|---|---|---|
| $\hat{w}$ | (37) | $\mathcal{O}(W^2N_T)$ | 1 |
| $\mathcal{A}^{[n+1]}$ | (25) | $\mathcal{O}(WN_T)$ | $N_{\mathrm{AAD}}+1$ |
| $\mathfrak{R}_{\mathbf{\Phi\Phi}}$ | (23) | $\mathcal{O}(W^2N_T^2\tilde{L}_{td})$ | 1 |
| $\hat{\mathbf{G}}_{\mathcal{A}}^{\mathrm{LS}}$ | (22) | $\mathcal{O}(W^3N_T^3N_R^3)$ | $N_{\mathrm{AAD}}+1$ |

MSE criterion, where the complexity order for Steps 1 to 11 in Algorithm 1 is dominated by that of the computations for $\mathfrak{R}_{\mathbf{XX}}$ and $\mathfrak{R}_{\mathbf{XX}}^{\dagger}$. We note that the covariance $\mathfrak{R}_{\mathbf{\Phi\Phi}}$ is a submatrix of $\mathfrak{R}_{\mathbf{XX}}$, however, $\mathfrak{R}_{\mathbf{\Phi\Phi}}^{\dagger}$ cannot be efficiently updated from $\mathfrak{R}_{\mathbf{XX}}^{\dagger}$. The complexity of the AIM-MMSE is hence at most $\mathcal{O}(W^2N_T^2\tilde{L}_{td} + 3W^3N_T^3N_R^3)$ when the $\ell 1$ MMSE-IA is performed at Step 12, because it is dominated by the process for obtaining $\mathfrak{R}_{\mathbf{XX}}$, $\mathfrak{R}_{\mathbf{XX}}^{\dagger}$, $\mathfrak{R}_{\mathbf{\Phi\Phi}_{\mathcal{D}_w}}^{\dagger}$, and $\mathfrak{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}^{-1}$.

## V. PERFORMANCE ANALYSIS

We, first of all, consider MSE performance of the LS channel estimation techniques. The MSE performance of the MMSE estimation algorithms is, then, investigated. We do not detail the MSE performance of the AIM-MMSE since it is given by the minimum of (50) and (51).

*1) The $\ell 1$ LS-IA :* We show the MSE performance of the $\ell 1$ LS-IA as a corollary obtained from Lemma 1. The slot timing index $l$ is omitted for the sake of conciseness.

**Lemma 1.** *For an active-set $\mathcal{A}$, the error vector of the conditional $\ell 1$ LS channel estimation given active-set $\mathcal{A}$, $\Delta\hat{\mathbf{h}}_{\mathcal{A}}^{\mathrm{LS}} \stackrel{\mathrm{def}}{=} \mathtt{vec}\{\hat{\mathbf{H}}_{\mathcal{A}}^{\mathrm{LS}} - \mathbf{H}\}$, can be written, as*

$$\Delta\hat{\mathbf{h}}_{\mathcal{A}}^{\mathrm{LS}} \;=\; \Delta_0^{\mathrm{LS}}(\mathcal{A}) + \Delta_{\mathbf{Z}}^{\mathrm{LS}}(\mathcal{A}) + \Delta_{\perp}^{\mathrm{LS}}(\mathcal{A}),$$

*where the three error component vectors are given by*

$$\begin{aligned}
\Delta_0^{\mathrm{LS}}(\mathcal{A}) &= \mathfrak{J}_{\mathcal{A}}\left(\mathfrak{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}^{\dagger}\mathfrak{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}} - \mathbf{I}_{|\mathcal{A}|N_R}\right)\mathbf{g}_{\mathcal{A}} \\
\Delta_{\mathbf{Z}}^{\mathrm{LS}}(\mathcal{A}) &= \mathfrak{J}_{\mathcal{A}}\mathfrak{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}^{\dagger}\mathtt{vec}\{\mathbf{R}_{\mathbf{ZX}}\mathbf{P}_{\mathcal{A}}\} \\
\Delta_{\perp}^{\mathrm{LS}}(\mathcal{A}) &= \left[\mathfrak{J}_{\mathcal{A}}\mathfrak{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}^{\dagger}\mathfrak{J}_{\mathcal{A}}^{\mathsf{T}}\mathfrak{R}_{\mathbf{XX}} - \mathbf{I}_{WN_TN_R}\right]\mathbf{h}_{\mathcal{A}}^{\perp}
\end{aligned}$$

*with* $\mathbf{R}_{\mathbf{ZX}} = \mathbf{Z}_t\mathbf{X}_t^{\mathsf{H}} + \hat{\mathbf{\Gamma}}(\mathbf{Z}_d - \mathbf{H}\Delta\hat{\mathbf{X}}_d)\mathbf{X}_d^{\mathsf{H}}$ *and* $\mathbf{h}_{\mathcal{A}}^{\perp} = \mathtt{vec}\{\mathbf{H}_{\mathcal{A}}^{\perp}\}$.

*Proof.* Because $\hat{\mathbf{h}}_{\mathcal{A}}^{\mathrm{LS}} = \mathfrak{J}_{\mathcal{A}}\mathtt{vec}\{\hat{\mathbf{G}}_{\mathcal{A}}\}$ with (22). □

**Corollary 1** (MSE bound of the $\ell 1$ LS-IA)**.**

$$\mathrm{MSE}(\hat{\mathbf{H}}_{\mathcal{A}}^{\mathrm{LS}} \mid w) \;\;\geq\;\; f_{\mathbf{\Phi}}(\sigma_z^2, \mathcal{A}) + \mathtt{tr}\{\mathbb{K}[\mathbf{H}_{\mathcal{A}}^{\perp}]\}.$$

*Proof.* In the $\ell 1$ LS-IA, $\Delta_0^{\mathrm{LS}}(\mathcal{A}) = \mathbf{0}$ holds since $\mathfrak{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}$ is invertible due to the cardinality regularization. By assuming ideally uncorrelated TSs, we have $\mathbb{E}[\|\Delta_{\mathbf{Z}}^{\mathrm{LS}}(\mathcal{A})\|^2] = f_{\mathbf{\Phi}}(\sigma_z^2, \mathcal{A})$ and $\mathbb{E}[\|\Delta_{\perp}^{\mathrm{LS}}(\mathcal{A})\|^2] \geq \mathtt{tr}\{\mathbb{K}[\mathbf{H}_{\mathcal{A}}^{\perp}]\}$. □

*2) The $\ell 1$ MMSE-IA:* Proof of Proposition 1 in Section IV-B is shown after describing a property of the subspace projection.

**Property 1.** *Let $\mathbf{Z} \in \mathbb{C}^{N\times M}$ be a random matrix satisfying $\mathbb{K}[\mathbf{Z}] \propto \mathbf{I}_M$. For a column sparse noise matrix $\mathbf{Z}_{\mathcal{A}} = \mathbf{Z}\cdot\mathbf{J}_{\mathcal{A}}\mathbf{J}_{\mathcal{A}}^{\mathsf{T}}$ with $\mathcal{A} \subseteq \{1 : M\}$ and an arbitrary subspace projection $\mathbf{\Pi} \in \mathbb{C}^{M\times M}$, we have*

$$\mathbb{E}\left[\|\mathbf{Z}_{\mathcal{A}}\mathbf{\Pi}\|^2\right] = \frac{\min\{\mathtt{rank}(\mathbf{\Pi}), |\mathcal{A}|\}}{|\mathcal{A}|}\mathbb{E}\left[\|\mathbf{Z}_{\mathcal{A}}\|^2\right].$$

*Proof.* If $|\mathcal{A}| > \mathtt{rank}(\mathbf{\Pi})$, $\mathbb{E}[\|\mathbf{Z}_{\mathcal{A}}\|^2] > \mathbb{E}[\|\mathbf{Z}_{\mathcal{A}}\mathbf{\Pi}\|^2] = \{\mathtt{rank}(\mathbf{\Pi})/|\mathcal{A}|\}\,\mathbb{E}\left[\|\mathbf{Z}_{\mathcal{A}}\|^2\right]$ holds. Otherwise, no subspace projection expands the domain. □

**Proof of Proposition 1** . By assuming the received signal length is long enough, we have $\bar{\mathfrak{R}}_{\mathbf{XX}}/\bar{L}_{td} \approx \mathbf{I}_{WN_TN_R}$, where $\bar{L}_{td} = \mathtt{tr}\{\bar{\mathfrak{R}}_{\mathbf{XX}}\}/(WN_TN_R)$. Since $\bar{\mathfrak{Q}}/\sqrt{\bar{L}_{td}} \approx \mathbf{I}_{WN_TN_R}$, we approximate subspace vectors as $\mathbf{V} \approx \tilde{\mathbf{V}}$. Thus, we have

$$\begin{aligned}
\Delta\hat{\mathbf{H}}_{\mathcal{A}}^{\mathrm{IA}} \stackrel{\mathrm{def}}{=} \hat{\mathbf{H}}_{\mathcal{A}}^{\mathrm{IA}} - \mathbf{H} \;&\approx\; \hat{\mathbf{H}}_{\mathcal{A}}^{\mathrm{LS}}\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}} - \mathbf{H} \\
&=\; \mathtt{mat}_{N_R}\{\Delta\hat{\mathbf{h}}_{\mathcal{A}}^{\mathrm{LS}}\}\mathbf{\Pi}_{\tilde{\mathbf{V}}} \;+\; \hat{\mathbf{H}}_{\mathcal{A}}^{\mathrm{LS}}\Delta\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}}, \qquad (52)
\end{aligned}$$

since $\mathbf{H} \approx \mathbf{H}\mathbf{\Pi}_{\tilde{\mathbf{V}}}$, where $\Delta\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}} = \hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}} - \mathbf{\Pi}_{\tilde{\mathbf{V}}}$ is the estimation error of the subspace projector. As long as the WSSUS assumption holds for $\forall L_M$, we have $\mathbb{E}[\|\Delta\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}}\|^2] \stackrel{L_M\to\infty}{=} 0$. Thereby,

$$\begin{aligned}
\mathrm{MSE}(\hat{\mathbf{H}}_{\mathcal{A}}^{\mathrm{IA}}) \;&=\; \mathbb{E}\left[\|\mathtt{mat}_{N_R}\{\Delta_{\mathbf{Z}}^{\mathrm{LS}}(\mathcal{A})\}\mathbf{\Pi}_{\tilde{\mathbf{V}}}\|^2\right] \\
&\quad+\; \mathbb{E}\left[\|\mathtt{mat}_{N_R}\{\Delta_{\perp}^{\mathrm{LS}}(\mathcal{A})\}\mathbf{\Pi}_{\tilde{\mathbf{V}}}\|^2\right]. \qquad (53)
\end{aligned}$$

The equation (42) in Proposition 1 is obtained by combining (53), Property 1, and Corollary 1. The inequality (43) is due to $\mathcal{A} \subseteq \bar{\mathcal{A}}_w$. □

## VI. NUMERICAL EXAMPLES

### A. Simulation Setups

*1) Channel models:* The path gains $g_r^{(k,n)}[t]$ in (5) are generated with the spatial channel model (SCM) [16], [31], where the roll-off factor $\alpha$ of the raised cosine filter $p[t]$ is set at 0.3. This paper assumes $4 \times 4$ MIMO channels, where the antenna element spacing at the base station and the mobile station are, respectively, set at 4 and 0.5 wavelength. Spatial parameters such as the direction of arrival (DoA) are randomly chosen. Moreover, six path fading channel realizations based on the Vehicular-A (VA) model [16] with a 30 km/h mobility (VA30) and the PB model [16] with a 3 km/h mobility (PB3) are assumed. The path positions $\tau_r^{(k,n)}/T_s$ of the VA and PB

TABLE III
TX FORMATS FOR A $4 \times 4$ MIMO SYSTEM.

| Format | $L_{\rm info}$ | $L_t$ | $L_{\rm G1}$ | $L_{\rm CP}$ | $L_d$ | $L_{\rm G2}$ | $R_c$ | $\eta$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 2048 | 127 | 0 | 0 | 512 | 0 | 1/2 | 1.6 |
| 2 | 2048 | 127 | 31 | 32 | 512 | 31 | 1/2 | 1.4 |
| 3 | 2048 | 127 | 0 | 32 | 480 | 0 | 8/15 | 1.6 |

models are respectively set at $\{1, 3.2, 6, 8.6, 13.1, 18.6\}$ and $\{1, 2.4, 6.6, 9.4, 17.1, 26.9\}$ symbol timings assuming that the TX bandwidth is 7 MHz with a carrier frequency of 2 GHz. However, the CIRs observed at the receiver can be distributed over more than 27 symbol duration due to the effect of the pulse shape filtering. The maximum CIR length $W$ is, hence, set at 31.

*2) TX formats:* Table III shows details of the TX format parameters, where TX formats 1, 2, and 3 correspond to Figs. 1(c), (b), and (a), respectively. First of all, the TX formats 1 and 2 are used to verify estimation accuracy of the proposed algorithms, where length $L_{\rm info} = 2048$ information bits after convolutional encoding with the generator polynomial $(g_1, g_2) = (7, 5)_8$ are transmitted over $N_{\rm S} = 2$ slots using the $4 \times 4$ MIMO system. The BER performance of the MIMO system using the proposed channel estimation is, then, investigated by assuming all the three TX formats.

### B. Channel Estimation Performance

*1) NMSE performance of the $\ell 2$ MMSE-IC:* Fig. 3(a) shows the normalized MSE (NMSE) of the channel estimates obtained using the $\ell 2$ MMSE-IC in the VA30 scenario, where TX format 1 without the GI is used. The normalized CRB (NCRB) is defined by $\mathrm{NCRB}_{\bar{L}}(\sigma_z^2) = \mathrm{CRB}_{\bar{L}}(\sigma_z^2)/\mathbb{E}[\|\mathbf{H}\|^2]$. As observed from Fig. 3(a), the NMSE performance using the conventional $\ell 2$ MMSE technique without the IBI cancelation diverges from the NCRB even after performing six iterations ($I_{\rm TB} = 6$), since the TX format 1 incurs the IBI problem.

The conventional CHATES [10] technique, referred to as $\ell 2$ CHATES, improves the IBI problem, however, it exhibits NMSE deterioration slightly at SNR = 30 dB even after six iterations. The $\ell 2$ MMSE-IC achieves the performance bound NCRB asymptotically in six iterations. This is because the $\ell 2$ MMSE-IC computes the covariance matrix $\mathbf{K}_{\mathbf{Z}_t,\ell 2}$ (45) correctly for the MIMO signal model, while the $\ell 2$ CHATES technique assumes $\mathbf{K}_{\mathbf{Z}_t,\ell 2} \approx \mathbf{I}_{\bar{L}_t}$.

*2) NMSE performance of the AIM-MMSE in a moderate CIR length scenario:* In the first iteration ($I_{\rm TB} = 1$), the $\ell 2$ MMSE-IC also suffers from MSE deterioration, since the replica signals of the transmitted sequence are not available. It is, hence, necessary to resort to the IBI avoidance strategy. As shown in Fig. 3(a), the NMSE performance of the AIM-MMSE achieves the analytical performance (42) and completely solves the MSE floor problem in the first turbo iteration, where $(N_{\rm AAD}, L_M, \alpha) = (1, 50, 0.7)$ is assumed.

Note that, we cannot attain the optimal performance (42) of the IBI avoidance by straightforwardly combining the $\ell 1$ LS-IA and the subspace method ($\ell 1$ LS-IA+SS). As shown in Fig. 3(a), the $\ell 1$ LS-IA+SS technique exhibits NMSE deterioration, where the PCA is performed for the covariance matrix $\hat{\mathbf{K}}_{\mathbf{H}_{\mathcal{A}}} = \mathbb{K}_{j=l}^{L_M}[\hat{\hat{\mathbf{H}}}_{\mathcal{A}}^{\rm LS}(j)]$ by assuming $\mathbb{E}[\mathfrak{R}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}}] \propto \mathbf{I}_{|\mathcal{A}|N_R}$. We can observe from Fig. 3(b) that the $\ell 1$ LS-IA+SS technique suffers from the subspace projection error $\mathbb{E}[\|\hat{\mathbf{H}}_{\mathcal{A}}^{\rm LS}\Delta\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}}\|^2] = \mathbb{E}[\|\mathbf{H}_{\mathcal{A}}\Delta\hat{\mathbf{\Pi}}_{\tilde{\mathbf{V}}}\|^2]$ in (52) significantly, although it obtains accurate CIR length $\hat{w}$ and rank $\bar{r}$ estimates in a high SNR regime. This is because the AAD (25) over-compresses the interested information under the BIC according to the LS criterion (17). Hence, the AAD (41) used in the AIM-MMSE algorithm adjusts the threshold not to distort the signal subspace under the MMSE criterion (28). At SNR = 0 dB, the $\ell 1$ LS-IA+SS suffers from another problem: $\mathrm{NMSE} = \sigma_{\mathbf{H}}^2$. This is because, the $\ell 1$ MMSE-IA by itself does not restore the covariance matrix $\hat{\mathbf{K}}_{\mathbf{H}_{\mathcal{A}}}$, if the estimated active-set has once converged into $\mathcal{A} \approx \emptyset$. However, the AIM-MMSE improves the problem by utilizing the past $\ell 2$ solutions.

*3) NMSE performance of the AIM-MMSE in a long CIR length scenario:* The IBI avoidance strategy does not always improve the MSE deterioration when the significant paths are distributed over a large number of CIR taps. As shown in Fig. 4(a), the $\ell 1$ LS-IA given $\mathcal{D} = \mathcal{D}_{\hat{w}}$ suffers from the MSE floor in the first turbo iteration, although it achieves the analytical performance after performing the sixth iteration. This is because, when $\hat{w} \approx W$, the TS matrix $\mathbf{\Phi}_{t,\mathcal{D}_{\hat{w}},[\hat{w}]}$ becomes a thin matrix, which deteriorates the accuracy of the initial delay profile $\hat{\mathbf{d}}_{\mathbf{H}}^{[0]}$ given $\mathcal{A}^{[0]} = \mathcal{D}_{\hat{w}}$ in the first turbo iteration.

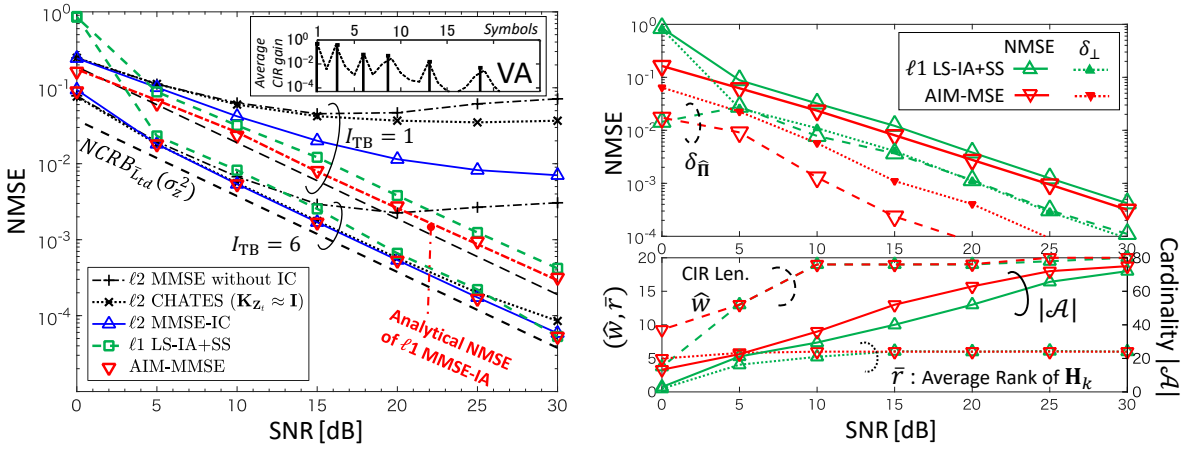
The problem can be improved by initializing $\mathcal{A}^{[0]}$ with $\bar{\mathcal{A}}_{\hat{w}} (\subseteq \mathcal{D}_{\hat{w}})$ obtained in the CIR length estimation (37). As observed from Fig. 4(a), the $\ell 1$ LS-IA given $\mathcal{D} = \bar{\mathcal{A}}_{\hat{w}}$ solves the MSE floor problem in the first iteration, since we can select the active-set (38) such that $|\bar{\mathcal{A}}_{\hat{w}}| \ll |\mathcal{D}_{\hat{w}}| \leq WN_T$ holds in the PB model. Therefore, the AIM-MMSE algorithm does not have the biased error due to the pseudo-inverse matrix. (i.e., $\mathbb{E}[\|\Delta_0^{\rm LS}(\bar{\mathcal{A}}_{\hat{w}})\|^2] = 0$ holds in Lemma 1.)

The biased error can, in general, be improved by using the cardinality regularization factor $\alpha$. The upper bound $\alpha_0$ of the factor (27) is calculated as $\alpha_0 = 0.8$ for the assumed system parameters $(W, N_T, L_t)$. Fig. 4(b) demonstrates the analytical NMSE performance in the length-$w$ uniform distribution channels with the density ratio $\rho = w/W$, where the biased error due to the pseudo-inverse matrix is computed according to Lemma 1. As illustrated in Fig. 4(b), the $\ell 1$ MMSE-IA technique diverges from the NCRB in the regime $\rho \geq \alpha_0 = 0.8$. Hence, in this section, the CIR length estimation is performed with $\alpha = 0.7 < \alpha_0$ such that the TS submatrix $\mathbf{\Phi}_{t,\bar{\mathcal{A}}_{\hat{w}},[\hat{w}]}$ becomes a fat matrix.

### C. The Algorithm Selection in the AIM-MMSE

This subsection describes how the AIM-MMSE algorithm adaptively selects the $\ell 1$ MMSE-IA and the $\ell 2$ MMSE-IC according to the LLR's accuracy. We, then, see that the $\ell 1$ MMSE-IA should not be performed alone by showing the NMSE tracking performance.

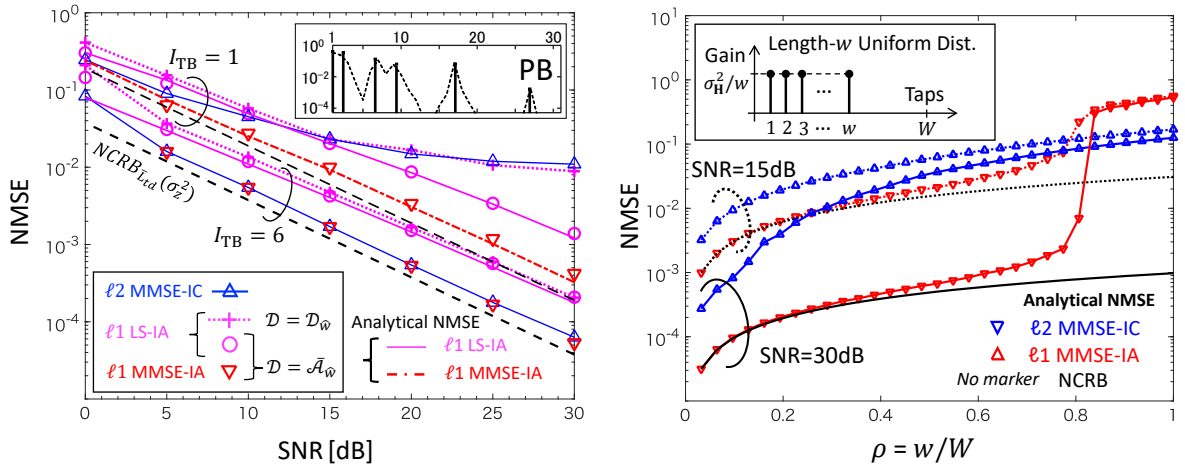
*1) MSE convergence v.s. the mutual information (MI):* We define the LLR's accuracy by the MI $\mathcal{I}_{\rm EQU}^a = \mathcal{I}(\lambda_{\rm EQU}^a; c)$ between the LLR $\lambda_{\rm EQU}^a$ and the coded bits $c$ at the transmitter by $\frac{1}{2}\sum_{\mathtt{m}=\pm 1}\int_{-\infty}^{+\infty} \mathrm{p_r}(\lambda_{\rm EQU}^a|\mathtt{m}) \ \log_2 \frac{\mathrm{p_r}(\lambda_{\rm EQU}^a|\mathtt{m})}{\mathrm{p_r}(\lambda_{\rm EQU}^a)} \ d\lambda_{\rm EQU}^a$, where




(a) NMSE performance in the VA scenario.

(b) Error analysis [32] in the 1st turbo iteration ($I_{\rm TB}=1$).

Fig. 3. NMSE performance (a) and the error analysis (b) in the first iteration, where the biased error $\delta_\perp$ due to inactive-set and the projection error $\delta_{\hat{\mathbf{\Pi}}}$ are respectively defined as $\delta_\perp=\mathbb{E}[\|\mathbf{H}_{\mathcal{A}}^{\perp}\|^2]$ and $\delta_{\hat{\mathbf{\Pi}}}=\mathbb{E}[\|\mathbf{H}_{\mathcal{A}}\Delta\hat{\mathbf{\Pi}}_{\hat{\mathbf{V}}}\|^2]$.




(a) NMSE performance in the PB scenario.

(b) Analytical NMSE in uniform distributed channels.

Fig. 4. NMSE performance (a) in the PB scenario and (b) in the uniform distributed channels.

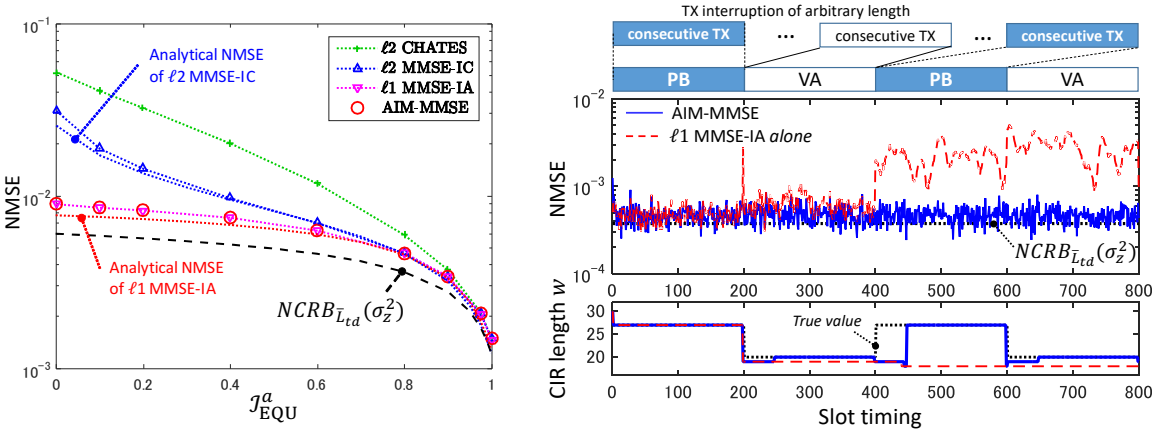
$\mathrm{p_r}(\lambda_{\rm EQU}^a|\mathfrak{m})$ is the conditional probability density function of $\lambda_{\rm EQU}^a$ given $\mathfrak{m}=1-2c$ [33].

Fig. 5(a) depicts NMSE performance over LLR's accuracy at SNR = 15 dB in the PB3 scenario, where we assume that the subspace projection matrix and the delay profile estimates have been obtained with ideal precision after performing sufficient turbo iterations. The analytical performance (43) is also shown. We can observe from Fig. 5(a) that, in a low to moderate MI regime, the $\ell 2$ MMSE-IC improves the NMSE performance over the conventional $\ell 2$ CHATES technique since the $\ell 2$ MMSE-IC computes the covariance matrix $\mathbf{K}_{\mathbf{z}_t,\ell 2}$ (45) correctly. In that MI regime, moreover, the $\ell 1$ MMSE-IA algorithm outperforms the $\ell 2$ MMSE-IC in the NMSE performance. However, all the techniques achieve the NCRB asymptotically as the MI tends to 1.0. The AIM-MMSE algorithm, hence, chooses the $\ell 1$ MMSE-IA in the MI range $0\le\mathcal{I}_{\rm EQU}^a\le 0.8$, and it switches to the $\ell 2$ MMSE-IC in the MI range $0.8<\mathcal{I}_{\rm EQU}^a\le 1.0$ in order to update the subspace projection matrix.

*2) NMSE tracking performance:* As mentioned above, the $\ell 1$ MMSE-IA can improve the MSE performance over the $\ell 2$ MMSE-IC when the LLR fed back from the decoder is not accurate enough. We note that, however, the improvement is guaranteed only when the $\ell 1$ MMSE-IA is performed jointly with the $\ell 2$ MMSE-IC. This subsection verifies the observation described above by investigating the NMSE tracking performance when the $\ell 1$ MMSE-IA is performed alone, where the delay profile vector (46) is computed by (40).

Fig. 5(b) shows the NMSE tracking performance over slot timings. We assume an intermittent transmission scenario [20], referred to as PB-VA, where, as illustrated in Fig. 5(b), CIRs in every other 200 slot-duration follow the PB and VA models, respectively. The received SNR is set at 20 dB. As shown in Fig. 5(b), the *true* effective CIR lengths $w$ at this SNR are 27 and 20 symbols in the the PB and VA models, respectively. Ideal subspace projection computed from the exact subspace vectors $\mathbf{V}_{\mathcal{A}}$ is used to investigate the impact of the CIR length estimation accuracy on the NMSE tracking performance.

We can observe from Fig. 5(b) that the $\ell 2$ MMSE-IC estimates the effective CIR length correctly except the first $L_M$ slots in every 200 slot-duration. The $\ell 1$ MMSE-IA can also estimate the effective CIR length correctly until the first 400




(a) NMSE performance over the LLR's $\mathcal{I}^a_{\mathrm{EQU}}$.

(b) NMSE tracking performance over slot timings.

Fig. 5. Verification of the AIM-MMSE algorithm: The PB3 scenario is used in (a), whereas an intermittent TX scenario PB-VA is assumed in (b). The tracking performance in TX interruptions are omitted. Ideal subspace projection is assumed.

slots. However, after the 400th slot, the CIR lengths obtained by the $\ell 1$ MMSE-IA alone are under-estimated, which causes serious MSE performance deterioration. This is because the CIR length estimation (37) using (40) can track the changes of reduction $(w = 27 \rightarrow 20)$ but fails for the expansion $(w = 20 \rightarrow 27)$, as detailed below.

The CIR length estimation problem (37) is formulated using the analytical MSE performance. As shown in Proposition 1, the analytical MSE is the sum of the unbiased CRB and the biased error involving $\mathtt{tr}\{\mathbb{K}[\hat{\mathbf{H}}^{\perp}_{\bar{\mathcal{A}}_w}]\}$. Note that $\mathrm{CRB}_{\bar{L}_{td,\ell 1}(w)}(\sigma_z^2)$ is proportional to the CIR length parameter $w$, while $\mathtt{tr}\{\mathbb{K}[\hat{\mathbf{H}}^{\perp}_{\bar{\mathcal{A}}_w}]\}$ decreases against $w$. The biased error $\mathtt{tr}\{\mathbb{K}[\hat{\mathbf{H}}^{\perp}_{\bar{\mathcal{A}}_w}]\}$ defined in Proposition 1 can practically be approximated by using the delay profile $\bar{\mathbf{d}}_{\mathbf{H}}$ (46) computed from the past $\ell 2$ channel estimates. However, we cannot accurately estimate $\mathtt{tr}\{\mathbb{K}[\hat{\mathbf{H}}^{\perp}_{\bar{\mathcal{A}}_w}]\}$ from (40) since the biased error has been lost, which makes (37) under-estimated. The AIM-MMSE algorithm is, therefore, necessary to perform the $\ell 1$ MMSE-IA jointly with the $\ell 2$ MMSE-IC in order to determine the optimal CIR length constraint.

### *D. Comparison between the PA-CoSaMP and AAD Algorithms*

Fig. 6(a) shows the NMSE performance of the PA-CoSaMP algorithm in exactly sparse channel realizations, where the CIRs follow the VA30 scenario. However, as shown in the right top conner of Fig. 6(a), the path positions are modified to $\{1, 3, 6, 9, 13, 19\}$, and the effect of the pulse shaping filtering is ignored in this toy example. As observed from Fig. 6(a), the PA-CoSaMP-based $\ell 1$ LS channel estimation techniques [11], [18] achieves the CRB in the exactly sparse channels if the initial active-set and the CIR length estimate are ideally given. This is because 1) the channel estimate obtained using the PA-CoSaMP can also be described by the temporally restricted LS estimate (22); and 2) in the exactly sparse channels, the LS estimate (22) achieves the CRB since the eigen space coincides with the temporal space.

However, we notice that the initial parameters of the PA-CoSaMP is computed by using adjusting factors determined empirically. As described in [11], initially, the PA-CoSaMP obtains a *rough* channel estimate $\hat{\mathbf{H}}_{cc}$ by performing a circular convolution. The PA-CoSaMP detects, then, the initial active-set $\mathcal{A}'_{[0]} = \{j \mid d_{cc}(j) > p_{\mathrm{th}}(\beta)\}$, where $d_{cc}(j)$ denotes the $j$-th entry of $\mathtt{diag}\{\hat{\mathbf{H}}^{\mathsf{H}}_{cc}\hat{\mathbf{H}}_{cc}\}$. The threshold $p_{\mathrm{th}}(\beta)$ may be defined as $p_{\mathrm{th}}(\beta) = \beta^2 \mathrm{MSE}(\mathbf{H}^{\mathrm{LS}}_{\ell 2})/(WN_T)$ according to the concept of the AAD approach, where $\beta$ denotes an adjusting factor. However, no systematic method for determining the optimal parameter $\beta$ is provided in [11], [18], [34]. Moreover, the PA-CoSaMP obtains the effective CIR length by using $\mathcal{A}'_{[0]}$. According to our experiments, nevertheless, the range of $\mathcal{A}'_{[0]}$ needs to be shrunken in order to avoid the last $L_s$ symbols. This is because, as shown in Fig. 6(b), the rough channel estimate $\hat{\mathbf{H}}_{cc}$ can suffer from the Gibbs phenomenon since the received TS itself does not have the circulant property in the assumed signal model.

It should be emphasized that, as observed from Fig. 6(a), the AAD algorithm can achieve the expected performance without empirical adjusting factors such as $\beta$ and $L_s$, since it is developed by exploiting the analytical MSEs.

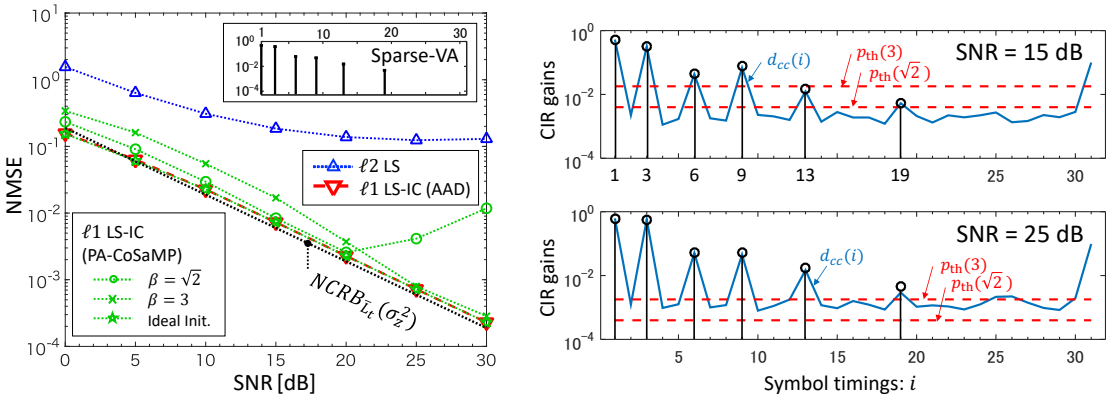
### *E. BER Performance*

The average SNR used in the following BER simulations is defined in association with the average energy per bit to noise density ratio $(\mathrm{E_b/N_0})$ as

$$\mathrm{SNR} = \sigma_x^2 \ (\sigma^2_{\mathbf{H}}/N_R) \ \eta \cdot \mathrm{E_b/N_0}, \tag{54}$$

where the spectral efficiency of the TX format structure is given by $\eta = L_{\mathrm{info}}/L_{\mathrm{frm}}$ with a frame length $L_{\mathrm{frm}} = L_{\mathrm{S}}N_{\mathrm{S}}$ in symbol. Moreover, the variances of a transmitted symbol and CIRs per TX stream are set at $(\sigma_x^2, \sigma^2_{\mathbf{H}}) = (1, 1)$.

The proposed techniques are verified from two viewpoints: 1) BER performance without a frame length constraint; 2) BER performance with a fixed frame length. The first verification aims to investigate whether the receiver using TX format 1 without the CP and GI sections improves the BER performance over that with TX format 2, when channel estimation is actually performed. The second item is motivated by a practical premise that the frame length is a constant




(a) NMSE performance at $I_{\mathrm{TB}} = 1$.

(b) Active-set detection using the circular convolution.

Fig. 6. Comparison between the PA-CoSaMP and AAD algorithms in the *sparse*-VA scenario with the effect of the pulse shaping filters ignored. The parameter $L_s$ is set at 2 to avoid the Gibbs phenomenon observed in the last several symbols.

in most communication systems. Note that the CHATUE is utilized for TX format 1 due to no CP-transmission, while the conventional turbo equalization with CP-transmission (TEQ-CP) is assumed for evaluating performance with TX formats 2 and 3.

*1) BER performance without a frame length constraint:* Since the definition of $\mathrm{E_b/N_0}$ (54) takes the frame length into account, we can compare TX format 1 with TX format 2 even when their frame lengths are different. Moreover, the expected BER gain with TX format 1 over the baseline format 2 is at most $10\log_{10}(\eta_2/\eta_1) = 0.6$ dB, where $\eta_N$ denotes the spectral efficiency of TX format $N$.

Fig. 7(a) compares the BER performance between the CHATUE and TEQ-CP receivers. The VA30 scenario is assumed. As observed from Fig. 7(a), in the first iteration, the CHATUE receiver using the $\ell 2$ MMSE-IC channel estimation exhibits BER deterioration compared to the TEQ-CP. This is because, as shown in Fig. 3(a), the $\ell 2$ MMSE-IC suffers from the MSE floor in the first iteration. As depicted in Fig. 7(a), however, the AIM-MMSE algorithm solves the problem completely. Moreover, the receiver using the AIM-MMSE algorithm achieves a BER gain of 0.5 dB over the TEQ-CP receiver after performing six turbo iterations.

Fig. 7(b) shows the BER convergence performance over turbo iterations at $\mathrm{E_b/N_0} = 11.5$ dB. Notice that the CHATUE receiver requires more iterations than the TEQ-CP, since it reconstructs the circulant property of channel parameters by utilizing the LLR fed back from the decoder. As observed from Fig. 7(b), the CHATUE receiver using the AIM-MMSE algorithm can obtain $\mathrm{BER} = 10^{-5}$ in four iterations while the $\ell 2$ MMSE-IC needs to perform more than five iterations. Therefore, by exploiting the AIM-MMSE channel estimation, we can obtain the expected BER gain due to the spectrally efficient frame format while improving the BER convergence performance in IBI channels.

*2) BER performance with a fixed frame length:* Fig. 8 shows BER performances with the CHATUE and TEQ-CP receivers in the PB3 scenario. TX formats 1 and 3 are used for the CHATUE and TEQ-CP techniques, respectively. Note that, in Table III, TX formats 1 and 3 have the same spectral efficiency $\eta$, since TX format 3 also transmits $L_{\mathrm{info}} = 2048$ information bits by encoding with an $R_c = 8/15$ puncutred code derived from the half-rate mother convolutional code $(7,5)_8$.

As shown in Fig. 8, the BCJR decoder using TX format 1 achieves a BER gain of 0.5 dB over that using the TX format 3 in a static AWGN channel. However, the gain significantly depends on the channel estimation accuracy. When the channel estimation is actually performed in the PB3 scenario, the CHATUE receiver assuming TX format 1 improves the BER gain more than 1 dB over the TEQ-CP with TX format 3. Moreover, we observe that the CHATUE receiver exploiting the AIM-MMSE channel estimation achieves almost the same BER performance as that of the TEQ-CP assuming known CIRs. Hence, by jointly utilizing the AIM-MMSE channel estimation and the CHATUE algorithms, the IBI problems incurred by TX format 1 without the CP nor GI sections can be solved perfectly. Thereby, the proposed algorithm enables us to pursue spectrally efficient transmission in practical MIMO systems.

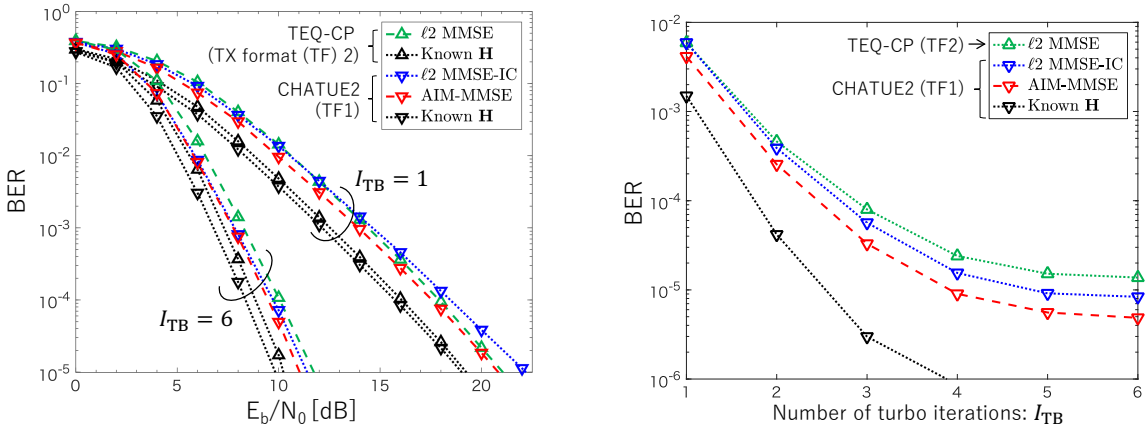
## VII. Conclusions

This paper has proposed a new IBI countermeasure, referred to as AIM-MMSE, which adaptively utilizes the IBI avoidance and cancelation strategies according to the analytical MSE performance. Specifically, the new turbo channel estimation algorithm avoids the IBI by performing the conditional $\ell 1$ MMSE channel estimation technique when the soft replica is not accurate enough, while solving the subspace-based $\ell 2$ MMSE problem to update the unbiased statistics. The simulation results shown in this paper verified that the AIM-MMSE algorithm solves the MSE floor problem and, hence, improves the BER performance.

## References

[1] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A survey on 5G networks for the internet of things: Communication technologies and challenges," *IEEE Access*, vol. 6, pp. 3619–3647, 2018.

[2] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "On the optimality of single-carrier transmission in large-scale antenna systems," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 276–279, Aug 2012.




(a) BER performance over $E_b/N_0$.

(b) BER performance over iterations.

Fig. 7. BER performance without constraint on the frame length. The $4 \times 4$ MIMO system in the VA30 scenario is assumed. TX formats 1 and 2 are used for the CHATUE and TEQ-CP techniques, respectively.

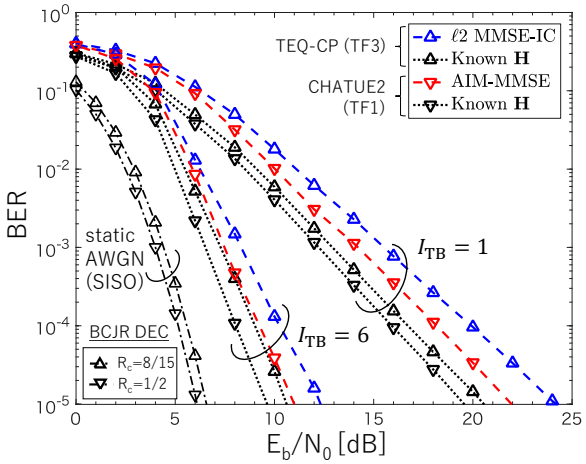



Fig. 8. BER performance in the PB3 scenario with a frame length constraint. For reference, BER performance with the BCJR decoder in the 1-path static AWGN SISO channel is also shown. The CHATUE receiver and the BCJR decoder with $R_c = 1/2$ use the TX format (TF) 1, whereas the TF3 is assumed for the TEQ-CP receiver and the BCJR decoder with $R_c = 8/15$.


[3] N. Michailow and G. Fettweis, "Low peak-to-average power ratio for next generation cellular systems with generalized frequency division multiplexing," in *2013 International Symposium on Intelligent Signal Process. and Commun. Sys.*, Nov 2013, pp. 651–655.

[4] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 120–128, February 2017.

[5] P. H. Lin, S. H. Lai, S. C. Lin, and H. J. Su, "On secrecy rate of the generalized artificial-noise assisted secure beamforming for wiretap channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1728–1740, September 2013.

[6] J. Karjalainen, K. Kansanen, N. Veselinovic, and T. Matsumoto, "Frequency domain joint-over-antenna MIMO turbo equalization," in *Signals, Systems and Computers, 2005. Conference Record of the Thirty-Ninth Asilomar Conference on*, Oct 2005, pp. 834–838.

[7] S. M. Kay, "Fundamentals of Statistical Signal Processing, volume i: Estimation Theory (v. 1)," 1993.

[8] K. Anwar, Z. Hui, and T. Matsumoto, "Chained turbo equalization for block transmission without guard interval," in *Vehcular Technology Conf. (VTC 2010-Spring), 2010 IEEE 71st*, May. 2010, pp. 1 –5.

[9] Y. Takano and H. J. Su, "Performance of frequency domain multiuser-MIMO turbo equalization without cyclic prefix," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, Oct 2017, pp. 1–6.

[10] Y. Takano, K. Anwar, and T. Matsumoto, "Spectrally efficient frame-format-aided turbo equalization with channel estimation," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1635–1645, May 2013.

[11] L. Dai, Z. Wang, and Z. Yang, "Compressive sensing based time domain synchronous OFDM transmission for vehicular communications," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 460–469, Sep 2013.

[12] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of communication systems: modeling, methodology and techniques*. Springer Science & Business Media, 2006.

[13] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 255–271, Apr 2010.

[14] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *IEEE Transactions on Communications*, vol. 60, no. 12, pp. 3705–3716, Dec 2012.

[15] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in ofdm systems using sparse bayesian learning," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3591–3603, Jul 2014.

[16] European Telecommunications Standards Institute (ETSI), "Spatial channel model for MIMO simulations (3GPP TR 25.996)," Jan. 2017.

[17] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *Journal of Machine Learning Research*, vol. 12, no. Nov, pp. 3371–3412, 2011.

[18] W. Ding, F. Yang, C. Pan, L. Dai, and J. Song, "Compressive sensing based channel estimation for OFDM systems under long delay channels," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 313–321, Jun 2014.

[19] A. Irawan, K. Anwar, and T. Matsumoto, "Chained turbo equalization for multiuser SIMO-OFDM systems without cyclic prefix," in *2012 International ITG Workshop on Smart Antennas (WSA)*, Mar 2012, pp. 301–306.

[20] Y. Takano, M. Juntti, and T. Matsumoto, "$\ell 1$ LS and $\ell 2$ MMSE-based hybrid channel estimation for intermittent wireless connections," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 314–328, Jan 2016.

[21] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. McGraw-Hill, 2008.

[22] M. Nicoli and U. Spagnolini, "Reduced-rank channel estimation for time-slotted mobile communication systems," *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 926–944, 2005.

[23] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284 – 287, Mar. 1974.

[24] M. Tüchler, A. Singer, and R. Koetter, "Minimum mean squared error equalization using a priori information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar 2002.

[25] V. Raghavan, J. H. Kotecha, and A. M. Sayeed, "Why does the kronecker model result in misleading capacity estimates?" *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4843–4864, Oct 2010.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.

[27] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse channel estimation with zero tap detection," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1743–1763, 2007.
[28] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461 – 464, 1978.
[29] S. Van Vaerenbergh, I. Santamaria, W. Liu, and J. Principe, "Fixed-budget kernel recursive least-squares," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, Mar 2010, pp. 1882–1885.
[30] G. Golub and C. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 1983.
[31] J. Salo, G. Del Galdo, J. Salmi, P. Kyosti, M. Milojevic, D. Laselva, and C. Schneider, "MATLAB implementation of the 3GPP spatial channel model (3GPP TR 25.996)," On-line, Jan. 2005, http://www.ist-winner.org/3gpp_scm.html.
[32] Y. Takano, M. Juntti, and T. Matsumoto, "Performance of an $\ell 1$ regularized subspace-based MIMO channel estimation with random sequences," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 112–115, Feb 2016.
[33] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727 –1737, Oct. 2001.
[34] F. Wan, W. P. Zhu, and M. N. S. Swamy, "Semi-blind most significant tap detection for sparse channel estimation of OFDM systems," *IEEE Trans. Circuits Syst. I*, vol. 57, no. 3, pp. 703–713, Mar 2010.

**Yasuhiro Takano** (S'11-M'16) received his Ph.D. and Dr.Sc. (Tech.) degrees, respectively, from Japan Advanced Institute of Science and Technology (JAIST), Japan, and University of Oulu, Finland, in 2016. He is currently with Kobe University, Japan. His research interests include signal processing for wireless communications.

**Hsuan-Jung Su** (S'90-M'99-SM'12) received the B.S. degree in Electronics Engineering from the National Chiao Tung University, Taiwan, in 1992, and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Maryland, College Park, in 1996 and 1999, respectively. From 1999 to 2000, he was a Postdoctoral Research Associate with the Institute for Systems Research, University of Maryland. From 2000 to 2003, he was with the Bell Laboratories, Lucent Technologies, Holmdel, New Jersey, where he received the Central Bell Labs Teamwork Award in 2002 and the Bell Labs President's Gold Award in 2003 for his contribution to the 3G wireless network design and standardization. In 2003, Dr. Su joined the Department of Electrical Engineering, National Taiwan University, where he is currently a Professor and the Director of the Graduate Institute of Communication Engineering. From 2014 to 2015, Dr. Su was a Visiting Fellow at Princeton University. Dr. Su is an Area Editor of the *Physical Communication (PHYCOM)* journal (Elsevier), and has guest edited special issues for journals such as *IEEE Access*. He has also served on the organizing committees and TPCs of many international conferences, including serving as the Finance Chair of IEEE ICASSP 2009, the Finance Co-Chair and a TPC Track Chair of IEEE VTC 2010 Spring, a TPC Co-Chair of WPMC 2012, a TPC Co-Chair of IEEE GreenCom 2014, the TPC Chair of WOCC 2015, and a TPC Co-Chair of IEEE Globecom 2020. Dr. Su was the Chair (2013-2015) of the Taipei Chapter of the IEEE Information Theory Society, the Secretary and Treasurer (2014-2015), the Vice Chair of the Technical Affairs Committee (2016-2017), and the Vice Chair of the Membership Development Committee (2018-2019) of the IEEE Communications Society Asia-Pacific Board. His research interests cover coding, modulation, signal processing, interference management, resource allocation, and MAC protocols of wireless communication, cognitive, M2M (IoT) and D2D networks.

**Markku Juntti** (S'93-M'98-SM'04) received his M.Sc. (EE) and Dr.Sc. (EE) degrees from University of Oulu, Oulu, Finland in 1993 and 1997, respectively. Dr. Juntti was with University of Oulu in 1992–98. In academic year 1994–95, he was a Visiting Scholar at Rice University, Houston, Texas. In 1999–2000, he was a Senior Specialist with Nokia Networks. Dr. Juntti has been a professor of communications engineering since 2000 at University of Oulu, Centre for Wireless Communications (CWC), where he leads the Communications Signal Processing (CSP) Research Group. He also serves as Head of CWC – Radio Technologies (RT) Research Unit. His research interests include signal processing for wireless networks as well as communication and information theory. He is an author or co-author in almost 500 papers published in international journals and conference records as well as in books *WCDMA for UMTS*, *Signal Processing Handbook* and *5G Wireless Technologies*. Dr. Juntti is also an Adjunct Professor at Department of Electrical and Computer Engineering, Rice University, Houston, Texas, USA. Dr. Juntti is an Editor of IEEE TRANSACTIONS ON COMMUNICATIONS and was an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY in 2002–2008. He was Secretary of IEEE Communication Society Finland Chapter in 1996–97 and the Chairman for years 2000–01. He has been Secretary of the Technical Program Committee (TPC) of the 2001 IEEE International Conference on Communications (ICC 2001), and the Co-Chair of the Technical Program Committee of 2004 Nordic Radio Symposium and 2006 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2006), and the General Chair of 2011 IEEE Communication Theory Workshop (CTW 2011). He has served as Co-Chair of the Signal Processing for Communications Symposium of Globecom 2014 Signal Processing for Communications Symposium, IEEE GlobalSIP 2016 Symposium on Transceivers and Signal Processing for 5G Wireless and mm-Wave Systems, and ACM NanoCom 2018.

**Tad Matsumoto** (S'84-SM'95-F'10) received his B.S., M.S., degrees, both under the mentorship by Prof. Shin-Ichi Takahashi, and Ph.D. degree under the supervision by Prof. Masao Nakagawa, from Keio University, Yokohama, Japan, in 1978, 1980, and 1991, respectively, all in electrical engineering. He joined Nippon Telegraph and Telephone Corporation (NTT) in April 1980. Since he engaged in NTT, he was involved in a lot of research and development projects, all for mobile wireless communications systems. In July 1992, he transferred to NTT DoCoMo, where he researched Code-Division Multiple-Access techniques for Mobile Communication Systems. In April 1994, he transferred to NTT America, where he served as a Senior Technical Advisor of a joint project between NTT and NEXTEL Communications. In March 1996, he returned to NTT DoCoMo, where he served as a Head of the Radio Signal Processing Laboratory until August of 2001; He worked on adaptive signal processing, multiple-input multiple-output turbo signal detection, interference cancellation, and space-time coding techniques for broadband mobile communications. In March 2002, he moved to University of Oulu, Finland, where he served as a Professor at Centre for Wireless Communications. In 2006, he served as a Visiting Professor at Ilmenau University of Technology, Ilmenau, Germany, funded by the German MERCATOR Visiting Professorship Program. Since April 2007, he has been serving as a Professor at Japan Advanced Institute of Science and Technology (JAIST), Japan, while also keeping a cross-appointment position at University of Oulu. He has led a lot of projects funded by Academy-of-Finland, European FP7, and Japan Society for the Promotion of Science as well as by Japanese private companies. Prof. Matsumoto has been appointed as a Finland Distinguished Professor for a period from January 2008 to December 2012, funded by the Finnish National Technology Agency (Tekes) and Finnish Academy, under which he preserves the rights to participate in and apply to European and Finnish national projects. Prof. Matsumoto is a recipient of IEEE VTS Outstanding Service Award (2001), Nokia Foundation Visiting Fellow Scholarship Award (2002), IEEE Japan Council Award for Distinguished Service to the Society (2006), IEEE Vehicular Technology Society James R. Evans Avant Garde Award (2006), and Thuringen State Research Award for Advanced Applied Science (2006), 2007 Best Paper Award of Institute of Electrical, Communication, and Information Engineers of Japan (2008), Telecom System Technology Award by the Telecommunications Advancement Foundation (2009), IEEE COMMUNICATION LETTERS Exemplary Reviewer (2011), Nikkei Wireless Japan Award (2013), IEEE VTS Recognition for Outstanding Distinguished Lecturer (2016), and IEEE TRANSACTIONS ON COMMUNICATIONS Exemplary Reviewer (2018),. He is a Fellow of IEEE and a Member of IEICE. He is serving as an IEEE Vehicular Technology Distinguished Speaker since July 2016.