Title	ℓ1 LS and ℓ2 MMSE-based Hybrid Channel Estimation for Intermittent Wireless Connections
Author(s)	Takano, Yasuhiro; Juntti, Markku; Matsumoto, Tad
Citation	IEEE Transactions on Wireless Communications, 15(1): 314-328
Issue Date	2015-08-25
Type	Journal Article
Text version	author
URL	http://hdl.handle.net/10119/12943
Rights	This is the author's version of the work. Copyright (C) 2015 IEEE. IEEE Transactions on Wireless Communications, 15(1), 2015, 314-328. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Description	

$\ell 1$ LS and $\ell 2$ MMSE-based Hybrid Channel Estimation for Intermittent Wireless Connections

Yasuhiro Takano, *Student Member, IEEE,* Markku Juntti, *Senior Member, IEEE,* and Tad Matsumoto, *Fellow, IEEE*

***Abstract*—Broadband wireless channels observed at a receiver cannot fully exhibit dense nature in a low to moderate signal-to-noise ratio (SNR) regime, if the channels follow a typical propagation scenario such as Vehicular-A or Pedestrian-B. It is hence expected that $\ell 1$ regularized channel estimation methods can improve channel estimation performance in the broadband wireless channels. However, it is well-known that the $\ell 2$ multi-burst (MB) channel estimation achieves the Cramér-Rao bound (CRB) asymptotically. This is because the $\ell 2$ MB technique formulated as a minimum mean square error (MMSE) problem improves the mean squared error (MSE) performance by utilizing the subspace projection. Performance analysis shows that $\ell 1$ regularized channel estimation does not improve the MSE performance significantly over the $\ell 2$ MB technique so far as the subspace channel model assumption is correct. We demonstrate, however, a receiver with $\ell 1$ regularized channel estimation can improve bit error rate (BER) performance if the assumption is not always correct. For this purpose, we focus on *intermittent* transmission (TX) scenario which is defined as a generalized TX sequence having arbitrary length interruption between two continuous TX bursts. A receiver with the $\ell 2$ MB method suffers from BER deterioration in an intermittent TX scenario having abrupt channel changes. As a solution to the problem, we propose a new algorithm which is a hybrid of $\ell 1$ regularized least squares (LS) and $\ell 2$ MMSE channel estimation techniques. Simulation results show that the receiver with the proposed algorithm achieves a significant BER gain over that of the $\ell 2$ MB technique in the intermittent TX scenario.**

***Index Terms*—Subspace-based channel estimation, turbo channel estimation, compressed sensing, orthogonal matching pursuit (OMP), Akaike information criterion (AIC), Bayesian information criterion (BIC).**

I. Introduction

COMPRESSED sensing (CS) [1]-based $\ell 1$ regularized channel estimation can improve estimation performance over ordinary $\ell 2$ channel estimation if a channel impulse response (CIR) observed at a receiver exhibits sparse structure having several tap weights close to zero [2], [3]. This happens often, e.g., in under-water communication channels [4]–[6]. Broadband wireless channels are, in general, not observed as sparse channels at a receiver due to the effect of transmit (Tx) and receive (Rx) filters required to perform discrete-time processing properly. However, they can be seen as *approximately* sparse channels in a low to moderate signal-to-noise ratio (SNR) regime if the channels follow a typical propagation scenario such as Vehicular-A (VA) or Pedestrian-B (PB) [7]. The dominant path components in such propagation scenarios are, as shown in Fig. 1, not uniformly distributed in the observation domain after the Tx/Rx filtering. Furthermore, some of the small path components can be completely buried under the noise in a low SNR regime. Therefore, as described in [8], CS-based channel estimation techniques are expected to improve estimation performance in broadband wireless channels as well.

However, an ordinary $\ell 2$ multi-burst (MB) channel estimation can achieve the Cramér-Rao bound (CRB) asymptotically in the multi-path channels following the subspace channel model assumption [9]–[12]. This is because the $\ell 2$ MB technique formulated as a minimum mean square error (MMSE) problem improves the mean squared error (MSE) performance by utilizing the subspace projection. It can be seen that the $\ell 2$ MB technique performs noise *compression* in eigen domain of the signal of interest. Therefore, this paper investigates if there are any advantages of $\ell 1$ regularized channel estimation over the $\ell 2$ MB method in broadband wireless channels. For this purpose, *intermittent* transmission (TX[1]) scenario is focused on. This paper defines the intermittent TX scenario as a generalized TX sequence which is constructed with a repetition[2] of a TX chunk and a TX interruption of arbitrary duration, where a TX chunk is a certain length continuous data TX duration. The two TX chunks do not always follow the identical channel model due to the TX interruption. Thereby, the $\ell 2$ MB technique may suffer from a tracking error problem, since the subspace channel model assumption can partially be incorrect at borders of the TX chunks. As a solution to the problem, we propose a new channel estimation algorithm which is a hybrid of $\ell 1$ least squares (LS) and $\ell 2$ MB techniques.

The communication system assumed in this paper is a turbo receiver framework over broadband multiple-input and multiple-output (MIMO) wireless channels due to the following motivations: it is well-known that MIMO communication systems can improve the spectral-efficiency and the transmission rate [13], [14]. However, channel estimation needed for practical MIMO systems has the problem that the

Y. Takano and T. Matsumoto are with Japan Advanced Institute of Science and Technology (JAIST) 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan (e-mail: {yace.takano; matumoto}@jaist.ac.jp).

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

This research was conducted under the financial support of the double degree program between JAIST and University of Oulu.

[1]This paper distinguishes TX (transmission) from Tx (transmit).

[2]The *repetition* applies to the TX scenario structure only. Each TX chunk transmits different data bursts.

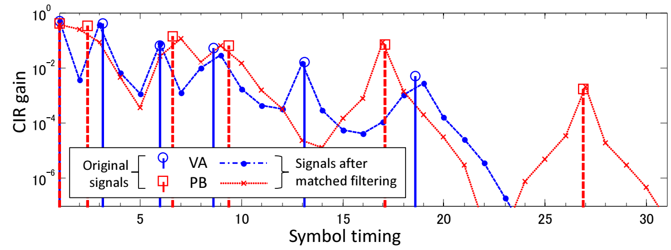

Fig. 1. Channel delay profiles of VA and PB channel realizations. We note that the receiver can observe CIRs only as that after the matched filtering. A transmission bandwidth of 7 MHz with a carrier frequency of 2 GHz is assumed. The implementation of the matched filter is described in Section V. The CIR gain vector is defined by $\mathtt{diag}\{\sum_{k=1}^{N_T}\mathbb{E}[\mathbf{H}_k^{\mathrm{H}}(l)\cdot\mathbf{H}_k(l)]\}/(\sigma_{\mathbf{H}}^2 N_T)$, where the notation follows the description in Section II.

number of the CIR parameters increases due to the spatial multiplexing. Hence, $\ell 1$ regularized channel estimation is expected to improve estimation performance in broadband MIMO wireless channels by *compressing* the number of parameters to be estimated. Furthermore, it is shown in [6], [15], [16] that a turbo receiver with an $\ell 1$ regularized channel estimation can achieve a bit error rate (BER) gain over that with an ordinary $\ell 2$ channel estimation. However, the channel estimation performance is not addressed in [6], [15], [16]. Therefore, this paper aims to clarify the MSE performance of $\ell 1$ regularized channel estimation techniques in a MIMO turbo receiver through theoretical analysis. Simulation results are also presented to verify the theoretical analysis.

This paper is organized as follows. Section II describes the system model assumed in this paper. Section III proposes new $\ell 1$ regularized MB and hybrid channel estimation algorithms. Section IV describes analytical performance bounds of the new techniques. Section V presents results of computer simulations conducted to verify the analytical performance. This paper is concluded in Section VI with some concluding remarks.

Notations: The bold lower-case $\mathbf{x}$ and upper-case $\mathbf{X}$ denote a vector and a matrix, respectively. For a matrix $\mathbf{X}$, its transpose and transposed conjugate are denoted as $\mathbf{X}^{\mathrm{T}}$ and $\mathbf{X}^{\mathrm{H}}$, respectively. $\mathtt{vec}(\mathbf{X})$ is a vectorization operator to produce an $MN\times 1$ vector by stacking the columns of an $M\times N$ matrix $\mathbf{X}$. An operator $\mathtt{diag}(\mathbf{X})$ forms a vector from the diagonal elements of its argument matrix $\mathbf{X}$. $\mathtt{svd}(\mathbf{X})=\mathbf{UDV}^{\mathrm{H}}$ is the singular value decomposition (SVD) of a matrix $\mathbf{X}\in\mathbb{C}^{M\times N}$, where $\mathbf{U}\in\mathbb{C}^{M\times M}$ and $\mathbf{V}\in\mathbb{C}^{N\times N}$ are unitary matrices and $\mathbf{D}\in\mathbb{C}^{M\times N}$ is a rectangular matrix with a square diagonal matrix on the left top corner. $\mathbf{X}|_{\mathcal{A}}$ is a submatrix composed of the column vectors in a matrix $\mathbf{X}$, the columns of which are defined by index set $\mathcal{A}$. Similarly, $\mathbf{x}|_{\mathcal{A}}$ is a subvector of a vector $\mathbf{x}$ which extracts the elements specified by index set $\mathcal{A}$ from the vector $\mathbf{x}$. The index set is assumed to be sorted in an ascending order and can be denoted by $\mathcal{A}=\{i:j\}=\{i,i+1,...,j\}$, when $\mathcal{A}$ is composed of a contiguous integer sequence with positive integers $i\le j$. $|\mathcal{A}|$ denotes the cardinality of the argument set $\mathcal{A}$. A weighted Frobenius norm is defined as $\|\mathbf{X}\|_{\mathbf{W}}^2=\mathrm{tr}\{\mathbf{X}^{\mathrm{H}}\mathbf{W}\mathbf{X}\}$ for a matrix $\mathbf{X}\in\mathbb{C}^{M\times N}$ with a positive definite matrix $\mathbf{W}\in\mathbb{C}^{M\times M}$. In the case of $\mathbf{W}=\mathbf{I}_M$, we simply denote $\|\mathbf{X}\|_{\mathbf{W}}^2=\|\mathbf{X}\|^2$, where $\mathbf{I}_M$ is an $M\times M$ identity matrix. An $\ell 1$ norm for a matrix $\mathbf{X}\in\mathbb{C}^{M\times N}$ is defined as $\|\mathbf{X}\|_1=\sum_{i=1}^{M}\sum_{j=1}^{N}|x_{ij}|$ where x_{ij} is the (i,j)-th element of the matrix $\mathbf{X}$.

II. System Model

This paper assumes a vertical-Bell laboratories layered space-time (V-BLAST) type spatial multiplexing MIMO system [17] as depicted in Fig. 2. A length N_{info} bit binary data information sequence $b(i)$, $1\le i\le N_{\mathrm{info}}$, is channel-encoded into a coded frame $c(i_c)$ by a rate R_c convolutional code (CC) with generator polynomials $(g_1,\cdots,g_{1/R_c})$ and is interleaved by an interleaver (Π). The interleaved coded frame $c_\Pi(j_c)$, $1\le j_c\le N_{\mathrm{info}}/R_c$, 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_B data blocks such that fading is assumed to be static over each burst. A data block is modulated into binary phase shift keyed (BPSK) symbols[3] $x_{d,k}(j_s;l)$ with variance σ_x^2 and the modulation multiplicity $M_b=1$. The k-th Tx antenna transmits data symbols $\mathbf{x}_{d,k}(l)=[x_{d,k}(1;l),\cdots,x_{d,k}(N_d;l)]^{\mathrm{T}}$ together with a length N_t symbol training sequence (TS) $\mathbf{x}_{t,k}(l)$ and a length N_{CP} symbol cyclic prefix (CP), using single carrier signaling, where l denotes the burst timing index. The data symbol length N_d in a burst is defined as $N_d=N_{\mathrm{info}}/(R_cN_TN_BM_b)$. As depicted in Fig. 2, the burst format has two length N_G symbol guard intervals (GIs) following the training and the data sequences, respectively, to avoid[4] inter-block-interference (IBI).

The receiver observes signal sequences $\mathbf{y}_n(l)$ with N_R receive antennas. The received signal suffers from inter-symbol-interference (ISI) due to fading frequency selectivity, and from complex additive white Gaussian noise (AWGN) as well. The ISI length is at most $L_{\mathrm{ISI}}=W-1$ symbols under the assumption that the maximum CIR length is W. The received signal can be described in a matrix form $\mathcal{Y}(l)$ as,

$$\mathcal{Y}(l)=\mathcal{H}(l)\mathcal{X}(l)+\mathcal{Z},\qquad(1)$$

where

$$\begin{aligned}\mathcal{Y}(l)&=[\mathbf{y}_1(l),\cdots,\mathbf{y}_{N_R}(l)]^{\mathrm{T}}&&\in\mathbb{C}^{N_R\times L_B},\\ \mathcal{X}(l)&=[\mathbf{X}_1^{\mathrm{T}}(l),\cdots,\mathbf{X}_{N_T}^{\mathrm{T}}(l)]^{\mathrm{T}}&&\in\mathbb{C}^{WN_T\times L_B},\\ \mathcal{H}(l)&=[\mathbf{H}_1(l),\cdots,\mathbf{H}_{N_T}(l)]&&\in\mathbb{C}^{N_R\times WN_T},\\ \mathcal{Z}&=[\mathbf{z}_1,\cdots,\mathbf{z}_{N_R}]^{\mathrm{T}}&&\in\mathbb{C}^{N_R\times L_B},\end{aligned}$$

and the burst length is $L_B=N_t+N_{\mathrm{CP}}+N_d+2N_G$. The $W\times L_B$ matrix $\mathbf{X}_k(l)$ is a Toeplitz matrix whose first row vector is $[\mathbf{x}_{t,k}^{\mathrm{T}}(l),\mathbf{0}_{N_G}^{\mathrm{T}},\mathbf{x}_{d,k}^{\mathrm{T}}(l)|_{(N_d-W+1):N_d},\mathbf{x}_{d,k}^{\mathrm{T}}(l),\mathbf{0}_{N_G}^{\mathrm{T}}]\in\mathbb{C}^{1\times L_B}$. The expected variance of the CIR matrix $\mathbf{H}_k(l)$ for the k-th transmission (TX) stream is $\mathbb{E}[\|\mathbf{H}_k(l)\|^2]=\sigma_{\mathbf{H}}^2$ with a constant $\sigma_{\mathbf{H}}^2$. Furthermore, the CIR satisfies a property that the spatial covariance matrix $\mathbb{E}[\mathbf{H}_k(l)\mathbf{H}_k(l)^{\mathrm{H}}]$ is of full-rank by assuming no unknown interferences [9], [12]. The noise vector at the n-th Rx antenna $\mathbf{z}_n$ follows $\mathcal{CN}(\mathbf{0},\sigma_z^2\mathbf{I}_{L_B})$.

[3] For the sake of simplicity, we assume binary modulation in this paper. However, extension to higher order modulation is straightforward [18].

[4] Although it is out of scope of this paper, the GIs can be eliminated by using the chained turbo estimation (CHATES) [19].

As depicted in Fig. 2, the receiver performs channel estimation (EST) jointly over the Rx antennas while also obtaining the extrinsic log-likelihood ratio (LLR) $\lambda^e_{\mathrm{EQU},k}$ for the k-th TX stream by means of frequency domain soft-cancellation and minimum mean-square-error (FD/SC-MMSE) MIMO turbo equalization [20] (EQU). The N_T LLRs $\lambda^e_{\mathrm{EQU},k}$ are parallel-to-serial (P/S)-converted to form an extrinsic LLR sequence λ^e_{EQU} corresponding to the interleaved coded frame $c_{\Pi}(j_c)$ at the transmitter. An *a priori* LLR λ^a_{DEC} for the channel decoder (CC^{-1}) is obtained by deinterleaving λ^e_{EQU}. The channel decoder performs decoding for λ^a_{DEC} by using the Bahl, Cocke, Jelinek and Raviv (BCJR) algorithm [21], and outputs the *a posteriori* LLR λ^p_{DEC}. After several iterations, CC^{-1} outputs the estimates of the transmitted sequence $\hat{\mathbf{b}}$ by making a hard decision on λ^p_{DEC}. Both EST and EQU utilize the soft replica[5] of the transmitted symbols $\hat{\mathbf{x}}_{d,k}$ which is generated from the equalizer's *a priori* LLR λ^a_{EQU}. We note that LLR λ^a_{EQU} is the interleaved version of the extrinsic LLR λ^e_{DEC} which is obtained as $\lambda^e_{\mathrm{DEC}} = \lambda^p_{\mathrm{DEC}} - \lambda^a_{\mathrm{DEC}}$ according to the turbo principle.

III. Channel Estimation Algorithms

This section proposes new $\ell1$ regularized MB and hybrid channel estimation algorithms after showing $\ell1$ regularized LS channel estimation. The computational complexity order required for the new techniques is discussed at the end of this section.

A. $\ell1$ Regularized LS Channel Estimation ($\ell1$ LS)

1) Problem formulation: By imposing an $\ell1$ regularizing term to an ordinary $\ell2$ LS problem, $\ell1$ LS channel estimation becomes

$$\hat{\mathcal{H}}^{LS}_{\ell1}(l) = \arg\min_{\mathcal{H}} \mathcal{L}_{td}(l,\mathcal{H}) + \lambda(l)\|\mathcal{H}\|_1 \tag{2}$$

with a Lagrange multiplier $\lambda(l)$ [22], [23]. Similar to [12], the equivalent negative log-likelihood function $\mathcal{L}_{td}(l,\mathcal{H})$ is defined as $\mathcal{L}_{td}(l,\mathcal{H}) = \mathcal{L}_t(l,\mathcal{H}) + \mathcal{L}_d(l,\mathcal{H})$, where we have

$$\mathcal{L}_t(l,\mathcal{H}) = \frac{1}{\sigma_z^2}\|\mathcal{Y}_t(l) - \mathcal{H}\mathcal{X}_t(l)\|^2, \tag{3}$$

$$\mathcal{L}_d(l,\mathcal{H}) = \frac{1}{\sigma_z^2}\|\mathcal{Y}_d(l) - \mathcal{H}\hat{\mathcal{X}}_d(l)\|^2_{\mathbf{\Gamma}(l)}. \tag{4}$$

Received signal matrices for the training and data sections are respectively defined as $\mathcal{Y}_t(l) = \mathcal{Y}(l)|_{1:\tilde{N}_t}$ and $\mathcal{Y}_d(l) = \mathcal{Y}(l)|_{(\eth+1):(\eth+\tilde{N}_d)}$, where input signal lengths are $\tilde{N}_t = N_t + W$ and $\tilde{N}_d = N_d$. The offset $\eth$ is chosen as $\eth = N_t + N_G + N_{\mathrm{CP}} + W$ so that the received data section avoids IBI from CP. Correspondingly, we define a Toeplitz matrix $\mathcal{X}_t(l) = \mathcal{X}(l)|_{1:\tilde{N}_t}$. $\hat{\mathcal{X}}_d(l)$ is the soft replica of $\mathcal{X}_d(l)$, where we denote $\mathcal{X}_d(l) = \mathcal{X}(l)|_{(\eth+1):(\eth+\tilde{N}_d)}$. The weight matrix $\mathbf{\Gamma}(l)$ is defined as $\mathbf{\Gamma}(l) = \sigma_z^2\left(\sigma_z^2\mathbf{I}_{N_R} + \Delta\sigma_d^2\mathbf{R}_{\mathcal{H}\mathcal{H}}(l)\right)^{-1}$, where we denote $\Delta\sigma_d^2 = \sum_{k=1}^{N_T}\mathbb{E}[\|\hat{\mathbf{x}}_{d,k}(l) - \mathbf{x}_{d,k}(l)\|^2]/(N_d N_T)$ and $\mathbf{R}_{\mathcal{H}\mathcal{H}}(l) = \mathcal{H}(l)\mathcal{H}(l)^{\mathrm{H}}$. The $\ell1$ regularized LS problem can be solved with the zero-tap detection (ZD) [3] or orthogonal matching pursuit [24] (OMP)-based algorithms. Before detailing a ZD-based algorithm, we briefly show a temporally restricted MIMO channel estimation technique which can be utilized commonly for the ZD and OMP-type $\ell1$ solvers.

2) Temporally restricted MIMO LS channel estimation: Let us assume the symbol timings of significant path components are specified in a column index set $\mathcal{A}$ of the CIR matrix $\mathcal{H}$. This paper refers to the index set $\mathcal{A}$ as *active-set* [25], hereafter. Moreover, this paper denotes a column-*shrunk* $N_R \times |\mathcal{A}|$ CIR matrix as $\mathcal{G}_{\mathcal{A}} = \mathcal{H}|_{\mathcal{A}}$, or equivalently $\mathcal{G}_{\mathcal{A}} = \mathcal{H}\mathbf{P}_{\mathcal{A}}$, where a $WN_T \times |\mathcal{A}|$ matrix $\mathbf{P}_{\mathcal{A}}$ is defined so that the (m,n)-th entry is set at 1 if the n-th element in $\mathcal{A}$ is m, otherwise, at zero.

The ZD and OMP-type algorithms determine an active-set $\mathcal{A}$ under a certain criterion. Simultaneously, the algorithms obtain a possible estimate $\hat{\mathcal{H}}_{\mathcal{A}}(l) = \hat{\mathcal{G}}_{\mathcal{A}}(l)\mathbf{P}^{\mathrm{T}}_{\mathcal{A}}$ by minimizing the conditional negative log-likelihood function, given the active-set $\mathcal{A}$, as

$$\hat{\mathcal{G}}_{\mathcal{A}}(l) = \arg\min_{\mathcal{G}} \mathcal{L}_{td}(l, \mathcal{G}\mathbf{P}^{\mathrm{T}}_{\mathcal{A}} \mid \mathcal{A}). \tag{5}$$

The problem (5) can be seen as an $\ell2$ LS channel estimation technique by using a temporally restricted (or row-shrunk) training $\mathbf{\Phi}_{t,\mathcal{A}} = \mathbf{P}^{\mathrm{T}}_{\mathcal{A}}\mathcal{X}_t = [\mathcal{X}^{\mathrm{T}}_t|_{\mathcal{A}}]^{\mathrm{T}}$ and data $\hat{\mathbf{\Phi}}_{d,\mathcal{A}} = [\hat{\mathcal{X}}^{\mathrm{T}}_d|_{\mathcal{A}}]^{\mathrm{T}}$ sequences.

Similar to the case of SIMO [12], a MIMO turbo receiver can obtain an LS estimate via its vectorization to take account of the weight matrix $\mathbf{\Gamma}(l)$. Specifically, for an active-set $\mathcal{A}$, a length $|\mathcal{A}|N_R$ *compressed* channel estimate vector $\hat{\mathbf{g}}_{\mathcal{A}} = \mathtt{vec}\{\hat{\mathcal{G}}_{\mathcal{A}}\}$ is described as

$$\hat{\mathbf{g}}_{\mathcal{A}} = \mathcal{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}} \cdot \mathtt{vec}\{\mathbf{R}_{\mathcal{Y}\mathbf{\Phi}_{\mathcal{A}}}\} \tag{6}$$

with $\mathcal{R}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}} = \mathcal{P}^{\mathrm{T}}_{\mathcal{A}}\mathcal{R}_{\mathcal{X}\mathcal{X}}\mathcal{P}_{\mathcal{A}}$ and $\mathbf{R}_{\mathcal{Y}\mathbf{\Phi}_{\mathcal{A}}} = \mathbf{R}_{\mathcal{Y}\mathcal{X}}\mathbf{P}_{\mathcal{A}}$, where we denote $\mathcal{P}_{\mathcal{A}} = \mathbf{P}_{\mathcal{A}} \otimes \mathbf{I}_{N_R}$ and omit the burst timing index l for the sake of simplicity. Furthermore, we define

$$\mathcal{R}_{\mathcal{X}\mathcal{X}} = \mathbf{R}^{\mathrm{T}}_{\mathcal{X}\mathcal{X}_t} \otimes \mathbf{I}_{N_R} + \hat{\mathbf{R}}^{\mathrm{T}}_{\mathcal{X}\mathcal{X}_d} \otimes \hat{\mathbf{\Gamma}}, \tag{7}$$

$$\mathbf{R}_{\mathcal{Y}\mathcal{X}} = \mathbf{R}_{\mathcal{Y}\mathcal{X}_t} + \hat{\mathbf{\Gamma}}\mathbf{R}_{\mathcal{Y}\mathcal{X}_d}, \tag{8}$$

where $\mathbf{R}_{\mathcal{X}\mathcal{X}_t} = \mathcal{X}_t\mathcal{X}^{\mathrm{H}}_t$, $\hat{\mathbf{R}}_{\mathcal{X}\mathcal{X}_d} = \hat{\mathcal{X}}_d\hat{\mathcal{X}}^{\mathrm{H}}_d$, $\mathbf{R}_{\mathcal{Y}\mathcal{X}_t} = \mathcal{Y}_t\mathcal{X}^{\mathrm{H}}_t$ and $\mathbf{R}_{\mathcal{Y}\mathcal{X}_d} = \mathcal{Y}_d\hat{\mathcal{X}}^{\mathrm{H}}_d$. The matrix $\hat{\mathbf{\Gamma}}$ is obtained as

$$\hat{\mathbf{\Gamma}} = \sigma_z^2\left(\sigma_z^2\mathbf{I}_{N_R} + \Delta\hat{\sigma}_d^2\hat{\mathbf{R}}_{\mathcal{H}\mathcal{H}}\right)^{-1}, \tag{9}$$

with $\Delta\hat{\sigma}_d^2 = \sigma_x^2 - \sum_{k=1}^{N_T}\|\hat{\mathbf{x}}_{d,k}(l)\|^2/(N_d N_T)$ and $\hat{\mathbf{R}}_{\mathcal{H}\mathcal{H}} = \hat{\mathcal{H}}^{(i-1)}(\hat{\mathcal{H}}^{(i-1)})^{\mathrm{H}}$, where $\hat{\mathcal{H}}^{(i-1)}$ is the channel estimate obtained by the previous (i-1)-th[6] turbo iteration. Finally, the solution to (5) is described as $\hat{\mathcal{G}}_{\mathcal{A}} = \mathtt{mat}_{N_R}\{\hat{\mathbf{g}}_{\mathcal{A}}\}$, where the operation $\mathtt{mat}_N(\mathbf{x})$ forms an $N \times M$ matrix from the argument vector $\mathbf{x} \in \mathbb{C}^{NM\times 1}$, so that $\mathbf{x} = \mathtt{vec}\{\mathtt{mat}_N\{\mathbf{x}\}\}$.

3) The $\ell1$ LS with adaptive active-set detection (AAD): Based on the MSE performance analysis shown in Section IV-A, a new ZD-type algorithm, *AAD*, can be formulated as

$$\mathcal{A} = \arg\min_{\mathcal{A}} \|\hat{\mathcal{G}}\mathbf{P}^{\mathrm{T}}_{\mathcal{A}} - \mathcal{H}\|^2, \tag{10}$$

[5]In the case of BPSK, as shown in [18], the i-th entry in $\hat{\mathbf{x}}_{d,k}$ is generated as $\hat{x}_{d,k}(i) = \sigma_x\tanh(\lambda^a_{\mathrm{EQU},k}(i)/2)$, where $\lambda^a_{\mathrm{EQU},k}(i)$ denotes the i-th S/P-converted the equalizer's *a priori* LLR for the k-th Tx stream.

[6]For the first turbo iteration, $i = 1$, the term $\mathbf{R}_{\mathcal{H}\mathcal{H}}$ is discarded in (9) since the channel estimation is performed with the TS only.

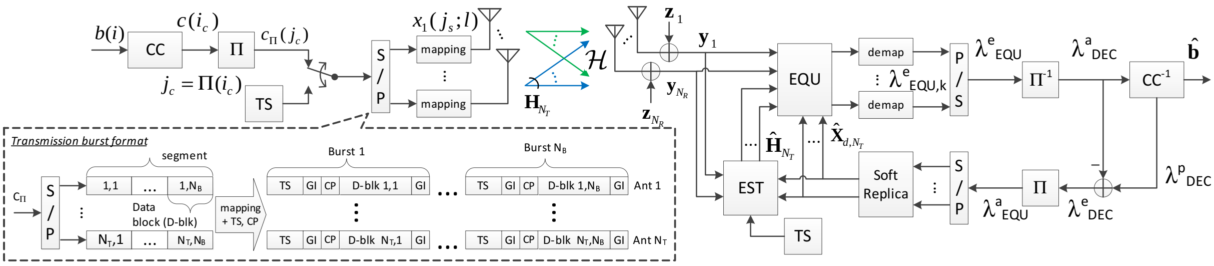

Fig. 2. The system model and the transmission burst format assumed in this paper.

where $\hat{\mathcal{G}}$ is the LS estimate given by (6). We can solve (10) if a channel delay profile $\mathbf{d}_{\mathcal{H}} = \mathtt{diag}\{\mathcal{H}^{\mathrm{H}}\mathcal{H}\}$ is given. In general, however, $\mathbf{d}_{\mathcal{H}}$ is not known since it requires the parameter $\mathcal{H}$ to be estimated. We show, thereby, Algorithm 1 to solve (10) with reasonable computational complexity.

In summary, Algorithm 1 solves the problems (5) and (10) alternately in N_{AAD} iterations. First of all, a possible solution to the problem (10) is obtained by the steps 5 and 6. Algorithm 1 approximates the delay profile by using a possible channel estimate $\hat{\mathcal{G}}_{[n]}$ obtained in the previous iteration, as

$$\hat{\mathbf{d}}_{\mathcal{H}}^{[n]} = \mathbf{P}_{[n]}\ \mathtt{diag}\{\hat{\mathcal{G}}_{[n]}^{\mathrm{H}} \cdot \hat{\mathcal{G}}_{[n]}\}, \tag{11}$$

where $\mathbf{P}_{[n]}$ denotes $\mathbf{P}_{\mathcal{A}_{[n]}}$. As detailed in Appendix A, the active-set can be detected by

$$\mathcal{A}_{[n+1]} = \left\{ j \;\middle|\; \begin{array}{l} \hat{d}_{\mathcal{H},j}^{[n]} > \left(f(\sigma_z^2, \mathcal{A}_{[n]}) + |\Delta \hat{\mathbf{d}}_{\mathcal{H}}^{[n]}| \right) / |\mathcal{A}_{[n]}|, \\ j \in \mathcal{A}_{[n]} \end{array} \right\}, \tag{12}$$

where $\hat{d}_{\mathcal{H},j}^{[n]}$ denotes the j-th entry in $\hat{\mathbf{d}}_{\mathcal{H}}^{[n]}$ and we define $f(\sigma_z^2, \mathcal{A}) = \sigma_z^2 \mathrm{tr}\{\mathcal{R}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}}^{-1}\}$. The absolute error of the delay profile estimation can also be approximated by $|\Delta \hat{\mathbf{d}}_{\mathcal{H}}^{[n]}| \approx f(\sigma_z^2, \mathcal{A}_{[n]})$. This is because, as shown in Section IV-A, $f(\sigma_z^2, \mathcal{A})$ is identical to the analytical MSE performance of the $\ell 1$ LS technique if the CIR $\mathcal{H}$ to be estimated is exactly supported with the active-set $\mathcal{A}$. Problem (5) is then solved at the step 7. Algorithm 1 obtains a possible estimate $\hat{\mathcal{G}}_{[n+1]}$ via (6) with the detected active-set $\mathcal{A}_{[n+1]}$. However, let $\hat{\mathcal{G}}_{[n+1]} = \mathbf{O}_{WN_T}$ if $\mathcal{A}_{[n+1]} = \emptyset$.

Algorithm 1 utilizes the Bayesian information criterion (BIC) [26] as a stopping tool of the iteration. Suppose that the CIR estimate is described as $\hat{\mathcal{H}} = \hat{\mathcal{G}}_{[n]}\mathbf{P}_{[n]}^{\mathrm{T}}$, the BIC can be defined for the complex matrix normal distribution $\mathcal{L}_{td}(\cdot)$, as

$$\mathtt{BIC}(\hat{\mathcal{G}}_{[n]}) = 2\mathcal{L}_{td}(l, \hat{\mathcal{G}}_{[n]}\mathbf{P}_{[n]}^{\mathrm{T}}) + K_{\mathrm{IC}} \cdot \log(N_{\mathrm{IC}}). \tag{13}$$

The number K_{IC} of free parameters in $\hat{\mathcal{G}}_{[n]}$ is $K_{\mathrm{IC}} = 2N_R|\mathcal{A}_{[n]}|$, where the factor 2 is to represent the freedom of the real and imaginary parts in a complex parameter. The length N_{IC} of input samples denotes $N_{\mathrm{IC}} = \tilde{N}_{td}$ with the input signal length $\tilde{N}_{td} = \tilde{N}_t + \tilde{N}_d$.

It should be noted that Algorithm 1 is a computational complexity-efficient version of the iterative detection/estimation with threshold by "structured" least squared channel estimation (ITDSE) [3]. Algorithm 1 determines thresholds *adaptively* according to the analytical MSE of the $\ell 1$ or $\ell 2$ LS channel estimation. Therefore, as demonstrated in Section V, Algorithm 1 can asymptotically achieve the analytical MSE performance even with the first iteration by setting $N_{\mathrm{AAD}} = 1$ except in a very low SNR regime.

Algorithm 1 The $\ell 1$ LS with the AAD.

Input: $\mathcal{Y}_t, \mathcal{Y}_d, \mathcal{X}_t, \hat{\mathcal{X}}_d$ and N_{AAD}.
1: Compute $\mathbf{R}_{\mathcal{Y}\mathcal{X}}$ (8), $\mathcal{R}_{\mathcal{X}\mathcal{X}}$ (7) and $\hat{\mathbf{\Gamma}}$ (9).
2: Obtain the $\ell 2$ LS estimate $\hat{\mathcal{G}}_{[0]} = \mathtt{mat}_{N_R}\{\hat{\mathbf{g}}_{\mathcal{A}_{[0]}}\}$ by (6) with $\mathcal{A}_{[0]} = \{1, \cdots, WN_T\}$.
3: $\beta(0) = \mathtt{BIC}(\hat{\mathcal{G}}_{[0]})$ by (13).
4: **for** $n = 0$ to $N_{\mathrm{AAD}} - 1$ **do**
5: Update the delay profile estimate $\hat{\mathbf{d}}_{\mathcal{H}}^{[n]}$ by (11).
6: Detect the active-set $\mathcal{A}_{[n+1]}$ by (12).
7: Obtain an estimate $\hat{\mathcal{G}}_{[n+1]} = \mathtt{mat}_{N_R}\{\hat{\mathbf{g}}_{\mathcal{A}_{[n+1]}}\}$ by (6) with $\mathcal{A}_{[n+1]}$.
8: $\beta(n+1) = \mathtt{BIC}(\hat{\mathcal{G}}_{[n+1]})$ by (13).
9: **if** $\beta(n+1) \geq \beta(n)$ **then**
10: Let $n = n - 1$ and terminate the iteration.
11: **end if**
12: **end for**
Output: $\hat{\mathcal{H}}_{\ell 1}^{LS} = \hat{\mathcal{G}}_{[n+1]}\mathbf{P}_{\mathcal{A}_{[n+1]}}^{\mathrm{T}}$.

B. $\ell 1$ Regularized Multi-burst Channel Estimation ($\ell 1$ MB)

1) Problem formulation: $\ell 1$ MB channel estimation is described as an MMSE problem with $\ell 1$ regularization:

$$\hat{\mathcal{H}}_{\ell 1}^{MB}(l) = \underset{\mathcal{H}(l)}{\arg\min} \frac{1}{L_M} \sum_{j \in \mathcal{J}_{L_M}(l)} \{\mathcal{L}_{td}(j, \mathcal{H}(j)) + \lambda(j)\|\mathcal{H}(j)\|_1\}, \tag{14}$$

where a consecutive index set $\mathcal{J}_{L_M}(l)$ is defined as $\{l - L_M + 1, \cdots, l\}$ with a burst-wise sliding window length L_M. To perform the principal component analysis (PCA) correctly, L_M is required to satisfy $L_M \geq W/N_R$.

This problem (14) can be solved by using the same concept as the simplified component technique-LASSO[7] (SCotLASS) [28]. A challenge of SCotLASS-based algorithms is that there is no certain method to determine $\lambda(j)$. We hence relax the problem (14) by introducing an assumption that $\lambda(j)$ can be approximately specified by the following active-sets (15). The

[7] Least absolute shrinkage and selection operator [27].

problem (14) can then be reduced into at most W problems without $\ell 1$ regularization.

Since CIRs can be assumed as the output of a finite impulse response (FIR) filter in general, they attenuate according to the elapse of time. Therefore, we notice that the $\ell 1$ regularization may be replaced by a CIR length constraint. It is hence sufficient to consider W active-sets defined as

$$\mathcal{A}_{[w]} = \bigcup_{k'=0}^{N_T-1} \{(1+k'W) : (w+k'W)\} \tag{15}$$

for $1 \leq w \leq W$. With (15), the problem (14) can be decomposed into at most W problems without $\ell 1$ regularization, as

$$\hat{\mathcal{H}}^{MB}_{[w]}(l) = \underset{\mathcal{H}_{[w]}(l,\mathbf{\Theta})}{\arg\min} \frac{1}{L_M} \sum_{j\in\mathcal{J}_{L_M}(l)} \mathcal{L}_{td}(j, \mathcal{H}_{[w]}(j,\mathbf{\Theta})), \tag{16}$$

where $\mathcal{H}_{[w]}(j,\mathbf{\Theta}) = \mathcal{G}_{[w]}(j,\mathbf{\Theta})\mathbf{P}^{\mathrm{T}}_{[w]}$ with $\mathcal{G}_{[w]}(j,\mathbf{\Theta}) = \mathcal{H}(j,\mathbf{\Theta})|_{\mathcal{A}_{[w]}}$ and $\mathbf{P}_{[w]} = \mathbf{P}_{\mathcal{A}_{[w]}}$. The parameter vector $\mathbf{\Theta}$ is defined as follows: the k-th TX-stream's CIR in $\mathcal{H}(j,\mathbf{\Theta}_{[w]}) = [\mathbf{H}_1(l,\theta_1),\cdots,\mathbf{H}_{N_T}(l,\theta_{N_T})]$ can be described as

$$\mathbf{H}_k(l,\theta_k) \;=\; \mathbf{B}_k(l)\mathbf{U}^{\mathrm{H}}_k, \tag{17}$$

when the CIR follows the subspace channel model assumption [9]. The parameter $\mathbf{\Theta}$ describes the CIR models (17) for N_T Tx streams in a vector as $\mathbf{\Theta} = [\theta_1^{\mathrm{T}},\cdots,\theta_{N_T}^{\mathrm{T}}]^T$, where $\theta_k = [\theta^{\mathrm{T}}_{\mathbf{B},k}, \mathtt{vec}\{\mathbf{U}_k\}^{\mathrm{T}}]^{\mathrm{T}}$. The subvector $\theta_{\mathbf{B},k}$ denotes $\theta_{\mathbf{B},k} = \left[\mathtt{vec}\{\mathbf{B}_k(j_1)\}^{\mathrm{T}},\cdots,\mathtt{vec}\{\mathbf{B}_k(j_{L_M})\}^{\mathrm{T}}\right]^{\mathrm{T}}$, for $j_n \in \mathcal{J}_{L_M}$. It should be noticed that the $N_R \times r_k$ matrix $\mathbf{B}_k(l)$ is burst-dependent. However, the $W \times r_k$ matrix $\mathbf{U}_k$ is independent of the burst timing since it represents a temporally invariant FIR filter. The parameter r_k denotes the rank of the temporal covariance matrix $\mathbb{K}^{L_M}_l[\mathbf{H}_k(l)]$, where the operation $\mathbb{K}^{L_M}_l[\cdot]$ denotes $\mathbb{K}^{L_M}_l[\mathbf{A}(l)] = \frac{1}{L_M}\sum_{j\in\mathcal{J}_{L_M}(l)}\mathbf{A}(j)^{\mathrm{H}}\mathbf{A}(j)$ for a matrix sequence $\mathbf{A}(l)$ and a sliding window $\mathcal{J}_{L_M}(l)$.

2) $\ell 1$ MB algorithm: By Appendix A in [10], the problems (16) is equivalent to minimizing $\Psi_{[w]}(l,\mathbf{\Theta}) = \sum_{k=1}^{N_T}\Psi_{[w],k}(l,\theta_k)$ with

$$\Psi_{[w],k}(l,\theta_k) = \frac{1}{L_M}\sum_{j\in\mathcal{J}_{L_M}(l)} \left\| \hat{\bar{\mathbf{g}}}^{LS}_{[w],k}(j) - \tilde{\mathbf{g}}_{[w],k}(j,\theta_k)\right\|^2, \tag{18}$$

where we define, for the noise whitening,

$$\begin{aligned}\hat{\bar{\mathbf{g}}}^{LS}_{[w],k}(j) \quad &= \bar{\mathbf{Q}}_{[w],kk}\cdot\hat{\mathbf{g}}^{LS}_{[w],k}(j) \\ &+ \sum_{i=k+1}^{N_T}\bar{\mathbf{Q}}_{[w],ki}\left\{\hat{\mathbf{g}}^{LS}_{[w],i}(j) - \mathbf{g}_{[w],i}(j,\theta_i)\right\} \end{aligned} \tag{19}$$

$$\tilde{\mathbf{g}}_{[w],k}(j,\theta_k) = \bar{\mathbf{Q}}_{[w],kk}\cdot\mathbf{g}_{[w],k}(j,\theta_k) \tag{20}$$

with a length wN_R CIR vector $\mathbf{g}_{[w],k}(j,\theta_k) = \mathtt{vec}\{\mathbf{H}_k(j,\theta_k)|_{1:w}\}$ for the k-th Tx stream's and its LS estimate $\hat{\mathbf{g}}^{LS}_{[w],k}(j)$. The $wN_R \times wN_R$ matrix $\bar{\mathbf{Q}}_{[w],ki}$ is the (k,i)-th block matrix of the Cholesky decomposition for $\bar{\mathcal{R}}_{\mathbf{\Phi\Phi}_{[w]}}$ (21). It should be noted that (18) is based on the approximation[8] that

$$\bar{\mathcal{R}}_{\mathbf{\Phi\Phi}_{[w]}} \triangleq \mathbb{E}\left[\mathcal{P}^{\mathrm{T}}_{[w]}\mathcal{R}_{\mathcal{X}\mathcal{X}}(j)\mathcal{P}_{[w]}\right] \approx \mathcal{P}^{\mathrm{T}}_{[w]}\mathcal{R}_{\mathcal{X}\mathcal{X}}(j)\mathcal{P}_{[w]} \tag{21}$$

for $\forall j \in \mathcal{J}_{L_M}(l)$, where $\mathcal{P}_{[w]} = \mathbf{P}_{\mathcal{A}_{[w]}} \otimes \mathbf{I}_{N_R}$. Therefore, the w-th active-set has to be independent of the burst timing j, such as (15).

As shown in Section III in [10], the minimization problems of (18) are solvable if we reduce them by descending order $k = N_T,\cdots,1$. Moreover, by following Section IV-C in [12], the solution $\hat{\mathbf{g}}^{MB}_{[w],k}(l)$ that minimizes (18) can be obtained by the PCA for the covariance matrix $\mathbb{K}^{L_M}_l[\mathtt{mat}_{N_R}\{\hat{\bar{\mathbf{g}}}^{LS}_{[w],k}(l)\}]$, where approximations $\mathbf{g}_{[w],i}(l,\theta_i) \approx \hat{\mathbf{g}}^{MB}_{[w],i}(l)$ for $i > k$ are used in (19). Correspondingly, the solution to (16) is described as $\hat{\mathcal{H}}^{MB}_{[w]}(l) = [\hat{\mathbf{G}}^{MB}_{[w],1}(l),\cdots,\hat{\mathbf{G}}^{MB}_{[w],N_T}(l)]\mathbf{P}^{\mathrm{T}}_{[w]}$ with $\hat{\mathbf{G}}^{MB}_{[w],k}(l) = \mathtt{mat}_{N_R}\{\hat{\mathbf{g}}^{MB}_{[w],k}(l)\}$. We finally choose the best solution to (14) from the W possible estimates as $\hat{\mathcal{H}}^{MB}_{\ell 1}(l) = \hat{\mathcal{H}}^{MB}_{[\hat{w}]}(l)$. The optimal CIR length $\hat{w}$ may be determined by Akaike information criterion (AIC) [29]: $\mathtt{AIC}(\hat{\mathcal{H}}^{MB}_{[w]}) = 2\mathcal{L}_{td}(l,\hat{\mathcal{H}}^{MB}_{[w]}) + 2K_{\mathrm{IC}}$, where the number of free parameters is modified as $K_{\mathrm{IC}} = 2\sum_{k=1}^{N_T}N_R r^{[w]}_k$ so that it describes the number of burst-dependent parameters in the CIR model (17). The rank $r^{[w]}_k$ of the temporal subspace is obtained together with the estimate vector $\hat{\mathbf{g}}^{MB}_{[w],k}(l)$ by the PCA, as shown in [10], [12].

C. Hybrid Channel Estimation

1) Problem statement: Later in Section V-D2, it is shown that the $\ell 1$ MB channel estimation can improve the tracking error problem. Nevertheless, as discussed in Section III-D, the $\ell 1$ MB channel estimation requires a higher complexity order than the ordinary $\ell 2$ MB channel estimation. We thereby propose a new hybrid channel estimation algorithm to improve robustness of the estimate with reasonable complexity.

2) Hybrid algorithm: In summary, the new hybrid algorithm performs the $\ell 1$ LS and the ordinary $\ell 2$ MB channel estimation[9] simultaneously, then selects *better* estimate under the Bayesian information criterion.

Specifically, the hybrid technique is shown in Algorithm 2, where the counter L_m is initialized to 0 before starting the hybrid channel estimation. The counter L_m is used to define the sliding window $\mathcal{J}_{L_m}(l)$ in the MMSE problem (14). The guard constant L_G in Algorithm 2 is set at $\lceil W/N_R \rceil$, because the PCA used in the $\ell 2$ MB channel estimation is numerically unstable for $L_m < \lceil W/N_R \rceil$, where $\lceil\cdot\rceil$ denotes the ceiling function. At the first step, the $\ell 1$ LS channel estimation is performed. The $\ell 2$ MB channel estimate can then be obtained efficiently by reusing the $\ell 2$ LS estimate $\hat{\mathcal{H}}_{[0]}$ computed in Algorithm 1.

The better estimate between the two possible solutions is then determined by the steps 4 to 15. At the step 4, Algorithm 2 monitors the tracking error by comparing the BIC of the $\ell 1$ LS channel estimate with that of the $\ell 2$ MB channel estimate. The tracking error can be detected based on a property that $\mathtt{BIC}(\hat{\mathcal{H}}^{LS}_{\ell 1}(l)) > \mathtt{BIC}(\hat{\mathcal{H}}^{MB}_{\ell 2}(l))$ is satisfied so far as CIRs follow the subspace channel model assumption. In the case the tracking error is detected, Algorithm 2 selects the channel estimate $\hat{\mathcal{H}}^{LS}_{\ell 1}$ as the output $\hat{\mathcal{H}}^{HB}$ of the hybrid estimation.

[8]The conventional $\ell 2$ MB techniques [9]–[12] also assume (21) with $\mathcal{P}_{[w]} = \mathbf{I}_{WN_TN_R}$.

[9]The $\ell 2$ MB channel estimation is formulated by (14) with $\lambda(j) = 0$.

Furthermore, at the step 7, Algorithm 2 resets the counter L_m when the $\ell 2$ MB channel estimation performed for more than L_G bursts. The counter reset is performed so that the covariance matrix in the PCA is adjusted to a change of channel models quickly. On the other hand, if the tracking error is not detected, Algorithm 2 selects the channel estimate $\hat{\mathcal{H}}_{\ell 2}^{MB}$ at the step 13. However, Algorithm 2 selects $\hat{\mathcal{H}}_{\ell 1}^{LS}$ at the step 11 if the counter L_m is less than L_G. This is because the channel estimate $\hat{\mathcal{H}}_{\ell 2}^{MB}$ is not accurate enough when $L_m < L_G$.

Algorithm 2 Hybrid channel estimation at the burst timing l

1: Perform the $\ell 1$ LS (2) and obtain $\hat{\mathcal{H}}_{\ell 1}^{LS}(l)$.
2: Update the counter as $L_m = \min(L_m + 1, L_M)$ per a burst.
3: Perform the $\ell 2$ MB (14) with the sliding window $\mathcal{J}_{L_m}(l)$ and $\lambda(j) = 0$. Obtain $\hat{\mathcal{H}}_{\ell 2}^{MB}(l)$.
4: **if** $\mathtt{BIC}(\hat{\mathcal{H}}_{\ell 1}^{LS}(l)) < \mathtt{BIC}(\hat{\mathcal{H}}_{\ell 2}^{MB}(l))$ **then**
5: $\quad \hat{\mathcal{H}}^{HB}(l) = \hat{\mathcal{H}}_{\ell 1}^{LS}(l)$
6: $\quad$ **if** $L_m \geq L_G$ **then**
7: $\qquad L_m = 0$
8: $\quad$ **end if**
9: **else**
10: $\quad$ **if** $L_m < L_G$ **then**
11: $\qquad \hat{\mathcal{H}}^{HB}(l) = \hat{\mathcal{H}}_{\ell 1}^{LS}(l)$
12: $\quad$ **else**
13: $\qquad \hat{\mathcal{H}}^{HB}(l) = \hat{\mathcal{H}}_{\ell 2}^{MB}(l)$
14: $\quad$ **end if**
15: **end if**

D. Computational Complexity Order

The computational complexity orders $\mathcal{O}(\cdot)$ required for the channel estimation techniques investigated in this paper are summarized in Table I. The complexity order required for the proposed hybrid algorithm is equivalent to the $\ell 2$ MB channel estimation when $N_{\mathrm{AAD}} = 1$. However, the $\ell 1$ MB channel estimation requires a larger complexity order than the $\ell 2$ MB by $\mathcal{O}(W^4 N_T^3 N_R^3)$.

1) The $\ell 1$ LS: The computational complexity orders required for each step in Algorithm 1 and its details are shown in Tables II(a) and (b), respectively. For example, the step 2 in Algorithm 1 performs an $|\mathcal{A}_{[0]}|N_R \times |\mathcal{A}_{[0]}|N_R$ matrix inversion and a matrix-vector product, the size of which is $[|\mathcal{A}_{[0]}|N_R \times |\mathcal{A}_{[0]}|N_R] \times [|\mathcal{A}_{[0]}|N_R \times 1]$ with $|\mathcal{A}_{[0]}| = WN_T$. The complexity order needed for these operations is shown in the row (iv) of Table II(b). It is, however, dominated by $\mathcal{O}(\{|\mathcal{A}_{[0]}|N_R\}^3) = \mathcal{O}(W^3 N_T^3 N_R^3)$, where we assume that an $M \times M$ matrix inverse requires the complexity order $\mathcal{O}(M^3)$ [30].

As shown in Table II(a), the complexity order needed for the steps 1 to 3 is dominated by $\mathcal{O}(W^2 N_R^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3)$. This is because $W^2 N_T^2 > W N_T N_R > N_R^2$ is satisfied in the assumed frequency selective fading channel, the CIR length of which is $W \gg N_R \geq N_T$. Moreover, the equivalent negative log-likelihood functions in (13) may be calculated by using the following equations:

$$\mathcal{L}_t(\hat{\mathcal{G}}_{\mathcal{A}}) = \frac{1}{\sigma_z^2}\|\mathcal{Y}_t - \hat{\mathcal{G}}_{\mathcal{A}}\mathbf{\Phi}_{t,\mathcal{A}}\|_F^2, \tag{22}$$

$$\mathcal{L}_d(\hat{\mathcal{G}}_{\mathcal{A}}) = \frac{1}{\sigma_z^2}\|\mathcal{Y}_d - \hat{\mathcal{G}}_{\mathcal{A}}\hat{\mathbf{\Phi}}_{d,\mathcal{A}}\|_{\mathbf{\Gamma}}^2, \tag{23}$$

where the $N_R \times |\mathcal{A}|$ CIR estimate matrix $\hat{\mathcal{G}}_{\mathcal{A}}$ is obtained via (6) for an active-set $\mathcal{A}$.

The complexity order required for the steps 5 to 11 is dominated by that of the steps 7 and 8. It should be noticed that, at the step 6, the matrix inverse $\mathbf{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}_{[n]}}}^{-1}$ in $f(\sigma_z^2, \mathcal{A})$ is already computed in the previous iteration. Furthermore, the matrix inverse $\mathbf{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}_{[n+1]}}}^{-1}$ can be efficiently updated from $\mathbf{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}_{[n]}}}^{-1}$. Specifically, the complexity order needed for the step 7 is dominated by $\mathcal{O}(\{|\mathcal{A}_{[n]}|^2|\Delta\mathcal{A}_{[n+1]}| + |\Delta\mathcal{A}_{[n+1]}|^3\}N_R^3)$, where $\Delta\mathcal{A}_{[n+1]} = \mathcal{A}_{[n]} \backslash \mathcal{A}_{[n+1]}$. This is because, as shown in [31], if the matrix inverse of an $M \times M$ Hermitian matrix is known, the complexity order needed to compute the matrix inverse of its arbitrary rank-1-downsized submatrix is $\mathcal{O}(M^2)$. By extending the algorithm in [31] straightforwardly, the matrix inverse of its arbitrary rank-N-downsized submatrix[10] can be computed with the complexity order $\mathcal{O}(M^2 N + N^3)$.

Algorithm 1 performs at most $\max(N_{\mathrm{AAD}}) = WN_T$ iterations since $WN_T \geq |\mathcal{A}_{[n]}| \geq |\mathcal{A}_{[n+1]}| \geq 0$ is guaranteed by (12). The complexity is, hence, maximized when WN_T iterations are performed without the termination at the step 10. This case happens when the active-sets are updated so that the cardinality changes $|\mathcal{A}_{[n]}| = WN_T - n$ at the n-th iteration. Therefore, the maximum complexity order required for Algorithm 1 becomes $\mathcal{O}(\{W^2 N_R^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3\} + \sum_{m=1}^{WN_T}\{m^2 N_R^3 + N_R^3 + ((m-1)N_R + N_R^2)\tilde{N}_{td}\}) = \mathcal{O}(W^2(N_T^2 N_R + N_R^2)\tilde{N}_{td} + W^3 N_T^3 N_R^3)$, where $m = WN_T - n$ is used. Especially for $N_{AAD} = 1$, the complexity order is at most $\mathcal{O}(W^2 N_R^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3)$ due to $|\mathcal{A}_{[n]}| \leq WN_T$.

2) The $\ell 2$ LS: The $\ell 2$ LS channel estimation requires the complexity order of at most $\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3)$ since it is identical to the steps 1 and 2 in Algorithm 1.

3) The $\ell 1$ MB: The $\ell 1$ MB algorithm performs a set of operations, which are CIR length-shrunk $\ell 2$ LS channel estimation (6) and the PCA, for at most W possible solutions. The complexity order required for obtaining the W LS channel estimates is, however, equivalent to that of the ordinary $\ell 2$ LS channel estimation. This is because, in (6), the matrices $\mathcal{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}$ and $\mathbf{R}_{\mathcal{Y}\mathbf{\Phi}_{\mathcal{A}}}$ are the submatrices of $\mathcal{R}_{\mathcal{X}\mathcal{X}}$ and $\mathcal{R}_{\mathcal{Y}\mathcal{X}}$, respectively. Furthermore, by using the matrix inverse downsizing or upsizing algorithm [31], the complexity order needed for the W matrix inverses $\mathcal{R}_{\mathbf{\Phi\Phi}_{[w]}}^{-1}$, $1 \leq w \leq W$, is equivalent to that of the single matrix inverse $\mathcal{R}_{\mathcal{X}\mathcal{X}}^{-1}$. Therefore, the complexity order required for obtaining the W LS channel estimates is dominated by $\mathcal{O}(W^3 N_T^3 N_R^3) = \mathcal{O}(W^3 N_T^3 N_R^3 + \sum_{w=1}^{W} w^2 N_T^2 N_R^2)$, where the summation term describes the

[10] Let $\mathbf{R}_n$ denote $\mathbf{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}_{[n]}}}$ after relevant permutations so that $\mathbf{R}_n = \begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{H}} \\ \mathbf{B} & \mathbf{R}_{n+1} \end{bmatrix}$. $\mathbf{R}_n^{-1} = \begin{bmatrix} \mathbf{E} & \mathbf{F}^{\mathrm{H}} \\ \mathbf{F} & \mathbf{G} \end{bmatrix} \in \mathbb{C}^{M \times M} \Rightarrow \mathbf{R}_{n+1}^{-1} = \mathbf{G} - \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^{\mathrm{H}}$, where the sizes of submatrices $\mathbf{E}$, $\mathbf{F}$ and $\mathbf{G}$ are $N \times N$, $(M - N) \times N$ and $(M - N) \times (M - N)$, respectively.

TABLE I
COMPUTATIONAL COMPLEXITY ORDERS FOR CHANNEL ESTIMATION ALGORITHMS

Algorithm	Computational complexity order	N_{AAD}
$\ell1$ LS	$\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3)$	1
	$\mathcal{O}(W^2 (N_T^2 N_R + N_R^2) \tilde{N}_{td} + W^3 N_T^3 N_R^3)$	WN_T
$\ell2$ LS	$\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3)$	
$\ell1$ MB	$\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^4 N_T^3 N_R^3)$	
$\ell2$ MB	$\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3)$	
Hybrid	$\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3)$	1
	$\mathcal{O}(W^2 (N_T^2 N_R + N_R^2) \tilde{N}_{td} + W^3 N_T^3 N_R^3)$	WN_T

complexity for matrix-vector products according to the row (iv) of Table II(b).

The complexity analysis of the PCA for the w-th possible solution is summarized as follows: the complexity orders required for the Cholesky decomposition[11] of $\bar{\mathcal{R}}_{\mathbf{\Phi\Phi}_{[w]}}$, the noise whitening (19) and the SVD[12] are $\mathcal{O}(w^3 N_T^3 N_R^3)$, $\mathcal{O}(w^2 N_T^2 N_R^2)$, and $\mathcal{O}(w^3 N_T)$, respectively, in total for N_T Tx streams. Consequently, the complexity order needed for the w-th PCA is dominated by $\mathcal{O}(w^3 N_T^3 N_R^3)$. The complexity order required for the $\ell1$ MB is, thereby, dominated by $\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^4 N_T^3 N_R^3) = \mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3) + \mathcal{O}(\sum_{w=1}^{W} w^3 N_T^3 N_R^3)$.

4) The $\ell2$ MB: The $\ell2$ MB technique performs the above-mentioned set of operations for the W-th possible solution, only once. Hence, the complexity order needed for the $\ell2$ MB channel estimation is dominated by $\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3) = \mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3) + \mathcal{O}(\sum_{w=W}^{W} w^3 N_T^3 N_R^3)$.

5) The hybrid algorithm: The hybrid algorithm performs the $\ell1$ LS and $\ell2$ MB techniques at a time. However, its complexity order is equivalent to that of the $\ell1$ LS, since $\mathcal{O}(\{W^2 (N_T^2 N_R + N_R^2) \tilde{N}_{td} + W^3 N_T^3 N_R^3\} + \{W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3\}) = \mathcal{O}(W^2 (N_T^2 N_R + N_R^2) \tilde{N}_{td} + W^3 N_T^3 N_R^3)$. The complexity order needed for the BIC of the $\ell2$ MB channel estimate is $\mathcal{O}((W N_T N_R + N_R^2) \tilde{N}_{td})$ and hence it is very minor. Especially for $N_{\text{AAD}} = 1$, the complexity *order* required for the hybrid algorithm is the same as that of the $\ell2$ MB technique, although the number of operations are increased slightly.

IV. PERFORMANCE ANALYSIS

A. MSE performance of the $\ell1$ LS

MSE performance of the $\ell1$ LS channel estimation can be given by the following (24) with the optimal active-set (30).

[11] We assume that the Cholesky decomposition for an $M \times M$ matrix requires the complexity order $\mathcal{O}(M^3)$ [30].

[12] As shown in [12], SVD for the $w \times w$ covariance matrices $\mathbb{K}_l^{L_M}[\hat{\hat{\mathbf{G}}}_{[w],k}^{LS}(l)]$ is performed to find the principal components of the CIR for the k-th TX stream, where $\hat{\hat{\mathbf{G}}}_{[w],k}^{LS}(l) = \mathtt{mat}_{N_R}\{\hat{\hat{\mathbf{g}}}_{[w],k}^{LS}(l)\}$. Hence, the complexity order becomes $\mathcal{O}(w^3 N_T)$ for N_T Tx streams, by assuming that an SVD operation for an $M \times M$ matrix needs $\mathcal{O}(M^3)$ [30]. Note that the complexity order needed for the covariance matrix $\mathbb{K}_l^{L_M}[\hat{\hat{\mathbf{G}}}_{[w],k}^{LS}(l)]$ is minor since it can be updated recursively: $\mathfrak{K}_l^{L_M} = (L_M \cdot \mathfrak{K}_{l-1}^{L_M} + \mathfrak{K}_l^1 - \mathfrak{K}_{l-L_M}^1)/L_M$, where we denote $\mathfrak{K}_l^{L_M} = \mathbb{K}_l^{L_M}[\hat{\hat{\mathbf{G}}}_{[w],k}^{LS}(l)]$ by omitting the subscripts w and k.

1) Analytical MSE: For a given $\mathcal{A}$, an $N_R \times W N_T$ sparse channel estimate matrix $\hat{\mathcal{H}}_{\mathcal{A}} = \hat{\mathcal{G}}_{\mathcal{A}} \mathbf{P}_{\mathcal{A}}^{\mathrm{T}}$ is vectorized as $\mathtt{vec}\{\hat{\mathcal{H}}_{\mathcal{A}}\} = \mathcal{P}_{\mathcal{A}} \cdot \hat{\mathbf{g}}_{\mathcal{A}}$, where the $N_R \times |\mathcal{A}|$ compressed channel estimate matrix $\hat{\mathcal{G}}_{\mathcal{A}}$ is obtained via the vectorized channel estimate $\hat{\mathbf{g}}_{\mathcal{A}}$ (6). The MSE of the $\ell1$ LS channel estimate can, thereby, be reduced to

$$\begin{aligned} \mathrm{MSE}(\hat{\mathcal{H}}_{\ell1}^{LS}, \sigma_z^2, \mathcal{A}) &= \mathbb{E}\left[\|\mathtt{vec}\{\hat{\mathcal{H}}_{\ell1}^{LS}(l) - \mathcal{H}(l)\}\|^2\right] \\ &= \sigma_z^2 \operatorname{tr}\left\{\mathbb{E}[\mathcal{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}^{-1}(l)]\right\} + \mathfrak{E}(\mathcal{A}), \end{aligned} \quad (24)$$

where we define $\mathfrak{E}(\mathcal{A}) = \mathbb{E}\left[\|\mathfrak{B}(\mathcal{A}, l) \cdot \mathtt{vec}\{\mathcal{H}_{\mathcal{A}}^{\perp}(l)\}\|^2\right]$ with $\mathfrak{B}(\mathcal{A}, l) = \mathcal{P}_{\mathcal{A}} \mathcal{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}^{-1}(l) \mathcal{P}_{\mathcal{A}}^{\mathrm{T}} \mathcal{R}_{\mathcal{XX}}(l) - \mathbf{I}_{W N_T N_R}$. We note that, when $\mathcal{A} = \emptyset$, the MSE of $\hat{\mathcal{H}}_{\ell1}^{LS}$ becomes $\mathfrak{E}(\mathcal{A}) = \mathbb{E}\left[\|\mathcal{H}(l)\|^2\right]$. The CIR unsupported with the active-set is denoted by $\mathcal{H}_{\mathcal{A}}^{\perp}(l) = \mathcal{H}(l) \mathbf{J}_{\mathcal{A}}^{\perp}$, where $\mathbf{J}_{\mathcal{A}}^{\perp} = \mathbf{I}_{W N_T} - \mathbf{J}_{\mathcal{A}}$ with $\mathbf{J}_{\mathcal{A}} = \mathbf{P}_{\mathcal{A}} \mathbf{P}_{\mathcal{A}}^{\mathrm{T}}$.

2) Optimal active-set: For an active-set $\mathcal{A}$, denote the MSE residual $\Delta_{\ell1\ell2}^{LS}(\mathcal{A})$, as

$$\Delta_{\ell1\ell2}^{LS}(\mathcal{A}) = \mathrm{MSE}(\hat{\mathcal{H}}_{\ell1}^{LS}, \sigma_z^2, \mathcal{A}) - \mathrm{MSE}(\hat{\mathcal{H}}_{\ell2}^{LS}, \sigma_z^2), \quad (25)$$

where $\mathrm{MSE}(\hat{\mathcal{H}}_{\ell2}^{LS}, \sigma_z^2) = \sigma_z^2 \mathrm{tr}\{\mathbb{E}[\mathcal{R}_{\mathcal{XX}}^{-1}(l)]\}$. The optimal active-set which minimizes the MSE performance (24) may be reduced via the minimization of (25). This is because $\mathrm{MSE}(\hat{\mathcal{H}}_{\ell2}^{LS}, \sigma_z^2)$ is independent of $\mathcal{A}$. However, we notice that, for any $\mathcal{A}$,

$$\begin{aligned} &\mathrm{MSE}(\hat{\mathcal{H}}_{\ell2}^{LS}, \sigma_z^2) = \\ &\quad \sigma_z^2 \left[\operatorname{tr}\left\{\mathbb{E}\left[\mathfrak{J}_{\mathcal{A}} \mathcal{R}_{\mathcal{XX}}^{-1}(l)\right]\right\} + \operatorname{tr}\left\{\mathbb{E}\left[\mathfrak{J}_{\mathcal{A}}^{\perp} \mathcal{R}_{\mathcal{XX}}^{-1}(l)\right]\right\}\right], \end{aligned} \quad (26)$$

where we denote $\mathfrak{J}_{\mathcal{A}} = \mathbf{J}_{\mathcal{A}} \otimes \mathbf{I}_{N_R}$ and $\mathfrak{J}_{\mathcal{A}}^{\perp} = \mathbf{I}_{W N_T N_R} - \mathfrak{J}_{\mathcal{A}}$. By Theorem 7.7.8 in [32],

$$\mathcal{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}^{-1}(l) \preceq \mathcal{P}_{\mathcal{A}}^{\mathrm{T}} \cdot \mathcal{R}_{\mathcal{XX}}^{-1}(l) \cdot \mathcal{P}_{\mathcal{A}} \quad (27)$$

is satisfied for $\forall \mathcal{A}$, where $\mathbf{A} \preceq \mathbf{B}$ denotes that a residual $\mathbf{B} - \mathbf{A}$ is a positive semidefinite matrix. We, hence, have

$$\mathbb{E}\left[\operatorname{tr}\left\{\mathcal{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}^{-1}(l)\right\} - \operatorname{tr}\left\{\mathfrak{J}_{\mathcal{A}} \mathcal{R}_{\mathcal{XX}}^{-1}(l)\right\}\right] \leq 0. \quad (28)$$

Substituting (24), (26) and (28) into (25) yeilds

$$\Delta_{\ell1\ell2}^{LS}(\mathcal{A}) \leq \mathfrak{E}(\mathcal{A}) - \sigma_z^2 \operatorname{tr}\left\{\mathbb{E}\left[\mathfrak{J}_{\mathcal{A}}^{\perp} \mathcal{R}_{\mathcal{XX}}^{-1}(l)\right]\right\}. \quad (29)$$

The MSE performance (24) is, thereby, minimized with the optimal active-set $\mathcal{A}^*$ given by

$$\mathcal{A}^* = \arg\min_{\mathcal{A}} \left[\mathfrak{E}(\mathcal{A}) - \sigma_z^2 \operatorname{tr}\left\{\mathbb{E}\left[\mathfrak{J}_{\mathcal{A}}^{\perp} \mathcal{R}_{\mathcal{XX}}^{-1}(l)\right]\right\}\right]. \quad (30)$$

Obviously, the problem (30) is a combinatorial optimization. The solution to (30) can be found from all possible $\sum_{k=0}^{W N_T} \binom{W N_T}{k}$ active-sets if the delay profile $\mathbb{E}[\mathbf{d}_{\mathcal{H}}(l)] = \mathbb{E}[\mathtt{diag}\{\mathcal{H}^{\mathrm{H}}(l) \cdot \mathcal{H}(l)\}]$ is known. The operation $\binom{n}{k}$ denotes the binomial coefficient.

B. MSE performance bound of the $\ell1$ MB

Since the $\ell2$ MB channel estimation techniques asymptotically achieve the CRB [9]–[12], MSE performance of the $\ell1$ MB algorithm is discussed through the CRB.

TABLE II
COMPLEXITY ORDER IN ALGORITHM 1

(a) COMPLEXITY ORDER FOR EACH STEP IN ALGORITHM 1

Step	Computational complexity order	Details
1:	$\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W N_T N_R \tilde{N}_{td} + N_R^3)$	(i, ii, iii)
2:	$\mathcal{O}(W^3 N_T^3 N_R^3)$	(iv)
3:	$\mathcal{O}(\ (W N_T N_R + N_R^2)\tilde{N}_{td}\ )$	(v, vi)
5:	$\mathcal{O}(\lvert\mathcal{A}_{[n]}\rvert^2 N_R)$	(vii)
6:	$\mathcal{O}(W N_T)$	
7:	$\mathcal{O}(\{\lvert\mathcal{A}_{[n]}\rvert^2 \lvert\Delta\mathcal{A}_{[n+1]}\rvert + \lvert\Delta\mathcal{A}_{[n+1]}\rvert^3\} N_R^3)$	(iv), [31]
8:	$\mathcal{O}(\ (\lvert\mathcal{A}_{[n+1]}\rvert\ N_R + N_R^2)\tilde{N}_{td}\ )$	(v, vi)

(b) DETAILS IN TABLE II(a)

	Symbol	Eqn.	Computational complexity order
(i)	$\hat{\mathbf{\Gamma}}$	(9)	$\mathcal{O}(W N_T N_R^2 + N_R^3)$
(ii)	$\mathcal{R}_{\mathcal{X}\mathcal{X}}$	(7)	$\mathcal{O}(W^2 N_T^2 \tilde{N}_{td})$
(iii)	$\mathbf{R}_{\mathcal{Y}\mathcal{X}}$	(8)	$\mathcal{O}(W N_T N_R \tilde{N}_{td})$
(iv)	$\hat{\mathbf{g}}_{\mathcal{A}}$	(6)	$\mathcal{O}(\lvert\mathcal{A}\rvert^3 N_R^3 + \lvert\mathcal{A}\rvert^2 N_R^2)$
(v)	$\mathcal{L}_t(\hat{\mathcal{G}}_{\mathcal{A}})$	(22)	$\mathcal{O}(\{\lvert\mathcal{A}\rvert N_R + N_R^2\}\tilde{N}_t)$
(vi)	$\mathcal{L}_d(\hat{\mathcal{G}}_{\mathcal{A}})$	(23)	$\mathcal{O}(\{\lvert\mathcal{A}\rvert N_R + N_R^2\}\tilde{N}_d + N_R^3)$
(vii)	$\hat{\mathcal{G}}_{\mathcal{A}}^{\mathrm{H}}\hat{\mathcal{G}}_{\mathcal{A}}$	(11)	$\mathcal{O}(\lvert\mathcal{A}\rvert^2 N_R)$

1) Definition of Unbiased- and Adaptive-Subspace: We define terminologies *unbiased*- and *adaptive*-subspace which are used to describe the performance bound of the new channel estimation algorithms. Note that the reference signal length[13] $\bar{N}$ is referred as to $\mathtt{tr}\{\mathbf{X}\mathbf{X}^{\mathrm{H}}\}/M$, where the $M \times M$ matrix $\mathbf{X}$ denotes the Toeplitz matrix used in a channel estimator.

Definition 1 (Unbiased-subspace). *An unbiased-subspace for CIRs $\mathbf{G}_k(j,w) = \mathbf{H}_k(j)|_{1:w}, \forall j \in \mathcal{J}_L(l)$ with $L \geq w/N_R$, is a subspace spanned by column vectors of $\mathbf{U}_k(l,w)|_{1:r_k}$, where the unitary matrix $\mathbf{U}_k(l,w)$ can be obtained from $\mathbf{U}_k(l,w) \cdot \mathbf{\Lambda}_k(l,w) \cdot \mathbf{U}_k(l,w)^{\mathrm{H}} = \mathtt{svd}\left\{\mathbb{K}_l^L[\mathbf{G}_k(l,w)]\right\}$ and r_k is the path number of a channel model assumed for the k-th TX stream.*

Definition 2 (Adaptive-subspace). *An adaptive-subspace for CIRs $\mathbf{G}_k(j,w)$ is spanned by column vectors of $\mathbf{U}_k(l,w)|_{1:r^a_{w,k}(\sigma_z^2,\bar{N})}$, where the parameter $r^a_{w,k}(\sigma_z^2,\bar{N})$ is defined as*

$$r^a_{w,k}(\sigma_z^2,\bar{N}) = \sum_{i=1}^{r_k} 1\left\{\lambda_k^i(l,w) > N_R\sigma_z^2/\bar{N}\right\} \quad (31)$$

for the ideally uncorrelated reference signal, the length of which is $\bar{N}$. The i-th largest singular value $\lambda_k^i(l,w)$ is obtained from $\mathbf{\Lambda}_k(l,w)$. The indicator function $1\{\mathcal{B}\}$ takes 1 if its argument Boolean $\mathcal{B}$ is true, otherwise 0.

It should be noted that the adaptive-subspace is an approximation of the unbiased-subspace in the noisy covariance matrix $\mathbb{K}_l^L[\mathbf{G}_k(l,w)] + (N_R\sigma_z^2/\bar{N})\mathbf{I}_w$. We define another terminology *complemental*-subspace as a subspace spanned by the column vectors of $\mathbf{U}_k(l,w)|_{r^a_{w,k}(\sigma_z^2,\bar{N})+1:r_k}$.

2) CRB: The CRB for an unbiased estimator in a MIMO channel can be derived as a sum of CRBs over N_T TX streams in SIMO channels or their vectorized SISO versions. This is because (19) is independent of θ_k. Therefore, by utilizing the CRB of SISO channel estimation in [10], the CRB of MIMO channel estimation can be described as $\mathrm{CRB}_{\bar{N},w}(\sigma_z^2,\mathbf{r}) = \mathrm{CRB}^{\mathcal{Z}}_{\bar{N}}(\sigma_z^2,\mathbf{r}) + \mathrm{CRB}^{\Pi}_{\bar{N},w}(\sigma_z^2,\mathbf{r})$, where we denote the unbiased-ranks of CIRs by a vector as $\mathbf{r} = [r_1, ..., r_{N_T}]^{\mathrm{T}}$ and define

$$\mathrm{CRB}^{\mathcal{Z}}_{\bar{N}}(\sigma_z^2,\mathbf{r}) = \sum_{k=1}^{N_T} \frac{N_R\sigma_z^2 r_k}{\bar{N}} \quad (32)$$

$$\mathrm{CRB}^{\Pi}_{\bar{N},w}(\sigma_z^2,\mathbf{r}) = \sum_{k=1}^{N_T} \frac{\sigma_z^2}{L_M\bar{N}}(w\, r_k - r_k^2) \quad (33)$$

under the assumption that the length $\bar{N}$ ideally uncorrelated sequence is used. The $\ell1$ MB channel estimation can decrease the projection error (33) by assuming a shorter CIR length w than W, so long as it does not distort the original r_k paths to perform the *unbiased* channel estimation. However, it should be noticed that (32) is independent of w. Therefore, the $\ell1$ MB can improve the projection error, nevertheless, it does not improve asymptotic MSE performance (32) when L_M tends to ∞.

3) Adaptive-CRB: We define a new performance bound *adaptive*-CRB (aCRB) to describe the performance bound of an unbiased channel estimation for the adaptive-subspace:

$$\mathrm{aCRB}_{\bar{N},w}(\sigma_z^2) = \mathrm{CRB}_{\bar{N},w}(\sigma_z^2,\mathbf{r}^a_w(\sigma_z^2,\bar{N})) + \|\mathbf{\Lambda}^c(w)\|, \quad (34)$$

with $\mathbf{r}^a_w(\sigma_z^2,\bar{N}) = [r^a_{w,1}(\sigma_z^2,\bar{N}),\cdots,r^a_{w,N_T}(\sigma_z^2,\bar{N})]^{\mathrm{T}}$. The sum of singular values in the complemental-subspace is denoted by $\|\mathbf{\Lambda}^c(w)\| = \sum_{k=1}^{N_T}\mathbb{E}[\|\mathbf{\Lambda}_k(l,w)|_{r^a_k(\sigma_z^2,\bar{N})+1:r_k}\|_1]$. By the definition, the aCRB has a property that

$$\mathrm{aCRB}_{\bar{N},w}(\sigma_z^2) \quad \leq \quad \mathrm{CRB}_{\bar{N},w}(\sigma_z^2,\mathbf{r}).$$

The equality holds in a high SNR regime such that $\sigma_z^2 \leq \mathbb{E}[\lambda_k^{r_k}(l,w)]\cdot\bar{N}/N_R$ for $\forall k \in \{1,\cdots,N_T\}$.

4) Asymptotic MSE performance of MB techniques: The MSE performance of the $\ell1$ MB is given by $\mathrm{MSE}(\hat{\mathcal{H}}^{MB}_{\ell1},\sigma_z^2) = \min_w \mathrm{aCRB}_{\bar{N},w}(\sigma_z^2)$. As mentioned above, however, the asymptotic MSE performance of the $\ell1$ MB with $L_M \to \infty$ is independent of the parameter w. The MSE performances of both the $\ell1$ and $\ell2$ MB algorithms are, hence, lower bounded by

$$\mathrm{aCRB}_{\bar{N}}(\sigma_z^2) = \mathrm{CRB}^{\mathcal{Z}}_{\bar{N}}(\sigma_z^2,\mathbf{r}^a_W(\sigma_z^2,\bar{N})) + \|\mathbf{\Lambda}^c(W)\|. \quad (35)$$

V. NUMERICAL EXAMPLES

After describing simulation setups, first of all, MSE performance of proposed techniques is shown. The NMSE convergence property of the new algorithms is then investigated. Moreover, tracking performance against channel changes is

[13]The notation $\bar{N}$ (bar over N) denotes a reference signal length of a channel estimation algorithm, in order to distinguish it from an input signal length $\tilde{N}$ (tilde over N) for the estimation algorithm.

demonstrated to show the robustness of the hybrid algorithm. BER performance of a MIMO turbo receiver with the new channel estimation algorithms is also presented at the end of this section.

A. Simulation Setups

1) Channel models: The CIRs are generated with the spatial channel model (SCM) [7], [33]. This paper assumes 4×4 MIMO channels, where the antenna element spacing at the base station (BS) and the mobile station (MS) are, respectively, set at 4 and 0.5 wavelength. Spatial parameters such as the direction of arrival (DoA) are randomly chosen per a TX chunk. Moreover, six path fading channel realizations based on the Pedestrian-B model with a 3 km/h (PB3) mobility and the Vehicular-A model [7] with a 30 km/h (VA30) mobility are assumed. The path positions of PB and VA are respectively at $\{1\ 2.4\ 6.6\ 9.4\ 17.1\ 26.9\}$ and $\{1\ 3.2\ 6\ 8.6\ 13.1\ 18.6\}$ symbol timings assuming that a transmission bandwidth is 7 MHz with a carrier frequency of 2 GHz.

The receiver can, however, observe CIRs only in the integer symbol timings due to the discrete-time signal processing. In practice, the CIRs are observed as resampled signals so that the original channel parameters at fractional path timings can be reconstructed as samples at the integer symbol timings without distortion. We assume that the resampling is performed by the matched filter (e.g., [34]) with a parameter set $\{\alpha, N_{ovs}, N_{flt}\} = \{0.3, 8, 6\}$, where the parameters denote the roll-off factor of the raised cosine filter, the over-sampling factor and the filter order in symbol, respectively. As shown in Fig. 1, the CIR length observed at the receiver can be around 30 symbols when it follows the PB channel model. The maximum CIR length is hence set at $W = 31$ symbols.

2) TX scenarios: As described in Section I, we focus on intermittent communication scenarios to verify robustness of $\ell 1$ regularized channel estimation. A length $L_C = 100$ burst TX chunk is transmitted continuously. However, as illustrated in Fig. 7, a TX interruption of arbitrary length is assumed between the TX chunks. Two scenarios VA-VA and PB-VA are defined as follows. In the VA-VA scenario, all TX chunks follow a single channel model VA30. The PB-VA scenario has a channel model transition $\{\text{PB3} \rightarrow \text{VA30} \rightarrow \text{PB3} \rightarrow \text{VA30} \rightarrow \ldots\}$ in the series of TX chunks. The variations of the two TX chunks do not always smoothly change due to the interruption, even in the VA-VA scenario.

3) System parameters: The 4×4 MIMO system transmits $N_{\text{info}} = 2048$ information bits. A data frame is encoded by the $R_c = 1/2$ convolutional code with the generator polynomials $(g_1, g_2) = (7, 5)_8$. The number N_B of bursts per a TX stream in a frame is determined such that $N_B = N_{\text{info}}/(N_T N_d)$. The burst format parameters are set at $(N_t, N_{\text{CP}}, N_G, N_d) = (127, W, W, 512)$. The TSs are generated with the pseudo noise (PN) sequence [35].

B. Normalized MSE Performance with LS channel estimation techniques

We define a normalized mean squared error (NMSE) of a channel estimate $\hat{\mathcal{H}}$ by $\text{NMSE}(\hat{\mathcal{H}}, \sigma_z^2) = \text{MSE}(\hat{\mathcal{H}}, \sigma_z^2)/\mathbb{E}[\|\mathcal{H}\|^2]$. Similarly, normalized aCRB (NaCRB) is denoted as $\text{NaCRB}_{\bar{N}_{td}}(\sigma_z^2) = \text{aCRB}_{\bar{N}_{td}}(\sigma_z^2)/\mathbb{E}[\|\mathcal{H}\|^2]$ with asymptotic aCRB (35), where the length $\bar{N}_{td}$ of reference signals composed of TS and data sequences is defined by $\bar{N}_{td} = \mathtt{tr}\{\mathcal{R}_{\mathcal{X}\mathcal{X}}\}/(W N_T N_R)$. The NaCRB for the PB-VA scenario is assumed as the mean of those for the PB and VA channel models.

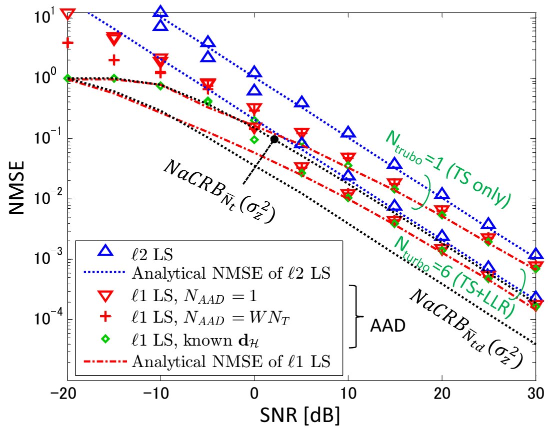

Fig. 3. NMSE performance with LS channel estimation techniques in the PB-VA scenario. N_{turbo} denotes the number of turbo iterations.

1) Comparison between the $\ell 1$ and $\ell 2$ LS techniques: Fig. 3 shows NMSE performance of the $\ell 1$ LS and $\ell 2$ LS channel estimation techniques in the 4×4 MIMO system. The PB-VA scenario is assumed. The channel estimation results are obtained after performing the first and the sixth turbo iteration ($N_{\text{turbo}} = 1, 6$). The maximum number of iteration in Algorithm 1 is set at $N_{\text{AAD}} = 1$ or WN_T. As observed from Fig. 3, the $\ell 1$ LS technique improves the NMSE significantly compared to the $\ell 2$ version in a low to moderate SNR regime. This is because the dominant CIRs above the noise level exist sparsely in the SNR regime. In a high SNR regime, however, the CIRs cannot be assumed as sparse channels. Thereby, in the PB-VA scenario, the $\ell 1$ LS does not improve NMSE performance over the $\ell 2$ version in the high SNR regime, although enough turbo iterations ($N_{\text{turbo}} = 6$) are performed.

In a very low SNR regime, the NMSE with the $\ell 1$ LS deviates from the analytical MSE performance. This is because, even though the NMSE performance is improved by setting $N_{\text{AAD}} > 1$, the active-set detection (12) can fail in the very low SNR regime such that $\text{NMSE}(\hat{\mathcal{H}}_{\ell 2}^{LS}, \sigma_z^2) \gg 1$ since the delay profile is approximated with LS estimates. As shown later, the problem is improved with the hybrid algorithm since the $\ell 2$ MB method estimates the path number correctly.

2) Comparison between $\ell 1$ solvers – AAD vs. OMP / SP: The following two subsections compare the AAD algorithm with well-known $\ell 1$ solvers. Before discussing the comparison, it should be noticed that we can straightforwardly extend $\ell 1$ solvers such as the OMP and subspace pursuit (SP) [36] algorithms for MIMO channel estimation by using (6). Although the algorithms are not explicitly shown due to the page limitation, we briefly describe a note for the MIMO extension. The active-set in the OMP-based algorithms can be constructed by using either of the following criteria: 1) $\underset{1\le j\le WN_TN_R}{\arg\max}\ (|\mathtt{vec}\{\Xi\}|)|_j$, or 2) $\underset{1\le j\le WN_T}{\arg\max}\ \mathtt{diag}\{\Xi^{\rm H}\Xi\}|_j$, where the residual correlation Ξ is defined as $\Xi = (\mathcal{Y}_t - \hat{\mathcal{H}}\mathcal{X}_t)\mathcal{X}_t^{\rm H} + \mathbf{\Gamma}(\mathcal{Y}_d - \hat{\mathcal{H}}\hat{\mathcal{X}}_d)\hat{\mathcal{X}}_d^{\rm H}$ for a possible estimate $\hat{\mathcal{H}}$ obtained in the OMP-based algorithms. In Figs. 4, the OMP algorithms with criteria 1) and 2) are referred to as *vec-OMP* and *OMP*, respectively. As observed from Figs. 4(a) and (b), channel estimation with the criterion 2) achieves better NMSE performance than the vec-OMP. This is because the diversity combining over N_R Rx antennas by the matrix product $\Xi^{\rm H}\Xi$ improves the accuracy of the active-set selection. We, hence, focus on the OMP with the criterion 2) hereafter.

Fig. 4(a) shows NMSE performance with the OMP, SP and AAD algorithms in the VA-VA scenario. Channel estimation is performed with the TS only. As observed from Fig. 4(a), the AAD achieves the same NMSE performance as that of the OMP and SP algorithms, where the degree of sparsity (DoS) for OMP and SP is given by the cardinality of the estimated active-set (12). In other words, the NMSE performance is not improved significantly by combining the AAD with the OMP and SP algorithms. If the DoS is known, of course, the NMSE performance with the OMP and SP algorithms is improved. However, the knowledge of the delay profile is required to determine the DoS correctly. It should be noted that, as shown in Figs. 3 and 4, the AAD algorithm achieves the analytical MSE performance of the $\ell 1$ LS exactly if the delay profile is known.

3) Comparison between $\ell 1$ solvers – AAD vs. ITDSE: The ITDSE [3] algorithm detects the active-set iteratively by increasing a threshold with a step-wise $\max_j \hat{d}^{[0]}_{\mathcal{H},j}/N_{RES}$, where N_{RES} denotes a resolution constant. As observed from Fig. 4(a), the NMSE with the ITDSE algorithm follows the analytical MSE performance if the resolution constant is set large enough. (e.g., $N_{RES} = 10^4$ is required in the VA-VA scenario for SNR $\ge$ 25 dB.) The original ITDSE has to perform N_{RES} iterations, although NMSE convergence performance shown in Fig. 4(c) suggests that the ITDSE may terminate the process before the N_{RES}-th iteration with a certain criterion. We note that, even with $N_{\rm AAD} = 1$, the AAD algorithm can detect the active-set very accurately since it decides the threshold adaptively according to the SNR. Therefore, the computational complexity required for the AAD algorithm is significantly decreased from that of the ITDSE.

4) Analytical NMSE performance of the $\ell 1$ LS: As shown in Figs. 3 and 4(a), the analytical NMSE of the $\ell 1$ LS channel estimation does not achieve the performance bound NaCRB in approximately sparse channels. As an exception, Fig. 4(b) shows the NMSE performance in sparse-VA channels, the path positions of which are set at integer symbol timings $\{1, 3, 6, 9, 13, 19\}$. Effect of Tx/Rx filters is also neglected. As observed from Fig. 4(b), the analytical NMSE of the $\ell 1$ LS technique coincides with the NaCRB in the sparse-VA scenario. This is because the eigen domain of the signal of interest is identical to the temporal domain in the exactly sparse channels.

C. Normalized MSE Performance with the MB and hybrid algorithms

Figs. 5 show NMSE with the MB channel estimation in the VA-VA (a) and PB-VA (b) scenarios. The MB sliding window length is set at $L_M = 50$ bursts. We note that $N_R L_M = 200$ is long enough so that NMSE converges. As shown in Fig. 5(a), both the $\ell 1$ and $\ell 2$ MB algorithms achieve the NaCRB asymptotically. This observation verifies the MSE performance analysis described in the Section IV-B. Furthermore, it should be emphasized that the NaCRB saturates at 1 if SNR $\le$ -15 dB. This is because in the very low SNR regime, the adaptive-rank (31) becomes $\mathbb{E}[r_k^a(l)] = 0$. The NMSEs with the $\ell 1$ and $\ell 2$ MB algorithms also saturate at 1 in the very low SNR regime, hence, they follow the NaCRB rather than the ordinary normalized CRB (NCRB). The NMSE with the hybrid algorithm follows that of the $\ell 2$ MB, where the maximum number of iteration in the AAD algorithm is set at 1.

Fig. 5(b) shows the case of PB-VA scenario. The $\ell 2$ MB exhibits NMSE deterioration from that in the VA-VA since the PB-VA scenario has abrupt channel changes. The $\ell 1$ MB algorithm improves the NMSE in the low SNR regime, nevertheless, the gain is slight even by using the oracle criterion which minimizes the squared error $\|\hat{\mathcal{H}} - \mathcal{H}\|^2$ between a possible channel estimate $\hat{\mathcal{H}}$ and a known CIR $\mathcal{H}$. The robustness with the $\ell 1$ regularization is investigated further in terms of NMSE convergence properties and BER performance in the subsequent sections.

D. NMSE Convergence Properties

1) Effect of LLR's accuracy onto NMSE: Figs. 6 depict NMSE performance over LLR's accuracy at SNR = 15 dB in the VA-VA (a) and PB-VA (b) scenarios. We define the LLR's accuracy by the mutual information (MI) $\mathcal{I}^a_{\rm EQU}$ between the LLR $\lambda^a_{\rm EQU}$ and the coded bits c at the transmitter, as

$$\begin{aligned}\mathcal{I}^a_{\rm EQU} &= \mathcal{I}(\lambda^a_{\rm EQU}; c) \\ &= \frac{1}{2}\sum_{\mathfrak{m}=\pm 1}\int_{-\infty}^{+\infty} {\rm P_r}(\lambda^a_{\rm EQU}|\mathfrak{m})\ \log_2 \frac{{\rm P_r}(\lambda^a_{\rm EQU}|\mathfrak{m})}{{\rm P_r}(\lambda^a_{\rm EQU})}\ d\lambda^a_{\rm EQU},\end{aligned} \tag{36}$$

where ${\rm P_r}(\lambda^a_{\rm EQU}|\mathfrak{m})$ is the conditional probability density of $\lambda^a_{\rm EQU}$ given $\mathfrak{m} = 1 - 2c$ [37].

It is observed from Fig. 6(a) that all channel estimation techniques improve the NMSE performance as MI increases. This is because the reference signal length $\bar{N}_{td}$ is proportional to the MI $\mathcal{I}^a_{\rm EQU}$, since $\bar{N}_{td} \approx N_t + \gamma\hat{\sigma}_d^2\{N_d - (W-1)/2\}$ with $\gamma = \sigma_z^2/(\sigma_z^2 + \Delta\hat{\sigma}_d^2 N_T \sigma_{\mathbf{H}}^2/N_R)$ holds when $\mathbf{R}_{\mathcal{H}\mathcal{H}} \approx (N_T\sigma_{\mathbf{H}}^2/N_R)\mathbf{I}_{N_R}$. The variance of $\lambda^a_{\rm EQU}$ tends to ∞ as $\mathcal{I}^a_{\rm EQU}$ converges to 1 [37], which gets $\|\hat{\mathbf{x}}_{d,k}\|^2/N_d$ and $\Delta\hat{\sigma}_d^2$ converged to σ_x^2 and 0, respectively.

The $\ell 1$ MB algorithm improves the NMSE over the $\ell 2$ MB channel estimation in the entire MI regime since it can decrease the projection error as discussed in Section IV-B. The hybrid algorithm is inferior to the $\ell 1$ MB in the VA-VA scenario since it behaves as the $\ell 2$ MB when CIRs follow a single channel model. Nevertheless, as shown in Fig. 6(b), the hybrid algorithm improves NMSE over the $\ell 1$ MB if there are

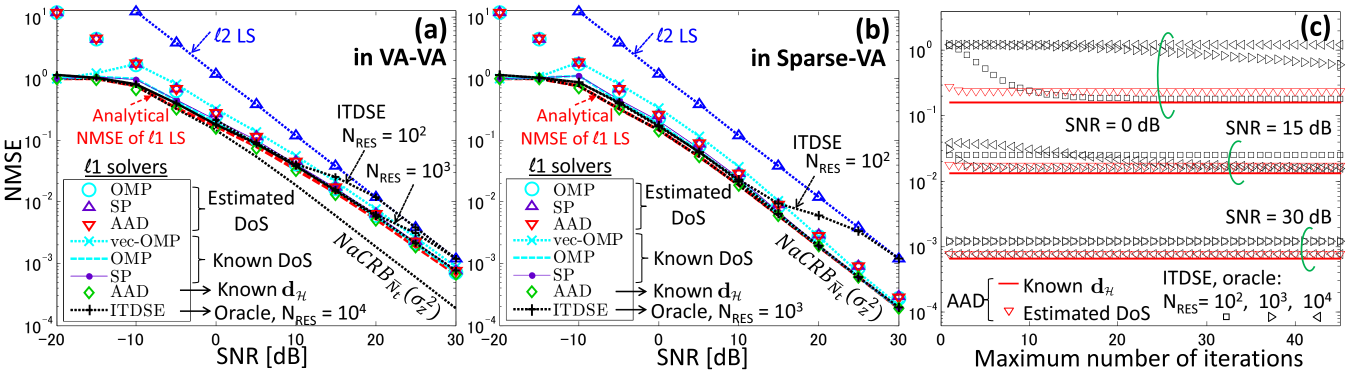

Fig. 4. Comparison between $\ell1$ solvers: NMSE performance over SNR with $\ell1$ LS channel estimation in the VA-VA (a) and sparse-VA (b) scenarios and NMSE convergence over iteration (c) in the VA-VA scenario. Channel estimation is performed with the TS only. For the OMP and SP algorithms, known and estimated DoSs are given by the cardinality of the optimal active-set (30) and the cardinality of the estimated active-set (12), respectively. The numbers of the maximum iterations for the vec-OMP, OMP, SP and AAD algorithms are set at WN_TN_R, WN_T, WN_T, and 1, respectively. The ITDSE in these figures determines the optimal solution from possible channel estimates by the oracle criterion.

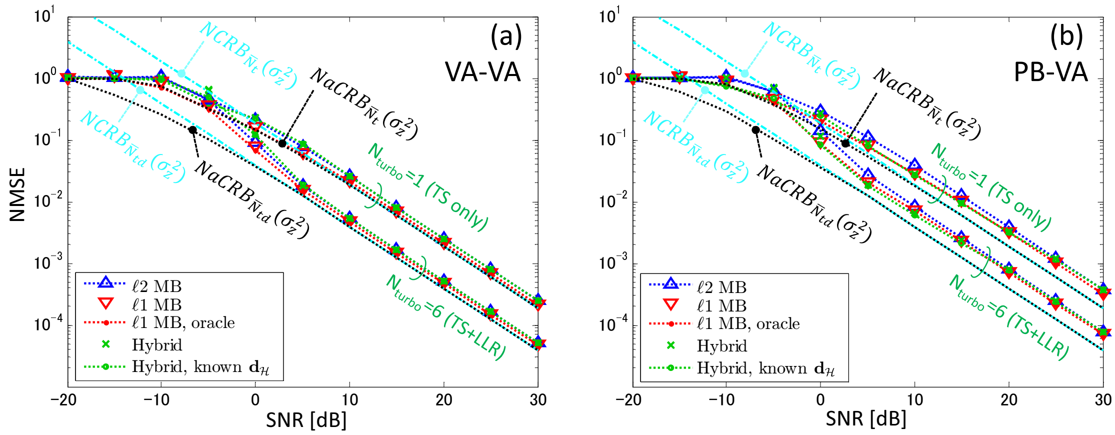

Fig. 5. NMSE performance with MB channel estimation techniques in the VA-VA (a) and PB-VA (b) scenarios. Normalized CRB is given by $NCRB_{\bar{N}}(\sigma_z^2) = CRB_{\bar{N}}^{z}(\sigma_z^2, \mathbf{r})/\mathbb{E}[\|\mathcal{H}\|^2]$ with (32), where all N_T entries of the rank vector $\mathbf{r}$ are set at 6 for the PB or VA channel models. In the hybrid algorithm, $N_{\rm AAD} = 1$ is assumed.

abrupt channel changes such as in the PB-VA scenario. The reason for the improvement is clarified by observing NMSE tracking performance.

2) Tracking performance : Fig. 7 shows the NMSE tracking performance in the PB-VA scenario. The $\ell2$ MB channel estimation suffers from the NMSE tracking errors as seriously as that causes bit errors at the boarders between the TX chunks. The $\ell1$ MB channel estimation also suffers from the NMSE tracking errors, however, improves bit errors at the boarders between the TX chunks. This is because, as described in Section IV-B, the $\ell1$ regularization decreases the projection error. Nevertheless, as observed from Fig. 7, the $\ell1$ MB cannot solve the NMSE tracking error problem completely. This is because the $\ell1$ MB estimate inherits the past CIRs' characteristics in the sliding window of the MMSE formulation.

On the other hand, the $\ell1$ LS channel estimation does not suffer from the NMSE tracking error problem since it detects the active-set for each burst independently. The hybrid algorithm can, therefore, avoid the tracking error problem by utilizing the $\ell1$ LS, while achieving the performance bound aCRB asymptotically by the $\ell2$ MB estimate when the tracking error problem is not observed.

E. BER Performance

The average SNR used in BER simulations is defined in association with the average energy per bit to noise density ratio (E_b/N_0) as $\text{SNR} = \sigma_x^2\ (\sigma_{\mathbf{H}}^2/N_R)\ \eta \cdot E_b/N_0$, where we assume that the variances of a transmitted symbol and CIRs per a TX stream are $\sigma_x^2 = 1$ and $\sigma_{\mathbf{H}}^2 = 1$, respectively. The spectrum efficiency η of the frame format structure is defined as $\eta = N_{\rm info}/L_{frm}$ with a frame length $L_{frm} = L_BN_B$ in symbol. It is hence reduced to $\eta = 1.4$ for the MIMO system used in the simulations.

Figs. 8 show BER performance with the receiver using the new channel estimation techniques in the 4×4 MIMO system. BERs with the receiver assuming known CIRs $\mathcal{H}(l)$ are also shown as the BER performance bound of the system. BER is obtained after performing the first and sixth turbo iterations. In the case the VA-VA scenario is assumed, as observed from

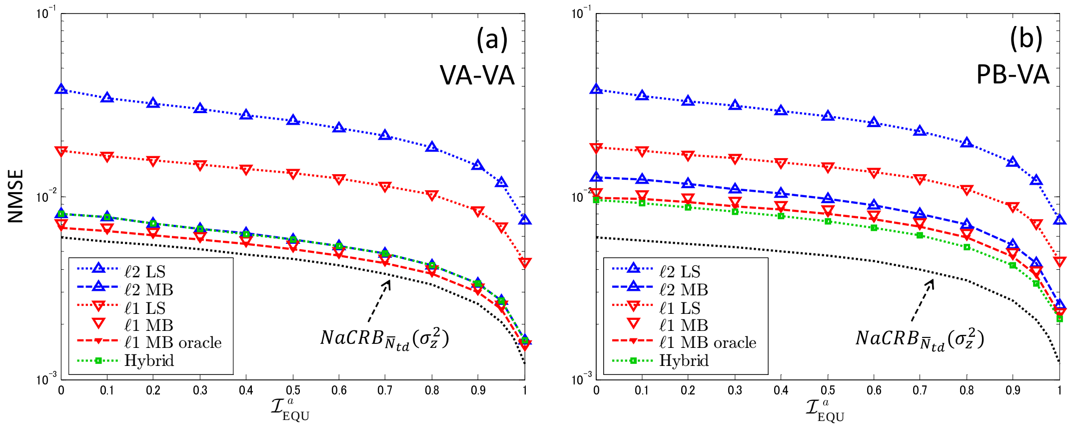

Fig. 6. NMSE convergence performance over the MI $\mathcal{I}^a_{\rm EQU}$ (36) in the VA-VA (a) and PB-VA (b) scenarios at SNR is set at 15 dB.

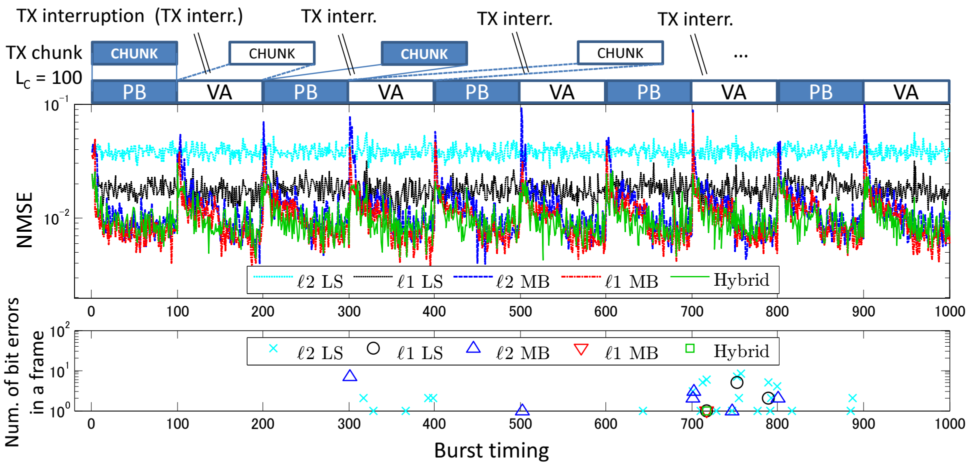

Fig. 7. NMSE tracking performance in the PB-VA scenario. The arbitrary length TX interruptions are omitted in NMSE tracking results. The channel estimation is performed with the TS only. SNR is set at 15 dB. In the second subfigure, the number (num.) of bit errors in the i-th frame is shown at the $\{(i-1)N_B+1\}$-th burst timing.

Fig. 8(a), the receiver using the $\ell 1$ MB achieves the BER performance bound asymptotically. However, even with the oracle criterion, the $\ell 1$ MB does not improve BER significantly over that of the $\ell 2$ MB technique.

In the PB-VA scenario, as shown in Fig. 8(b), the BER performance with the $\ell 2$ MB deviates from that of known $\mathcal{H}$ by 4 dB at BER = 10^{-5}, even after performing the sixth turbo iteration. This is because, as shown in Fig. 7, the $\ell 2$ MB suffers from the tracking error problem. The receiver with the $\ell 1$ MB improves the tracking error problem, however, its BER performance is still away from the bound by roughly 2.5 dB.

As shown in Fig. 8(b), the $\ell 2$ MB can, of course, improve the tracking error problem by resetting the length L_m of the MMSE sliding window at the start timing of each TX chunk. Nevertheless, as observed from Fig. 8(a), the $\ell 2$ MB with the L_m resetting suffers from BER degradation if there is no tracking problem. This is because MSE performance of the $\ell 2$ MB is unstable for $\lceil W/N_R \rceil$ bursts after the L_m resetting. The proposed hybrid algorithm compensates the MSE deterioration by utilizing the $\ell 1$ LS channel estimate for the unstable duration. Moreover, the hybrid algorithm resets the sliding window length only when the tracking error is detected. The receiver with the hybrid algorithm can, therefore, achieve roughly a 2 dB gain in E_b/N_0 at BER = 10^{-5} over that of the $\ell 2$ MB method in the PB-VA scenario, while obtaining the BER performance bound asymptotically in the VA-VA scenario.

VI. Conclusions

This paper has studied the performance of $\ell 1$ regularized turbo channel estimation algorithms in broadband MIMO wireless channels, via theoretical analysis supported with simulation results. The $\ell 1$ LS channel estimation does not achieve MSE performance bound in broadband wireless channels since the CIRs at the receiver are, in general, not observed as exactly sparse channels due to the effect of Tx/Rx filters. The MSE performance of both the $\ell 1$ MB and $\ell 2$ MB algorithms are bounded by the aCRB defined in this paper. Moreover, the $\ell 1$ MB technique does not improve MSE significantly over the $\ell 2$ MB if the following three assumptions are correct: 1) CIRs

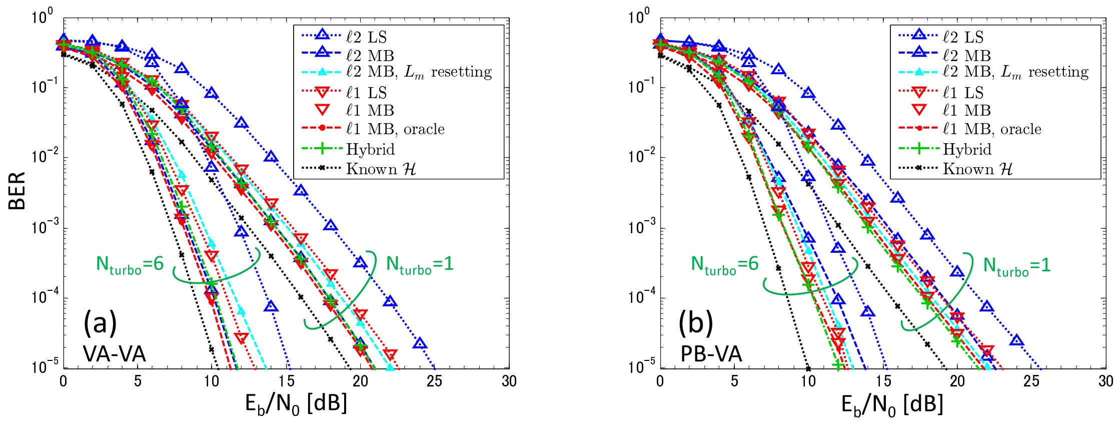

Fig. 8. BER performance with the 4×4 MIMO system in the VA-VA (a) and PB-VA (b) scenarios.

follow the subspace channel model. 2) The sliding window length in the MMSE formulation is set long enough. 3) The reference signals are ideally uncorrelated.

However, the $\ell 2$ MB technique suffers from deterioration in the channel estimation performance if the three assumptions are partially incorrect. By focusing on intermittent TX scenarios which do not always satisfy the first assumption, this paper has demonstrated robustness with $\ell 1$ regularization. Simulation results shows that, due to the tracking error problem, the receiver with the $\ell 2$ MB exhibits BER degradation in the PB-VA scenario although enough number of turbo iterations are performed. The $\ell 1$ MB improves the tracking error by decreasing the projection error, however, it requires a larger complexity order than the $\ell 2$ MB.

The hybrid algorithm proposed in this paper solves the tracking error problem completely. Therefore, the receiver with the proposed algorithm achieves a significant BER gain over the $\ell 2$ MB technique in the PB-VA scenario, while obtaining the BER performance bound asymptotically in the VA-VA scenario. It should be noted that the computational complexity order required for the hybrid algorithm is equivalent to that of the $\ell 2$ MB if the number of the maximum iteration of the AAD algorithm is set at 1.

Appendix A

1) Approximation of the MSE (24)*:* For the sake of simplicity, the burst timing index l is omitted hereafter. If both the training and data signals are ideally uncorrelated sequences, $\mathbf{R}_{\mathcal{X}\mathcal{X}_t}/\bar{N}_t \approx \mathbf{I}_{WN_T}$ and $\hat{\mathbf{R}}_{\mathcal{X}\mathcal{X}_d}/\bar{N}_d \approx \mathbf{I}_{WN_T}$, where $\bar{N}_t = N_t$ and $\bar{N}_d = \hat{\sigma}_d^2\{N_d - (W-1)/2\}$. Hence, $\mathcal{R}_{\mathcal{X}\mathcal{X}}/\bar{N}_{td} \approx \mathbf{I}_{WN_TN_R}$ with $\Delta\hat{\sigma}_d^2 \approx 0$. Accordingly, we have approximations

$$\mathrm{tr}\{\mathcal{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}}}^{-1}\} \approx \frac{|\mathcal{A}|}{WN_T}\mathrm{tr}\{\mathcal{R}_{\mathcal{X}\mathcal{X}}^{-1}\} \tag{37}$$

and $\mathfrak{E}(\mathcal{A}) \approx \mathbb{E}\left[\|\mathcal{H}_{\mathcal{A}}^{\perp}\|^2\right]$. The analytical MSE (24) is, therefore, approximated by

$$\mathrm{MSE}(\hat{\mathcal{H}}_{\ell 1}^{LS}, \sigma_z^2, \mathcal{A}) \approx |\mathcal{A}|\frac{\mathrm{MSE}(\hat{\mathcal{H}}_{\ell 2}^{LS}, \sigma_z^2)}{WN_T} + \mathbb{E}[\|\mathcal{H}_{\mathcal{A}}^{\perp}\|^2] \tag{38}$$

$$= \mathbb{E}[\|\mathcal{H}\|^2] + \sum_{j\in\mathcal{A}}\left\{\frac{\mathrm{MSE}(\hat{\mathcal{H}}_{\ell 2}^{LS}, \sigma_z^2)}{WN_T} - \bar{d}_{\mathcal{H},j}\right\}, \tag{39}$$

since $\mathbb{E}[\|\mathcal{H}_{\mathcal{A}}^{\perp}\|^2] = \mathbb{E}[\|\mathcal{H}\|^2] - \sum_{j\in\mathcal{A}}\bar{d}_{\mathcal{H},j}$, where $\bar{d}_{\mathcal{H},j}$ denotes the j-th entry in the delay profile $\mathbb{E}[\mathbf{d}_{\mathcal{H}}]$. The problem (30) can also be approximated by

$$\begin{aligned}\mathcal{A}^* &\approx \arg\min_{\mathcal{A}} \sum_{j\in\mathcal{A}}\left\{\frac{\mathrm{MSE}(\hat{\mathcal{H}}_{\ell 2}^{LS}, \sigma_z^2)}{WN_T} - \bar{d}_{\mathcal{H},j}\right\} \\ &= \left\{ j \;\middle|\; \begin{array}{l}\bar{d}_{\mathcal{H},j} > \mathrm{MSE}(\hat{\mathcal{H}}_{\ell 2}^{LS}, \sigma_z^2)/(WN_T), \\ j = 1, \cdots, WN_T\end{array}\right\}. \end{aligned}\tag{40}$$

2) Derivation of the AAD: It is reasonable to assume that $\|\hat{\mathcal{H}}_{\ell 1}^{LS} - \mathcal{H}\|^2 \approx \mathbb{E}[\|\hat{\mathcal{H}}_{\ell 1}^{LS} - \mathcal{H}\|^2]$, when the reference signal length is long enough. Under this assumption, the problem (10) can be seen as an approximated version of the minimization of (25). Hence, (10) can be reduced to a solution corresponding to (40). Accordingly, the AAD algorithm approximates the delay profile $\mathbb{E}[\mathbf{d}_{\mathcal{H}}]$ by using the channel estimate obtained in the previous iteration. The approximation error is dominated by the first term of (38) if the active-set can be selected so that $\|\mathcal{H}_{\mathcal{A}}^{\perp}\|^2$ is very minor. It should be noticed that

$$\mathrm{tr}\{\mathcal{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}^*}}^{-1}\} \lessapprox \frac{|\mathcal{A}^*|}{|\mathcal{A}_{[n]}|}\mathrm{tr}\{\mathcal{R}_{\mathbf{\Phi\Phi}_{\mathcal{A}_{[n]}}}^{-1}\} \lessapprox \frac{|\mathcal{A}^*|}{WN_T}\mathrm{tr}\{\mathcal{R}_{\mathcal{X}\mathcal{X}}^{-1}\} \tag{41}$$

is satisfied for $\mathcal{A}^* \subseteq \mathcal{A}_{[n]} \subseteq \{1, \cdots, WN_T\}$ by (27) and (37). Thereby, the active-set detection (12) is an extension of (40) so that it takes account of the delay profile approximation error. Furthermore, the recursive formula (12) aims to improve detection accuracy by the inequality (41). However, even with $N_{\mathrm{AAD}} = 1$, Algorithm 1 can detect the active-set accurately when ideally uncorrelated reference signals are used. This is because the equalities in (41) holds when $\mathcal{R}_{\mathcal{X}\mathcal{X}}/\bar{N}_{td} = \mathbf{I}_{WN_TN_R}$.

REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] S. Cotter and B. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, 2002.

[3] 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.

[4] C. Berger, S. Zhou, J. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1708–1721, 2010.

[5] C. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, 2010.

[6] J. Huang, J. Huang, C. Berger, S. Zhou, and P. Willett, "Iterative sparse channel estimation and decoding for underwater MIMO-OFDM," in *OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges*, 2009, pp. 1–8.

[7] ETSI, "Universal Mobile Telecommunications System (UMTS); Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (3GPP TR 25.996 version 12.0.0 Release 12)," Sep. 2014.

[8] J. Haupt, W. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5862–5875, 2010.

[9] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of fast-varying space-time communication channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1184 – 1195, may 2003.

[10] ——, "Multislot estimation of frequency-selective fast-varying channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1337–1347, Aug 2003.

[11] M. Nicoli, S. Ferrara, and U. Spagnolini, "Soft-iterative channel estimation: Methods and performance analysis," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2993–3006, 2007.

[12] S. Cai, T. Matsumoto, and K. Yang, "SIMO channel estimation using space-time signal subspace projection and soft information," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4219–4235, 2012.

[13] A. Paulraj, D. Gore, R. Nabar, and H. Bolcskei, "An overview of MIMO communications – a key to gigabit wireless," *Proc. IEEE*, vol. 92, no. 2, pp. 198–218, Feb 2004.

[14] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge university press, 2005.

[15] C. Komninakis and R. Wesel, "Joint iterative channel estimation and decoding in flat correlated rayleigh fading," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1706–1717, 2001.

[16] R. Iltis, "Iterative joint decoding and sparse channel estimation for single-carrier modulation," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008, pp. 2689–2692.

[17] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Signals, Systems, and Electronics, 1998. ISSSE 98. 1998 URSI International Symposium on*, 1998, pp. 295–300.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] S. Chen and D. Donoho, "Basis pursuit," in *Signals, Systems and Computers, 1994. 1994 Conference Record of the Twenty-Eighth Asilomar Conference on*, vol. 1, Oct 1994, pp. 41–44 vol.1.

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

[24] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[25] H. Zou, T. Hastie, and R. Tibshirani, "On the degrees of freedom of the Lasso," *The Annals of Statistics*, vol. 35, no. 5, pp. 2173 – 2192, 2007.

[26] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461 – 464, 1978.

[27] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[28] I. T. Jolliffe, N. T. Trendafilov, and M. Uddin, "A modified principal component technique based on the LASSO," *Journal of Computational and Graphical Statistics*, vol. 12, no. 3, pp. 531–547, 2003.

[29] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716 – 723, Dec. 1974.

[30] G. Golub and C. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 1983.

[31] 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*, March 2010, pp. 1882–1885.

[32] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge university press, 2012.

[33] 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.

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

[35] S. Haykin, *Adaptive Filter Theory*, ser. Prentice Hall Information and System Science Series. Prentice Hall, 2002.

[36] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[37] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727 –1737, Oct. 2001.