IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 27, 2020, accepted February 17, 2020, date of publication February 27, 2020, date of current version March 17, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.2976553*

# Performance Analysis of a SAGE-Based Semi-Blind Channel Estimator for Pilot Contaminated MU Massive MIMO Systems

**KHUSHBOO MAWATWAL[1], (Student Member, IEEE), DEBARATI SEN[2], (Senior Member, IEEE), AND RAJARSHI ROY[1], (Senior Member, IEEE)**
[1]Electronics and Electrical Communication Engineering (E&ECE) Department, IIT Kharagpur, Kharagpur 721302, India
[2]G.S. Sanyal School of Telecommunications (GSSST), IIT Kharagpur, Kharagpur 721302, India

Corresponding author: Khushboo Mawatwal (khushbumawatwal@gmail.com)

**ABSTRACT** Massive multiple-input multiple-output (M-MIMO) is one of the key ingredients in the upcoming 5G technology. The benefits associated with M-MIMO rely largely on the accuracy of channel state information (CSI) available at the base station (BS). The existing literature mostly employs pilot-aided schemes which require additional pilots for improving their CSI accuracy; additionally, pilot contamination degrades their performance to a large extent. In this paper, we design a Space-Alternating Generalized Expectation-Maximization (SAGE) based semi-blind estimator for pilot contaminated multi-user (MU) M-MIMO systems. It utilizes both pilot and a few data symbols for CSI estimation through two stages, namely initialization and iterative SAGE (ISAGE). We obtain an initial channel estimate with the help of a pilot-aided linear minimum mean squared error (LMMSE) estimator in the initialization stage. The acquired initial estimate from the former stage is then iteratively updated by SAGE algorithm with the joint usage of pilot and a few data symbols in ISAGE stage. The inclusion of data information in ISAGE stages' estimation process aids in simultaneous improvement of CSI accuracy and spectral efficiency (SE) of a M-MIMO system; which is unlikely for a pilot based estimation scheme. Through simulations, we show that our estimator obtains a considerable improvement over the existing pilot-aided schemes in terms of mean squared error (MSE), bit error rate (BER), SE, and energy efficiency (EE) at a nominal increase in complexity. Besides, on average, it achieves convergence in almost two iterations. We also derive modified Cramer-Rao lower bound (MCRLB) to validate the estimation efficacy of our estimator. We evaluate a closed-form expression for lower bound on UL achievable rate of MU M-MIMO systems under both perfect and imperfect CSI scenario. We also discuss the trade-off between SE and EE.

**INDEX TERMS** Massive MIMO, semi-blind estimator, SAGE, modified CRLB, pilot contamination.

## I. INTRODUCTION

Massive MIMO (M-MIMO) is an integral part of the upcoming 5G system [1], [2]. It is a large scale version of regular MIMO [3] which obtains the benefits such as increased data rate, energy efficiency (EE), interference reduction on a massive scale [4]. M-MIMO prefers time-division duplexing (TDD) over the frequency-division duplexing (FDD) for the communication between base station (BS) and user terminals (UTs) [5]. The channel reciprocity assumption of TDD mode reduces the load of CSI acquisition at both ends but escalates the importance of channel estimation. In TDD mode, BS uses uplink (UL) training session to obtain the channel state information (CSI) and uses it for beam-forming, power allocation, and pre-coding operation during downlink (DL) transmission session. The UL training session transmits orthogonal pilot sequences for CSI evaluation. *However, the availability of a finite coherence interval puts a limit on the pilot length. In view of this, the same set of pilot sequences are used across all the cells, which causes severe inter-cell interference during the UL training session. This phenomenon is known as pilot contamination [5] which makes CSI acquisition a challenging task for the practical realization of multi-user (MU) M-MIMO [2], [4].*

Design of a channel estimator for MU M-MIMO is an interesting area for the research community, and is

The associate editor coordinating the review of this manuscript and approving it for publication was Zhen Gao.

well addressed in [6]–[20]. The propagation channel of M-MIMO can be modeled by two popular approach, namely, rich-scattered Rayleigh channel model [14] and sparsely-scattered geometric channel model [15]–[17]. Moreover, these prevailing channel models are associated with the use-cases of M-MIMO systems in Sub 6 GHz and millimeter wave (mmWave) bands [21], [22]. The measurements at Sub 6 GHz band advocates Rayleigh based channel model, whereas the mmWave band uses a geometric channel model. Since our work concentrates on the favorable propagation condition of the Sub 6 GHz band, we will discuss the literature concerning Sub 6 GHz M-MIMO systems. The existing pilot based channel estimation schemes [6], [7], [9], [11]–[13] improve CSI accuracy at the cost of increased pilot length which further reduces system spectral efficiency (SE). In addition, the availability of limited coherence interval and pilot contamination degrades the performance of pilot based schemes to a large extent. These shortcomings have laid the need of semi-blind [8], [14], [23]–[26] and blind [10] algorithms for CSI acquisition in massive MIMO. A pilot-aided Bayesian channel estimator is discussed in [7], [11]. Authors of [8] have proposed a data-aided estimator that uses pilot, some *a priori* data information, and partially decoded data for CSI acquisition. However, the system model uses a simplified scenario of one user per cell. A least squared semi-blind estimator is presented in [24], which sequentially detects UL data of different users in the target cell and formulates a constrained minimization problem for each user. However, it suffers from high computational complexity and requires large UL data in the estimation process. An expectation-maximization (EM) based semi-blind estimator is proposed in [26] without considering the effect of pilot contamination in CSI acquisition. An eigenvalue-decomposition (EVD) based blind method is reported in [10], which uses the asymptotic orthogonal property of channels in large antenna systems. But, it suffers from high computational complexity. Recently, a semi-blind scheme is presented in [14] with relaxed pilot contamination condition. From the literature survey, we notice that the estimation schemes of [7]–[9], [11], [12], [14], [26] assume full knowledge of the interfering cells' channel covariance matrix. However, this underlying assumption is impractical and involves huge overhead [6]. In our work, we exclude this assumption.

We propose an iterative space-alternating generalized expectation-maximization (SAGE) based semi-blind estimator for pilot contaminated MU M-MIMO systems. The SAGE algorithm is well known for its simple implementation and faster convergence [27], and are reported for conventional MIMO systems in [28], [29]. However, they do not address the multi-cell approach and the pilot contamination issue, which restricts their direct application on M-MIMO systems. Additionally, the definition of complete and incomplete data space in [28], [29] is different from our approach. Therefore, the implementation of classical SAGE algorithm [27] for CSI acquisition process in our work is distinct from the methods of [28], [29]. In this context, our earlier SAGE algorithm [30] also lacks the consideration of intra-cell user interference in its system model and is not applicable to a realistic environment. This proposed scheme utilizes few data symbols along with the training sequences in CSI acquisition. The inclusion of data symbols in the estimation process simultaneously improves the CSI accuracy and the SE, unlike pilot-aided methods. Also, the impact of pilot contamination on CSI accuracy decreases with the increased data length usage in our algorithm, which is not possible in pilot-aided estimators. Our semi-blind scheme comprises of two stages: a) Stage-I (Initialization stage)- We obtain the initial channel estimate with the help of pilot based estimation scheme, b) Stage-II (Iterative SAGE (ISAGE) stage)- It iteratively updates the initial CSI estimate of stage-I through the SAGE algorithm. To our knowledge, this is the first attempt to design a SAGE-based semi-blind channel estimator in pilot contaminated MU M-MIMO systems. The key novelty of our proposed scheme over the existing one is the non-requirement of full knowledge of interfering cells' channel covariance matrix or their users' large-scale fading coefficients. Hence, it removes the unnecessary overheads involved with the coordination from the adjacent neighboring cells. We obtain an estimate for the sum of interfering cells' large scale fading coefficients from the received signals and use it in our estimation process. Furthermore, we derive modified Cramer-Rao lower bound (MCRLB) [31], [32] to determine the estimation efficacy of the proposed scheme. We also evaluate a theoretical bound on UL achievable rate, SE, energy efficiency (EE) and computational complexity of the proposed estimator. We evaluate the performance of our estimator in terms of mean-squared error (MSE), bit-error rate (BER), SE, EE, and complexity order. We compare the results with existing schemes of [6], [7], [14], [24], [25]. Major contributions of our estimator are outlined as:

- An iterative SAGE based semi-blind algorithm is designed for pilot contaminated MU M-MIMO systems, which simultaneously improves the CSI accuracy and SE with minimum pilot length [33]. It obtains considerable improvement in terms of MSE, BER, SE and EE from the existing schemes [6], [7]. Furthermore, it reduces the effect of pilot contamination by increasing the data length used in the estimation process instead of introducing additional pilots in the system.
- It obviates the need of a priori knowledge of interfering cells' large-scale fading coefficients during the estimation process, and thus removes huge overhead involved with it. It evaluates an estimate for the sum of the interfering cell users' large-scale fading coefficients from the received symbols. This simple technique adds negligible complexity to the system.
- We derive a MCRLB on the MSE of semi-blind estimators in M-MIMO systems. Through simulations, we note that our scheme approaches the MCRLB even at lower signal-to-noise ratios (SNRs) with large BS antennas and data length used in the iteration stage (i.e., $N_d'$).

- We obtain a theoretical lower bound on the UL achievable rate of M-MIMO systems for maximal ratio combining (MRC) and zero-forcing (ZF) receivers under perfect and imperfect CSI. The derived lower bounds are tight for both the receivers, and CSI availability with the BS. The designed semi-blind scheme approaches the SE performance of perfect CSI for large $N_d'$ and/or low SNR. Furthermore, we evaluate a closed form expression for the SE and EE of the semi-blind estimator. We also discuss the trade-off between EE and SE for both perfect and imperfect CSI through numerical and analytical results.
- We calculate an upper bound on the computational complexity of the semi-blind algorithm in terms of Big O notation. The complexity order of our scheme is $N_i N_d'$ times higher than the pilot-aided estimator [7], where $N_i$ denotes the number of iterations needed for convergence. From extensive simulations, we notice that our scheme requires around two iterations i.e., $N_i$=2 for attaining convergence in the given scenarios which makes the complexity comparable with the existing pilot-aided methods. In addition, we provide theoretical proof on the convergence of our algorithm.

The paper is organized into nine sections: Sec. II provides the system model for pilot contaminated MU M-MIMO. Sec. III and IV describes the proposed semi-blind estimator and theoretical lower bound on its UL achievable rate, respectively. Sec. V discusses the trade-off between SE and EE, whereas MCRLB calculation is given in Sec. VI. The convergence proof of our scheme is provided in Sec. VII. The simulation results are shown in Sec. VIII and conclusions are drawn in Sec IX.

*Notations:* Matrices and vectors are shown by capital bold and small bold letters respectively. $Re\{.\}$, $Im\{.\}$, $(.)^T$, $(.)^H$, $(.)^*$ and $\mathbb{E}\{.\}$ denote real, imaginary, transpose, Hermitian, conjugate, and expectation operation respectively. $\mathbb{C}^{a\times b}$ and $\mathbb{R}^{a\times b}$, represents complex valued, and real valued matrix, respectively of size $a \times b$. $tr\{\mathbf{A}\}$, $\mathbf{0}_{m\times n}$, and $\mathbf{I}_n$ denotes trace of matrix $\mathbf{A}$, zero matrix of size $m \times n$, and identity matrix of size $n$ respectively. $\mathcal{CN}(\mu, \sigma^2)$ is the complex Gaussian distribution with mean $\mu$ and variance $\sigma^2$; $\mathcal{O}$ denotes Big-O notation for computational complexity.

## II. SYSTEM MODEL

We consider an MU M-MIMO system, which contains $L$ time-synchronized cells sharing the same time-frequency-pilot resources. Each cell of the system consists of one BS with $M$ antennas serving $K$ single-antenna UTs such that $M \gg K$ [7], [8]. We consider an uplink (UL) transmission session, which uses the same set of pilot sequences for all the $L$ cells. The unity pilot reuse factor causes pilot contamination, where CSI of the desired cell gets contaminated from the channels of $L-1$ interfering cells. We assume cell 1 as the desired cell and remaining $L-1$ cells as interfering cells for simple system formulation. An example of UL transmission in seven-cell based MU M-MIMO system is shown in Fig. 1.

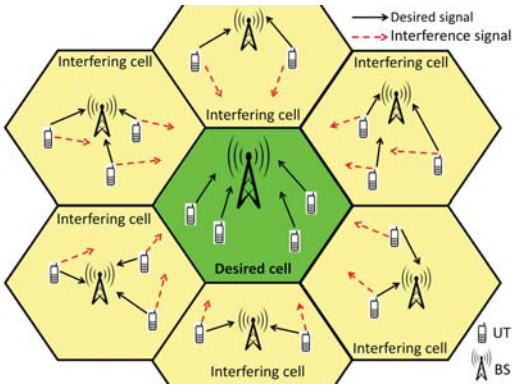


**FIGURE 1. Schematic diagram of a seven-cell ($L = 7$) based pilot contaminated multi-user massive MIMO system.**

Thus, received signal at the BS of cell 1 (BS 1) is given as

$$\mathbf{y} = \sqrt{E_s}\mathbf{G}_1\mathbf{D}_1^{\frac{1}{2}}\mathbf{x}_1 + \sqrt{E_s}\sum_{l=2}^{L}\mathbf{G}_l\mathbf{D}_l^{\frac{1}{2}}\mathbf{x}_l + \mathbf{w}. \quad (1)$$

$\mathbf{G}_l \in \mathbb{C}^{M\times K}$, and $\mathbf{D}_l \in \mathbb{R}^{K\times K}$ are the small-scale and large-scale fading matrices of wireless channel between BS 1 and users of cell $l$ respectively for $l = 1, \ldots, L$. The elements of $\mathbf{G}_l$ are independent and identically distributed $(i.i.d.)$ with $\mathcal{CN}(0, 1)$. Moreover, $\mathbf{G}_l = [\mathbf{g}_{l1}, \ldots, \mathbf{g}_{lK}]$ where $\mathbf{g}_{lk} \in \mathbb{C}^{M\times 1}$ denotes channel vector between $k^{th}$ user of cell $l$ and BS 1 for $k = 1, \ldots, K$. $\mathbf{D}_l = diag\,(\beta_{l1}, \ldots, \beta_{lK})$ is a diagonal matrix; $\beta_{lk}$ denotes the large-scale fading coefficient between $k^{th}$ UT of cell $l$ and BS 1. Each BS has knowledge of its cell users large-scale fading coefficients i.e., $\mathbf{D}_l$ is known to $l^{th}$ cell BS only. In (1), $\mathbf{x}_l$ represents $K \times 1$ transmit vector from $K$ UTs of cell $l$. Further, $\mathbf{w} \in \mathbb{C}^{M\times 1}$ is the received additive noise whose elements are i.i.d. with $\mathcal{CN}(0, N_0)$; $N_0$ denotes noise variance. $E_s$ and $\frac{E_s}{N_0}$ indicate the average transmit power per user and the SNR at the transmitter side, respectively; for clarity we term transmit SNR as SNR.

## III. CHANNEL ESTIMATION IN MASSIVE MIMO SYSTEMS

The UL session comprises of training and data transmission, where all UTs of the $L$ cells simultaneously transmit a transmission frame to their respective BSs. The transmission frame comprises of $N$ $(N = N_p + N_d)$ symbols where $N_p$ pilot symbols are followed by $N_d$ data symbols. We assume the channel to remain constant over each transmission frame and vary independently across the frames i.e., quasi-static. Thus, the received signal at BS 1 is given by

$$\mathbf{Y} = \sqrt{E_s}\mathbf{G}_1\mathbf{D}_1^{\frac{1}{2}}\mathbf{X}_1 + \sqrt{E_s}\sum_{l=2}^{L}\mathbf{G}_l\mathbf{D}_l^{\frac{1}{2}}\mathbf{X}_l + \mathbf{W}. \quad (2)$$

$\mathbf{Y} = \begin{bmatrix}\mathbf{Y}_p & \mathbf{Y}_d\end{bmatrix}$; $\mathbf{Y}_p \in \mathbb{C}^{M\times N_p}$ and $\mathbf{Y}_d \in \mathbb{C}^{M\times N_d}$ denotes pilot and data matrices respectively. Similarly, $\mathbf{X}_l = \begin{bmatrix}\mathbf{X}_{lp} & \mathbf{X}_{ld}\end{bmatrix} \in \mathbb{C}^{K\times N}$ represents transmit matrix from the users of $l^{th}$ cell, where $\mathbf{X}_{lp} \in \mathbb{C}^{K\times N_p}$ and $\mathbf{X}_{ld} \in \mathbb{C}^{K\times N_d}$ are the transmit pilot and data matrices respectively. Besides, we represent

**TABLE 1.** Notations list.

| Symbol | Description |
|---|---|
| $M$ | Number of BS antennas |
| $K$ | Number of UTs per cell |
| $N$ | Number of symbols in a transmission frame |
| $N_p$ | Number of pilot symbols in a transmission frame |
| $N_d$ | Number of data symbols in a transmission frame |
| $N'_d$ | Number of data symbols used in the ISAGE stage of proposed scheme |
| $N_0$ | Noise variance |
| $E_s$ | Transmit power of each user per symbol |
| $\mathbf{Y}$ | Received matrix at cell 1 |
| $\mathbf{Y}'$ | Received matrix used during the CSI estimation process |
| $\mathbf{X}_l$ | Transmitted matrix of cell $l$ |
| $\mathbf{W}$ | Received noise at the BS of cell 1 |
| $\mathbf{Y}_p$ | Received pilot matrix at cell 1 |
| $\mathbf{X}_{lp}$ | Transmitted pilot matrix of cell $l$ |
| $\mathbf{W}_p$ | Received noise at BS 1 corresponding to pilot matrix $\mathbf{Y}_p$ |
| $\mathbf{Y}_d$ | Received data matrix at the BS of cell 1 |
| $\mathbf{Y}'_d$ | Received data matrix of BS 1 used for the CSI acquisition |
| $\mathbf{X}_{ld}$ | Transmitted data matrix of cell $l$ |
| $\mathbf{W}_d$ | Received noise at BS 1 corresponding to data matrix $\mathbf{Y}_d$ |
| $\mathbf{y}_m$ | Received vector at $m^{th}$ BS antenna of cell 1 |
| $\mathbf{y}_{mp}$ | Received pilot vector at $m^{th}$ BS antenna |
| $\mathbf{y}_{md}$ | Received data vector at $m^{th}$ BS antenna |
| $\mathbf{x}_{lk}$ | Transmit vector from $k^{th}$ UT of cell $l$ |
| $\mathbf{x}_{lkp}$ | Transmit pilot vector from $k^{th}$ UT of cell $l$ |
| $\mathbf{x}_{lkd}$ | Transmit data vector of cell $l$s' $k^{th}$ user |
| $\mathbf{G}_l$ | Small-scale fading of channel matrix between the UTs of cell $l$ and the BS of cell 1 |
| $\mathbf{g}_{lk}$ | Small-scale fading channel vector between the $k^{th}$ UT of cell $l$ and BS of cell 1 |
| $\mathbf{D}_l$ | Large-scale fading matrix between the UTs of cell $l$ and the BS of cell 1 |
| $\beta_{lk}$ | Large-scale fading coefficient between the $k^{th}$ UT of cell $l$ and the BS of cell 1 |
| $\mathbf{X}'_{1d}$ | Transmit data matrix of cell 1 needed during the ISAGE stage for the estimation of $\mathbf{G}_1$ |
| $\mathbf{S}$ | Estimate of $\mathbf{X}'_{1d}$ |
| $\hat{\mathbf{G}}_1^{(0)}$ | Initial estimate of $\mathbf{G}_1$ |
| $\hat{\mathbf{G}}_1^{(i)}$ | $i^{th}$ iteration estimate of $\mathbf{G}_1$ |
| $\hat{\mathbf{g}}_{1k}^{(0)}$ | Initial estimate of channel vector $\mathbf{g}_{1k}$ |
| $N_i$ | Number of iterations |

$\mathbf{X}_l = \left[\mathbf{x}_{l1}^H, \ldots, \mathbf{x}_{lK}^H\right]^H$ where $\mathbf{x}_{lk} \in \mathbb{C}^{1\times N}$ denotes transmit vector from $k^{th}$ UT of cell $l$ for $l = 1, \ldots, L$, and $k = 1, \ldots, K$. The $\mathbf{x}_{lk} = \left[\mathbf{x}_{lkp}\ \mathbf{x}_{lkd}\right]$ with transmit pilot $\mathbf{x}_{lkp} \in \mathbb{C}^{1\times N_p}$ and data $\mathbf{x}_{lkd} \in \mathbb{C}^{1\times N_d}$ vectors. In addition, we denote $\mathbf{W} = \left[\mathbf{W}_p\ \mathbf{W}_d\right]$ as the received noise matrix. For simplicity, we consider same transmit power for both pilot and data symbols. The received pilot at $m^{th}$ antenna of BS 1 is given as

$$\mathbf{y}_{mp} = \sqrt{E_s}\mathbf{g}_{1m}\mathbf{D}_1^{\frac{1}{2}}\mathbf{X}_p + \sqrt{E_s}\sum_{l=2}^{L}\mathbf{g}_{lm}\mathbf{D}_l^{\frac{1}{2}}\mathbf{X}_p + \mathbf{w}_{mp} \quad (3)$$

for $m = 1, \ldots, M$. The $\mathbf{g}_{lm} \in C^{1\times K}$ indicates channel vector between the UTs of $l^{th}$ cell and $m^{th}$ antenna of BS 1. Due to pilot contamination, all cells use the same pilot set i.e. $\mathbf{X}_{lp} = \mathbf{X}_p$ for $l = 1, \ldots, L$. Furthermore, the transmit pilot sequences are known at the BS and maintains orthogonality within the cell i.e., $\mathbf{X}_p\mathbf{X}_p^H = N_p\mathbf{I}_K$. The $\mathbf{w}_{mp}$ represents noise corresponding to received pilot $\mathbf{y}_{mp}$. The received data at $m^{th}$ antenna of BS 1 is

$$\mathbf{y}_{md} = \sqrt{E_s}\mathbf{g}_{1m}\mathbf{D}_1^{\frac{1}{2}}\mathbf{X}_{1d} + \sqrt{E_s}\sum_{l=2}^{L}\mathbf{g}_{lm}\mathbf{D}_l^{\frac{1}{2}}\mathbf{X}_{ld} + \mathbf{w}_{md}. \quad (4)$$

The transmit data matrix $\mathbf{X}_{ld}$ are unknown to their BSs and are independent of each other i.e. $\mathbb{E}\left\{\mathbf{X}_{l_1d}\mathbf{X}_{l_2d}^H\right\} = N_d\delta(l_1 - l_2)\mathbf{I}_K$ where $\delta(l_1 - l_2) = 1$ for $l_1 = l_2$ otherwise zero; $l_1, l_2 \in \{1, \ldots, L\}$. Moreover, $\mathbb{E}\left\{\mathbf{x}_{l_1id}\mathbf{x}_{l_2jd}^H\right\} = N_d\delta(l_1 - l_2, i - j)$ where $\delta(l_1 - l_2, i - j) = 1$ for $i = j$ and $l_1 = l_2$ otherwise zero; $i, j \in \{1, \ldots, K\}$. We consider $\mathbb{E}\{x_{lkd}(n)\} = 0$, and $\mathbb{E}\left\{|\, x_{lkd}(n)\,|^2\right\} = 1$ for $n = 1, \ldots, N_d$. In addition, $\mathbf{Y}_p = \left[\mathbf{y}_{1p}^T \ldots \mathbf{y}_{Mp}^T\right]^T$, $\mathbf{Y}_d = \left[\mathbf{y}_{1d}^T \ldots \mathbf{y}_{Md}^T\right]^T$, $\mathbf{y}_{md} = [y_{md}(1) \ldots y_{md}(N_d)]$, $\mathbf{X}_{ld} = [\mathbf{x}_{ld}(1) \ldots \mathbf{x}_{ld}(N_d)]$, and $\mathbf{x}_{ld}(n) = [x_{l1d}(n) \ldots x_{lKd}(n)]^T \in \mathbb{C}^{K\times 1}$. The $\mathbf{w}_{md}$ represents noise corresponding to received and data vector $\mathbf{y}_{md}$.

### A. IMPACT OF PILOT CONTAMINATION

The phenomenon of pilot contamination arises due to unity pilot reuse factor and limited coherence interval, which acts as a bottleneck for the massive MIMO. The effect of pilot contamination is shown with the help of following example.

For ease of exposition, we assume single user per cell (i.e., $K = 1$) in a $L$-cell system (3). Furthermore, we employ

a simple pilot-aided least squared (LS) estimator [7] for CSI acquisition at BS 1 as

$$\hat{\mathbf{g}}_1^{(LS)} = \frac{\mathbf{Y}_p \mathbf{x}_p^H}{\sqrt{E_s \beta_1} N_p} = \underbrace{\mathbf{g}_1}_{\text{desired cell}} + \underbrace{\sum_{l=2}^{L} \sqrt{\frac{\beta_l}{\beta_1}} \mathbf{g}_l}_{\text{Interfering cells}} + \frac{\mathbf{W}_p \mathbf{x}_p^H}{\sqrt{E_s \beta_1} N_p}. \quad (5)$$

Due to pilot contamination, we note that the estimated channel vector of cell 1 i.e., $\hat{\mathbf{g}}_1^{(LS)}$ in (5) contains the CSI of $L-1$ interfering cells. Moreover, the impact of inter-cell interference in CSI estimate $\hat{\mathbf{g}}_1^{(LS)}$ does not diminish with the increase in BS antennas ($M$). This phenomenon is observed with every pilot based estimators. Thus, pilot contamination limits the performance of massive MIMO systems to a large extent.

The above constraint restricts the usage of training based schemes and advocates the necessity of semi-blind approach in channel estimation process of pilot contaminated MU M-MIMO systems. In view of this, we present our semi-blind channel estimator in the following subsection.

### B. PROPOSED ITERATIVE SAGE BASED SEMI-BLIND CHANNEL ESTIMATOR

The designed semi-blind estimator uses known pilot as well as unknown data symbols for CSI acquisition. It is an iteration based scheme where the number of iterations ($N_i$) are kept fixed to the minimum value needed to achieve convergence. It consist of two-stages: Initialization and Iteration stage. The former stage utilizes a pilot based scheme to obtain the initial CSI estimate which is then passed to the ISAGE stage. We employ SAGE algorithm in the ISAGE stage to iteratively improve the initial CSI estimate accuracy with the help of $N_d'$ data symbols such that $N_d' < N_d$. The ISAGE stage is also known as semi-blind stage. The initialization stage uses $\mathbf{Y}_p$ whereas ISAGE stage requires $\mathbf{Y}' = \left[\mathbf{Y}_p \mathbf{Y}_d'\right] \in \mathbb{C}^{M \times N'}$ for channel estimation; $N' = N_p + N_d'$.

$$\mathbf{Y}_p = \sum_{k=1}^{K} \sqrt{E_s \beta_{1k}} \mathbf{g}_{1k} \mathbf{x}_{kp} + \sqrt{E_s} \sum_{l=2}^{L} \mathbf{G}_l \mathbf{D}_l^{\frac{1}{2}} \mathbf{X}_p + \mathbf{W}_p. \quad (6)$$

$$\mathbf{Y}_d' = \sqrt{E_s} \mathbf{G}_l \mathbf{D}_l^{\frac{1}{2}} \mathbf{X}_{1d}' + \sqrt{E_s} \sum_{l=2}^{L} \mathbf{G}_l \mathbf{D}_l^{\frac{1}{2}} \mathbf{X}_{ld}' + \mathbf{W}_d', \quad (7)$$

where $\mathbf{Y}_d' = \left[\mathbf{y}_d(1), \ldots, \mathbf{y}_d(N_d')\right] \in \mathbb{C}^{M \times N_d'}$ such that $\mathbf{y}_d(n) = [y_{1d}(n), \ldots, y_{Md}(n)]^T$ for $n = 1, \ldots, N_d'$. Moreover, the received vector for CSI acquisition at $m^{th}$ antenna of BS 1 is given by $\mathbf{y}_m' = \left[\mathbf{y}_{mp} \mathbf{y}_{md}'\right]$ where $\mathbf{y}_{md}' = \left[y_{md}(1), \ldots, y_{md}(N_d')\right]$. Similarly, $\mathbf{W}_d' = \left[\mathbf{w}_d(1), \ldots, \mathbf{w}_d(N_d')\right]$ where $\mathbf{w}_d(n) \in C^{M \times 1}$ represents noise vector corresponding to data symbols at $n^{th}$ time instant for $n = 1, \ldots, N_d'$.

#### 1) INITIALIZATION STAGE

We use a pilot-aided linear minimum-mean squared error (LMMSE) estimation algorithm for obtaining the initial estimate of channel $\mathbf{G}_1$ (i.e., $\hat{\mathbf{G}}_1^{(0)}$) from received pilot matrix $\mathbf{Y}_p$. Thus, the initial estimate of $\mathbf{g}_{1k}$ is given as

$$\hat{\mathbf{g}}_{1k}^{(0)} = \sqrt{E_s \beta_{1k}} \left( \sum_{l=1}^{L} E_s \beta_{lk} + \frac{N_0}{N_p} \right)^{-1} \frac{\mathbf{Y}_p \mathbf{x}_{kp}^H}{N_p} \quad (8)$$

for $k = 1, \ldots, K$. The inverse term in (8) demands the perfect knowledge of interfering cell's large-scale fading coefficients with BS 1, however, this assumption is impractical and incurs huge overhead. In order to avoid the unnecessary complexity, we derive an estimate of the inverse term in (8) with the help of $\mathbf{Y}_p$. We define

$$\mathbf{q}_k = \frac{\mathbf{Y}_p \mathbf{x}_{kp}^H}{N_p} = \sqrt{E_s \beta_{1k}} \mathbf{g}_{1k} + \sum_{l=2}^{M} \sqrt{E_s \beta_{lk}} \mathbf{g}_{lk} + \frac{\mathbf{W}_p \mathbf{x}_{kp}^H}{N_p} \quad (9)$$

as a $M \times 1$ random vector with $\mathbb{E}\{\mathbf{q}_k\} = \mathbf{0}_{M \times 1}$ and $\mathbb{E}\left\{\mathbf{q}_k \mathbf{q}_k^H\right\} = \left(E_s \beta_{1k} + \sum_{l=2}^{L} E_s \beta_{lk} + \frac{N_0}{N_p}\right) \mathbf{I}_M$. By utilizing the law of large numbers on $\mathbf{q}_k$, we obtain

$$\frac{\mathbf{q}_k^H \mathbf{q}_k}{M} \to E_s \beta_{1k} + \sum_{l=2}^{L} E_s \beta_{lk} + \frac{N_0}{N_p} \quad (10)$$

for $M \to \infty$. By replacing the inverse term of (8) with (10), the initial estimate of $\mathbf{g}_{1k}$ results in

$$\hat{\mathbf{g}}_{1k}^{(0)} = \sqrt{E_s \beta_{1k}} \left( \frac{\mathbf{q}_k^H \mathbf{q}_k}{M} \right)^{-1} \mathbf{q}_k. \quad (11)$$

#### 2) ITERATION STAGE

It uses a SAGE based semi-blind estimator to iteratively update the initial CSI estimate $\hat{\mathbf{G}}_1^{(0)}$. For CSI acquisition, we denote $\left\{\mathbf{Y}'\right\}$ and $\left\{\mathbf{Y}', \mathbf{X}_1'\right\}$ as incomplete and complete data space, respectively; $\mathbf{Y}' = \left[\mathbf{y}_1'^T, \ldots, \mathbf{y}_M'^T\right]^T$ and $\mathbf{X}_1' = \left[\mathbf{X}_p \; \mathbf{X}_{1d}'\right] \in \mathbb{C}^{K \times N'}$.

The SAGE algorithm [27] uses hidden data space instead of complete data space to obtain the maximum likelihood estimator (MLE) through iterations. It divides the estimation parameters in groups such that if one group is being updated the remaining groups remain fixed to their current revised values. This aids in faster convergence as compared to other iteration based schemes such as [8]. For channel estimation, $\mathbf{g}_{1m} \in \mathbb{C}^{1 \times K}$ denotes the parameter under consideration and rest of the channel vectors are combined under $\mathbf{g}_{1\tilde{m}} \in \mathbb{C}^{(M-1) \times K}$ where $m \neq \tilde{m}$ and $m, \tilde{m} \in \{1, \ldots, M\}$. From (3) and (4), we note that channel vector $\mathbf{g}_{1m}$ depends solely on $\mathbf{y}_m'$. Thus, we denote $\left\{\mathbf{y}_m', \mathbf{X}_1'\right\}$ as its hidden data space.

$$\mathbf{y}_m' = \sqrt{E_s} \mathbf{g}_{1m} \mathbf{D}_1^{\frac{1}{2}} \mathbf{X}_1' + \sqrt{E_s} \sum_{l=2}^{L} \mathbf{g}_{lm} \mathbf{D}_l^{\frac{1}{2}} \mathbf{X}_l' + \mathbf{w}_m'. \quad (12)$$

Since $\mathbf{X}'_{1d} = \left[\mathbf{x}_{1d}(1), \ldots, \mathbf{x}_{1d}(N'_d)\right] \in \mathbb{C}^{K\times N'_d}$ is unknown at BS 1, we obtain an estimate of $\mathbf{X}'_{1d}$ from $\mathbf{Y}'_d$ to replace it in the ISAGE stage. We employ LMMSE receiver which is a sufficient statistic [32] of $\mathbf{X}'_{1d}$ for its estimate

$$\mathbf{S} = \left\{\mathbf{I}_K + \left(\sum_{l=2}^{L}\sum_{k=1}^{K}\beta_{lk} + \frac{N_0}{E_s}\right) \left(\mathbf{D}_1^{-\frac{1}{2}}\left(\mathbf{G}_1^H\mathbf{G}_1\right)^{-1}\mathbf{D}_1^{-\frac{1}{2}}\right)\right\}^{-1} \sqrt{\frac{1}{E_s}}\mathbf{D}_1^{-\frac{1}{2}}\left(\mathbf{G}_1^H\mathbf{G}_1\right)^{-1}\mathbf{G}_1^H\mathbf{Y}'_d; \tag{13}$$

the inverse term contains large scale fading coefficients of the $L-1$ interfering cells, which are not available with BS 1. Therefore, we derive an estimate of the inverse term in (13) with the help of

$$\mathbf{Z} = \mathbf{D}_1^{-\frac{1}{2}}\left(\mathbf{G}_1^H\mathbf{G}_1\right)^{-1}\mathbf{G}_1^H\mathbf{Y}'_d \in \mathbb{C}^{K\times N'_d}. \tag{14}$$

$\mathbf{Z} = \left[\mathbf{z}(1), \ldots, \mathbf{z}(n), \ldots, \mathbf{z}(N'_d)\right]$ is a random matrix where $\mathbf{z}(n) \in \mathbb{C}^{K\times 1}$ has $\mathbb{E}\{\mathbf{z}(n)\} = \mathbf{0}_{K\times 1}$ and $\mathbb{E}\left\{\mathbf{z}(n)\mathbf{z}^H(n)\right\} = E_s\mathbf{I}_K + \left(E_s\sum_{l=2}^{L}\sum_{k=1}^{K}\beta_{lk} + N_0\right)\mathbf{D}_1^{-\frac{1}{2}}\left(\mathbf{G}_1^H\mathbf{G}_1\right)^{-1}\mathbf{D}_1^{-\frac{1}{2}}$ for $n = 1, \ldots, N'_d$. By using the law of large numbers on $\mathbf{Z}$ for $N'_d \to \infty$, we obtain

$$\frac{\mathbf{Z}\mathbf{Z}^H}{N'_d} \to \mathbf{I}_K + \left(\sum_{l=2}^{L}\sum_{k=1}^{K}\beta_{lk} + \frac{N_0}{E_s}\right)\left(\mathbf{D}_1^{-\frac{1}{2}}\left(\mathbf{G}_1^H\mathbf{G}_1\right)^{-1}\mathbf{D}_1^{-\frac{1}{2}}\right). \tag{15}$$

Thus, the resultant estimate of $\mathbf{X}'_{1d}$ in (13) reduces to

$$\mathbf{S} = \sqrt{E_s}N'_d\left(\mathbf{Z}\mathbf{Z}^H\right)^{-1}\mathbf{Z}. \tag{16}$$

Due to the non-availability of $\mathbf{X}'_{1d}$ with BS 1, we will replace all the instances of $\mathbf{X}'_{1d}$ in the ISAGE stage with its estimate $\mathbf{S}$ of (16).

In this stage, we represent $\hat{\mathbf{G}}_1^{(i)}$ and $\hat{\mathbf{g}}_{1m}^{(i)}$ as the $i^{th}$ iteration estimate of $\mathbf{G}_1$ and $\mathbf{g}_{1m}$ respectively where $i = 0, 1, \ldots, N_i$. Moreover, each iteration of the SAGE algorithm undergoes expectation and maximization steps to update the initial CSI estimate $\hat{\mathbf{G}}_1^{(0)}$.

***Expectation step:*** It evaluates conditional expectation of the hidden data space log-likelihood function (LLF) given $\mathbf{g}_{1m}$ over the incomplete data space and the current estimate of $\mathbf{G}_1$ as

$$\begin{aligned} Q_m\left(\mathbf{g}_{1m}|\hat{\mathbf{G}}_1^{(i)}\right) &= \mathbb{E}\left\{\ln f\left(\mathbf{y}'_m, \mathbf{X}'_1|\mathbf{g}_{1m}, \hat{\mathbf{g}}_{1\bar{m}}^{(i)}\right)|\mathbf{Y}', \hat{\mathbf{G}}_1^{(i)}\right\} \\ &= \mathbb{E}\left\{\ln f\left(\mathbf{y}_{mp}|\mathbf{X}_p, \mathbf{g}_{1m}\right)|\mathbf{Y}_p, \hat{\mathbf{G}}_1^{(i)}\right\} \\ &\quad + \mathbb{E}\left\{\ln f\left(\mathbf{y}'_{md}|\mathbf{X}'_{1d}, \mathbf{g}_{1m}\right)|\mathbf{Y}'_d, \hat{\mathbf{G}}_1^{(i)}\right\}, \end{aligned} \tag{17}$$

where $\ln f\left(\mathbf{y}_{mp}|\mathbf{X}_p, \mathbf{g}_{1m}\right)$ and $\ln f\left(\mathbf{y}'_{md}|\mathbf{X}'_{1d}, \mathbf{g}_{1m}\right)$ denotes LLF of $\mathbf{g}_{1m}$ given $\mathbf{y}_{mp}$ and $\mathbf{y}'_{md}$ respectively. For LLF, we associate the inter-cell interference with the received noise to form a random variable which follows complex Gaussian distribution. This consideration holds true because: *a*) Large-scale fading coefficients of interfering cells remain constant over various transmission frames, *b*) Channel coefficients of interfering cells follow $\mathcal{CN}(0, 1)$, *c*) Transmit data symbols of the interfering cell users have zero mean and unit average power, *d*) The elements of received noise follow $\mathcal{CN}(0, N_0)$.

From (3), we obtain the LLF of $\mathbf{g}_{1m}$ given $\mathbf{y}_{mp}$ as

$$\ln f\left(\mathbf{y}_{mp}|\mathbf{X}_p, \mathbf{g}_{1m}\right) = \left(\mathbf{y}_{mp} - \sqrt{E_s}\mathbf{g}_{1m}\mathbf{D}_1^{\frac{1}{2}}\mathbf{X}_p\right)\mathbf{C}_{v_{mp}}^{-1} \left(\mathbf{y}_{mp} - \sqrt{E_s}\mathbf{g}_{1m}\mathbf{D}_1^{\frac{1}{2}}\mathbf{X}_p\right)^H; \tag{18}$$

$\mathbf{C}_{v_{mp}}$ is the covariance matrix of $\mathbf{v}_{mp}$ where $\mathbf{v}_{mp} = \sum_{l=2}^{L}\sqrt{E_s}\mathbf{g}_{lm}\mathbf{D}_l^{\frac{1}{2}}\mathbf{X}_p + \mathbf{w}_{mp}$. Here, $\mathbf{w}_{mp}$ represents received noise vector at $m^{th}$ antenna of BS 1 with $\mathbf{w}_{mp} \sim \mathcal{CN}(\mathbf{0}_{1\times N_p}, N_0\mathbf{I}_{N_p})$. The elements of $\mathbf{v}_{mp}$ have zero mean as $g_{lmk} \sim \mathcal{CN}(0, 1)$ and $w_{mp}(n) \sim \mathcal{CN}(0, N_0)$ for $l = 2, \ldots, L$, $k = 1, \ldots, K$, $n = 1 \ldots, N_p$. Furthermore, its covariance matrix is given by

$$\mathbf{C}_{v_{mp}} = E_s\sum_{l=2}^{L}\mathbf{X}_p^H\mathbf{D}_l\mathbf{X}_p + N_0\mathbf{I}_{N_p} = \mathbf{C}_p. \tag{19}$$

Similarly, we obtain the LLF of $\mathbf{g}_{1m}$ given $\mathbf{y}'_{md}$ from (4) as

$$\ln f\left(\mathbf{y}'_{md}|\mathbf{X}'_{1d}, \mathbf{g}_{1m}\right) = \left(\mathbf{y}'_{md} - \sqrt{E_s}\mathbf{g}_{1m}\mathbf{D}_1^{\frac{1}{2}}\mathbf{X}'_{1d}\right)\mathbf{C}_{v_{md}}^{-1} \left(\mathbf{y}'_{md} - \sqrt{E_s}\mathbf{g}_{1m}\mathbf{D}_1^{\frac{1}{2}}\mathbf{X}'_{1d}\right)^H. \tag{20}$$

Here, $\mathbf{C}_{v_{md}}$ is the covariance matrix of $\mathbf{v}_{md}$ where $\mathbf{v}_{md} = \sum_{l=2}^{L}\sqrt{E_s}\mathbf{g}_{lm}\mathbf{D}_l^{\frac{1}{2}}\mathbf{X}'_{ld} + \mathbf{w}'_{md} \in \mathbb{C}^{1\times N'_d}$. We indicate the received noise at $m^{th}$ antenna of BS 1 by $\mathbf{w}'_{md} \sim \mathcal{CN}(\mathbf{0}_{1\times N'_d}, N_0\mathbf{I}_{N'_d})$. The elements of $\mathbf{v}_{md}$ have zero mean as both $\mathbf{w}_{md}$ and $\mathbf{g}_{lm}$ are zero mean vectors. Moreover, its covariance matrix is given by

$$\mathbf{C}_{v_{md}} = \left(E_s\sum_{l=2}^{L}\sum_{k=1}^{K}\beta_{lk} + N_0\right)\mathbf{I}_{N'_d} = c_d\mathbf{I}_{N'_d}. \tag{21}$$

Therefore, by replacing the LLFs of (17) with their respective expressions, we obtain the resultant $Q_m\left(\mathbf{g}_{1m}|\hat{\mathbf{G}}_1^{(i)}\right)$ for the next step.

*Maximization Step:* In this step, we update the $i^{th}$ iteration estimate of $\mathbf{g}_{1m}$ to its' $i+1^{th}$ iteration revised value i.e., $\hat{\mathbf{g}}_{1m}^{(i+1)}$, by maximizing the expectation step.

$$\hat{\mathbf{g}}_{1m}^{(i+1)} = \underset{\mathbf{g}_{1m}}{\arg\max}\ Q_m\left(\mathbf{g}_{1m}|\hat{\mathbf{G}}_1^{(i)}\right) \tag{22}$$

We evaluate $\hat{\mathbf{g}}_{1m}^{(i+1)}$ by differentiating $Q_m\left(\mathbf{g}_{1m}|\hat{\mathbf{G}}_1^{(i)}\right)$ with respect to $\mathbf{g}_{1m}^H$ and equating the resultant term to zero.

Therefore,

$$\hat{\mathbf{g}}_{1m}^{(i+1)} = \sqrt{\frac{1}{E_s}} \left( \mathbf{y}_{mp} \mathbf{C}_p^{-1} \mathbf{X}_p^H + \frac{\mathbf{y}'_{md} \boldsymbol{\eta}_1^{(i)H}}{c_d} \right) \left( \mathbf{X}_p \mathbf{C}_p^{-1} \mathbf{X}_p^H + \frac{\boldsymbol{\eta}_2^{(i)}}{c_d} \right)^{-1} \mathbf{D}_1^{-\frac{1}{2}}; \quad (23)$$

$\boldsymbol{\eta}_1^{(i)} = \mathbb{E}\left\{ \mathbf{X}'_{1d} | \mathbf{Y}'_d, \hat{\mathbf{G}}_1^{(i)} \right\}$, $\boldsymbol{\eta}_2^{(i)} = \mathbb{E}\left\{ \mathbf{X}'_{1d} \mathbf{X}'^{H}_{1d} | \mathbf{Y}'_d, \hat{\mathbf{G}}_1^{(i)} \right\}$ denotes $i^{th}$ iteration *a posteriori* expectations of $\mathbf{X}'_{1d}$ and $\mathbf{X}'_{1d}\mathbf{X}'^{H}_{1d}$, respectively. Due to the unavailability of $\mathbf{X}'_{1d}$ with BS 1, we substitute it with $\mathbf{S}$ of (16). Further, we replace channel $\mathbf{G}_1$ with $\hat{\mathbf{G}}_1^{(i)}$ in (16) to obtain the $i^{th}$ iteration of the above-mentioned *a posteriori* expectations.

We provide pseudo code for the semi-blind estimator in Algorithm 1.

**Algorithm 1** Pseudo code of SAGE based semi-blind scheme

1: **Initialization stage:**
2: **for** $k = 1, 2, \ldots, K$ **do**
3: Calculate $\mathbf{z}_k = \frac{\mathbf{Y}_p \mathbf{x}_{kp}^H}{N_p}$.
4: Obtain the initial estimate of $\mathbf{g}_{1k}$ as
$$\hat{\mathbf{g}}_{1k}^{(0)} = \sqrt{E_s \beta_{1k}} \left( \frac{\mathbf{z}_k^H \mathbf{z}_k}{M} \right)^{-1} \mathbf{z}_k$$
5: **end for**
6: **Iterative SAGE stage:**
7: $N'_d \leftarrow$number of data symbols used in the ISAGE stage
8: **for** $i = 0, 1, 2, \ldots$ **do**
9: **for** $m = 1, \ldots, M$ **do**
10: Calculate $\boldsymbol{\eta}_1^{(i)}$ and $\boldsymbol{\eta}_2^{(i)}$ from $\mathbf{S}$ of (16) by replacing $\mathbf{G}_1$ with $\hat{\mathbf{G}}_1^{(i)}$.
11: Obtain the $i+1^{th}$ iteration estimate of $\mathbf{g}_{1m}$ as
$$\hat{\mathbf{g}}_{1m}^{(i+1)} = \frac{1}{\sqrt{E_s}} \left( \mathbf{y}_{mp} \mathbf{C}_p^{-1} \mathbf{X}_p^H + \frac{\mathbf{y}'_{md} \boldsymbol{\eta}_1^{(i)H}}{c_d} \right) \left( \mathbf{X}_p \mathbf{C}_p^{-1} \mathbf{X}_p^H + \frac{\boldsymbol{\eta}_2^{(i)}}{c_d} \right)^{-1} \mathbf{D}_1^{-\frac{1}{2}}$$
12: Update $\hat{\mathbf{g}}_{1m}^{(i)} = \hat{\mathbf{g}}_{1m}^{(i+1)}$
13: **end for**
14: **end for**

### C. COVARIANCE MATRIX OF THE PROPOSED SEMI-BLIND CHANNEL ESTIMATOR

*Initialization stage:* For covariance matrix of $\hat{\mathbf{g}}_{1k}^{(0)}$, we utilize (8) instead of (11) as both (11) and (8) will asymptotically lead to the same value. Therefore, the covariance matrix of $\hat{\mathbf{g}}_{1k}^{(0)}$ is given by

$$\mathbb{E}\left\{ \hat{\mathbf{g}}_{1k}^{(0)} \hat{\mathbf{g}}_{1k}^{(0)H} \right\} = \frac{\beta_{1k} N_p}{\beta_{1k} N_p + \sum_{l=2}^{L} \beta_{lk} N_p + \frac{N_0}{E_s}} \mathbf{I}_M \quad (24)$$

for $k = 1, \ldots, K$. The elements of $\hat{\mathbf{g}}_{1k}^{(0)}$ has variance

$$\mathbb{E}\left\{ | \hat{g}_{1mk}^{(0)} |^2 \right\} = \frac{\beta_{1k} N_p}{\beta_{1k} N_p + \sum_{l=2}^{L} \beta_{lk} N_p + \frac{N_0}{E_s}} = \sigma_{kp}^2, \quad (25)$$

where $m = 1, \ldots, M$.

In (25), we note that the initial estimate $\hat{\mathbf{G}}_1^{(0)}$ approaches the perfect CSI i.e. $\sigma_{kp}^2 \simeq 1$ for $N_p \gg \sum_{l=2}^{L} \frac{\beta_{lk}}{\beta_{1k}} N_p + \frac{N_0}{E_s \beta_{1k}}$. However, availability of limited coherence interval in wireless environment restricts the increment in pilot length and thus its CSI accuracy. Besides, variance of its estimation error is given by $\sigma_{ekp}^2 = \mathbb{E}\left\{ |\hat{g}_{1mk}^{(0)} - g_{1mk}|^2 \right\} = 1 - \sigma_{kp}^2$.

*ISAGE Stage:* For covariance calculation of the semi-blind stage estimate, we consider $\mathbf{X}'_{1d}$ and $\mathbf{X}'_{1d}\mathbf{X}'^{H}_{1d}$ in place of *a posteriori* expectation $\boldsymbol{\eta}_1^{(i)}$, and $\boldsymbol{\eta}_2^{(i)}$, respectively in (23). This substitution holds true as we use LMMSE receiver for the detection of $\mathbf{X}'_{1d}$ from received matrix $\mathbf{Y}'_d$ in (13), which is a sufficient statistic for $\mathbf{X}'_{1d}$ [32]. We also remove iteration index $i$ from $\hat{\mathbf{g}}_{1m}^{(i+1)}$ for clarity and rewrite (23) in (26), as shown at the bottom of the next page. Thus, its covariance matrix is given by

$$\mathbb{E}\left\{ \hat{\mathbf{g}}_{1m}^H \hat{\mathbf{g}}_{1m} \right\} = \mathbf{I}_K + \left( \mathbf{D}_1^{-\frac{1}{2}} \right) \left( \frac{\mathbf{X}_p \mathbf{C}_p^{-1} \mathbf{X}_p^H}{E_s} + \frac{\mathbf{X}'_{1d} \mathbf{X}'^{H}_{1d}}{E_s c_d} \right)^{-1} \mathbf{D}_1^{-\frac{1}{2}}, \quad (27)$$

where the second term on right side of the equality is a non-diagonal matrix. However, for large number of interfering users i.e. $(L-1)K \gg 1$, the diagonal elements of $\mathbf{C}_p$ in (27) becomes significantly large in comparison to its off-diagonal terms such that

$$\mathbf{C}_p \simeq \left( \sum_{l=2}^{L} \sum_{i=1}^{K} E_s \beta_{li} + N_0 \right) \mathbf{I}_{N_p} = c_p \mathbf{I}_{N_p}. \quad (28)$$

Similarly, with large $N'_d$ we obtain

$$\mathbf{X}'_{1d} \mathbf{X}'^{H}_{1d} \simeq N'_d \mathbf{I}_K. \quad (29)$$

Therefore, from (28) and (29) the covariance matrix of $\hat{\mathbf{g}}_{1m}$ (27) will asymptotically result in a diagonally dominant matrix

$$\mathbb{E}\left\{ \hat{\mathbf{g}}_{1m}^H \hat{\mathbf{g}}_{1m} \right\} \simeq \mathbf{I}_K + \mathbf{D}_1^{-\frac{1}{2}} \left( \frac{\mathbf{X}_p \mathbf{X}_p^H}{E_s c_p} + \frac{N'_d \mathbf{I}_K}{E_s c_d} \right)^{-1} \mathbf{D}_1^{-\frac{1}{2}} \simeq \mathbf{I}_K + \frac{c}{E_s \left( N_p + N'_d \right)} \mathbf{D}_1^{-1}; \quad (30)$$

where $c = \sum_{l=2}^{L} \sum_{i=1}^{K} E_s \beta_{li} + N_0$ as $c_p = c_d$. Furthermore, variance of the semi-blind stage estimator is

$$\mathbb{E}\left\{ | \hat{g}_{1mk} |^2 \right\} = 1 + \frac{c}{E_s \beta_{1k} \left( N_p + N'_d \right)} = \sigma_{kd}^2 \quad (31)$$

for $m = 1, \ldots, M$ and $k = 1, \ldots, K$.

From (31), we observe that for large $N'_d$, variance of the ISAGE stage estimator will converge to perfect CSI i.e., $\sigma^2_{kd} \simeq 1$. Moreover, its estimation accuracy improves with the increase in $N'_d$ without adding extra pilots in the system unlike pilot-aided estimators. Also, the impact of pilot contamination reduces with the increase in $N'_d$. We denote $\sigma^2_{ekd} = \mathbb{E}\left\{|\hat{g}_{1mk} - g_{1mk}|^2\right\} = \sigma^2_{kd} - 1$ as variance of its estimation error.

### D. COMPLEXITY ANALYSIS

We characterize complexity of an algorithm by the number of complex additions and multiplications involved in it. We use Big O notation to define an upper bound on the increase in computational complexity.

**Complexity of pilot-aided scheme [7]:**

$$C_{pilot} = \mathcal{O}\left(MKN_p\right). \tag{32}$$

**Complexity of the proposed semi-blind estimator:**

$$\begin{aligned} C_{pro} &= \mathcal{O}\left(MKN_p + MK^2N_iN' + MKN_iN_p^2\right) \\ &\simeq \mathcal{O}\left(MK^2N_iN'\right). \end{aligned} \tag{33}$$

The proposed scheme compises of two stages: intialization and iteration stage; the former stage uses only pilots for CSI acquisition, whereas the later stage uses both pilot and data symbols. Therefore, the complexity order of our estimator comprises of $N_p$ and $N'$ terms in the expression (33). Since $N' = N_p + N'_d$, we notice that the complexity order of (33) is largely dependent on $N_i$ and $N'_d$. Moreover, for $N'_d > N_p$, the computational complexity $C_{pro}$ is upper bounded by $\mathcal{O}\left(MK^2N_iN'\right)$ which is approximately $N_iN'_d$ times more than the complexity order of pilot-aided scheme [7].

## IV. ACHIEVABLE RATE ANALYSIS

In this section, we evaluate a theoretical lower bound on ergodic achievable rate [34] of MU M-MIMO systems under perfect and imperfect CSI. Thus, we detect the transmitted data of cell 1 UTs from the received data at BS 1

$$\mathbf{y_d} = \sqrt{E_s}\mathbf{G}_1\mathbf{D}_1^{\frac{1}{2}}\mathbf{x}_{1d} + \sqrt{E_s}\sum_{l=2}^{L}\mathbf{G}_l\mathbf{D}_l^{\frac{1}{2}}\mathbf{x}_{ld} + \mathbf{w}_d \tag{34}$$

### A. WITH PERFECT CSI

Here, we presume the availability of perfect CSI i.e., $\mathbf{G_1}$ at BS 1. We employ linear detector $\mathbf{A} \in \mathbb{C}^{M\times K}$ to detect transmitted data $\mathbf{x}_{1d}$ from $\mathbf{y}_d$ as

$$\mathbf{r}_d = \mathbf{A}^H\mathbf{y}_d = \sqrt{E_s}\sum_{l=1}^{L}\mathbf{A}^H\mathbf{G}_l\mathbf{D}_l^{\frac{1}{2}}\mathbf{x}_{ld} + \mathbf{A}^H\mathbf{w}_d. \tag{35}$$

The data of $k^{th}$ user is given by

$$\begin{aligned} r_{kd} = \mathbf{a}_k^H\mathbf{y}_d &= \sqrt{E_s\beta_{1k}}\mathbf{a}_k^H\mathbf{g}_{1k}x_{1kd} + \mathbf{a}_k^H\mathbf{w}_d \\ &+ \sum_{\substack{i=1,\\ i\neq k}}^{K}\sqrt{E_s\beta_{1i}}\mathbf{a}_k^H\mathbf{g}_{1i}x_{1id} \\ &+ \sum_{l=2}^{L}\sum_{i=1}^{K}\sqrt{E_s\beta_{li}}\mathbf{a}_k^H\mathbf{g}_{li}x_{lid}, \end{aligned} \tag{36}$$

where $\mathbf{a}_k$ and $\mathbf{g}_{1k}$ are the $k^{th}$ columns of matrix $\mathbf{A}$ and $\mathbf{G}_1$, respectively for $k = 1, \ldots, K$. In (36), the first term corresponds to desired signal and rest of the terms correspond to noise, intra-cell interference, and inter-cell interference, respectively. These noise and interference (from cell 1 and the $L-1$ interfering cells) components are undesirable while recovering data at BS 1; we add them up to form the aggregate noise. For a fixed $\mathbf{G}_1$, we model the noise plus interference term i.e., aggregate noise as worst case Gaussian noise [33], [35] and obtain the UL achievable rate for $k^{th}$ user in (37), as shown at the bottom of the next page. We employ linear receivers (Rxs) such as maximum-ratio combiner (MRC) and zero-forcing (ZF) for data detection.

#### 1) MRC RECEIVER

Here, $\mathbf{A}^H = \mathbf{G}_1^H$ and $\mathbf{a}_k = \mathbf{g}_{1k}$. Thus, the ergodic achievable rate for MRC receiver (Rx) is given in (38), as shown at the bottom of the next page. With the help of Jensen's inequality and convex property of $\log_2\left(1+\frac{1}{x}\right)$, we obtain a lower bound on the UL achievable rate in (39), as shown at the bottom of the next page.

*Proposition 1:* With MRC Rx, the lower bound on UL achievable rate of cell 1 $k^{th}$ user under perfect CSI is

$$\hat{R}^{mrc}_{p,k} = \log_2\left(1 + \frac{M\beta_{1k}}{\sum_{\substack{i=1\\ i\neq k}}^{K}\beta_{1i} + \sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li} + \frac{N_0}{E_s}}\right) \tag{40}$$

$$\begin{aligned} \hat{\mathbf{g}}_{1m} &= \sqrt{\frac{1}{E_s}}\left(\mathbf{y}_{mp}\mathbf{C}_p^{-1}\mathbf{X}_p^H + \frac{\mathbf{y}'_{md}\mathbf{X}'^H_{1d}}{c_d}\right)\left(\mathbf{X}_p\mathbf{C}_p^{-1}\mathbf{X}_p^H + \frac{\mathbf{X}'_{1d}\mathbf{X}'^H_{1d}}{c_d}\right)^{-1}\mathbf{D}_1^{-\frac{1}{2}} \\ &= \mathbf{g}_{1m} + \left(\sum_{l=2}^{L}\mathbf{g}_{lm}\mathbf{D}_l^{\frac{1}{2}}\left(\mathbf{X}_p\mathbf{C}_p^{-1}\mathbf{X}_p^H + \frac{\mathbf{X}_{ld}\mathbf{X}'^H_{1d}}{c_d}\right) + \left(\frac{\mathbf{w}_{mp}\mathbf{C}_p^{-1}\mathbf{X}_p^H}{\sqrt{E_s}} + \frac{\mathbf{w}'_{md}\mathbf{X}'^H_{1d}}{\sqrt{E_s}c_d}\right)\right) \\ &\quad \left(\mathbf{X}_p\mathbf{C}_p^{-1}\mathbf{X}_p^H + \frac{\mathbf{X}'_{1d}\mathbf{X}'^H_{1d}}{c_d}\right)^{-1}\mathbf{D}_1^{-\frac{1}{2}} \end{aligned} \tag{26}$$

*Proof:* See Appendix A.

#### 2) ZF RECEIVER

The ZF Rx is given by $\mathbf{A}^H = \left(\mathbf{G}_1^H \mathbf{G}_1\right)^{-1} \mathbf{G}_1^H$ which satisfies $\mathbf{A}^H \mathbf{G}_1 = \mathbf{I}_K$ i.e., $\mathbf{a}_k^H \mathbf{g}_{1i} = 1$ for $i = k$ otherwise $\mathbf{a}_k^H \mathbf{g}_{1i} = 0$ where $k, i \in \{1, \ldots, K\}$. So, the UL achievable rate of (37) results in

$$R_{p,k}^{zf} = \mathbb{E}\left\{\log_2\left(1+ \frac{\beta_{1k}}{\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}|\mathbf{a}_k^H \mathbf{g}_{li}|^2 + \frac{N_0}{E_s}||\mathbf{a}_k||^2}\right)\right\}. \tag{41}$$

From Jensen's inequality, we obtain a lower bound on (41) as

$$R_{p,k}^{zf} \geq \hat{R}_{p,k}^{zf} = \log_2\left(1 + \frac{\beta_{1k}}{\mathbb{E}\left\{\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}|\mathbf{a}_k^H \mathbf{g}_{li}|^2 + \frac{N_0||\mathbf{a}_k||^2}{E_s}\right\}}\right) \tag{42}$$

*Proposition 2:* For ZF Rx, the UL achievable rate of $k^{th}$ user with perfect CSI is lower bounded by

$$\hat{R}_{p,k}^{zf} = \log_2\left(1 + \frac{\beta_{1k}\,(M-K)}{\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li} + \frac{N_0}{E_s}}\right) \tag{43}$$

*Proof:* See Appendix B.

### B. WITH IMPERFECT CSI

In practical scenario, BS 1 uses an estimator to obtain the CSI. However, the acquired channel $\hat{\mathbf{G}}_1$ is an estimate of true CSI $\mathbf{G}_1$ with estimation error $\mathbf{E}_1 = \hat{\mathbf{G}}_1 - \mathbf{G}_1$. We assume $\mathbf{E}_1$ to be independent of $\hat{\mathbf{G}}_1$ [33]. Thus, BS 1 treats $\hat{\mathbf{G}}_1$ as true CSI for the data recovery of cell 1 UTs with the linear receiver $\hat{\mathbf{A}}$ from received data $\mathbf{y}_d$ (34) as

$$\hat{\mathbf{r}}_d = \hat{\mathbf{A}}^H \mathbf{y}_d = \sqrt{E_s}\hat{\mathbf{A}}^H\left(\hat{\mathbf{G}}_1 - \mathbf{E}_1\right)\mathbf{D}_1^{\frac{1}{2}}\mathbf{x}_{1d} + \sqrt{E_s}\sum_{l=2}^{L}\hat{\mathbf{A}}^H\mathbf{G}_l\mathbf{D}_l^{\frac{1}{2}}\mathbf{x}_{ld} + \hat{\mathbf{A}}^H\mathbf{w}_d. \tag{44}$$

The recovered data of $k^{th}$ user is given by

$$\begin{aligned}\hat{r}_{kd} &= \sqrt{E_s}\hat{\mathbf{a}}_k^H\left(\sqrt{\beta_{1k}}\hat{\mathbf{g}}_{1k}x_{1kd} - \sum_{i=1}^{K}\sqrt{\beta_{1i}}\mathbf{e}_{1i}x_{1id} + \sum_{\substack{i=1\\ i\neq k}}^{K}\sqrt{\beta_{1i}}\hat{\mathbf{g}}_{1i}x_{1id} + \sum_{l=2}^{L}\sum_{i=1}^{K}\sqrt{\beta_{li}}\mathbf{g}_{li}x_{lid}\right) + \hat{\mathbf{a}}_k^H\mathbf{w}_d \\ &= \sqrt{E_s}\hat{\mathbf{a}}_k^H\left(\sqrt{\beta_{1k}}\hat{\mathbf{g}}_{1k}x_{1kd} - \sqrt{\beta_{1k}}\mathbf{e}_{1k}x_{1kd} + \sum_{\substack{i=1\\ i\neq k}}^{K}\sqrt{\beta_{1i}}\mathbf{g}_{1i}x_{1id} + \sum_{l=2}^{L}\sum_{i=1}^{K}\sqrt{\beta_{li}}\mathbf{g}_{li}x_{lid}\right) + \hat{\mathbf{a}}_k^H\mathbf{w}_d;\end{aligned} \tag{45}$$

$\hat{\mathbf{a}}_k$, $\hat{\mathbf{g}}_{1k}$, $\mathbf{g}_{lk}$ and $\mathbf{e}_{1k}$ denote $k^{th}$ column vectors of $\hat{\mathbf{A}}$, $\hat{\mathbf{G}}_1$, $\mathbf{G}_l$, $\mathbf{E}_1$, respectively. Furthermore, the first term on right side of the equality in (45) corresponds to desired component and the last four terms account for estimation error, intra-cell interference, inter-cell interference, and additive noise, respectively. To obtain a lower bound on the achievable rate, we consider the estimation error ($\mathbf{e}_{1k}$) of the channel estimate $\hat{\mathbf{g}}_{1k}$ as a variant of additive noise such that the resulting noise is replaced by a worst case noise (i.e., Gaussian distribution) [33]. For a given $\hat{\mathbf{G}}_1$, we sum up the interference, estimation error and noise terms to form the aggregate noise [33]. Since we consider channel estimate $\hat{\mathbf{G}}_1$ as the true CSI, the ergodic UL achievable rate of $k^{th}$ user under imperfect CSI is given in (46), as shown at the bottom of the next page and the details of its derivation is given in Appendix C. We consider MRC and ZF Rxs for the data recovery of cell 1 UTs in the following subsections:

#### 1) MRC RECEIVER

We have $\hat{\mathbf{A}}^H = \hat{\mathbf{G}}_1^H$ and $\hat{\mathbf{a}}_k = \hat{\mathbf{g}}_{1k}$ for $k = 1, \ldots, K$. So, the UL achievable rate of $k^{th}$ UT with MRC Rx i.e., $R_{im,k}^{mrc}$ is given in (47), as shown at the bottom of the next page. From Jensen's inequality and convex property of $\log\left(1+\frac{1}{x}\right)$

$$R_{p,k} = \mathbb{E}\left\{\log_2\left(1 + \frac{E_s\beta_{1k}|\mathbf{a}_k^H\mathbf{g}_{1k}|^2}{E_s\sum_{i=1,i\neq k}^{K}\beta_{1i}|\mathbf{a}_k^H\mathbf{g}_{1i}|^2 + E_s\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}|\mathbf{a}_k^H\mathbf{g}_{li}|^2 + N_0||\mathbf{a}_k||^2}\right)\right\} \tag{37}$$

$$R_{p,k}^{mrc} = \mathbb{E}\left\{\log_2\left(1 + \left(\frac{E_s\sum_{i=1,i\neq k}^{K}\beta_{1i}|\mathbf{g}_{1k}^H\mathbf{g}_{1i}|^2 + E_s\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}|\mathbf{g}_{1k}^H\mathbf{g}_{li}|^2 + N_0||\mathbf{g}_{1k}||^2}{E_s\beta_{1k}||\mathbf{g}_{1k}||^4}\right)^{-1}\right)\right\} \tag{38}$$

$$R_{p,k}^{mrc} \geq \hat{R}_{p,k}^{mrc} = \log_2\left(1 + \left(\mathbb{E}\left\{\frac{\sum_{i=1,i\neq k}^{K}\beta_{1i}|\mathbf{g}_{1k}^H\mathbf{g}_{1i}|^2 + \sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}|\mathbf{g}_{1k}^H\mathbf{g}_{li}|^2 + \frac{N_0}{E_s}||\mathbf{g}_{1k}||^2}{\beta_{1k}||\mathbf{g}_{1k}||^4}\right\}\right)^{-1}\right) \tag{39}$$

function, we achieve a lower bound on $R^{mrc}_{im,k}$ in (48), as shown at the bottom of this page.

*Proposition 3:* By using MRC Rx, the UL achievable rate of cell 1 $k^{th}$ user under imperfect CSI is lower bounded by

$$\hat{R}^{mrc}_{im,k} = \log_2\left(1+\frac{\beta_{1k}M\sigma^2_{ku}}{\sum_{i=1}^{K}IE_i+\sum_{\substack{i=1\\ i\neq k}}^{K}IG_i+\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}+\frac{N_0}{E_s}}\right); \tag{49}$$

$IE_i = \beta_{1i}\sigma^2_{eiu}$, and $IG_i = \beta_{1i}\sigma^2_{iu}$ for $u \in \{p, d\}$ where $p$, and $d$ stand for pilot-aided, and semi-blind estimators, respectively; $\sigma^2_{eiu}$ and $\sigma^2_{iu}$ represent variance of the estimation error, and CSI estimate of the estimator. For the proposed scheme, $\sigma^2_{ip}$ and $\sigma^2_{id}$ is obtained from (25) and (31), respectively.

*Proof:* Refer Appendix D.

For $E_s \to 0$, the lower bound on achievable UL rate for MRC Rx tends to $\hat{R}^{mrc}_{im,k} \to \log_2\left(1+\frac{M}{N'}\right)$.

#### 2) ZF RECEIVER

It is given by $\hat{\mathbf{A}}^H = \left(\hat{\mathbf{G}}_1^H\hat{\mathbf{G}}_1\right)^{-1}\hat{\mathbf{G}}_1^H$ which satisfies $\hat{\mathbf{A}}^H\hat{\mathbf{G}}_1 = \mathbf{I}_K$ i.e. $\hat{\mathbf{a}}_k^H\hat{\mathbf{g}}_{1i} = 1$ if $k = i$, otherwise zero for $k, i \in \{1, \ldots K\}$. Thus, we denote $R^{zf}_{im,k}$ as the $k^{th}$ user UL achievable rate for ZF Rx and it is given in (50), as shown at the bottom of the next page. From Jensen's inequality, we obtain a lower bound on $R^{zf}_{im,k}$ which is shown in (51), as shown at the bottom of the next page.

*Proposition 4:* With ZF Rx, the UL achievable rate of $k^{th}$ user with imperfect CSI is lower bounded by

$$\hat{R}^{zf}_{im,k} = \log_2\left(1+\frac{(M-K)\,\beta_{1k}\sigma^2_{ku}}{\sum_{i=1}^{K}\beta_{1i}\sigma^2_{eiu}+\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}+\frac{N_0}{E_s}}\right). \tag{52}$$

*Proof:* Refer Appendix E.

For $E_s \to 0$, the lower bound on achievable UL rate for ZF Rx tends to $\hat{R}^{zf}_{im,k} \to \log_2\left(1+\frac{M-K}{N'}\right)$.

## V. SPECTRAL AND ENERGY EFFICIENCY ANALYSIS

In this section, we study the relationship between SE and EE such that the service providers can select an operating point as per the traffic demand [34]. Therefore, we define the SE of M-MIMO systems under perfect and imperfect CSI as

$$S^B_p = \sum_{k=1}^{K}\hat{R}^B_{p,k} \quad \text{and} \quad S^B_{im} = \frac{N-N_p}{N}\sum_{k=1}^{K}\hat{R}^B_{im,k}, \tag{53}$$

respectively, where $B \in \{mrc, zf\}$. Furthermore, we describe EE [34] of M-MIMO systems for perfect and imperfect CSI as

$$\eta^B_p = \frac{S^B_p}{E_s}, \quad \text{and} \quad \eta^B_{im} = \frac{S^B_{im}}{E_s}, \tag{54}$$

respectively.

We discuss the trade-off between SE and EE of the semi-blind estimator under imperfect CSI condition. For simplicity, we use $\mathbf{D}_1 = \mathbf{I}_K$ and $\mathbf{D}_l = \beta\mathbf{I}_K$ for $l \in \{2, \ldots, L-1\}$ [6], [34]. Thus, the variance of our estimator and its estimation error becomes $\sigma^2_{kd} = 1 + \frac{E_sK\beta(L-1)+N_0}{E_sN'}$ and $\sigma^2_{ekd} = \frac{E_sK\beta(L-1)+N_0}{E_sN'}$, respectively for $k = 1, \ldots, K$.

### A. MRC RECEIVER

We obtain the SE and EE of our semi-blind scheme in (55), as shown at the bottom of the next page. Furthermore, upper and lower limits on the EE of (55) is given by

$$\lim_{E_s\to 0}\eta^{mrc}_{im} = \lim_{E_s\to 0}\frac{S^{mrc}_{im}}{E_s} = \infty, \tag{56}$$

$$\lim_{E_s\to \infty}\eta^{mrc}_{im} = \lim_{E_s\to \infty}\frac{S^{mrc}_{im}}{E_s} = 0, \tag{57}$$

respectively.

$$R_{im,k} = \mathbb{E}\left\{\log_2\left(1+\frac{\beta_{1k}|\hat{\mathbf{a}}_k^H\hat{\mathbf{g}}_{1k}|^2}{\sum_{i=1}^{K}\beta_{1i}|\hat{\mathbf{a}}_k^H\mathbf{e}_{1i}|^2+\sum_{i=1,i\neq k}^{K}\beta_{1i}|\hat{\mathbf{a}}_k^H\hat{\mathbf{g}}_{1i}|^2+\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}|\hat{\mathbf{a}}_k^H\mathbf{g}_{li}|^2+\frac{N_0}{E_s}||\hat{\mathbf{a}}_k||^2}\right)\right\} \tag{46}$$

$$R^{mrc}_{im,k} = \mathbb{E}\left\{\log_2\left(1+\left(\frac{\sum_{i=1}^{K}\beta_{1i}|\hat{\mathbf{g}}_{1k}^H\mathbf{e}_{1i}|^2+\sum_{\substack{i=1\\ i\neq k}}^{K}\beta_{1i}|\hat{\mathbf{g}}_{1k}^H\hat{\mathbf{g}}_{1i}|^2+\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}|\hat{\mathbf{g}}_{1k}^H\mathbf{g}_{li}|^2+\frac{N_0}{E_s}||\hat{\mathbf{g}}_{1k}||^2}{\beta_{1k}||\hat{\mathbf{g}}_{1k}||^4}\right)^{-1}\right)\right\} \tag{47}$$

$$R^{mrc}_{im,k} \geq \hat{R}^{mrc}_{im,k} = \log_2\left(1+\left(\mathbb{E}\left\{\frac{\sum_{i=1}^{K}\beta_{1i}|\hat{\mathbf{g}}_{1k}^H\mathbf{e}_{1i}|^2+\sum_{\substack{i=1\\ i\neq k}}^{K}\beta_{1i}|\hat{\mathbf{g}}_{1k}^H\hat{\mathbf{g}}_{1i}|^2+\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}|\hat{\mathbf{g}}_{1k}^H\mathbf{g}_{li}|^2+\frac{N_0}{E_s}||\hat{\mathbf{g}}_{1k}||^2}{\beta_{1k}||\hat{\mathbf{g}}_{1k}||^4}\right\}\right)^{-1}\right) \tag{48}$$

### B. ZF RECEIVER

From (52), we obtain SE and EE of our semi-blind scheme in (58), as shown at the bottom of this page. The upper and lower limits on EE of the semi-blind scheme are shown as

$$\lim_{E_s\to 0}\eta_{im}^{zf}=\lim_{E_s\to 0}\frac{S_{im}^{zf}}{E_s}=\infty, \tag{59}$$

$$\lim_{E_s\to \infty}\eta_{im}^{zf}=\lim_{E_s\to \infty}\frac{S_{im}^{zf}}{E_s}=0, \tag{60}$$

respectively. From the above equations we infer that the EE of our algorithm decreases with an increase in SE and vice versa. This indicates a trade-off between its' EE and SE which is similar to the perfect CSI.

## VI. MCRLB FOR AN UNBIASED SEMI-BLIND ESTIMATOR IN MU MASSIVE MIMO SYSTEMS

Cramer-Rao lower bound (CRLB) provides the lowest achievable MSE for an unbiased estimator. However, its calculation in the presence of unknown parameters is difficult to compute due to which we use a looser bound known as MCRLB. Thus, we derive a MCRLB for semi-blind based unbiased estimator of pilot contaminated M-MIMO systems.

### A. MODIFIED CRAMER-RAO LOWER BOUND CALCULATION

For MCRLB, we consider full knowledge of $\mathbf{X}'_{1d}$ with BS 1 and $N'$ as the observation interval for CSI estimation. Moreover, we denote $\mathbf{F}$ as the Fisher information matrix (FIM) of channel $\mathbf{G}_1$. The small-scale fading channel coefficients corresponding to $m^{th}$ antenna of BS 1 i.e., $\mathbf{g}_{1m}$, depends only on the received signal by the same antenna $\mathbf{y}'_m$. Therefore, its FIM $\mathbf{F}$ becomes block diagonal with $\mathbf{F}_m$ as the FIM for $\mathbf{g}_{1m}$. The parameter vector for $\mathbf{F}_m$ is $\boldsymbol{\eta}_m=[Re\{\mathbf{g}_{1m}\}\, Im\{\mathbf{g}_{1m}\}]\in\mathbb{R}^{1\times 2K}$. For further calculations, we rewrite (3) and (4) as

$$\mathbf{y}_{mp}=\sqrt{E_s}\mathbf{g}_{1m}\mathbf{D}_1^{\frac{1}{2}}\mathbf{X}_p+\mathbf{v}_{mp}=\mathbf{u}_{mp}+\mathbf{v}_{mp}, \tag{61}$$

$$\mathbf{y}'_{md}=\sqrt{E_s}\mathbf{g}_{1m}\mathbf{D}_1^{\frac{1}{2}}\mathbf{X}'_{1_d}+\mathbf{v}_{md}=\mathbf{u}_{md}+\mathbf{v}_{md}, \tag{62}$$

respectively. We model $\mathbf{v}_{mp}$ and $\mathbf{v}_{md}$ as $\mathcal{CN}\left(\mathbf{0}_{1\times N_p},\mathbf{C}_{v_{mp}}\right)$ and $\mathcal{CN}\left(\mathbf{0}_{1\times N'_d},\mathbf{C}_{v_{md}}\right)$, respectively for $\mathbf{C}_{v_{mp}}=\mathbf{C}_p$ (19) and $\mathbf{C}_{v_{md}}=c_d\mathbf{I}_{N'_d}$ (21). Thus, the received vector $\mathbf{y}_{mp}$, and $\mathbf{y}'_{md}$ follow $\mathcal{CN}\left(\mathbf{u}_{mp},\mathbf{C}_{v_{mp}}\right)$, and $\mathcal{CN}\left(\mathbf{u}_{md},\mathbf{C}_{v_{md}}\right)$, respectively. The LLF of parameter $\boldsymbol{\eta}_m$ is

$$\ln f\left(\mathbf{y}'_m|\mathbf{X}'_1,\boldsymbol{\eta}_m\right)=\left(\mathbf{y}_{mp}-\mathbf{u}_{mp}\right)\mathbf{C}_p^{-1}\left(\mathbf{y}_{mp}-\mathbf{u}_{mp}\right)^H+\frac{\left(\mathbf{y}'_{md}-\mathbf{u}_{md}\right)\left(\mathbf{y}'_{md}-\mathbf{u}_{md}\right)^H}{c_d}. \tag{63}$$

The FIM of parameter $\boldsymbol{\eta}_m$ is given as

$$\mathbf{F}_m=2E_s\begin{bmatrix}Re[\mathbf{U}] & -Im[\mathbf{U}]\\ Im[\mathbf{U}] & Re[\mathbf{U}]\end{bmatrix}, \tag{64}$$

where $\mathbf{U}=\mathbf{D}_1^{\frac{1}{2}}\mathbf{X}_p\mathbf{C}_p^{-1}\mathbf{X}_p^H\mathbf{D}_1^{\frac{1}{2}}+\frac{1}{c_d}\mathbf{D}_1^{\frac{1}{2}}\mathbf{X}'_{1d}\mathbf{X}'^H_{1d}\mathbf{D}_1^{\frac{1}{2}}$.

Since $\mathbf{X}_p\mathbf{X}_p^H=N_p\mathbf{I}_K$ and $\mathbb{E}\left\{\mathbf{X}'_{1d}\mathbf{X}'^H_{1d}\right\}=N'_d\mathbf{I}_K$, the off-diagonal terms of (64) reduces to zero. Therefore, the resultant FIM of $\boldsymbol{\eta}_m$ is

$$\mathbf{F}_{mk}=2E_s\beta_{1k}\begin{bmatrix}u_k & 0\\ 0 & u_k\end{bmatrix}, \tag{65}$$

where $u_k=\mathbf{x}_{kp}\mathbf{C}_p^{-1}\mathbf{x}_{kp}^H+\frac{\mathbf{x}'_{1kd}\mathbf{x}'^H_{1kd}}{c_d}$ for $k=1,\ldots,K$. We obtain the MCRLB of $g_{1mk}$ by performing inverse operation on $\mathbf{F}_{mk}$ as

$$MCRLB(g_{1mk})=[1\;\; j]\frac{1}{2E_s\beta_{1k}}\begin{bmatrix}\frac{1}{u_k} & 0\\ 0 & \frac{1}{u_k}\end{bmatrix}\begin{bmatrix}1\\ -j\end{bmatrix}. \tag{66}$$

Hence, the MCRLB of $g_{1mk}$ is given by

$$\mathbb{E}\left\{|\hat{g}_{1mk}-g_{1mk}|^2\right\}\geq\frac{1}{E_s\beta_{1k}\left(\mathbf{x}_{kp}\mathbf{C}_p^{-1}\mathbf{x}_{kp}^H+\frac{N'_d}{c_d}\right)}, \tag{67}$$

From (67), we note that the MCRLB for unbiased semi-blind estimator is inversely proportional to the amount of data incorporated in the channel estimation process ($N'_d$), i.e. larger the $N'_d$ lower the MSE.

$$R_{im,k}^{zf}=\mathbb{E}\left\{\log_2\left(1+\left(\frac{\sum_{i=1}^K\beta_{1i}|\hat{\mathbf{a}}_k^H\mathbf{e}_{1i}|^2+\left(\sum_{l=2}^L\sum_{i=1}^K\beta_{li}+\frac{N_0}{E_s}\right)||\hat{\mathbf{a}}_k||^2}{\beta_{1k}}\right)^{-1}\right)\right\} \tag{50}$$

$$R_{im,k}^{zf}\geq\hat{R}_{im,k}^{zf}=\log_2\left(1+\frac{\beta_{1k}}{\mathbb{E}\left\{\sum_{i=1}^K\beta_{1i}|\hat{\mathbf{a}}_k^H\mathbf{e}_{1i}|^2+\left(\sum_{l=2}^L\sum_{i=1}^K\beta_{li}+\frac{N_0}{E_s}\right)||\hat{\mathbf{a}}_k||^2\right\}}\right) \tag{51}$$

$$S_{im}^{mrc}=\frac{N-N_p}{N}K\log_2\left(1+\frac{E_s\left(\beta K(L-1)+N'\right)+N_0}{E_s\left(\beta K(L-1)(2K-1+N')+N'(K-1)\right)+N_0(2K-1+N')}\right)\text{ and }\eta_{im}^{mrc}=\frac{S_{im}^{mrc}}{E_s} \tag{55}$$

$$S_{im}^{zf}=\frac{N-N_p}{N}K\log_2\left(1+\frac{(M-K)\left(E_s\left(K\beta(L-1)+N'\right)+N_0\right)}{E_sK\beta(L-1)(K+N')+N_0(K+N')}\right),\quad and\quad \eta_{im}^{zf}=\frac{S_{im}^{zf}}{E_s} \tag{58}$$

### B. TEST OF UNBIASEDNESS FOR THE DESIGNED SEMI-BLIND ESTIMATOR

For bias calculation of the proposed estimator, we assume full information of transmit matrix $X'_{1d}$ with BS 1 (26). Therefore,

$$\mathbb{E}\left\{\hat{\mathbf{g}}_{1m}\right\} = \mathbf{g}_{1m} + \mathbb{E}\left\{\left(\sum_{l=2}^{L}\mathbf{g}_{lm}\mathbf{D}_l^{\frac{1}{2}}\left(\mathbf{X}_p\mathbf{C}_p^{-1}\mathbf{X}_p^H + \frac{\mathbf{X}_{ld}\mathbf{X}_{1d}'^H}{c_d}\right) + \left(\frac{\mathbf{w}_{mp}\mathbf{C}_p^{-1}\mathbf{X}_p^H}{\sqrt{E_s}} + \frac{\mathbf{w}'_{md}\mathbf{X}_{1d}'^H}{\sqrt{E_s}c_d}\right)\right)\left(\mathbf{X}_p\mathbf{C}_p^{-1}\mathbf{X}_p^H + \frac{\mathbf{X}'_{1d}\mathbf{X}_{1d}'^H}{c_d}\right)^{-1}\mathbf{D}_1^{-\frac{1}{2}}\right\}. \quad (68)$$

Since noise component $\left[\mathbf{w}_{mp}\mathbf{w}'_{md}\right]$ is independent of $\left[\mathbf{X}_p\mathbf{X}'_{1d}\right]$, the expectation term in right side of the equality is approximated to zero. Thus, $\mathbb{E}\left\{\hat{\mathbf{g}}_{1m}\right\} \simeq \mathbf{g}_{1m}$ which implies that our estimator is asymptotically unbiased.

## VII. CONVERGENCE ANALYSIS OF THE PROPOSED SEMI-BLIND ESTIMATOR

In this section, we will prove the convergence property of the proposed semi-blind estimator which utilizes a SAGE algorithm for pilot contaminated multi-user massive MIMO systems following [27]. It shows that if initial values of parameters using a SAGE algorithm are initialized in a region suitably close to the local maximum, then over iterations, the sequence of estimates will monotonically converge to the norm of it. Since we implement SAGE algorithm in channel estimation problem of a linear system model (1), convergence of the proposed semi-blind estimator is guaranteed.

From previous section, we know that SAGE algorithm uses hidden data space and incomplete data space to obtain maximum likelihood estimator (MLE) through iterations. It divides the estimation parameters in groups such that if one group is being updated the remaining groups remain fixed to their current revised values. In our work, $\mathbf{g}_{1m} \in \mathbb{C}^{1\times K}$ denotes the parameter under consideration and rest of the channel vectors are combined under $\mathbf{g}_{1\tilde{m}} \in \mathbb{C}^{(M-1)\times K}$ where $m \neq \tilde{m}$ and $m, \tilde{m} \in \{1, \ldots, M\}$. We indicate $\mathcal{S}^m$ and $\mathbf{Y}'$ as hidden data and incomplete data space, respectively where $\mathcal{S}^m = \left\{\mathbf{y}'_m, \mathbf{X}'_1\right\}$. For convergence proof, we have made the following assumptions [27]:

- The regularity condition for convergence of likelihood and the sequence of estimates are satisfied.
- The maximum point of the likelihood is an interior point of the parameter space.

From (12), we note that the mean of hidden data space $\mathcal{S}^m$ is linear with $\mathbf{g}_{1m}$ i.e., $\mathcal{S}^m \sim \mathcal{CN}\left(\mathbf{g}_{1m}\mathbf{X}'_1, \mathbf{C}_{\mathcal{S}^m}\right)$;

$$\mathbf{C}_{\mathcal{S}^m} = \begin{bmatrix} \mathbf{C}_{vmp} & \mathbf{0}_{N_p\times N'_d} \\ \mathbf{0}_{N'_d\times N_p} & \mathbf{C}_{vmd} \end{bmatrix} \in \mathbb{C}^{N'\times N'}; \quad (69)$$

$\mathbf{C}_{vmp}$ and $\mathbf{C}_{vmd}$ matrices are given in expression (19) and (21), respectively.

Therefore, convergence of the proposed SAGE based semi-blind algorithm can be proved by satisfying the following theorem [27].

*Theorem 1:* $\mathcal{U} \subset \mathcal{G}$ is a region of monotone convergence in norm if there exists a nonsingular matrix $\mathbf{T}$ such that $\mathcal{U}$ is an open ball with respect to the norm $||.||_{\mathbf{T}}$ and

1). For $m = 1, \ldots, M$, the norm of composite matrix $||\mathcal{M}^m\left(\mathbf{g}_{1m}, \hat{\mathbf{G}}_1^{(i)}\right)||_T \leq 1$ for all $\hat{\mathbf{G}}_1^{(i)} \in \mathcal{U}$ and $\mathbf{g}_{1m} \in Q_m\left(\mathbf{g}_{1m}|\hat{\mathbf{G}}_1^{(i)}\right)$,

2). For any $\hat{\mathbf{g}}_{11}^{(i)}, \ldots, \hat{\mathbf{g}}_{1M}^{(i)} \in \mathcal{U}$ and $\mathbf{g}_{1m} \in Q_m\left(\mathbf{g}_{1m}|\hat{\mathbf{G}}_1^{(i)}\right)$ with $m = 1, \ldots, M$

$$||\mathcal{M}^1(\mathbf{g}_{11}, \hat{\mathbf{G}}_1^{(i)}) \ldots \mathcal{M}^M(\mathbf{g}_{1M}, \hat{\mathbf{G}}_1^{(i)})|| \leq 1; \quad (70)$$

$\mathcal{G}$ represents subset of $\mathbb{C}^{M\times K}$ and $i$ indicates the iteration number.

*Proof:* To establish the first statement, we need the Hessian of the problem i.e., $\mathcal{H} = -\nabla^2 Q_m\left(\mathbf{g}_{1m}|\hat{\mathbf{G}}_1^{(i)}\right)$. We note that the Hessian matrix $\mathcal{H}$ of (17) is the Fischer information matrix (FIM) of $\mathbf{g}_{1m}$ (i.e., $\mathbf{F}_m$), which is given in (64). Moreover, $\mathbf{F}_m$ is a positive definite matrix which defines $Q_m\left(\mathbf{g}_{1m}|\hat{\mathbf{G}}_1^{(i)}\right)$ as a strictly concave objective of $\hat{\mathbf{G}}_1$. This guarantees the existence of a non-empty region of monotone convergence in norm $\mathcal{U}$ which satisfies the statements of theorem 1 for $\mathbf{T} = (\mathbf{F}_m)^{\frac{1}{2}}$ [27]. Also, for the SAGE algorithm $\mathcal{M}^m = \mathbf{T}^{-1}\mathcal{P}_m\mathbf{T}$ [27]. Since $\mathbf{T}$ is a non-singular matrix,

$$\begin{aligned} ||\mathcal{M}^m\left(\mathbf{g}_{1m}, \hat{\mathbf{G}}_1^{(i)}\right)||_T &= ||\mathbf{T}\mathcal{M}^m\left(\mathbf{g}_{1m}, \hat{\mathbf{G}}_1^{(i)}\right)\mathbf{T}^{-1}|| \\ &= ||\mathbf{T}\mathbf{T}^{-1}\mathcal{P}_m\mathbf{T}\mathbf{T}^{-1}|| \\ &= ||\mathcal{P}_m|| \end{aligned} \quad (71)$$

where $\mathcal{P}_m$ is the orthogonal projection onto the $m^{th}$ column of $\mathbf{T}$. Since an orthogonal projection is non-expansive, $||\mathcal{M}^m\left(\mathbf{g}_{1m}, \hat{\mathbf{G}}_1\right)||_T = ||\mathcal{P}_m|| \leq 1$ satisfies the first condition of Theorem 1.

To prove the second statement of theorem 1, we consider the Gauss-Siedel method which is an iterative technique that converges for a linear problem $\mathbf{Ax} = \mathbf{b}$, if $\mathbf{A}$ is well conditioned (i.e., positive definite or diagonally dominant). In our work, we utilize $\mathbf{X}'_1 = \left[\mathbf{X}_p\mathbf{X}'_{1d}\right]$ as $\mathbf{A}$ matrix for estimation of $\mathbf{g}_{1m}$. Since we employ Zadoff-Chu (ZC) codes [36] for generating pilot sequences and i.i.d. unit energy symbols for data sequence, the transmit matrix $\mathbf{X}'_1$ forms a positive definite matrix. Therefore, for all the $M$ antennas we have

$$||\mathcal{M}^1(\mathbf{g}_{11}, \hat{\mathbf{G}}_1) \ldots \mathcal{M}^M(\mathbf{g}_{1M}, \hat{\mathbf{G}}_1)||_T \leq 1. \quad (72)$$

From (71) and (72), we note that the proposed SAGE based semi-blind estimator converges to the norm of $\hat{\mathbf{G}}_1^{(i)}$.

The convergence proof will conclude the following statements [27], [37]:

- If a SAGE algorithm is initialized in a region suitably close to a local maximum in the interior of parameter space, then the sequence of estimates will converge monotonically in norm to it.

- For suitably regular objective function, the region of monotone convergence in norm is guaranteed to be non-empty.
- The asymptotic convergence rate of a SAGE algorithm will be improved if one chooses a less informative hidden data space.

## VIII. RESULTS AND DISCUSSION

In this section, we simulate, and discuss the MSE, BER, SE and EE performance of the semi-blind estimator against the existing schemes of pilot contaminated M-MIMO systems.

*Simulation Environment:* We consider a four cell system i.e. $L = 4$, where each BS serves $K = 4$ single-antenna UTs. The transmission frame comprises of $N = 200$ symbols with four pilot ($N_p = 4$) and 196 data symbols ($N_d = 196$). The consideration of 200 symbols for one transmission frame goes in-line with a typical LTE communication scenario [3], [8], [34], [38], [39]. This ensures that the transmission frame is within the coherence interval of 1 ms [3], [8], [34]. Furthermore, we use the minimum pilot length, i.e., $N_p = K$ [33] for channel estimation in all the simulations. Our estimator uses $N_p = 4$ and $N'_d = 20$ symbols for CSI acquisition, where $N'_d = 20$ accounts only 10% of the transmission frame. Moreover, the approximation using law of large numbers holds true in (10) and (15) for $M \geq 16$ and $N'_d \geq 20$, respectively. We consider a fixed case of large-scale fading coefficients [6], [34], where $\beta_{1k} = 1$ ($\mathbf{D}_1 = \mathbf{I}_K$) and $\beta_{lk} = \beta$ ($\mathbf{D}_l = \beta\mathbf{I}_K$) such that $0 \leq \beta < 1$ for $l = 2, \ldots, L$, and $k = 1, \ldots, K$. The $\beta$ quantifies pilot contamination intensity on desired cell (i.e., cell 1) as the impact of the same increases with $\beta$. The noise variance ($N_0$) is taken as 1. We employ Zadoff-Chu code [36] and BPSK modulation for generating pilot, and data sequences, respectively. We compare our scheme against the existing estimators [6], [7] whose results are regenerated as per the simulation environment. We label the ISAGE stage estimate as proposed with iteration number in legend of all the figures. We follow the above-mentioned parameters for all the simulations unless specified.

### A. MSE AND BER PERFORMANCE ANALYSIS

Fig. 2 shows MSE performance of $M = 32$ and $K = 4$ M-MIMO system, where the proposed scheme achieves a significant improvement in MSE compared to the estimators of [6], [7] with the increase in SNR. *This enhancement is attributed to the addition of $N'_d$ data symbols in the CSI acquisition process which reduces the effect of pilot contamination and additive noise.* We obtain a maximum MSE gain of about 6 dB from [6], [7] for SNR$\geq$10 dB. Furthermore, the CSI accuracy of the semi-blind estimator improves from iteration 1 to 2, but the difference becomes negligible between iteration 2 and 3. This indicates that it converges in two iterations. So, we plot the rest of the results of our algorithm with two iterations only. In addition, the MSE of SAGE iteration 2 tends toward the derived MCRLB with a difference of 0.5 dB between them for SNR$\geq$ 0 dB.

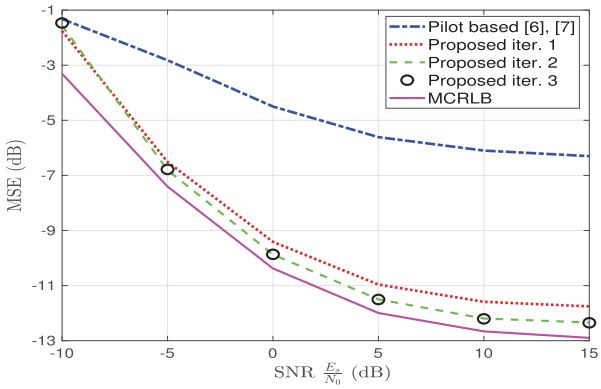


**FIGURE 2.** MSE vs. SNR for $M = 32$, and $K = 4$ massive MIMO system with $\beta = 0.1$, $N'_d = 20$.

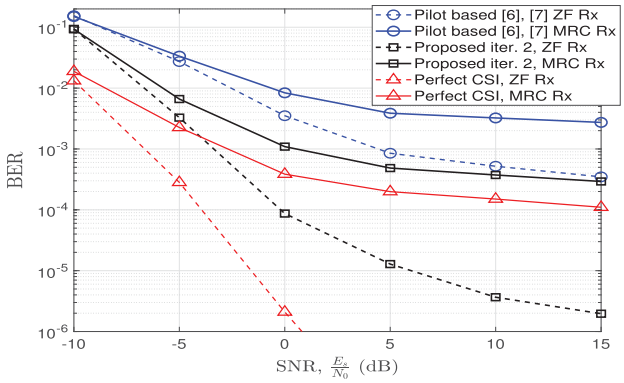


**FIGURE 3.** BER performance for $M = 32$, and $K = 4$ massive MIMO system with $\beta = 0.1$, $N'_d = 20$.

We also plot BER performance of the same configuration for both ZF and MRC Rx in Fig. 3. At a BER of $3 \times 10^{-3}$, the semi-blind scheme obtains an SNR gain of about 12 dB and 5 dB from the estimators of [6], [7] while using MRC and ZF Rx, respectively. Moreover, the BER performance of our scheme improves considerably with the increase in SNR. Besides, at BER of $10^{-3}$, ISAGE iteration 2 trails by 3 dB in SNR from the perfect CSI case for both MRC and ZF Rx. Thus, we conclude that the semi-blind scheme obtains considerable improvement from the existing pilot-aided estimators in terms of MSE and BER at the cost of a nominal increase in complexity.

The convergence of our semi-blind estimator in two iterations can be intuitively explained as:

- The proposed algorithm uses two stages to obtain a CSI estimate, namely, initialization and iteration stage. In the initialization stage, we use a pilot based LMMSE method for an initial estimate of channel parameters, which is then refined in the iteration stage (ISAGE stage) through the SAGE algorithm. The iteration stage uses both known pilot and unknown data symbols for updating the initial CSI estimate. Since data information is unavailable with the BS, we use an LMMSE detector for decoding the unknown data sent from users of the

desired cell. For data recovery, we treat the available channel estimate as true CSI and replace all the instances of true CSI with the channel estimate. Therefore, first iteration of the proposed scheme uses pilot based initial estimate (i.e., $\hat{\mathbf{G}}_1^{(0)}$) for obtaining an estimate of $\mathbf{X}'_{1d}$ from received $\mathbf{Y}'_d$ with the help of (16). The recovered data of $N'_d$ length is then combined with the known pilot sequence of $N_p$ length for refining the initial CSI estimate $\hat{\mathbf{G}}_1^{(0)}$ during the first iteration of SAGE algorithm. Due to the incorporation of $N'_d$ data symbols in the estimation process, CSI accuracy of channel estimate obtained from the first SAGE iteration (i.e., $\hat{\mathbf{G}}_1^{(1)}$) is significantly improved from the initial CSI estimate ($\hat{\mathbf{G}}_1^{(0)}$).

- During the second iteration of ISAGE stage, we utilize the channel estimate of the first iteration (i.e., $\hat{\mathbf{G}}_1^{(1)}$) for detecting $\mathbf{X}'_{1d}$ from received $\mathbf{Y}'_d$. As the first iteration estimate $\hat{\mathbf{G}}_1^{(1)}$ is more accurate than the initial estimate $\hat{\mathbf{G}}_1^{(0)}$, the $K$ users data information is correctly recovered with $\hat{\mathbf{G}}_1^{(1)}$ than $\hat{\mathbf{G}}_1^{(0)}$. Hence, we infer that the estimated CSI accuracy is highly dependent on how correctly the algorithm can detect the transmit data $\mathbf{X}'_{1d}$. However, during the third iteration, the data detection of $\mathbf{X}_{1d'}$ employs estimated channel coefficients of second iteration (i.e., $\hat{\mathbf{G}}_1^{(2)}$) which obtains limited improvement in CSI accuracy from iteration 1s' updated CSI ($\hat{\mathbf{G}}_1^{(1)}$). Additionally, usage of large BS antennas aids in the accurate recovery of $\mathbf{X}'_{1d}$ with fewer iterations.

We display the impact of $N'_d$ on CSI accuracy of the semi-blind estimator in Fig. 4. We note that its' MSE equals the derived MCRLB for $N'_d \geq 60$ with all the given SNRs. Moreover, the CSI accuracy improves with the increase in $N'_d$. We observe that, it requires about 84, and 48 $N'_d$ to obtain a MSE of -16 dB at 0 dB, and 10 dB SNR, respectively. Thus, the semi-blind estimator demands less $N'_d$ with the increased SNR which further reduces its computational complexity. For the given frame length i.e., $N = 200$, we achieve a maximum MSE gain of about 15 dB and 16 dB with $N'_d = 196$ from [6], [7] at 0 dB and 5 dB SNR, respectively. Therefore, we conclude that for large $N'_d$ our semi-blind estimator attains the MCRLB even at low SNRs. From the above discussion, we conclude that the estimation accuracy of the estimator depends largely on data length used in the channel estimation process (i.e., $N'_d$). Thus, the choice of $N'_d$ should be made as per the systems' objective. For example, if a system demands an accurate CSI, we choose a higher number for $N'_d$ with the help of (31); whereas a system with a constraint on signal processing ends may not opt for larger $N'_d$ and can be evaluated from the joint optimization of (31) and (33).

The effect of pilot contamination on the proposed estimator and [6], [7] scheme is shown in Fig. 5. We note that the semi-blind estimator obtains an appreciable gain in MSE over the schemes of [6], [7] for $\beta < 0.45$. The MSE of semi-blind scheme improves by doubling $M$ from 32 to 64 for $\beta > 0.1$; however, the CSI accuracy of pilot based schemes [6], [7]

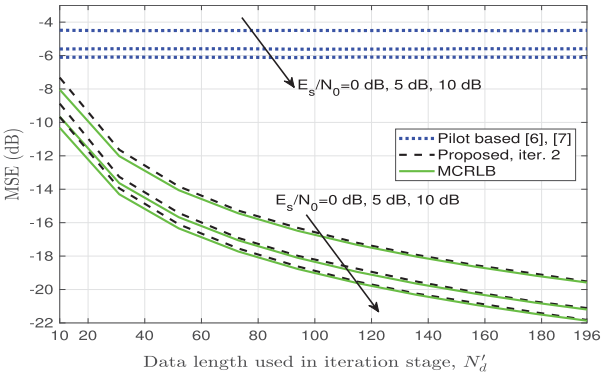


**FIGURE 4.** MSE vs. $N'_d$ plot for $M = 32$, and $K = 4$ M-MIMO system with $\beta = 0.1$.

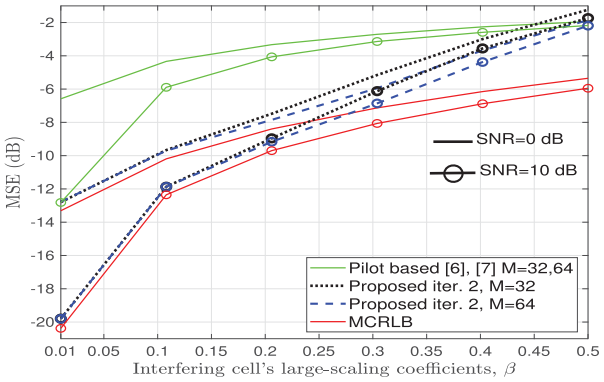


**FIGURE 5.** MSE vs. $\beta$ curve for $M = 32, 64$, and $K = 4$ M-MIMO system with $N'_d = 20$.

does not show any improvement with the increase in $M$. Therefore, the semi-blind estimator is more robust to the pilot contamination in comparison to the pilot based methods.

To determine the effectiveness of our proposed scheme against the existing semi-blind estimators [14], [24], [25], we have shown MSE performance in Fig. 6 and Fig. 7. For a fair comparison, we have simulated our algorithm as per their simulation environment. As a result, Fig. 6 displays analogy with the MSE of data-aided scheme [14], whereas Fig. 7 shows with the estimators of [24], [25]. The description on both these figures are provided as:

- From Fig. 6, we note that the MSE curve of our scheme converges at 0.172 for $M \geq 128$ in $L = 19$ cells with $K = 10$ massive MIMO configuration. It attains a maximum MSE gain of about 0.07 from the data-aided method [14] at $M = 8$. Moreover, it guarantees a minimum MSE gain of around 0.03 for rest of the given BS antennas ($M$). Therefore, the proposed semi-blind estimator outperforms the existing data-aided scheme [14] for all the given $M$ with the above system configuration.
- The effect of $N'_d$ on MSE performance of our algorithm and estimators of [24], [25] for a seven-cell system ($L = 7$) with $M = 60$, $K = 4$ massive MIMO structure is shown in Fig. 7. We observe that the MSE of proposed estimator improves with the increase in $N'_d$,

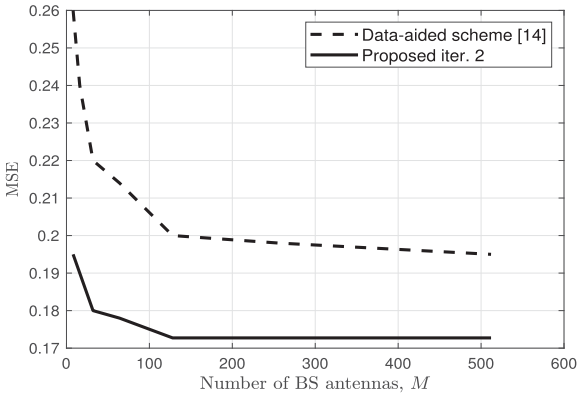


**FIGURE 6.** MSE vs. $M$ performance for $L = 19$ and $K = 10$ M-MIMO configuration at 0 dB SNR; $N'_d = 100$, $N_p = 10$,and $\beta = 0.1$.

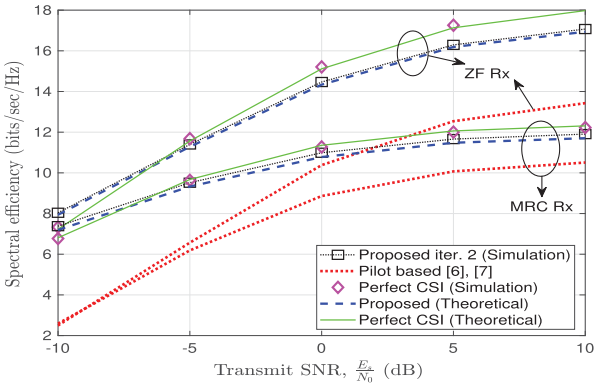


**FIGURE 8.** SE vs. SNR plot for $M = 32$, and $K = 4$ M-MIMO systems; $\beta = 0.1$, $N'_d = 20$.

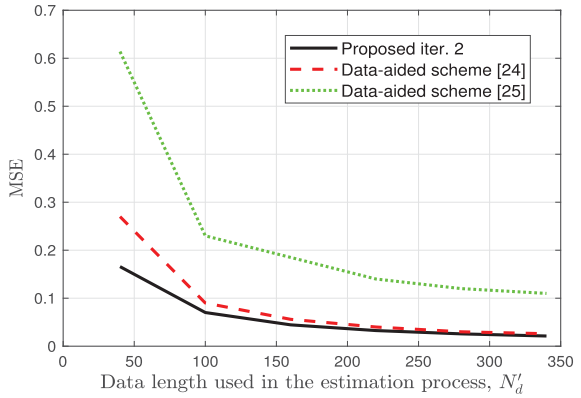


**FIGURE 7.** MSE vs. $N'_d$ plot for $L = 7$, $M = 60$, and $K = 4$ M-MIMO system at 10 dB SNR; $N_p = 8$ and $\beta = 0.3$.

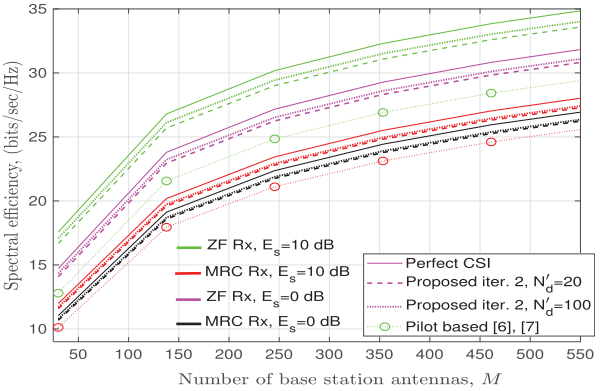


**FIGURE 9.** SE vs. $M$ curve for $K = 4$, $\beta = 0.1$, and $N'_d = 20$ M-MIMO system.

but it ceases at 0.02 for $N'_d \geq 300$. However, the MSE of data-aided scheme [24] converges to 0.03 for $N'_d \geq 300$. In addition, our estimator attains a considerable gain of about 0.12 and 0.45 in MSE from the methods of [24] and [25], respectively at $N'_d = 40$. We notice that the difference between the MSE of our algorithm and [24], [25] reduces with the increase in $N'_d$, despite the proposed scheme ensures a MSE gain of about 0.1 and 0.01 against the data-aided methods of [25] and [24], respectively. Thus, the proposed estimator surpasses the existing data-aided techniques of [24], [25] for all the given $N'_d$. Note that algorithm [24] requires six iterations to converge and scheme [25] utilises computationally demanding singular-value-decomposition (SVD) for CSI acquisition. Hence, the proposed algorithm is less computationally complex than the existing data-aided schemes of [24], [25].

### B. SPECTRAL EFFICIENCY PERFORMANCE ANALYSIS

We ascertain the tightness of the derived bound on the UL achievable rate of MRC and ZF Rx in Fig. 8 for $M = 32$ and $K = 4$ M-MIMO system. We note that the derived bound on SE of both perfect CSI and imperfect CSI is tight for all the given SNRs. Additionally, we notice that our scheme attains a minimum gain of 4 bits/sec/Hz and 1.2 bits/sec/Hz in SE against [6], [7] with ZF and MRC Rx, respectively. Therefore, the semi-blind scheme is more spectrally efficient than the existing methods [6], [7]. Since the theoretically derived bound closely matches the simulation-based results, the rest of the figures will use the analytical bounds for comparisons.

Fig. 9 shows the effect of increased BS antennas ($M$) on the SE of both the semi-blind and pilot-aided [6], [7] schemes. We note that for $M \geq 200$ difference between SE of the perfect CSI and the proposed scheme with 20 $N'_d$ remains constant at about 1.4 bits/sec/Hz and 0.8 bits/sec/Hz with ZF and MRC Rx, respectively. Besides, this difference reduces to 0.8 bits/sec/Hz by increasing $N'_d$ to 100 for ZF Rx. The SE of [6], [7] with 10 dB SNR lies below the proposed schemes' SE of 0 dB SNR for all the given $M$. Hence, the semi-blind estimator is power-efficient in comparison to existing schemes [6], [7] as it demands less SNR for the same SE.

The impact of pilot contamination on SE of $M = 50, 100$ and $K = 4$ massive MIMO system for ZF and MRC Rx is plotted in Fig. 10.

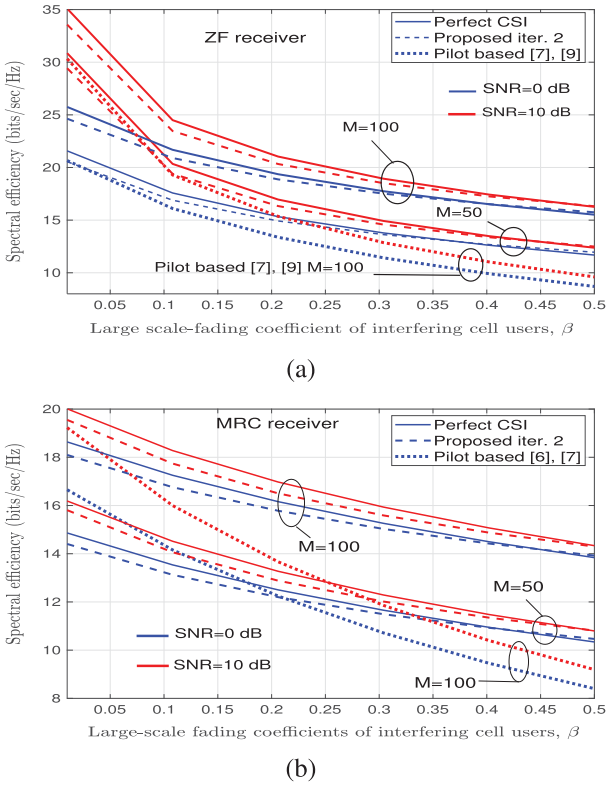


**FIGURE 10.** SE vs. $\beta$ performance for (a) ZF and (b) MRC receiver; $K = N_p = 4$, $N'_d = 20$.

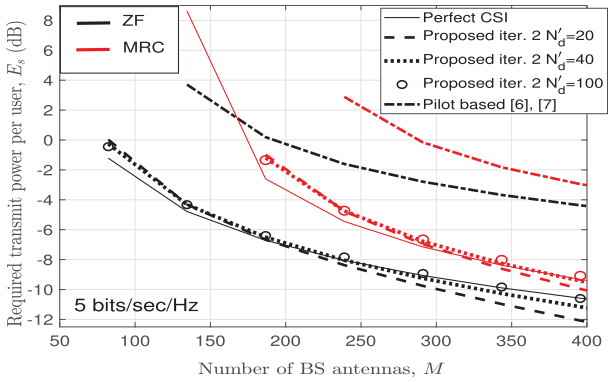


**FIGURE 11.** $E_s$ vs. $M$ plot for a SE of 5 bits/sec/Hz per user with $K = N_p = 4$, and $N'_d = 20, 40, 100$.

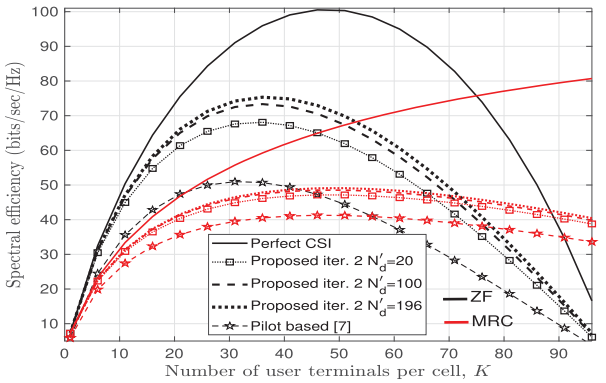


**FIGURE 12.** SE vs. $K$ plot for $M = 100$ and $K = N_p$ M-MIMO system with $\beta = 0.1$ and SNR=5 dB.

*ZF Receiver:* In Fig. 10 (a), we note that the semi-blind scheme approaches the SE of perfect CSI for $\beta \geq 0.35$. Moreover, its SE tends toward 16 and 12.5 bits/sec/Hz for $M = 50$ and $M = 100$, respectively. In addition, the SE of semi-blind scheme with $M = 50$ and SNR= 0 dB outperforms the SE of [6], [7] with $M = 100$ and SNR= 10 dB for $\beta > 0.25$. Moreover, at 0 dB SNR and $M = 100$, our scheme achieves a considerable gain in SE from estimators [6], [7] for all the given values of $\beta$.

*MRC Receiver:* Similarly in Fig. 10 (b), the SE of semi-blind scheme approaches the SE of perfect CSI for $\beta > 0.4$ with MRC Rx. Further, its' SE with $M = 50$ and $M = 100$ advances towards 10.5 and 14.5 bits/sec/Hz respectively. Besides, SE curve of estimator [7] with $M = 100$ and SNR=10 dB lags behind our scheme with fewer BS antennas and lower SNR i.e., $M = 50$ and SNR=0 dB for $\beta > 0.35$. From Fig. 10, we notice that the semi-blind scheme requires fewer BS antennas and low SNRs to achieve a given SE against the pilot-aided estimators. So, our scheme reduces the hardware complexity, cost, and circuit power consumption involved with a large number of BS antennas. Besides, we note that the ZF Rx performs fairly well against the MRC for all the given $M$, $\beta$ and SNRs. However, complexity involved with the ZF Rx is much higher compared to the MRC.

Fig. 11 plots the $E_s$ required to achieve a SE of 5 bits/sec/Hz per user as a function of $M$. For ZF Rx, we observe that the $E_s$ vs. $M$ curve of proposed estimator ($M > 180$ and $N'_d = 100$) coincides with the same of perfect CSI. For MRC Rx, we require $M \geq 300$ and $N'_d = 100$ to attain the performance of perfect CSI. In addition, by doubling the number of BS antennas from $M = 100$ to $M = 200$, the $E_s$ of the proposed estimator reduces by 5 dB using ZF Rx. Similarly, with MRC Rx the $E_s$ of proposed scheme reduces by about 8.5 dB while increasing $M$ from 200 to 400. We also note that the MRC Rx requires more $E_s$ in comparison to ZF to achieve the same SE per user for all the given $M$ and $N'_d$. Hence, we conclude that our schemes' power requirement matches with the perfect CSI condition for large $N'_d$.

The relationship between the number of users per cell ($K$) and SE for $M = 100$ based M-MIMO is shown in Fig. 12. We note that the proposed estimator closely follows perfect CSI condition for $\frac{M}{K} \geq 10$ whereas the difference increases for $\frac{M}{K} < 10$. For ZF Rx, the estimator can obtain a maximum SE of 68, 73 and 75 bits/sec/Hz for $N'_d = 20, 100, 196$, respectively with only $K = 36$ users per cell. However, the estimators of [6], [7] using ZF Rx can achieve only 51 bits/sec/Hz with 36 users per cell. Moreover, for ZF Rx, the maximum number of users which can be served simultaneously at an SE of 40 bits/sec/Hz is 72 (90 with MRC)

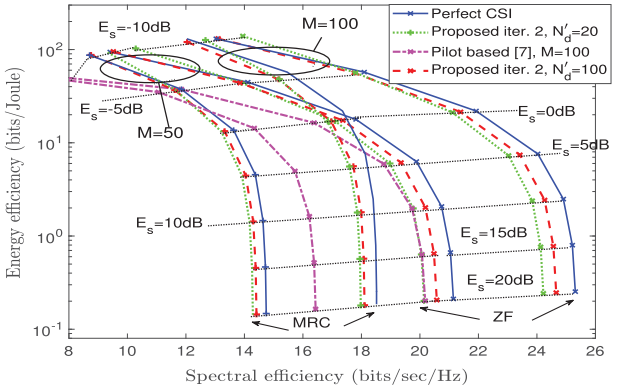


**FIGURE 13.** EE vs. SE performance for $K = N_p = 4$ based M-MIMO system.

and 56 (64 with MRC) for the proposed and pilot aided [7] estimator, respectively. Therefore, the semi-blind scheme can accommodate more number of users in comparison to pilot aided estimators for a given SE and BS antenna constraint.

### C. ENERGY EFFICIENCY VERSUS SPECTRAL EFFICIENCY TRADE-OFF

Fig. 13 shows the trade-off between EE and SE for $M = 50, 100$ and $K = 4$ based M-MIMO system configuration. We obtain the points shown on the curves by calculating the EE and SE for the given values of $E_s$. We note that the ZF Rx outperforms the MRC for all the given configurations; moreover, the difference between them increases with the increase in $E_s$. With MRC Rx, the proposed estimator achieves a gain of about 2 bits/sec/Hz in SE and 4 bits/Joule in EE by doubling the BS station antenna from $M = 50$ to $M = 100$ at $E_s = 0$ dB. Similarly, we obtain a gain of about 4 bits/sec/Hz in SE and 4 bits/Joule in EE with ZF Rx. Furthermore, the proposed scheme approaches perfect CSI condition with the increase in $N'_d$. In addition, for $E_s \leq 0$ dB the semi-blind scheme with MRC Rx and $M = 100$ obtains a considerable improvement in EE and SE from estimator [7] using ZF Rx and $M = 100$.

### D. PERFORMANCE SUMMARY OF THE PROPOSED ESTIMATOR

In this subsection, we summarize the performance of our estimator in the following points:

- The CSI accuracy of our algorithm significantly improves with the inclusion of more $N'_d$ in the estimation process against the existing pilot-aided schemes.
- The proposed scheme requires on an average two iterations for attaining convergence under the given scenarios which makes the complexity comparable with the existing pilot-aided methods.
- The MSE of our estimator attains MCRLB with the increased $N'_d$ even at lower SNRs.
- The minimum transmit power needed to attain a given spectral efficiency by our algorithm matches with the perfect CSI condition for large $N'_d$ and BS antennas $M$.
- The estimation process of proposed method serves more number of users in comparison to existing pilot-aided estimators by increasing the $N'_d$ for a given spectral efficiency, SNR, and BS antenna constraint.
- The energy efficiency and spectral efficiency of our estimator considerably improves with the increase in $N'_d$.
- The proposed estimator is practically realizable as the minimum number of BS antennas and $N'_d$ needed to satisfy the approximations using law of large numbers in the estimation process is 16 and 20, respectively.
- The usage of a large number of data symbols in the estimation process ($N'_d$) induces proportional increase in the computational complexity order of the proposed scheme.

## IX. CONCLUSION

In this paper, we have presented a SAGE-based semi-blind channel estimator for pilot contaminated MU M-MIMO systems. The algorithm obtains a maximum MSE gain of about 6 dB from the existing pilot-aided estimators by incorporating only 20 data symbols ($N'_d$) in the estimation process. Further, it is more robust to the pilot contamination in comparison to the existing pilot based schemes. In addition, the CSI accuracy of the proposed estimator coincides with the derived MCRLB for large values of $N'_d$. The estimator also gains at least 4 bits/sec/Hz and 1.2 bits/sec/Hz in SE from the existing pilot-aided estimators with ZF and MRC receiver, respectively. Moreover, it achieves the SE of perfect CSI condition with large $M$, large $N'_d$, or in the case of the high pilot contaminated region. It can also accommodate more number of users in comparison to pilot aided estimators for a given SE and BS antenna constraint. Therefore, the proposed estimator improves the overall performance of an MU M-MIMO system at a nominal increase in computational complexity.

## APPENDIXES

## APPENDIX A

## PERFECT CSI: UL ACHIEVABLE RATE WITH MRC RECEIVER

The lower bound on $R^{mrc}_{p,k}$ of (39) can be rewritten as (73) which is shown at the top of the next page; where $\tilde{g}_{1i} = \frac{\mathbf{g}^H_{1k}\mathbf{g}_{1i}}{||\mathbf{g}_{1k}||}$ and $\tilde{g}_{li} = \frac{\mathbf{g}^H_{1k}\mathbf{g}_{li}}{||\mathbf{g}_{1k}||}$. The random variable $\tilde{g}_{1i}$ conditioned on $\mathbf{g}_{1k}$ follows $\mathcal{CN}(0,1)$ for $i = 1, \ldots, k-1, k+1, \ldots, K$. Similarly, $\tilde{g}_{li} \sim \mathcal{CN}(0,1)$. By using $\mathbb{E}\left\{||\mathbf{g}_{1k}||^2\right\} = M$, $\mathbb{E}\left\{|\tilde{g}_{1i}|^2\right\} = 1$ and $\mathbb{E}\left\{|\tilde{g}_{li}|^2\right\} = 1$ in (73), as shown at the the top of the next page, we obtain the closed form expression in (40).

## APPENDIX B

## PERFECT CSI: UL ACHIEVABLE RATE WITH ZF RECEIVER

Since $\mathbb{E}\left\{|g_{lim}|^2\right\} = 1$, the expectation term $\mathbb{E}\left\{|\mathbf{a}^H_k\mathbf{g}_{li}|^2\right\}$ reduces to $\mathbb{E}\left\{||\mathbf{a}_k||^2\right\}$ in (42) for $l = 2, .., L$, $i = 1, .., K$ and $m = 1, .., M$. Further,

$$\mathbb{E}\left\{||\mathbf{a}_k||^2\right\} = \mathbb{E}\left\{\left[\left(\mathbf{G}^H_1\mathbf{G}_1\right)^{-1}\right]_{kk}\right\} = \frac{1}{K}\,\mathbb{E}\left\{tr\left[\left(\mathbf{G}^H_1\mathbf{G}_1\right)^{-1}\right]\right\} \quad (74)$$

$$\hat{R}_{p,k}^{mrc} = \log_2\left(1+\left(\mathbb{E}\left\{\frac{\sum_{i=1}^{K}\beta_{1i}|\tilde{g}_{1i}|^2+\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}|\tilde{g}_{li}|^2+\frac{N_0}{E_s}}{\beta_{1k}||\mathbf{g}_{1k}||^2}\right\}\right)^{-1}\right) \tag{73}$$

$$\hat{R}_{im,k}^{mrc} = \log_2\left(1+\left(\mathbb{E}\left\{\frac{\sum_{i=1}^{K}\beta_{1i}|\tilde{\mathbf{e}}_{1i}|^2+\sum_{\substack{i=1\\i\neq k}}^{K}\beta_{1i}|\tilde{\hat{\mathbf{g}}}_{1i}|^2+\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}|\tilde{\mathbf{g}}_{li}|^2+\frac{N_0}{E_s}}{\beta_{1k}||\hat{\mathbf{g}}_{1k}||^2}\right\}\right)^{-1}\right) \tag{77}$$

$$\hat{R}_{im,k}^{zf} = \log_2\left(1+\frac{\beta_{1k}}{\left(\sum_{i=1}^{K}\beta_{1i}\sigma_{eiu}^2+\sum_{l=2}^{L}\sum_{i=1}^{K}\beta_{li}+\frac{N_0}{E_s}\right)\mathbb{E}\left\{||\hat{\mathbf{a}}_k||^2\right\}}\right). \tag{78}$$

As $\mathbf{G}_1$ follows zero mean complex Gaussian distribution, $\mathbb{E}\left\{tr\left(\mathbf{G}_1^H\mathbf{G}_1\right)^{-1}\right\} = \frac{K}{M-K}$ forms a $K \times K$ central complex Wishart matrix with $M$ $(M > K)$ degrees of freedom [34]. Therefore,

$$\mathbb{E}\left\{\mathbf{a}_k^H\mathbf{a}_k\right\} = \frac{1}{M-K} \tag{75}$$

for $M \geq K+1$. In (42), we replace $\mathbb{E}\left\{|\mathbf{a}_k^H\mathbf{g}_{li}|^2\right\}$ and $\mathbb{E}\left\{\mathbf{a}_k^H\mathbf{a}_k\right\}$ with $\frac{1}{M-K}$ to obtain the analytical expression for lower bound on the UL achievable rate $R_{p,k}^{zf}$ in (43).

## APPENDIX C
## IMPERFECT CSI: UL ACHIEVABLE RATE

From (45), we obtain the desired signal power as $E_s\beta_{1k}|\hat{\mathbf{a}}_k^H\hat{\mathbf{g}}_{1k}|^2$. We represent the power of inter-cell interference and intra-cell interference, CSI estimation error and additive noise as

$$Int_k = E_s\beta_{1k}|\hat{\mathbf{a}}_k^H\mathbf{e}_{1k}|^2 + \sum_{i=1,i\neq k}^{K} E_s\beta_{1i}|\hat{\mathbf{a}}_k^H\hat{\mathbf{g}}_{1i}|^2 + \sum_{l=2}^{L}\sum_{i=1}^{K} E_s\beta_{li}|\hat{\mathbf{a}}_k^H\hat{\mathbf{g}}_{li}|^2 + N_0||\hat{\mathbf{a}}_k^H||^2. \tag{76}$$

We obtain ergodic UL achievable rate of $k^{th}$ user in (46) by considering the $Int_k$ term as worst-case uncorrelated additive noise [33].

## APPENDIX D
## IMPERFECT CSI: UL ACHIEVABLE RATE WITH MRC RECEIVER

By central limit theorem, elements of the proposed estimate $\hat{\mathbf{g}}_{1k}$ can be approximated as a complex Gaussian distribution with $\mathcal{CN}\left(0,\sigma_{iu}^2\right)$. Therefore, $\hat{R}_{im,k}^{mrc}$ of (48) can be rewritten as (77) which is shown at the top of this page, where $\tilde{e}_{1i} = \frac{\hat{\mathbf{g}}_{1k}^H\mathbf{e}_{1i}}{||\hat{\mathbf{g}}_{1k}||}$, $\tilde{\hat{g}}_{1i} = \frac{\hat{\mathbf{g}}_{1k}^H\hat{\mathbf{g}}_{1i}}{||\hat{\mathbf{g}}_{1k}||}$ and $\tilde{g}_{1i} = \frac{\hat{\mathbf{g}}_{1k}^H\mathbf{g}_{li}}{||\hat{\mathbf{g}}_{1k}||}$.

The $\tilde{\mathbf{e}}_{1i}$ is a Gaussian random variable conditioned on $\hat{\mathbf{g}}_{1k}$ with zero mean and $\sigma_{eiu}^2$ variance. Similarly, $\tilde{\hat{\mathbf{g}}}_{1i}$ and $\tilde{\mathbf{g}}_{1i}$ has zero mean with variance $\sigma_{iu}^2$ and one, respectively for $u \in \{p,d\}$. By substituting these values in (77), we obtain the lower bound on UL achievable rate $\hat{R}_{im,k}^{mrc}$ in (49).

## APPENDIX E
## IMPERFECT CSI: UL ACHIEVABLE RATE WITH ZF RECEIVER

As $\mathbf{e}_{1i}$ and $\hat{\mathbf{g}}_{1k}$ are independent of each other, $|\hat{\mathbf{a}}_k^H\mathbf{e}_{1i}|^2$ of (51) becomes $||\hat{\mathbf{a}}_k||^2\sigma_{eiu}^2$. Thus, UL achievable rate $\hat{R}_{im,k}^{zf}$ of (51) reduces to (78) which is shown at the top of this page.

Moreover, $\mathbb{E}\left\{||\hat{\mathbf{a}}_k||^2\right\}$ can be written as

$$\mathbb{E}\left\{\left[\left(\hat{\mathbf{G}}_1^H\hat{\mathbf{G}}_1\right)^{-1}\right]_{kk}\right\} = \frac{1}{K}\mathbb{E}\left\{tr\left[\left(\hat{\mathbf{G}}_1^H\hat{\mathbf{G}}_1\right)^{-1}\right]\right\} = \frac{1}{(M-K)\,\sigma_{ku}^2}. \tag{79}$$

By using complex Wishart matrix property of (79) in (78), we obtain the resultant lower bound of UL achievable rate with ZF Rx in (52).

## REFERENCES


[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[5] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[6] A. Khansefid and H. Minn, "On channel estimation for massive MIMO with pilot contamination," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1660–1663, Sep. 2015.

[7] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[8] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 4626–4631.

[9] T. E. Bogale and L. Bao Le, "Pilot optimization and channel estimation for multiuser massive MIMO systems," in *Proc. 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2014, pp. 1–6.

[10] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 3249–3252.

[11] C.-K. Wen, S. Jin, K.-K. Wong, J.-C. Chen, and P. Ting, "Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1356–1368, Mar. 2015.
[12] M. Marques da Silva, R. Dinis, and J. Guerreiro, "Implicit pilots for an efficient channel estimation in simplified massive MIMO schemes with precoding," *Int. J. Antennas Propag.*, vol. 2019, pp. 1–11, Jan. 2019.
[13] E. O. Asiedu and R. Ahmed, "Spectral efficiency analysis and achievable rates in massive MIMO systems with LS channel estimation," in *Proc. 6th ACM/ACIS Int. Conf. Appl. Comput. Inf. Technol. (ACIT)*, 2018, pp. 43–48.
[14] A. S. Alwakeel and A. H. Mehana, "Data-aided channel estimation for multiple-antenna users in massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10752–10760, Nov. 2019.
[15] Z. Gao, L. Dai, S. Han, C.-L. I, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, Jun. 2018.
[16] A. Liao, Z. Gao, H. Wang, S. Chen, M.-S. Alouini, and H. Yin, "Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8329–8345, Dec. 2019.
[17] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing based adaptive active user detection and channel estimation: Massive access meets massive MIMO," 2019, *arXiv:1906.09867*. [Online]. Available: http://arxiv.org/abs/1906.09867
[18] A. Wang, R. Yin, and C. Zhong, "Channel estimation for uniform rectangular array based massive MIMO systems with low complexity," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2545–2556, Mar. 2019.
[19] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.
[20] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.
[21] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in Sub-6 GHz and mmWave: Physical, practical, and use-case differences," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 100–108, Apr. 2019.
[22] S. Buzzi and C. D'Andrea, "Massive MIMO 5G cellular networks: Mmwave vs. $\mu$-wave frequencies," 2017, *arXiv:1702.07187*. [Online]. Available: http://arxiv.org/abs/1702.07187
[23] D. Neumann, M. Joham, and W. Utschick, "Channel estimation in massive MIMO systems," 2015, *arXiv:1503.08691*. [Online]. Available: http://arxiv.org/abs/1503.08691
[24] D. Hu, L. He, and X. Wang, "Semi-blind pilot decontamination for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 525–536, Jan. 2016.
[25] R. R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.
[26] E. Nayebi and B. D. Rao, "Semi-blind channel estimation for multiuser massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 540–553, Jan. 2018.
[27] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.
[28] J. Shen and M. Wu, "A MAP-based SAGE channel estimation and data detection joint algorithm for MIMO system," in *Proc. IEEE 2nd Int. Conf. Softw. Eng. Service Sci.*, Jul. 2011, pp. 82–85.
[29] T.-H. Pham, A. Nallanathan, and Y.-C. Liang, "Joint channel and frequency offset estimation in distributed MIMO flat-fading channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 648–656, Feb. 2008.
[30] K. Mawatwal, D. Sen, and R. Roy, "A semi-blind channel estimation algorithm for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 70–73, Feb. 2017.
[31] M. Moeneclaey, "On the true and the modified cramer-rao bounds for the estimation of a scalar parameter in the presence of nuisance parameters," *IEEE Trans. Commun.*, vol. 46, no. 11, pp. 1536–1544, Nov. 1998.
[32] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, 1st ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[33] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
[34] P. SinghParihar, R. Saraswat, and S. Maheshwari, "Energy and spectral efficiency of very large multiuser MIMO systems," *Int. J. Comput. Appl.*, vol. 111, no. 5, pp. 4–7, Feb. 2015.
[35] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
[36] D. Chu, "Polyphase codes with good periodic correlation properties (Corresp)," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 531–532, Jul. 1972.
[37] P. Jung Chung and J. F. Bohme, "Comparative convergence analysis of EM and SAGE algorithms in DOA estimation," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 2940–2949, Dec. 2001.
[38] N. Aerts, I. Avram, J. Van Hecke, H. Bruneel, and M. Moeneclaey, "Iterative SAGE-based channel estimation in a block fading amplify-and-forward relaying network," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1742–1753, Apr. 2014.
[39] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

**KHUSHBOO MAWATWAL** (Student Member, IEEE) received the B.Tech. degree in electronics and telecommunication engineering from the Biju Patnaik University of Technology, Rourkela, India, in 2011, and the M.Tech. degree in communication and networks from the National Institute of Technology, Rourkela, in 2014. She is currently pursuing the Ph.D. degree with the Department of Electronics and Electrical Communication Engineering, IIT Kharagpur, India, under the joint supervision of Prof. R. Roy and Prof. D. Sen. Her research interests include the areas of channel estimation and resource allocation for massive MIMO systems.

**DEBARATI SEN** (Senior Member, IEEE) received the Ph.D. degree in telecommunication engineering from IIT Kharagpur, Kharagpur, India. She is currently an Associate Professor with IIT Kharagpur. Her research interests include wireless and optical communication systems, mostly on MB-OFDM, synchronization, equalization, UWB, BAN, green communications, 60-GHz communications, and baseband algorithm design for coherent optical communications. She was a Postdoctoral Research Fellow of the Chalmers University of Technology, Sweden, and a Senior Chief Engineer with the Samsung Research and Development Institute India, Bengaluru (SRI-B), India. She is an editorial board member of two international journals. She was the recipient of the Best Paper Award from the Samsung Technology Conference, in 2010, and the IEI Young Engineers Award, in 2010.

**RAJARSHI ROY** (Senior Member, IEEE) received the B.E. degree (Hons.) in electronics and telecommunication engineering from Jadavpur University, Kolkata, India, in 1992, the M.Sc. degree in engineering from the Department of Electrical Communication Engineering, IISc, Bengaluru, India, in 1995, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Polytechnic University, Brooklyn, NY, USA (currently known as New York University, Tandon School of Engineering), in 2001. He is currently working as an Associate Professor with the Department of Electronics and Electrical Communication Engineering, IIT Kharagpur, Kharagpur, India. He has served as a Summer Ph.D. Intern at Lucent Bell labs, Holmdel, NJ, USA, in 1997, and as an Adjunct Teacher at Polytechnic University, Brooklyn, NY, USA, in Summer 2000. He served as an Academic Visitor with the Helsinki University of Technology, in 2004, and the Indian Statistical Institute Kolkata as a Visiting Scientist, in 2001. He was with Comverse, MA, USA, from 2000 to 2001, and Lucent, Bengaluru, in 2002, as a Research and Development Engineer. His research interests include queuing theory, Markov decision theory, 5G/6G wireless communications, complex communication networks, MIMO, network coding, cooperative communication, cyber-physical systems, cognitive technology, and network science. He serves on the Editorial Board of *IETE Technical Review* and the *IETE Journal of Research*.