Joint Semi-blind Channel Estimation and Finite Alphabet Signal Recovery Detection for large-scale MIMO systems

Hajji Zahran, Abdeldjalil Aissa El Bey, Karine Amis

Joint Semi-blind Channel Estimation and Finite Alphabet Signal Recovery Detection for large-scale MIMO systems

Zahran Hajji, *Student Member IEEE*, Abdeldjalil Aïssa-El-Bey, *Senior Member IEEE* and Karine Amis, *Member IEEE*

***Abstract*—In this paper, we consider large-scale MIMO systems and we address the channel estimation problem. We propose an iterative receiver consisting of the cascade of a semi-blind least-squares channel estimation algorithm with a simplicity-based detection algorithm for finite-alphabet signals (FAS and FAS-SAC). A minimum number of pilot sequences is used to get an initial channel estimation. The detection algorithm outputs are then used to refine it gradually. Two feeding methods are studied. The first one uses raw detection outputs. The second one is based on hard decisions and enables better performance. Theoretical MSEs are calculated in both cases. Simulations assess the efficiency of the proposed iterative procedure compared to the state-of-the-art methods and show that it performs very close to the ideal scenario where all the communication frame sequences are known.**

***Index Terms*—Large-scale MIMO, Channel estimation, finite-alphabet, simplicity, semi-blind, Joint channel estimation and detection.**

I. Introduction

The predicted exponential growth in the number of mobile connected machines and the traffic of data they represent motivate 5G designers to look for new technologies and approaches to address the mounting demand. It was theoretically demonstrated that usual schemes cannot achieve the sum-rate capacity of multiuser wireless systems and the maximum number of supported users is limited by the total amount of orthogonal resources [1]. To overcome this problem and in order to support massive connectivity of users and devices, enhanced technologies are needed.

The massive MIMO (Multiple-Input Multiple-Output) concept is considered to address technological challenges raised by high-volumes of traffic together with the continuously-increasing number of connected devices in communication systems such as 5G and beyond, internet of things (IoT) and wireless sensor networks (WSN) [2], [3]. The idea is to implement a large number of antennas to better exploit the spatial diversity so as to provide higher throughput under spectrum limitations. Massive MIMO [4] (also called Large-Scale Antenna Systems) refers to a wireless communication system, and can be seen as a particular large-scale multiple-input multiple-output system that involve high dimensions.

Paper submitted on April 9, 2021 and revised on July 7, 2021. This work was supported in part by the Brittany Region and in part by the Pôle de Recherche Avancée en Communications (PRACOM).

Z. Hajji, A. Aïssa-El-Bey and K. Amis are with IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238, France.

The performance of large-scale MIMO systems may highly depend on the accuracy of channel state information (CSI). The effect of imperfect CSI on the achievable capacity is investigated in several papers, where a lower bound on capacity is defined as a function of the Cramer-Rao bound (CRB) [5]–[8]. In [9], the authors showed that the capacity gain of the perfect CSI case over the imperfect one decreases when the SNR increases. This is due to the fact that the two cases share the same optimal input covariance matrix which tends to the identity matrix for high SNR. However, when CSI is required at the transmitter to determine the precoding scheme, the degradation due to erroneous channel estimates increases with the SNR and leads to the saturation of the effective SNR.

CSI is usually estimated thanks to known training sequences inserted in the data frame. This approach is well investigated in several works. It consists in dividing the transmission into a training phase and a data phase. In the training phase, known pilot sequences at the transmitter and receiver sides are transmitted in order to calculate an estimate at the receiver. This estimate can be obtained using ML [10] or MMSE criterion [11] and then it is used for detection in the data phase. When a small number of pilots is considered, the throughput loss due to pilots would be less at the price of an inaccurate estimate of the channel which degrades the detection performance. On the other hand, when large pilot sequences are considered, the quality of the channel estimate improves but the time for data transmission is reduced comparatively.

The problem of the minimum number of pilots is addressed in point-to-point frequency flat and selective channels [12], [13], multiuser MIMO channels [14]. In [12], the authors showed that their optimal number can be made equal to the number of transmit antennas by adjusting the power level used to transmit pilot sequences (higher than for data). With uniform power allocation between pilots and data, the optimal number can be much larger than the number of transmit antennas, which reduces the spectrum efficiency and thus limits the benefit of large-scale MIMO systems. The multiuser MIMO channel estimation is investigated in [14] together with the question of the required number of training sequences. Given the coherence time and the number of BS antennas, the optimal number of pilot sequences and the optimal number of users to serve simultaneously are fixed by maximizing a lower bound on the sum-rate on the downlink.

To avoid the spectral efficiency loss, all resources can be allocated to the data and the channel can be blindly

References	Channel estimation	Detection algorithm	Complexity	Performance	Spectral efficiency	System Configuration
E. Nayebi *et al.* [11]	EM	MMSE	High	+	Poor	$n > N$
C. H. Aldana *et al.* [20]	EM	MMSE	Very high	++	Very poor	$\forall(N, n) \in \mathbb{N}^2$
E. de Carvalho *et al.* [23]	EM	MMSE	Very high	++	Poor	$N = 1$ and $n \geqslant 2$
M. Abuthinien *et al.* [24]	ML	MMSE	High	++	Poor	$n \geqslant N$
Proposed method	LS	FAS	Medium	+++	Good	$\forall(N, n) \in \mathbb{N}^2$

TABLE I: Semi-blind channel estimation algorithms comparison for N inputs and n outputs systems.

estimated from the received data signal only [15], [16] (blind estimation techniques). However, in such cases, the channel may be identified within some ambiguities and errors which can degrade the quality of the CSI estimates. This may have a direct impact on the achieved throughput of the MIMO system and the performance of detection algorithms.

To reach a compromise between spectral efficiency and CSI estimate accuracy, an alternative solution is to adopt a semi-blind channel estimation procedure [17]–[19]. The idea is to use a limited number of training sequences in order to get an initial channel estimate and then improve its quality by considering data symbols estimates together with pilot sequences.

This method has been well investigated in several papers [11], [20]–[24]. In [20], the authors studied the channel and signal identifiability conditions for an underdetermined MIMO system. They proposed a channel estimation algorithm expectation-maximization (EM) in the frequency domain with a discrete random variable model for the unknown data. The authors in [21] proposed two iterative EM-based channel estimation algorithms. In [23], two semi-blind channel estimation algorithms for single input multiple output (SIMO) systems are studied with deterministic and Gaussian models. The performance of these algorithms is investigated when the frame length is infinitely large. In [24], a semi-blind channel estimation technique is proposed using a two-level optimization loop for joint channel estimation and data detection. In [11], the authors investigate two different semi-blind channel estimation schemes based on the EM algorithm. The first algorithm considers a Gaussian distribution of the data symbols and the second one is derived by using additional channel priors to improve the channel estimation quality.

In [25], we focused on detection and we proposed simplicity-based algorithms, assuming perfect CSI knowledge. We showed that the proposed simplicity-based algorithms offer good performance in terms of detection. This is why we decided to study their integration in an iterative semi-blind channel estimation scheme. The principle is to use pilot sequences to get an initial channel estimate that will be used for the data detection based on FAS and FAS-SAC algorithms. Data estimates are then provided to a ML-based channel estimator to improve the CSI accuracy. We first empirically validated the key principle in [26] in the coded case. In this paper, we focus on the non-coded case and our purpose is a theoretical analysis of performance and limitations of the iterative scheme.

The strategy used to feed the CSI estimate from detection output is crucial. In most papers, raw data detection output (soft decisions) are used with the risk of error propagation, yielding degradation of the CSI estimate. In this paper, we propose to mitigate this weakness by providing hard decisions (alphabet symbols prior) instead. Both strategies are investigated with the mean squared error (MSE) as a performance criterion. For the first strategy based on soft-decision outputs, we compute the CRBs as lower bounds of the MSEs. For the second strategy which introduces a non-linear function, the CRB is not defined (due to the non-linearity) and thus, the theoretical expression of the MSE is derived instead. The bit error rate is also assessed.

Our contributions are: ***(i)*** the exploitation of FAS and FAS-SAC detection outputs to gradually refine the CSI estimate within an iterative channel estimation algorithm. We impose a limited number of training sequences to obtain the initial CSI estimate. We show that the proposed approach can also be applied in underdetermined systems contrary to the state-of-the art methods. ***(ii)*** an analytical expression of the CRBs when the channel estimator is fed by raw detection outputs, ***(iii)*** a second feeding strategy based on hard decisions on data from detection outputs for further CSI estimate accuracy ***(iv)*** and the theoretical computation of corresponding asymptotic MSE.

The paper is organized as follows. Section II describes the large-scale MIMO system model. In Section III, we present an overview of channel estimation techniques and simplicity-based detection algorithms. The Cramer-Rao bound is also investigated for semi-blind channel estimation. In Section IV, we propose semi-blind channel estimation algorithms based on the output of FAS and FAS-SAC detection algorithms in the uncoded case. CRBs and asymptotic MSEs are also calculated. Finally, simulation results are presented in Section V and we conclude the paper in Section VI.

Notations: Boldface upper case letters and boldface lower case letters denote matrices and vectors, respectively. For the transpose, transpose conjugate and conjugate operations we use $(.)^T$, $(.)^H$ and $(.)^*$, respectively. $\otimes$ is the Kronecker product. $\boldsymbol{I}_k$ is the $k \times k$ identity matrix and $\mathbf{1}_k$ is the all-one size-k vector. Let $\boldsymbol{z} \in \mathbb{C}^k$ be a complex-valued vector of size k. We denote by $\underline{\boldsymbol{z}} \in \mathbb{R}^{2k}$ its real-valued transformed vector which can be defined by $\underline{\boldsymbol{z}} = \begin{pmatrix}\mathrm{Re}\,(\boldsymbol{z}) & \mathrm{Im}\,(\boldsymbol{z})\end{pmatrix}^T$. Let also $\boldsymbol{H} \in \mathbb{C}^{n\times N}$ a complex-valued matrix with size $n \times N$, we denote by $\underline{\boldsymbol{H}} \in \mathbb{R}^{2n\times 2N}$ its real-valued matrix version, which is defined by $\underline{\boldsymbol{H}} = \begin{pmatrix}\mathrm{Re}\,(\boldsymbol{H}) & -\mathrm{Im}\,(\boldsymbol{H}) \\ \mathrm{Im}\,(\boldsymbol{H}) & \mathrm{Re}\,(\boldsymbol{H})\end{pmatrix}$. Given two integer values k and i, $\delta_{k,i} = 1$ if $k = i$ and $\delta_{k,i} = 0$ if $k \neq i$

II. System model

Let us consider the uplink transmission in a communication system with TDD protocol where the base station is equipped with n antennas and each of the N users has a single antenna. We assume a block fading channel model, where the channel coefficients are constant during a block of T symbols and change to independent values at the next block transmission. Each transmitted frame consists of T_p pilot vectors and T_d data vectors ($T = T_p + T_d$). We also consider unit-variance channel coefficients, which is equivalent either to neglect the large-scale fading (the users are supposed to be located at similar distances from the base station) or to assume perfect power control at the user side. When N and n take high values, the resulting model amounts to a large-scale MIMO system with N inputs and n outputs.

Under the above assumptions, the received signal can be modelled as:

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{Z}. \tag{1}$$

$\boldsymbol{Y} = (\boldsymbol{Y}_p, \boldsymbol{Y}_d)$ is the received signal matrix. $\boldsymbol{Y}_p$ and $\boldsymbol{Y}_d$ are the $n \times T_p$ pilot received matrix and the $n \times T_d$ data received matrix respectively. $\boldsymbol{H}$ is the $n \times N$ complex channel matrix. $\boldsymbol{X}$ stands for the transmitted signal matrix. This $N \times T$ complex-valued matrix can be decomposed as $\boldsymbol{X} = (\boldsymbol{X}_p, \boldsymbol{X}_d)$. $\boldsymbol{X}_p$ and $\boldsymbol{X}_d$ are the $N \times T_p$ pilot transmitted matrix and the $N \times T_d$ data transmitted matrix respectively.

We denote by $\boldsymbol{x}(t)$ the t-th column of $\boldsymbol{X}$ which is the transmitted vector at time t. Its k-th element x_k belongs to $\mathcal{B} = \{\beta_1, \beta_2, .., \beta_M\}$ such that its real and imaginary parts take values on $\mathcal{F} = \{\alpha_1, \alpha_2, .., \alpha_p\}$ where $p = \sqrt{M}$. Pilot symbols are assumed to be known at the BS. The rows of the pilot transmitted matrix are assumed to be mutually orthogonal, that is to say $\boldsymbol{X}_p \boldsymbol{X}_p^H = T_p \boldsymbol{I}_N$. Transmitted data symbols are independent and identically distributed (i.i.d.): $\mathbb{E}[\boldsymbol{x}(t)\boldsymbol{x}(t)^H] = \boldsymbol{I}_N$ for all $\boldsymbol{x}(t)$ with $t = T_p, \ldots, T$. Note that the channel use at time t, for $t = 1, \ldots, T$, corresponds to the received vector given by:

$$\boldsymbol{y}(t) = \boldsymbol{H}\boldsymbol{x}(t) + \boldsymbol{z}(t). \tag{2}$$

The Maximum Likelihood (ML) estimate of $\boldsymbol{H}$ based on both training and data signals is given by

$$\hat{\boldsymbol{H}}_{ML} = \arg\max_{\boldsymbol{H}} \log p(\boldsymbol{Y}|\boldsymbol{H}), \tag{3}$$

where $p(\boldsymbol{Y}|\boldsymbol{H})$ is the probability of $\boldsymbol{Y}$ conditionally to $\boldsymbol{H}$.

III. Overview on channel estimation and simplicity-based detection algorithms

A. Channel estimation

In this section, we present the pilot-based ML estimator which only uses the training sequences as well as a full data-based ML estimator which assumes perfect data estimation. The last one can serve as a lower bound on the performance of semi-blind estimation (which consists of a first step of initialization of the channel estimate thanks to pilots followed by estimate refinement with data decisions).

1) ML estimation based on training Pilot Sequences [10]: The ML estimate of the channel matrix $\boldsymbol{H}$ based on pilot sequences $\boldsymbol{X}_p$ is given by:

$$\hat{\boldsymbol{H}}_{\text{ML}}^{\text{training}} = \frac{1}{T_p} \boldsymbol{Y}_p \boldsymbol{X}_p^H. \tag{4}$$

To minimize the MSE subject to the transmit power, the training sequences must be orthogonal, i.e $\boldsymbol{X}_p \boldsymbol{X}_p^H = T_p \boldsymbol{I}_N$. The corresponding mean square error (MSE) is then computed as

$$\mathbb{E}\left[\left\|\boldsymbol{H} - \hat{\boldsymbol{H}}_{\text{ML}}^{\text{training}}\right\|_F^2\right] = \frac{nN\sigma^2}{T_p} \tag{5}$$

The reliability of the channel estimate based on pilot sequences highly depends on the number of orthogonal sequences.

2) ML estimation based on full data [10]: The full data-based ML estimator assumes that all data symbols are known at the receiver side and the performance of such estimator serves as a lower bound of the performance of semi-blind estimation. The channel estimate based on perfect knowledge of pilot matrix $\boldsymbol{X}_p$ and data matrix $\boldsymbol{X}_d$ is denoted by $\hat{\boldsymbol{H}}_{\text{ML}}^{\text{full}}$ and is computed as:

$$\hat{\boldsymbol{H}}_{\text{ML}}^{\text{full}} = (\boldsymbol{Y}\boldsymbol{X}^H)(\boldsymbol{X}\boldsymbol{X}^H)^{-1}. \tag{6}$$

Its corresponding mean square error (MSE) equals

$$\mathbb{E}\left[\left\|\boldsymbol{H} - \hat{\boldsymbol{H}}_{\text{ML}}^{\text{full}}\right\|_F^2\right] = n\,\sigma^2 \operatorname{tr}\left(\mathbb{E}\left[\left(\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\right]\right). \tag{7}$$

3) EM algorithm [5]: As data symbols are not known, the ML problem cannot be analytically solved in practice. It is necessary to use iterative algorithms that converge to the solution of (3). Among them, the EM algorithm [5] updates the channel estimate based on an old one in the following manner:

$$\hat{\boldsymbol{H}}_{i+1} = \arg\max_{\boldsymbol{H}} \mathbb{E}_{\Pr(\boldsymbol{X}_d|\boldsymbol{Y},\hat{\boldsymbol{H}}_i)}\left[\log \Pr\left(\boldsymbol{Y}, \boldsymbol{X}_d|\boldsymbol{H}\right)\right]. \tag{8}$$

As we can see, the algorithm involves an expectation step and a maximization one. The maximization step can be simplified and the updated estimate of the channel matrix can be written as

$$\begin{aligned}\hat{\boldsymbol{H}}_{i+1} \quad &= \quad \frac{1}{T_p}\left(\boldsymbol{Y}_p\boldsymbol{X}_p^H + \boldsymbol{Y}_d\,\mathbb{E}\left[\boldsymbol{X}_d|\boldsymbol{Y},\hat{\boldsymbol{H}}_i\right]^H\right) \\ &\qquad \left(\boldsymbol{I}_N + \frac{1}{T_p}\mathbb{E}\left[\boldsymbol{X}_d\boldsymbol{X}_d^H|\boldsymbol{Y},\hat{\boldsymbol{H}}_i\right]^H\right)^{-1}. \end{aligned} \tag{9}$$

To compute the estimate (9), the expectation step (E-step) must be defined. In practice, the data symbols are discrete random variables, which leads to complex E-step whose complexity grows exponentially with N. To overcome this problem, in [11], the authors propose to use an approximation and they assume that the data symbols are Gaussian. Thus $\mathbb{E}\left[\boldsymbol{X}_d|\boldsymbol{Y},\hat{\boldsymbol{H}}_i\right]^H$ and $\mathbb{E}\left[\boldsymbol{X}_d\boldsymbol{X}_d^H|\boldsymbol{Y},\hat{\boldsymbol{H}}_i\right]^H$ can be computed from the conditional probability density function of circularly symmetric Gaussian random vectors and we get the updated estimate as follows:

$$\begin{aligned}\hat{\boldsymbol{H}}_{i+1} \quad &= \quad \left(\boldsymbol{Y}_p\boldsymbol{X}_p^H + \sum_{t=T_p+1}^{T} \boldsymbol{y}(t)\left(\hat{\boldsymbol{x}}(t)\right)^H\right) \\ &\qquad \left(\boldsymbol{X}_p\boldsymbol{X}_p^H + \sum_{t=T_p+1}^{T}\left(\hat{\boldsymbol{x}}(t)\left(\hat{\boldsymbol{x}}(t)\right)^H + \boldsymbol{\Sigma}\right)\right)^{-1} \end{aligned} \tag{10}$$

where

$$\hat{\boldsymbol{x}}(t)=\left(\hat{\boldsymbol{H}}_i^H\hat{\boldsymbol{H}}_i+\sigma^2\boldsymbol{I}_N\right)^{-1}\hat{\boldsymbol{H}}_i^H\boldsymbol{y}(t), \tag{11}$$

and

$$\boldsymbol{\Sigma}=\sigma^2\left(\hat{\boldsymbol{H}}_i^H\hat{\boldsymbol{H}}_i+\sigma^2\boldsymbol{I}_N\right)^{-1} \tag{12}$$

4) Cramer-Rao bound of semi-blind channel estimation: In this section, we derive the CRB [5], the lower bound of covariance, for semi blind channel estimation. We assume that deterministic model in which the data signal is modeled as an unknown deterministic quantity.

The following theorem gives the CRB of any unbiased semi-blind channel estimation based on least-squares algorithm.

Theorem III.1. *The deterministic Cramer-Rao bound of the covariance matrix of any unbiased semi-blind estimate of $\boldsymbol{H}$ is given by:*

$$\mathrm{CRB}(\boldsymbol{H})=\left(\boldsymbol{\Pi}-\boldsymbol{\Pi}_{\mathrm{er}}\right)^{-1}, \tag{13}$$

where the matrix $\boldsymbol{\Pi}$ is defined as:

$$\boldsymbol{\Pi}=\frac{2}{\sigma^2}(\underline{\boldsymbol{X}}^T\underline{\boldsymbol{X}})\otimes\boldsymbol{I}_{2N} \tag{14}$$

with $\underline{\boldsymbol{X}}$ is the real-valued matrix version of $\boldsymbol{X}$,

$$\boldsymbol{\Pi}_{\mathrm{er}}=\sum_{t=T_p+1}^{T}\boldsymbol{\Omega}_t^T\left(\underline{\boldsymbol{H}}^T\underline{\boldsymbol{H}}\right)\boldsymbol{\Omega}_t, \tag{15}$$

with $\underline{\boldsymbol{H}}$ is the real-valued matrix version of $\boldsymbol{H}$, $\boldsymbol{\Omega}_t=\underline{\boldsymbol{x}}(t)^T\otimes\underline{\boldsymbol{H}}$ for $t=T_p+1,\dots,T$ and $\underline{\boldsymbol{x}}(t)$ is the real-valued vector version of $\boldsymbol{x}(t)$.

Proof: see Appendix.

Remark III.1. *When all symbols are correctly detected, $\sum_{t=T_p+1}^{T}\boldsymbol{\Omega}_t^T\left(\underline{\boldsymbol{H}}^T\underline{\boldsymbol{H}}\right)\boldsymbol{\Omega}_t=0$ and $\mathrm{CRB}(\boldsymbol{H})=\boldsymbol{\Pi}^{-1}$ which corresponds to the covariance matrix obtained with the full-data ML, leading to a MSE equal to $\mathrm{tr}(\boldsymbol{\Pi}^{-1})=\mathbb{E}\left[\left\|\boldsymbol{H}-\hat{\boldsymbol{H}}_{ML}^{full}\right\|_F^2\right]$ as defined in (7).*

B. Overview of simplicity-based detection algorithms

In this section, we provide a concise description of FAS and FAS-SAC algorithms. The interested reader can refer to [25] and [27] for further details. Both of them are compressed-sensing algorithms that exploit the finite alphabet property.

They require a real-valued formulation of the problem which is given hereinafter.

Let consider at time t, the following large-scale MIMO system:

$$\boldsymbol{y}=\boldsymbol{H}\,\boldsymbol{x}+\boldsymbol{z},\qquad \boldsymbol{x}\in\mathcal{B}^N \tag{16}$$

The equivalent real-valued system can be then written as:

$$\underline{\boldsymbol{y}}=\underline{\boldsymbol{H}}\,\underline{\boldsymbol{x}}+\underline{\boldsymbol{z}},\qquad \underline{\boldsymbol{x}}\in\mathcal{F}^{2N}. \tag{17}$$

Let us now describe the FAS and FAS-SAC algorithms.

1) FAS algorithm [25]: The vector $\boldsymbol{x}$ is recovered through its real valued-transformation $\underline{\boldsymbol{x}}$. We briefly describe the detection technique previously proposed in [25]. For that purpose, we introduce the following definition:

Definition III.1. ***Simplicity [28]*** *A given vector $\underline{\boldsymbol{x}}\in[\alpha_1,\alpha_p]^{\,2N}$ is called k-simple if it has exactly k entries different from α_1 and α_p.*

The simplicity property of $\mathcal{F}$ is exploited to define an optimization problem whose complexity is independent of the constellation size and is lower compared with [29], while performing the same in terms of error rate. The vector $\underline{\boldsymbol{x}}$ is simple and its components are minored by α_1 and majored by α_p. It can be decomposed as $\underline{\boldsymbol{x}}=\boldsymbol{B}_\alpha\boldsymbol{r}$ where $\boldsymbol{B}_\alpha=\boldsymbol{I}_{2N}\otimes[\alpha_1;\alpha_p]$ and $\boldsymbol{r}\in[0,1]^{4N}$. We used the previous decomposition to define the simplicity-based optimization problem given by [25]

$$\begin{aligned}\arg\min_{\boldsymbol{r}}\|\underline{\boldsymbol{y}}-\underline{\boldsymbol{H}}\boldsymbol{B}_\alpha\boldsymbol{r}\|_2 \quad &\text{subject to}\\ \boldsymbol{B}_1\boldsymbol{r}&=\boldsymbol{1}_{2N},\\ \boldsymbol{r}&\geq 0,\end{aligned} \tag{18}$$

where $\boldsymbol{B}_1=\boldsymbol{I}_{2N}\otimes\boldsymbol{1}_2^T$. The FAS detection amounts to the resolution of the optimization problem defined by (18), which can be solved by an interior point method [30] among others.

2) FAS-SAC algorithm [27]: In the FAS detection method, all sources are detected at once and some decisions may be less reliable than others. In [27], the authors propose a reliability measure based on the shadow area principle [31], [32] that exploits the output statistics. The authors first define the *centers* as the elements of $\mathcal{F}$. The principle is to take decision on components $\underline{x}_k$ such that $\hat{\underline{x}}_k$ is close enough to one center and cancel their contribution in the observation $\underline{\boldsymbol{y}}$ so as to proceed a novel detection iteration. To do so, they propose to take into account the reliabilities of the output $\hat{\underline{x}}_k$ based on the output distribution. Then, the algorithm takes decisions on reliable $\hat{\underline{x}}_k$, cancels their contribution from $\underline{\boldsymbol{y}}$ and proceeds another detection as shown in Fig. 1. Adjacent to shadow areas, the high-reliability intervals are defined as intervals of length 2η and are centered on the different symbols of $\mathcal{F}$. Let us denote by $\mathcal{A}$ the set of indices k such that $\hat{\underline{x}}_k$ is considered as reliable. The decision on $\hat{\underline{x}}_k, k\in\mathcal{A}$ is taken as the nearest symbol value in $\mathcal{F}$. We denote by $\tilde{\underline{\boldsymbol{x}}}_{\mathcal{A}}$ the resulting decision vector. The equivalent notations for unreliable elements (falling in shadow areas) are respectively $\overline{\mathcal{A}}$ for the set of indexes and v_N for its cardinality. The algorithm is detailed in Algorithm 1.

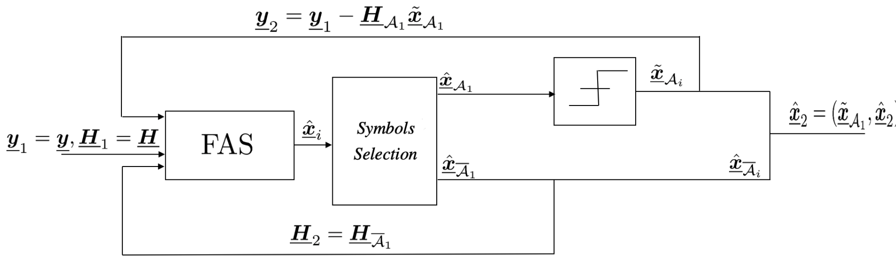

Fig. 1: FAS-SAC algorithm flowchart

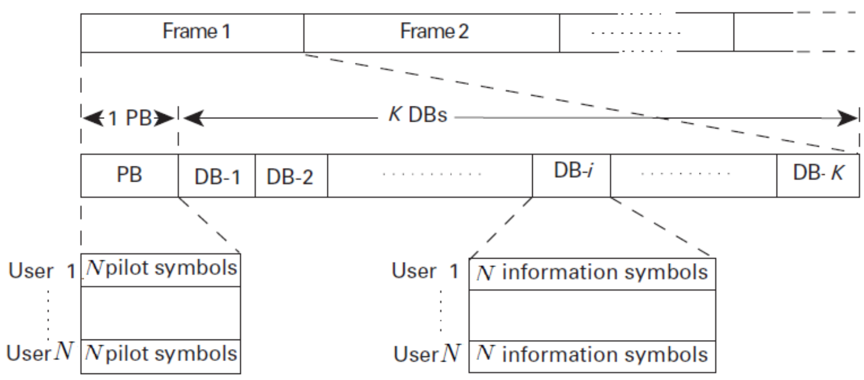

Fig. 2: Equivalent frame structure used at the receiver side for N simultaneous transmissions of identically-formatted frames.

Algorithm 1 Shadow Area Constrained (SAC) - FAS detection

Input: $\boldsymbol{H}$, $\boldsymbol{y}$

$\boldsymbol{r} = \arg\min_{\boldsymbol{r}} \|\underline{\boldsymbol{y}} - \underline{\boldsymbol{H}}\boldsymbol{B}_\alpha \boldsymbol{r}\|_2$ subject to $\boldsymbol{B}_1\boldsymbol{r} = \boldsymbol{1}_{2N}$ and $\boldsymbol{r} \geq 0$.

Compute $\hat{\underline{\boldsymbol{x}}} = \boldsymbol{B}_\alpha\, \boldsymbol{r}$.

Define $\mathcal{A} = \left\{ k \mid \min_{\alpha_i \in \mathcal{F}} |\hat{\underline{x}}_k - \alpha_i| \leq \eta,\ k \in \{1, \ldots, 2N\} \right\}$, $v_N = \mathrm{card}(\overline{\mathcal{A}})$.

Compute $\tilde{\underline{\boldsymbol{x}}}_{\mathcal{A}}$ by $\tilde{\underline{x}}_k = \arg\min_{\alpha_i \in \mathcal{F}} |\hat{\underline{x}}_k - \alpha_i|$, $k \in \mathcal{A}$ and $\tilde{\underline{\boldsymbol{y}}} = \underline{\boldsymbol{y}} - \underline{\boldsymbol{H}}_{\mathcal{A}} \tilde{\underline{\boldsymbol{x}}}_{\mathcal{A}}$.

$\tilde{\boldsymbol{r}} = \arg\min_{\tilde{\boldsymbol{r}}} \|\tilde{\underline{\boldsymbol{y}}} - \underline{\boldsymbol{H}}_{\overline{\mathcal{A}}} \tilde{\boldsymbol{B}}_\alpha \tilde{\boldsymbol{r}}\|_2$ subject to $\tilde{\boldsymbol{B}}_1 \tilde{\boldsymbol{r}} = \boldsymbol{1}_{v_N}$ and $\tilde{\boldsymbol{r}} \geq 0$.

Compute $\tilde{\underline{\boldsymbol{x}}}_{\overline{\mathcal{A}}} = \tilde{\boldsymbol{B}}_\alpha \tilde{\boldsymbol{r}}$.

Output: $\tilde{\underline{\boldsymbol{x}}}$.

The FAS-SAC detection algorithm can be implemented with a large number of iterations as shown in [27]. In the remainder of the paper, we consider the case with two iterations in order to use the theoretical study demonstrated in [27]. To the best of our knowledge, there doesn't exist other large-scale MIMO detection algorithms based on the simplicity principle. Among the best iterative large-scale MIMO detection algorithms, we can mention RTS and LAS algorithms which are outperformed by FAS-SAC (see [27] for detailed comparison).

IV. Semi-blind uplink channel estimation for large-scale MIMO systems

In this section, we develop a joint semi-blind channel estimation and detection schemes based on the simplicity-based algorithms (FAS) and (FAS-SAC) described above.

A. Proposed Semi-blind uplink channel estimation algorithms

We consider N users transmitting simultaneously identically-formatted frames. The channel is assumed to be constant over one frame duration. At the receiver side, the N transmissions can be modelled as a single transmission from N inputs. The equivalent frame structure is depicted in Fig. 2. It consists of a pilot block followed by K data blocks. The pilot block is the vertical stacking of N length-N training sequences (one per user). Each data block is the vertical stacking of N length-N data sequences (one per user). The system model introduced in Section II applies with $T_p = N$ and $T_d = KN$.
An iterative process consisting of the cascade of CSI-estimation and data detection is carried out. Data blocks are detected using either FAS or FAS-SAC detection algorithm. An initial CSI estimate is obtained by a pilot-based channel estimation. The detected data blocks are then used to gradually refine the channel estimates thanks to a least-squares algorithm. The steps of the iterative process are detailed hereinafter.

1) Initial channel estimate during pilot phase: Let $\boldsymbol{x}_p^u$ denote the transmitted pilot symbol vector from user u. Let $\boldsymbol{X}_p = (\boldsymbol{x}_p^1, \boldsymbol{x}_p^2, \ldots, \boldsymbol{x}_p^N)$ denote the $N \times N$ pilot matrix formed by the pilot symbol vectors transmitted by all users in the pilot transmission phase. The received signal matrix at BS, $\boldsymbol{Y}_p$, is given by:

$$\boldsymbol{Y}_p = \boldsymbol{H}\boldsymbol{X}_p + \boldsymbol{Z}_p, \tag{19}$$

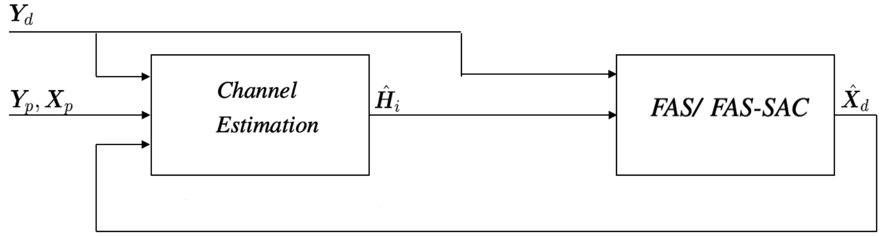

Fig. 3: Channel estimation algorithm based on detection FAS/FAS-SAC outputs

where $\boldsymbol{Z}_p$ is the noise matrix. Choosing orthogonal pilot sequences, we get $\hat{\boldsymbol{H}}_0$, the first estimate, as:

$$\begin{aligned} \hat{\boldsymbol{H}}_0 &= \frac{1}{T_p}\boldsymbol{Y}_p\boldsymbol{X}_p^H \qquad (20) \\ &= \boldsymbol{H} + \frac{1}{T_p}\boldsymbol{Z}_p\boldsymbol{X}_p^H \end{aligned}$$

This choice is adopted in order to avoid matrix inversion at first iteration and then reducing computation complexity.

2) Data detection using initial channel estimate: During data transmission phase, the received signal matrix at BS, $\boldsymbol{Y}_d$, is given by:

$$\boldsymbol{Y}_d = \boldsymbol{H}\boldsymbol{X}_d + \boldsymbol{Z}_d, \qquad (21)$$

where $\boldsymbol{X}_d$ is the concatenation of different data blocks. The received vector for the channel use at time t, for $t = T_p + 1, \ldots, T$ is

$$\boldsymbol{y}(t) = \boldsymbol{H}\boldsymbol{x}(t) + \boldsymbol{z}(t), \qquad (22)$$

The initial channel estimate $\hat{\boldsymbol{H}}_0$ obtained from (20) is used to detect the transmitted data vectors using FAS and FAS-SAC algorithms described in previous sections to get an estimate for the transmitted data matrix $\boldsymbol{X}_d$ denoted by $\hat{\boldsymbol{X}}_d$.
The initial CSI error $\hat{\boldsymbol{H}}_0 - \boldsymbol{H}$ is used to calculate the statistics needed to study the performance of the proposed schemes in the uncoded cases and to interface the proposed detectors with FEC decoder and channel estimation block in the coded case. We then get, at channel use t, $\boldsymbol{y}(t) - \hat{\boldsymbol{H}}_0\boldsymbol{x}(t)$, the updated noise vector at first iteration with zero mean and variance $(1+\frac{N}{T_p})\sigma^2$.

B. Channel estimation based on detection algorithms outputs

1) Channel estimation refinement based on FAS/FAS-SAC soft-decision output: Let define $\hat{\boldsymbol{X}} = (\hat{\boldsymbol{x}}(T_p+1), \ldots, \hat{\boldsymbol{x}}(T))$; where $\hat{\boldsymbol{x}}(t)$, for $t = T_p + 1, ..., T$, is the detected complex-valued FAS/FAS-SAC soft-decision output at time t based on the channel estimate at i-th iteration. The iterative receiver is illustrated in Fig. 3. The channel estimate at the $(i+1)$-th iteration is denoted by $\hat{\boldsymbol{H}}_{i+1}^{\mathrm{SD}}$ and defined as

$$\hat{\boldsymbol{H}}_{i+1}^{\mathrm{SD}} = \left(\boldsymbol{Y}_p\boldsymbol{X}_p^H + \boldsymbol{Y}_d\hat{\boldsymbol{X}}^H\right)\left(\boldsymbol{X}_p\boldsymbol{X}_p^H + \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^H\right)^{-1}. \qquad (23)$$

In order to compute the CRB of the proposed channel estimation algorithm based on FAS detection, let us remember that at time t, the components of the real-valued vector $\underline{\hat{\boldsymbol{x}}}(t)$ provided by the FAS algorithm can be classified into two sets [25]. The first set is the set of reliable elements which are exactly equal to the transmitted symbols. This set was referred to as Λ_t. In case of QPSK, its cardinality follows the binomial distribution with parameters $2N$ and $\frac{1}{2}$. The second set referred to as $\bar{\Lambda}_t$ contains the remaining noisy components. Its cardinality follows the same distribution as Λ_t.

Given $t = T_p + 1, \ldots, T$, let $\boldsymbol{G}_t^{\mathrm{FAS}}$ stand for the matrix defined from $\underline{\boldsymbol{H}}$ by annulling the columns whose indices are in Λ_t and the matrix $\boldsymbol{\Omega}_t^{\mathrm{FAS}}$ as $\boldsymbol{\Omega}_t^{\mathrm{FAS}} = \underline{\boldsymbol{x}}(t)^T \otimes \boldsymbol{G}_t^{\mathrm{FAS}}$.
Let us also introduce $\boldsymbol{C}_t^{\mathrm{FAS}} = \left(\underline{\boldsymbol{H}}_{\bar{\Lambda}_t}^T \underline{\boldsymbol{H}}_{\bar{\Lambda}_t}\right)^{-1}$. We define the matrix $\mathfrak{C}_t^{\mathrm{FAS}}$ such that $\mathfrak{C}_t^{\mathrm{FAS}}\left(\bar{\Lambda}_t(i), \bar{\Lambda}_t(j)\right) = \boldsymbol{C}_t^{\mathrm{FAS}}(i,j)$ for $(i,j) \in \{1, \ldots, |\bar{\Lambda}_t|\}^2$ and the other entries are equal to zero.

As mentioned in Section III-B2, the FAS-SAC algorithm is itself iterative (to make the difference with the iterations involved in the cascade of estimation and detection, we will refer to as inner iterations for the FAS-SAC). Analytical expressions of defining and performance parameters of FAS-SAC are available up to two inner iterations. This is why we fix to two the inner iteration number in order to determine the CRB of the channel estimation algorithm. As detailed in Section III-B2, the components of the FAS-SAC output after one inner iteration can be split into two subsets. The set $\mathcal{A}_t$ contains the components presumed to be reliable by the algorithm. It can be itself divided into two subsets: the sets of correct decisions and erroneous decisions denoted by $(\mathcal{A}_c)_t$ and $(\mathcal{A}_e)_t$ respectively and such that $\mathcal{A}_t = (\mathcal{A}_c)_t \cup (\mathcal{A}_e)_t$. The complementary set $\bar{\mathcal{A}}_t$ contains the symbols that are re-estimated in the second iteration. The updated noise vector in the second iteration is $\tilde{\boldsymbol{z}}(t)$ with variance $\sigma_{\tilde{z}}^2(\eta)$ calculated in [27].

Given $t = T_p + 1, \ldots, T$, let $\boldsymbol{G}_t^{\mathrm{FSAC}}$ be the matrix defined from $\underline{\boldsymbol{H}}$ by annulling the entries of columns whose indices are in $\mathcal{A}_t$ and let the matrix $\boldsymbol{\Omega}_t^{\mathrm{FSAC}} = \underline{\boldsymbol{x}}(t)^T \otimes \boldsymbol{G}_t^{\mathrm{FSAC}}$.

Let us also introduce $\boldsymbol{C}_t^{\mathrm{FSAC}} = \left(\underline{\boldsymbol{H}}_{\bar{\mathcal{A}}_t}^T \underline{\boldsymbol{H}}_{\bar{\mathcal{A}}_t}\right)^{-1}$. We define the matrix $\mathfrak{C}_t^{\mathrm{FSAC}}$ such that $\mathfrak{C}_t^{\mathrm{FSAC}}\left(\bar{\mathcal{A}}_t(i), \bar{\mathcal{A}}_t(j)\right) = \boldsymbol{C}_t^{\mathrm{FSAC}}(i,j)$ for $(i,j) \in \{1, \ldots, |\bar{\mathcal{A}}_t|\}^2$ and the other entries are equal to zero, for $t = T_p + 1, \ldots, T$. Then the CRB of both schemes is given by the following theorem.

Theorem IV.1. *The deterministic CRB of the channel estimation based on FAS detection algorithm with soft decision outputs is defined as*

$$\mathrm{CRB}(\boldsymbol{H}) = \left(\boldsymbol{\Pi} - \boldsymbol{\Pi}_{\mathrm{er}}^{\mathrm{FAS}}\right)^{-1}, \qquad (24)$$

where

$$\mathbf{\Pi}_{\text{er}}^{\text{FAS}} = \frac{2}{\sigma^2} \sum_{t=T_p+1}^{T} (\mathbf{\Omega}_t^{\text{FAS}})^T \mathfrak{C}_t^{\text{FAS}} (\mathbf{\Omega}_t^{\text{FAS}}).$$

Similarly, the deterministic CRB of the channel estimation based on FAS-SAC detection algorithm is defined as

$$\text{CRB}(\boldsymbol{H}) = \left(\mathbf{\Pi} - \mathbf{\Pi}_{\text{er}}^{\text{FSAC}}\right)^{-1}, \tag{25}$$

where

$$\mathbf{\Pi}_{\text{er}}^{\text{FSAC}} = \frac{2}{\sigma_{\tilde{z}}^2(\eta)} \sum_{t=T_p+1}^{T} (\mathbf{\Omega}_t^{\text{FSAC}})^T \mathfrak{C}_t^{\text{FSAC}} (\mathbf{\Omega}_t^{\text{FSAC}}).$$

Proof: see Appendix.

2) Channel estimation refinement based on FAS/FAS-SAC hard decision output: Let $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{x}}(T_p+1), \ldots, \tilde{\boldsymbol{x}}(T))$, where $\tilde{\boldsymbol{x}}(t)$ is the hard decision of the detected complex-valued FAS/FAS-SAC output at time t based on the channel estimate at the i-th iteration. The channel estimate at the $(i+1)$-th iteration is denoted by $\hat{\boldsymbol{H}}_{i+1}^{\text{HD}}$ and is computed as

$$\hat{\boldsymbol{H}}_{i+1}^{\text{HD}} = \left(\boldsymbol{Y}_p \boldsymbol{X}_p^H + \boldsymbol{Y}_d \tilde{\boldsymbol{X}}^H\right)\left(\boldsymbol{X}_p \boldsymbol{X}_p^H + \tilde{\boldsymbol{X}} \tilde{\boldsymbol{X}}^H\right)^{-1}. \tag{26}$$

The following theorem gives the asymptotic MSE of the proposed channel estimation schemes based on hard decisions of proposed detection algorithm outputs. Let us consider 4-QAM complex-valued symbols yielding BPSK real-valued symbols in the real-equivalent system. The estimated data matrix $\underline{\tilde{\boldsymbol{X}}}$ based on hard decisions simply writes $\underline{\tilde{\boldsymbol{X}}} = \underline{\boldsymbol{X}} + \underline{\Delta_{\boldsymbol{X}}}$ where $\underline{\Delta_{\boldsymbol{X}}}$ is the error matrix with entries in the set $\{-2, 0, 2\}$.

Theorem IV.2. *Let us consider QPSK complex-valued alphabet. The asymptotic MSE of the channel estimation combined with hard decision outputs detection equals:*

$$\begin{aligned} \text{MSE} &= \mathbb{E}\left[\left\|\boldsymbol{H} - \hat{\boldsymbol{H}}\right\|_F^2\right] \qquad (27) \\ &= \frac{2N\sigma^2}{T} + \frac{\sigma^2}{T^2}\text{tr}\left(\mathbb{E}\left[\underline{\Delta_{\boldsymbol{X}}}^T \underline{\boldsymbol{X}}\, \underline{\boldsymbol{H}}^T \underline{\boldsymbol{H}}\, \underline{\boldsymbol{X}}^T \underline{\Delta_{\boldsymbol{X}}}\right]\right) \end{aligned}$$

Proof: see Appendix.

V. Simulation results and complexity analysis

A. Simulation results

1) Comparison with EM algorithm: In Fig. 4, we compare the MSE of the proposed iterative channel estimation algorithm based on soft decision FAS-output under one iteration given in (23) to the ML estimators and the EM algorithm with two iterations described in Section III for an overdetermined system with $N = 8$ and $n = 64$. We show that the same performance at the second iteration can be achieved by the proposed algorithm.
In Fig. 5, we consider a determined system with $N = n = 64$. It is shown that the EM algorithm exhibits no improvement compared to ML training-based estimation in such configuration. However, the proposed algorithm based on soft decision FAS output presents a gain over the ML training-based one beyond SNR=11dB. This gain is of about 2dB at 10^{-3} MSE.

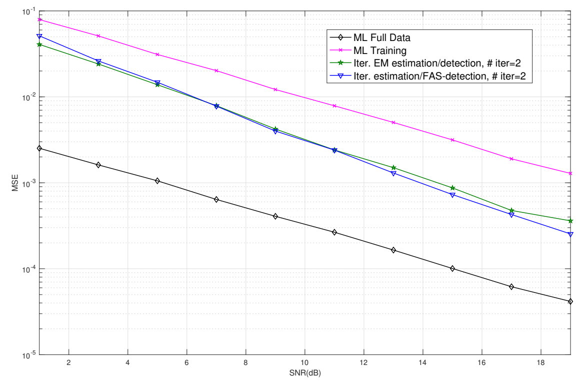

Fig. 4: MSE versus SNR with uncoded QPSK, $N = 8$, $n = 64$, $T_p = 16$ and $T = 512$.

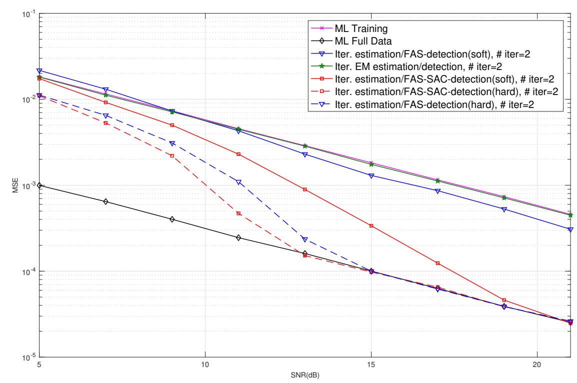

Fig. 5: MSE versus SNR with uncoded QPSK, $n = N = 64$, $T_p = 64$ and $T = 1280$.

The second proposed algorithm based on soft decision FAS-SAC output outperforms EM algorithm and soft decision FAS-output algorithm over the whole SNR range. It achieves the same MSE as the full data based estimation beyond SNR=19dB. Let us now compare the two proposed feeding strategies based on raw detection outputs (soft-decisions) on one hand and on hard decisions on the other hand. In Fig. 5, we show that the second strategy based on hard decisions outperforms the first one with a gain of about $5dB$ for FAS detection and $2.7dB$ for FAS-SAC detection algorithm at 10^{-3} MSE.

2) Comparison with theoretical bounds: Let us first study the feeding strategy based on soft decisions. We consider the proposed scheme with FAS detection with $N = 8$ and $n = 64$ (overdetermined system) in Fig. 6 and with FAS-SAC detection and $n = 64$ and $N = 8$ (determined system) in Fig. 7. In both cases, we observe that the proposed algorithm performs close to its CRB.

As for the second feeding strategy based on hard decisions, considering both FAS and FAS-SAC algorithm with $N = n = 64$ in Fig. 8, we observe that the empirical MSE is the same as the asymptotic MSE derived in (27) and converges towards the lower bound (full data-based ML) in both cases, which

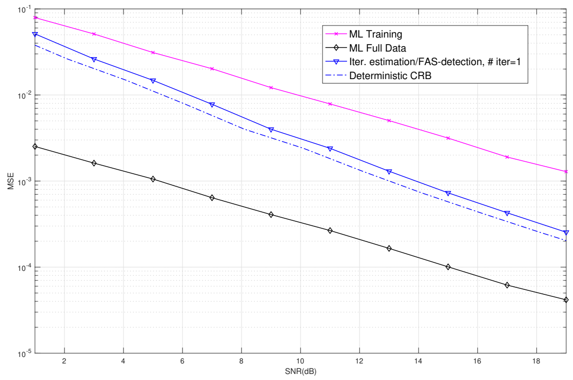

Fig. 6: MSE versus SNR with uncoded QPSK, $n = 64$, $N = 8$, $T_p = 16$ and $T = 512$ (overdetermined system, soft decision FAS output-based scheme).

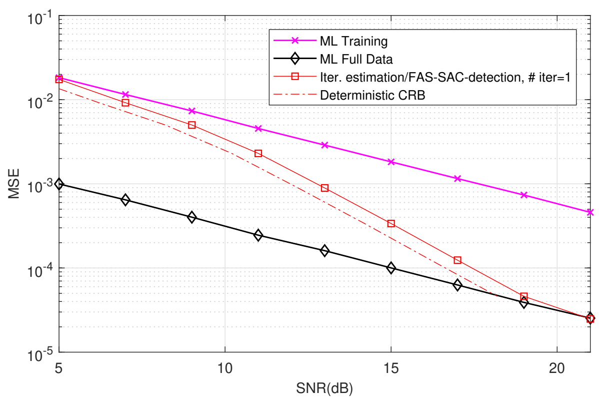

Fig. 7: MSE versus SNR with uncoded QPSK, $n = N = 64$, $T_p = 64$ and $T = 1280$ (determined system, soft decision FAS-SAC output-based scheme).

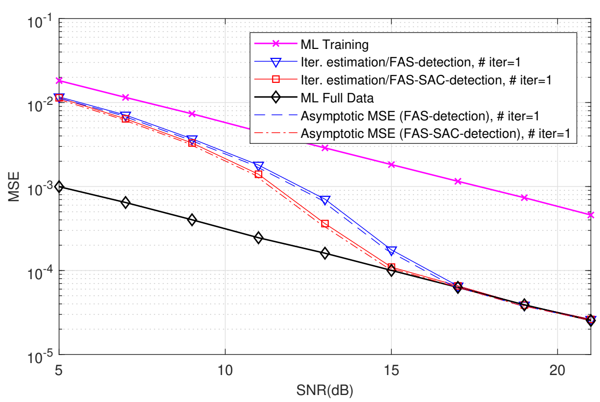

Fig. 8: MSE versus SNR with uncoded QPSK, $n = N = 64$, $T_p = 64$ and $T = 1280$ (hard decision FAS and FAS-SAC outputs-based schemes).

assesses the validity of the theoretical analysis.

3) Impact of iterations on the estimation accuracy: We consider the second feeding strategy based on hard decisions and both FAS and FAS-SAC detection and we measure the

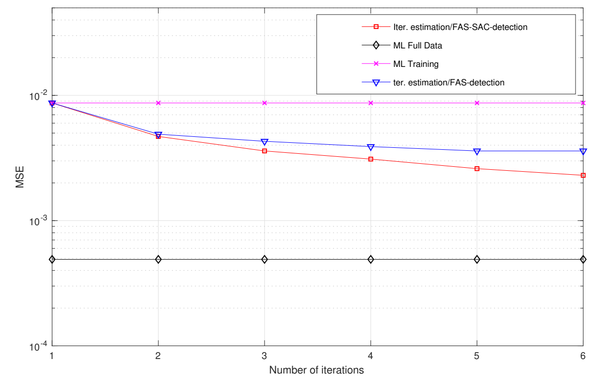

Fig. 9: MSE versus number of iterations with uncoded QPSK, $N = 8$, $n = 64$, $T_p = 16$ and $T = 512$, $SNR = 8$ dB.

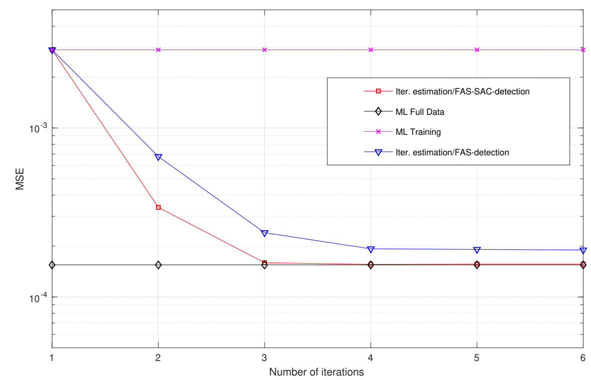

Fig. 10: MSE versus number of iterations with uncoded QPSK, $n = N = 64$, $T_p = 64$ and $T = 1280$ - $SNR = 8$ dB.

MSE evolution as a function of the iterations with a SNR value equal to 8 dB in Fig. 9 and equal to 13 dB in Fig. 10. For the lowest SNR value, the MSE has converged to a steady state from fourth iteration with FAS, while it keeps on decreasing until the sixth iteration with FAS-SAC. For the highest SNR value, the convergence is achieved after fewer iterations in both cases and the lower bound is even reached with FAS-SAC.

4) Superposition of MSE and BER: We consider the second feeding strategy based on hard decisions. Our purpose is to investigate the impact of inner detection iteration number on the MSE and BER achieved by the proposed scheme in Fig. 11 and Fig. 12 with $N = n = 64$. It can be seen that the MSE performance can be improved for an increased number of iterations between channel estimation and detection for both schemes. It can also be seen that with just two iterations of the channel estimation/detection procedure, we can get a BER close to perfect channel detection BER. We show that the FAS and FAS-SAC based-iterative schemes achieve 10^{-3} BER within 0.5dB and 0.9dB of the perfect channel knowledge respectively.

In Fig. 13 and Fig. 14, we show that the proposed channel estimation/detection schemes are efficient in underdetermined systems where $N = 64$ and $n = 50$. It can be confirmed

	Computational cost
ML training	$\mathcal{O}(nNT_p)$
ML full data	$\mathcal{O}(nNT)$
EM	$\mathcal{O}(\max((T_d + NQ)n^2), ((T + N)nNQ))$
Proposed algorithms (FAS-based)	$\mathcal{O}(T_d Q N^3)$
Proposed algorithms (FAS-SAC-based)	$\mathcal{O}(T_d(2 - Z_\eta)QN^3)$

TABLE II: Computational cost with the interior point (iteration number: Q).

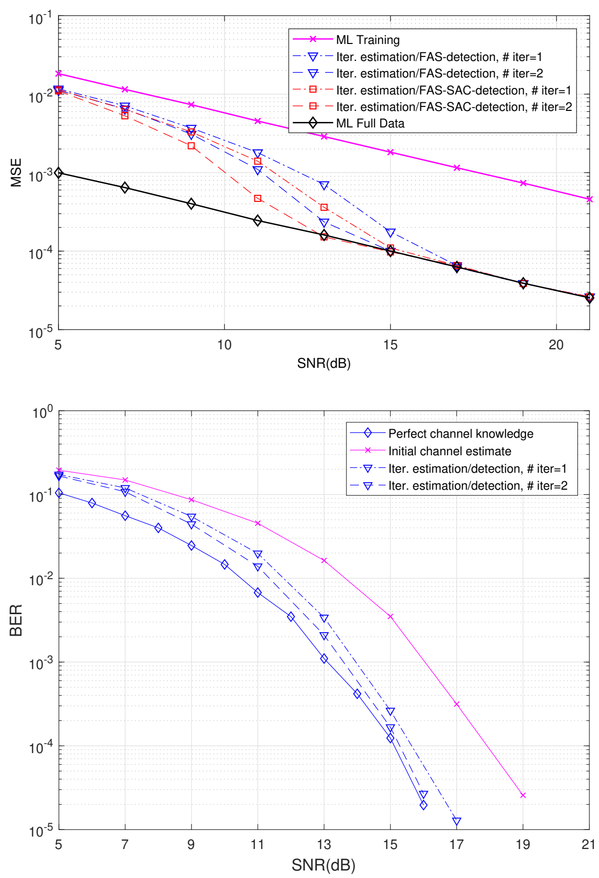

Fig. 11: MSE versus SNR (hard decision FAS and FAS-SAC output-based schemes), BER versus SNR (hard decision FAS output-based schemes) with uncoded QPSK, $n = N = 64$, $T_p = 64$ and $T = 1280$.

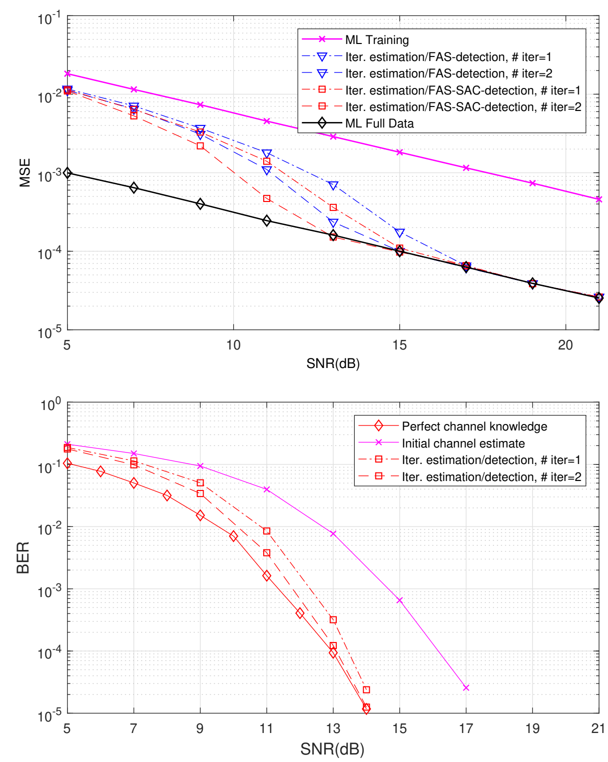

Fig. 12: MSE versus SNR (hard decision FAS and FAS-SAC output-based schemes), BER versus SNR (hard decision FAS-SAC output-based scheme) with uncoded QPSK, $n = N = 64$, $T_p = 64$ and $T = 1280$.

as for the determined system that the MSE performance can be improved increasing the number of iterations. It is the same for the BER where we get a gain of about 2.7dB and 2.8dB over the initial channel estimation detection for FAS and FAS-SAC hard decisions-based channel estimation/detection schemes respectively at the second iteration.

B. Complexity analysis

We now compare the computational complexity of proposed semi-blind channel estimation with the ML estimators and the EM algorithms detailed in Section III. Calculation of ML training-based channel estimation consists of matrix multiplications with dominant factor of $\mathcal{O}(nNT_p)$ and a matrix inversion with complexity $\mathcal{O}(N^3)$. Therefore, the whole complexity order is $\mathcal{O}(nNT_p)$. Similarly we can show that the complexity of full data based channel estimation is about $\mathcal{O}(nNT)$. Let denote Q the number of iterations taken into account in the iterative algorithms. We get that the EM channel estimation algorithm has a complexity of order $\mathcal{O}(\max((T_d + NQ)n^2), ((T + N)nNQ))$. The complexity orders of proposed algorithms based on FAS soft decision output and FAS hard decision output are the same and equal $\mathcal{O}(T_d Q N^3)$. The channel estimation based on FAS SAC with both soft and hard decisions outputs represents a complexity order of $\mathcal{O}(T_d(2 - Z_\eta)QN^3)$. Finally, it can be mentioned that all the proposed iterative algorithms represent the same order of complexity as EM algorithm. The computational complexities of the different algorithms are reported in Table II.

VI. Conclusion

In this paper we have addressed the problem of imperfect CSI and we have proposed semi-blind channel estimation algorithms in large MIMO systems with finite alphabets assuming

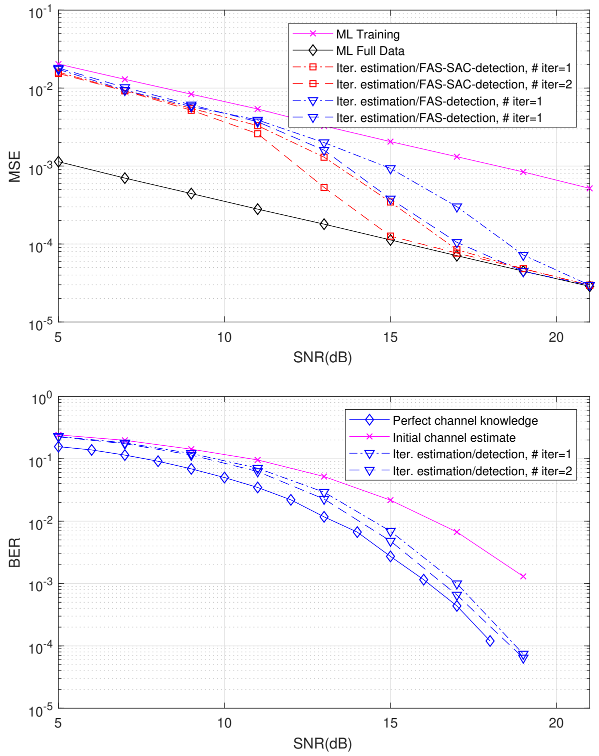

Fig. 13: MSE versus SNR (hard decision FAS and FAS-SAC output-based schemes), BER versus SNR (hard decision FAS output-based scheme) with uncoded QPSK, $n = 64$, $N = 50$, $T_p = 64$ and $T = 1280$.

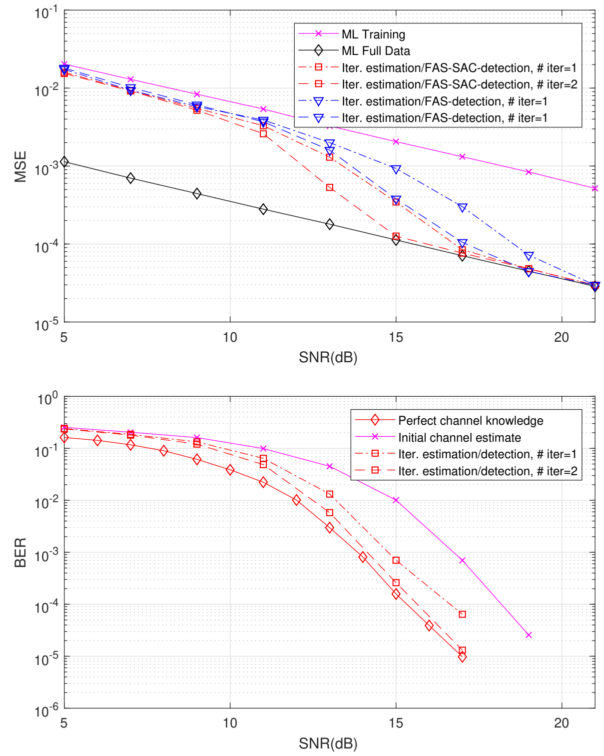

Fig. 14: MSE versus SNR (hard decision FAS and FAS-SAC output-based schemes), BER versus SNR (hard decision FAS-SAC output-based scheme) with uncoded QPSK, $n = 64$, $N = 50$, $T_p = 64$ and $T = 1280$.

limited pilot sequence length. We have proposed channel estimation schemes based on soft and hard decisions outputs of FAS and FAS-SAC algorithms. We have shown that taking into account a number of pilot sequences equal to the number of users is sufficient. Theoretical studies for both algorithms are established and we have determined the CRBs when soft decisions are considered and the asymptotic MSEs when hard decisions are used. Simulation results showed their validity and the efficiency of the proposed schemes which perform close to the ML full-data lower bound, with a superiority of the ones based on hard decisions. The whole work is done assuming non-selective channels. Future works will focus on the spectral efficiency by optimizing pilot sequences number to avoid pilot contamination and how the proposed schemes can be extended to frequency-selective channels and to higher order modulation.

APPENDIX

Proof of Theorem III.1*: The log likelihood function of the received signal is given by:*

$$\mathcal{L} = \text{const} - \sum_{t=1}^{T} \frac{1}{\sigma^2} \left\| \underline{\boldsymbol{y}}(t) - \underline{\boldsymbol{H}}\, \underline{\boldsymbol{x}}(t) \right\|_2^2 \tag{28}$$

The CRB of both channel coefficients and unknown data symbols is computed as:

$$\mathrm{CRB}(\boldsymbol{X}_d, \boldsymbol{H}) = \left(\mathbb{E}[\mathcal{J}\mathcal{J}^T] \right)^{-1}, \tag{29}$$

where

$$\mathcal{J} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{u}} \tag{30}$$

with $\boldsymbol{u} = \left(\underline{\boldsymbol{x}}^T(T_p+1), ..., \underline{\boldsymbol{x}}^T(T), \underline{\boldsymbol{h}}_1^T, ..., \underline{\boldsymbol{h}}_{2N}^T \right)$

Following similar steps as in the proof of [33], for $t = T_p + 1, \ldots, T$ *and* $k = 1, \ldots, 2N$, *we can show that*

$$\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{x}}(t)} = \frac{2}{\sigma^2} \underline{\boldsymbol{H}}^T \underline{\boldsymbol{z}}(t) \tag{31}$$

$$\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} = \frac{2}{\sigma^2} \sum_{t=1}^{T} \underline{\boldsymbol{z}}(t) \underline{x}_k(t) \tag{32}$$

where $\underline{\boldsymbol{h}}_k$ *is the* k*-th column of the matrix* $\underline{\boldsymbol{H}}$ *and* $\underline{\boldsymbol{z}}(t)$ *is the noise vector at time* t.
We get then

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{x}}(q)} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{x}}(p)} \right)^T \right] = \frac{2}{\sigma^2} \underline{\boldsymbol{H}}^T \underline{\boldsymbol{H}}\, \delta(q-p) \tag{33}$$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{x}}(q)} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} \right)^T \right] = \frac{2}{\sigma^2} \underline{\boldsymbol{H}}^T \underline{x}_k(q) \tag{34}$$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_i} \right)^T \right] = \frac{2}{\sigma^2} \sum_{t=1}^{T} \underline{x}_k(t) \underline{x}_i(t) \tag{35}$$

Substituting (33), (34) and (35) in (29), we get

$$\mathrm{CRB}(\boldsymbol{X}_d, \boldsymbol{H})^{-1} = \begin{pmatrix} \frac{2\underline{\boldsymbol{H}}^T\underline{\boldsymbol{H}}}{\sigma^2} & \dots & 0 & \boldsymbol{\Omega}_{(T_p+1)} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \frac{2\underline{\boldsymbol{H}}^T\underline{\boldsymbol{H}}}{\sigma^2} & \boldsymbol{\Omega}_{(T)} \\ \boldsymbol{\Omega}^T_{(T_p+1)} & \dots & \boldsymbol{\Omega}^T_{(T)} & \boldsymbol{\Pi} \end{pmatrix} \quad (36)$$

The CRB of the channel matrix $\boldsymbol{H}$

$$\mathrm{CRB}(\boldsymbol{H}) = \left(\boldsymbol{\Pi} - \sum_{t=T_p+1}^{T} \boldsymbol{\Omega}_t^T \left(\underline{\boldsymbol{H}}^T \underline{\boldsymbol{H}} \right) \boldsymbol{\Omega}_t \right)^{-1} \quad (37)$$

■

***Proof of Theorem IV.1**:*

Following similar steps as in the proof of [11], [33], for $t = T_p + 1, \dots, T$ *and* $k = 1, \dots, 2N$ *applied to the FAS algorithm, we get*

- *for* $k \in \bar{\Lambda}_t$

$$\frac{\partial \mathcal{L}}{\partial \underline{x}_k(t)} = \frac{2}{\sigma^2} \underline{\boldsymbol{h}}_k^T \underline{\boldsymbol{z}}(t) \quad (38)$$

- *for* $k \in \Lambda_t$

$$\frac{\partial \mathcal{L}}{\partial \underline{x}_k(t)} = 0 \quad (39)$$

- *for* $k \in \{1, \dots, 2N\}$

$$\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} = \frac{2}{\sigma^2} \sum_{t=1}^{T} \underline{\boldsymbol{z}}(t) \underline{x}_k(t) \quad (40)$$

We then get

- *for* $(k', k'') \in (\bar{\Lambda}_{t'} \times \bar{\Lambda}_{t''})$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{x}_{k'}(t')} \left(\frac{\partial \mathcal{L}}{\partial \underline{x}_{k''}(t'')} \right)^T \right] = \frac{2}{\sigma^2} \underline{\boldsymbol{h}}_{k'}^T \underline{\boldsymbol{h}}_{k''} \delta_{t',t''} \quad (41)$$

- *for* $(k', k'') \in (\Lambda_{t'} \times \Lambda_{t''})$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{x}_{k'}(t')} \left(\frac{\partial \mathcal{L}}{\partial \underline{x}_{k''}(t'')} \right)^T \right] = 0 \quad (42)$$

so

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{x}}(t')} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{x}}(t'')} \right)^T \right] = \frac{2}{\sigma^2} \left(\boldsymbol{G}_{t'}^{\mathrm{FAS}} \right)^T \boldsymbol{G}_{t'}^{\mathrm{FAS}} \delta_{t',t''} \quad (43)$$

- *for* $k \in \bar{\Lambda}_t$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{x}_k(t')} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} \right)^T \right] = \frac{2}{\sigma^2} \underline{\boldsymbol{h}}_k^T \underline{x}_k(t') \quad (44)$$

- *for* $k \in \Lambda_t$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{x}_k(t')} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} \right)^T \right] = 0 \quad (45)$$

so

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{x}}(t')} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} \right)^T \right] = \frac{2}{\sigma^2} \left(\boldsymbol{G}_{t'}^{\mathrm{FAS}} \right)^T \underline{x}_k(t') \quad (46)$$

- *for* $k \in \{1, \dots, 2N\}$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_i} \right)^T \right] = \frac{2}{\sigma^2} \sum_{t=1}^{T} \underline{x}_k(t) \underline{x}_i(t) \quad (47)$$

Similarly, considering the different sets defined above, we get with the FAS-SAC algorithm the results below:

- *for* $k \in \bar{\mathcal{A}}$

$$\frac{\partial \mathcal{L}}{\partial \underline{x}_k(t)} = \frac{2}{\sigma^2} \underline{\boldsymbol{h}}_k^T \underline{\boldsymbol{z}}(t) \quad (48)$$

- *for* $k \in \mathcal{A}$

$$\frac{\partial \mathcal{L}}{\partial \underline{x}_k(t)} = 0 \quad (49)$$

- *for* $k \in \{1, \dots, 2N\}/\mathcal{A}_e$

$$\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} = \frac{2}{\sigma^2_{\tilde{z}}(\eta)} \sum_{t=1}^{T} \underline{\tilde{\boldsymbol{z}}}(t) \underline{x}_k(t) \quad (50)$$

- *for* $k \in \mathcal{A}_e$

$$\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} = \frac{2}{\sigma^2_{\tilde{z}}(\eta)} \sum_{t=1}^{T} \underline{\tilde{\boldsymbol{z}}}(t) (\underline{x}_k(t) + \delta \underline{x}_k(t)) \quad (51)$$

Note that $\delta \underline{x}_k(t)$ *takes values in* $\{-2, 2\}$.
We then obtain

- *for* $(k', k'') \in (\bar{\mathcal{A}}_{t'} \times \bar{\mathcal{A}}_{t''})$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{x}_{k'}(t')} \left(\frac{\partial \mathcal{L}}{\partial \underline{x}_{k''}(t'')} \right)^T \right] = \frac{2}{\sigma^2} \underline{\boldsymbol{h}}_{k'}^T \underline{\boldsymbol{h}}_{k''} \delta_{t',t''} \quad (52)$$

- *for* $(k', k'') \in (\mathcal{A}_{t'} \times \mathcal{A}_{t''})$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{x}_{k'}(t')} \left(\frac{\partial \mathcal{L}}{\partial \underline{x}_{k''}(t'')} \right)^T \right] = 0 \quad (53)$$

- *for* $k \in \bar{\mathcal{A}}_t$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{x}_k(t')} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} \right)^T \right] = \frac{2}{\sigma^2} \underline{\boldsymbol{h}}_k^T \underline{x}_k(t') \quad (54)$$

- *for* $k \in \mathcal{A}_t$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{x}_k(t')} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} \right)^T \right] = 0 \quad (55)$$

- *for* $(k, i) \in (\bar{\mathcal{A}}_e)^2$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_i} \right)^T \right] = \frac{2}{\sigma^2_{\tilde{z}}(\eta)} \sum_{t=1}^{T} \underline{x}_k(t) \, \underline{x}_i(t) \quad (56)$$

- *for* $(k, i) \in (\bar{\mathcal{A}}_e \times \mathcal{A}_e)$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k} \left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_i} \right)^T \right] = \frac{2}{\sigma^2_{\tilde{z}}(\eta)} \sum_{t=1}^{T} \underline{x}_k(t) \, \underline{u}_i(t) \quad (57)$$

with $\underline{u}_i(t) = \underline{x}_i(t) + \delta \underline{x}_i(t)$.

- *for* $(k,i) \in (\mathcal{A}_e \times \bar{\mathcal{A}}_e)$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k}\left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_i}\right)^T\right] = \frac{2}{\sigma^2_{\tilde{z}}(\eta)}\sum_{t=1}^{T}\underline{u}_k(t)\,\underline{x}_i(t) \tag{58}$$

with $\underline{u}_k(t) = \underline{x}_k(t) + \delta\underline{x}_k(t)$.

- *for* $(k,i) \in (\mathcal{A}_e)^2$

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_k}\left(\frac{\partial \mathcal{L}}{\partial \underline{\boldsymbol{h}}_i}\right)^T\right] = \frac{2}{\sigma^2_{\tilde{z}}(\eta)}\sum_{t=1}^{T}\underline{u}_k(t)\,\underline{u}_i(t) \tag{59}$$

*Following the last steps of [**Proof of Theorem III.1**], we get the final theorem result.* ■

Proof of Theorem IV.2: *The real-valued matrix version of the channel estimate can be written as follows:*

$$\begin{aligned}\underline{\hat{\boldsymbol{H}}}^{\mathrm{HD}} &= \left(\underline{\boldsymbol{Y}}\underline{\tilde{\boldsymbol{X}}}^T\right)\left(\underline{\tilde{\boldsymbol{X}}}\underline{\tilde{\boldsymbol{X}}}^T\right)^{-1} \\ &= \frac{1}{T}(\underline{\boldsymbol{H}}\underline{\boldsymbol{X}} + \underline{\boldsymbol{Z}})(\underline{\boldsymbol{X}} + \underline{\Delta_{\boldsymbol{X}}})^T \\ &= \frac{1}{T}(\underline{\boldsymbol{H}}\underline{\boldsymbol{X}}\underline{\boldsymbol{X}}^T + \underline{\boldsymbol{H}}\underline{\boldsymbol{X}}\underline{\Delta_{\boldsymbol{X}}}^T + \underline{\boldsymbol{Z}}\underline{\boldsymbol{X}}^T + \underline{\boldsymbol{Z}}\underline{\Delta_{\boldsymbol{X}}}^T)\end{aligned} \tag{60}$$

Assuming that the frame size is very large (i.e $2N << T$*), the covariance matrix of the frame block can be approximated by* $\underline{\boldsymbol{X}}\underline{\boldsymbol{X}}^T \approx T\boldsymbol{I}_{2N}$. *The real-valued matrix version of the channel error can then be written as:*

$$\begin{aligned}\underline{\Delta\boldsymbol{H}} &= \underline{\hat{\boldsymbol{H}}}^{HD} - \underline{\boldsymbol{H}} \\ &= \frac{1}{T}(\underline{\boldsymbol{Z}}\underline{\boldsymbol{X}}^T + \underline{\boldsymbol{Z}}\underline{\Delta_{\boldsymbol{X}}}^T + \underline{\boldsymbol{H}}\underline{\boldsymbol{X}}\underline{\Delta_{\boldsymbol{X}}}^T)\end{aligned} \tag{61}$$

Assuming that the channel noise is independent of the transmitted symbols and symbols errors we get:

$$\mathbb{E}\left[\underline{\Delta\boldsymbol{H}}\right] = \frac{1}{T}(\mathbb{E}\left[\underline{\boldsymbol{H}}\right]\mathbb{E}\left[\underline{\boldsymbol{X}}\underline{\Delta_{\boldsymbol{X}}}^T\right]) \tag{62}$$

As the channel is a zero-mean random matrix, the estimation bias is always zero.
The MSE of the channel estimation reads:

$$\begin{aligned}\mathbb{E}\left[\left\|\boldsymbol{H} - \hat{\boldsymbol{H}}^{HD}\right\|_F^2\right] &= \mathbb{E}\left[\mathrm{tr}(\underline{\boldsymbol{H}} - \underline{\hat{\boldsymbol{H}}}^{\mathrm{HD}})(\underline{\boldsymbol{H}} - \underline{\hat{\boldsymbol{H}}}^{\mathrm{HD}})^T\right] \\ &= \frac{1}{T^2}\mathbb{E}\left[\mathrm{tr}\left(\boldsymbol{E}\boldsymbol{E}^T\right)\right]\end{aligned} \tag{63}$$

where $\boldsymbol{E} = \underline{\boldsymbol{Z}}\underline{\boldsymbol{X}}^T + \underline{\boldsymbol{Z}}\underline{\Delta_{\boldsymbol{X}}}^T + \underline{\boldsymbol{H}}\underline{\boldsymbol{X}}\underline{\Delta_{\boldsymbol{X}}}^T$.
Let us now calculate the error terms of (63). The first error term equals:

$$\begin{aligned}\mathbb{E}\left[\mathrm{tr}\left(\left(\underline{\boldsymbol{Z}}\underline{\boldsymbol{X}}^T\right)\left(\underline{\boldsymbol{Z}}\underline{\boldsymbol{X}}^T\right)^T\right)\right] &= T\sigma^2\,\mathrm{tr}(\boldsymbol{I}_{2N}) \\ &= 2NT\sigma^2\end{aligned} \tag{64}$$

The second term is calculated as:

$$\begin{aligned}\mathbb{E}\left[\mathrm{tr}\left((\underline{\boldsymbol{Z}}\underline{\Delta_{\boldsymbol{X}}}^T)\left(\underline{\boldsymbol{Z}}\underline{\boldsymbol{X}}^T\right)^T\right)\right] &= \sigma^2\mathrm{tr}\left(\mathbb{E}\left[(\underline{\Delta\boldsymbol{X}})^T\underline{\boldsymbol{X}}\right]\right) \\ &= -4\sigma^2 NF,\end{aligned} \tag{65}$$

where F *is the number of errors in the detected symbols block* $\hat{\boldsymbol{X}}$.

The third term which considers the noise, the data block and the hard decisions errors can be calculated as follows:

$$\begin{aligned}\mathbb{E}\left[\mathrm{tr}\left(\left(\underline{\boldsymbol{Z}}\underline{\boldsymbol{X}}^T\right)(\underline{\boldsymbol{Z}}\underline{\Delta_{\boldsymbol{X}}}^T)^T\right)\right] &= \sigma^2\mathrm{tr}\left(\mathbb{E}\left[\underline{\boldsymbol{X}}^T(\underline{\Delta_{\boldsymbol{X}}})\right]\right) \\ &= -4\sigma^2 NF\end{aligned} \tag{66}$$

Another term is computed as:

$$\begin{aligned}\mathbb{E}\left[\mathrm{tr}\left((\underline{\boldsymbol{Z}}\underline{\Delta_{\boldsymbol{X}}}^T)(\underline{\boldsymbol{Z}}\underline{\Delta_{\boldsymbol{X}}}^T)^T\right)\right] &= \sigma^2\mathrm{tr}\left(\mathbb{E}\left[\underline{\Delta_{\boldsymbol{X}}}^T\underline{\Delta_{\boldsymbol{X}}}\right]\right) \\ &= 8\sigma^2 NF\end{aligned} \tag{67}$$

We also get,

$$\mathrm{tr}\left(\mathbb{E}\left[\underline{\Delta_{\boldsymbol{X}}}^T\underline{\boldsymbol{X}}\right] + \mathbb{E}\left[\underline{\boldsymbol{X}}^T(\underline{\Delta_{\boldsymbol{X}}})\right] + \mathbb{E}\left[\underline{\Delta_{\boldsymbol{X}}}^T\underline{\Delta_{\boldsymbol{X}}}\right]\right) = 0. \tag{68}$$

Last term in (63) is computed as:

$$\begin{aligned}&\mathbb{E}\left[\mathrm{tr}\left((\underline{\boldsymbol{H}}\underline{\boldsymbol{X}}\underline{\Delta_{\boldsymbol{X}}}^T)(\underline{\boldsymbol{H}}\underline{\boldsymbol{X}}\underline{\Delta_{\boldsymbol{X}}}^T)^T\right)\right] \\ &= \sigma^2\mathrm{tr}\left(\mathbb{E}\left[\underline{\Delta_{\boldsymbol{X}}}^T\underline{\boldsymbol{X}}\underline{\boldsymbol{H}}^T\underline{\boldsymbol{H}}\underline{\boldsymbol{X}}^T\underline{\Delta_{\boldsymbol{X}}}\right]\right).\end{aligned} \tag{69}$$

We finally get (27). ■

References

[1] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *77th IEEE Vehicular Technology Conference (VTC Spring)*, June 2013, pp. 1–5.

[2] D. Ciuonzo, P. S. Rossi, and S. Dey, "Massive MIMO channel-aware decision fusion," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 604–619, 2015.

[3] I. Dey, D. Ciuonzo, and P. S. Rossi, "Wideband collaborative spectrum sensing using massive MIMO decision fusion," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5246–5260, 2020.

[4] T. Marzetta, "Massive mimo: An introduction," *Bell Labs technical Journal*, vol. 20, pp. 11–22, March 2015.

[5] T. Moon and W. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prenctice Hall Inc., 1999.

[6] L. Berriche, K. Abed-Meraim, and J. Belfiore, "Investigation of the channel estimation error on MIMO system performance," in *13th European Signal Processing Conference*, September 2005, pp. 1–4.

[7] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Transactions on Signal Processing*, vol. 50, no. 12, pp. 3055–3069, December 2002.

[8] L. Berriche, K. Abed-Meraim, and J. C. Belfiore, "Cramer-Rao bounds for MIMO channel estimation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, May 2004, pp. 397–400.

[9] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[10] A. Sayed, *Adaptive Filters*, ser. IEEE Press. John Wiley & Sons Inc., 2008.

[11] E. Nayebi and B. D. Rao, "Semi-blind channel estimation for multiuser massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 540–553, January 2018.

[12] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[13] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 453–466, March 2005.

[14] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Asilomar Conference on Signals, Systems and Computers*, October 2006, pp. 359–363.

[15] C. Shin, R. W. Heath, and E. J. Powers, "Blind channel estimation for MIMO-OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 2, pp. 670–685, March 2007.

[16] Y.-S. Chen and C.-A. Lin, "Blind channel identification for MIMO single carrier zero padding block transmission systems," in *7th IEEE Workshop on Signal Processing Advances in Wireless Communications*, July 2006, pp. 1–5.

[17] Y.-S. Chen, "Semiblind channel estimation for MIMO single carrier with frequency-domain equalization systems," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 1, pp. 53–62, January 2010.

[18] A. Ladaycia, A. Belouchrani, K. Abed-Meraim, and A. Mokraoui, "EM-based semi-blind MIMO-OFDM channel estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 3899–3903.

[19] O. Rekik, A. Ladaycia, K. Abed-Meraim, and A. Mokraoui, "Performance bounds analysis for semi-blind channel estimation with pilot contamination in massive MIMO-OFDM systems," in *26th European Signal Processing Conference (EUSIPCO)*, September 2018, pp. 1267–1271.

[20] C. H. Aldana, E. de Carvalho, and J. M. Cioffi, "Channel estimation for multicarrier multiple input single output systems using the EM algorithm," *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3280–3292, December 2003.

[21] X. Wautelet, C. Herzet, A. Dejonghe, J. Louveaux, and L. Vandendorpe, "Comparison of EM-based algorithms for MIMO channel estimation," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 216–226, January 2007.

[22] E. De Carvalho and D. T. M. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, April 1997, pp. 129–132.

[23] E. de Carvalho and D. T. M. Slock, "Asymptotic performance of ML methods for semi-blind channel estimation," in *Asilomar Conference on Signals, Systems and Computers (Cat. No.97CB36136)*, vol. 2, November 1997, pp. 1624–1628.

[24] M. Abuthinien, S. Chen, and L. Hanzo, "Semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems," *IEEE Signal Processing Letters*, vol. 15, pp. 202–205, January 2008.

[25] Z. Hajji, A. Aïssa-El-Bey, and K. Amis, "Simplicity-based recovery of finite-alphabet signals for large-scale MIMO systems," *Digital Signal Processing*, vol. 80, pp. 70–82, September 2018.

[26] Z. Hajji, K. Amis, and A. Aïssa-El-Bey, "Joint channel estimation and simplicity-based detection for large-scale MIMO FEC-coded systems," in *10th IEEE International Symposium on Turbo Codes Iterative Information Processing (ISTC)*, December 2018, pp. 1–5.

[27] ——, "Iterative receivers for large-scale MIMO systems with finite-alphabet simplicity-based detection," *IEEE Access*, vol. 8, pp. 21 742–21 758, January 2020.

[28] D. L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Discrete & Computational Geometry*, vol. 43, no. 3, pp. 522–541, 2010.

[29] Y. Fadlallah, A. Aïssa-El-Bey, K. Amis, D. Pastor, and R. Pyndiah, "New decoding strategy for underdetermined MIMO transmission using sparse decomposition," in *21st European Signal Processing Conference (EUSIPCO)*, September 2013, pp. 1–5.

[30] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, 1994.

[31] P. Li, R. C. de Lamare, and R. Fa, "Multiple feedback successive interference cancellation detection for multiuser MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2434–2439, August 2011.

[32] P. Li and R. C. D. Lamare, "Adaptive decision-feedback detection with constellation constraints for MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 853–859, February 2012.

[33] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.