JAIST Repository
https://dspace.jaist.ac.jp/

Title	シングルキャリア広帯域無線通信のためのスペクトラム利用効率に優れたターボ受信技術
Author(s)	高野，泰洋
Citation	
Issue Date	2016-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/13514
Rights	
Description	Supervisor:松本　正，情報科学研究科，博士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Spectrally Efficient Turbo Reception Technologies for Single-Carrier Broadband Wireless Communications

Yasuhiro Takano

March, 2016

Academic dissertation to be presented with the assent of the Doctoral Training Committee of School of Information Science of Japan Advanced Institute Science and Technology (JAIST) for public defense in Collaboration Room 7 (I-56), on 12 February 2016, at 3 p.m.

Supervised by
Professor Markku Juntti
Professor Tadashi Matsumoto

Reviewed by
Assistant Professor Monica Nicoli
Associate Professor Brian Kurkoski
Professor Christoph Mecklenbräuker
Professor Takeo Ohgane
Professor Mineo Kaneko

The dissertation is made based on a curriculum that is organized by the Collaborative Education Program of Japan Advanced Institute Science and Technology (JAIST), Japan, and the Centre for Wireless Communications (CWC), University of Oulu.

JAIST Press, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan.
Tel: +81-761-1980, FAX:+81-761-1199
e-mail: jaistpress@jaist.ac.jp
http://www.jaist.ac.jp/library/jaist-press

ISBN978-4-903092-42-3

Takano, Yasuhiro, Spectrally efficient turbo reception technologies for single-carrier broadband wireless communications.
Japan Advanced Institute Science and Technology, School of Information Science, Information Theory and Signal Processing Laboratory;
University of Oulu Graduate School; University of Oulu, Faculty of Information Technology and Electrical Engineering, Department of Communications Engineering; Centre for Wireless Communications; Infotech Oulu;
JAIST Press, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan.

Abstract

Future broadband wireless communication systems are expected to increase both their transmission (TX) rate and their spectrum efficiency under the constraints of low TX power and a low computational complexity. In general, a data sequence is transmitted together with overheads such as training sequence (TS) required to perform energy- and computationally-efficient reception techniques. We hence have a trade-off between the spectral efficiency and the receiver performance. The objective of this thesis is to enhance robustness of the receiving algorithms with reasonable complexity, aiming to improve the trade-off.

For this purpose, $\ell 1$ regularized channel estimation techniques are studied under an assumption that broadband wireless channels observed at a receiver does not fully exhibit dense nature in a low to moderate signal-to-noise ratio (SNR) regime. This thesis proposes a novel conditional $\ell 1$ regularized minimum mean square error (MMSE) channel estimation and chained turbo estimation (CHATES) algorithms to solve the inter-block-interference (IBI) problem incurred as the result of pursuing spectral efficiency. A new $\ell 1$ least squares (LS) and $\ell 2$ MMSE-based hybrid channel estimation algorithm is also proposed to solve the tracking error problem often observed with intermittent transmission. Moreover, performance analysis shows that an $\ell 1$ regularized MMSE channel estimation algorithm can achieve the Cramér-Rao bound (CRB) asymptotically even when random TSs are used.

This thesis further studies frequency domain turbo equalization techniques without cyclic prefix (CP) transmission to improve the spectral efficiency. The previously-proposed chained turbo equalization, referred to as CHATUE1, allows us to use a lower rate code. However, it can suffer from the noise enhancement problem at the equalizer output. As a solution to the problem, this thesis proposes a new algorithm, CHATUE2. The theoretical analysis supported with simulation results shows that the proposed CHATUE2 can solve the problem after performing enough turbo iterations by utilizing a new composite replica constructed with the conventional soft replica and received signals.

Keywords: Subspace-based channel estimation, compressive sensing, turbo channel estimation, turbo equalization, spectral efficiency.

Contents

List of Figures

List of Tables

Notations

$\mathbf{x}$	A vector.
$\mathbf{X}$	A matrix.
$\mathbf{X}^{\mathrm{T}}$	The transpose of matrix $\mathbf{X}$.
$\mathbf{X}^{\mathrm{H}}$	The conjugate transpose of matrix $\mathbf{X}$.
$\mathbf{X}^{-1}$	The matrix inverse of matrix $\mathbf{X}$.
$\mathbf{X}^{\dagger}$	The Moore-Penrose pseudoinverse of matrix $\mathbf{X}$.
$\mathcal{A}$	An index set sorted in an ascending order. The set 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 $i \leq j$.
$\lvert\mathcal{A}\rvert$	The cardinality of the argument set $\mathcal{A}$.
$\mathbf{X}\lvert_{\mathcal{A}}$	A submatrix composed of the column vectors in a matrix $\mathbf{X}$, the columns of which are defined by index set $\mathcal{A}$.
$\mathbf{x}\lvert_{\mathcal{A}}$	A subvector of a vector $\mathbf{x}$ which extracts the elements specified by index set $\mathcal{A}$ from the vector $\mathbf{x}$.
$\lVert\mathbf{X}\rVert^2_{\mathbf{W}}$	A weighted Frobenius norm defined by $\mathrm{tr}\{\mathbf{X}^{\mathrm{H}}\mathbf{W}\mathbf{X}\}$ 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 $\lVert\mathbf{X}\rVert^2_{\mathbf{W}} = \lVert\mathbf{X}\rVert^2$, where $\mathbf{I}_M$ is an $M \times M$ identity matrix.
$\lVert\mathbf{X}\rVert_1$	An $\ell 1$ norm for a matrix $\mathbf{X} \in \mathbb{C}^{M\times N}$ defined by $\sum_{i=1}^{M}\sum_{j=1}^{N} \lvert x_{ij}\rvert$, where x_{ij} is the (i,j)-th element of the matrix $\mathbf{X}$.
$\mathtt{vec}(\mathbf{X})$	A vectorization operator to produce an $MN \times 1$ vector by stacking the columns of an $M \times N$ matrix $\mathbf{X}$.
$\mathtt{mat}_N(\mathbf{x})$	An operation to form an $N \times M$ matrix from the argument vector $\mathbf{x} \in \mathbb{C}^{NM\times 1}$, so that $\mathbf{x} = \mathtt{vec}\{\mathtt{mat}_N\{\mathbf{x}\}\}$.
$\mathtt{diag}(\mathbf{X})$	An operation to form a vector from the diagonal elements of its argument matrix $\mathbf{X}$.

$\texttt{DIAG}(\mathbf{x})$	An operation to form a diagonal matrix from its argument vector $\mathbf{x}$.
$\texttt{rank}(\mathbf{X})$	The rank of matrix $\mathbf{X}$
$\texttt{svd}(\mathbf{X})$	The singular value decomposition (SVD) of matrix $\mathbf{X} \in \mathbb{C}^{M\times N}$: $\texttt{svd}(\mathbf{X}) = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{H}}$, where $\mathbf{U} \in \mathbb{C}^{M\times M}$ and $\mathbf{V} \in \mathbb{C}^{N\times N}$ are unitary matrices. $\mathbf{D} \in \mathbb{C}^{M\times N}$ is a rectangular matrix with a square diagonal matrix on the left top corner.
$\mathbb{E}_{j\in\mathcal{J}}\left[\mathbf{X}(j)\right]$	The expectation matrix of matrix sequence $\mathbf{X}(j)$ defined by $\frac{1}{\lvert\mathcal{J}\rvert}\sum_{j\in\mathcal{J}}\mathbf{X}(j)$ for the argument index set $\mathcal{J}$.
$\mathbb{E}[\mathbf{X}(j)]$	Asymptotic expectation matrix: $\mathbb{E}_{j\in\mathcal{J}}[\mathbf{X}(j)]$ with $\lvert\mathcal{J}\rvert \to \infty$.
$\mathbb{K}_{j\in\mathcal{J}}\left[\mathbf{X}(j)\right]$	The covariance matrix of matrix sequence $\mathbf{X}(j)$ defined by $\frac{1}{\lvert\mathcal{J}\rvert}\sum_{j\in\mathcal{J}}\mathbf{X}^{\mathrm{H}}(j)\ \mathbf{X}(j)$.
$\mathbb{K}[\mathbf{X}(j)]$	Asymptotic covariance matrix: $\mathbb{K}_{j\in\mathcal{J}}[\mathbf{X}(l)]$ with $\lvert\mathcal{J}\rvert \to \infty$.
$\mathbb{P}_{\Pi}(\mathbf{U})$	Projection matrix defined by $\mathbf{U}\mathbf{U}^{\dagger}$.
$\texttt{mod}(n, m)$	n modulo m for integers n and m.
$\mathbb{C}^{M\times N}$	Complex number field of dimension $M \times N$.
$\mathbb{R}^{M\times N}$	Real number field of dimension $M \times N$.
$\mathcal{O}(f_N)$	Computational complexity order of f_N by the big O notation.
$\mathfrak{R}(\mathbf{v})$	The real part of complex vector $\mathbf{v}$.
$\mathcal{CN}(\mu, \sigma^2)$	Complex normal distribution with mean μ and variance σ^2.
$\odot$	Hadamard product.
$\otimes$	Kronecker product.
$1\{\mathcal{B}\}$	The indicator function takes 1 if its argument Boolean $\mathcal{B}$ is true, otherwise 0.

Acknowledgements

The research presented in this thesis was conducted under the financial support of, in part, the double degree program between JAIST and University of Oulu, in part by the Japan Society for the Promotion of Science (JSPS) Grant under the Scientific Research (C) No. 22560367, in part by KDDI R&D Laboratories, Inc., and in part by Koden Electronics Co., Ltd, Japan.

Chapter 1

Introduction

BROADBAND wireless communication systems are expected to increase both their transmission (TX) rate and their spectral efficiency [1, 2] under the constraints of low TX power and a low computational complexity. For the reliable data transmission, channel parameters must be estimated online in practical systems, since the channel parameters can change in the middle of communications. The training sequence (TS) [3] is, therefore, transmitted together with a data sequence in general. The TS length has to be long enough to perform channel estimation accurately.

However, the TS is an overhead from the viewpoint of data transmission because the TS is a known data at the receiver. Moreover, as shown in Fig. 1.1, transmitters need to transmit cyclic prefix (CP) and guard interval (GI) sections which are necessary for receivers to perform low-complexity frequency domain equalization (FDE) [4] and to avoid inter-block-interference (IBI), respectively. Notice that the CP and GIs are also overheads for data transmission.

The objective of this thesis is, in summary, to ameliorate the trade-off between the spectral efficiency and the receiver performance. Specifically, this thesis aims to eliminate or reduce the overheads for data transmission by enhancing reception performance with reasonable computational complexities. This trade-off is detailed in Section 1.2 after briefly reviewing the background of this thesis.

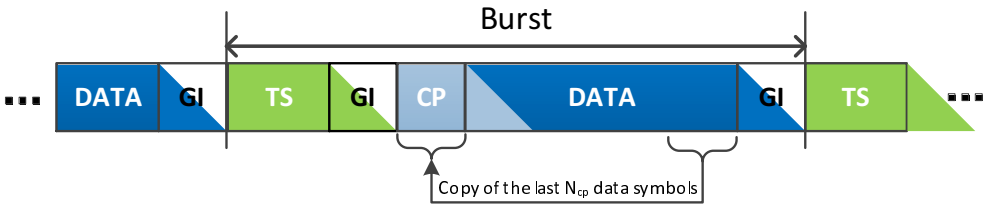

Fig. 1.1: An example of TX burst format. The triangle parts illustrate the IBIs.

1.1 Background

An overview of channel parameters' property is shown as background knowledge. Necessity of the above-mentioned overheads is, then, discussed.

1.1.1 Channel parameters in broadband wireless communications

The higher the symbol rate the data signals are transmitted with, the broader the bandwidth the received signals are spread over. In broadband wireless communications, thereby, the received signals can experience frequency selective fading. It should be noticed that the frequency selectivity is caused by multipath propagation, the propagation distance or time of which can be determined by geometric properties of propagation paths. By observing the frequency selectivity in the temporal domain, therefore, the channel parameters can be described as complex coefficients of a finite impulse response (FIR) filter, the order of which depends on the TX bandwidth. The channel parameter is, hence, referred to as channel impulse response (CIR).

1.1.2 The ISI and IBI problems

Received signals in broadband wireless communications can suffer from the problem of inter-symbol-interference (ISI) due to multipath propagation. Notice that the ISI can leak to the next TX block, as illustrated in Fig. 1.1. The IBI can, hence, be regarded as a block-wise ISI problem. This subsection, therefore, focuses on the ISI problem.

For the sake of conciseness, let us consider a signal model in a single-input single-output (SISO) system. Suppose that a length N_d data $\mathbf{x}_d$ is transmitted over a length W channel $\mathbf{h}$, the length $\tilde{N}_d = N_d + W - 1$ received

signal $\mathbf{y}_d$ can be described as a convolution of $\mathbf{h}$ and $\mathbf{x}_d$ because the received signal can be assumed as output of an FIR filter. Concretely,

$$y_d(n) = z(n) + \sum_{j=1}^{W} h(j)x(n-j+1), \tag{1.1}$$

where $y_d(n)$, $h(n)$ and $x_d(n)$ denote the n-th entry of the vectors $\mathbf{y}_d$, $\mathbf{h}$ and $\mathbf{x}_d$, respectively. The noise $z(n)$ follows the complex normal distribution $\mathcal{CN}(0, \sigma_z^2)$ with mean μ and variance σ_z^2. The problem of the ISI is that the n-th received sample $y_d(n)$ is interfered by the past $W-1$ transmitted signals: $x_d(n-1), \cdots, x_d(n-W+1)$.

1.1.3 Strategies for the ISI problem

We can take either of the following two strategies for the ISI problem.

- *ISI avoidance*: ISI can be avoided by decreasing the symbol rate of TX signals. In other words, we may transmit signals every W symbol so that the interference becomes $[x(n-1), \cdots, x(n-W+1)] = [0, \cdots, 0]$. Alternatively, a symbol may be assigned a long duration so that the distortion due to the ISI becomes very minor.

- *ISI cancelation*: in a turbo receiver framework, feedback information from a decoder is available. The receiver can hence cancel the interference by using estimates of $x(n-1), \cdots, x(n-W+1)$.

Based on the first strategy, orthogonal frequency division multiple access (OFDMA) avoids the ISI problem by using a low orthogonal frequency division multiplexing (OFDM) symbol rate and CP-transmission to be mentioned later. An OFDMA receiver requires, therefore, a low computational complexity because there is no ISI. However, an OFDMA transmitter has another problem: its radio frequency (RF) amplifier has to satisfy a high peak-to-average-power ratio (PAPR) requirement which needs a high back-off power (e.g., [4, 5]).

For long battery life of mobile terminals, thereby, single-carrier transmission is preferable, if an uplink receiver–usually a base station (BS)–is capable of performing the ISI cancelation. Turbo equalization (e.g., [6, 7]) is known as one of the most promising techniques to solve the ISI problem. In the aspect of pursuing a low computational complexity system, a criticism is

that turbo equalization techniques centralize the complexity required for the whole system into the receiver side. However, by assuming CP-transmission, frequency domain turbo equalization (FD-TEQ) algorithm [8–11] can reduce the computational complexity significantly. It is well-known that order of complexity of the FD-TEQ is the same as that of the OFDMA receiver.

1.1.4 Necessity of CP-transmission for circulant structured channels

In either case of an OFDMA or single-carrier frequency division multiple access (SC-FDMA) system, low-complexity FDE algorithms assume CP-transmission. A mathematical background is summarized as follows: (1.1) in a vector form is,

$$\mathbf{y}_d \;=\; \mathbf{H}\mathbf{x}_d + \mathbf{z}, \tag{1.2}$$

with $\mathbf{z} = [z(1), \cdots, z(\tilde{N}_d)]^{\mathrm{T}}$, where $\mathbf{H} \in \mathbb{C}^{\tilde{N}_d \times N_d}$ is a Toeplitz matrix whose first column is $[\mathbf{h}^{\mathrm{T}}\ \mathbf{0}_{1\times(N_d-1)}]^{\mathrm{T}}$. The CP which is a copy of the last W TX data symbols can be denoted as $\mathbf{x}_{\mathrm{CP}} = \mathbf{x}_d|_{N_d-W+1:N_d}$. As is well-known, the identity (e.g., [9, 10])

$$(\tilde{\mathbf{H}}\tilde{\mathbf{x}}_d)|_{W:\tilde{N}_d} \;=\; \mathbf{H}_c\mathbf{x}_d \tag{1.3}$$

holds for $\tilde{\mathbf{x}}_d = [\mathbf{x}_{\mathrm{CP}}^{\mathrm{T}}\ \mathbf{x}_d^{\mathrm{T}}]^{\mathrm{T}}$, where $\tilde{\mathbf{H}} \in \mathbb{C}^{(\tilde{N}_d+W-1)\times\tilde{N}_d}$ is a Toeplitz matrix whose first column is $[\mathbf{h}^{\mathrm{T}}\ \mathbf{0}_{1\times(\tilde{N}_d-1)}]^{\mathrm{T}}$, however, $\mathbf{H}_c \in \mathbb{C}^{N_d\times N_d}$ is a circulant matrix, the first column of which is $\tilde{\mathbf{h}} = [\mathbf{h}^{\mathrm{T}}\ \mathbf{0}_{1\times(N_d-W)}]^{\mathrm{T}}$. Specifically, they can be written as

$$\tilde{\mathbf{H}} \;=\; \begin{bmatrix} h(1) & & & \\ \vdots & h(1) & & \\ h(W) & \vdots & \ddots & \\ & h(W) & \vdots & h(1) \\ & & \ddots & \vdots \\ & & & h(W) \end{bmatrix} \;\in\; \mathbb{C}^{(\tilde{N}_d+W-1)\times\tilde{N}_d}$$

and

$$\mathbf{H}_c \;=\; \begin{bmatrix} \tilde{h}(1) & \tilde{h}(N_d) & \dots & \tilde{h}(2) \\ \tilde{h}(2) & \tilde{h}(1) & & \tilde{h}(3) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{h}(N_d) & \tilde{h}(N_d-1) & \dots & \tilde{h}(1) \end{bmatrix} \;\in\; N_d \times N_d,$$

where $\tilde{h}(i)$ denotes the i-th entry of the vector $\tilde{\mathbf{h}}$. Note that the operation on the left-hand side (LHS) of (1.3) is a so-called *CP-removal* at a receiver.

By a property of the circulant matrix [12], the matrix product

$$\mathbf{\Xi}_c = \mathbf{F}\mathbf{H}_c\mathbf{F}^{\mathrm{H}}$$

is a diagonal matrix, the diagonal entry of which is $\mathbf{F}\tilde{\mathbf{h}}$, where $\mathbf{F}$ denotes an $N_d \times N_d$ discrete Fourier transform (DFT) matrix whose $(i+1, j+1)$-th entry is

$$\exp\left[-2\pi ij\sqrt{-1}/N_d\right]/\sqrt{N_d} \tag{1.4}$$

with integer indexes $0 \le i, j \le N_d - 1$. Practically, the matrix multiplication with $\mathbf{F}$ can be computed by using a fast Fourier transform (FFT) algorithm (e.g., [13]), the complexity order of which is $\mathcal{O}(N_d \log N_d)$. The complexity order needed to equalize the received signals in the diagonal structured channel $\mathbf{\Xi}_c$ is $\mathcal{O}(N_d)$ since equalization is performed with element-by-element operations. FDE algorithms can, hence, reduce the computational complexity significantly.

1.1.5 Required lengths of the TS and the GI

In a turbo receiver framework, a soft replica of the transmitted sequence can be generated using feedback information from the decoder. Turbo channel estimation techniques [14–16] perform channel estimation using the TS and the soft replica jointly. They can, hence, improve estimation accuracy even with a short TS. However, we cannot eliminate the TS completely since channel estimation has to be performed with the TS only in the first iteration. It should be noted that, moreover, the received signals corresponding to the transmitted TS should not suffer from IBI. The GIs in Fig. 1.1 are, thereby, needed to avoid the IBIs in the received TS.

We look into the required TS length N_t and GI length N_G by observing a basic example of least squares (LS) channel estimation (e.g., [16, 17]) in the

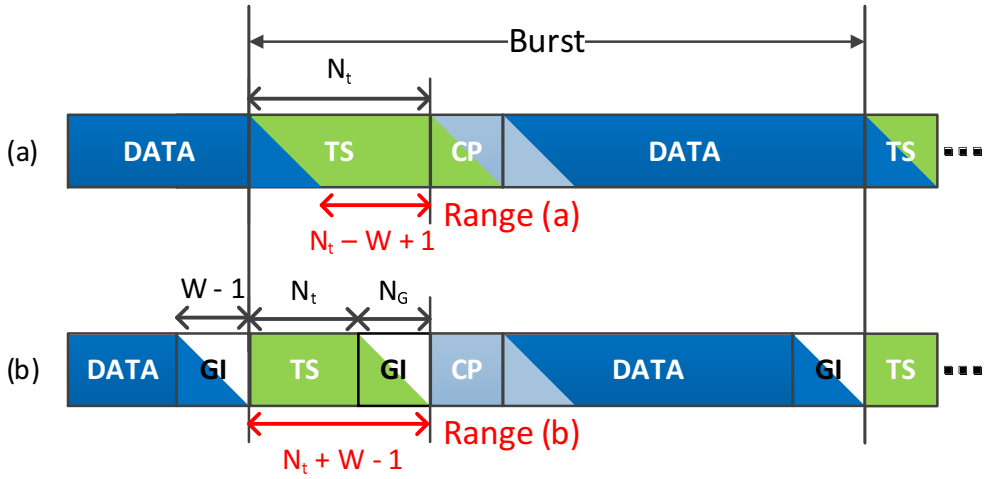

Fig. 1.2: Received TS ranges without the GI (a) and with the GI (b). The triangle parts illustrate the IBIs. The received TS (a) suffers from the IBIs, whereas the one (b) avoids the problem by the GIs.

SISO system. The received signal corresponding to the transmitted TS can be described as

$$\mathbf{y}_t \;=\; \mathbf{X}_t\mathbf{h} + \mathbf{z} \tag{1.5}$$

for the length-W SISO channel $\mathbf{h}$, where $\mathbf{y}_t$ denotes a length $N_t + 2N_G - W + 1$ input signals, the range of which is defined as either (a) with $N_G = 0$ or (b) with $N_G = W - 1$ in Fig. 1.2. Let $\mathbf{X}_t \in \mathbb{C}^{(N_t+2N_G-W+1)\times W}$ be a Toeplitz matrix whose first column vector is either $\mathbf{x}_t|_{W:N_t}$ for the range (a) or $[\mathbf{x}_t^{\mathrm{T}}\ \mathbf{0}_{1\times(W-1)}]^{\mathrm{T}}$ for the range (b), where $\mathbf{x}_t$ denotes a length N_t TS. The LS estimate $\hat{\mathbf{h}}$ for (1.5) can be described generally as

$$\hat{\mathbf{h}} \;=\; \mathbf{X}_t^{\dagger}\mathbf{y}_t, \tag{1.6}$$

where $\mathbf{X}_t^{\dagger}$ denotes the Moore-Penrose pseudoinverse of the matrix $\mathbf{X}_t$. Notice that the solution to (1.6) exists for any TS matrix $\mathbf{X}_t$. Nevertheless, it is required that

- the rank of $\mathbf{X}_t$ is greater than W, or equivalently,
- $\mathbf{R}_{\mathbf{X}\mathbf{X}_t} = \mathbf{X}_t^{\mathrm{H}}\mathbf{X}_t$ is invertible,

in order to obtain the length-W channel estimate $\hat{\mathbf{h}}$ with a certain precision.

To confirm the requirement, Fig. 1.3 shows a normalized mean squared error (MSE) performance of the LS estimator (1.6), α_{N_t}MSE, where the normalization factor α_{N_t} denotes $\alpha_{N_t} = N_t/W$. CIRs follow the Vehicular-A (VA) [18] channel model with 30 km/h mobility (VA30). The received TS range (a) without the GI is assumed. The TSs are generated randomly. Notice that $\alpha_{N_t}\text{MSE} \propto \sigma_z^2$ since $\text{MSE} \triangleq \mathbb{E}[\|\hat{\mathbf{h}} - \mathbf{h}\|^2] \propto \sigma_z^2 W/N_t$ is expected for the LS estimator with the *ideally uncorrelated TS* (e.g., [17]), where the ideally uncorrelated TS is referred to as the TS such that $\mathbf{R}_{\mathbf{X}\mathbf{X}_t}/N_t = \mathbf{I}_W$. Therefore, the *asymptotic performance* in Fig. 1.3 is given by the noise variance σ_z^2 which is independent of the parameters W and N_t. We can observe from Fig. 1.3 that the normalized performance with $N_t = W$ does not follow the noise variance σ_z^2 even in the very high signal-to-noise ratio (SNR) regime. In addition, Fig. 1.4 shows the α_{N_t}MSE performance of the LS estimator (1.6), where the received SNR is set at 30 dB. As observed from Fig. 1.4, the α_{N_t}MSE performance deteriorates seriously for the range $N_t < 2W$. This is because the condition number of the matrix $\mathbf{R}_{\mathbf{X}\mathbf{X}_t}$ is much greater than 1 when $N_t < 2W$, and hence, the LS estimate (1.6) using the pseudoinverse computation based on the SVD algorithm becomes inaccurate. In practice, thereby, the LS estimator (1.6) requires $N_t \geq 2W$ in the case the received TS ranges without the GI (a) is assumed.

By assuming that $\mathbf{R}_{\mathbf{X}\mathbf{X}_t}$ is invertible, (1.6) can be rewritten[1] as

$$\hat{\mathbf{h}} \;=\; \mathbf{R}_{\mathbf{X}\mathbf{X}_t}^{-1} \cdot \mathbf{X}_t^{\mathrm{H}} \mathbf{y}_t. \tag{1.7}$$

As mentioned above, in the case of the range (a) without GI, $N_t \geq 2W$ is required so that the matrix $\mathbf{R}_{\mathbf{X}\mathbf{X}_t}$ becomes invertible. In the case of the range (b) with GI, the TS can be minimized to $N_t = W$, however, the length of the GI should be $N_G \geq W - 1$ to avoid IBIs. Notice that a TX format with GI can decrease the total TX power since the GI is a duration transmitted nothing. However, the GI decreases the spectral efficiency. Therefore, this thesis focuses on the TX format (b) with GI hereafter, and, eventually, eliminates the GI by improving channel estimation techniques.

[1]We introduce the LS estimator (1.7) with the matrix inverse because this thesis uses the formulation mainly rather than (1.6) with the Moore-Penrose pseudoinverse, in order to develop channel estimation techniques in MIMO systems.

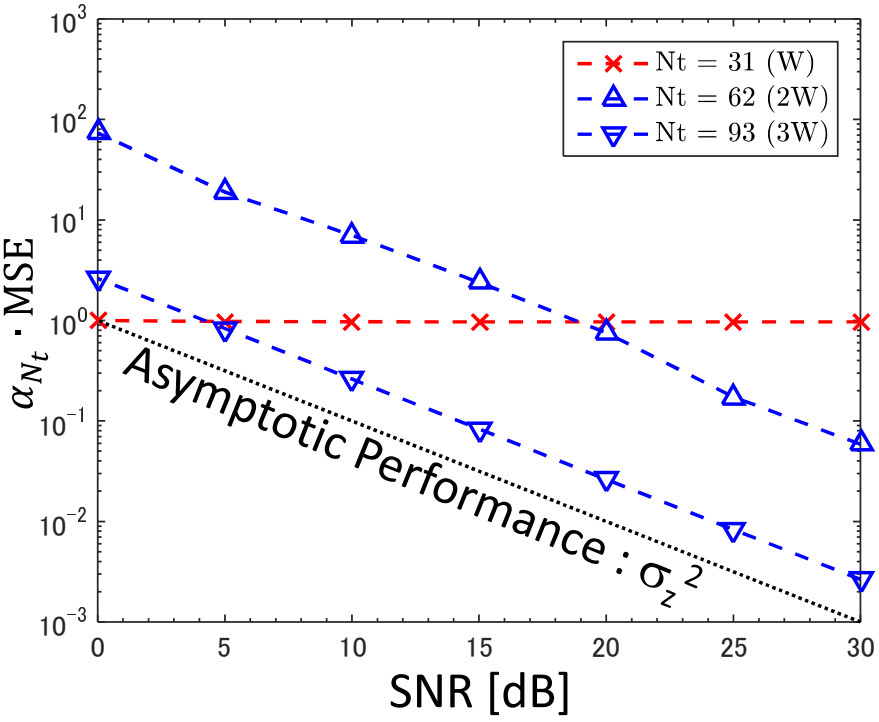

Fig. 1.3: The normalized performance of the LS estimator (1.6): α_{N_t}MSE in the VA30 scenario, where the normalization factor α_{N_t} denotes $\alpha_{N_t} = N_t/W$. The TS range without the GI (a) is assumed. The CIR length W is set at 31.

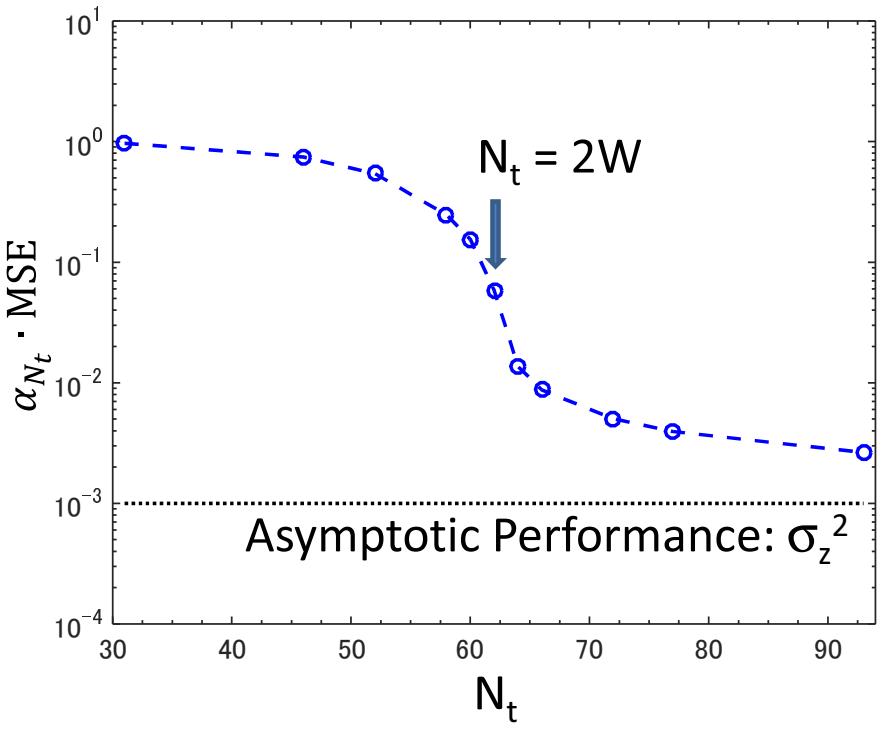

Fig. 1.4: The normalized performance of the LS estimator (1.6): α_{N_t}MSE against different TS lengths N_t. The received SNR is set at 30 dB. The TS range without the GI (a) is assumed. The CIR length W is set at 31.

1.2 Motivation

The trade-off mentioned in the beginning of this chapter can be specifically described by a contradiction between the spectral efficiency and the channel estimation performance.

1.2.1 Definition of the spectral efficiency

Definition 1 (Spectral efficiency)**.** *The spectral efficiency η of the frame format structure is defined as*

$$\eta \;=\; N_{\text{info}}/L_{\text{frm}} \tag{1.8}$$

for single user communications, where N_{info} and L_{frm} denote the number of information bits in a frame and the frame length in symbol, respectively.

The frame length L_{frm} includes the above-mentioned lengths of TS, GI, CP and data sections. By the definition, the spectral efficiency η can be improved by

(η-i) decreasing L_{frm} by reducing the overheads of transmission;

(η-ii) increasing N_{info} by multiple-input multiple-output (MIMO) transmission and/or multi-level modulation techniques.

1.2.2 Channel estimation performance

The MSE performance of the unbiased channel estimation (e.g., [17]) is expected to have a property that

$$\text{MSE}(\sigma_z^2) \;\propto\; \frac{N_{\text{param}}}{N_t}\sigma_z^2 \tag{1.9}$$

under an assumption that the TS is ideally uncorrelated. The MSE performance (1.9) can hence be improved by

($\mathcal{M}$-i) increasing N_t, the TS length;

($\mathcal{M}$-ii) decreasing N_{param}, the number of parameters to be estimated.

Notice that $N_{\text{param}} \leq WN_TN_R$, where N_T and N_R denote the number of transmit (Tx) and receive (Rx) antennas, respectively. This is because the CIR for a Tx-Rx link can be assumed as an FIR filter of order W. Therefore, **($\mathcal{M}$-ii)** is possible when not all WN_TN_R parameters are dominant.

1.2.3 Trade-off between the spectral efficiency and the receiver performance

The TS length should be shortened for **(η-i)**, however, this deteriorates the MSE performance due to **($\mathcal{M}$-i)**. Moreover, MIMO transmission techniques for **(η-ii)** increases N_{param} due to spatial multiplexing, nevertheless, this contradicts **($\mathcal{M}$-ii)**.

1.2.3.1 An example of the trade-off between the throughput performance and the TS length

Fig. 1.5 shows the throughput performance of the SISO system (1.5) in the VA30 scenario. Two TS lengths $N_t = W, 3W$ are examined. The other parameters are assumed as $(N_{\text{CP}}, N_d, N_{\text{G}}, W, N_{\text{turbo}}, T_s) = (32, 512, 31, 31, 1, (7\times 10^6)^{-1})$, where N_{turbo} and T_s denote the maximum number of turbo iterations[2] and the symbol rate in second, respectively. A half-rate convolutional code is used. Further details of the system is postponed to Section 2.1.

As observed from Fig. 1.5, the throughput performance with LS channel estimation is degraded from that with known CIR $\mathbf{h}$ in the moderate $\mathrm{E_b/N_0}$ regime. The throughput can be improved by using a long TS. The asymptotic throughput performance in the high $\mathrm{E_b/N_0}$ regime can, however, be decreased according to the TS length N_t.

This observation can be supported from Fig. 1.6 which shows the throughput performance against different TS lengths N_t. As depicted in Fig. 1.6, the best throughput performance with LS channel estimation can be achieved with $N_t = 62$ $(2W)$ when $\mathrm{E_b/N_0} = 15$ dB. Comparing the throughput performances at $\mathrm{E_b/N_0} = 15$ and 30 dB, however, we notice that, in the moderate $\mathrm{E_b/N_0}$ regime, the throughput performance with channel estimation has room for improvement. Therefore, this thesis pursues ameliorating channel estimation performance with the aim of improving the trade-off.

1.2.4 Approaches for improving channel estimation performance

As seen in Section 1.2.2, channel estimation performance (1.9) can be enhanced by the two approaches **($\mathcal{M}$-i)** and **($\mathcal{M}$-ii)**. Turbo channel estimation

[2]Since $N_{\text{turbo}} = 1$, the receiver in this example does not use the feedback information from the decoder.

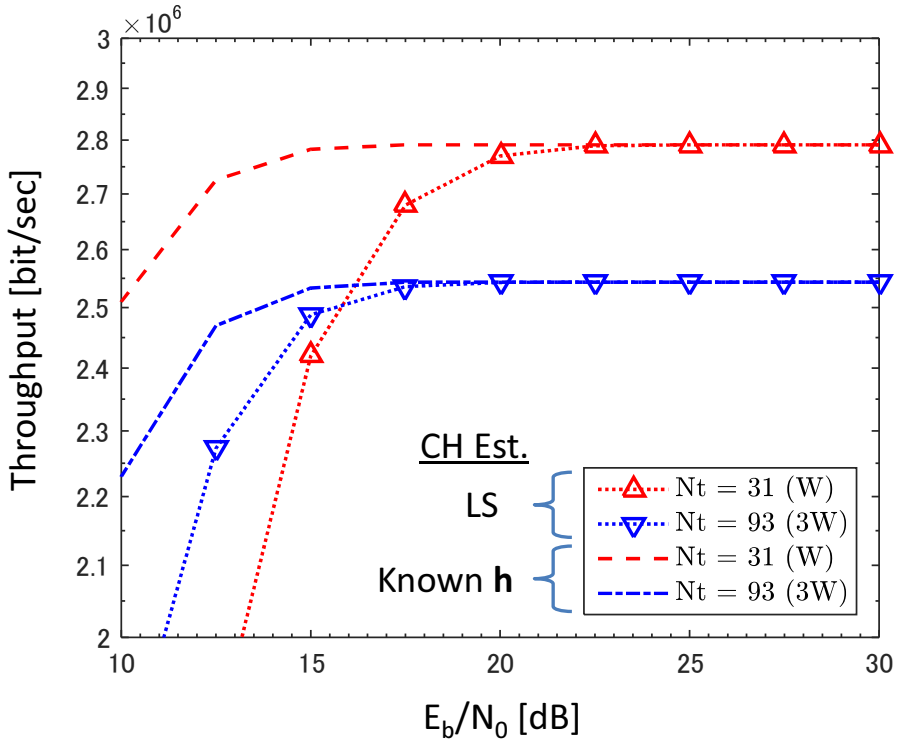

Fig. 1.5: Throughput performance in the SISO VA30 scenario. Two TS lengths $N_t = W, 3W$ are examined.

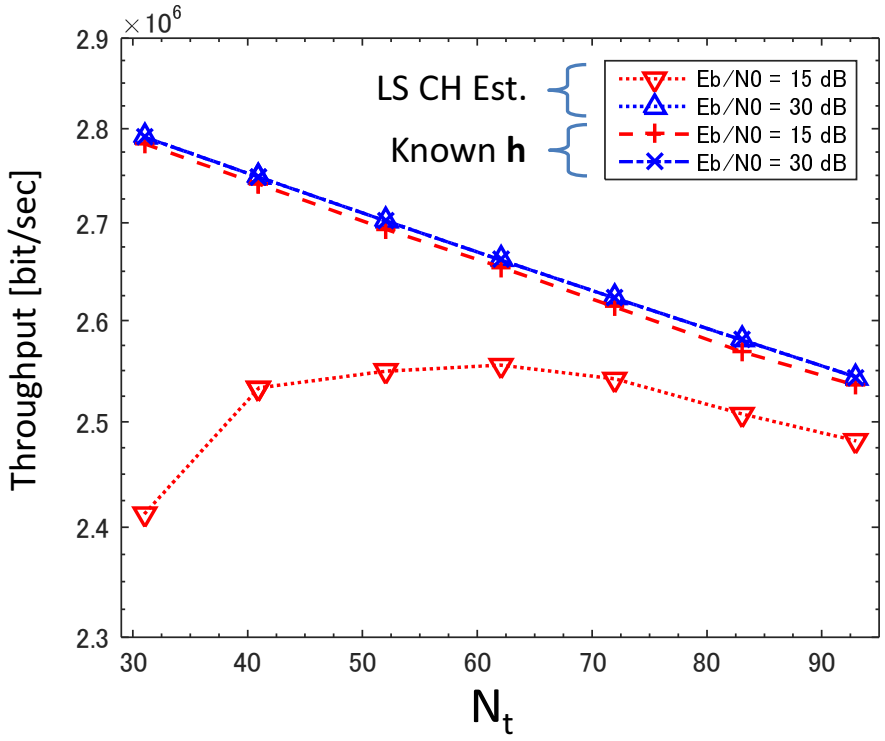

Fig. 1.6: Throughput performance against different N_t setups. The SISO VA30 scenario is assumed. E_b/N_0 is set at 15 or 30 dB.

techniques [14–16] take the first approach (**$\mathcal{M}$-i**) since they extend the TS length virtually by utilizing the log-likelihood ratio (LLR) of transmitted data fed back from the decoder.

The second approach (**$\mathcal{M}$-ii**) is referred to as *reduced-rank* channel estimation [19–21], too. Specifically, subspace-based channel estimation techniques [16,21–26] perform noise reduction under an assumption that the rank r of significant parameters in eigen-domain is less than the CIR length W.

1.3 Thesis Outline

After this introductory chapter, Chapter 2 studies $\ell 1$ regularized channel estimation algorithms in addition to the approaches (**$\mathcal{M}$-i**) and (**$\mathcal{M}$-ii**), aiming to further improve channel estimation performance. Chapter 3 explores, then, spectrally efficient turbo receiving techniques by extending the $\ell 1$ regularized channel estimation algorithms shown in Chapter 2 for frame formats having small overheads. Moreover, Chapter 3 shows a new FDE technique without assuming the CP-transmission. Chapter 4 summarizes concluding remarks of this thesis.

1.4 Contributions

Chapter 2 is described based on the first and second publications shown below. Chapter 3 includes the spectrally efficient turbo receiving techniques presented in the third publication, and provides technical evidence for the fourth patent.

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

2. Y. Takano, M. Juntti, and T. Matsumoto, "Performance of an $\ell 1$ regularized subspace-based MIMO channel estimation with random sequences," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 112–115, Feb 2016.

3. 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.

4. T. Matsumoto, Y. Hatakawa, S. Konishi, Y. Takano, K. Anwar, and T. Matsumoto, "Receiver and receiving techniques", Japanese Patent Application No. 2013-058999., Oct 2014.

Core contributions presented in the above publications are summarized, respectively, as

1. proposals of new techniques such as $\ell 1$ LS, $\ell 1$ minimum mean square error (MMSE) and hybrid channel estimation algorithms, and verification of MSE and bit error rate (BER) performances with the proposed algorithms in intermittent transmission scenarios;

2. clarification of MSE performance with the $\ell 1$ MMSE channel estimation algorithm when random training sequences are assumed;

3. proposals of new channel equalization and channel estimation techniques using a spectrally efficient frame format without the CP and GI sections.

Chapter 2

$\ell 1$ Regularized Channel Estimation Algorithms

Compressive sensing (CS) [27]-based $\ell 1$ regularized channel estimation can improve estimation performance over ordinary $\ell 2$ channel estimation if a CIR observed at a receiver exhibits sparse structure having several tap weights close to zero [28, 29]. This happens often, e.g., in under-water communication channels [30–32]. Broadband wireless channels are, in general, not observed as sparse channels at a receiver due to the effect of Tx[1] and Rx filters required to perform discrete-time processing properly. However, they can be seen as *approximately* sparse channels in a low to moderate SNR regime if the channels follow a typical propagation scenario such as VA or Pedestrian-B (PB) [18]. The dominant path components in such propagation scenarios are, as shown in Fig. 2.1, 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 [33], 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 [16, 23, 24, 26]. This is because the $\ell 2$ MB technique formulated as an MMSE problem improves the MSE performance by utilizing the subspace projection. It can be seen that

[1]We distinguish Tx (transmit) from TX (transmission).

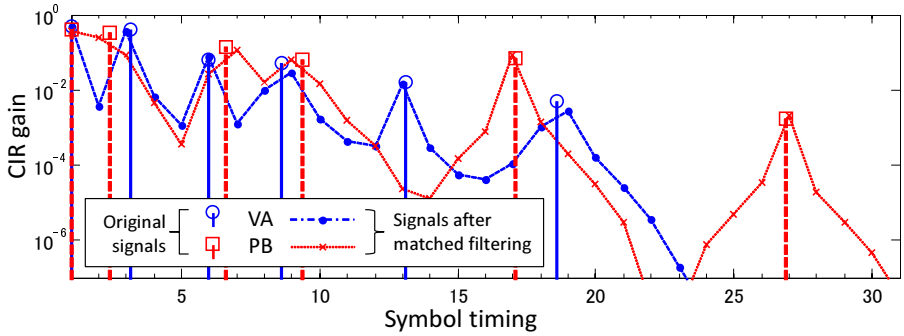

Fig. 2.1: Channel delay profiles of VA and PB channel realizations. We note that the receiver can observe CIRs only as that after the matched filtering. A transmission bandwidth of 7 MHz with a carrier frequency of 2 GHz is assumed. The implementation of the matched filter is described in Section 2.4.

the $\ell 2$ MB technique performs noise *compression* in eigen domain of the signal of interest. Therefore, this chapter 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* TX scenario is focused on. As illustrated in Fig. 2.2, this chapter 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 $\ell 1$ LS and $\ell 2$ MB techniques.

The communication system assumed in this thesis is a turbo receiver framework over broadband 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 [34, 35]. However, channel estimation needed for practical MIMO systems has the problem that the number of the CIR parameters increases due to the spatial multiplexing. Hence, $\ell 1$ regularized channel estimation is expected to improve estimation

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

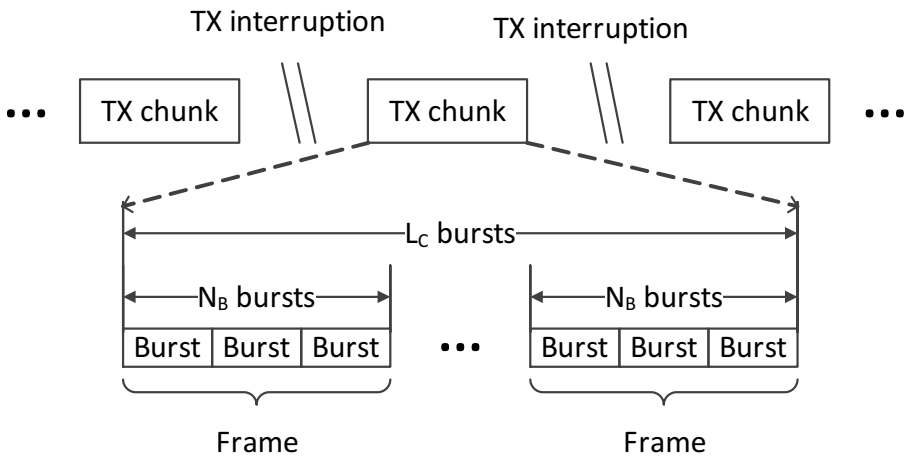

Fig. 2.2: Intermittent TX scenario having arbitrary length TX interruptions. A *TX chunk* is referred to as a continuous data duration which is composed of L_c/N_B frames, where a *frame* is a data unit for forward error correction (FEC). A *burst* is a short data duration, the channel parameters of which are assumed to be constant.

performance in broadband MIMO wireless channels by *compressing* the number of parameters to be estimated. Furthermore, it is shown in [32, 36, 37] that a turbo receiver with an $\ell 1$ regularized channel estimation can achieve a BER gain over that with an ordinary $\ell 2$ channel estimation. However, the channel estimation performance is not addressed in [32, 36, 37]. Therefore, this chapter 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 chapter is organized as follows. Section 2.1 describes the system model assumed in this chapter. Section 2.2 proposes new $\ell 1$ regularized MB and hybrid channel estimation algorithms. Section 2.3 describes analytical performance bounds of the new techniques. Section 2.4 presents results of computer simulations conducted to verify the analytical performance. This chapter is concluded in Section 2.5 with some concluding remarks.

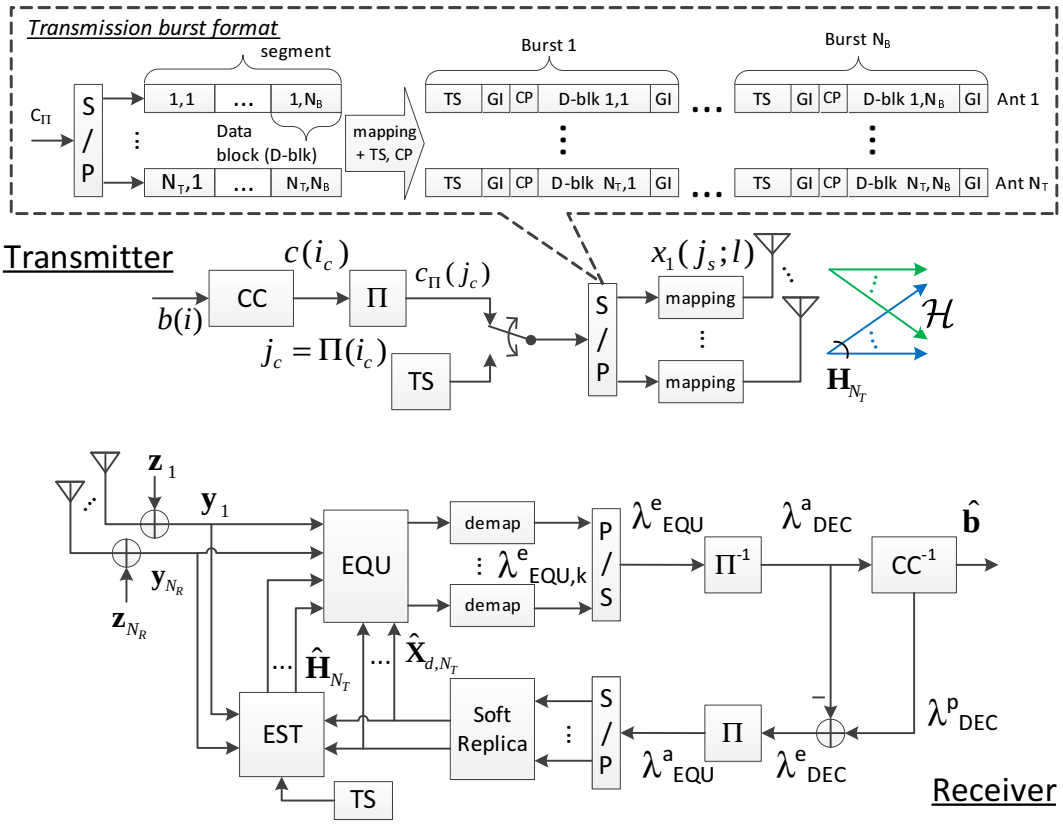

Fig. 2.3: The system model and the transmission burst format assumed in this thesis.

2.1 System Model

This thesis assumes a vertical-Bell laboratories layered space-time (V-BLAST) type spatial multiplexing MIMO system [38] as depicted in Fig. 2.3.

2.1.1 Transmitter

A length N_{info} bit binary data information sequence $b(i)$, $1 \leq i \leq N_{\text{info}}$, 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 $c_\Pi(j_c)$, $1 \leq j_c \leq N_{\text{info}}/R_c$, 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 TS $\mathbf{x}_{t,k}(l)$ and a length-N_{CP} symbol CP, using single carrier signaling, where l denotes the burst timing index. The data symbol length N_d in a burst is defined as $N_d = N_{\mathrm{info}}/(R_c N_T N_B M_b)$. As depicted in Fig. 2.3, the burst format has two length-N_G symbol GIs following the training and the data sequences, respectively, to avoid[4] the IBI problem.

2.1.2 Signal Model

The receiver observes signal sequences $\mathbf{y}_n(l)$ with N_R Rx antennas. The received signal suffers from 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,

$$\mathcal{Y}(l) \;=\; \mathcal{H}(l)\mathcal{X}(l) + \mathcal{Z}, \tag{2.1}$$

where

$$\begin{array}{rcll} \mathcal{Y}(l) & = & [\mathbf{y}_1(l), \cdots, \mathbf{y}_{N_R}(l)]^{\mathrm{T}} & \in \; \mathbb{C}^{N_R \times L_B}, \\ \mathcal{X}(l) & = & [\mathbf{X}_1^{\mathrm{T}}(l), \cdots, \mathbf{X}_{N_T}^{\mathrm{T}}(l)]^{\mathrm{T}} & \in \; \mathbb{C}^{WN_T \times L_B}, \\ \mathcal{H}(l) & = & [\mathbf{H}_1(l), \cdots, \mathbf{H}_{N_T}(l)] & \in \; \mathbb{C}^{N_R \times WN_T}, \\ \mathcal{Z} & = & [\mathbf{z}_1, \cdots, \mathbf{z}_{N_R}]^{\mathrm{T}} & \in \; \mathbb{C}^{N_R \times L_B}, \end{array} \tag{2.2}$$

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 TX stream is

$$\mathbb{E}[\|\mathbf{H}_k(l)\|^2] \;=\; \sigma_{\mathbf{H}}^2 \tag{2.3}$$

[3]For the sake of simplicity, we assume binary modulation in this thesis. However, extension to higher order modulation is straightforward [39].

[4]The GIs can be eliminated by using the chained turbo estimation (CHATES) [40] shown in Chapter 3.

with a constant $\sigma^2_{\mathbf{H}}$. Furthermore, the CIR satisfies a property that the spatial covariance matrix is of full-rank by assuming no unknown interferences [23, 26]:

$$\mathtt{rank}\left\{\mathbb{E}[\mathbf{H}_k(l)\mathbf{H}_k(l)^{\mathrm{H}}]\right\} \;=\; N_R, \tag{2.4}$$

where the operation $\mathtt{rank}\{\mathbf{M}\}$ denotes the rank of matrix $\mathbf{M}$. The noise vector at the n-th Rx antenna $\mathbf{z}_n$ follows $\mathcal{CN}(\mathbf{0}, \sigma_z^2\mathbf{I}_{L_B})$ and has the spatially uncorrelated property: $\mathbb{E}[\mathbf{z}_{n_1}^{\mathrm{H}}\mathbf{z}_{n_2}] = 0$ for $n_1 \neq n_2$.

2.1.2.1 Semi-WSSUS model assumption

As mentioned in the beginning of this chapter, the intermittent TX scenarios are assumed to investigate performance of $\ell 1$ regularized channel estimation techniques. Due to the arbitrary length TX interruptions, we note that CIRs have the following properties in addition to (2.3) and (2.4).

- CIRs in a TX chunk follow the wide-sense stationary uncorrelated scattering (WSSUS) model assumption (e.g., [41]). Hence, it is assumed that the CIRs in a TX chunk are generated from a single channel model such as the PB or VA [18] channel model with a certain doppler frequency (or mobility).
- However, two CIRs in different TX chunks do not always follow the identical channel model, as illustrated in Fig. 2.4.

We refer to the above properties as *semi*-WSSUS model assumption. Moreover, due to the first property, the CIR $\mathcal{H}(l)$ is assumed to be a constant[5] matrix in the L_B symbol duration at the burst timing l. However, $\mathcal{H}(l_1) \neq \mathcal{H}(l_2)$ if $l_1 \neq l_2$.

2.1.3 Receiver

As depicted in Fig. 2.3, the receiver performs channel estimation (EST) jointly over the Rx antennas while also obtaining the extrinsic LLR $\lambda^e_{\mathrm{EQU},k}$ for the k-th TX stream by means of frequency domain soft-cancelation and MMSE (FD/SC-MMSE) MIMO turbo equalization [10] (EQU). The N_T LLRs $\lambda^e_{\mathrm{EQU},k}$ are parallel-to-serial (P/S)-converted to form an extrinsic LLR

[5]The CIR can change very slowly compared to the duration of the burst length L_B.

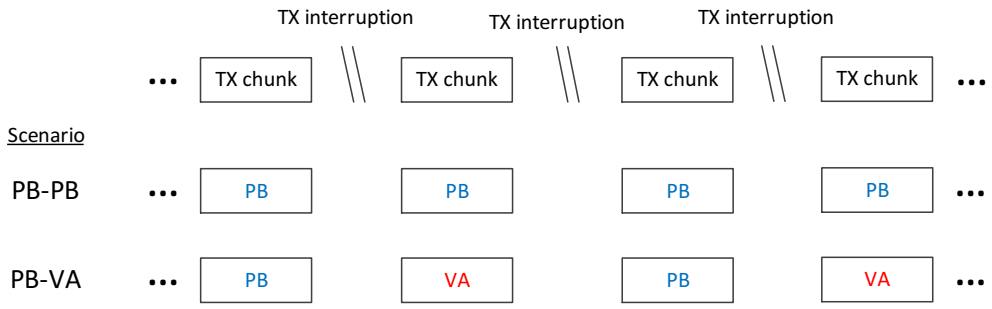

Fig. 2.4: Examples of intermittent TX scenarios following the semi-WSSUS model assumption. CIRs in a TX chunk follow the WSSUS model assumption and, hence, are generated from a single channel model. However, two CIRs in different TX chunks do not always follow the identical channel model. The *PB-PB* scenario shows an example that all CIRs follows the PB channel model, whereas the *PB-VA* scenario shows the case that the channel models change between the TX chunks.

sequence $\lambda^{e}_{\mathrm{EQU}}$ corresponding to the interleaved coded frame $c_{\Pi}(j_c)$ at the transmitter. An *a priori* LLR $\lambda^{a}_{\mathrm{DEC}}$ for the channel decoder (CC^{-1}) is obtained by deinterleaving $\lambda^{e}_{\mathrm{EQU}}$. The channel decoder performs decoding for $\lambda^{a}_{\mathrm{DEC}}$ by using the Bahl, Cocke, Jelinek and Raviv (BCJR) algorithm [42], and outputs the *a posteriori* LLR $\lambda^{p}_{\mathrm{DEC}}$. After several iterations, CC^{-1} outputs the estimates of the transmitted sequence $\hat{\mathbf{b}}$ by making a hard decision on $\lambda^{p}_{\mathrm{DEC}}$. Both EST and EQU utilize the soft replica[6] of the transmitted symbols $\hat{\mathbf{x}}_{d,k}$ which is generated from the equalizer's *a priori* LLR $\lambda^{a}_{\mathrm{EQU}}$. We note that LLR $\lambda^{a}_{\mathrm{EQU}}$ is the interleaved version of the extrinsic LLR $\lambda^{e}_{\mathrm{DEC}}$ which is obtained as $\lambda^{e}_{\mathrm{DEC}} = \lambda^{p}_{\mathrm{DEC}} - \lambda^{a}_{\mathrm{DEC}}$ according to the turbo principle [7].

2.2 Channel Estimation Algorithms

This section proposes new $\ell 1$ regularized MB and hybrid channel estimation algorithms after showing $\ell 1$ regularized LS channel estimation. The computational complexity order required for the new techniques is discussed at the end of this section.

[6]In the case of BPSK, as shown in [39], the i-th entry of $\hat{\mathbf{x}}_{d,k}$ 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.

2.2.1 $\ell 1$ regularized LS channel estimation ($\ell 1$ LS)

2.2.1.1 Problem formulation

By imposing an $\ell 1$ regularizing term to an ordinary $\ell 2$ LS problem, $\ell 1$ LS channel estimation becomes

$$\hat{\mathcal{H}}_{\ell 1}^{LS}(l) = \arg\min_{\mathcal{H}} \mathcal{L}_{td}(l,\mathcal{H}) + \lambda(l)\|\mathcal{H}\|_1 \tag{2.5}$$

with a Lagrange multiplier $\lambda(l)$ [43, 44]. Similar to [26], 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

$$\mathcal{L}_t(l,\mathcal{H}) = \frac{1}{\sigma_z^2}\|\mathcal{Y}_t(l) - \mathcal{H}\mathcal{X}_t(l)\|^2, \tag{2.6}$$

$$\mathcal{L}_d(l,\mathcal{H}) = \frac{1}{\sigma_z^2}\|\mathcal{Y}_d(l) - \mathcal{H}\hat{\mathcal{X}}_d(l)\|_{\mathbf{\Gamma}(l)}^2. \tag{2.7}$$

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 $\mathcal{Y}_d(l) = \mathcal{Y}(l)|_{(\mathfrak{d}+1):(\mathfrak{d}+\tilde{N}_d)}$, where input signal lengths are $\tilde{N}_t = N_t + W$ and $\tilde{N}_d = N_d$. The offset $\mathfrak{d}$ is chosen as $\mathfrak{d} = N_t + N_G + N_{\mathrm{CP}} + W$ so that the received data section avoids IBI from CP. Correspondingly, we define a Toeplitz matrix

$$\mathcal{X}_t(l) = \mathcal{X}(l)|_{1:\tilde{N}_t}. \tag{2.8}$$

$\hat{\mathcal{X}}_d(l)$ is the soft replica of $\mathcal{X}_d(l)$, where we denote

$$\mathcal{X}_d(l) = \mathcal{X}(l)|_{(\mathfrak{d}+1):(\mathfrak{d}+\tilde{N}_d)}. \tag{2.9}$$

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[7] $\Delta\sigma_d^2 \triangleq \sum_{k=1}^{N_T}\mathbb{E}[\|\hat{\mathbf{x}}_{d,k}(l) - \mathbf{x}_{d,k}(l)\|^2]/(N_d N_T)$ and $\mathbf{R}_{\mathcal{H}\mathcal{H}}(l) \triangleq \mathcal{H}(l)\mathcal{H}(l)^{\mathrm{H}}$. The $\ell 1$ regularized LS problem can be solved with the zero-tap detection (ZD) [29] or orthogonal matching pursuit (OMP) [45, 46]-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 $\ell 1$ solvers.

[7] Here we define $\Delta\sigma_d^2$ and $\mathbf{R}_{\mathcal{H}\mathcal{H}}(l)$ using the parameters to be estimated (the transmitted data $\mathbf{x}_{d,k}(l)$ and the CIR $\mathcal{H}$), in order to describe the negative log-likelihood function (2.7) correctly. As shown later in (2.15) and (2.16), however, we exploit the approximations [16, 26] which do not assume the parameters to be estimated.

2.2.1.2 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 thesis refers to the index set $\mathcal{A}$ as *active-set* [47], hereafter. Moreover, we 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 the notation $\mathbf{M}|_{\mathcal{I}}$ describe a submatrix composed of the column vectors in a matrix $\mathbf{M}$, the columns of which are defined by index set $\mathcal{I}$. The $WN_T \times |\mathcal{A}|$ 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 $\hat{\mathcal{H}}_{\mathcal{A}}(l) = \hat{\mathcal{G}}_{\mathcal{A}}(l)\mathbf{P}_{\mathcal{A}}^{\mathrm{T}}$ by minimizing the conditional negative log-likelihood function, given the active-set $\mathcal{A}$, as

$$\hat{\mathcal{G}}_{\mathcal{A}}(l) = \arg\min_{\mathcal{G}} \mathcal{L}_{td}(l, \mathcal{G}\mathbf{P}_{\mathcal{A}}^{\mathrm{T}} \mid \mathcal{A}). \tag{2.10}$$

The problem (2.10) can be seen as an $\ell 2$ LS channel estimation technique by using a temporally restricted (or row-shrunk) training $\mathbf{\Phi}_{t,\mathcal{A}} = \mathbf{P}_{\mathcal{A}}^{\mathrm{T}}\mathcal{X}_t = [\mathcal{X}_t^{\mathrm{T}}|_{\mathcal{A}}]^{\mathrm{T}}$ and data $\hat{\mathbf{\Phi}}_{d,\mathcal{A}} = [\hat{\mathcal{X}}_d^{\mathrm{T}}|_{\mathcal{A}}]^{\mathrm{T}}$ sequences.

Similar to the case of single-input multi-output (SIMO) [26], 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

$$\hat{\mathbf{g}}_{\mathcal{A}} = \mathcal{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}} \cdot \mathtt{vec}\{\mathbf{R}_{\mathcal{Y}\mathbf{\Phi}_{\mathcal{A}}}\} \tag{2.11}$$

with $\mathcal{R}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}} = \mathcal{P}_{\mathcal{A}}^{\mathrm{T}}\mathcal{R}_{\mathcal{X}\mathcal{X}}\mathcal{P}_{\mathcal{A}}$ 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

$$\mathcal{R}_{\mathcal{X}\mathcal{X}} = \mathbf{R}_{\mathcal{X}\mathcal{X}_t}^{\mathrm{T}} \otimes \mathbf{I}_{N_R} + \hat{\mathbf{R}}_{\mathcal{X}\mathcal{X}_d}^{\mathrm{T}} \otimes \hat{\mathbf{\Gamma}}, \tag{2.12}$$

$$\mathbf{R}_{\mathcal{Y}\mathcal{X}} = \mathbf{R}_{\mathcal{Y}\mathcal{X}_t} + \hat{\mathbf{\Gamma}}\mathbf{R}_{\mathcal{Y}\mathcal{X}_d}, \tag{2.13}$$

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

$$\hat{\mathbf{\Gamma}} = \sigma_z^2 \left(\sigma_z^2\mathbf{I}_{N_R} + \Delta\hat{\sigma}_d^2\hat{\mathbf{R}}_{\mathcal{H}\mathcal{H}}\right)^{-1}, \tag{2.14}$$

with the approximations [16, 26]

$$\Delta\hat{\sigma}_d^2 \approx \sigma_x^2 - \sum_{k=1}^{N_T} \|\hat{\mathbf{x}}_{d,k}(l)\|^2/(N_d N_T), \tag{2.15}$$

$$\hat{\mathbf{R}}_{\mathcal{H}\mathcal{H}} \approx \hat{\mathcal{H}}^{(i-1)}(\hat{\mathcal{H}}^{(i-1)})^{\mathrm{H}}, \tag{2.16}$$

where $\hat{\mathcal{H}}^{(i-1)}$ is the channel estimate obtained by the previous (i-1)-th[8] turbo iteration. Finally, the solution to (2.10) 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 $\mathbf{x} = \mathtt{vec}\{\mathtt{mat}_N\{\mathbf{x}\}\}$.

2.2.1.3 The $\ell 1$ LS with adaptive active-set detection (AAD)

Based on the MSE performance analysis shown in Section 2.3.1, a new ZD-type algorithm, AAD, can be formulated as

$$\hat{\mathcal{A}} = \arg\min_{\mathcal{A}} \|\hat{\mathcal{G}}\mathbf{P}_{\mathcal{A}}^{\mathrm{T}} - \mathcal{H}\|^2, \tag{2.17}$$

where $\hat{\mathcal{G}}$ is the LS estimate given by (2.11). It should be noted that the problem (2.17) is equivalent to

$$\hat{\mathcal{A}} = \arg\min_{\mathcal{A}} \mathcal{L}_{td}(l, \mathcal{G}\mathbf{P}_{\mathcal{A}}^{\mathrm{T}} \mid \mathcal{G}) \tag{2.18}$$

by Appendix A in [24], when the CIRs unsupported with $\hat{\mathcal{A}}$ are very minor and both the training and data signals are ideally uncorrelated sequences.

We can solve (2.17) 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 (2.17) with reasonable computational complexity. In summary, Algorithm 1 solves the problems (2.10) and (2.17) alternately in N_{AAD} iterations. First of all, a possible solution to the problem (2.17) 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

$$\hat{\mathbf{d}}_{\mathcal{H}}^{[n]} = \mathbf{P}_{[n]}\ \mathtt{diag}\{\hat{\mathcal{G}}_{[n]}^{\mathrm{H}} \cdot \hat{\mathcal{G}}_{[n]}\}, \tag{2.19}$$

[8]For the first turbo iteration, $i = 1$, the term $\mathbf{R}_{\mathcal{H}\mathcal{H}}$ is discarded in (2.14) since the channel estimation is performed with the TS only.

where $\mathbf{P}_{[n]}$ denotes $\mathbf{P}_{\mathcal{A}_{[n]}}$. As detailed in Appendix 2.A, the active-set can be detected by

$$\mathcal{A}_{[n+1]} = \left\{ j \;\middle|\; \hat{d}^{[n]}_{\mathcal{H},j} > \left(f(\sigma_z^2, \mathcal{A}_{[n]}) + |\Delta \hat{\mathbf{d}}^{[n]}_{\mathcal{H}}| \right) / |\mathcal{A}_{[n]}|,\ j \in \mathcal{A}_{[n]} \right\}, \tag{2.20}$$

where $\hat{d}^{[n]}_{\mathcal{H},j}$ denotes the j-th entry of $\hat{\mathbf{d}}^{[n]}_{\mathcal{H}}$ and we define $f(\sigma_z^2, \mathcal{A}) = \sigma_z^2 \mathrm{tr}\{\mathcal{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}}\}$. The absolute error of the delay profile estimation can also be approximated by $|\Delta \hat{\mathbf{d}}^{[n]}_{\mathcal{H}}| \approx f(\sigma_z^2, \mathcal{A}_{[n]})$. This is because, as shown in Section 2.3.1, $f(\sigma_z^2, \mathcal{A})$ 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 (2.10) is then solved at the step 7. Algorithm 1 obtains a possible estimate $\hat{\mathcal{G}}_{[n+1]}$ via (2.11) 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) [48] as a stopping tool of the iteration. Suppose that the CIR estimate is described as $\hat{\mathcal{H}} = \hat{\mathcal{G}}_{[n]} \mathbf{P}^{\mathrm{T}}_{[n]}$, the BIC can be defined for the complex matrix normal distribution $\mathcal{L}_{td}(\cdot)$, as

$$\mathtt{BIC}(\hat{\mathcal{G}}_{[n]}) = 2\mathcal{L}_{td}(l, \hat{\mathcal{G}}_{[n]} \mathbf{P}^{\mathrm{T}}_{[n]}) + K_{\mathrm{IC}} \cdot \log(N_{\mathrm{IC}}). \tag{2.21}$$

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 degrees of 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) [29]. 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 2.4, 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}.

Compute $\mathbf{R}_{\mathcal{Y}\mathcal{X}}$ (2.13), $\mathcal{R}_{\mathcal{X}\mathcal{X}}$ (2.12) and $\hat{\mathbf{\Gamma}}$ (2.14).
Obtain the $\ell 2$ LS estimate $\hat{\mathcal{G}}_{[0]} = \mathtt{mat}_{N_R}\{\hat{\mathbf{g}}_{\mathcal{A}_{[0]}}\}$ by (2.11) with $\mathcal{A}_{[0]} = \{1, \cdots, WN_T\}$.
$\beta(0) = \mathtt{BIC}(\hat{\mathcal{G}}_{[0]})$ by (2.21).
for $n = 0$ to $N_{\mathrm{AAD}} - 1$ **do**
  Update the delay profile estimate $\hat{\mathbf{d}}_{\mathcal{H}}^{[n]}$ by (2.19).
  Detect the active-set $\mathcal{A}_{[n+1]}$ by (2.20).
  Obtain an estimate $\hat{\mathcal{G}}_{[n+1]} = \mathtt{mat}_{N_R}\{\hat{\mathbf{g}}_{\mathcal{A}_{[n+1]}}\}$ by (2.11) with $\mathcal{A}_{[n+1]}$.
  $\beta(n+1) = \mathtt{BIC}(\hat{\mathcal{G}}_{[n+1]})$ by (2.21).
  if $\beta(n+1) \geq \beta(n)$ **then**
    Let $n = n - 1$ and terminate the iteration.
  end if
end for

Output: $\hat{\mathcal{H}}_{\ell 1}^{LS} = \hat{\mathcal{G}}_{[n+1]}\mathbf{P}_{\mathcal{A}_{[n+1]}}^{\mathrm{T}}$.

2.2.2 $\ell 1$ regularized multi-burst channel estimation ($\ell 1$ MB)

2.2.2.1 Problem formulation

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

$$\hat{\mathcal{H}}_{\ell 1}^{MB}(l) = \arg\min_{\mathcal{H}(l)} \mathop{\mathbb{E}}_{j \in \mathcal{J}_{L_M}(l)} \left[\mathcal{L}_{td}(j, \mathcal{H}(j)) + \lambda(j)\|\mathcal{H}(j)\|_1\right], \tag{2.22}$$

where the operation $\mathbb{E}_{j\in\mathcal{J}}[a(j)]$ is the expectation of sequence $a(j)$ defined by $\frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} a(j)$ for the argument index set $\mathcal{J}$. The 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, which denotes the last L_M burst timings from the latest timing l. To perform the principal component analysis (PCA) correctly, L_M is required to satisfy $L_M \geq W/N_R$.

This problem (2.22) can be solved by using the same concept as the simplified component technique-LASSO (SCotLASS) [49] which is a version of the least absolute shrinkage and selection operator (LASSO) [50]. A challenge

of SCotLASS-based algorithms is that there is no certain method to determine $\lambda(j)$. Notice that the number of possible active-sets corresponding to $\lambda(j)$ becomes $\sum_{k=0}^{WN_T} \binom{WN_T}{k}$, where the operation $\binom{n}{k}$ denotes the binomial coefficient. In practice, however, a receiver is very difficult to examine all possible $\sum_{k=0}^{WN_T} \binom{WN_T}{k}$ active-sets. We hence relax the problem (2.22) by introducing an assumption that $\lambda(j)$ can be approximately specified by a CIR length constraint. The problem (2.22) can then be reduced into at most W problems without $\ell 1$ regularization.

Since CIRs can be assumed as the output of an 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

$$\mathcal{A}_{[w]} = \bigcup_{k'=0}^{N_T-1} \{(1+k'W) : (w+k'W)\} \tag{2.23}$$

for $1 \leq w \leq W$. With (2.23), the problem (2.22) can be decomposed into at most W problems without $\ell 1$ regularization, as

$$\hat{\mathcal{H}}_{[w]}^{MB}(l) = \underset{\mathcal{H}_{[w]}(l,\boldsymbol{\Theta})}{\arg\min} \underset{j \in \mathcal{J}_{L_M}(l)}{\mathbb{E}} \left[\mathcal{L}_{td}(j, \mathcal{H}_{[w]}(j, \boldsymbol{\Theta})) \right], \tag{2.24}$$

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

$$\mathbf{H}_k(l, \theta_k) \;=\; \mathbf{B}_k(l) \mathbf{U}_k^{\mathrm{H}}, \tag{2.25}$$

when the CIR follows the subspace channel model assumption [23]. The parameter $\boldsymbol{\Theta}$ describes the CIR models (2.25) for N_T Tx streams in a vector as $\boldsymbol{\Theta} = [\theta_1^{\mathrm{T}}, \cdots, \theta_{N_T}^{\mathrm{T}}]^T$, where $\theta_k = [\theta_{\mathbf{B},k}^{\mathrm{T}}, \mathtt{vec}\{\mathbf{U}_k\}^{\mathrm{T}}]^{\mathrm{T}}$. 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 $\mathbb{K}_{j \in \mathcal{J}_{L_M}(l)}[\mathbf{H}_k(j)]$, where the operation $\mathbb{K}_{j \in \mathcal{J}}[\cdot]$ denotes $\mathbb{K}_{j \in \mathcal{J}}[\mathbf{A}(j)] = \frac{1}{L_M} \sum_{j \in \mathcal{J}} \mathbf{A}(j)^{\mathrm{H}} \mathbf{A}(j)$ for the argumented matrix sequence $\mathbf{A}(j)$ and the index set $\mathcal{J}$.

2.2.2.2 $\ell 1$ MB algorithm

By Appendix A in [24], the problems (2.24) is equivalent to minimizing $\Psi_{[w]}(l,\boldsymbol{\Theta}) = \sum_{k=1}^{N_T} \Psi_{[w],k}(l,\theta_k)$ with

$$\Psi_{[w],k}(l,\theta_k) = \mathop{\mathbb{E}}_{j\in\mathcal{J}_{L_M}(l)} \left[\left\| \hat{\tilde{\mathbf{g}}}^{LS}_{[w],k}(j) - \tilde{\mathbf{g}}_{[w],k}(j,\theta_k) \right\|^2 \right], \tag{2.26}$$

where we define, for the noise whitening,

$$\begin{aligned} \hat{\tilde{\mathbf{g}}}^{LS}_{[w],k}(j) &= \bar{\mathbf{Q}}_{[w],kk} \cdot \hat{\mathbf{g}}^{LS}_{[w],k}(j) \\ &\quad + \sum_{i=k+1}^{N_T} \bar{\mathbf{Q}}_{[w],ki} \left\{ \hat{\mathbf{g}}^{LS}_{[w],i}(j) - \mathbf{g}_{[w],i}(j,\theta_i) \right\} \qquad (2.27) \\ \tilde{\mathbf{g}}_{[w],k}(j,\theta_k) &= \bar{\mathbf{Q}}_{[w],kk} \cdot \mathbf{g}_{[w],k}(j,\theta_k) \qquad (2.28) \end{aligned}$$

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 $\bar{\mathcal{R}}_{\boldsymbol{\Phi}\boldsymbol{\Phi}_{[w]}}$ (2.29). It should be noted that (2.26) is based on the approximation[9] that

$$\bar{\mathcal{R}}_{\boldsymbol{\Phi}\boldsymbol{\Phi}_{[w]}} \triangleq \mathbb{E}\left[\mathcal{P}^{\mathrm{T}}_{[w]} \mathcal{R}_{\mathcal{X}\mathcal{X}}(j) \mathcal{P}_{[w]}\right] \approx \mathcal{P}^{\mathrm{T}}_{[w]} \mathcal{R}_{\mathcal{X}\mathcal{X}}(j) \mathcal{P}_{[w]} \tag{2.29}$$

for $\forall j \in \mathcal{J}_{L_M}(l)$, where $\mathcal{P}_{[w]} = \mathbf{P}_{\mathcal{A}_{[w]}} \otimes \mathbf{I}_{N_R}$. In other words, the whitening matrix $\bar{\mathbf{Q}}_{[w],ki}$ obtained from (2.29) is independent of the burst timing j. Therefore, the w-th active-set is required to be independent of the burst timing j, such as (2.23).

As shown in Section III in [24], the minimization problems of (2.26) for $1 \le w \le W$ are solvable if we reduce them by descending order $k = N_T, \cdots, 1$. Moreover, by following Section IV-C in [26], the solution $\hat{\mathbf{g}}^{MB}_{[w],k}(l)$ that minimizes (2.26) can be obtained by the PCA for the covariance matrix $\mathbb{K}_{j\in\mathcal{J}_{L_M}(l)}[\mathtt{mat}_{N_R}\{\hat{\tilde{\mathbf{g}}}^{LS}_{[w],k}(j)\}]$, where approximations $\mathbf{g}_{[w],i}(l,\theta_i) \approx \hat{\mathbf{g}}^{MB}_{[w],i}(l)$ for $i > k$ are used in (2.27). Correspondingly, the solution to (2.24) is described as

$$\hat{\mathcal{H}}^{MB}_{[w]}(l) = [\hat{\mathbf{G}}^{MB}_{[w],1}(l), \cdots, \hat{\mathbf{G}}^{MB}_{[w],N_T}(l)]\mathbf{P}^{\mathrm{T}}_{[w]}$$

[9]The conventional $\ell 2$ MB techniques [16, 23, 24, 26] also assume (2.29) with $\mathcal{P}_{[w]} = \mathbf{I}_{WN_TN_R}$.

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 (2.22) 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) [51]:

$$\mathtt{AIC}(\hat{\mathcal{H}}^{MB}_{[w]}) = 2\mathcal{L}_{td}(l, \hat{\mathcal{H}}^{MB}_{[w]}) + 2K_{\mathrm{IC}},$$

where the number of free parameters is modified as $K_{\mathrm{IC}} = 2\sum_{k=1}^{N_T} N_R r_k^{[w]}$ so that it describes the number of burst-dependent parameters in the CIR model (2.25). The rank $r_k^{[w]}$ of the temporal subspace is obtained together with the estimate vector $\hat{\mathbf{g}}^{MB}_{[w],k}(l)$ by the PCA, as shown in [24, 26].

This thesis exploits the temporal-subspace only to focus on developing the $\ell 1$ regularized channel estimation techniques. Notice that estimation performance of the MB channel estimation techniques can be further improved by using the spatial-subspace projection [21, 24, 26], too. Under the assumption (2.4) that the CIRs are of spatially full-rank, however, it is expected that the spatial-subspace projection matrix becomes an identity matrix. Nevertheless, the assumption is not always correct in multi-user MIMO systems. The extension of the proposed technique by jointly utilizing the temporal- and spatial-subspace projections is a future work of this thesis.

2.2.3 Hybrid channel estimation

2.2.3.1 Problem statement

Later in Section 2.4.4.2, it is shown that the $\ell 1$ MB channel estimation can improve the tracking error problem. Nevertheless, as discussed in Section 2.2.4, the $\ell 1$ MB channel estimation requires a higher complexity order than that of the ordinary $\ell 2$ MB channel estimation. We thereby propose a new hybrid channel estimation algorithm to improve robustness of the estimate for abrupt channel changes with reasonable complexity.

2.2.3.2 Hybrid algorithm

In summary, the new hybrid algorithm performs the $\ell 1$ LS and the ordinary $\ell 2$ MB channel estimation[10] simultaneously, then selects *better* estimate under the Bayesian information criterion.

[10]The $\ell 2$ MB channel estimation is formulated by (2.22) with $\lambda(j) = 0$.

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 (2.22). 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 $\texttt{BIC}(\hat{\mathcal{H}}_{\ell 1}^{LS}(l)) > \texttt{BIC}(\hat{\mathcal{H}}_{\ell 2}^{MB}(l))$ 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 $\hat{\mathcal{H}}_{\ell 1}^{LS}$ as the output $\hat{\mathcal{H}}^{HB}$ of the hybrid estimation. 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 $L_m < L_G$.

2.2.4 Computational complexity order

The computational complexity orders $\mathcal{O}(\cdot)$ required for the channel estimation techniques investigated in this chapter are summarized in Table 2.1. 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 $\mathcal{O}(W^4 N_T^3 N_R^3)$.

2.2.4.1 The $\ell 1$ LS

The computational complexity orders required for each step in Algorithm 1 and its details are shown in Tables 2.2(a) and (b), respectively. For example,

Algorithm 2 Hybrid channel estimation at the burst timing l

1: Perform the $\ell 1$ LS (2.5) 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 (2.22) with the sliding window $\mathcal{J}_{L_m}(l)$ and $\lambda(j) = 0$. Obtain $\hat{\mathcal{H}}_{\ell 2}^{MB}(l)$.
4: **if** $\texttt{BIC}(\hat{\mathcal{H}}_{\ell 1}^{LS}(l)) < \texttt{BIC}(\hat{\mathcal{H}}_{\ell 2}^{MB}(l))$ **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: $\quad\quad L_m = 0$
8: $\quad$ **end if**
9: **else**
10: $\quad$ **if** $L_m < L_G$ **then**
11: $\quad\quad \hat{\mathcal{H}}^{HB}(l) = \hat{\mathcal{H}}_{\ell 1}^{LS}(l)$
12: $\quad$ **else**
13: $\quad\quad \hat{\mathcal{H}}^{HB}(l) = \hat{\mathcal{H}}_{\ell 2}^{MB}(l)$
14: $\quad$ **end if**
15: **end if**

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 2.2(b). It is, however, dominated by $\mathcal{O}(\{|\mathcal{A}_{[0]}|N_R\}^3) = \mathcal{O}(W^3N_T^3N_R^3)$, where we assume that an $M \times M$ matrix inverse requires the complexity order $\mathcal{O}(M^3)$ [52].

As shown in Table 2.2(a), the complexity order needed for the steps 1 to 3 is dominated by $\mathcal{O}(W^2N_R^2\tilde{N}_{td} + W^3N_T^3N_R^3)$. This is because $W^2N_T^2 > WN_TN_R > N_R^2$ 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 (2.21) may be calculated by using the following equations:

$$\mathcal{L}_t(\hat{\mathcal{G}}_{\mathcal{A}}) = \frac{1}{\sigma_z^2}\|\mathcal{Y}_t - \hat{\mathcal{G}}_{\mathcal{A}}\mathbf{\Phi}_{t,\mathcal{A}}\|_F^2, \tag{2.30}$$

$$\mathcal{L}_d(\hat{\mathcal{G}}_{\mathcal{A}}) = \frac{1}{\sigma_z^2}\|\mathcal{Y}_d - \hat{\mathcal{G}}_{\mathcal{A}}\hat{\mathbf{\Phi}}_{d,\mathcal{A}}\|_{\mathbf{\Gamma}}^2, \tag{2.31}$$

where the $N_R \times |\mathcal{A}|$ CIR estimate matrix $\hat{\mathcal{G}}_{\mathcal{A}}$ is obtained via (2.11) 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 $\mathbf{R}^{-1}_{\mathbf{\Phi\Phi}_{\mathcal{A}_{[n]}}}$ in $f(\sigma_z^2, \mathcal{A})$ is already computed in the previous iteration. Furthermore, the matrix inverse $\mathbf{R}^{-1}_{\mathbf{\Phi\Phi}_{\mathcal{A}_{[n+1]}}}$ can be efficiently updated from $\mathbf{R}^{-1}_{\mathbf{\Phi\Phi}_{\mathcal{A}_{[n]}}}$. 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 $\Delta\mathcal{A}_{[n+1]} = \mathcal{A}_{[n]} \setminus \mathcal{A}_{[n+1]}$. This is because, as shown in [53], 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 [53] straightforwardly, the matrix inverse of its arbitrary rank-N-downsized submatrix[11] can be computed with the complexity order $\mathcal{O}(M^2N+N^3)$.

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 (2.20). 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 $\mathcal{O}(\{W^2N_R^2\tilde{N}_{td} + W^3N_T^3N_R^3\} + \sum_{m=1}^{WN_T}\{m^2N_R^3 + N_R^3 + ((m-1)N_R + N_R^2)\tilde{N}_{td}\}) = \mathcal{O}(W^2(N_T^2N_R + N_R^2)\tilde{N}_{td} + W^3N_T^3N_R^3)$, where $m = WN_T - n$ is used. Especially for $N_{AAD} = 1$, the complexity order is at most $\mathcal{O}(W^2N_R^2\tilde{N}_{td} + W^3N_T^3N_R^3)$ due to $|\mathcal{A}_{[n]}| \leq WN_T$.

2.2.4.2 The $\ell 2$ LS

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

---

[11] 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.

2.2.4.3 The $\ell 1$ MB

The $\ell 1$ MB algorithm performs a set of operations, which are CIR length-shrunk $\ell 2$ LS channel estimation (2.11) 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 (2.11), the matrices $\mathcal{R}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}}$ 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 [53], the complexity order needed for the W matrix inverses $\mathcal{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_{[w]}}$, $1 \leq w \leq W$, is equivalent to that of the single matrix inverse $\mathcal{R}^{-1}_{\mathcal{X}\mathcal{X}}$. 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 complexity for matrix-vector products according to the row (iv) of Table 2.2(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[12] of $\bar{\mathcal{R}}_{\mathbf{\Phi}\mathbf{\Phi}_{[w]}}$, the noise whitening (2.27) and the SVD[13] 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 $\ell 1$ MB is, thereby, dominated by $\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^4 N_T^3 N_R^3) = \mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3) + \mathcal{O}(\sum_{w=1}^{W} w^3 N_T^3 N_R^3)$.

2.2.4.4 The $\ell 2$ MB

The $\ell 2$ MB technique performs the above-mentioned set of operations for the W-th possible solution, only once. Hence, the complexity order needed for

[12] We assume that the Cholesky decomposition for an $M \times M$ matrix requires the complexity order $\mathcal{O}(M^3)$ [52].

[13] As shown in [26], SVD for the $w \times w$ covariance matrices $\mathbb{K}_{j \in \mathcal{J}_{L_M}(l)}[\hat{\tilde{\mathbf{G}}}^{LS}_{[w],k}(j)]$ is performed to find the principal components of the CIR for the k-th TX stream, where $\hat{\tilde{\mathbf{G}}}^{LS}_{[w],k}(l) = \mathtt{mat}_{N_R}\{\hat{\tilde{\mathbf{g}}}^{LS}_{[w],k}(l)\}$. 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 $\mathcal{O}(M^3)$ [52]. Note that the complexity order needed for the covariance matrix $\mathbb{K}_{j \in \mathcal{J}_{L_M}(l)}[\hat{\tilde{\mathbf{G}}}^{LS}_{[w],k}(l)]$ 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 $\mathfrak{K}_l^{L_M} = \mathbb{K}_{j \in \mathcal{J}_{L_M}(l)}[\hat{\tilde{\mathbf{G}}}^{LS}_{[w],k}(j)]$ by omitting the subscripts w and k.

Table 2.1: Computational complexity orders for channel estimation algorithms

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

the $\ell 2$ MB channel estimation is dominated by $\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3) = \mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3) + \mathcal{O}(\sum_{w=W}^{W} w^3 N_T^3 N_R^3)$.

2.2.4.5 The hybrid algorithm

The hybrid algorithm performs the $\ell 1$ LS and $\ell 2$ MB techniques at a time. However, its complexity order is equivalent to that of the $\ell 1$ LS, since

$$\begin{aligned} &\mathcal{O}(\{W^2 (N_T^2 N_R + N_R^2) \tilde{N}_{td} + W^3 N_T^3 N_R^3\} + \{W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3\}) \\ &= \mathcal{O}(W^2 (N_T^2 N_R + N_R^2) \tilde{N}_{td} + W^3 N_T^3 N_R^3). \end{aligned}$$

The complexity order needed for the BIC of the $\ell 2$ 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 $N_{\mathrm{AAD}} = 1$, the complexity *order* required for the hybrid algorithm is the same as that of the $\ell 2$ MB technique, although the number of operations are increased slightly.

2.3 Performance Analysis

2.3.1 MSE performance of the $\ell 1$ LS

MSE performance of the $\ell 1$ LS channel estimation is discussed in the following two subsections: 1) *Analytical MSE given active-set* and 2) *Optimal active-set*. This is because the $\ell 1$ LS channel estimation is formulated as the conditional LS problem (2.10).

Table 2.2: 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), [53]
8:	$\mathcal{O}(\ (\lvert\mathcal{A}_{[n+1]}\rvert\ N_R + N_R^2)\tilde{N}_{td}\ )$	(v, vi)

(b) Details in Table 2.2(a)

	Symbol	Eqn.	Computational complexity order
(i)	$\hat{\mathbf{\Gamma}}$	(2.14)	$\mathcal{O}(W N_T N_R^2 + N_R^3)$
(ii)	$\mathcal{R}_{\mathcal{X}\mathcal{X}}$	(2.12)	$\mathcal{O}(W^2 N_T^2 \tilde{N}_{td})$
(iii)	$\mathbf{R}_{\mathcal{Y}\mathcal{X}}$	(2.13)	$\mathcal{O}(W N_T N_R \tilde{N}_{td})$
(iv)	$\hat{\mathbf{g}}_{\mathcal{A}}$	(2.11)	$\mathcal{O}(\lvert\mathcal{A}\rvert^3 N_R^3 + \lvert\mathcal{A}\rvert^2 N_R^2)$
(v)	$\mathcal{L}_t(\hat{\mathcal{G}}_{\mathcal{A}})$	(2.30)	$\mathcal{O}(\{\lvert\mathcal{A}\rvert N_R + N_R^2\}\tilde{N}_t)$
(vi)	$\mathcal{L}_d(\hat{\mathcal{G}}_{\mathcal{A}})$	(2.31)	$\mathcal{O}(\{\lvert\mathcal{A}\rvert N_R + N_R^2\}\tilde{N}_d + N_R^3)$
(vii)	$\hat{\mathcal{G}}_{\mathcal{A}}^{\mathrm{H}}\hat{\mathcal{G}}_{\mathcal{A}}$	(2.19)	$\mathcal{O}(\lvert\mathcal{A}\rvert^2 N_R)$

2.3.1.1 Analytical MSE given active-set

For a given $\mathcal{A}$, an $N_R \times WN_T$ 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 $\hat{\mathbf{g}}_{\mathcal{A}}$ (2.11). The MSE of the $\ell 1$ LS channel estimate can, thereby, be reduced to

$$\begin{aligned} \mathrm{MSE}(\hat{\mathcal{H}}_{\ell 1}^{LS}, \sigma_z^2, \mathcal{A}) &= \mathbb{E}\left[\|\mathtt{vec}\{\hat{\mathcal{H}}_{\ell 1}^{LS}(l) - \mathcal{H}(l)\}\|^2\right] \\ &= \sigma_z^2 \,\mathrm{tr}\left\{\mathbb{E}[\mathcal{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}}(l)]\right\} + \mathfrak{E}(\mathcal{A}), \end{aligned} \tag{2.32}$$

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

$$\mathfrak{B}(\mathcal{A}, l) = \mathcal{P}_{\mathcal{A}}\mathcal{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}}(l)\mathcal{P}_{\mathcal{A}}^{\mathrm{T}}\mathcal{R}_{\mathcal{X}\mathcal{X}}(l) - \mathbf{I}_{WN_TN_R}.$$

We note that, when $\mathcal{A} = \emptyset$, the MSE of $\hat{\mathcal{H}}_{\ell 1}^{LS}$ 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 $\mathbf{J}_{\mathcal{A}}^{\perp} = \mathbf{I}_{WN_T} - \mathbf{J}_{\mathcal{A}}$ with $\mathbf{J}_{\mathcal{A}} = \mathbf{P}_{\mathcal{A}}\mathbf{P}_{\mathcal{A}}^{\mathrm{T}}$.

2.3.1.2 Optimal active-set

For an active-set $\mathcal{A}$, denote the MSE residual $\Delta_{\ell 1\ell 2}^{LS}(\mathcal{A})$, as

$$\Delta_{\ell 1\ell 2}^{LS}(\mathcal{A}) = \mathrm{MSE}(\hat{\mathcal{H}}_{\ell 1}^{LS}, \sigma_z^2, \mathcal{A}) - \mathrm{MSE}(\hat{\mathcal{H}}_{\ell 2}^{LS}, \sigma_z^2), \tag{2.33}$$

where $\mathrm{MSE}(\hat{\mathcal{H}}_{\ell 2}^{LS}, \sigma_z^2) = \sigma_z^2 \mathrm{tr}\{\mathbb{E}[\mathcal{R}_{\mathcal{X}\mathcal{X}}^{-1}(l)]\}$. The optimal active-set which minimizes the MSE performance (2.32) may be reduced via the minimization of (2.33). This is because $\mathrm{MSE}(\hat{\mathcal{H}}_{\ell 2}^{LS}, \sigma_z^2)$ is independent of $\mathcal{A}$. However, we notice that, for any $\mathcal{A}$,

$$\mathrm{MSE}(\hat{\mathcal{H}}_{\ell 2}^{LS}, \sigma_z^2) = \sigma_z^2\left[\mathrm{tr}\left\{\mathbb{E}\left[\mathfrak{J}_{\mathcal{A}}\mathcal{R}_{\mathcal{X}\mathcal{X}}^{-1}(l)\right]\right\} + \mathrm{tr}\left\{\mathbb{E}\left[\mathfrak{J}_{\mathcal{A}}^{\perp}\mathcal{R}_{\mathcal{X}\mathcal{X}}^{-1}(l)\right]\right\}\right], \tag{2.34}$$

where we denote $\mathfrak{J}_{\mathcal{A}} = \mathbf{J}_{\mathcal{A}} \otimes \mathbf{I}_{N_R}$ and $\mathfrak{J}_{\mathcal{A}}^{\perp} = \mathbf{I}_{WN_TN_R} - \mathfrak{J}_{\mathcal{A}}$. By Theorem 7.7.8 in [54],

$$\mathcal{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}}(l) \preceq \mathcal{P}_{\mathcal{A}}^{\mathrm{T}} \cdot \mathcal{R}_{\mathcal{X}\mathcal{X}}^{-1}(l) \cdot \mathcal{P}_{\mathcal{A}} \tag{2.35}$$

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

$$\mathbb{E}\left[\mathrm{tr}\left\{\mathcal{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}}(l)\right\} - \mathrm{tr}\left\{\mathfrak{J}_{\mathcal{A}}\mathcal{R}_{\mathcal{X}\mathcal{X}}^{-1}(l)\right\}\right] \leq 0. \tag{2.36}$$

Substituting (2.32), (2.34) and (2.36) into (2.33) yields

$$\Delta^{LS}_{\ell 1 \ell 2}(\mathcal{A}) \;\leq\; \mathfrak{E}(\mathcal{A}) - \sigma_z^2 \,\mathrm{tr}\left\{\mathbb{E}\left[\mathfrak{J}^{\perp}_{\mathcal{A}} \mathcal{R}^{-1}_{\mathcal{X}\mathcal{X}}(l)\right]\right\}. \tag{2.37}$$

The MSE performance (2.32) is, thereby, minimized with the optimal active-set $\mathcal{A}^*$ given by

$$\mathcal{A}^* \;=\; \arg\min_{\mathcal{A}} \left[\mathfrak{E}(\mathcal{A}) - \sigma_z^2 \,\mathrm{tr}\left\{\mathbb{E}\left[\mathfrak{J}^{\perp}_{\mathcal{A}} \mathcal{R}^{-1}_{\mathcal{X}\mathcal{X}}(l)\right]\right\}\right]. \tag{2.38}$$

Obviously, the problem (2.38) is a combinatorial optimization. The solution to (2.38) can be found from all possible $\sum_{k=0}^{WN_T} \binom{WN_T}{k}$ active-sets if the delay profile

$$\mathbb{E}[\mathbf{d}_{\mathcal{H}}(l)] \;=\; \mathbb{E}[\mathtt{diag}\{\mathcal{H}^{\mathrm{H}}(l) \cdot \mathcal{H}(l)\}] \tag{2.39}$$

is known.

Consequently, the analytical MSE performance of the $\ell 1$ LS channel estimation (2.10) is given by (2.32) with the optimal active-set (2.38).

2.3.2 MSE performance bound of the $\ell 1$ MB

Since the $\ell 2$ MB channel estimation techniques asymptotically achieve the CRB [16, 23, 24, 26], MSE performance of the $\ell 1$ MB algorithm is discussed through the CRB.

2.3.2.1 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[14] $\bar{N}$ is defined by $\mathtt{tr}\{\mathbf{X}\mathbf{X}^{\mathrm{H}}\}/M$, where the $M \times M$ matrix $\mathbf{X}$ denotes the Toeplitz matrix used in a channel estimator.

Definition 2 (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.*

[14] 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.

Definition 3 (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*

$$r^a_{w,k}(\sigma_z^2,\bar{N}) = \sum_{i=1}^{r_k} 1\left\{\lambda_k^i(l,w) > N_R\sigma_z^2/\bar{N}\right\} \tag{2.40}$$

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 $\mathbf{U}_k(l,w)|_{r^a_{w,k}(\sigma_z^2,\bar{N})+1:r_k}$.

2.3.2.2 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 (2.27) is independent of θ_k. Therefore, by utilizing the CRB of SISO channel estimation in [24], 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

$$\mathrm{CRB}^{\mathcal{Z}}_{\bar{N}}(\sigma_z^2,\mathbf{r}) = \sum_{k=1}^{N_T} \frac{N_R\sigma_z^2 r_k}{\bar{N}} \tag{2.41}$$

$$\mathrm{CRB}^{\Pi}_{\bar{N},w}(\sigma_z^2,\mathbf{r}) = \sum_{k=1}^{N_T} \frac{\sigma_z^2}{L_M\bar{N}}(w\,r_k - r_k^2) \tag{2.42}$$

under the assumption that the length $\bar{N}$ ideally uncorrelated sequence is used. The $\ell 1$ MB channel estimation can decrease the projection error (2.42) 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 (2.41) is independent of w. Therefore, the $\ell 1$ MB can improve the projection error, nevertheless, it does not improve asymptotic MSE performance (2.41) when L_M tends to ∞.

2.3.2.3 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:

$$\mathrm{aCRB}_{\bar{N},w}(\sigma_z^2) = \mathrm{CRB}_{\bar{N},w}(\sigma_z^2, \mathbf{r}_w^a(\sigma_z^2, \bar{N})) + \|\mathbf{\Lambda}^c(w)\|, \tag{2.43}$$

with $\mathbf{r}_w^a(\sigma_z^2, \bar{N}) = [r_{w,1}^a(\sigma_z^2, \bar{N}), \cdots, r_{w,N_T}^a(\sigma_z^2, \bar{N})]^{\mathrm{T}}$. The sum of singular values in the complemental-subspace is denoted by

$$\|\mathbf{\Lambda}^c(w)\| = \sum_{k=1}^{N_T} \mathbb{E}[\|\mathbf{\Lambda}_k(l,w)|_{r_k^a(\sigma_z^2,\bar{N})+1:r_k}\|_1].$$

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 $\forall k \in \{1, \cdots, N_T\}$.

2.3.2.4 Asymptotic MSE performance of MB techniques

The MSE performance of the $\ell 1$ MB is given by

$$\mathrm{MSE}(\hat{\mathcal{H}}_{\ell 1}^{MB}, \sigma_z^2) = \min_w \mathrm{aCRB}_{\bar{N},w}(\sigma_z^2).$$

As mentioned above, however, the asymptotic MSE performance of the $\ell 1$ MB with $L_M \to \infty$ is independent of the parameter w. The MSE performances of both the $\ell 1$ and $\ell 2$ MB algorithms are, hence, lower bounded by

$$\mathrm{aCRB}_{\bar{N}}(\sigma_z^2) = \mathrm{CRB}_{\bar{N}}^{\mathcal{Z}}(\sigma_z^2, \mathbf{r}_W^a(\sigma_z^2, \bar{N})) + \|\mathbf{\Lambda}^c(W)\|. \tag{2.44}$$

2.4 Numerical Examples

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

2.4.1 Simulation setups

2.4.1.1 Channel models

The CIRs are generated with the spatial channel model (SCM) [18,55]. This section assumes 4×4 MIMO channels, where the antenna element spacing at the 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 PB model with 3 km/h (PB3) mobility and the VA model [18] with 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., [56]) 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. 2.1, 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.

2.4.1.2 TX scenarios

We focus on intermittent communication scenarios to verify robustness of $\ell 1$ regularized channel estimation. A length $L_C = 100$ burst TX chunk is trans-

Table 2.3: Examples of initial registers (in hexadecimal) for PN sequences.

Combination	TX1	TX2	TX3	TX4
1	45	69	71	78
2	0E	31	5F	79
3	06	3A	4D	5C
4	0A	19	27	56
5	0D	1C	64	7F
6	12	22	52	70
7	03	1D	26	6E
8	2B	51	74	7B
9	2F	3C	43	6C
10	02	33	49	4B

mitted continuously. However, as illustrated in Fig. 2.4, 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 $\{\text{PB3} \rightarrow \text{VA30} \rightarrow \text{PB3} \rightarrow \text{VA30} \rightarrow ...\}$ 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.

2.4.1.3 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 [13] with the generator polynomial $1 + x^6 + x^7$. Specifically, they are obtained by initializing the shift register with the least significant 7 bits of hexadecimal initial values[15] shown in Table 2.3 so that the cross-correlations between TX streams are ideally low.

[15]There are more than 10 combinations, but not all possible $\binom{127}{4}$ combinations.

2.4.2 Normalized MSE performance with LS channel estimation techniques

We define the NMSE of a channel estimate $\hat{\mathcal{H}}$ by

$$\mathrm{NMSE}(\hat{\mathcal{H}}, \sigma_z^2) = \mathrm{MSE}(\hat{\mathcal{H}}, \sigma_z^2)/\mathbb{E}[\|\mathcal{H}\|^2].$$

Similarly, normalized aCRB (NaCRB) is denoted as

$$\mathrm{NaCRB}_{\bar{N}_{td}}(\sigma_z^2) = \mathrm{aCRB}_{\bar{N}_{td}}(\sigma_z^2)/\mathbb{E}[\|\mathcal{H}\|^2] \tag{2.45}$$

with asymptotic aCRB (2.44), where the length $\bar{N}_{td}$ of reference signals composed of TS and data sequences is defined by $\bar{N}_{td} = \mathtt{tr}\{\mathcal{R}_{\mathcal{X}\mathcal{X}}\}/(WN_TN_R)$. The NaCRB for the PB-VA scenario is assumed as the mean of those for the PB and VA channel models.

2.4.2.1 Comparison between the $\ell 1$ and $\ell 2$ LS techniques

Fig. 2.5 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[16] turbo iterations ($N_{\mathrm{turbo}} = 1, 6$). The maximum number of iterations in Algorithm 1 is set at $N_{\mathrm{AAD}} = 1$ or WN_T. As observed from Fig. 2.5, 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 ($N_{\mathrm{turbo}} = 6$) 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 $N_{\mathrm{AAD}} > 1$, the active-set detection (2.20) can fail in the very low SNR regime such that $\mathrm{NMSE}(\hat{\mathcal{H}}_{\ell 2}^{LS}, \sigma_z^2) \gg 1$ 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.

[16]The turbo iteration is stopped before the maximum number of iterations N_{turbo} if $|\mathcal{A}| = 0$ is detected. This is because the equalization is performed incorrectly for the null channel estimate $\hat{\mathcal{H}} = \mathbf{O}$.

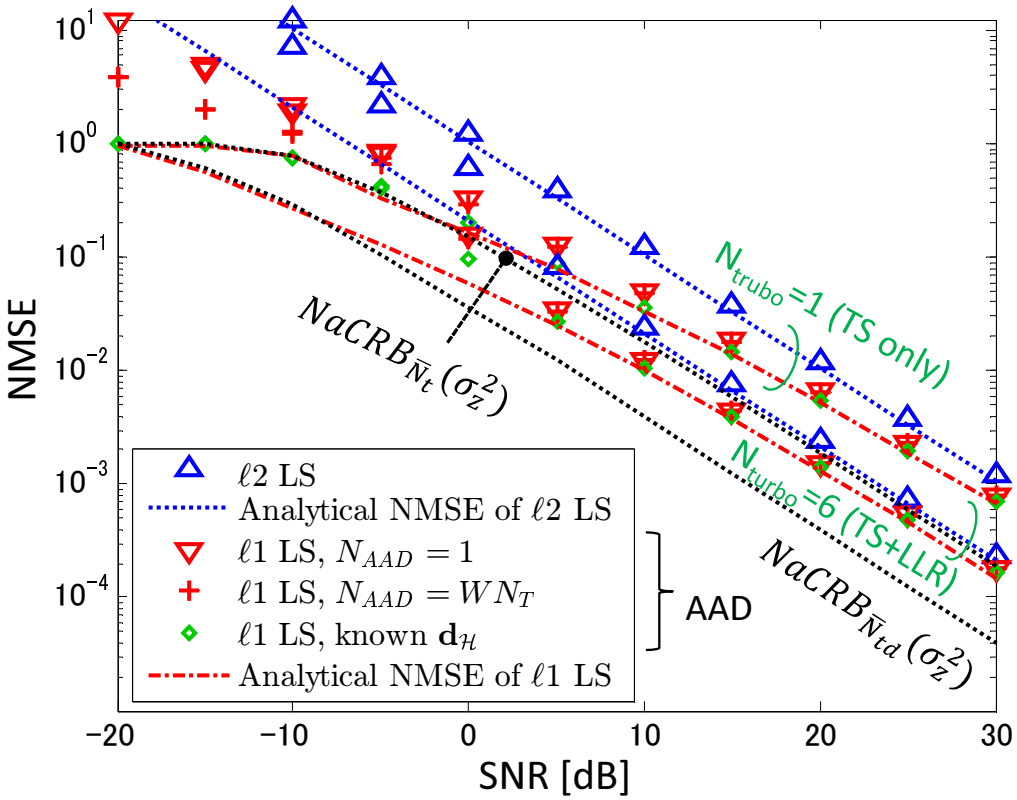

Fig. 2.5: NMSE performance with LS channel estimation techniques in the PB-VA scenario. N_{turbo} denotes the number of turbo iterations.

2.4.2.2 Illustration for the NMSE gain of the $\ell 1$ LS channel estimation over the $\ell 2$ LS technique

As detailed in Section 2.A, the analytical NMSE performance (2.32) can be approximated by

$$\mathrm{MSE}(\hat{\mathcal{H}}_{\ell 1}^{LS}, \sigma_z^2, \mathcal{A}) \;\approx\; \frac{|\mathcal{A}|}{N_T} \cdot \frac{\mathrm{MSE}(\hat{\mathcal{H}}_{\ell 2}^{LS}, \sigma_z^2)}{W} + \mathbb{E}[\|\mathcal{H}_{\mathcal{A}}^{\perp}\|^2] \qquad (2.46)$$

when the ideally uncorrelated reference signal is used. Hence, the NMSE gain with the $\ell 1$ LS over the $\ell 2$ LS technique is determined by the trade-off between the cardinality $|\mathcal{A}|$ and the bias $\mathbb{E}[\|\mathcal{H}_{\mathcal{A}}^{\perp}\|^2]$ due to *inactive*-set, the complement of the active-set.

Fig. 2.6 illustrates the approximated MSE performance (2.46) of the $\ell 1$ LS technique, where the active-set is chosen optimally by (2.38). The VA-VA scenario and $N_{\mathrm{turbo}} = 1$ are assumed. As observed from the second subfigure, the cardinality $|\mathcal{A}|$ is proportional to the SNR. In the case the VA-VA scenario, $|\mathcal{A}|/N_T \ll W$ is expected in a moderate SNR regime. The $\ell 1$ LS channel estimation can, hence, improve the NMSE performance by this property. In a very low SNR regime, the first term in (2.46) vanishes due to $|\mathcal{A}| = 0$, however, the second term increases up to $\mathbb{E}[\|\mathcal{H}\|^2]$. Therefore, the NMSE performance of the $\ell 1$ LS saturates at 1 in the very low SNR regime.

Figs. 2.7(a), (b) and (c) illustrate the optimal active-set selection (2.38). The SNR is assumed at 15 dB in the VA-VA scenario. The errors due to the noise in Figs. 2.7(b) and (c) represent the entry-wise NMSE performance given by a length-W symbol vector:

$$\mathbb{E}\left[\sum_{k=1}^{N_T} \mathtt{diag}\{\Delta\hat{\mathbf{H}}_k^{\mathrm{H}} \Delta\hat{\mathbf{H}}_k\}\right] \;/\; \mathbb{E}\left[\sum_{k=1}^{N_T} \|\mathbf{H}_k\|^2\right], \qquad (2.47)$$

where $\Delta\hat{\mathbf{H}}_k = \hat{\mathbf{H}}_k - \mathbf{H}_k$. Intuitively, the active-set is the index of significant CIR taps above the threshold according to a noise level. In the case of the LS-based channel estimation techniques, the noise level can be determined by $\sigma_z^2 \mathrm{tr}\{\mathcal{R}_{\mathcal{X}\mathcal{X}}\}/(WN_T)$. This is because, as depicted in Fig. 2.7(b), the symbol-wise error of the $\ell 2$ LS channel estimate is distributed uniformly based on the fact that the correlation (2.11)–the LS estimate itself–computes the expectation of the noise component consequently, together with the channel estimate. As depicted in Fig. 2.7(c), the $\ell 1$ LS channel estimation aims to decrease estimation errors by discarding the insignificant CIR taps.

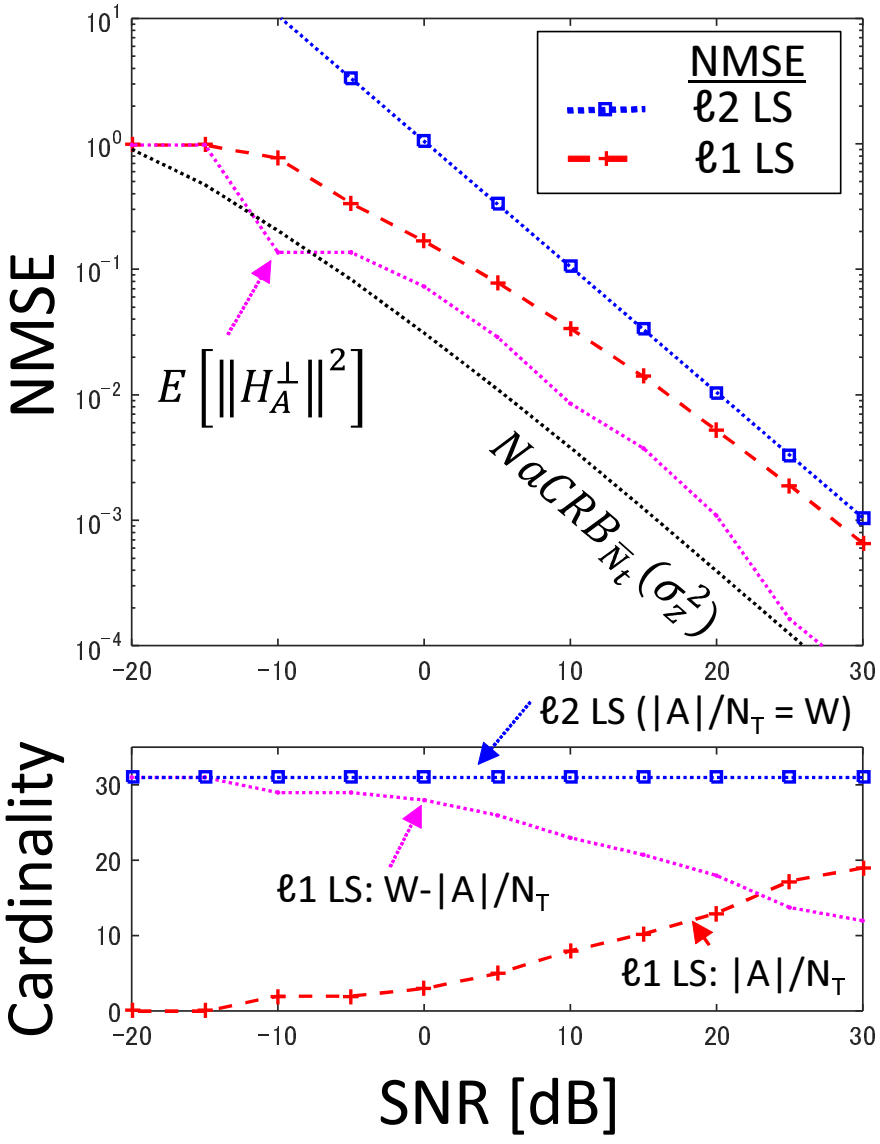

Fig. 2.6: Detalis of the NMSE gain with the $\ell 1$ LS channel estimation. The VA-VA scenario and $N_{\text{turbo}} = 1$ are assumed. The second figure shows the normalized cardinalities of the active-set ($|\mathcal{A}|/N_T$) and the inactive-set ($W - |\mathcal{A}|/N_T$) in the $\ell 1$ LS technique. Notice that the $\ell 2$ LS always assumes $|\mathcal{A}|/N_T = W$.

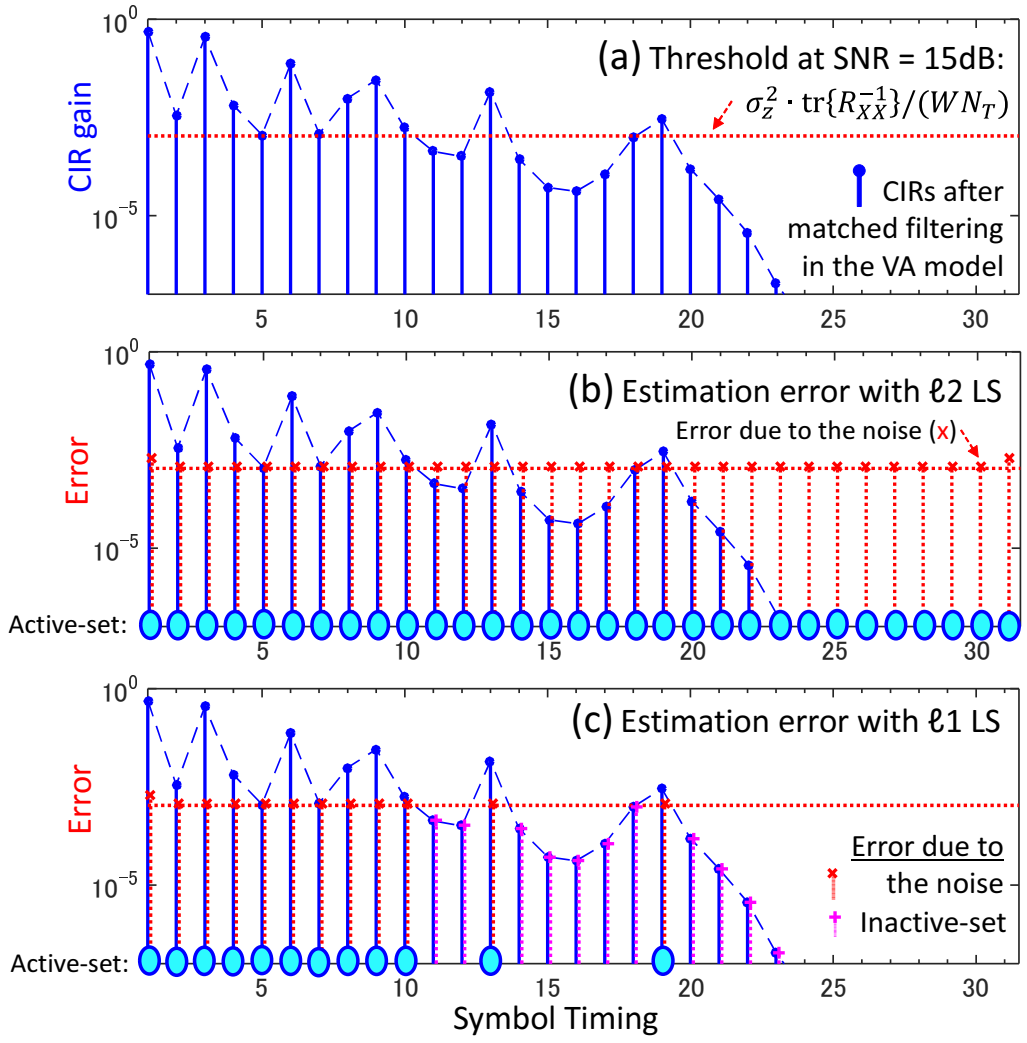

Fig. 2.7: Active-sets and estimation errors of LS channel estimation techniques. The SNR is assumed at 15 dB in the VA-VA scenario. The errors due to the noise in (b) and (c) are the entry-wise NMSE (2.47). Note that the area of the estimation errors in (c) is decreased from that in (b).

2.4.2.3 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) [57] algorithms for MIMO channel estimation by using (2.11). Although the OMP algorithms is detailed later in Section 2.B, 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 two criteria:

1. $\underset{1\le j\le WN_TN_R}{\arg\max}\ (|\mathtt{vec}\{\mathbf{\Xi}\}|)|_j$,
2. $\underset{1\le j\le WN_T}{\arg\max}\, \mathtt{diag}\{\mathbf{\Xi}^{\mathrm{H}}\mathbf{\Xi}\}|_j$,

where the residual correlation $\mathbf{\Xi}$ is defined as

$$\mathbf{\Xi} \;=\; (\mathcal{Y}_t - \hat{\mathcal{H}}\mathcal{X}_t)\mathcal{X}_t^{\mathrm{H}} + \hat{\mathbf{\Gamma}}(\mathcal{Y}_d - \hat{\mathcal{H}}\hat{\mathcal{X}}_d)\hat{\mathcal{X}}_d^{\mathrm{H}} \tag{2.48}$$

for a possible estimate $\hat{\mathcal{H}}$ obtained in the OMP-based algorithms. In Figs. 2.8 and 2.9, the OMP algorithms with criteria 1) and 2) are referred to as vectorized OMP (vec-OMP) and *OMP*, respectively. As observed from Figs. 2.8 and 2.9, 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 $\mathbf{\Xi}^{\mathrm{H}}\mathbf{\Xi}$ improves the accuracy of the active-set selection. We, hence, focus on the OMP with the criterion 2) hereafter.

Fig. 2.8 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. 2.8, 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 (2.20). 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. 2.5, 2.8 and 2.9, the AAD algorithm achieves the analytical MSE performance of the $\ell 1$ LS exactly if the delay profile is known.

2.4.2.4 Comparison between $\ell 1$ solvers – AAD vs. ITDSE

The ITDSE [29] 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. 2.8, 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 $\geq$ 25 dB.) The original ITDSE has to perform N_{RES} iterations, although NMSE convergence performance shown in Fig. 2.10 suggests that the ITDSE may terminate the process before the N_{RES}-th iteration with a certain criterion. We note that, even with $N_{\mathrm{AAD}} = 1$, 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.

2.4.2.5 Analytical NMSE performance of the $\ell 1$ LS

As shown in Figs. 2.5 and 2.8, 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. 2.9 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. 2.9, 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.

2.4.3 Normalized MSE performance with the MB and hybrid algorithms

Fig. 2.11 shows NMSE with the MB channel estimation in the VA-VA scenario. 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. 2.11, both the $\ell 1$ and $\ell 2$ MB algorithms achieve the NaCRB asymptotically. This observation verifies the MSE performance analysis described in the Section 2.3.2. Furthermore, it should be emphasized that the NaCRB saturates at 1 if SNR $\leq$ -15 dB. This is because in the very low SNR regime, the adaptive-rank (2.40) becomes $\mathbb{E}[r^a_k(l)] = 0$. The NMSEs with the $\ell 1$ and

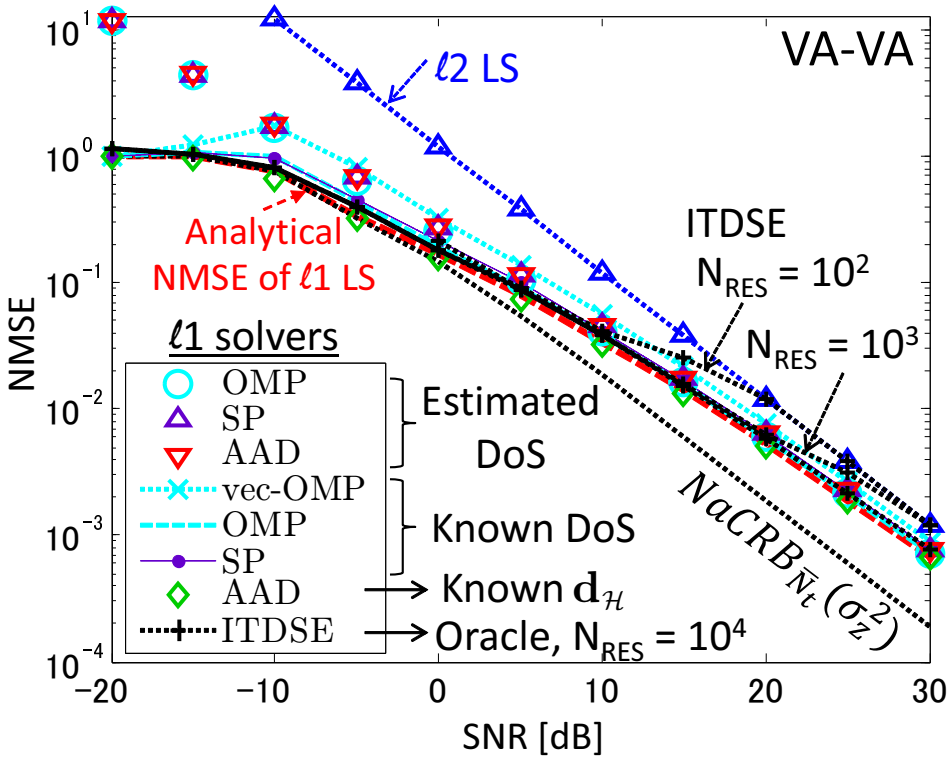

Fig. 2.8: Comparison between $\ell 1$ solvers: NMSE performance over SNR with $\ell 1$ LS channel estimation in the VA-VA scenario.

In Figures 2.8 to 2.10, 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 (2.38) and the cardinality of the estimated active-set (2.20), 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 [29] in these figures determines the optimal solution from possible channel estimates by the oracle criterion.

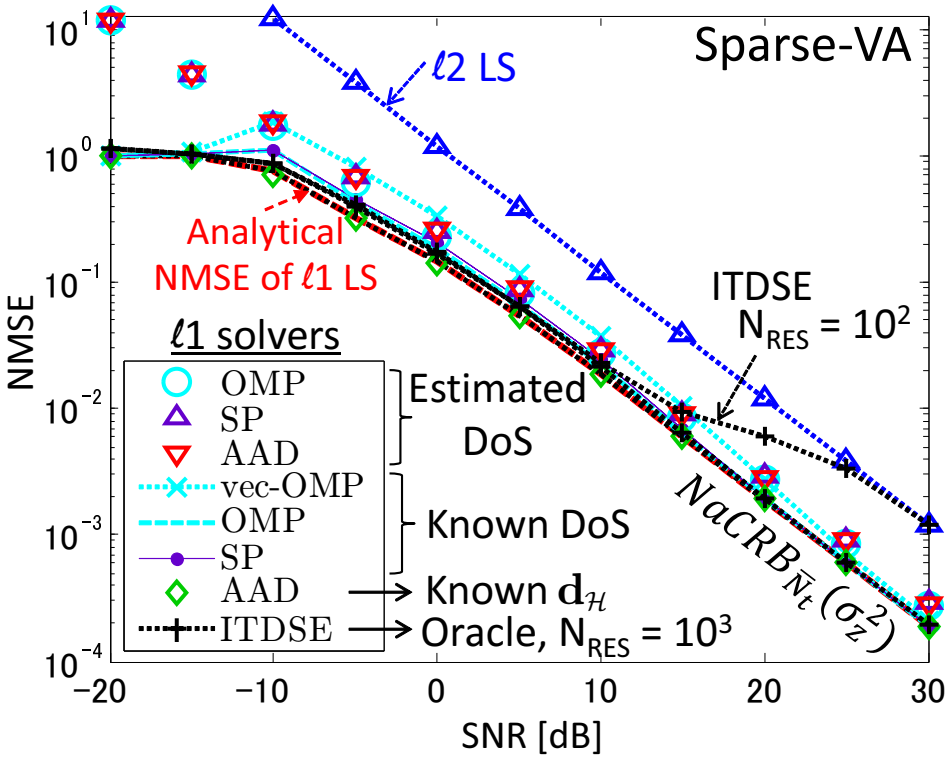

Fig. 2.9: Comparison between $\ell 1$ solvers: NMSE performance over SNR with $\ell 1$ LS channel estimation in the sparse-VA scenario. It should be noted that the analytical NMSE of the $\ell 1$ LS technique coincides with the NaCRB in exactly sparse channels.

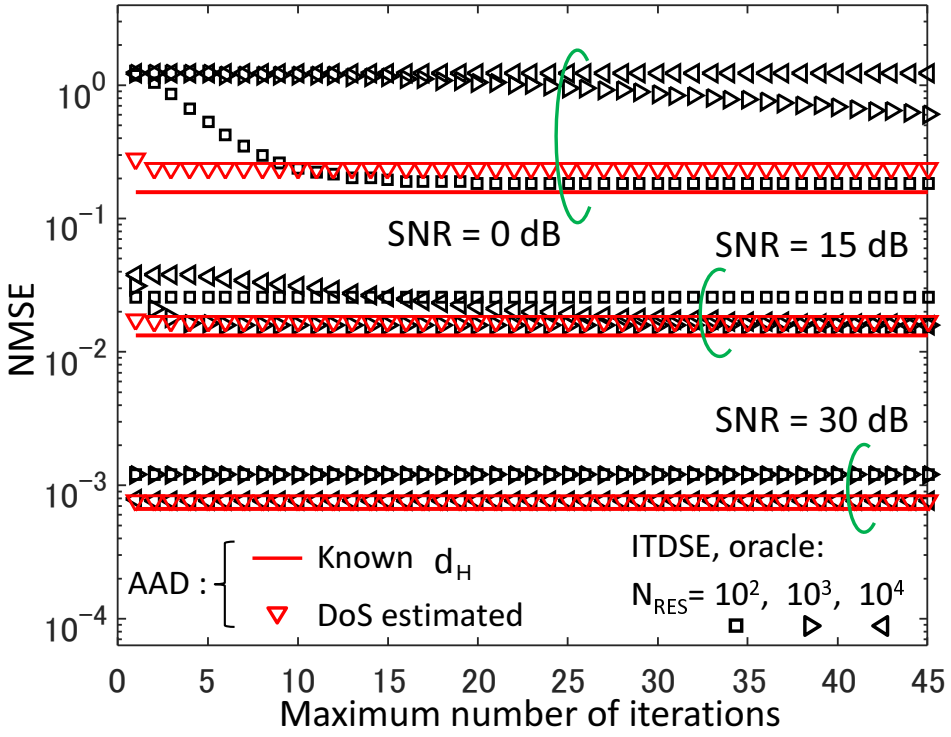

Fig. 2.10: Comparison between the AAD and ITDSE algorithms: NMSE convergence over iteration in the VA-VA scenario.

$\ell 2$ 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 iterations in the AAD algorithm is set at 1.

Fig. 2.12 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.

2.4.4 NMSE convergence properties

2.4.4.1 Effect of LLR's accuracy onto NMSE

Figs. 2.13 and 2.14 depict NMSE performance over LLR's accuracy at SNR = 15 dB in the VA-VA and PB-VA scenarios, respectively. We define the LLR's accuracy by the mutual information (MI) $\mathcal{I}^a_{\mathrm{EQU}}$ between the LLR λ^a_{EQU} and the coded bits c at the transmitter, as

$$\begin{aligned}\mathcal{I}^a_{\mathrm{EQU}} &= \mathcal{I}(\lambda^a_{\mathrm{EQU}}; c) \\ &= \frac{1}{2}\sum_{\mathfrak{m}=\pm 1}\int_{-\infty}^{+\infty} \mathrm{P_r}(\lambda^a_{\mathrm{EQU}}|\mathfrak{m})\ \log_2 \frac{\mathrm{P_r}(\lambda^a_{\mathrm{EQU}}|\mathfrak{m})}{\mathrm{P_r}(\lambda^a_{\mathrm{EQU}})}\, d\lambda^a_{\mathrm{EQU}}, \end{aligned} \tag{2.49}$$

where $\mathrm{P_r}(\lambda^a_{\mathrm{EQU}}|\mathfrak{m})$ is the conditional probability density of λ^a_{EQU} given $\mathfrak{m} = 1 - 2c$ [58].

It is observed from Fig. 2.13 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 $\mathcal{I}^a_{\mathrm{EQU}}$, since $\bar{N}_{td} \approx N_t + \gamma\hat{\sigma}_d^2\{N_d - (W-1)/2\}$ with $\gamma = \sigma_z^2/(\sigma_z^2 + \Delta\hat{\sigma}_d^2 N_T \sigma^2_{\mathbf{H}}/N_R)$ holds when $\mathbf{R}_{\mathcal{H}\mathcal{H}} \approx (N_T\sigma^2_{\mathbf{H}}/N_R)\mathbf{I}_{N_R}$. The variance of λ^a_{EQU} tends to ∞ as $\mathcal{I}^a_{\mathrm{EQU}}$ converges to 1 [58], 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 2.3.2. 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

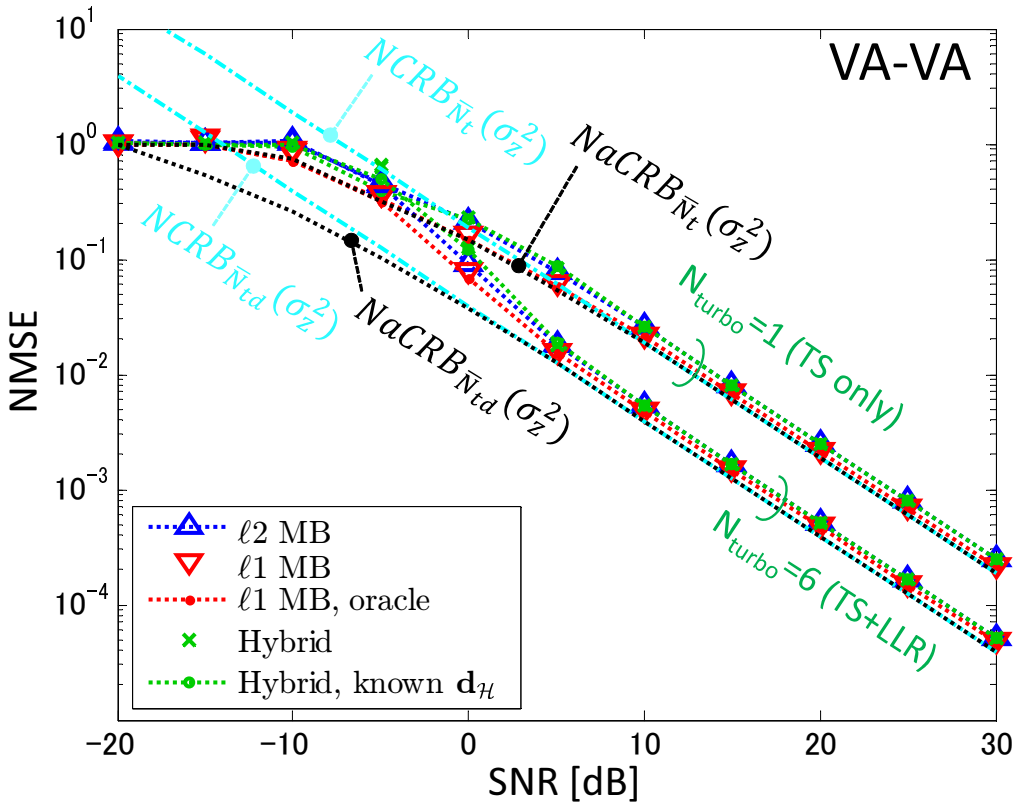

Fig. 2.11: NMSE performance with MB channel estimation techniques in the VA-VA scenario. Normalized CRB is given by $\mathrm{NCRB}_{\bar{N}}(\sigma_z^2) = \mathrm{CRB}_{\bar{N}}^{\mathcal{Z}}(\sigma_z^2, \mathbf{r})/\mathbb{E}[\|\mathcal{H}\|^2]$ with (2.41), where all N_T entries of the rank vector $\mathbf{r}$ are set at 6 for the PB or VA channel models. In the hybrid algorithm, $N_{\mathrm{AAD}} = 1$ is assumed.

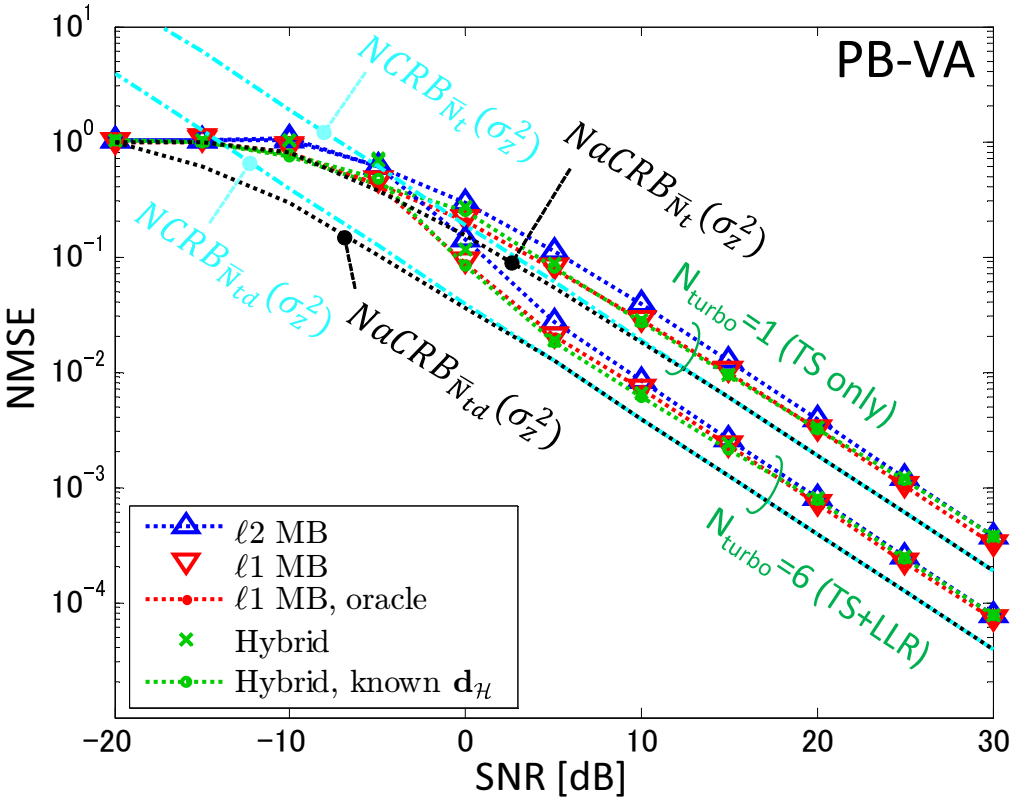

Fig. 2.12: NMSE performance with MB channel estimation techniques in the PB-VA scenario. In the hybrid algorithm, $N_{\mathrm{AAD}} = 1$ is assumed.

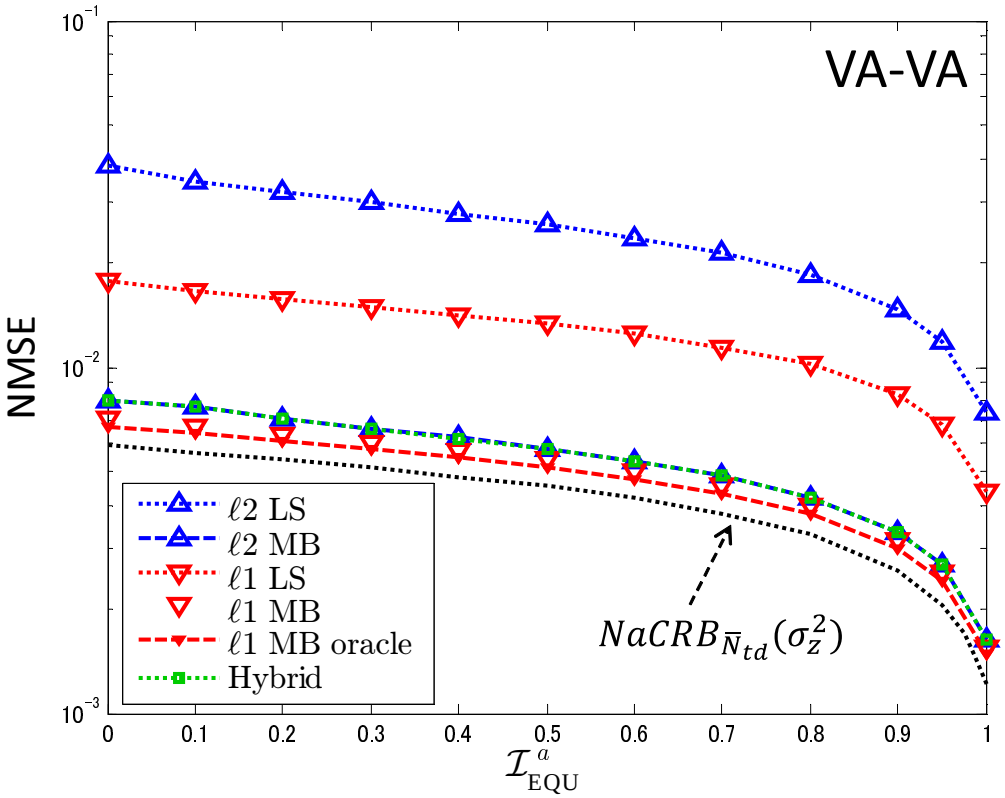

Fig. 2.13: NMSE convergence performance over the MI $\mathcal{I}^{a}_{\mathrm{EQU}}$ (2.49) in the VA-VA scenario at SNR is set at 15 dB.

channel model. Nevertheless, as shown in Fig. 2.14, the hybrid algorithm improves NMSE over the $\ell 1$ MB if there are abrupt channel changes such as in the PB-VA scenario. The reason for the improvement is clarified by observing NMSE tracking performance.

2.4.4.2 Tracking performance

Fig. 2.15 shows the NMSE tracking performance in the PB-VA scenario. The $\ell 2$ MB channel estimation suffers from the NMSE tracking errors as seriously as that causes bit errors at the boarders between the TX chunks. The $\ell 1$ MB channel estimation also suffers from the NMSE tracking errors, however,

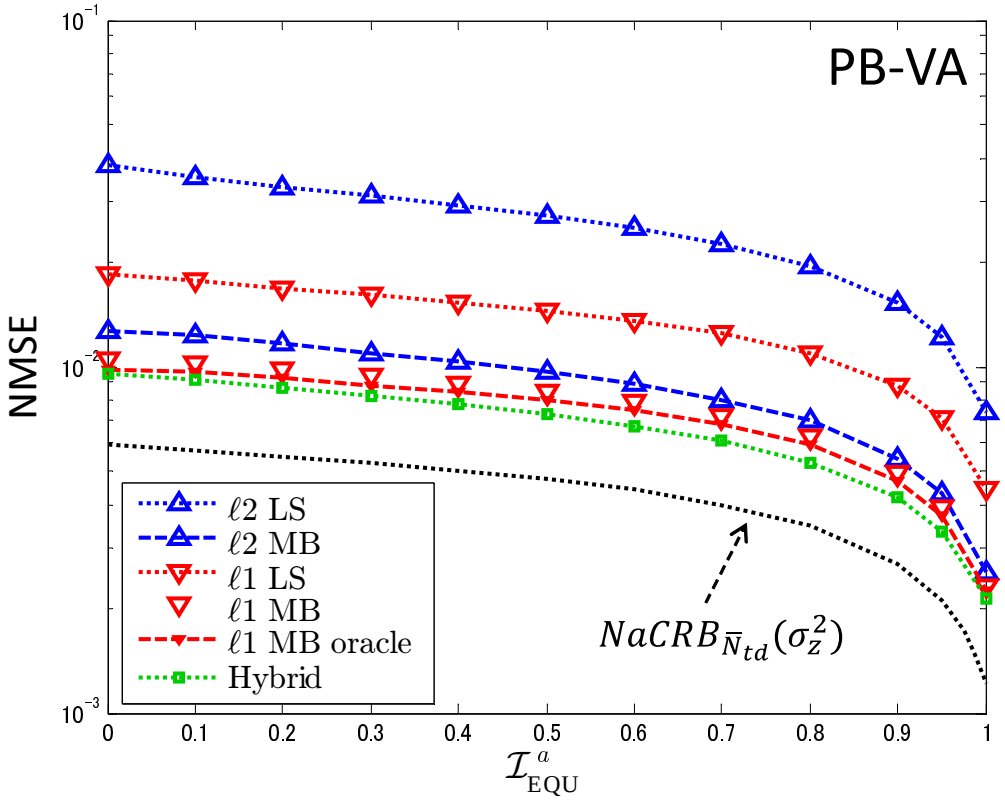

Fig. 2.14: NMSE convergence performance over the MI $\mathcal{I}^{a}_{\mathrm{EQU}}$ (2.49) in the PB-VA scenario at SNR is set at 15 dB.

improves bit errors at the boarders between the TX chunks. This is because, as described in Section 2.3.2, the $\ell 1$ regularization decreases the projection error. Nevertheless, as observed from Fig. 2.15, the $\ell 1$ MB cannot solve the NMSE tracking error problem completely. This is because the $\ell 1$ MB estimate inherits the past CIRs' characteristics in the sliding window of the MMSE formulation.

On the other hand, the $\ell 1$ 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 $\ell 1$ LS, while achieving the performance bound aCRB asymptotically by the $\ell 2$ MB estimate when the tracking error problem is not observed.

2.4.5 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

$$\mathrm{SNR} = \sigma_x^2 \ (\sigma_{\mathbf{H}}^2/N_R) \ \eta \cdot E_b/N_0, \tag{2.50}$$

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 spectral efficiency η of the frame format structure is defined as $\eta = N_{\mathrm{info}}/L_{frm}$ with a frame length $L_{frm} = L_B N_B$ in symbol. It is hence reduced to $\eta = 1.4$ for the MIMO system used in the simulations.

Figs. 2.16 and 2.17 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. 2.16, 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. 2.17, 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. 2.15, 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.

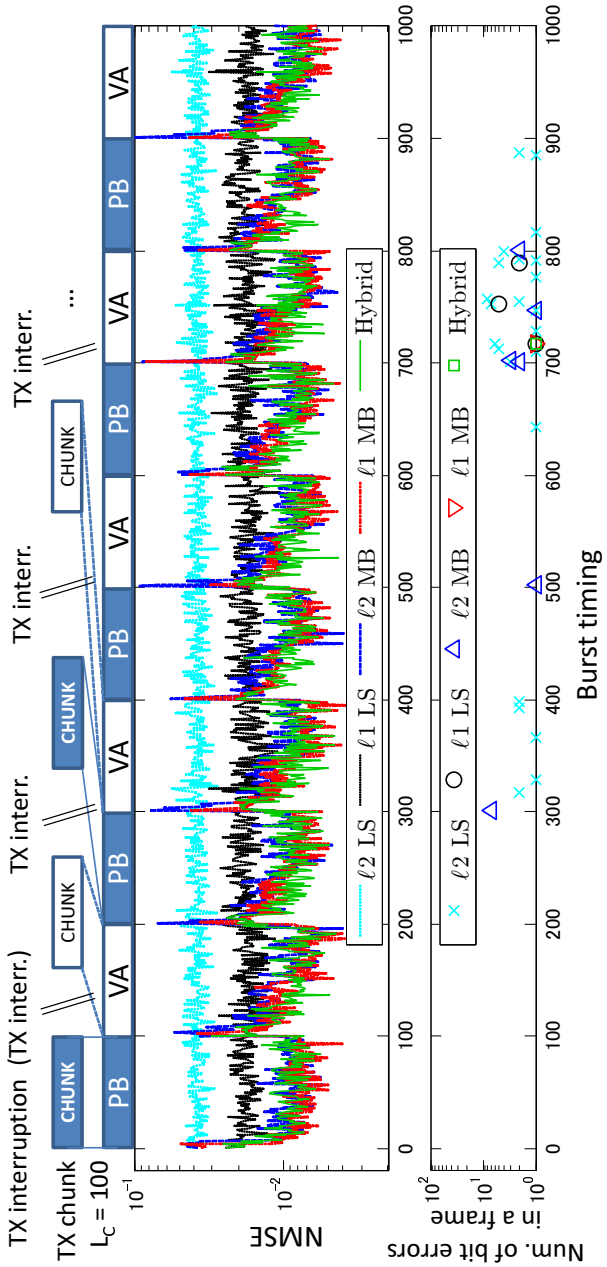

Fig. 2.15: NMSE tracking performance in the PB-VA scenario. The arbitrary length TX interruptions are omitted in NMSE tracking results. The channel estimation is performed with the TS only. SNR is set at 15 dB. In the second subfigure, the number (num.) of bit errors in the i-th frame is shown at the $\{(i-1)N_B+1\}$-th burst timing.

As shown in Fig. 2.17, 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. 2.16, 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.

2.5 Summary

This chapter 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 the MSE performance bound of 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 chapter. Moreover, the $\ell 1$ MB technique does not improve MSE significantly over the $\ell 2$ MB if the following four assumptions hold:

1. CIRs follow the subspace channel model.
2. The reference signals are ideally uncorrelated.
3. The MMSE formulation follows the complex normal distribution.
4. The sliding window length in the MMSE formulation is long enough.

However, the $\ell 2$ MB technique suffers from deterioration of the channel estimation performance if the four assumptions only partially hold. By focusing on intermittent TX scenarios which do not always satisfy the first assumption,[17] this chapter has demonstrated robustness using $\ell 1$ regulariza-

[17] Appendix 2.C and Section 3.2.3 show the cases that, respectively, the second and third assumptions are not always correct.

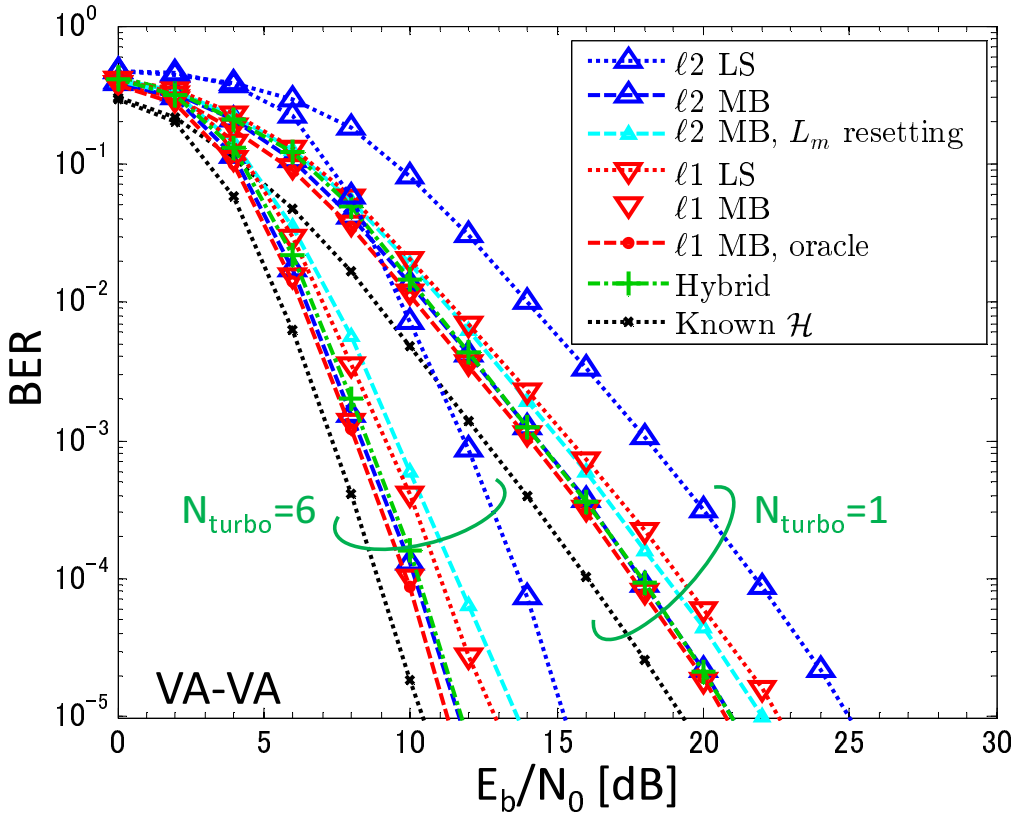

Fig. 2.16: BER performance with the 4×4 MIMO system in the VA-VA scenario.

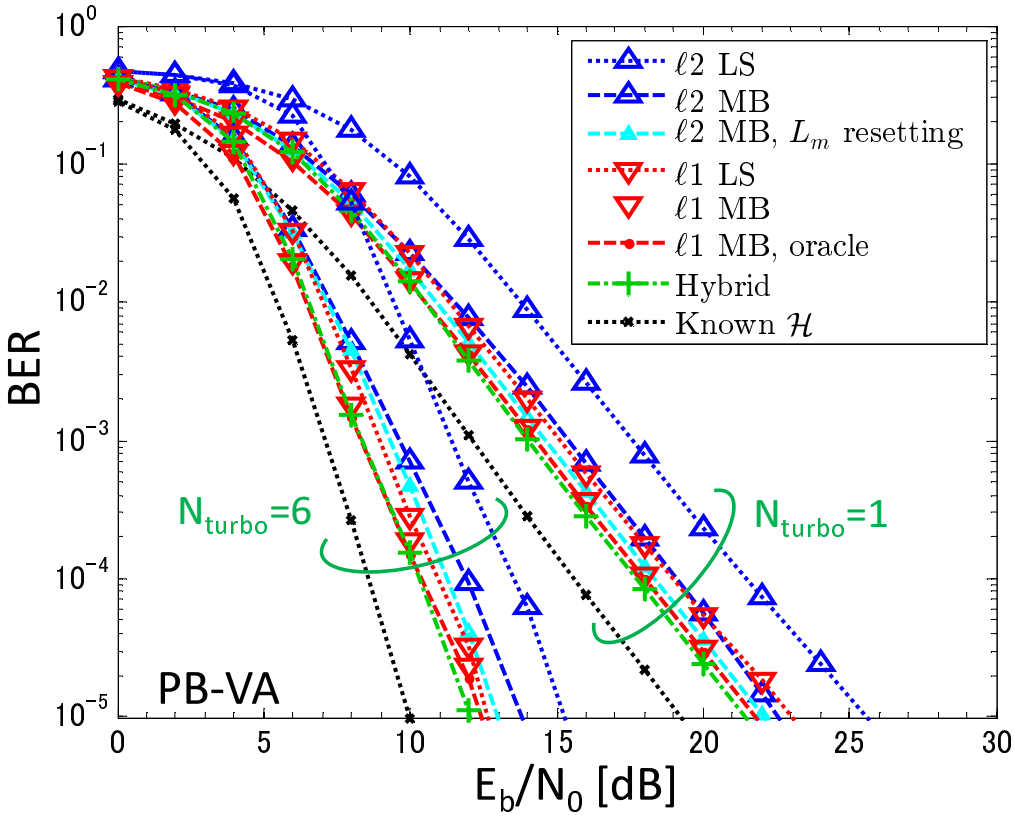

Fig. 2.17: BER performance with the 4×4 MIMO system in the PB-VA scenario.

tion. Simulation results shows that, due to the tracking error problem, the receiver using the $\ell 2$ MB exhibits BER degradation in the PB-VA scenario even though 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 chapter 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 equal to that of the $\ell 2$ MB if the number of the maximum iteration of the AAD algorithm is set to 1.

Appendix

2.A Derivation of the AAD Algorithm

2.A.1 Approximation of the MSE (2.32)

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

$$\mathrm{tr}\{\mathcal{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}}}\} \approx \frac{|\mathcal{A}|}{WN_T}\mathrm{tr}\{\mathcal{R}^{-1}_{\mathcal{X}\mathcal{X}}\} \tag{2.51}$$

and $\mathfrak{E}(\mathcal{A}) \approx \mathbb{E}\left[\|\mathcal{H}^{\perp}_{\mathcal{A}}\|^2\right]$. The analytical MSE (2.32) is, therefore, approximated by

$$\begin{aligned}\mathrm{MSE}(\hat{\mathcal{H}}^{LS}_{\ell 1}, \sigma_z^2, \mathcal{A}) &\approx |\mathcal{A}|\frac{\mathrm{MSE}(\hat{\mathcal{H}}^{LS}_{\ell 2}, \sigma_z^2)}{WN_T} + \mathbb{E}[\|\mathcal{H}^{\perp}_{\mathcal{A}}\|^2] &(2.52)\\ &= \mathbb{E}[\|\mathcal{H}\|^2] + \sum_{j\in\mathcal{A}}\left\{\frac{\mathrm{MSE}(\hat{\mathcal{H}}^{LS}_{\ell 2}, \sigma_z^2)}{WN_T} - \bar{d}_{\mathcal{H},j}\right\}, &(2.53)\end{aligned}$$

since $\mathbb{E}[\|\mathcal{H}^{\perp}_{\mathcal{A}}\|^2] = \mathbb{E}[\|\mathcal{H}\|^2] - \sum_{j\in\mathcal{A}}\bar{d}_{\mathcal{H},j}$, where $\bar{d}_{\mathcal{H},j}$ denotes the j-th entry of the delay profile $\mathbb{E}[\mathbf{d}_{\mathcal{H}}]$. The problem (2.38) can also be approximated by

$$\begin{aligned}\mathcal{A}^* &\approx \arg\min_{\mathcal{A}} \sum_{j\in\mathcal{A}}\left\{\frac{\mathrm{MSE}(\hat{\mathcal{H}}^{LS}_{\ell 2}, \sigma_z^2)}{WN_T} - \bar{d}_{\mathcal{H},j}\right\}\\ &= \left\{ j \;\middle|\; \begin{array}{l}\bar{d}_{\mathcal{H},j} > \mathrm{MSE}(\hat{\mathcal{H}}^{LS}_{\ell 2}, \sigma_z^2)/(WN_T),\\ j = 1, \cdots, WN_T\end{array} \right\}. &(2.54)\end{aligned}$$

2.A.2 Derivation of the AAD

It is reasonable to assume that $\|\hat{\mathcal{H}}^{LS}_{\ell 1} - \mathcal{H}\|^2 \approx \mathbb{E}[\|\hat{\mathcal{H}}^{LS}_{\ell 1} - \mathcal{H}\|^2]$, when the reference signal length is long enough. Under this assumption, the problem (2.17) can be seen as an approximated version of the minimization of (2.33). Hence, (2.17) can be reduced to a solution corresponding to (2.54). 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 (2.52) if the active-set can be selected so that $\|\mathcal{H}_{\mathcal{A}}^{\perp}\|^2$ is very minor. It should be noticed that

$$\mathrm{tr}\{\mathcal{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}^*}}\} \lessapprox \frac{|\mathcal{A}^*|}{|\mathcal{A}_{[n]}|}\mathrm{tr}\{\mathcal{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_{\mathcal{A}_{[n]}}}\} \lessapprox \frac{|\mathcal{A}^*|}{WN_T}\mathrm{tr}\{\mathcal{R}^{-1}_{\mathcal{X}\mathcal{X}}\} \tag{2.55}$$

is satisfied for $\mathcal{A}^* \subseteq \mathcal{A}_{[n]} \subseteq \{1, \cdots, WN_T\}$ by (2.35) and (2.51). Thereby, the active-set detection (2.20) is an extension of (2.54) so that it takes account of the delay profile approximation error.

Specifically, at the first iteration ($n = 0$), Algorithm 1 performs

$$\hat{\mathcal{A}}_{[0+1]} = \left\{ j \;\middle|\; \begin{array}{l} \hat{d}^{[0]}_{\mathcal{H},j} > \mathrm{MSE}(\hat{\mathcal{H}}^{LS}_{\ell 2}, \sigma_z^2)/|\mathcal{A}_{[0]}| + |\Delta\hat{d}^{[0]}_{\mathcal{H},j}|, \\ j \in \mathcal{A}_{[0]} \end{array} \right\}, \tag{2.56}$$

with $\mathcal{A}_{[0]} = \{1, \cdots, WN_T\}$, where $\hat{d}^{[n]}_{\mathcal{H},j}$ and $\Delta\hat{d}^{[n]}_{\mathcal{H},j}$ denote the j-th entries of a delay profile estimation vector $\hat{\mathbf{d}}^{[n]}_{\mathcal{H}}$ and its estimation error $\Delta\hat{\mathbf{d}}^{[n]}_{\mathcal{H}} = \hat{\mathbf{d}}^{[n]}_{\mathcal{H}} - \mathbf{d}_{\mathcal{H}}$. For the first iteration $n = 0$, (2.19) becomes $\hat{\mathbf{d}}^{[0]}_{\mathcal{H}} = \mathbf{P}_{\mathcal{A}_{[0]}} \cdot \mathtt{diag}\{\hat{\mathcal{G}}^{\mathrm{H}}_{[0]}\hat{\mathcal{G}}_{[0]}\}$ with $\mathbf{P}_{\mathcal{A}_{[0]}} = \mathbf{I}_{WN_T}$ and $\hat{\mathcal{G}}_{[0]} = \hat{\mathcal{H}}^{LS}_{\ell 2}$. Moreover, by denoting $\hat{\mathcal{H}}^{LS}_{\ell 2} = \mathcal{H} + \Delta\hat{\mathcal{H}}^{LS}_{\ell 2}$,

$$\begin{aligned} \Delta\hat{\mathbf{d}}^{[0]}_{\mathcal{H}} &= \mathtt{diag}\left\{(\Delta\hat{\mathcal{H}}^{LS}_{\ell 2})^{\mathrm{H}}\Delta\hat{\mathcal{H}}^{LS}_{\ell 2} + \mathcal{H}^{\mathrm{H}}\Delta\hat{\mathcal{H}}^{LS}_{\ell 2} + (\Delta\hat{\mathcal{H}}^{LS}_{\ell 2})^{\mathrm{H}}\mathcal{H}\right\} \\ &\overset{(a)}{\approx} \mathtt{diag}\left\{(\Delta\hat{\mathcal{H}}^{LS}_{\ell 2})^{\mathrm{H}}\Delta\hat{\mathcal{H}}^{LS}_{\ell 2}\right\} \\ &\overset{(b)}{\approx} \left(\mathrm{MSE}(\hat{\mathcal{H}}^{LS}_{\ell 2}, \sigma_z^2)/|\mathcal{A}_{[0]}|\right) \cdot \mathbf{1}_{|\mathcal{A}_{[0]}|}, \end{aligned}$$

where the approximations (a) and (b) are due, respectively, to (a) $\mathbb{E}[\mathcal{H}^{\mathrm{H}}\Delta\hat{\mathcal{H}}^{LS}_{\ell 2}] = \mathbf{O}$ and (b) the estimation error of the $\ell 2$ LS estimate is distributed uniformly over all symbol timings. Thereby, Algorithm 1 assumes

$$\Delta\hat{d}^{[0]}_{\mathcal{H},j} = \mathrm{MSE}(\hat{\mathcal{H}}^{LS}_{\ell 2}, \sigma_z^2)/|\mathcal{A}_{[0]}|$$

for $\forall j \in \mathcal{A}_{[0]}$. After the second iteration ($n \geq 1$), the recursive formula (2.20) aims to improve detection accuracy by the inequality (2.55). 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 (2.55) holds when $\mathcal{R}_{\mathcal{X}\mathcal{X}}/\bar{N}_{td} = \mathbf{I}_{WN_TN_R}$.

2.B $\ell 1$ LS Channel Estimation Techniques with the OMP and ITDSE Algorithms

The $\ell 1$ LS channel estimation techniques with the OMP and/or ZD-based algorithms can also be described with (2.11). However, the active-set detection is different from (2.17).

2.B.1 $\ell 1$ LS channel estimation techniques with the OMP algorithm

In the OMP algorithm, the active-set is selected so that the residual correlation $\boldsymbol{\Xi}$ (2.48) is maximized. The active-set update can, specifically, be described as either of the following two strategies:

1. maximizing the vectorized residual correlation:

$$\mathcal{S} \leftarrow \mathcal{S} \cup \left\{ \underset{1 \le j \le WN_T N_R}{\arg\max} \left(|\mathtt{vec}\{\boldsymbol{\Xi}\}| \right)|_j \right\}, \tag{2.57}$$

$$\mathcal{A} = \{ \mathtt{mod}(s-1, WN_T) + 1 \mid \forall s \in \mathcal{S} \} \tag{2.58}$$

2. maximizing the Rx diversity combined residual correlation:

$$\mathcal{A} \leftarrow \mathcal{A} \cup \left\{ \underset{1 \le j \le WN_T}{\arg\max} \, \mathtt{diag}\{\boldsymbol{\Xi}^{\mathrm{H}} \boldsymbol{\Xi}\}|_j \right\}, \tag{2.59}$$

where the index sets $\mathcal{S}$ and $\mathcal{A}$ are initialized to $\emptyset$ before performing the iteration of the OMP algorithm. The operation $\mathtt{mod}(n, m)$ denotes that n modulo m.

Algorithm 3 shows the $\ell 1$ LS channel estimation with the OMP (e.g., [46]). It should be noted that Algorithm 3 requires the DoS parameter N_{OMP} which may be given by the cardinality of the optimal active-set (2.38) if the delay profile (2.39) of the CIRs is known.

2.B.2 The $\ell 1$ LS channel estimation with the ITDSE algorithm

Algorithm 4 summarizes the $\ell 1$ LS channel estimation with the ITDSE [29]. The ITDSE algorithm updates the active-set by a step-wise threshold determined by the maximum entry of the $\ell 2$ channel estimate. Specifically, the

Algorithm 3 The $\ell 1$ LS with the OMP.

Input: $\mathcal{Y}_t, \mathcal{Y}_d, \mathcal{X}_t$, $\hat{\mathcal{X}}_d$ and $N_{\rm OMP}$.

1: Compute $\mathbf{R}_{\mathcal{YX}}$ (2.13), $\mathcal{R}_{\mathcal{XX}}$ (2.12) and $\hat{\mathbf{\Gamma}}$ (2.14).
2: Initialize: $\mathbf{\Xi} = \mathbf{R}_{\mathcal{YX}}$, $\mathcal{A} = \emptyset$ and $\mathcal{S} = \emptyset$.
3: **for** $|\mathcal{A}| < N_{\rm OMP}$ **do**
4: Update the active-set $\mathcal{A}$ by (2.58) or (2.59).
5: Obtain an estimate $\hat{\mathcal{H}} = \mathtt{mat}_{N_R}\{\hat{\mathbf{g}}_{\mathcal{A}}\} \cdot \mathbf{P}_{\mathcal{A}}^{\rm T}$ by (2.11) with the updated $\mathcal{A}$.
6: Update the residual correlation $\mathbf{\Xi}$ (2.48).
7: **end for**

Output: $\hat{\mathcal{H}}_{\ell 1}^{LS} = \hat{\mathcal{H}}$.

step-wise Δ is defined by

$$\Delta \;=\; \max_j \{\hat{d}_{\mathcal{H}_{[0]},j}\}/N_{\rm RESO} \tag{2.60}$$

with the resolution constant $N_{\rm RESO}$. The parameter $\hat{d}_{\mathcal{H}_{[0]},j}$ denotes the j-th entry of $\hat{\mathbf{d}}_{\mathcal{H}_{[0]}} = \mathtt{diag}\{\hat{\mathcal{H}}_{[0]}^{\rm H} \cdot \hat{\mathcal{H}}_{[0]}\}$, where the $\ell 2$ LS channel estimate $\hat{\mathcal{H}}_{[0]}$ can be obtained by (2.11) with $\mathcal{A}_{[0]} = \{1, \cdots, WN_T\}$. At the n-th iteration, the active-set $\mathcal{A}_{[n]}$ is updated, as

$$\mathcal{A}_{[n]} \;=\; \left\{ j \;\middle|\; \hat{d}_{\mathcal{H}_{\rm TMP},j} > n\Delta,\ j \in \mathcal{A}_{[n-1]} \right\}, \tag{2.61}$$

where $\hat{d}_{\mathcal{H}_{\rm TMP},j}$ denotes the j-th entry of the delay profile vector

$$\hat{\mathbf{d}}_{\mathcal{H}_{\rm TMP}} \;=\; \mathtt{diag}\{\hat{\mathcal{H}}_{\rm TMP}^{\rm H} \cdot \hat{\mathcal{H}}_{\rm TMP}\} \tag{2.62}$$

with the temporary channel estimate matrix $\hat{\mathcal{H}}_{\rm TMP}$ obtained in the iteration.

Notice that $\hat{\mathcal{H}}_{\rm TMP}$ and $\mathcal{A}_{[n]}$ are mutually dependent. Thereby, the ITDSE algorithm has to perform the second loop, the iteration number of which is pre-defined by $N_{\rm FINE}$, in order to re*fine* accuracy of the active-set update. According to [29], the constant $N_{\rm FINE}$ is typically limited to 3. Nevertheless, no certain method to determine the resolution constant $N_{\rm RESO}$ can be found in [29]. As discussed in Sections 2.2.1.3 and 2.4.2.4, the AAD algorithm adaptively determines the threshold: $n\Delta$ according to the analytical MSE performance of the LS channel estimation techniques.

Algorithm 4 The $\ell 1$ LS with the ITDSE.

Input: $\mathcal{Y}_t, \mathcal{Y}_d, \mathcal{X}_t$, $\hat{\mathcal{X}}_d$, N_{RESO}, N_{FINE} and a small positive constant ϵ.

Compute $\mathbf{R}_{\mathcal{YX}}$ (2.13), $\mathcal{R}_{\mathcal{XX}}$ (2.12) and $\hat{\mathbf{\Gamma}}$ (2.14).
Determine the step-wise of the threshold by (2.60) according to the $\ell 2$ LS estimate $\hat{\mathcal{H}}_{[0]}$.
for $n = 1$ to N_{RESO} **do**
  $\mathcal{H}_{\mathrm{TMP}} = \hat{\mathcal{H}}_{[n-1]}$.
  for $k = 1$ to N_{FINE} **do**
    Update the delay profile (2.62) with $\mathcal{H}_{\mathrm{TMP}}$.
    Update the active-set $\mathcal{A}_{[n]}$ by (2.61).
    Obtain an estimate $\hat{\mathcal{H}}_{[n]} = \mathtt{mat}_{N_R}\{\hat{\mathbf{g}}_{\mathcal{A}_{[n]}}\} \cdot \mathbf{P}^{\mathrm{T}}_{\mathcal{A}_{[n]}}$ by (2.11).
    if $\|\hat{\mathcal{H}}_{[n]} - \hat{\mathcal{H}}_{\mathrm{TMP}}\|^2 < \epsilon$ **then**
      Break the loop of the counter k.
    end if
    $\hat{\mathcal{H}}_{\mathrm{TMP}} = \hat{\mathcal{H}}_{[n]}$.
  end for
end for

Output: $\hat{\mathcal{H}}^{LS}_{\ell 1} = \hat{\mathcal{H}}_{[\hat{n}]}$, where $\hat{n} = \arg\min_n \mathcal{L}_{td}(\hat{\mathcal{H}}_{[n]})$.

2.C Performance of the $\ell 1$ MB Estimation with Random Sequences

Conventional $\ell 2$ MB channel estimation techniques [16,23,24,26] can achieve the CRB asymptotically under the four assumptions described in Section 2.5. However, as mentioned in Section 2.5, this claim is potentially not true if the four assumptions only partially hold. It should be noticed that finding optimal TS combinations is a non-polynomial (NP) hard problem in a massive MIMO system, since binomial coefficients increase in factorial orders. Moreover, the number of the ideally uncorrelated sequences with a given bandwidth is limited, which can cause the pilot contamination problem [59]. This appendix studies, therefore, performance of the MB algorithm where the second assumption does not always hold. Specifically, random TS is assumed as a typical *moderately* uncorrelated sequence.

This appendix shows that the $\ell 2$ MB technique can suffer from the noise enhancement problem when the noise whitening in the MB algorithm is not accurate enough. However, the $\ell 1$ regularized MB channel estimation [60] can improve the problem by a CIR length constraint. This appendix clarifies the reason for the improvement. Furthermore, asymptotic channel estimation performance with very long TSs and/or massive transmission streams is discussed from the viewpoint of the noise whitening accuracy.

2.C.1 $\ell 1$ MB channel estimation with TSs only

We rewrite the $\ell 1$ MB method for channel estimation with TSs only, in order to concentrate on its basic performance analysis. As described in Section 2.2.2, the $\ell 1$ MB estimation performs the subspace projection per a TX stream and it obtains $N_R \times w$ channel estimate matrices $\hat{\mathbf{G}}^{MB}_{[w]k}(l)$, $1 \leq k \leq N_T$, for each TX stream. The w-th possible solution corresponding to the length w CIR constraint is, hence, described as $\hat{\mathcal{H}}^{MB}_{[w]}(l) = [\hat{\mathbf{G}}^{MB}_{[w]1}(l), \cdots, \hat{\mathbf{G}}^{MB}_{[w]N_T}(l)]\mathcal{P}^{\mathrm{T}}_{[w]}$, where $\mathcal{P}_{[w]} = \mathbf{I}_{N_T} \otimes \mathbf{P}_{[w]}$ with the $W \times w$ matrix $\mathbf{P}_{[w]} = [\mathbf{I}_w \ \mathbf{O}]^{\mathrm{T}}$. The operator $\otimes$ denotes the Kronecker product.

In the case the channel estimation is performed with TSs only, the $N_R \times w$ estimated matrix $\hat{\mathbf{G}}^{MB}_{[w]k}(l)$ is given by

$$\hat{\mathbf{G}}^{MB}_{[w]k}(l) \ = \ \hat{\bar{\mathbf{G}}}^{LS}_{[w]k}(l) \cdot \hat{\bar{\mathbf{\Pi}}}_{[w]k} \cdot \bar{\mathbf{Q}}^{-\mathrm{H}}_{[w]kk} \tag{2.63}$$

for the k-th TX stream, where the $w \times w$ matrix $\bar{\mathbf{Q}}_{[w]ij}$ denotes the (i,j)-th block matrix of $\bar{\mathbf{R}}^{1/2}_{\mathcal{X}\mathcal{X}_t[w]}$ with $\bar{\mathbf{R}}_{\mathcal{X}\mathcal{X}_t[w]} = \mathbb{E}_{j\in\mathcal{J}_{L_M}(l)}[\mathcal{P}^{\mathrm{T}}_{[w]}\mathbf{R}_{\mathcal{X}\mathcal{X}_t}(j)\mathcal{P}_{[w]}]$. The index set $\mathcal{J}_{L_M}(l)$ denote the length L_M sliding window in the MB algorithm. Moreover,

$$\begin{aligned}\hat{\tilde{\mathbf{G}}}^{LS}_{[w]k}(l) \triangleq \hat{\mathbf{G}}^{LS}_{[w]k}(l)\cdot\bar{\mathbf{Q}}^{\mathrm{H}}_{[w]kk} \\ + \sum_{i=k+1}^{N_T}\left\{\hat{\mathbf{G}}^{LS}_{[w]i}(l) - \mathfrak{G}_{[w]i}(l)\right\}\bar{\mathbf{Q}}^{\mathrm{H}}_{[w]ki}.\end{aligned} \tag{2.64}$$

with $\mathfrak{G}_{[w]i}(l) = \hat{\mathbf{G}}^{MB}_{[w]i}(l)$, where $\hat{\mathbf{G}}^{LS}_{[w]k}(l)$ is the LS channel estimate corresponding to an $N_R \times w$ CIR matrix $\mathbf{G}_{[w]k}(l) = \mathbf{H}_k(l)\mathbf{P}_{[w]}$. The projection matrix $\hat{\tilde{\mathbf{\Pi}}}_{[w]k}$ denotes $\hat{\tilde{\mathbf{V}}}_{[w]k}|_{1:r_k}(\hat{\tilde{\mathbf{V}}}_{[w]k}|_{1:r_k})^{\dagger}$, where the unitary matrix $\hat{\tilde{\mathbf{V}}}_{[w]k}$ is the singular vectors of the covariance matrix $\mathbb{K}_{j\in\mathcal{J}_{L_M}(l)}[\hat{\tilde{\mathbf{G}}}^{LS}_{[w]k}(j)]$. The parameter r_k denotes the number of paths for the k-th TX stream.

It should be noticed that (2.64) is performed for the noise whitening. Let us denote $\Delta\hat{\tilde{\mathbf{G}}}^{LS}_{[w]k}(l) = \hat{\tilde{\mathbf{G}}}^{LS}_{[w]k}(l) - \tilde{\mathbf{G}}_{[w]k}(l)$ with $\tilde{\mathbf{G}}_{[w]k}(l) = \mathbf{G}_{[w]k}(l)\bar{\mathbf{Q}}^{\mathrm{H}}_{[w]kk}$ and concatenate the N_T residual matrices into an $N_R \times wN_T$ matrix as $\Delta\hat{\tilde{\mathcal{G}}}^{LS}_{[w]}(l) = [\Delta\hat{\tilde{\mathbf{G}}}^{LS}_{[w]1}(l), \cdots, \Delta\hat{\tilde{\mathbf{G}}}^{LS}_{[w]N_T}(l)]$. Suppose $\mathfrak{G}_{[w]i}(l) = \mathbf{G}_{[w]i}(l)$ in (2.64), we observe that

$$\begin{aligned}\underset{j\in\mathcal{J}_{L_M}(l)}{\mathbb{K}}[\Delta\hat{\tilde{\mathcal{G}}}^{LS}_{[w]}(j)] &= \sigma_z^2 N_R \bar{\mathbf{R}}^{1/2}_{\mathcal{X}\mathcal{X}_t[w]}\cdot\underset{j\in\mathcal{J}_{L_M}(l)}{\mathbb{E}}\left[\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}_t[w]}(j)\right]\cdot\bar{\mathbf{R}}^{\mathrm{H}/2}_{\mathcal{X}\mathcal{X}_t[w]} \\ &\approx \sigma_z^2 N_R \mathbf{I}_{WN_T}\end{aligned} \tag{2.65}$$

holds when the TSs are fixed to a consistent sequence or the TSs are ideally uncorrelated $\mathbf{R}_{\mathcal{X}\mathcal{X}_t}(l)/N_t \approx \mathbf{I}_{WN_T}$ for $\forall l$.

2.C.2 MSE analysis

The burst index l is omitted for the sake of simplicity.

Theorem 1. *Denote the channel estimation error $\hat{\mathbf{H}}^{MB}_{[w]k} - \mathbf{H}_{[w]k}$ by $\Delta\hat{\mathbf{H}}^{MB}_{[w]k}$. The MSE for the $\ell 1$ MB estimate $\hat{\mathbf{H}}^{MB}_{[w]k}$ can be decomposed into the following three terms:*

$$\begin{aligned}\mathbb{E}\left[\|\Delta\hat{\mathbf{H}}^{MB}_{[w]k}\|^2\right] &= \mathbb{E}\left[\|\mathbf{H}^{\perp}_{[w]k}\|^2\right] \\ &\quad + \mathbb{E}\left[\|\epsilon_{\mathcal{Z},k}(w)\|^2\right] + \mathbb{E}\left[\|\epsilon_{\Pi,k}(w)\|^2\right],\end{aligned} \tag{2.66}$$

where the discarded part of CIR $\mathbf{H}^{\perp}_{[w]k}$ *due to the CIR length constraint, the residual noise* $\epsilon_{\mathcal{Z},k}(w)$ *and the projection error* $\epsilon_{\Pi,k}(w)$ *are respectively defined as*

$$\begin{aligned}
\mathbf{H}^{\perp}_{[w]k} &= \mathbf{H}_k(\mathbf{I}_W - \mathbf{P}_{[w]}\mathbf{P}^{\mathrm{T}}_{[w]}), && (2.67)\\
\epsilon_{\mathcal{Z},k}(w) &= \Delta\hat{\tilde{\mathbf{G}}}^{LS}_{[w]k} \cdot \hat{\tilde{\mathbf{\Pi}}}_{[w]k} \cdot \bar{\mathbf{Q}}^{-\mathrm{H}}_{[w]kk}, && (2.68)\\
\epsilon_{\Pi,k}(w) &= \tilde{\mathbf{G}}_{[w]k} \cdot \Delta\hat{\tilde{\mathbf{\Pi}}}_{[w]k} \cdot \bar{\mathbf{Q}}^{-\mathrm{H}}_{[w]kk}. && (2.69)
\end{aligned}$$

Furthermore, $\Delta\hat{\tilde{\mathbf{\Pi}}}_{[w]k} = \hat{\tilde{\mathbf{\Pi}}}_{[w]k} - \tilde{\mathbf{\Pi}}_{[w]k}$*, where* $\tilde{\mathbf{\Pi}}_{[w]k}$ *is obtained from the first* r_k *singular vectors of* $\mathbb{K}[\tilde{\mathbf{G}}_{[w]k}]$*.*

Proof. Obviously, $\mathbb{E}\left[\|\Delta\hat{\mathbf{H}}^{MB}_{[w]k}\|^2\right] = \mathbb{E}\left[\|\mathbf{H}^{\perp}_{[w]k}\|^2\right] + \mathbb{E}\left[\|\Delta\hat{\mathbf{G}}^{MB}_{[w]k}\|^2\right]$, where $\Delta\hat{\mathbf{G}}^{MB}_{[w]k} = \epsilon_{\mathcal{Z},k}(w) + \epsilon_{\Pi,k}(w)$. Moreover, $\mathrm{tr}\{\mathbb{E}[\epsilon^{\mathrm{H}}_{\Pi,k}(w) \cdot \epsilon_{\mathcal{Z},k}(w)]\} = 0$ since $\mathbb{E}[\tilde{\mathbf{G}}^{\mathrm{H}}_{[w]k} \cdot \Delta\hat{\tilde{\mathbf{G}}}^{LS}_{[w]k}] = \mathbf{O}$. □

<u>*Remark*</u>: For TSs satisfying $\mathbb{E}_{j\in\mathcal{J}_{L_M}(l)}[\mathbf{R}_{\mathcal{X}\mathcal{X}_t}(j)]/N_t = \mathbf{I}_{WN_T}$, we have $\bar{\mathbf{Q}}^{-1/2}_{[w]kk} = \mathbf{I}_w/\sqrt{N_t}$. Hence,

$$\begin{aligned}
\mathbb{E}[\|\epsilon_{\mathcal{Z},k}(w)\|^2] &= \frac{1}{N_t}\mathrm{tr}\left\{\underset{j\in\mathcal{J}_{L_M}(l)}{\mathbb{K}}\left[\Delta\hat{\tilde{\mathbf{G}}}^{LS}_{[w]k}(j)\right]\right\}\frac{r_k}{w}\\
&= \sigma_z^2 N_R \frac{\omega(w)}{N_t} r_k, && (2.70)
\end{aligned}$$

where we define *whitening ratio* $\omega(w)$ [61] as

$$\begin{aligned}
\omega(w) &= \mathrm{tr}\left\{\bar{\mathbf{R}}_{\mathcal{X}\mathcal{X}_t[w]} \cdot \underset{j\in\mathcal{J}_{L_M}(l)}{\mathbb{E}}[\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}_t[w]}(j)]\right\} \,/\, \mathrm{tr}\{\mathbf{I}_{wN_T}\}\\
&= N_t \cdot \mathrm{tr}\left\{\underset{j\in\mathcal{J}_{L_M}(l)}{\mathbb{E}}[\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}_t[w]}(j)]\right\} \,/\, wN_T. && (2.71)
\end{aligned}$$

It should be noted that $\mathbb{E}_{j\in\mathcal{J}_{L_M}(l)}[\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}_t[w]}(j)] = \mathbf{I}_{wN_T}/N_t$ is not always satisfied although $\bar{\mathbf{R}}_{\mathcal{X}\mathcal{X}_t[w]} \approx N_t\mathbf{I}_{wN_T}$. This is because $(\mathbf{A}+\mathbf{B})^{-1} = (\mathbf{A}^{-1}+\mathbf{B}^{-1})$ does not hold in general for arbitrary invertible matrices $\mathbf{A}$ and $\mathbf{B}$.

2.C.3 Numerical examples

The same MIMO system as that in Section 2.4.1 are used. However, the path number r_k is assumed to be known in order to focus on analysis of the residual error (2.68).

2.C.3.1 NMSE performance of the $\ell 1$ MB

Fig. 2.18 shows NMSE performance with random TSs, where the NMSE is defined by $\mathbb{E}[\|\hat{\mathcal{H}}^{MB}_{[\hat{w}]} - \mathcal{H}\|^2]/\mathbb{E}[\|\mathcal{H}\|^2]$. The TS length and the sliding window length in the MB algorithm are set at $N_t = 127$ and $L = 50$, respectively. The TSs are re-generated every burst timing so that $\bar{\mathbf{R}}_{\mathcal{X}\mathcal{X}_t}/N_t = \mathbf{I}_{WN_T}$ holds. As shown in Fig. 2.18, the NMSE with the $\ell 2$ MB is 8 dB away from the performance bound, NCRB, given by $NCRB(\sigma_z^2) = N_R\sigma_z^2 \sum_{k=1}^{N_T} r_k/(N_t\mathbb{E}[\|\mathcal{H}\|^2])$. This is because the whitening ratio with the random TSs becomes $\omega(W) = 6.4 \gg 1$ and, thereby, the $\ell 2$ MB suffers from the noise enhancement in (2.70). As observed from Fig. 2.18, the NMSE with the $\ell 1$ MB can be improved significantly over that of the $\ell 2$ MB. The reason for the improvement is detailed in the next section 2.C.3.2.

It should be noticed that the problem of the noise whitening can be avoided by using a fixed TS pattern so that $\mathbb{E}_{j\in\mathcal{J}_{L_M}(l)}[\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}_t}(j)] = \bar{\mathbf{R}}^{-1}_{\mathcal{X}\mathcal{X}_t}$. Fig. 2.19 shows the NMSE performance with a fixed TS pattern. However, as shown in Fig. 2.19, the NMSE with the $\ell 2$ MB is not improved due to $\bar{\mathbf{R}}_{\mathcal{X}\mathcal{X}_t[W]}/N_t = \mathbf{R}_{\mathcal{X}\mathcal{X}_t[W]}(l)/N_t \neq \mathbf{I}_{WN_T}$ for a fixed random TS. Another potential solution can be to process the noise whitening with $\tilde{\mathbf{R}}^{1/2}_{\mathcal{X}\mathcal{X}_t[W]} \triangleq \left\{\mathbb{E}_{j\in\mathcal{J}_{L_M}(l)}[\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}_t[W]}(j)]\right\}^{-1/2}$ so that the covariance matrix $\mathbb{K}[\Delta\hat{\tilde{\mathcal{G}}}^{LS}_{[W]}(j)]$ yields (2.65) correctly. Nevertheless, the $\ell 2$ MB does not improve the NMSE performance significantly due to $\tilde{\mathbf{R}}_{\mathcal{X}\mathcal{X}_t[W]}/N_t \neq \mathbf{I}_{WN_T}$, although simulation results are omitted for the sake of conciseness.

After all, ideally uncorrelated TSs are needed to essentially solve the problem of the noise whitening. As observed from Fig. 2.11 in Section 2.4.3, both the $\ell 1$ MB and $\ell 2$ MB channel estimation techniques achieve the NCRB asymptotically with the PN sequences.

In a large-scale MIMO system, nevertheless, finding the optimal sequence combinations is an NP hard problem. The Gold sequence [62] is known as one of the most promising solutions to the problem, although it can be inferior

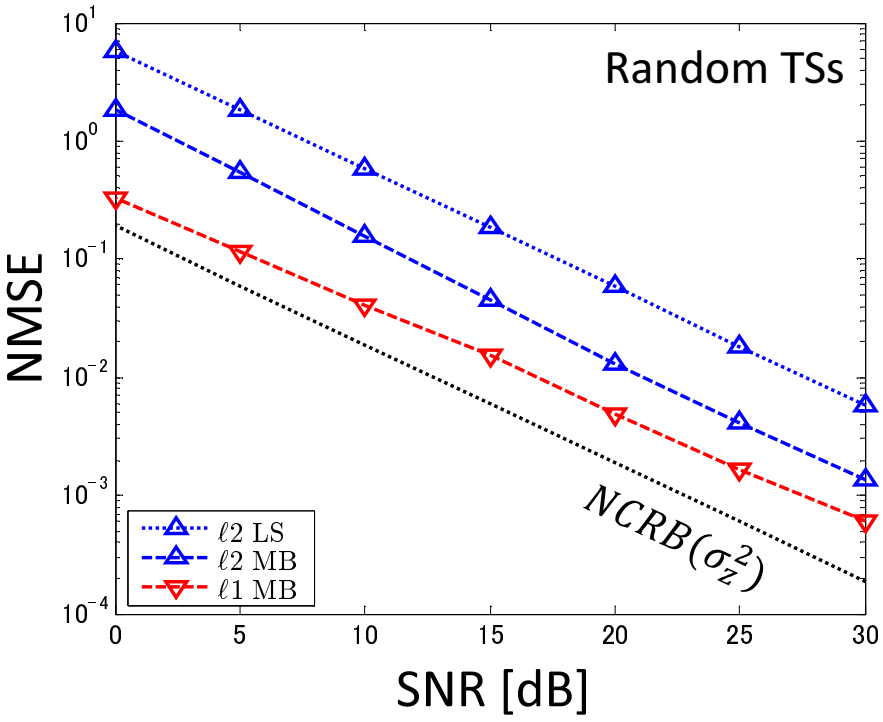

Fig. 2.18: The NMSE performance of the 4×4 MIMO system in the VA30 scenario by using the random TSs. The TSs are changed every burst timing.

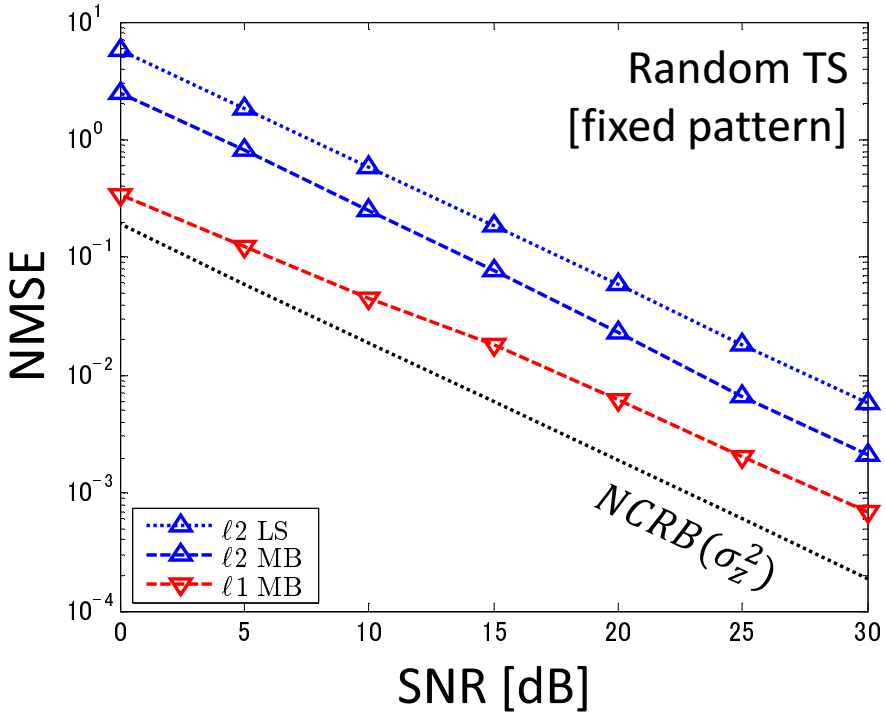

Fig. 2.19: The NMSE performance of the 4×4 MIMO system in the VA30 scenario by using the random TSs. The TSs are fixed to a certain sequence in to avoid the noise whitening problem.

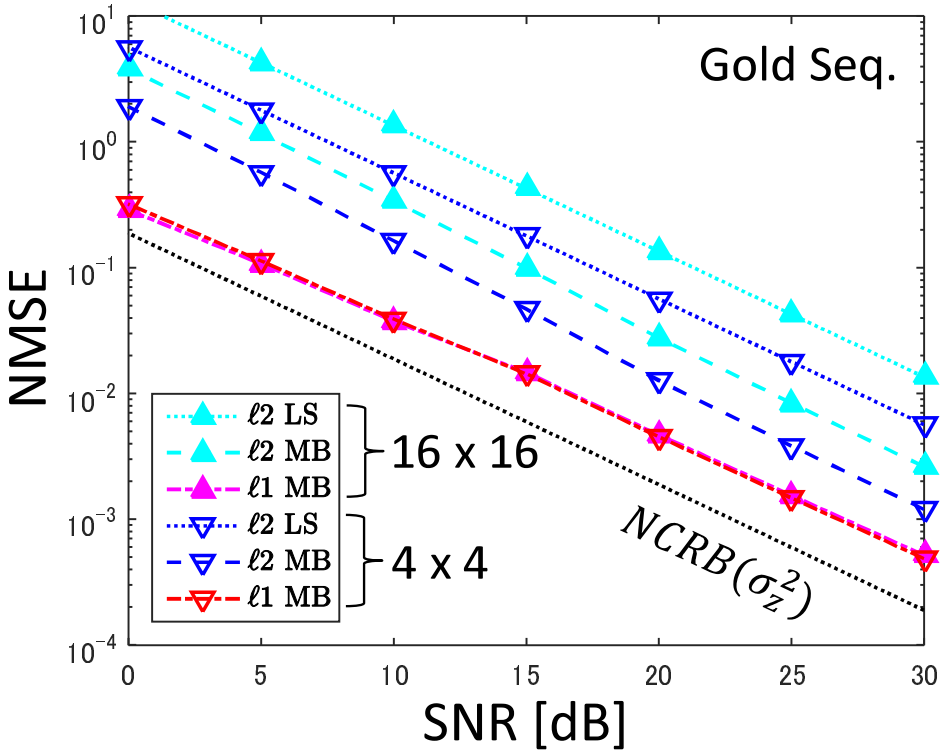

Fig. 2.20: The NMSE performance comparison between the 4×4 and 16×16 MIMO systems. The Gold sequences are generated by initializing the two shift registers with the indexes of the frame timing and the TX stream, where the generator polynomials are $\{1+x^3+x^7,\ 1+x+x^2+x^3+x^7\}$ and $\{1+x^4+x^9,\ 1+x^3+x^4+x^6+x^9\}$ for $N_t = 127$ and 511, respectively.

to the ideally chosen PN sequence.[18] It is worth noting that, as shown in Fig. 2.20, the NMSE improvement of the $\ell 1$ MB over the $\ell 2$ MB technique becomes significant in a large-scale 16×16 MIMO system, where $N_t = 511$ is assumed.

[18]It is expected that there exist ideally uncorrelated Gold sequences. However, the number of combinations for a length $N_t = 511$ Gold sequence in a 16×16 MIMO system becomes $\binom{511^2}{16} = 2.2 \times 10^{73}$. We are, hence, very difficult to find the optimal sequence combinations in a large-scale MIMO system, even though an off-line process.

2.C.3.2 Error analysis

Figs. 2.21 and 2.22 show the NMSE performance for possible CIR lengths w, $r_k < w \leq W$, where the random and PN TSs are used in Figs. 2.21 and 2.22, respectively. The received SNR is set at 15 dB. As observed from Figs. 2.21 and 2.22, $\bar{\delta}(w) = \bar{\delta}^{\perp}(w) + \bar{\delta}_{\mathcal{Z}}(w) + \bar{\delta}_{\Pi}(w)$ is satisfied, according to Theorem 1, where we define $\bar{\delta}(w) = \sum_{k=1}^{N_T} \mathbb{E}[\|\Delta\hat{\mathbf{H}}_{[w]k}^{MB}\|^2] / \mathbb{E}[\|\mathcal{H}\|^2]$. $\bar{\delta}^{\perp}(w)$, $\bar{\delta}_{\mathcal{Z}}(w)$ and $\bar{\delta}_{\Pi}(w)$ are defined similarly corresponding to the variances of (2.67), (2.68) and (2.69), respectively.

In the case the random TSs are used, as shown in Fig. 2.21, the $\ell 1$ MB can improve the NMSE of channel estimates significantly by selecting the CIR length as $\arg\min_w \{\bar{\delta}^{\perp}(w) \ll \bar{\delta}(w)\}$. In the case the TSs are generated with the PN sequences, however, the improvement by the CIR length constraint is very slight as shown in Fig. 2.22. This is because the whitening ratio becomes $\omega(w) = 1$ for $\forall w$ when the TSs are ideally uncorrelated sequences.

It should be emphasized that the NMSE of channel estimates is dominated by $\bar{\delta}_{\mathcal{Z}}(w)$ in the CIR length range $\{w \mid \bar{\delta}^{\perp}(w) \ll \bar{\delta}(w)\}$. Furthermore, in that CIR length range, the NMSE $\bar{\delta}_{\mathcal{Z}}(w)$ follows the analytical curve given by (2.70). In other words, the NMSE performance of the $\ell 1$ MB algorithm can be described via the whitening ratio (2.71). The next subsection shows, therefore, asymptotic property of the whitening ratio for system setups assuming very long training lengths and/or massive TX streams.

2.C.3.3 Asymptotic property of the whitening ratio

Fig. 2.23 illustrates asymptotic property of the whitening ratio for the length N_t of random TSs. The maximum CIR length W and the number of TX streams N_T are fixed at 31 and 4, respectively. As observed from Fig. 2.23, the whitening ratio becomes much greater than 1 for a short training length $N_t = WN_T$. However, because of (2.55), the whitening ratio can be decreased significantly by the CIR length constraint. Specifically,

$$\exists w \leq W,\ \mathrm{tr}\{\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}_{t[w]}}(l)\}/w \leq \mathrm{tr}\{\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}_{t[W]}}(l)\}/W \tag{2.72}$$

holds by Theorem 7.7.8 in [54]. In the case the training length is long enough, nevertheless, the $\ell 1$ MB cannot improve NMSE performance over the $\ell 2$ MB algorithm since $\omega(w) \approx 1$ for any CIR length constraint $\forall w$.

Fig. 2.24 depicts the whitening ratio (2.71) for massive numbers of the TX streams. The training length is set at $N_t = WN_T$ for the number N_T

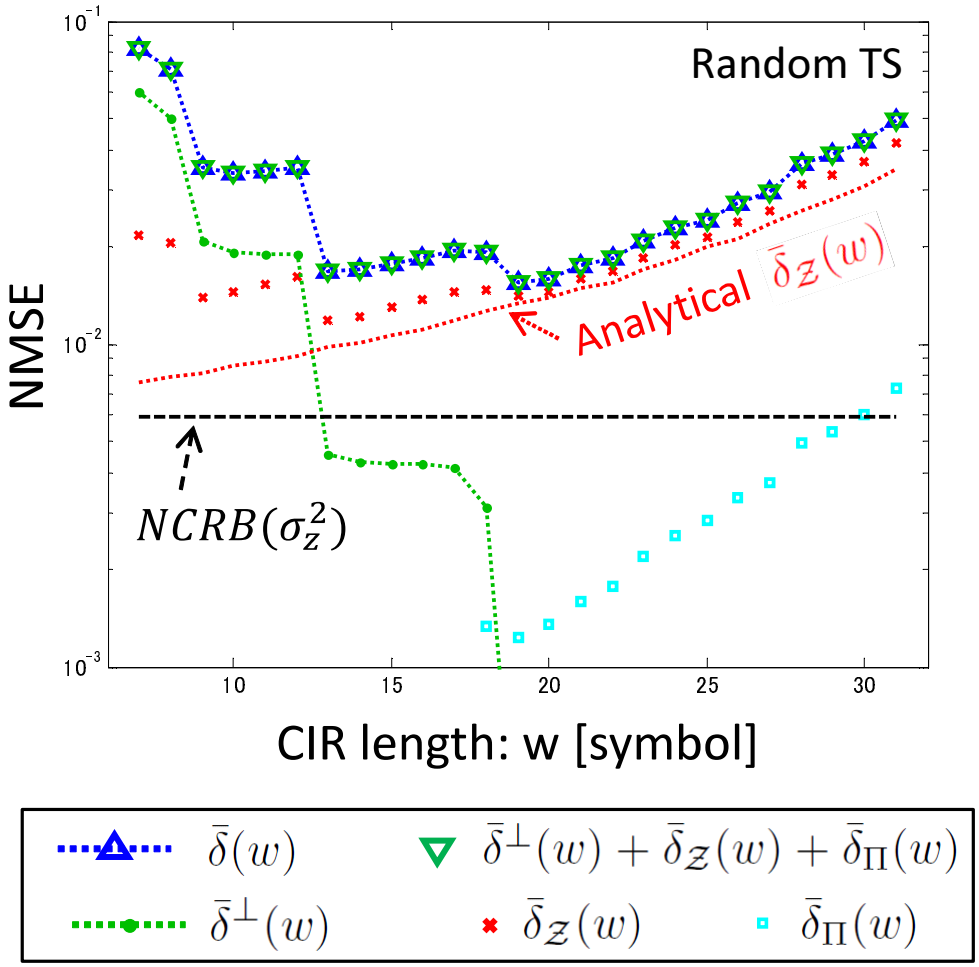

Fig. 2.21: The NMSE performance for possible CIR lengths w, where the random TSs are used. $\bar{\delta}(w)$ denotes the NMSE of the channel estimate $\hat{\mathcal{H}}_{[w]}^{MB}$. $\bar{\delta}^{\perp}(w), \bar{\delta}_{\mathcal{Z}}(w)$ and $\bar{\delta}_{\Pi}(w)$ are normalized variances of (2.67), (2.68) and (2.69), respectively. The red dotted curve *Analytical* $\bar{\delta}_{\mathcal{Z}}(w)$ is the NMSE normalized (2.70) with $\mathbb{E}[\|\mathcal{H}\|^2]$.

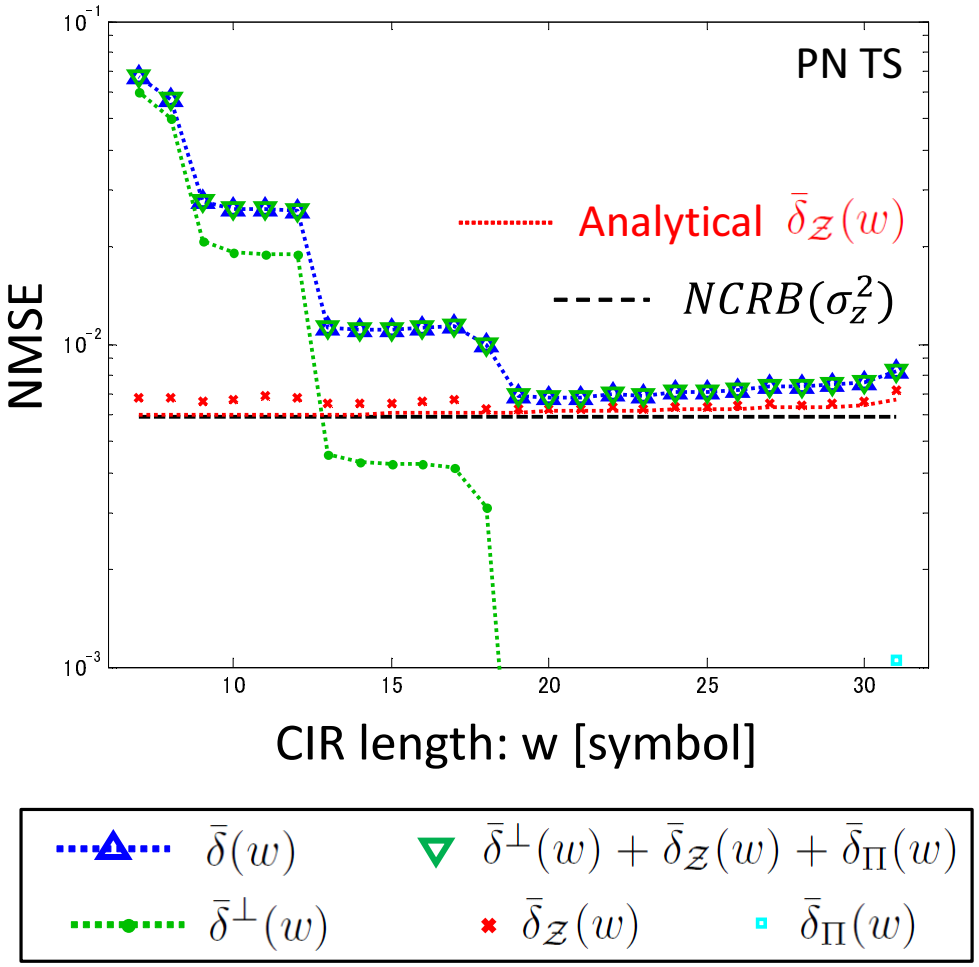

Fig. 2.22: The NMSE performance for possible CIR lengths w, where the PN TSs are used. $\bar{\delta}(w)$ denotes the NMSE of the channel estimate $\hat{\mathcal{H}}_{[w]}^{MB}$. $\bar{\delta}^{\perp}(w), \bar{\delta}_{\mathcal{Z}}(w)$ and $\bar{\delta}_{\Pi}(w)$ are normalized variances of (2.67), (2.68) and (2.69), respectively. The red dotted curve *Analytical* $\bar{\delta}_{\mathcal{Z}}(w)$ is the NMSE normalized (2.70) with $\mathbb{E}[\|\mathcal{H}\|^2]$.

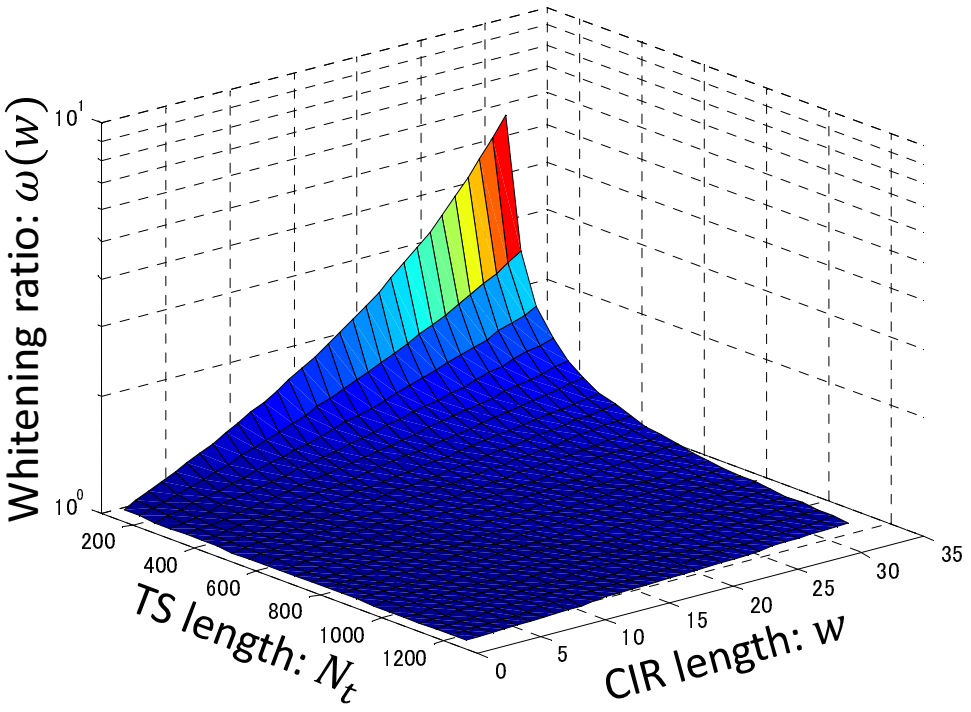

Fig. 2.23: The whitening ratio $\omega(w)$ (2.71) for the TS length. Random TSs are assumed.

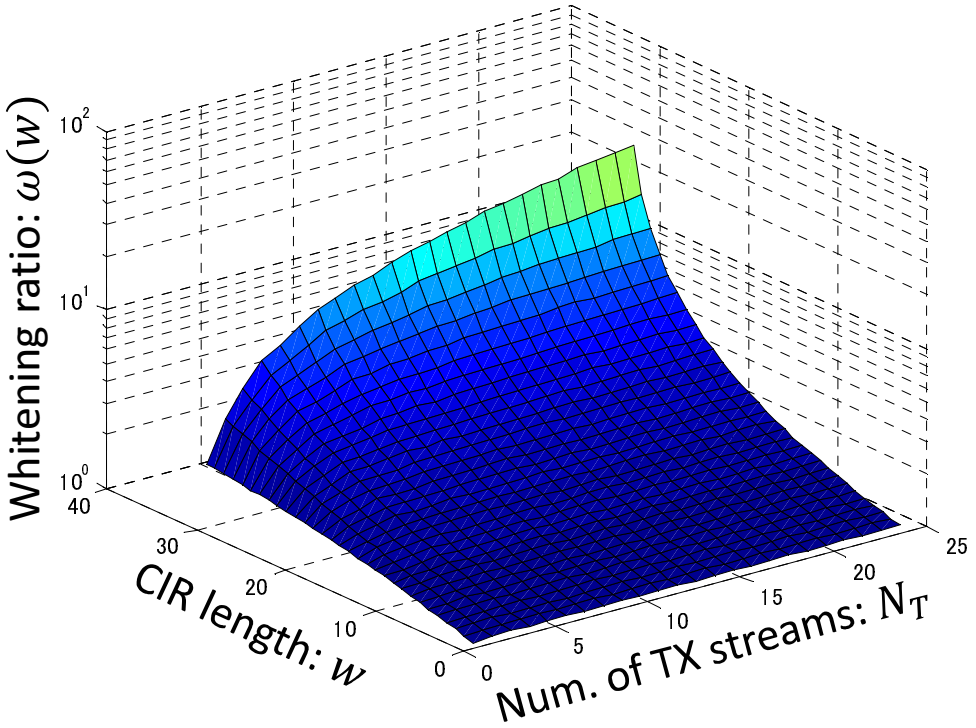

Fig. 2.24: The whitening ratio $\omega(w)$ (2.71) for the number of the TX streams. Random TSs are assumed.

of the TX streams, where the maximum CIR length W is fixed at 31. As shown in Fig. 2.24, the whitening ratio deteriorates as the number of TX streams increases. Therefore, the $\ell 1$ MB algorithm is expected to improve NMSE performance significantly in a massive MIMO system [63] when ideally uncorrelated TSs are not used. In the case $N_T = 24$ for example, the $\ell 1$ MB has a possibility to achieve up to 14 dB of NMSE gain over the $\ell 2$ MB. However, in a SISO or SIMO system, the NMSE gain becomes at most 3 dB since $\omega(w) \leq 2$ for $\forall w \leq W$.

2.C.4 Summary

In the case the ideally uncorrelated TSs are not used, the subspace-based $\ell 2$ MB technique can suffer from the noise enhancement. This is because the noise whitening in the MB algorithm is not accurate enough. The $\ell 1$ MB algorithm can, however, compensate for the problem according to the property (2.72), if the length w of the effective CIRs above the noise level is shorter than the maximum CIR length W assumed in the system.

Furthermore, this appendix has discussed the asymptotic NMSE performance of the $\ell 1$ MB algorithm via the whitening ratio $\omega(w)$. The whitening ratio deteriorates as the TS length decreases or the number of TX streams increases. The $\ell 1$ MB algorithm can, therefore, improve the NMSE performance over the conventional $\ell 2$ MB technique in a massive MIMO system when the TSs are not long enough and not ideally uncorrelated.

Chapter 3

Spectrally Efficient Frame Format–Aided Turbo Receiving Techniques

Cyclic prefix-aided block transmission has been recently gaining popularity in block transmission systems such as in SC-FDMA and/or OFDMA. One of the benefits of utilizing CP is to reduce the computational complexity for signal detection while keeping the robustness against fading frequency selectivity. The CP-transmission, on the other hand, imposes an overhead in the transmission format structure. It is hence preferable to minimize the length of the CP to improve the transmission energy- and spectrum-efficiencies. However, it causes serious degradation in BER performance if the length of the CP is shorter than the actual length of the CIR.

Chained turbo equalization (CHATUE) proposed in [64] provides a solution to this problem. CHATUE makes it possible to perform the frequency domain equalization processing, even without a CP, while requiring the same order of computational complexity as that of conventional frequency domain turbo equalization with CP-transmission (TEQ-CP) [9–11]. Since CHATUE requires no CP-transmission, it provides us with more design flexibility in terms of energy- and spectral-efficiency tradeoff. In other words, CHATUE enables us to transmit more information bits or to use a lower rate code by utilizing the time duration allocated for the CP-transmission. Thereby, CHATUE has a potential to improve performance over TEQ-CP, as detailed in [65], in terms of required SNR or throughput efficiency. Never-

theless, the previously-proposed CHATUE–referred to as CHATUE version 1 (CHATUE1)–has the following two problems, which are the consequence of eliminating CP-transmission.

1. *Latency*: the CHATUE algorithms studied so far in [64], [65], [66] require a processing latency three times that of TEQ-CP, since it performs iterations over at least three blocks (past, current and future blocks) to cancel the IBI. On the other hand, TEQ-CP performs turbo iterations within the current block alone.

2. *Noise Enhancement*: CHATUE1 utilizes a so called **J**-matrix [67] to retrieve the circulant structure of the channel matrix. However, a part of the signal after the transformation suffers from noise enhancement because of the multiplication of the **J**-matrix, as detailed in Section 3.1.3. The SNR at the equalizer output of CHATUE1, as a consequence, is decreased compared to that of TEQ-CP.

This chapter shows that Problem 1) can be easily solved under a practical assumption on the training sequence transmission. For Problem 2), this chapter proposes a novel algorithm, CHATUE version 2 (CHATUE2).

Furthermore, this chapter proposes a new channel estimation technique, chained turbo estimation (CHATES), that inherits the CHATUE concept, to pursue further improvement of the spectrum efficiency. The required length N_t of the TS is determined according to the length W of CIR. Conventional LS-based estimation techniques requires $N_t \geq W(N_T+1)$ to achieve accurate channel estimates if the transmission format does not have a GI between the TS and its neighboring segments. However, the CHATES requires a TS length of only $N_t = WN_T$, while it achieves the CRB asymptotically.

This chapter is organized as follows. Section 3.1 reviews the conventional CHATUE1 technique and discusses the above-mentioned problems in detail. The new CHATUE2 algorithm is also shown in Section 3.1. Section 3.2 proposes the new turbo channel estimation technique, CHATES. Section 3.3 presents results of computer simulations conducted to verify the effectiveness of the proposed techniques. Specifically, BER performance versus $\mathrm{E_b/N_0}$ is shown to validate if the proposed techniques improve the spectral efficiency over the conventional techniques. Section 3.4 summarizes this chapter with concluding remarks.

3.1 Channel Equalization

3.1.1 Signal model for channel equalization

The signal model (2.1) can be re-formulated to develop MIMO channel equalization algorithms. Assuming a TS is transmitted at the head of every burst, the vectorized received data segment $\mathtt{y}_d(l) \in \mathbb{C}^{N_R\tilde{N}_d \times 1}$ for the transmitted burst in the current burst timing l can be described as

$$\mathtt{y}_d(l) \;=\; \mathtt{H}(l)\mathtt{s}_d(l) + \mathtt{H}'(l)\mathtt{s}'_d(l) + \mathtt{H}''(l+1)\mathtt{s}''_d(l+1) + \mathtt{z}_d, \qquad (3.1)$$

where the noise vector $\mathtt{z}_d$ follows $\mathcal{CN}(\mathbf{0}_{N_R\tilde{N}_d}, \sigma_z^2\mathbf{I}_{N_R\tilde{N}_d})$ with $\tilde{N}_d = N_d + L$. The length N_TN_d signal vectors are defined as

$$\begin{aligned}
\mathtt{s}_d(l) &= [\mathbf{x}_{d,1}^{\mathrm{T}}(l), \cdots, \mathbf{x}_{d,N_T}^{\mathrm{T}}(l)]^{\mathrm{T}}, \\
\mathtt{s}'_d(l) &= [\mathtt{s}'_{d,1}(l)^{\mathrm{T}}, \cdots, \mathtt{s}'_{d,N_T}(l)^{\mathrm{T}}]^{\mathrm{T}}, \\
\mathtt{s}''_d(l) &= [\mathtt{s}''_{d,1}(l+1)^{\mathrm{T}}, \cdots, \mathtt{s}''_{d,N_T}(l+1)^{\mathrm{T}}]^{\mathrm{T}},
\end{aligned}$$

where $\mathtt{s}'_{d,k}(l) = [\mathbf{0}_{1\times(N_d-N_t)} \;\; \mathbf{x}_{t,k}^{\mathrm{T}}(l)]^{\mathrm{T}}$ and $\mathtt{s}''_{d,k}(l+1) = [\mathbf{x}_{t,k}^{\mathrm{T}}(l+1) \;\; \mathbf{0}_{1\times(N_d-N_t)}]^{\mathrm{T}}$. The $N_R\tilde{N}_d \times N_TN_d$ CIR matrix $\mathtt{H}(l)$ is defined as

$$\mathtt{H}(l) \;=\; \begin{bmatrix} \mathtt{H}_{\{1,1\}}(l) & \cdots & \mathtt{H}_{\{1,N_T\}}(l) \\ \vdots & \ddots & \vdots \\ \mathtt{H}_{\{N_R,1\}}(l) & \cdots & \mathtt{H}_{\{N_R,N_T\}}(l) \end{bmatrix}, \qquad (3.2)$$

where the (i,j)-th submatrix $\mathtt{H}_{\{i,j\}}(l)$ denotes the $\tilde{N}_d \times N_d$ Toeplitz matrix whose first column vector is $[\mathbf{h}_{i,j}^{\mathrm{T}}(l) \; \mathbf{0}_{1\times(\tilde{N}_d-W)}]^{\mathrm{T}}$ with $\mathbf{h}_{i,j}(l)$ being the length-W symbol CIR vector for the (i,j)-th Rx-Tx link. The $N_R\tilde{N}_d \times N_TN_d$ CIR matrices $\mathtt{H}'(l)$ and $\mathtt{H}''(l+1)$ are, similar to (3.2), structured with submatrices given by

$$\mathtt{H}'_{\{i,j\}}(l) \;=\; \begin{bmatrix} \mathbf{O}_{L\times(N_d-L)} & \mathtt{H}'_{\nabla\{i,j\}}(l) \\ \mathbf{O}_{N_d\times(N_d-L)} & \mathbf{O}_{N_d\times L} \end{bmatrix} \qquad (3.3)$$

and

$$\mathtt{H}''_{\{i,j\}}(l+1) \;=\; \begin{bmatrix} \mathbf{O}_{N_d\times L} & \mathbf{O}_{N_d\times(N_d-L)} \\ \mathtt{H}''_{\Delta\{i,j\}}(l+1) & \mathbf{O}_{L\times(N_d-L)} \end{bmatrix}, \qquad (3.4)$$

respectively, where we define

$$\mathtt{H}'_{\nabla\{i,j\}}(l) \;=\; \begin{bmatrix} h_{i,j}(W;l) & h_{i,j}(W-1;l) & \cdots & h_{i,j}(2;l) \\ & h_{i,j}(W;l) & \cdots & h_{i,j}(3;l) \\ & & \ddots & \vdots \\ 0 & & & h_{i,j}(W;l) \end{bmatrix},$$

and

$$\mathtt{H}''_{\Delta\{i,j\}}(l) \;=\; \begin{bmatrix} h_{i,j}(1;l) & & & 0 \\ h_{i,j}(2;l) & h_{i,j}(1;l) & & \\ \vdots & \vdots & \ddots & \\ h_{i,j}(L;l) & h_{i,j}(L-1;l) & \cdots & h_{i,j}(1;l) \end{bmatrix}.$$

3.1.2 CHATUE1

By combining CHATUE1 [66] with the MIMO turbo equalization shown in [10,11], the equalizer output of MIMO CHATUE1 for the k-th TX stream can be described as

$$\begin{aligned} \mathtt{q}_k^{[1]}(l) \;&=\; \left(\mathbf{I}_{N_d} + \mathbf{\Gamma}_k(l)\hat{\mathtt{S}}_k(l)\right)^{-1} \\ &\quad \cdot \left[\mathbf{\Gamma}_k(l)\hat{\mathtt{s}}_{d,k}(l) + \mathbf{F}_{N_R}^{\mathrm{H}}\hat{\mathbf{\Phi}}_k^{\mathrm{H}}(l)\mathbf{\Omega}_k^{-1}(l)\mathbf{F}_{N_R}\tilde{\mathtt{r}}_d(l)\right], \end{aligned} \tag{3.5}$$

where the residual $\tilde{\mathtt{r}}_d(l) \in \mathbb{C}^{N_R N_d \times 1}$ is given by

$$\begin{aligned} \tilde{\mathtt{r}}_d(l) \;&=\; \mathtt{r}_d(l) - \hat{\mathtt{r}}_d(l) && (3.6) \\ &=\; \tilde{\mathtt{J}}\mathtt{y}_d(l) - \tilde{\mathtt{J}}\hat{\mathtt{y}}_d(l) && (3.7) \end{aligned}$$

with

$$\hat{\mathtt{y}}_d(l) \;=\; \hat{\mathtt{H}}(l)\hat{\mathtt{s}}_d(l) + \hat{\mathtt{H}}'(l)\hat{\mathtt{s}}'_d(l) + \hat{\mathtt{H}}''(l+1)\hat{\mathtt{s}}''_d(l+1). \tag{3.8}$$

The $\tilde{\mathtt{J}}$-matrix denotes $\mathbf{I}_{N_R} \otimes \mathtt{J}$, where the $\mathtt{J}$-matrix proposed in [67] are defined as

$$\mathtt{J} \;=\; \begin{pmatrix} \mathbf{O}_{(N_d-L)\times L} & \mathbf{I}_{N_d} \\ \mathbf{I}_L & \end{pmatrix} \in \mathbb{R}^{N_d \times (N_d+L)}. \tag{3.9}$$

Moreover, $\hat{\mathtt{S}}_k(l) = \mathtt{DIAG}\{\hat{\mathtt{s}}_{d,k}(l) \odot \hat{\mathtt{s}}^*_{d,k}(l)\} \approx (\|\hat{\mathtt{s}}_{d,k}(l)\|^2/N_d)\mathbf{I}_{N_d}$ and

$$\hat{\mathbf{\Phi}}_k(l) \;=\; \mathbf{F}_{N_R}\tilde{\mathtt{J}}\hat{\mathtt{H}}_k(l)\mathbf{F}^{\mathrm{H}} \tag{3.10}$$

is an $N_R N_d \times N_d$ sub-block-wise diagonalized matrix[1], where $\hat{\mathtt{H}}_k(l) \in \mathbb{C}^{N_R \tilde{N}_d \times N_d}$ denotes a channel estimate matrix for the k-th TX stream which corresponds to the k-th column block in (3.2). $\mathtt{DIAG}(\mathbf{x})$ forms a diagonal matrix from its argument vector $\mathbf{x}$. $\mathbf{F}_N$ denotes $\mathbf{I}_N \otimes \mathbf{F}$, where $\mathbf{F} \in \mathbb{C}^{N_d \times N_d}$ is the DFT matrix whose entry is defined by (1.4). The matrix $\mathbf{\Omega}(l) \in \mathbb{C}^{N_R N_d \times N_R N_d}$ in (3.5) is given by

$$\mathbf{\Omega}_k(l) \;=\; \mathbf{F}_{N_R} \mathbf{\Sigma}_k(l) \mathbf{F}_{N_R}^{\mathrm{H}}, \tag{3.11}$$

where

$$\begin{aligned} \mathbf{\Sigma}_k(l) \;=\;& \tilde{\mathtt{J}}\hat{\mathtt{H}}_k(l) \mathbf{\Lambda}_k(l) \{\tilde{\mathtt{J}}\hat{\mathtt{H}}_k(l)\}^H \\ &+ \tilde{\mathtt{J}}\hat{\mathtt{H}}'_k(l) \mathbf{\Lambda}'_k(l) \{\tilde{\mathtt{J}}\hat{\mathtt{H}}'_k(l)\}^{\mathrm{H}} \\ &+ \tilde{\mathtt{J}}\hat{\mathtt{H}}''_k(l+1) \mathbf{\Lambda}''_k(l+1) \{\tilde{\mathtt{J}}\hat{\mathtt{H}}''_k(l+1)\}^{\mathrm{H}} \\ &+ \sigma_n^2 \tilde{\mathtt{J}}\tilde{\mathtt{J}}^{\mathrm{H}} \end{aligned} \tag{3.12}$$

with

$$\begin{aligned} \mathbf{\Lambda}_k(l) \;&=\; \mathbb{E}\left[\{\hat{\mathtt{s}}_{d,k}(l) - \mathbf{s}_{d,k}(l)\} \{\hat{\mathtt{s}}_{d,k}(l) - \mathbf{s}_{d,k}(l)\}^{\mathrm{H}} \right], \\ \mathbf{\Lambda}'_k(l) \;&=\; \mathbb{E}\left[\{\hat{\mathtt{s}}'_{d,k}(l) - \mathbf{s}'_{d,k}(l)\} \{\hat{\mathtt{s}}'_{d,k}(l) - \mathbf{s}'_{d,k}(l)\}^{\mathrm{H}} \right] \end{aligned}$$

and

$$\mathbf{\Lambda}''_k(l+1) \;=\; \mathbb{E}\left[\{\hat{\mathtt{s}}''_{d,k}(l+1) - \mathbf{s}''_{d,k}(l+1)\} \cdot \{\hat{\mathtt{s}}''_{d,k}(l+1) - \mathbf{s}''_{d,k}(l+1)\}^{\mathrm{H}} \right].$$

However, taking account of $\mathbf{\Lambda}'_k(l) = \mathbf{\Lambda}''_k(l+1) = \mathbf{O}$, because $\hat{\mathtt{s}}'_{d,k}(l)$ and $\hat{\mathtt{s}}''_{d,k}(l+1)$ are the known training sequence, (3.11) is reduced to (3.14):

$$\mathbf{\Omega}_k(l) \;=\; \mathbf{F}_{N_R} \left\{ \tilde{\mathtt{J}}\hat{\mathtt{H}}_k(l) \mathbf{\Lambda}_k(l) (\tilde{\mathtt{J}}\hat{\mathtt{H}}_k(l))^{\mathrm{H}} + \sigma_z^2 \tilde{\mathtt{J}}\tilde{\mathtt{J}}^{\mathrm{H}} \right\} \mathbf{F}_{N_R}^{\mathrm{H}} \tag{3.13}$$

$$\approx\; \hat{\mathbf{\Phi}}_k(l) \mathbf{\Delta}_k(l) \hat{\mathbf{\Phi}}_k(l)^{\mathrm{H}} + \sigma_z^2 \frac{N_d + L}{N_d} \mathbf{I}_{N_R N_d}, \tag{3.14}$$

with approximations (3.15) and (3.16) proposed in [9–11] and [66], respectively:

$$\mathbf{\Delta}_k(l) = \left(\sigma_x^2 - \|\hat{\mathtt{s}}_{d,k}(l)\|^2 / N_d\right) \mathbf{I}_{N_R N_d} \;\approx\; \mathbf{F}_{N_T} \mathbf{\Lambda}_k(l) \mathbf{F}_{N_T}^{\mathrm{H}}, \tag{3.15}$$

$$\sigma_z^2 \frac{N_d + L}{N_d} \mathbf{I}_{N_R N_d} = \sigma_z^2 \frac{\mathrm{tr}(\tilde{\mathtt{J}}\tilde{\mathtt{J}}^{\mathrm{H}})}{N_R N_d} \mathbf{I}_{N_R N_d} \;\approx\; \sigma_z^2 \mathbf{F}_{N_R} \tilde{\mathtt{J}}\tilde{\mathtt{J}}^{\mathrm{H}} \mathbf{F}_{N_R}^{\mathrm{H}}. \tag{3.16}$$

[1]The n-th $N_d \times N_d$ sub-block of $\hat{\mathbf{\Phi}}_k(l)$ is a diagonal matrix, the diagonal entry of which represents the frequency response corresponding to the time domain CIR vector $\mathbf{h}_{n,k}$ for the (n,k)-th Rx-Tx link.

Similarly, $\mathbf{\Gamma}_k(l) \in \mathbb{C}^{N_d \times N_d}$ is approximated by (3.18).

$$\begin{aligned} \mathbf{\Gamma}_k(l) &= \mathtt{diag}\left[(\tilde{\mathtt{J}}\hat{\mathtt{H}}_k(l))^{\mathrm{H}}\mathbf{\Sigma}_k^{-1}(l)\tilde{\mathtt{J}}\hat{\mathtt{H}}_k(l)\right] && (3.17) \\ &\approx \frac{1}{N_d}\mathrm{tr}\left[\hat{\mathbf{\Phi}}_k^{\mathrm{H}}(l)\mathbf{\Omega}_k^{-1}(l)\hat{\mathbf{\Phi}}_k(l)\right]\mathbf{I}_{N_d}. && (3.18) \end{aligned}$$

We assume the final output of CHATUE1 $\mathtt{q}_k^{[1]}(l)$ can be approximated as an equivalent Gaussian channel output [6], [58] having input $\mathtt{s}_{d,k}(l)$, as

$$\mathtt{q}_k^{[1]}(l) = \mu_{\mathtt{q}[1],k}(l)\mathtt{s}_{d,k}(l) + \mathtt{z}_{\mathtt{q}[1],k}(l), \tag{3.19}$$

where

$$\begin{aligned} \mu_{\mathtt{q}[1],k}(l) &= \frac{1}{N_d}\mathrm{tr}\left\{\mathbb{E}[\mathtt{q}_k^{[1]}(l)\mathtt{s}_{d,k}^{\mathrm{H}}(l)]\right\} \\ &= \frac{\|\mathtt{s}_{d,k}(l)\|^2}{N_d^2}\mathrm{tr}\left\{(\mathbf{I}_{N_d} + \mathbf{\Gamma}_k(l)\hat{\mathtt{S}}_k(l))^{-1}\mathbf{\Gamma}_k(l)\right\} && (3.20) \end{aligned}$$

and $\mathtt{z}_{\mathtt{q}[1],k}(l) \sim \mathcal{CN}(\mathbf{0}_{N_d}, \sigma^2_{\mathtt{q}[1],k}(l)\mathbf{I}_{N_d})$ with

$$\sigma^2_{\mathtt{q}[1],k}(l) = \mu_{\mathtt{q}[1],k}(l)(1 - \mu_{\mathtt{q}[1],k}(l)). \tag{3.21}$$

We finally convert the equalizer output $\mathtt{q}_k^{[1]}(l)$ into its corresponding extrinsic LLR, as

$$\lambda^e_{\mathrm{EQU},k}(l) = \frac{4\Re(\mathtt{q}_k^{[1]}(l))}{1 - \mu_{\mathtt{q}[1],k}(l)}, \tag{3.22}$$

where $\Re(\mathbf{v})$ denotes the real part of the complex vector $\mathbf{v}$.

3.1.3 Noise enhancement with CHATUE1

By utilizing the $\tilde{\mathtt{J}}$-matrix, CHATUE1 has the potential to improve the spectral- and/or energy-efficiencies while keeping the computational complexity order equivalent to that of TEQ-CP. However, CHATUE1 inevitably incurs a noise enhancement problem shown as follows, where the subscript k of TX stream is omitted for the sake of simplicity.

After enough iterations, we can assume $\mathbb{E}[||\hat{\mathbf{s}}_d(l)||^2]/N_d \to 1$ at a certain SNR.[2] The mean (3.20) converges to

$$\mu_{\mathbf{q}[1]} \;\to\; \frac{N_d}{N_d + (N_d + L)\sigma_z^2}, \tag{3.23}$$

as described in Appendix. The variance of the equivalent Gaussian channel output (3.21) also converges into

$$\sigma^2_{\mathbf{q}[1]}(l) \;\to\; \frac{N_d(N_d + L)\sigma_z^2}{\{N_d + (N_d + L)\sigma_z^2\}^2}. \tag{3.24}$$

According to [9], the mean $\mu_{\mathbf{q}[\mathrm{CP}]}$ and the variance $\sigma^2_{\mathbf{q}[\mathrm{CP}]}$ of the output of TEQ-CP converge into:

$$\mu_{\mathbf{q}[\mathrm{CP}]} \;\to\; \frac{1}{1 + \sigma_z^2}, \tag{3.25}$$

$$\sigma^2_{\mathbf{q}[\mathrm{CP}]} \;\to\; \frac{\sigma_z^2}{\{1 + \sigma_z^2\}^2}, \tag{3.26}$$

respectively, when $\mathrm{E}[||\hat{\mathbf{s}}_d(l)||^2] \to 1$.

The asymptotic SNR, $\mathrm{SNR}_{\mathbf{q}[1]}$, of the equalizer output with CHATUE1 is reduced to

$$\mathrm{SNR}_{\mathbf{q}[1]} = \frac{\mu^2_{\mathbf{q}[1]}}{\sigma^2_{\mathbf{q}[1]}} \;\to\; \frac{N_d}{(N_d + L)\sigma_z^2}. \tag{3.27}$$

Similarly, the asymptotic SNR, $\mathrm{SNR}_{\mathbf{q}[\mathrm{CP}]}$, of the equalizer output with TEQ-CP is reduced to

$$\mathrm{SNR}_{\mathbf{q}[\mathrm{CP}]} = \frac{\mu^2_{\mathbf{q}[\mathrm{CP}]}}{\sigma^2_{\mathbf{q}[\mathrm{CP}]}} \;\to\; \frac{1}{\sigma_z^2}. \tag{3.28}$$

The SNR ratio at the equalizer output of CHATUE1 to that of TEQ-CP is, hence,

$$\frac{1}{2} \;\le\; \frac{\mathrm{SNR}_{\mathbf{q}[1]}}{\mathrm{SNR}_{\mathbf{q}[\mathrm{CP}]}} \;=\; \frac{N_d}{N_d + L} \;\le\; 1. \tag{3.29}$$

The inequality (3.29) is because $N_d \ge L \ge 0$. The final output (3.5) of CHATUE1, thereby, suffers from the noise enhancement of up to 3 dB over TEQ-CP as the IBI length L increases. Fig. 3.1 illustrates the noise enhancement problem of the CHATUE1 algorithm.

[2]The required SNR falls into the issue of matching between the equalizer and decoder's EXIT curves. However, it is out of the scope of this thesis.

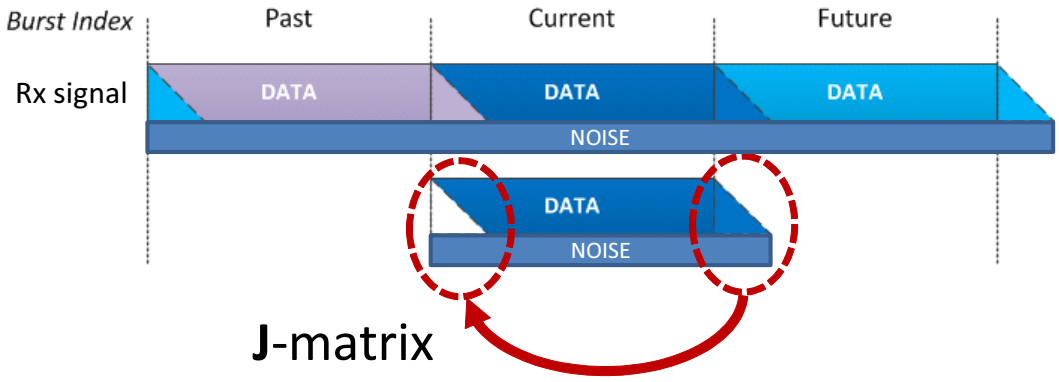

Fig. 3.1: Noise enhancement problem of CHATUE1 algorithm. Intuitively, the operation with **J**-matrix adds the leaked current signal to the future block onto the head part of the current signal, which causes the noise enhancement problem.

3.1.4 CHATUE2

A motivation of utilizing the $\tilde{\mathtt{J}}$-matrix in CHATUE1 is to reduce the computational complexity by restoring the circulant structure of the channel matrix. Although $\mathtt{H}_k \in \mathbb{C}^{N_R(N_d+L)\times N_d}$ is a Toepliz matrix, $\tilde{\mathtt{J}}\mathtt{H}_k \in \mathbb{C}^{N_R N_d \times N_d}$ becomes a circulant matrix. Thereby, it is possible to reduce the complexity by exploiting frequency domain processing, since $\mathbf{F}_{N_R}\tilde{\mathtt{J}}\mathtt{H}_k\mathbf{F}^{\mathrm{H}}$ is a sub-block-wise diagonalized matrix. On the other hand, the CHATUE1 incurs the noise enhancement problem due to the exploitation of the $\tilde{\mathtt{J}}$-matrix, as detailed in Section 3.1.3. To cope with the noise enhancement problem, we propose a new algorithm, CHATUE2, by introducing a new circulant property restoration method, as follows.

$$\mathtt{r}_d(l) \;\approx\; \bar{\mathtt{r}}_d(l) \triangleq \tilde{\mathtt{J}}_L(1-\beta)\mathtt{y}_d(l) + \tilde{\mathtt{G}}_L(\beta)\hat{\mathtt{y}}_d(l), \tag{3.30}$$

with $\tilde{\mathtt{J}}_L(1-\beta) = \mathbf{I}_{N_R} \otimes \mathtt{J}_L(1-\beta)$ and $\tilde{\mathtt{G}}_L(1-\beta) = \mathbf{I}_{N_R} \otimes \mathtt{G}_L(1-\beta)$ where $N_d \times (N_d + L)$ matrices $\mathtt{J}_L$ and $\mathtt{G}_L$ are respectively defined as

$$\mathtt{J}_L(1-\beta) \;=\; \begin{pmatrix} \mathbf{O}_{(N_d-L)\times L} & \mathbf{I}_{N_d} \\ (1-\beta)\mathbf{I}_L & \end{pmatrix}, \tag{3.31}$$

$$\mathtt{G}_L(\beta) \;=\; \begin{pmatrix} \mathbf{O}_{N_d} & \mathbf{O}_{(N_d-L)\times L} \\ & \beta\mathbf{I}_L \end{pmatrix}. \tag{3.32}$$

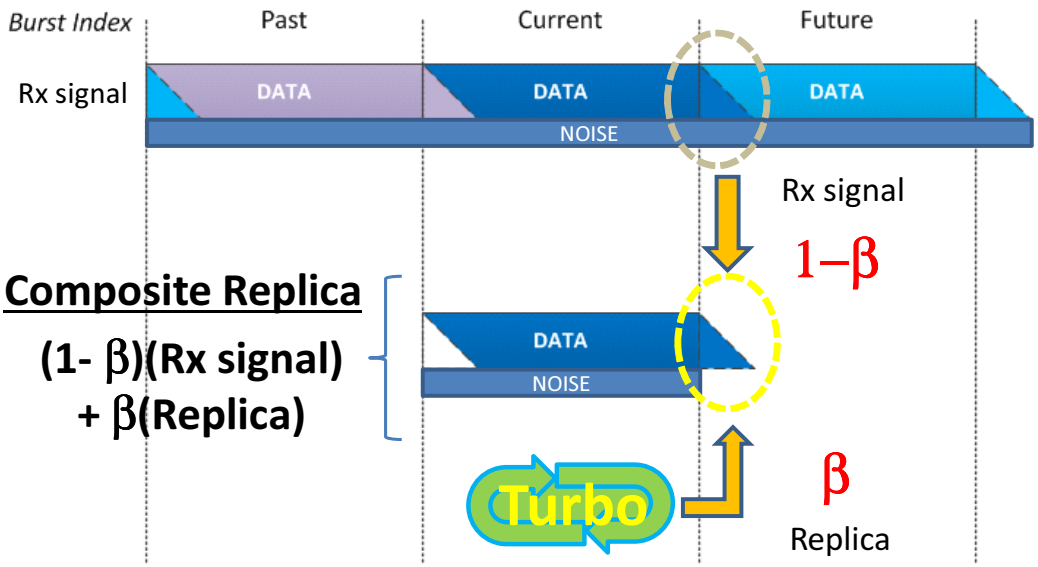

Fig. 3.2: Composite replica. Note that the operation with **J**-matrix does not cause the noise enhancement problem, if the leaked current signal to the future block does not contain the noise.

Note that the original J-matrix (3.9) is identical to $\mathtt{J}_L(1)$. Moreover, the entries for the i-th Rx antenna in (3.30) can be described as

$$\bar{\mathtt{r}}_d(l)|_{1+\mathfrak{d}:N_d+\mathfrak{d}} = \begin{bmatrix} y_d(L+1+\mathfrak{d};l) \\ \vdots \\ y_d(N_d+\mathfrak{d};l) \\ y_d(1+\mathfrak{d};l) \quad + \bar{y}_d(N_d+1+\mathfrak{d};l,\beta) \\ \vdots \\ y_d(L+\mathfrak{d};l) \quad + \bar{y}_d(N_d+L+\mathfrak{d};l,\beta) \end{bmatrix} \tag{3.33}$$

with the index offset $\mathfrak{d} = N_d(i-1)$, where the composite replica $\bar{y}(k;l,\beta)$ is defined as

$$\bar{y}(k;l,\beta) = (1-\beta)y_d(k;l) + \beta\hat{y}_d(k;l). \tag{3.34}$$

Fig. 3.2 shows a schematic diagram of the composite replica. We define the factor β such that the squared error between $\bar{\mathtt{y}}_d(l,\beta)$ and $\mathtt{c}_d(l) = \mathtt{H}(l)\mathtt{s}_d(l) + \mathtt{H}'(l)\mathtt{s}'_d(l) + \mathtt{H}''(l)\mathtt{s}''_d(l)$ is minimized, which can be formulated as

$$\beta = \arg\min_{\beta} \|\mathtt{c}_d(l) - \bar{\mathtt{y}}_d(l,\beta)\|^2, \tag{3.35}$$

where $\bar{\mathtt{y}}_d(l,\beta)$ denotes the vector version of (3.34), defined as $\bar{\mathtt{y}}_d(l,\beta) = (1-\beta)\mathtt{y}_d(l) + \beta\hat{\mathtt{y}}_d(l)$. By taking $\|\mathtt{c}_d(l) - \bar{\mathtt{y}}_d(l,\beta)\|^2 \geq 0$ into account, the problem (3.35) can be reduced by solving

$$\frac{\partial}{\partial\beta}\|\mathtt{c}_d(l) - \bar{\mathtt{y}}_d(l,\beta)\|^2 \;=\; 0. \tag{3.36}$$

Since $\mathtt{c}_d(l) = \mathtt{y}_d(l) - \mathtt{z}_d$, the solution to (3.35) is, therefore,

$$\beta \;=\; \frac{\sigma_z^2 N_R \tilde{N}_d}{\|\mathtt{y}_d(l) - \hat{\mathtt{y}}_d(l)\|^2}. \tag{3.37}$$

Accordingly, we rewrite (3.14) as,

$$\begin{aligned}\boldsymbol{\Omega}(l) \;&=\; \mathbf{F}_{N_R}\Big\{\tilde{\mathtt{J}}_L(1)\hat{\mathtt{H}}_k(l)\boldsymbol{\Lambda}_k(l)\{\tilde{\mathtt{J}}_L(1)\hat{\mathtt{H}}_k(l)\}^{\mathrm{H}} + \sigma_z^2\tilde{\mathtt{J}}_L(1-\beta)\tilde{\mathtt{J}}_L(1-\beta)^{\mathrm{H}}\Big\}\mathbf{F}_{N_R}^{\mathrm{H}} \\ &\approx\; \hat{\boldsymbol{\Phi}}_k(l)\boldsymbol{\Delta}_k(l)\hat{\boldsymbol{\Phi}}_k(l)^{\mathrm{H}} + \sigma_z^2\frac{N_d + (1-\beta)L}{N_d}\mathbf{I}_{N_R N_d}.\end{aligned} \tag{3.38}$$

3.1.5 Improvement of the noise enhancement by CHATUE2

The proposed CHATUE2 algorithm using (3.30) and (3.38) is expected to have the following advantageous points: At the first iteration, (3.30) with $\beta \approx 0$ is equivalent to the original $\mathtt{r}_d(l) = \tilde{\mathtt{J}}_L(1)\mathtt{y}_d(l)$ and CHATUE2 works in the same way as in CHATUE1. After enough iterations are performed, it is expected to satisfy both $\beta \to 1$ and $\mathbb{E}\left[||\hat{\mathtt{H}} - \mathtt{H}||^2\right] < \epsilon + \mathrm{aCRB}_{\bar{N}}(\sigma_z^2)$ with an arbitrary small positive value ϵ. The lower bound $\mathrm{aCRB}_{\bar{N}}(\sigma_z^2)$ of the estimation accuracy is given by (2.44) in Section 2.3.2.4. The channel matrix in $\bar{\mathtt{r}}_d$ approaches a matrix having a circulant structure when the estimate $\hat{\mathtt{H}}$

is accurate. Concretely, for the $\{i,j\}$-th Rx-Tx link, it can be observed that

$$\mathbf{J}_L(1-\beta)\mathtt{H}_{d\{i,j\}} + \mathbf{G}_L(\beta)\hat{\mathtt{H}}_{d\{i,j\}} \overset{\beta\to 1}{=}$$
$$\begin{bmatrix} h(W) & \cdots & h_{\{i,j\}}(2) & h_{\{i,j\}}(1) & & & \\ & \ddots & \vdots & h_{\{i,j\}}(2) & h_{\{i,j\}}(1) & & \\ & & h_{\{i,j\}}(W) & \vdots & h_{\{i,j\}}(2) & \ddots & \\ & & & h_{\{i,j\}}(W) & \vdots & \ddots & h_{\{i,j\}}(1) \\ \hat{h}_{\{i,j\}}(1) & & & & h_{\{i,j\}}(W) & & h_{\{i,j\}}(2) \\ \vdots & \ddots & & & & \ddots & \vdots \\ \hat{h}_{\{i,j\}}(L) & \cdots & \hat{h}_{\{i,j\}}(1) & & & & h_{\{i,j\}}(W) \end{bmatrix}, \tag{3.39}$$

where $h_{\{i,j\}}(w)$ and $\hat{h}_{\{i,j\}}(w)$ denote the w-th entries of the CIR vector $\mathbf{h}_{\{i,j\}}$ and its estimate $\hat{\mathbf{h}}_{\{i,j\}}$, respectively. The burst index l is omitted for the sake of simplicity. The convergence $\beta \to 1$ contributes to reducing the noise variance (3.21), through (3.18), (3.20) and (3.38). The mean $\mu_{\mathtt{q}[2]}$ and the variance $\sigma^2_{\mathtt{q}[2]}$ of the equalizer output with CHATUE2, respectively, converge into:

$$\mu_{\mathtt{q}[2]} \;\to\; \frac{N_d}{N_d + (N_d + (1-\beta)L)\sigma_z^2}, \tag{3.40}$$

$$\sigma^2_{\mathtt{q}[2]} \;\to\; \frac{N_d(N_d + (1-\beta)L)\sigma_z^2}{\{N_d + (N_d + (1-\beta)L)\sigma_z^2\}^2}, \tag{3.41}$$

when $\mathbb{E}[||\hat{\mathtt{s}}(l)||^2] \to 1$. Thereby, CHATUE2 improves the signal to noise power ratio $\mathrm{SNR}_{\mathtt{q}[2]}$ at the final equalizer output and it approaches that with TEQ-CP when $\beta \to 1$, as

$$\begin{aligned} \mathrm{SNR}_{\mathtt{q}[2]} &= \frac{\mu^2_{\mathtt{q}[2]}}{\sigma^2_{\mathtt{q}[2]}} \\ &\to \frac{N_d}{\{N_d + (1-\beta)L\}\,\sigma_z^2} \overset{\beta\to 1}{\longrightarrow} \frac{1}{\sigma_z^2} \;=\; \mathrm{SNR}_{\mathtt{q}[\mathrm{CP}]}. \end{aligned} \tag{3.42}$$

3.1.6 Computational complexity order

The computational complexity orders $\mathcal{O}(\cdot)$ required for the TEQ-CP, CHATUE1 and CHATUE2 techniques are summarized in Table 3.1. It should be emphasized that we have $\mathcal{O}(N_T N_R N_d \log N_d + N_T N_R W^2) = \mathcal{O}(N_T N_R N_d \log N_d)$

Table 3.1: Computational complexity orders for equalization algorithms

Algorithm	Complexity order	Details (i, $\cdots$, xiii)
TEQ-CP	$\mathcal{O}(N_T N_R N_d \log N_d)$	all except (viii, ix)
CHATUE1	$\mathcal{O}(N_T N_R N_d \log N_d + N_T N_R W^2)$	all except (viii, ix)
CHATUE2	$\mathcal{O}(N_T N_R N_d \log N_d + N_T N_R W^2)$	all (i, $\cdots$, xiii)

when $W \ll N_d$. Therefore, the complexity *order* needed for CHATUE2 becomes the same as that of the conventional TEQ-CP technique. The complexity order analysis is discussed further after detailing the complexity order needed for (3.5) which can describe the equalizer outputs of all the TEQ-CP, CHATUE1 and CHATUE2 techniques.

3.1.6.1 Details of the computational complexity order for the equalizer output (3.5)

Table 3.2 shows details of the complexity orders required for the equalizer output (3.5). The complexity orders indicated by the rows (i,ii,iii,v,viii,x) in Table 3.2 can be derived easily by noticing that the sub-block-wise diagonalized matrices $\hat{\mathbf{\Phi}}_k(l)$ and $\mathbf{\Omega}_k(l)$ are sparse. For (xi, xii), we assume that the FFT and inverse fast Fourier transform (IFFT) operations requires the complexity order of $\mathcal{O}(N \log N)$ for a length N input signal. The reasons for (iv, vi, vii, ix, xiii) are as follows:

(iv) The inversion of an $N \times N$ matrix requires the complexity order of $O(N^3)$, generally. However, by noticing the sparse structure of the $N_R N_d \times N_R N_d$ matrix $\mathbf{\Omega}_k(l)$, the calculation of $\mathbf{\Omega}_k^{-1}(l)$ is reduced to N_d problems of an $N_R \times N_R$ matrix inversion.

(vi) The complexity order required for (3.18) is dominated by the multiplication of the sparse matrices: $(\hat{\mathbf{\Phi}}_k^{\mathrm{H}}(l)\mathbf{\Omega}_k^{-1}(l)) \cdot \hat{\mathbf{\Phi}}_k(l)$.

(vii) The complexity order required for (3.8) is dominated by the term $\hat{\mathtt{H}}(l)\hat{\mathbf{s}}_d(l)$. The matrix-vector multiplication needs $\mathcal{O}(N_R N_T N_d^2)$ if it is computed in the time-domain purely. Similar to (3.5), however, notice

Table 3.2: Details of the complexity orders $\mathcal{O}(\cdot)$ in the equalizer output (3.5)

	Symbol	Eqn.	$\mathcal{O}(\cdot)$ for N_T TX streams
(i)	$\hat{\mathbf{\Phi}}_k(l)$	(3.10)	$\mathcal{O}(N_T \cdot N_R N_d \log N_d)$
(ii)	$\mathbf{\Delta}_k(l)$	(3.15)	$\mathcal{O}(N_T \cdot N_d)$
(iii)	$\mathbf{\Omega}_k(l)$	(3.14)	$\mathcal{O}(N_T \cdot N_R N_d)$
(iv)	$\mathbf{\Omega}_k^{-1}(l)$	(3.5)	$\mathcal{O}(N_T \cdot N_R^3 N_d)$
(v)	$\hat{\mathbf{\Phi}}_k^{\mathrm{H}}(l) \cdot \mathbf{\Omega}_k^{-1}(l)$	(3.5)	$\mathcal{O}(N_T \cdot N_R^2 N_d)$
(vi)	$\mathbf{\Gamma}_k(l)$	(3.18)	$\mathcal{O}(N_T \cdot N_R^2 N_d)$
(vii)	$\hat{\mathbf{y}}_d(l)$	(3.8)	$\mathcal{O}((N_R + N_T)N_d \log N_d + N_T N_R W^2)$
(viii)	β	(3.35)	$\mathcal{O}(N_R \tilde{N}_d)$
(ix)	$\bar{\mathtt{r}}_d$	(3.30)	$\mathcal{O}(N_R W)$
(x)	$\tilde{\mathtt{r}}_d$	(3.7)	$\mathcal{O}(N_R N_d)$
(xi)	$\mathbf{F}_{N_R} \cdot \tilde{\mathtt{r}}_d(l)$	(3.5)	$\mathcal{O}(N_T \cdot N_R \tilde{N}_d \log(N_R \tilde{N}_d))$
(xii)	$\mathbf{F}_{N_R}^{\mathrm{H}} \cdot \left\{ \hat{\mathbf{\Phi}}_k^{\mathrm{H}}(l) \mathbf{\Omega}_k^{-1}(l) \mathbf{F}_{N_R} \tilde{\mathtt{r}}_d(l) \right\}$	(3.5)	$\mathcal{O}(N_T \cdot N_R N_d \log(N_R N_d))$
(xiii)	$\left(\mathbf{I}_{N_d} + \mathbf{\Gamma}_k(l) \hat{\mathtt{S}}_k(l) \right)^{-1}$	(3.5)	$\mathcal{O}(N_T)$

that we may utilize the frequency-domain operations. Concretely,

$$\begin{aligned}\hat{\mathtt{H}}(l)\hat{\mathtt{s}}_d(l) &= \left\{\tilde{\mathtt{K}}\tilde{\mathtt{J}}\hat{\mathtt{H}}(l) - \left(\hat{\mathtt{H}}'(l) + \hat{\mathtt{H}}''(l)\right)\right\}\hat{\mathtt{s}}_d(l) \quad (3.43)\\ &= \tilde{\mathtt{K}}\mathbf{F}_{N_R}^{\mathrm{H}}\left[\hat{\mathbf{\Phi}}_1(l), \cdots, \hat{\mathbf{\Phi}}_{N_T}(l)\right]\mathbf{F}_{N_T}^{\mathrm{H}}\hat{\mathtt{H}}(l)\hat{\mathtt{s}}_d(l)\\ &\quad - \left(\hat{\mathtt{H}}'(l) + \hat{\mathtt{H}}''(l)\right)\hat{\mathtt{s}}_d(l), \quad (3.44)\end{aligned}$$

where $\tilde{\mathtt{K}} = \mathbf{I}_{N_R} \otimes \mathtt{K}$ with

$$\mathtt{K} = \begin{pmatrix} \mathbf{O}_{L\times(N_d-L)} & \mathbf{I}_L \\ & \mathbf{I}_{N_d} \end{pmatrix} \in \mathbb{R}^{\tilde{N}_d \times N_d}. \quad (3.45)$$

The complexity order needed for the first line of (3.44) is dominated by the FFT and IFFT operations:

$$\mathcal{O}(N_R N_d \log N_R N_d + N_T N_d \log N_T N_d) = \mathcal{O}((N_R + N_T) N_d \log N_d)$$

since $\hat{\mathbf{\Phi}}_k(l)$ is the sub-block-wise diagonalized matrix. The complexity order needed for the second line of (3.44) is $\mathcal{O}(N_R N_T W^2)$. Notice that the total complexity order of (vii) becomes

$$\mathcal{O}((N_R + N_T) N_d \log N_d + N_T N_R W^2) = \mathcal{O}((N_R + N_T) N_d \log N_d)$$

when $W \ll N_d$.

(ix) As discussed in (3.33), the operations with the $\tilde{\mathtt{J}}_L$ and $\tilde{\mathtt{G}}_L$ matrices performs the additions only for the $L = W - 1$ symbols in the N_R received signals.

(xiii) The matrix $(\mathbf{I}_{N_d} + \mathbf{\Gamma}_k(l)\hat{\mathbf{S}}_k(l))$ is proportional to the identity matrix. Thereby, the complexity order needed for the matrix inverse is $\mathcal{O}(1)$ for a TX stream.

3.1.6.2 TEQ-CP

The conventional TEQ-CP technique performs the equations in Table 3.2 indicated by all the rows (i, $\cdots$, xiii) except (viii, ix). The output of TEQ-CP is also described with (3.5) by assuming $\mathtt{J} = \mathbf{I}_{N_d}$. The matrix $\mathbf{\Omega}_k(l)$ for TEQ-CP is, therefore, denoted by (3.14) with $L = 0$. Notice that (vii) for TEQ-CP becomes $\mathcal{O}((N_R+N_T)N_d \log N_d)$. This is because $\mathtt{H}(l)$ is a circulant matrix due to the CP-transmission, and thereby, the second line of (3.44) is eliminated.

3.1.6.3 CHATUE1 and CHATUE2

The CHATUE2 algorithm performs all the equations of the rows (i, $\cdots$, xiii) in Table 3.2, whereas the CHATUE1 technique computes all the 13 equations except (viii, ix) which are needed to construct the composite replica. Note that the complexity order required for (viii, ix) is very minor. Hence, the complexity *order* required for the CHATUE2 algorithm is the same as that of CHATUE1, although the number of operations is increased slightly.

3.2 Channel Estimation

Turbo channel estimation can estimate the CIR accurately even though the TS length is short, since it extends the reference signal by utilizing the LLR of the transmitted data, fed back from the decoder. Obviously, a shorter TS is preferable from the viewpoint of the spectral efficiency. In practice, the TS length is designed such that $N_t \geq WN_T$ to estimate length W CIRs over N_T TX streams. With $N_t = WN_T$, however, the estimation accuracy is degraded because the input signal to the estimator suffers from IBI, as we can observe from the input data range (b) in Fig. 1.2 when $N_{G1} = N_{G2} = 0$.

To cope with this problem, we propose a new chained turbo channel estimation (CHATES) technique which performs IBI cancelation for channel estimation. The proposed technique is based on the concept of the CHATUE and improves the spectral efficiency without sacrificing the estimation accuracy after enough turbo iterations. It should be noted, however, CHATES can be applied to the transmission format with a CP as well.

3.2.1 Signal model for channel estimation

We re-formulate the signal model to develop channel estimation algorithms. The received signals $\mathcal{Y}_t(l)$ corresponding to the transmitted TS section can be described, as

$$\mathcal{Y}_t(l) \;=\; \mathcal{H}_c(l)\mathcal{X}_c(l) + \mathcal{H}_p(l-1)\mathcal{X}_p(l-1) + \mathcal{H}_f(l)\mathcal{X}_f(l) + \mathcal{Z}_t \quad (3.46)$$

where the matrices $\mathcal{H}_c(l)$ and $\mathcal{X}_c(l)$ denote the CIR matrix $\mathcal{H}(l)$ (2.2) and the TS matrix $\mathcal{X}_t(l)$ (2.8), respectively, for the current block l. The second and third terms on the right-hand side (RHS) of (3.46) are the IBI from the past and future blocks, where we denote the CIR matrices by $\mathcal{H}_p(l-1) = \mathcal{H}(l-1)$

and $\mathcal{H}_f(l) = \mathcal{H}(l)$ by assuming that the burst format shown in Fig. 2.3 is used. The matrices $\mathcal{X}_p(l-1)$ and $\mathcal{X}_f(l)$ are composed of a submatrix of the data matrix $\mathcal{X}_d$ (2.9). Specifically, they are defined by

$$\begin{aligned} \mathcal{X}_p(l-1) &= [\mathbf{X}_{p,1}^{\mathrm{T}}(l-1), \cdots, \mathbf{X}_{p,N_T}^{\mathrm{T}}(l-1)]^{\mathrm{T}}, & (3.47) \\ \mathcal{X}_f(l) &= [\mathbf{X}_{f,1}^{\mathrm{T}}(l), \cdots, \mathbf{X}_{f,N_T}^{\mathrm{T}}(l)]^{\mathrm{T}} & (3.48) \end{aligned}$$

with $W \times \tilde{N}_t$ submatrices

$$\begin{aligned} \mathbf{X}_{p,k}(l-1) &= \left[\tilde{\mathbf{X}}_{d,k}(l-1)|_{(\tilde{N}_D+N'_{G2}-W+2):\tilde{N}_D},\ \mathbf{O}_{W\times(N_t+N'_{G2})}\right] \\ \mathbf{X}_{f,k}(l) &= \left[\mathbf{O}_{W\times(N_t+N'_{G1})},\ \tilde{\mathbf{X}}_{d,k}(l)|_{1:(W-N'_{G1}-1)}\right], \end{aligned}$$

where $\tilde{\mathbf{X}}_{d,k}(l)$ is the $W \times \tilde{N}_D$ Toeplitz matrix, the first row vector of which is $[\tilde{\mathbf{x}}_{d,k}^{\mathrm{T}}(l), \mathbf{0}_{W-1}^{\mathrm{T}}]$ with

$$\tilde{\mathbf{x}}_{d,k}(l) = \begin{cases} [\mathbf{x}_{\mathrm{CP},k}^{\mathrm{T}}(l), \mathbf{x}_{d,k}^{\mathrm{T}}(l)]^{\mathrm{T}} & (N_{\mathrm{CP}} > 0) \\ \mathbf{x}_{d,k}(l) & (N_{\mathrm{CP}} = 0) \end{cases} \quad (3.49)$$

and $\tilde{N}_D = N_{\mathrm{CP}} + N_d + W - 1$. We denote, moreover, $N'_{G1} = \min(N_{G1}, W-1)$ and $N'_{G2} = \min(N_{G2}, W-1)$. Notice that the IBI components $\mathcal{X}_p$ and $\mathcal{X}_f$ vanish when both the guard interval lengths (N_{G1}, N_{G2}) are greater than $W-1$ since $\mathbf{X}_{p,k}(l-1) = \mathbf{X}_{f,k}(l) = \mathbf{O}_{W\times\tilde{N}_t}$. The j-th row vector in $\mathcal{Z}_t \in \mathbb{C}^{N_R\times\tilde{N}_t}$ follows $\mathcal{CN}(\mathbf{0}_{\tilde{N}_t}^{\mathrm{T}}, \sigma_z^2\mathbf{I}_{\tilde{N}_t})$.

3.2.2 CHATES

We define an IBI canceled version of the received training sequence $\tilde{\mathcal{Y}}_t^{[i]}(l) \in \mathbb{C}^{N_R\times\tilde{N}_t}$ for the current burst l at the i-th iteration as follows:

$$\tilde{\mathcal{Y}}_t^{[i]}(l) = \mathcal{Y}_t(l) - \left\{\hat{\mathcal{H}}_p^{[i-1]}(l-1)\,\hat{\mathcal{X}}_p^{[i-1]}(l-1) + \hat{\mathcal{H}}_f^{[i-1]}(l)\,\hat{\mathcal{X}}_f^{[i-1]}(l)\right\}, \quad (3.50)$$

where $\hat{\mathcal{H}}_p^{[i-1]}(l-1)$, $\hat{\mathcal{X}}_p^{[i-1]}(l-1)$, $\hat{\mathcal{H}}_f^{[i-1]}(l)$ and $\hat{\mathcal{X}}_f^{[i-1]}(l)$ are obtained as the result of the $(i-1)$-th turbo iteration. Any channel estimation algorithm shown in Chapter 2 can be performed with (3.50).

3.2.3 Self-supervised $\ell 1$ MB channel estimation (s-$\ell 1$ MB)

The channel estimation technique with (3.50) is expected to improve the estimation accuracy after performing enough turbo iterations. It should be noticed that, however, in the first iteration $i = 1$, the receiver does not have the soft replica matrix $\hat{\mathcal{X}}_f^{[i-1]}(l)$ of the current l-th burst's TX data sequence. For this problem, the $\ell 1$ MB channel estimation shown in Section 2.2.2 may be utilized to improve IBIs by the CIR length constraint. Nevertheless, the $\ell 1$ MB technique also requires the soft replica $\hat{\mathcal{X}}_f^{[i-1]}(l)$ in order to accurately determine the best CIR length with the AIC. Therefore, this section proposes a new *self-supervised* $\ell 1$ MB (s-$\ell 1$ MB) channel estimation algorithm [68] under the subspace channel model assumption.

3.2.3.1 Problem formulation

The s-$\ell 1$ MB channel estimation is a version of the $\ell 1$ MB technique which can be formulated by a conditional MMSE problem of (2.24), given $\boldsymbol{\Theta}_{\mathbf{U}} \triangleq [\mathtt{vec}\{\mathbf{U}_1\}^{\mathrm{T}}, \cdots, \mathtt{vec}\{\mathbf{U}_{N_T}\}^{\mathrm{T}}]^{\mathrm{T}}$, as

$$\hat{\mathcal{H}}_{[w]}^{\text{s-MB}}(l) \;=\; \underset{\mathcal{H}_{[w]}(l,\boldsymbol{\Theta})}{\arg\min} \underset{j \in \mathcal{J}_{L_M}(l)}{\mathbb{E}} \left[\mathcal{L}_{td}(j, \mathcal{H}_{[w]}(j, \boldsymbol{\Theta}) \;\mid\; \boldsymbol{\Theta}_{\mathbf{U}}) \right]. \tag{3.51}$$

The CIR length w is determined by

$$\hat{w} = \max_{1 \le k \le N_T} \hat{w}_k, \tag{3.52}$$

where

$$\hat{w}_k \;=\; \arg\min_{w} \mathbb{E}\left[\left\| \hat{\mathbf{H}}_{[w],k}^{MB}(l) - \mathbf{H}_k(l) \right\|^2 \right] \tag{3.53}$$

for the k-th TX stream.

3.2.3.2 Solution to (3.51)

It should be noticed that (3.51) aims to derive the optimal burst dependent parameter $\boldsymbol{\Theta}_{\mathbf{B}} = [\theta_{\mathbf{B},1}^{\mathrm{T}}, \cdots, \theta_{\mathbf{B},N_T}^{\mathrm{T}}]^{\mathrm{T}}$ if the CIR parameter description (2.25)

is correct. Hence, as detailed in [24], problem (3.51) is reduced to an LS problem, the solution of which can be given by $\hat{\mathcal{H}}^{\text{s-}MB}_{[w]}(l) = [\hat{\mathbf{H}}^{\text{s-}MB}_{[w,i],1}(l), \cdots, \hat{\mathbf{H}}^{\text{s-}MB}_{[w,i],N_T}(l)]$ with

$$\hat{\mathbf{H}}^{\text{s-}MB}_{[w,i],k}(l) \;=\; \hat{\mathbf{H}}^{LS}_{[w,i],k}(l) \cdot \mathbb{P}_{\Pi}(\mathbf{U}_k) \tag{3.54}$$

and $\mathbb{P}_{\Pi}(\mathbf{U}_k) = \mathbf{U}_k\mathbf{U}_k^{\dagger}$, where the subscripts $[w,i]$ describe the CIR length constraint w and the i-th turbo iteration, respectively.

However, the parameter $\mathbf{U}_k$ is not known in general. Nevertheless, $\mathbf{U}_k$ is independent of the burst timings under the subspace channel model assumption. Thereby, $\mathbb{P}_{\Pi}(\mathbf{U}_k)$ may be approximated by using the projection matrix $\hat{\bar{\mathbf{\Pi}}}_{[W,N_{\text{turbo}}],k}$ obtained as the result of the N_{turbo} turbo iteration in the previous frame.[3] The channel estimate matrix of the s-ℓ1 MB can concretely be written, as

$$\hat{\mathbf{H}}^{\text{s-}MB}_{[w,i],k}(l) \;=\; \hat{\bar{\mathbf{H}}}^{LS}_{[w,i],k}(l) \cdot \hat{\bar{\mathbf{\Pi}}}_{[W,N_{\text{turbo}}],k}\bar{\mathbf{Q}}^{-\text{H}}_{[W,N_{\text{turbo}}],kk} \tag{3.55}$$

for the k-th TX stream, where the $W \times W$ whitening matrix $\bar{\mathbf{Q}}_{[W,N_{\text{turbo}}],kj}$ is the (k,j)-th block matrix of the Cholesky decomposition for $\bar{\mathcal{R}}_{\mathbf{\Phi}\mathbf{\Phi}_{[W,N_{\text{turbo}}]}}$. The matrix $\hat{\bar{\mathbf{H}}}^{LS}_{[w,i],k}(l)$ is defined similarly by (2.27) with the ℓ1 LS channel estimate and $\bar{\mathbf{Q}}_{[W,N_{\text{turbo}}],kj}$.

As demonstrated later, (3.55) improves[4] channel estimation accuracy significantly over the ℓ2 MB channel estimate:

$$\hat{\mathbf{H}}^{MB}_{[W,i],k}(l) \;=\; \hat{\mathbf{H}}^{LS}_{[W,i],k}(l) \cdot \hat{\bar{\mathbf{\Pi}}}_{[W,i],k}\bar{\mathbf{Q}}^{-\text{H}}_{[W,i],kk} \tag{3.56}$$

in IBI scenarios for the first turbo iteration ($i = 1$). This is because the PCA to derive $\hat{\bar{\mathbf{\Pi}}}_{[w,i],k}$ cannot be performed accurately when the cancel terms in (3.50) are not available.

[3] Therefore, the s-ℓ1 MB is a *self-supervised* technique.

[4] The improvement is expected when $N_{\text{turbo}} > 1$. In other words, (3.55) with $N_{\text{turbo}} = 1$ does not improve channel estimation performance since it becomes identical to (3.56).

3.2.3.3 Solution to (3.52)

Since $\hat{\mathbf{H}}^{MB}_{[w],k}(l) = \hat{\mathbf{G}}^{MB}_{[w],k}(l)\mathtt{P}^{\mathrm{T}}_{[w]}$ with $\mathtt{P}_{[w]} = \mathbf{I}_W|_{1:w}$, the solution can be obtained via (3.57):

$$\begin{aligned}\mathbb{E}\left[\left\|\hat{\mathbf{G}}^{MB}_{[w],k}\mathtt{P}^{\mathrm{T}}_{[w]} - \mathbf{H}_k\right\|^2\right] &= \mathbb{E}\left[\left\|\left\{\hat{\mathbf{G}}^{LS}_{[w],k}\mathtt{P}^{\mathrm{T}}_{[w]} - \mathbf{H}_k\right\}\mathbb{P}_{\Pi}(\mathbf{U}_k)\right\|^2\right] \\ &= \frac{r_k}{w}\mathbb{E}\left[\left\|\hat{\mathbf{G}}^{LS}_{[w],k} - \mathbf{G}_k\right\|^2\right] + \mathbb{E}\left[\left\|\mathbf{H}^{\perp}_{[w],k}\right\|^2\right], \quad (3.57)\end{aligned}$$

where $\mathbf{G}_k(l) = \mathbf{H}_k(l)\mathtt{P}_{[w]}$, $\mathbf{H}^{\perp}_{[w],k}(l) = \mathbf{H}_k(l) - \mathbf{G}_k(l)\mathtt{P}^{\mathrm{T}}_{[w]}$ and the burst timing l is omitted for the sake of simplicity. Notice that the second term of (3.57) may be approximated as $\mathbb{E}[\|\mathbf{H}^{\perp}_{[w],k}\|^2] \approx \mathbb{E}[\|\hat{\mathbf{H}}^{MB}_{[W,N_{\mathrm{turbo}}],k}(\mathbf{I}_W - \mathtt{P}_{[w]}\mathtt{P}^{\mathrm{T}}_{[w]})\|^2]$ by using the $\ell 2$ MB channel estimate $\hat{\mathbf{H}}^{MB}_{[W,N_{\mathrm{turbo}}],k}$ obtained after N_{turbo} turbo iterations.

We investigate, therefore, the MSE of the $\ell 1$ LS in IBI channels to clarify the first term of (3.57). Moreover, the MSE analysis is focused on the channel estimation with the TS only, because the s-$\ell 1$ MB aims to improve channel estimation accuracy for the first turbo iteration.

Since the received signal (3.46) suffers from IBIs, the MSE of the $\ell 1$ LS for N_T TX streams becomes

$$\mathbb{E}\left[\|\hat{\mathcal{H}}^{LS}_{[w]} - \mathcal{H}\|^2\right] = \mathbb{E}\left[\|\tilde{\mathcal{Z}}\mathcal{X}^{\mathrm{H}}_c\mathbf{P}_{[w]}\mathbf{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_t[w]}\|^2\right] + \mathbb{E}\left[\|\mathcal{H}^{\perp}_{[w]}\|^2\right] \quad (3.58)$$

where $\tilde{\mathcal{Z}} = \mathcal{Z} + \mathcal{H}_p\mathcal{X}_p + \mathcal{H}_f\mathcal{X}_f$, $\mathbf{P}_{[w]} = \mathbf{I}_{N_T} \otimes \mathtt{P}_{[w]}$, $\mathbf{R}_{\mathbf{\Phi}\mathbf{\Phi}_t[w]} = \mathbf{P}^{\mathrm{T}}_{[w]}\mathbf{R}_{\mathcal{X}\mathcal{X}_t}\mathbf{P}_{[w]}$ and $\mathcal{H}^{\perp}_{[w]} = \mathcal{H}_{[w]}(\mathbf{I}_{WN_T} - \mathbf{P}_{[w]}\mathbf{P}^{\mathrm{T}}_{[w]})$. The first term on the RHS can be reduced to the following: by $\mathbb{E}[\mathbf{R}_{\mathbf{\Phi}\mathbf{\Phi}_t[w]}]/\bar{N}_t = \mathbf{I}_{wN_T}$,

$$\begin{aligned}&\mathbb{E}\left[\|\tilde{\mathcal{Z}}\mathcal{X}^{\mathrm{H}}_c\mathbf{P}_{[w]}\mathbf{R}^{-1}_{\mathbf{\Phi}\mathbf{\Phi}_t[w]}\|^2\right] \\ &= \frac{1}{\bar{N}_t^2}\mathbb{E}\left[\|\mathcal{Z}\mathcal{X}^{\mathrm{H}}_c\mathbf{P}_{[w]}\|^2 + \|(\mathcal{H}_p\mathcal{X}_p + \mathcal{H}_f\mathcal{X}_f)\mathcal{X}^{\mathrm{H}}_c\mathbf{P}_{[w]}\|^2\right] \quad (3.59) \\ &= \mathbf{1}^{\mathrm{T}}_{wN_T}\left[\mathbf{1}_{N_T} \otimes \left\{\frac{N_R}{\bar{N}_t}\sigma_z^2\mathbf{I}_w + \mathbf{P}^{\mathrm{T}}_{[w]}\mathfrak{e}_{\mathrm{IBI}}\right\}\right], \quad (3.60)\end{aligned}$$

where we define

$$\mathfrak{e}_{\mathrm{IBI}} = \frac{\sigma_x^4}{\bar{N}_t^2}\mathbf{T}_c\left(\mathbf{T}_p(N_{G2}) + \mathbf{T}_f(N_{G1})\right)^{\mathrm{T}}\sum_{k=1}^{N_T}\bar{\mathbf{d}}_{\mathcal{H},k} \quad (3.61)$$

with the delay profile vector for the k-th TX stream

$$\bar{\mathbf{d}}_{\mathcal{H},k} \;=\; \mathtt{diag}\{\mathbb{K}[\mathbf{H}_k]\}. \tag{3.62}$$

The $W \times \tilde{N}_t$ matrix $\mathbf{T}_c$ denotes a Toeplitz matrix, the first row vector of which is $[\mathbf{1}_{N_t}^{\mathrm{T}},\ \mathbf{0}_{W-1}^{\mathrm{T}}] \in \mathbb{R}^{1\times\tilde{N}_t}$. We define, moreover, that $\mathbf{T}_p(N_{G2}) = \left[\mathbf{T}_\Delta|_{(1+N_{G2}):W},\ \mathbf{O}_{W\times(\tilde{N}_t-(W-N_{G2}))}\right]$ and $\mathbf{T}_f(N_{G1}) = \left[\mathbf{O}_{W\times(\tilde{N}_t-(W-N_{G1}))},\ \mathbf{T}_\Delta^{\mathrm{T}}|_{1:(W-N_{G1})}\right]$ with

$$\mathbf{T}_\Delta \;=\; \begin{bmatrix} 0 & & & 0 \\ 1 & \ddots & & \\ \vdots & \ddots & \ddots & \\ 1 & \cdots & 1 & 0 \end{bmatrix} \;\in\; \mathbb{R}^{W\times W}.$$

The derivation of (3.60) is detailed in Appendix 3.B. Therefore, the MSE for the k-th TX stream is reduced to

$$\begin{aligned} \mathbb{E}\left[\|\hat{\mathbf{H}}_{[w],k}^{LS} - \mathbf{H}_k\|^2\right] &= \mathbb{E}\left[\left\|\hat{\mathbf{G}}_{[w],k}^{LS} - \mathbf{G}_k\right\|^2\right] + \mathbb{E}\left[\left\|\mathbf{H}_{[w],k}^{\perp}\right\|^2\right] \\ &= N_R\frac{w}{\bar{N}_t}\sigma_z^2 + \mathbf{1}_w^{\mathrm{T}}(\mathfrak{e}_{\mathrm{IBI}}|_{1:w}) + \mathbb{E}\left[\left\|\mathbf{H}_{[w],k}^{\perp}\right\|^2\right]. \end{aligned} \tag{3.63}$$

By the definition of the matrices $\mathbf{T}_c$, $\mathbf{T}_p$ and $\mathbf{T}_f$, we notice that

$$\mathfrak{e}_{\mathrm{IBI}} \;=\; \mathfrak{e}_{\mathrm{IBI},p} + \mathfrak{e}_{\mathrm{IBI},f}, \tag{3.64}$$

where the i-th entries of $\mathfrak{e}_{\mathrm{IBI},p}$ and $\mathfrak{e}_{\mathrm{IBI},f}$ are, respectively,

$$\begin{aligned} \mathfrak{e}_{\mathrm{IBI},p}|_i &= \frac{\sigma_x^4}{\bar{N}_t^2}\sum_{j=i+N_{G2}+1}^{W}(j-i-N_{G2})\sum_{k=1}^{N_T}\bar{d}_{\mathcal{H},k,j}, \qquad (1\le i\le W) \\ \mathfrak{e}_{\mathrm{IBI},f}|_i &= \begin{cases} 0, & (i\le N_{G1}) \\ \dfrac{\sigma_x^4}{\bar{N}_t^2}\displaystyle\sum_{j=1}^{i-N_{G1}-1}(i-j-N_{G1})\sum_{k=1}^{N_T}\bar{d}_{\mathcal{H},k,j}, & (1+N_{G1}\le i\le W) \end{cases} \end{aligned}$$

with $\bar{d}_{\mathcal{H},k,j}$ denoting the j-th entry of the delay profile vector $\bar{\mathbf{d}}_{\mathcal{H},k}$.

Consequently, problem (3.53) is reduced to

$$\hat{w}_k = \underset{1\le w\le W}{\arg\min}\left\{\hat{r}_k\left(\frac{N_R}{\bar{N}_t}\sigma_z^2 + \frac{\mathbf{1}_w^{\mathrm{T}}(\mathfrak{e}_{\mathrm{IBI}}|_{1:w})}{w}\right) + \mathbf{1}_{W-w}^{\mathrm{T}}(\bar{\mathbf{d}}_{\mathcal{H},k}|_{(w+1):W})\right\}, \tag{3.65}$$

which can be solved easily with the delay profile vector $\bar{\mathbf{d}}_{\mathcal{H},k}$ given by (3.62). The rank $\hat{r}_k$ of the CIR $\mathbf{H}_k$ can be approximated with that for the channel estimate $\hat{\mathbf{H}}^{MB}_{[W,N_{\rm turbo}]}(l_p)$ obtained by the $\ell 2$ MB technique at the last $N_{\rm turbo}$-th iteration, where $l_p = \lfloor (l-1)/N_B \rfloor N_B$ denotes the last N_B-th burst timing in the previous frame. The operation $\lfloor \cdot \rfloor$ is the floor function.

3.2.3.4 The s-$\ell 1$ MB algorithm

Algorithm 5 summarizes the above-mentioned s-$\ell 1$ MB technique, where the self-supervised subspace projection is applied for the first turbo iteration only. It should be noted that, as demonstrated in Section 3.3.2.3, Algorithm 5 can achieve the asymptotic MSE performance in two iterations, as long as the subspace projection obtained in the previous frame is accurate enough. The s-$\ell 1$ MB technique may, however, be applied after the second iteration as well, when the soft replica signals is not accurate enough. Convergence performance improvement including such the approach is a future work of this thesis.

Algorithm 5 The CHATES with the s-$\ell 1$ MB techniques.

1: The counter i denotes the i-th turbo iteration.
2: **if** $i = 1$ **then**
3: Estimate the CIR length $\hat{w}$ by (3.53).
4: Obtain $\hat{\mathcal{H}}^{LS}_{[\hat{w}]}$ by performing the $\ell 1$ LS (2.5) with the estimated CIR length $\hat{w}$.
5: Obtain $\hat{\mathcal{H}}^{\text{s-}MB}_{[\hat{w}]}$ (3.55) with $\hat{\mathcal{H}}^{LS}_{[\hat{w}]}$, where the subspace projection and whitening matrices are that obtained in the previous frame.
6: **else**
7: Perform the $\ell 2$ MB (2.22) with $\lambda(j) = 0$ for the IBI canceled received signal (3.50).
8: **end if**

3.2.4 Computational complexity order

The computational complexity orders $\mathcal{O}(\cdot)$ required for the CHATES techniques are summarized in Table 3.3. The CHATES techniques combined with the $\ell 2$ MB and/or s-$\ell 1$ MB algorithms require the same complexity order as that of the ordinary $\ell 2$ MB channel estimation assumed the length-W symbol

Table 3.3: Computational complexity orders for the CHATES algorithms

Algorithm	Computational complexity order
$\ell 2$ MB (GI assumed)	$\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3)$
CHATES with $\ell 2$ MB	$\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3)$
CHATES with s-$\ell 1$ MB	$\mathcal{O}(W^2 N_T^2 \tilde{N}_{td} + W^3 N_T^3 N_R^3)$

GI. This is because the complexity order required for the IBI canceling (3.50) is $\mathcal{O}(W^2 N_T N_R)$ by noticing the sparse structure of the matrices $\hat{\mathcal{X}}_p$ and $\hat{\mathcal{X}}_f$. Moreover, the complexity order needed for the CIR length estimation (3.52) is very minor, $\mathcal{O}(W(W - N_{G,\min})N_T)$ with $N_{G,\min} = \min(N_{G1}, N_{G2})$, since the complexity is dominated by that of (3.64).

3.3 Numerical Examples

First of all, performance of the proposed CHATUE2 and CHATES techniques are shown in Section 3.3.1 and 3.3.2, respectively. Performance of the system combining these new methods is, then, presented in 3.3.3.

Note that this section assumes the same system model as that in Chapter 2 except that the soft replica symbols $\hat{\mathbf{x}}_d, k$ for channel estimation algorithms is generated from the *a posteriori* LLR λ^p_{DEC} fed back from the decoder. This is because, as detailed in Section 3.3.2.3, the CHATES method with the *a posteriori* LLR improves MSE convergence performance over that with the extrinsic LLR.

3.3.1 Channel equalization performance

Performance of CHATUE2 is discussed by comparing with that of the conventional TEQ-CP and/or CHATUE1 techniques. This subsection assumes a SISO system with known CIRs in order to focus on the verification for the noise enhancement analysis shown in Section 3.1.3.

The parameters used in this subsection are detailed in Table 3.4. Burst format 11 is used for both CHATUE1 and CHATUE2, whereas Burst format 12 is used for TEQ-CP. In the CHATUE algorithms, a data frame encoded by a convolutional code $(g_1, g_2) = (7, 5)_8$ with code rate $R_c = 1/2$ was divided into $N_B = 10$ bursts. The information bits in TEQ-CP, the length of which is

Table 3.4: Burst Formats for a SISO system.

Format No.	N_t	N_{G1}	N_{CP}	N_d	N_{G2}	R_c	η
11	63	0	0	256	0	1/2	0.4
12	63	0	64	192	0	2/3	0.4

the same as the one in the CHATUE algorithms, is encoded with a code with rate $R_c = 2/3$ derived from a half-rate mother convolutional code $(g_1, g_2) = (7, 5)_8$ by using a puncturing matrix

$$\mathbf{P}_{\mathbf{x},n/(n-1)} = \begin{bmatrix} 1 & \cdots & & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \in \mathbb{R}^{2\times(n/2)} \tag{3.66}$$

with $n = 4$ for puncturing rate 4/3. It should be noted the spectral efficiency is $\eta = 0.4$ in both Burst format 11 for the CHATUE algorithms and Burst format 12 for the TEQ-CP. Thereby, the following comparisons are fair.

3.3.1.1 EXIT analysis

Convergence property of CHATUE2 is shown by using extrinsic information transfer (EXIT) charts. Burst format 11 described in Table 3.4 is used for both the CHATUE1 and CHATUE2 algorithms, whereas Burst format 12 is used for TEQ-CP.

Fig. 3.3 shows EXIT curves of the CHATUE1 and CHATUE2 algorithms as well as TEQ-CP. The equalizer's EXIT curves were obtained, in all the system setups tested, for a 64-path frequency selective Rayleigh fading channel realization with average SNR = 2.4 dB. Ideal channel estimation is assumed. The MI $\mathcal{I}^e_{\mathrm{EQU}}$ between the LLR λ^e_{EQU} (3.22) and the coded bits c at the transmitter is defined by (2.49).

It is found from Fig. 3.3 that the equalizer's EXIT curve of CHATUE1 is located below the TEQ-CP's EXIT curve over entire value range of *a priori* MI $\mathcal{I}^a_{\mathrm{EQU}}$. This is because of the noise enhancement described in Section 3.1.3. In contrast, CHATUE2 improves $\mathcal{I}^e_{\mathrm{EQU}}$ and achieves almost the same point as that with TEQ-CP when $\mathcal{I}^a_{\mathrm{EQU}} = 1$, although its left most point at $\mathcal{I}^a_{\mathrm{EQU}} = 0$ is almost the same as that of CHATUE1. This observation verifies the asymptotic perfect elimination of the noise enhancement with the CHATUE2 algorithm.

A trajectory of turbo equalization with CHATUE2 is also presented in Fig. 3.3. The trajectory reaches a point very close to $\mathcal{I}^e_{DEC} = 1$ without intersection in the channel realization used and hence the MI between the *a posteriori* LLR of decoder λ^p_{DEC} and the binary source information approaches 1. This is because of two reasons: 1) CHATUE2 improves the equalizer's EXIT curve by eliminating the noise enhancement; 2) the CHATUE algorithms allows us to use a lower rate code by utilizing the time duration allocated for the CP. On the other hand, the EXIT curves of the CHATUE1 and TEQ-CP algorithms have the intersection at (0.98, 0.8) and (0.92, 0.85), respectively. Thereby the trajectories of the CHATUE1 and TEQ-CP techniques can rarely approach points very close to $\mathcal{I}^e_{DEC} = 1$ for an SNR of 2.4 dB, although they are not presented in Fig. 3.3 to avoid too dense a representation. This is because CHATUE1 incurs the noise enhancement at the equalizer output or TEQ-CP can not use a lower rate code with the same spectral efficiency due to CP-transmission.

3.3.1.2 BER performance with known CIR

The average SNR used in BER simulations is defined in association with the average $\mathrm{E_b/N_0}$ (2.50). Burst format 11 described in Table 3.4 was used for both the CHATUE1 and 2 algorithms, whereas Burst format 12 was used for TEQ-CP.

Verification for the noise enhancement analysis: In Fig. 3.4, the BER performance of turbo equalization for a single path static AWGN channel are presented to verify the noise enhancement analysis described in Section 3.1.3, even though equalization is not needed in single path channels. For reference, the BER performance of BCJR decoders with the parameters mentioned above are also presented.

The BER with TEQ-CP is the same as that with a) BCJR decoder ($R_c = 2/3$) used in TEQ-CP, as depicted in Fig. 3.4. However, the BER with CHATUE1 is degraded compared to c) BCJR decoder ($R_c = 1/2$) due to the noise enhancement detailed in Section 3.1.3. The BER with CHATUE1 is identical to that with b) BCJR decoder ($R_c = 1/2$) assumed the noise enhancement to its input before interleaving.[5] The noise enhancement local-

[5]The noise power of input signal to the BCJR decoder b) is intentionally enhanced to reproduce the noise enhancement problem incurred by CHATUE1. The noise power of the input signal to the BCJR decoder b) is increased to $2\sigma_z^2$ for the first L bits. The BCJR

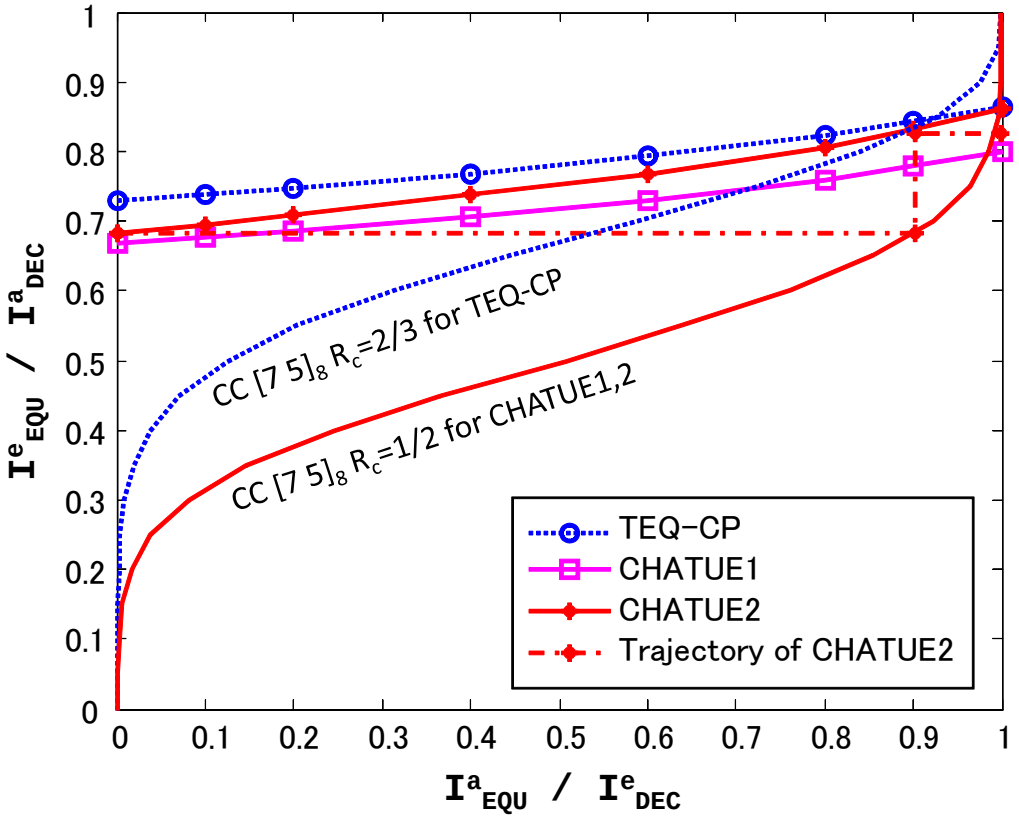

Fig. 3.3: EXIT charts and trajectory of iterative processing over a 64-path Rayleigh fading at average SNR = 2.4 dB.

Table 3.5: Burst Formats for a 4×4 MIMO system.

Format No.	N_t	N_{G1}	N_{CP}	N_d	N_{G2}	R_c	η
41	127	0	0	512	0	1/2	1.6
42	127	31	32	512	31	1/2	1.4
43	127	0	32	480	0	8/15	1.6

ized in the L symbols is not uniformly distributed over a frame even after interleaving and hence it degrades the performance of a BCJR decoder more than expected (0.97 dB), as shown in (3.29).

The BER with CHATUE2, on the other hand, achieves exactly the same as that with c) BCJR decoder ($R_c = 1/2$), in the same way as for TEQ-CP. It should be noted that the proposed CHATUE2 algorithm can fully exploit the time duration made available by eliminating the CP, which allows for the use of a lower rate code ($R_c = 1/2$) when the channel estimate is accurate enough.

BER performance in a fading channel: Fig. 3.4 also shows the BER performance of turbo equalization for the PB3 channel realizations. The path positions are at {0, 3, 12, 18, 34.5, 55.5} symbol timings assuming that a transmission bandwidth of 15 MHz. The CHATUE2 algorithm improves the BER over CHATUE1 by 0.5 dB at BER = 10^{-5} since the proposed technique with the composite replica improves the SNR at the equalizer output. Moreover, CHATUE2 achieves a gain of 1.5 dB or more over TEQ-CP at BER = 10^{-5} because the CHATUE algorithms allows of the use of lower rate codes when the spectral efficiency η is fixed.

3.3.2 Channel estimation performance

This subsection verifies the channel estimation performance by using the proposed joint IBI cancelation and channel estimation technique, CHATES. The same 4×4 MIMO system and its setups as that in Section 2.4 are used. However, two Burst formats 41 and 42 described in Table 3.4 are assumed because of the reasons as follows: Burst format 41, which can cause IBIs in

decoder b) decodes the noise enhanced input signal following interleaving. The BCJR decoder b) itself is the same as BCJR decoder c).

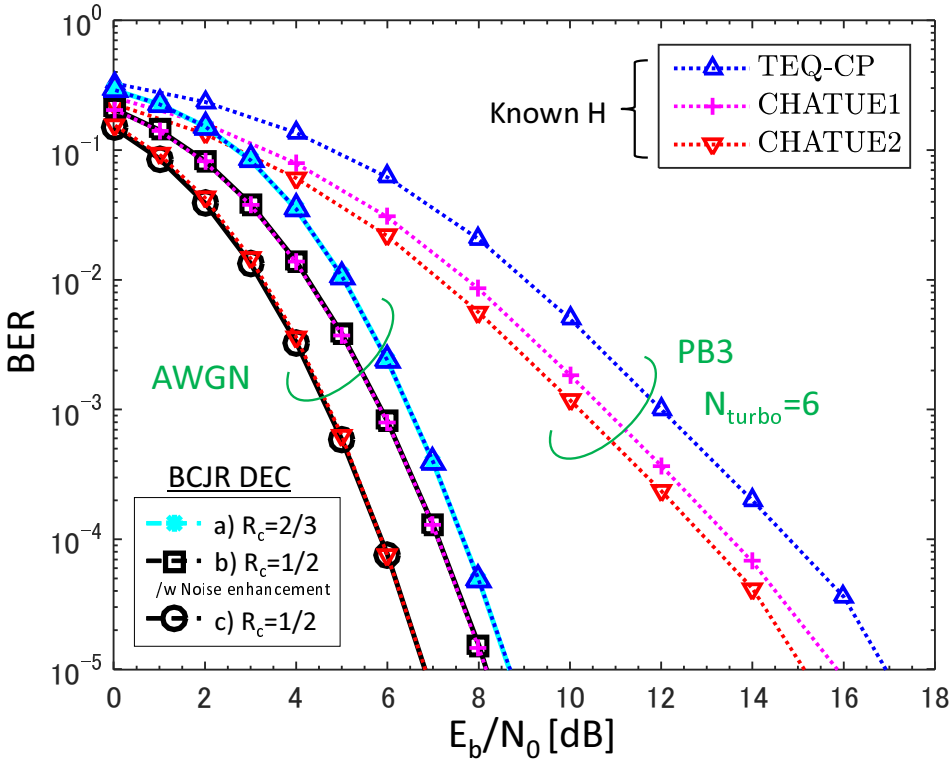

Fig. 3.4: BER performance in the 1-path static AWGN and in the PB3 scenario. The spectral efficiency is fixed to $\eta = 0.4$. The CHATUE1 and CHATUE2 algorithms use Burst format 11 with the coding rate $R_c = 1/2$, whereas TEQ-CP uses Burst format 12 with $R_c = 2/3$. The known CIR `H` and six turbo iterations are assumed.

the TS section, is used to verify CHATES with the s-ℓ1 MB algorithm; Burst format 42 has GIs on both the sides in time of TS such that the TS section does not suffer from the IBIs due to the neighboring data sections. The reason for presenting the simulation result with Burst format 42 is to provide a basis for the performance comparison of the IBI cancelation, although its spectral efficiency is less than that of Burst format 41.

3.3.2.1 NMSE performance of CHATES with ℓ2 MB

Fig. 3.5 shows the NMSE of the channel estimate with CHATES with the ℓ2 MB technique in the VA30 scenario. Similar to the ℓ2 MB channel estimation assuming Burst format 42, CHATES using Burst format 41 achieves the performance bound NaCRB (2.45) asymptotically after the sixth turbo iteration. In the case Burst format 41 without the IBI cancelation technique is assumed, the estimation accuracy diverges from the NaCRB due to the IBI even after performing enough iterations.

3.3.2.2 NMSE performance of CHATES with s-ℓ1 MB

In the first iteration, however, the CHATES with the ℓ2 MB technique also suffers from MSE deterioration due the IBI. As observed from Fig. 3.5, the s-ℓ1 MB algorithm can improve the MSE deterioration significantly. This is because the s-ℓ1 MB algorithm can avoid the IBI by utilizing the ℓ1 regularization onto the CIR length. Of course, the s-ℓ1 MB algorithm cannot always improve the MSE deterioration when the significant paths are distributed over a long CIR. As shown in Fig. 3.6, however, the MSE gain can be observed in the PB3 scenario, too. This is because, as shown in Fig. 2.1, the significant paths of the PB channel model do not exist in the entire W symbol duration when a low to moderate SNR regime is assumed.

It should be noted that CHATES with both the ℓ2 MB and s-ℓ1 MB algorithms can achieve the performance bound NaCRB after performing enough turbo iterations. Improvement of NMSE convergence performance with the s-ℓ1 MB algorithm is verified in the next subsection.

3.3.2.3 NMSE convergence performance

Fig. 3.7 shows NMSE convergence performance of the CHATES algorithms in the VA30 scenario at SNR = 15 dB. Burst format 41 is assumed. As mentioned in the beginning of Section 3.3, we assume that the soft replica

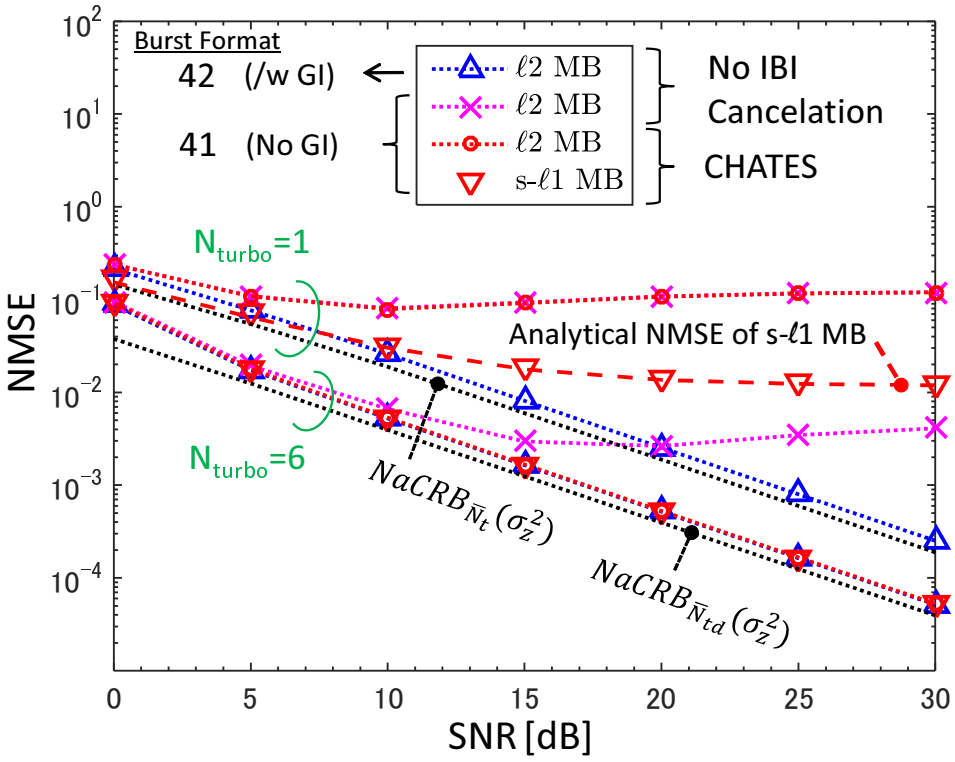

Fig. 3.5: NMSE performance in the VA30 scenario. The analytical NMSE of the s-ℓ1 MB algorithm for the 1st iteration is the normalized version of (3.57), where the normalization factor is $\mathbb{E}[\|\mathcal{H}\|^2]$, similar to the performance bound NaCRB (2.45).

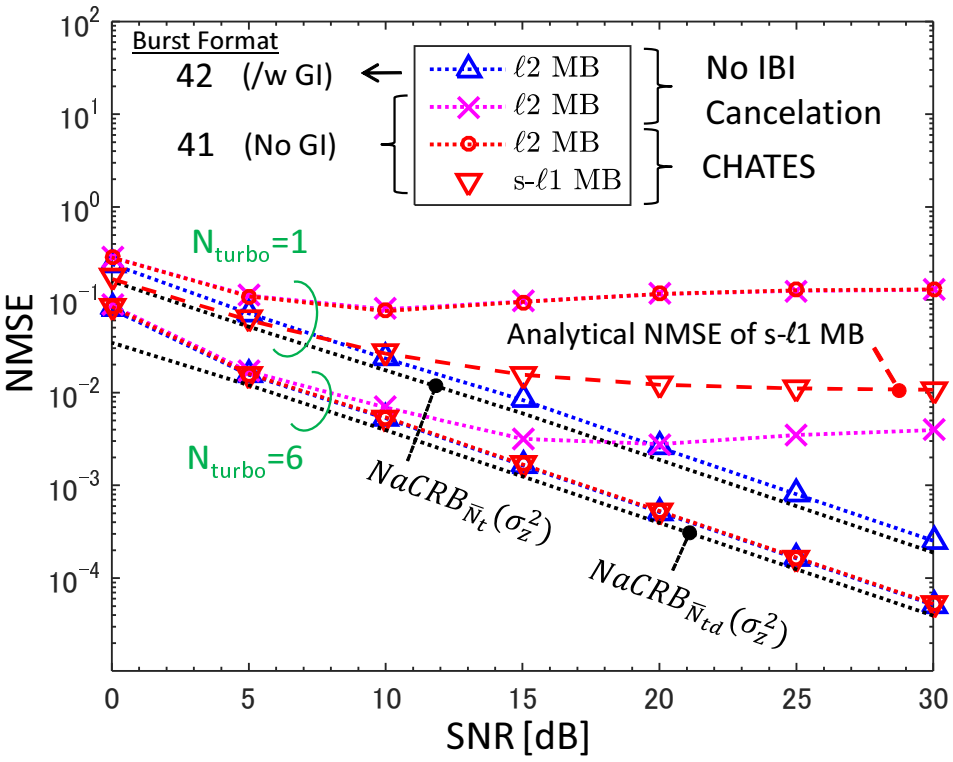

Fig. 3.6: NMSE performance in the PB3 scenario.

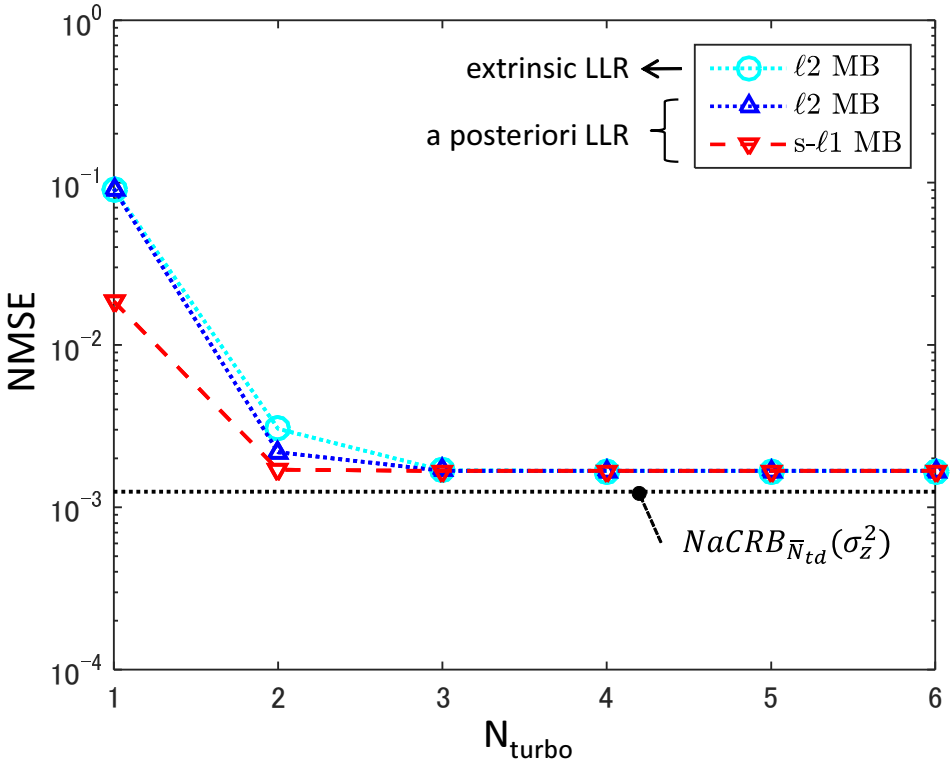

Fig. 3.7: NMSE convergence performance of the CHATES algorithms over turbo iterations. SNR is set at 15 dB. The VA30 scenario and Burst format 41 are assumed.

for the channel estimation algorithm is generated from the *a posteriori* LLR $\lambda^{p}_{\mathrm{DEC}}$. We can observe from Fig. 3.7 that CHATES with the $\ell 2$ MB technique assuming the *a posteriori* LLR improves NMSE convergence performance over that with the extrinsic LLR.

It should be noticed that CHATES with the s-$\ell 1$ MB algorithm improves NMSE convergence performance further. Specifically, CHATES with the s-$\ell 1$ MB algorithm asymptotically achieves the performance bound NaCRB in two iterations whereas the one with the $\ell 2$ MB technique requires three iterations in this simulation setup.

3.3.3 System performance

This subsection presents BER performance of the CHATUE2 receiver with the CHATES algorithm in the 4×4 MIMO system. The proposed techniques can be verified from the following two viewpoints: 1) BER performance without a frame length constraint; 2) BER performance with a fixed frame length. The first verification aims to investigate whether the receiver with the burst format 41 removed the CP and GIs can improve BER performance over that with Burst format 42 when channel estimation is actually performed. The second item verifies effectiveness of the CHATUE algorithm over the conventional TEQ-CP receiver by using Burst formats 41 and 43, respectively. This is because the frame length should be a constant in a practical communication system.

3.3.3.1 BER performance without a frame length constraint

Since the definition of $\mathrm{E_b/N_0}$ (2.50) takes account of the frame length, we can compare Burst format 41 with the format 42 although their frame lengths are different. It should be noticed that, moreover, expected BER gains between burst formats having the same coding rate R_c can be calculated from the spectral efficiencies. Specifically, the expected BER gain with Burst format 41 over the baseline format 42 becomes at most

$$10 \log_{10}(\eta_{42}/\eta_{41}) \;=\; 0.6 \text{ dB},$$

in $\mathrm{E_b/N_0}$, where η_N denotes the spectral efficiency of the burst format N.

Fig. 3.8 shows a BER performance comparison between Burst formats 41 and 42. The VA30 scenario is assumed. As observed from Fig. 3.8, CHATUE2 with the $\ell 2$ MB technique achieves a BER gain of 0.5 dB over TEQ-CP after performing six turbo iterations. Therefore, the receiver with Burst format 41 can improve BER performance over that with Burst format 42 even when channel estimation is performed.

However, in the first iteration, CHATUE2 with the $\ell 2$ MB technique suffers from an error floor. It should be noticed that, as depicted in Fig. 3.8, the proposed s-$\ell 1$ MB algorithm can solve the problem completely. The CHATUE2 receiver with the s-$\ell 1$ MB algorithm can, hence, improve BER convergence performance over that with the $\ell 2$ MB channel estimation technique. Specifically, as observed from Fig. 3.9, the receiver with the s-$\ell 1$ MB algorithm achieves BER $= 10^{-5}$ in four iterations, whereas the one with the $\ell 2$ MB technique requires six iterations at $\mathrm{E_b/N_0} = 11.5$ dB.

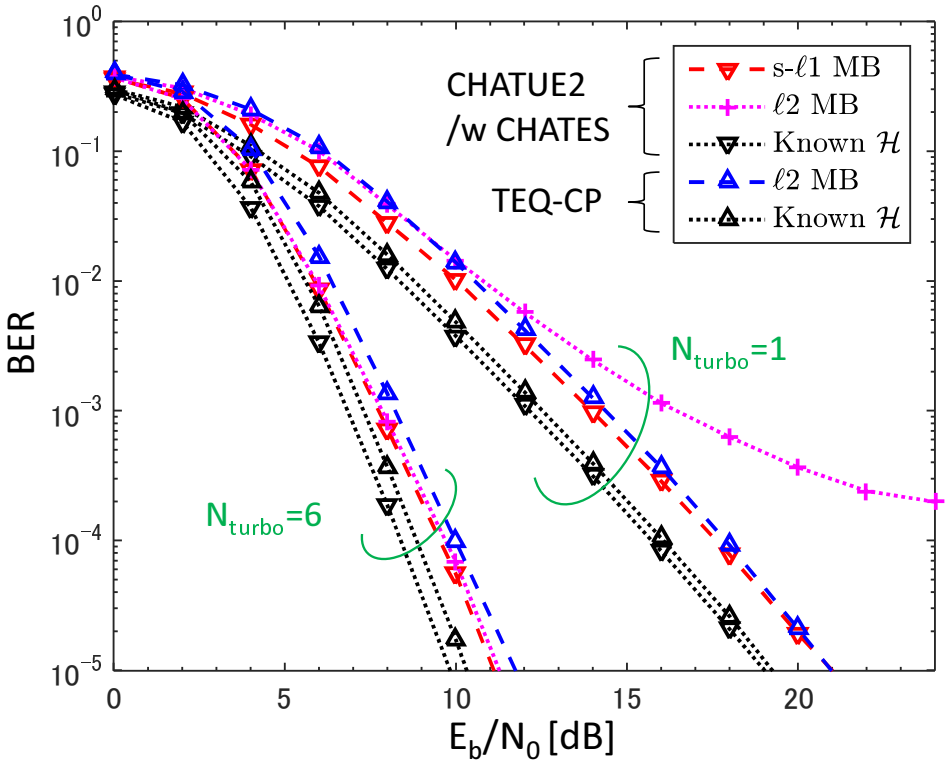

Fig. 3.8: BER performance in the VA30 scenario without constraint on the frame length. Burst formats 41 and 42 are used for the CHATUE and TEQ-CP techniques, respectively.

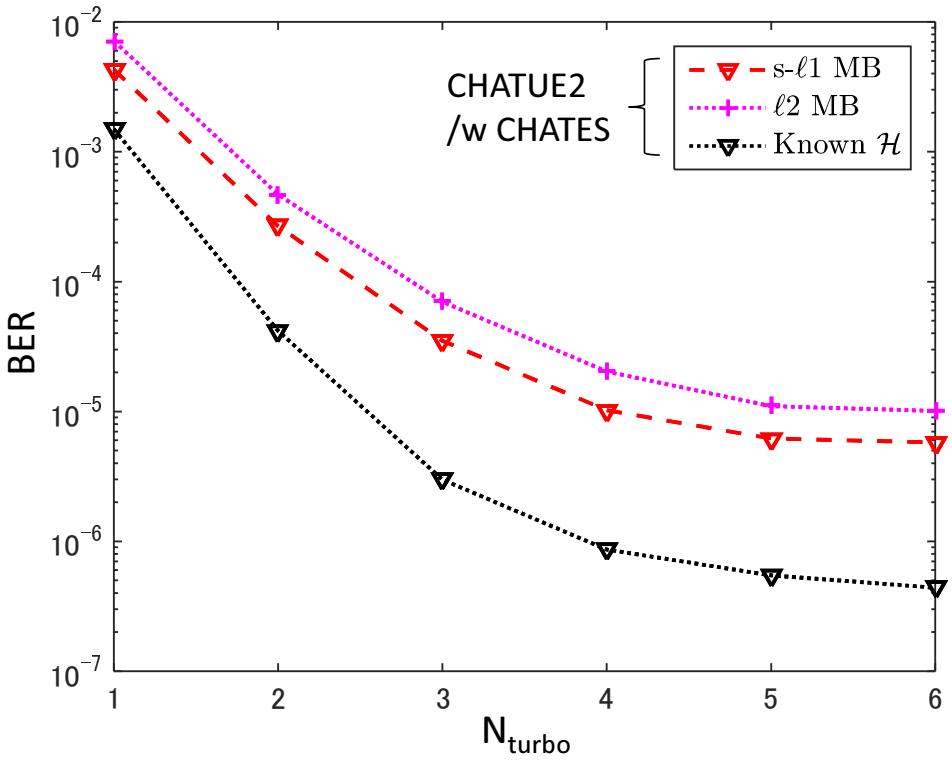

Fig. 3.9: BER convergence performance over turbo iterations at E_b/N_0 = 11.5 dB. The CHATUE2 receiver with Burst format 41 is used in the VA30 scenario.

3.3.3.2 BER performance with a fixed frame length

Fig. 3.10 shows BER performances with the CHATUE2 and TEQ-CP receivers in the PB3 scenario. Burst formats[6] 41 and 43 are used for the CHATUE2 and TEQ-CP techniques, respectively. As depicted in Fig. 3.10, the CHATUE2 algorithm improves BER performance over TEQ-CP by 1 dB after performing six iterations. Relevance of the BER gain can be supported by observing BER performance of the BCJR decoders in the single path static AWGN SISO channel shown in Fig. 3.10.

Similar to Fig. 3.8, we can observe from Fig. 3.10 that CHATUE2 with the $\ell 2$ MB estimation exhibits an error floor in the first iteration. However, the proposed s-$\ell 1$ MB algorithm can improve the problem in the PB3 channel scenario, too, although its CIR length is longer than that of the VA30 scenario.

3.4 Summary

The primary objective of this chapter has been to provide solutions to the problems inherent in chained turbo equalization techniques, which are: 1) the latency due to the time-concatenation of equalization, and 2) noise enhancement at the equalizer output. This chapter has showed that Problem 1) can easily be solved with a practical and reasonable assumption that the training sequence is transmitted in every burst. To cope with Problem 2), this chapter has proposed chained turbo equalization version 2, CHATUE2. By utilizing the composite replica to retrieve the circulant structure of the channel matrix in the received signal, the CHATUE2 algorithm improves the equalizer output SNR to the same level as that with TEQ-CP.

Furthermore, this chapter has proposed a new IBI cancelation technique for channel estimation, CHATES, to improve spectral efficiency. The conventional channel estimation technique can suffer from an MSE floor problem if the signals are transmitted with a burst format having no GIs. The CHATES algorithm can, however, achieve the CRB asymptotically even in IBI scenarios after performing enough turbo iterations. CHATES can be applied to the CHATUE1, CHATUE2 and TEQ-CP receivers, although CHATES inherits the CHATUE concept in the sense that the cancelation of IBI occurring in the TS section utilizes the LLR of transmitted data, fed back from the decoder,

[6]Burst format 43 uses the puncturing matrix (3.66) with $n = 16$.

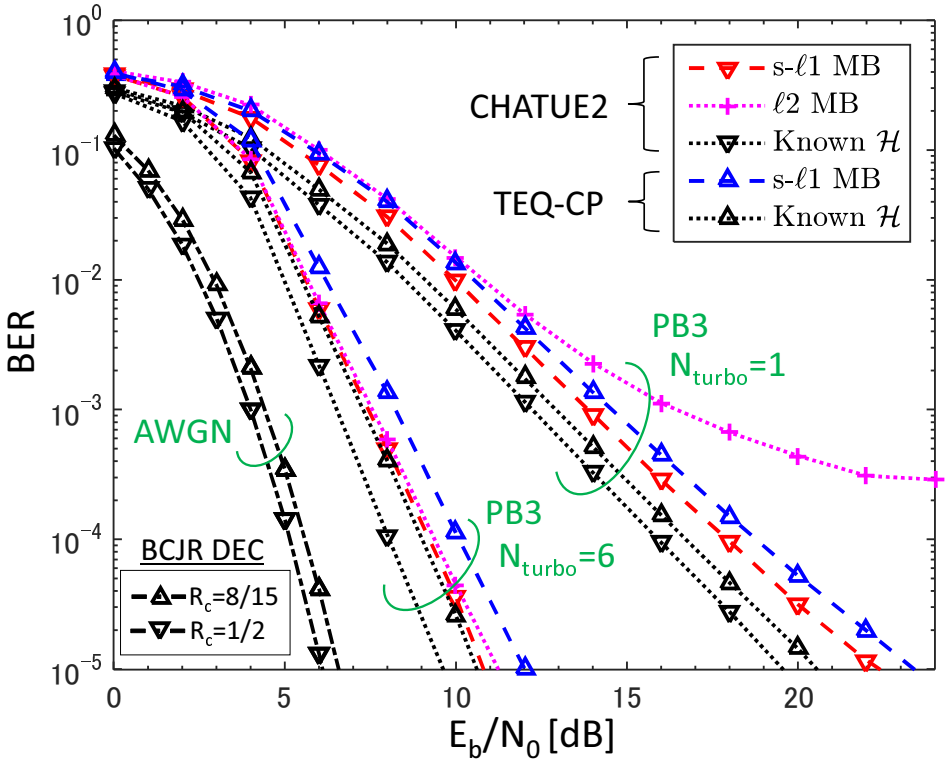

Fig. 3.10: BER performance in the PB3 scenario with a frame length constraint. For reference, BER performance with the BCJR decoder in the 1-path static AWGN SISO channel is also shown. The CHATUE2 receiver and the BCJR decoder with $R_c = 1/2$ use Burst format 41, whereas the TEQ-CP receiver and the BCJR decoder with $R_c = 8/15$ assume Burst format 43.

not only in the current but also in the past bursts.

A receiver with the CHATES algorithm can, nevertheless, suffer from a bit error floor in the first turbo iteration. This is because the feedback information from the decoder needed to perform the IBI cancelation is not available in the first turbo iteration. As a solution to the problem, this chapter has proposed a novel s-ℓ1 MB channel estimation which can be described as a conditional ℓ1 MMSE formulation, under the subspace channel model assumption. The s-ℓ1 MB algorithm can improve the IBI problem by simultaneously exploiting the CIR length constraint and the subspace projection obtained at the previous frame. The receiver with the s-ℓ1 MB algorithm can, therefore, completely solve the bit error floor problem. Simulation results have shown that, moreover, it can improve BER convergence performance over turbo iterations.

Appendix

3.A Derivation of the Asymptotic Mean (3.23)

Assuming that $\mathrm{E}[||\hat{\mathtt{s}}_{d,k}(l)||^2] \to 1$ after enough iterations, the noise covariance matrix (3.14) converges to

$$\mathbf{\Omega}_k(l) \;\; \to \;\; \sigma_z^2 \frac{N_d + L}{N_d} \mathbf{I}_{N_R N_d}. \tag{3.67}$$

Hence, the equation (3.18) converges to

$$\mathbf{\Gamma}_k(l) \;\; \to \;\; \frac{N_d}{(N_d + L)\sigma_z^2} \mathbf{I}_{N_d}, \tag{3.68}$$

under the assumption $\mathrm{E}[||\hat{\mathtt{H}}_k(l)||^2] = \mathrm{E}[||\mathtt{H}_k(l)||^2] = 1$. The asymptotic mean (3.23) is reduced by substituting (3.67), (3.68) and $\hat{\mathtt{S}}(l) \to \mathbf{I}_{N_d}$ into (3.20).

3.B Derivation of the MSE (3.60)

For the sake of conciseness, the case of $w = W$ is shown. The second term in (3.59) can be written as

$$\mathbb{E}\left[\|(\mathcal{H}_p \mathcal{X}_p + \mathcal{H}_f \mathcal{X}_f)\mathcal{X}_c^{\mathrm{H}}\|^2\right] \;\; = \;\; \mathbf{1}_{WN_T}^{\mathrm{T}} \mathtt{diag}\left\{\mathbb{E}[\mathcal{X}_c \mathbf{D} \mathcal{X}_c^{\mathrm{H}}]\right\}, \tag{3.69}$$

where $\mathbf{D} = \mathcal{X}_p^{\mathrm{H}} \mathbf{K}_{\mathcal{H}\mathcal{H}_p} \mathcal{X}_p + \mathcal{X}_f^{\mathrm{H}} \mathbf{K}_{\mathcal{H}\mathcal{H}_f} \mathcal{X}_f$ with $\mathbf{K}_{\mathcal{H}\mathcal{H}_p} = \mathcal{H}_p^{\mathrm{H}} \mathcal{H}_p$ and $\mathbf{K}_{\mathcal{H}\mathcal{H}_f} = \mathcal{H}_f^{\mathrm{H}} \mathcal{H}_f$. By noticing that $\mathbb{E}[\mathbf{D}]$ is a diagonal matrix, we have

$$\mathtt{diag}\left\{\mathbb{E}[\mathcal{X}_c \mathbf{D} \mathcal{X}_c^{\mathrm{H}}]\right\} \;\; = \;\; \mathbf{1}_{N_T} \otimes \left(\sigma_x^2 \mathbf{T}_c \mathtt{diag}\{\mathbb{E}[\mathbf{D}]\}\right), \tag{3.70}$$

since non-zero entries in $\mathcal{X}_c$ follow $\mathcal{CN}(0, \sigma_x^2)$. Moreover,

$$\mathtt{diag}\{\mathbb{E}[\mathbf{D}]\} \;\; = \;\; \sigma_x^2 (\mathbf{T}_p + \mathbf{T}_f)^{\mathrm{T}} \sum_{k=1}^{N_T} \bar{\mathbf{d}}_{\mathcal{H},k}, \tag{3.71}$$

since, according to the definition (3.47) of $\mathcal{X}_p$,

$$\begin{aligned} \mathtt{diag}\{\mathbb{E}[\mathcal{X}_p^{\mathrm{H}} \mathbf{K}_{\mathcal{H}\mathcal{H}_p} \mathcal{X}_p]\} \;\; &= \;\; \mathtt{diag}\left\{\mathbb{E}\left[\sum_{k=1}^{N_T} \mathbf{X}_{p,k}^{\mathrm{H}} \mathbf{K}_{\mathcal{H}\mathcal{H}_p,k} \mathbf{X}_{p,k}\right]\right\} \\ &= \;\; \sum_{k=1}^{N_T} \sigma_x^2 \mathbf{T}_p^{\mathrm{T}} \mathtt{diag}\left\{\mathbb{E}\left[\mathbf{K}_{\mathcal{H}\mathcal{H},k}\right]\right\}, \end{aligned}$$

where the $W \times W$ matrix $\mathbf{K}_{\mathcal{H}\mathcal{H}_p,k}$ denotes the (k,k)-th block matrix in $\mathbf{K}_{\mathcal{H}\mathcal{H}_p}$ and $\mathbb{E}\left[\mathbf{K}_{\mathcal{H}\mathcal{H}_p}\right] = \mathbb{E}\left[\mathbf{K}_{\mathcal{H}\mathcal{H}}\right]$ is used. The future term in (3.71) can be reduced similarly. The second term in the brackets $\{\cdot\}$ of (3.60) is obtained by plugging (3.70) and (3.71) into (3.69). Derivation of the first term in the brackets $\{\cdot\}$ is obvious.

Chapter 4

Conclusions and Future Work

4.1 Conclusions

GOALS of pursuing the spectral efficiency and improving receiving performance cannot simultaneously be achieved under constraints of low TX power and a low computational complexity, since overheads in a frame format required for energy- and computationally-efficient reception techniques decrease the spectral efficiency.

This thesis has studied $\ell 1$ regularized channel estimation algorithms with the aim of solving the problem. Chapter 2 has shown that, however, $\ell 1$ regularized channel estimation algorithms do not improve their MSE performance beyond the aCRB which is the performance bound for the conventional $\ell 2$ channel estimation techniques. In the case that the assumptions required for the conventional $\ell 2$ techniques do not always hold, nevertheless, the $\ell 1$ regularized algorithms can compensate the performance deterioration from which the $\ell 2$ techniques can suffer seriously. Therefore, the $\ell 1$ regularized algorithms are *robust* in the sense that they can achieve the expected estimation performance even under the deficient assumptions. Based on the theoretical analysis, this thesis has proposed new robust channel estimation algorithms as follows:

- the $\ell 1$ LS and $\ell 2$ MMSE-based hybrid algorithm [60] to solve the tracking error problem in intermittent TX scenarios;
- the $\ell 1$ MB algorithm [60, 61] to improve the noise whitening problem when TSs are not ideally uncorrelated;

- the CHATES algorithm [40] to solve the IBI problem after performing *sufficient* turbo iterations required for a turbo receiver to obtain BER convergence;

- the s-$\ell 1$ MB algorithm, which is formulated as a conditional $\ell 1$ MMSE problem [68], to improve the IBI problem in the first turbo iteration.

These new algorithms were detailed in Section 2.2.3, 2.C, 3.2.2 and 3.2.3, respectively.

Moreover, this thesis has proposed a new frequency domain turbo equalization technique, CHATUE2 [40]. The CHATUE algorithms do not assume CP-transmission and allows us to use a lower rate code by utilizing the time duration allocated for a CP. Section 3.1.3 has shown that, however, the previously-proposed CHATUE1 algorithm can suffer from the noise enhancement at the equalizer output. The theoretical analysis supported with simulation results shows that the CHATUE2 algorithm proposed in Section 3.1.5 can solve the noise enhancement problem after performing sufficient turbo iterations by utilizing the composite replica.

A communication system utilizing the proposed channel estimation and equalization algorithms can improve the spectral- and/or energy-efficiencies with reasonable computational complexity, as demonstrated in Section 3.3. It should be noticed that, however, as mentioned in Section 1.2.1, the spectral efficiency can be improved by MIMO transmission and/or multi-level modulation techniques. Therefore, as shown in Section 3.3.3, a significant BER performance gain is not always promised in a massive MIMO system.

4.2 Future Work

This thesis has concentrated on exploring spectrally efficient wireless communications for single-user SISO and MIMO systems. The proposed channel estimation algorithms can, however, be extended for multi-user MIMO systems with/without unknown interferences by combining spatial-subspace projection techniques [21, 24, 26].

The proposed $\ell 1$ regularized algorithms has been verified in a turbo reception framework with the single carrier signaling, however, they can be applicable to other transmission schemes. For example, as shown in Section 2.C, the $\ell 1$ MB channel estimation algorithm can improve the pilot contamination problem [59] in massive MIMO systems. It should be noticed

that the pilot contamination problem can happen both in the SC-FDMA and OFDMA systems. Application of the proposed algorithms to OFDMA systems is, therefore, an interesting topic.

Bibliography

[1] G. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communications: tutorial, survey, and open issues," *Wireless Communications, IEEE*, vol. 18, no. 6, pp. 28–35, December 2011.

[2] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology [Invited Paper]," *Wireless Communications, IEEE*, vol. 17, no. 3, pp. 10–22, June 2010.

[3] L. Tong, B. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *Signal Processing Magazine, IEEE*, vol. 21, no. 6, pp. 12–25, Nov 2004.

[4] D. Falconer, S. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58 –66, Apr. 2002.

[5] H. Myung, J. Lim, and D. Goodman, "Single carrier FDMA for uplink wireless transmission," *Vehicular Technology Magazine, IEEE*, vol. 1, no. 3, pp. 30–38, Sept 2006.

[6] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046 –1061, Jul. 1999.

[7] M. Tühler, R. Koetter, and A. Singer, "Turbo equalization: principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754 –767, May. 2002.

[8] M. Tuüchler and J. Hagenauer, "Linear time and frequency domain turbo equalization," in *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd*, vol. 2, 2001, pp. 1449–1453 vol.2.

[9] K. Kansanen, "Wireless broadband single-carrier systems with MMSE turbo equalization receivers," *University of Oulu, Oulu Finland*, 2005.

[10] 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.

[11] K. Kansanen and T. Matsumoto, "An analytical method for MMSE MIMO turbo equalizer EXIT chart computation," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 59 –63, Jan. 2007.

[12] P. J. Davis, *Circulant matrices*. American Mathematical Soc., 1979.

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

[14] M. Tüchler, R. Otnes, and A. Schmidbauer, "Performance of soft iterative channel estimation in turbo equalization," in *Communications, 2002. ICC 2002. IEEE International Conference on*, vol. 3, 2002, pp. 1858–1862 vol.3.

[15] T. Abe and T. Matsumoto, "Space-time turbo equalization in frequency-selective mimo channels," *Vehicular Technology, IEEE Transactions on*, vol. 52, no. 3, pp. 469–475, May 2003.

[16] 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.

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

[18] European Telecommunications Standards Institute (ETSI), "Spatial channel model for MIMO simulations (3GPP TR 25.996)," Sep. 2014.

[19] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *Signal Processing, IEEE Transactions on*, vol. 44, no. 12, pp. 3069–3078, Dec 1996.

[20] E. Lindskog and C. Tidestav, "Reduced rank channel estimation," in *Vehicular Technology Conference, 1999 IEEE 49th*, vol. 2, Jul 1999, pp. 1126–1130 vol.2.

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

[22] S. Bensley and B. Aazhang, "Subspace-based channel estimation for code division multiple access communication systems," *Communications, IEEE Transactions on*, vol. 44, no. 8, pp. 1009–1020, Aug 1996.

[23] 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.

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

[25] S. Ferrara, T. Matsumoto, M. Nicoli, and U. Spagnolini, "Soft Iterative Channel Estimation With Subspace and Rank Tracking," *IEEE Signal Process. Lett.*, vol. 14, no. 1, pp. 5–8, Jan. 2007.

[26] 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.

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

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

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

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

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

[42] 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.

[43] 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.

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

[45] Y. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, 1993, pp. 40–44 vol.1.

[46] 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.

[47] 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.

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

[49] 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.

[50] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

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

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

[53] 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.

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

[55] 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.

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

[57] 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.

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

[59] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 8, no. 5, pp. 742–758, Oct 2014.

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

[61] ——, "Performance of an $\ell 1$ regularized subspace-based MIMO channel estimation with random sequences," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 112–115, Feb 2016.

[62] R. Gold, "Optimal binary sequences for spread spectrum multiplexing (corresp.)," *Information Theory, IEEE Transactions on*, vol. 13, no. 4, pp. 619–621, October 1967.

[63] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 186–195, February 2014.

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

[65] K. Anwar and T. Matsumoto, "Low complexity time-concatenated turbo equalization for block transmission without guard interval: Part 1 - the concept," *Wireless Personal Comm., Springer, DOI 10.1007/s11277-012-0563-0*, Apr. 2012.

[66] Z. Hui, K. Anwar, and T. Matsumoto, "Low complexity time-concatenated turbo equalization for block transmission without guard interval: Part 2 - application to SC-FDMA," *Wireless Personal Comm., Springer, DOI 10.1007/s11277-011-0409-1*, Sep. 2011.

[67] D. Wang, C. Wei, Z. Pan, X. You, C. H. Kyu, and J. B. Jang, "Low-complexity turbo equalization for single-carrier systems without cyclic prefix," in *Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on*, Nov. 2008, pp. 1091 –1095.

[68] Y. Takano, M. Juntti, and T. Matsumoto, "A conditional $\ell 1$ regularized MMSE channel estimation technique for IBI channels," to be submitted.

Abbreviations

AAD adaptive active-set detection. 25, 27, 48, 49, 53, 63, 64

aCRB *adaptive*-CRB. 40, 43, 58, 60, 121

AIC Akaike information criterion. 30, 98

AWGN additive white Gaussian noise. 20, 105, 116

BCJR Bahl, Cocke, Jelinek and Raviv. 22, 105, 107, 116

BER bit error rate. 15, 18, 41, 53, 58, 60, 63, 82, 105, 107, 113, 116, 118, 122

BIC Bayesian information criterion. 26, 31, 35

BPSK binary phase shift keyed. 20, 22

BS base station. 3, 41

CHATES chained turbo estimation. 20, 83, 96, 102, 103, 107, 109, 112, 113, 116, 118, 122

CHATUE chained turbo equalization. 82, 83, 96, 103–105, 107, 113, 114, 116, 122

CHATUE1 CHATUE version 1. 83, 85, 87–89, 91–93, 96, 103–105, 107, 108, 116

CHATUE2 CHATUE version 2. 83, 89, 91–93, 96, 103–105, 107, 108, 113, 115–117, 122

CIR channel impulse response. 2, 14, 17, 20, 21, 23, 24, 26, 28–32, 34, 37–39, 41, 43, 53, 58, 60, 69–71, 76, 80–84, 92, 96, 98, 99, 102, 103, 109, 116, 118

CP cyclic prefix. 1, 10, 15, 20, 23, 82, 83, 96, 105, 107, 113, 122

CRB Cramér-Rao bound. 16, 38, 39, 69, 83, 116

DFT discrete Fourier transform. 5, 86

DoS degree of sparsity. 48, 66

EXIT extrinsic information transfer. 104, 105

FD-TEQ frequency domain turbo equalization. 4

FD/SC-MMSE frequency domain soft-cancelation and MMSE. 21

FDE frequency domain equalization. 1, 4, 5, 14

FEC forward error correction. 18

FFT fast Fourier transform. 5, 93, 95

FIR finite impulse response. 2, 10, 28

GI guard interval. 1, 5, 7, 10, 15, 20, 83, 103, 109, 113, 116

IBI inter-block-interference. 1, 2, 5, 7, 20, 23, 83, 88, 96–100, 102, 107, 109, 116, 118, 122

IFFT inverse fast Fourier transform. 93, 95

ISI inter-symbol-interference. 2, 3, 20

ITDSE iterative detection/estimation with threshold by "structured" least squared channel estimation. 26, 49, 66–68

LASSO least absolute shrinkage and selection operator. 27

LHS left-hand side. 5

LLR log-likelihood ratio. 14, 21, 22, 53, 87, 96, 103–105, 112, 116

LS least squares. 5–7, 15, 17, 22–27, 29–35, 37, 38, 43, 45, 48, 49, 58, 60, 66–68, 70, 83, 99, 100, 102, 121

MB multi-burst. 16, 17, 27, 30–32, 34, 35, 38, 40, 49, 53, 56, 58, 60, 63, 69, 70, 72, 76, 80, 81, 98, 99, 102, 109, 112, 113, 121

MI mutual information. 53, 104, 105

MIMO multiple-input multiple-output. 10, 11, 17–19, 21, 23, 24, 30, 39, 41–43, 48, 58, 60, 69, 72, 80, 81, 84, 85, 107, 113, 122

MMSE minimum mean square error. 15, 16, 27, 31, 58, 60, 98, 118, 121, 122

MS mobile station. 41

MSE mean squared error. 7, 10, 11, 15, 16, 18, 25, 26, 35, 37, 38, 40, 41, 43, 45, 48, 49, 60, 64, 67, 70, 100, 101, 103, 109, 116

NaCRB normalized aCRB. 43, 49, 53, 109, 112

NCRB normalized CRB. 53, 72

NMSE normalized MSE. 41, 43, 45, 48, 49, 53, 56, 58, 72, 76, 80, 81, 109, 112

NP non-polynomial. 69, 72

OFDM orthogonal frequency division multiplexing. 3

OFDMA orthogonal frequency division multiple access. 3, 4, 82, 123

OMP orthogonal matching pursuit. 23, 24, 48, 66, 67

P/S parallel-to-serial. 21

PAPR peak-to-average-power ratio. 3

PB Pedestrian-B. 16, 21, 22, 41, 43, 109, 116

PCA principal component analysis. 27, 29, 31, 34, 99

PN pseudo noise. 42, 72, 76

RF radio frequency. 3

RHS right-hand side. 96, 100

Rx receive. 10, 16, 20, 21

s-ℓ1 MB *self-supervised* ℓ1 MB. 98, 100, 102, 109, 112, 113, 118, 122

S/P serial-to-parallel. 19

SC-FDMA single-carrier frequency division multiple access. 4, 82, 123

SCM spatial channel model. 41

SCotLASS simplified component technique-LASSO. 27

SIMO single-input multi-output. 24, 39, 80

SISO single-input single-output. 2, 6, 39, 80, 103, 116, 122

SNR signal-to-noise ratio. 7, 16, 26, 40, 43, 49, 53, 58, 82, 83, 88, 104, 105, 109, 116

SP subspace pursuit. 48

SVD singular value decomposition. ix, 7, 34

TEQ-CP turbo equalization with CP-transmission. 82, 83, 87, 88, 92, 93, 103–105, 107, 108, 113, 114, 116, 117

TS training sequence. 1, 5–7, 10, 11, 14, 20, 42, 43, 48, 69–72, 76, 80, 81, 83, 84, 96, 100, 109

TX transmission. 1–4, 7, 20, 21, 28, 34, 38, 41, 42, 56, 58, 60, 69, 70, 76, 80, 85, 86, 95, 96, 98, 99, 101, 121

Tx transmit. 10, 16, 20, 29, 34

V-BLAST vertical-Bell laboratories layered space-time. 19

VA Vehicular-A. 7, 16, 21, 41, 43, 109, 113, 116

vec-OMP vectorized OMP. 48

WSSUS wide-sense stationary uncorrelated scattering. 21, 22

ZD zero-tap detection. 23–25, 66