Impact of Pilot Overhead and Channel Estimation on the Performance of Massive MIMO

Atchutaram K. Kocharlakota, *Student Member, IEEE*, Karthik Upadhya, *Member, IEEE*, and Sergiy A. Vorobyov, *Fellow, IEEE*

***Abstract*— This paper studies the impact of additional pilot overhead for covariance matrix estimation in a time-division duplexed (TDD) massive multiple-input multiple-output (MIMO) system. We choose average uplink (UL) and downlink (DL) spectral efficiencies (SEs) as performance metrics for the massive MIMO system, and derive closed form expressions for them in terms of the additional pilot overhead. The expressions are derived by considering linear minimum mean squared error (LMMSE)-type and element-wise LMMSE-type channel estimates that represent LMMSE and element-wise LMMSE with estimated covariance matrices, respectively. Using these expressions, a detailed theoretical analysis of SE behavior as a function of pilot overhead for both LMMSE-type and element-wise LMMSE-type channel estimation are presented, followed by simulations, which also demonstrate and validate theoretical results.**

***Index Terms*— Spectral efficiency, massive multiple-input multiple-output (MIMO), covariance estimation, channel estimation, pilot contamination.**

I. Introduction

A MULTI-USER massive multiple-input multiple-output (MIMO) system comprises multiple cells, each having a base station (BS) with a large number of antennas (hundreds) to serve multiple users (tens) within the cell. It is considered to be one of the key technologies for the fifth-generation (5G) cellular systems due to the considerable improvement in spectral efficiency (SE) through spatial multiplexing [1]–[5] achieved with low computational complexity [1], [6], [7]. However, acquiring channel state information (CSI) at the base station (BS) is essential to realize the benefits of a massive MIMO system.

We consider a time-division duplexing (TDD) massive MIMO system where the CSI is acquired through uplink (UL) pilots. In time-variant channels, the channels in two different coherence blocks, which is a collection of symbols within a coherence time and bandwidth, are uncorrelated. Consequently, the channel has to be estimated in each coherence block. The number of orthogonal pilots available for channel estimation in a coherence block is limited by the number of available symbols in the coherence block that are not reserved for UL data and DL data, and as a result, UL pilot sequences need to be reused by users across the cells, causing the pilot contamination problem [1], [8], [9].

Despite the presence of pilot contamination, under the assumption that the covariance matrices of interfering users are asymptotically linearly independent to each other, the sum rate of the massive MIMO system has been proven to be unbounded [10]. However, the authors assume that contamination-free covariance matrices of individual users are available at the BS, while, in practice, these covariance matrices also have to be estimated at the BS. Covariance matrix estimation in a multi-cell TDD massive MIMO system is a non-trivial task because the channel estimates from which the covariance matrix estimates are obtained are themselves contaminated. Naively utilizing the contaminated channel estimates in a sample covariance estimator will result in the target user covariance matrix estimate containing the covariance matrices of the interference users. The algorithm that estimates the target covariance matrix in such a setup needs additional information from the users to isolate the target user covariance from the contaminated covariance; this is typically done using additional pilots.

Methods for estimating the individual covariance matrices in the presence of pilot contamination have been recently studied in [11]–[16]. In all these works, the authors assume that the channel covariance matrices are constant across multiple coherence blocks, and then, the observations from a few of these coherence blocks are used to estimate the covariance matrices. In [11], the authors first estimate the angle-delay power spread function from the contaminated channel estimates of multiple coherence blocks, then use this function for supervised/unsupervised clustering of the multipath components belonging to the target user. Finally, they use the clusters to estimate the spatial covariance matrix of the target user. In [14], the authors develop a method where the pilot allocation is changed in each coherence block. The channel estimates obtained from these blocks are then used to obtain a maximum-likelihood estimate of the contamination-free covariance matrix. Work [15] presents two methods which avoid contamination in the covariance matrices by utilizing dedicated orthogonal pilots for each user for estimating its individual spatial covariance matrix. In [16], a new pilot structure and a covariance matrix estimation method are developed that offer higher throughput and lower mean squared

Manuscript received March 12, 2021; revised August 1, 2021; accepted September 1, 2021. Date of publication September 13, 2021; date of current version December 16, 2021. This work was supported by the Academy of Finland under grant 319822. The associate editor coordinating the review of this article and approving it for publication was A. García Armada. *(Corresponding author: Sergiy A. Vorobyov.)*

Atchutaram K. Kocharlakota and Sergiy A. Vorobyov are with the Department of Signal Processing and Acoustics, Aalto University, 00076 Aalto, Finland (e-mail: kameswara.kocharlakota@aalto.fi; svor@ieee.org).

Karthik Upadhya was with the Department of Signal Processing and Acoustics, Aalto University, 00076 Aalto, Finland. He is now with Nokia Bell Labs, 02610 Espoo, Finland (e-mail: karthik.upadhya@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2021.3112213.

Digital Object Identifier 10.1109/TCOMM.2021.3112213

error (MSE) of the channel estimates compared to the method in [15]. Although [16] requires additional pilots for estimating the individual covariance matrices of each user, it does not assume any additional structures on the covariance matrices of the users unlike [11]–[13], and it does not require backhaul communication between the neighboring cells unlike [14]. Moreover, since the additional pilots in [16] are not dedicated to each user as in [15], the number of additional pilots in [16] does not grow with the total number of users in the entire system. Therefore, in this paper, we choose [16] to study the performance of covariance estimation method in a massive MIMO system. In particular, we emphasize the impact of pilot overhead or choice of channel estimation method on the performance of a massive MIMO system.

Utilization of the estimated covariance matrices for channel estimation results in a trade-off for the system performance. Indeed, increase of the number of additional pilots for estimating the covariance matrices will not only improve the quality of the covariance estimate (and hence, the channel estimate) but also increase the pilot overhead. Consequently, choosing the additional pilot overhead related to estimating the covariance matrices becomes a key trade-off problem for the system performance.

Except [11], in all the covariance estimation papers mentioned above, the authors derive SE expressions corresponding to a single realization of covariance estimate. Such an SE is achievable for a practical receiver which does not have perfect covariance information. However, one can notice, these papers use a numerically computed average SE as a performance metric of covariance estimation method in a massive MIMO system. Therefore, in this paper, we utilize the SE value that is averaged over multiple realizations (ensemble average) of the covariance estimates as a performance metric of the covariance estimation method in [16]. We first derive average SE expressions for two types of channel estimation methods namely: (1) LMMSE-type and (2) element-wise LMMSE-type channel estimation methods, that use estimated covariance matrices.[1] Note that, in this paper, we use *LMMSE-type/element-wise LMMSE-type* to denote the channel estimation with estimated covariance matrices, and *LMMSE/element-wise LMMSE* to denote channel estimation with true covariance matrices. Using the derived expressions, we demonstrate the impact of pilot overhead and channel estimation method on the performance of the massive MIMO system with imperfect covariance information at BSs.

The following are the contributions of this paper.

- We first derive closed-form expressions for the average UL and DL spectral efficiencies when the LMMSE-type and element-wise LMMSE-type channel estimates are used in a maximum ratio combiner (MRC) UL receiver, and in a matched filter precoding DL transmitter.
- Using theoretical and simulation studies on the derived SE expressions, we establish the fact that the number of additional pilots for covariance estimation is a key trade-off parameter that needs to be chosen optimally.
- Using these expressions, we then quantitatively compare the performance of the element-wise LMMSE-type channel estimate with the LMMSE-type channel estimate. To the best of our knowledge, this is the first work that quantitatively compares the average UL/DL SE obtained with LMMSE-type and element-wise LMMSE-type estimates.

The paper is organized as follows. In Section II, we describe the system model along with a brief review of the channel estimation and covariance matrices estimation methods we want to study. Section III reports the derivations of the closed-form expressions for the UL and DL SEs for both the channel estimation techniques described above. In Section IV, we present a detailed theoretical discussion on the impact of pilot overhead and channel estimation technique on massive MIMO. In Section V we provide the simulation results that demonstrate the theoretical conclusions made in Section IV. We conclude this work in Section VI. Technical proofs of lemmas and theorems in the paper appear in appendices at the end of the paper.

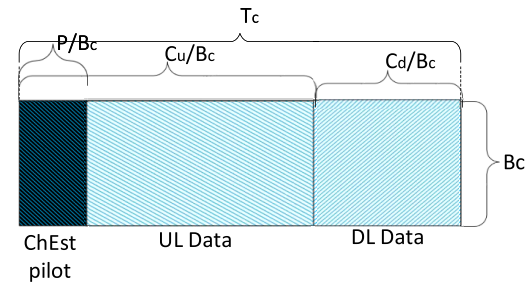
Notation: We use boldface capital letters for matrices, and boldface lowercase letters for vectors. The superscripts $(\cdot)^*$, $(\cdot)^\intercal$, and $(\cdot)^H$ denote element-wise conjugate, transpose, and Hermitian transpose operations, respectively. Moreover, $\mathcal{CN}(\mathbf{m}, \mathbf{R})$ denotes (circularly symmetric) complex Gaussian random vector with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{R}$, while $\mathcal{W}(N, \mathbf{R})$ denotes Wishart random matrix with N degrees of freedom and $\mathbf{R}$ is the covariance matrix that corresponds to underlying Gaussian random vectors. In addition, $\mathcal{U}[x_1, x_2]$ stands for the uniform distribution between x_1 and x_2. The element in i^{th} row and j^{th} column of the matrix $\mathbf{A}$ is denoted as $[\mathbf{A}]_{ij}$, $\mathbf{I}$ stands for an identity matrix (of appropriate size), $\mathrm{diag}(\mathbf{A})$ is a diagonal matrix whose diagonal elements are same as the diagonal elements of the matrix $\mathbf{A}$. We use $\mathrm{tr}(\cdot)$ to denote trace of a matrix, $\|\cdot\|$ to denote l_2 norm of a vector or a matrix, i.e., Frobenius norm, and $\mathbb{E}\{\cdot\}$ stands for the mathematical expectation. Finally, the symbol δ_{ij} is the Kronecker delta such that $\delta_{ij} = 1$ if $i = j$, and 0 otherwise.

II. System Model

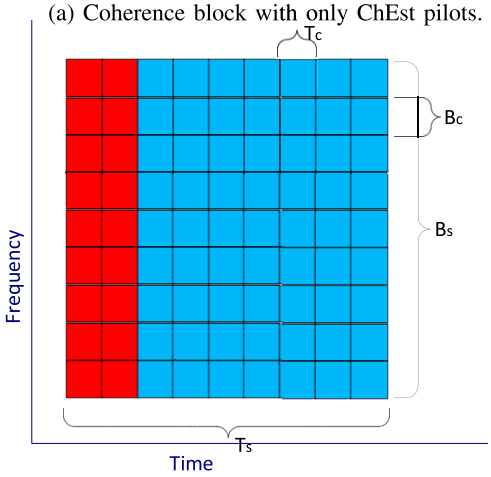
We consider a massive MIMO system with L cells, each having a BS with M antennas and serving K single-antenna users. We make a realistic assumption that the channels between users and BSs are spatially correlated [18]. The channel between user (l, k) (k^{th} user in l^{th} cell) and BS in the j^{th} cell is denoted as $\mathbf{h}_{jlk} \in \mathbb{C}^M$ and is assumed to be distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{R}_{jlk})$, where $\mathbf{R}_{jlk} \triangleq \mathbb{E}\{\mathbf{h}_{jlk}\mathbf{h}_{jlk}^H\}$ is the spatial covariance matrix. We consider the block-fading model where the channel is assumed to be constant over the coherence bandwidth B_c and coherence time T_c. In other words, the channel is assumed to be constant over a coherence block containing $\tau_c = B_c T_c$ symbols.

We consider TDD transmission and each coherence block is divided into slots for UL pilots, UL and DL data. The number of data symbols in the UL and DL time slot is denoted as C_u and C_d, respectively. The channel is assumed to be reciprocal, i.e., the DL channel between BS j and user (l, k) can be written as $\mathbf{h}_{jlk}^*$. This is represented in Fig. 1(a).

[1] Some preliminary results are also reported in [17].



(a) Coherence block with only ChEst pilots.



(b) Grid of coherence blocks with coherent covariance matrices.

Fig. 1. Time frequency grid and pilot positioning.

We consider two types of UL pilots, namely, (i) pilots for estimating the channel (also referred to as ChEst pilots) and (ii) pilots for estimating the covariance matrix (referred to as CovEst pilots). Both ChEst pilots and CovEst pilots are assumed to be of length P symbols.

The spatial covariance matrices are assumed to be constant over a considerably longer time-interval and bandwidth than a single coherence block [11], [14]–[16], [19].[2] Specifically, we assume that the covariance matrices are coherent over the time-interval T_s and system bandwidth B_s, which implies that they can be assumed to be constant over $\tau_s = B_s T_s/\tau_c$ coherence blocks (usually several tens of thousands of blocks in practice). This time-frequency grid over which the second-order statistics of the channel are assumed to be constant is illustrated in Fig. 1(b).

Each of the τ_s coherence blocks contain ChEst pilots for channel estimation, whereas only N_R out of the τ_s coherence

[2]Note that, according to [19], this assumption is valid of urban and rural environment. Howeever, this is questionable for indoor scenarios.

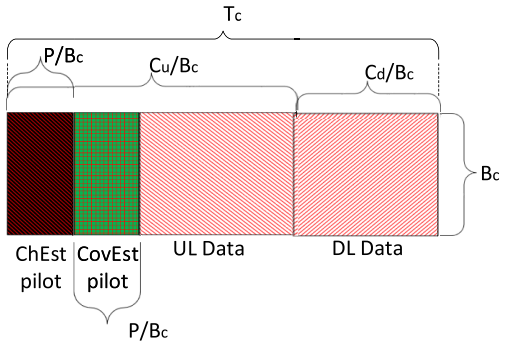


Fig. 2. Coherence block with additional CovEst pilots.

blocks contain CovEst pilots in addition to the ChEst pilots (as can be seen in Fig. 1(b)). The coherence blocks that contain the CovEst pilots are depicted in Fig. 2.

The UL received signal, $\mathbf{Y}_j[n] \in \mathbb{C}^{M\times C_u}$, in the n^{th} coherence block at BS j is given as

$$\mathbf{Y}_j[n] = \sum_{l=1}^{L}\sum_{k=1}^{K} \sqrt{\mu}\mathbf{h}_{jlk}\mathbf{x}_{lk}^{\mathsf{T}}[n] + \mathbf{N}_j[n] \tag{1}$$

where $\mathbf{x}_{lk} \in \mathbb{C}^{C_u}$ is the signal transmitted by user (l,k), $\mathbf{N}_j \in \mathbb{C}^{M\times C_u}$ is the additive white Gaussian noise (AWGN) at the BS, and μ is the UL transmit power. The transmitted data $\mathbf{x}_{lk}$ is assumed to be distributed as $\mathbf{x}_{lk} \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$ whereas the elements of $\mathbf{N}_j$ are assumed to be identically and independently distributed (i.i.d) as $\mathcal{CN}(0,1)$.

In the DL, the received signal $\mathbf{z}_{ju}[n] \in \mathbb{C}^{C_d}$ at user (j,u) in the n^{th} coherence block can be written as

$$\mathbf{z}_{ju}[n] = \sum_{l=1}^{L}\sum_{k=1}^{K} \sqrt{\lambda}(\mathbf{h}_{jlu}^{H}\mathbf{b}_{lk})\mathbf{d}_{lk}[n] + \mathbf{e}[n]$$

where $\mathbf{d}_{lk} \in \mathbb{C}^{C_d}$ is the payload data from BS l to its user (l,k), $\mathbf{b}_{lk} \in \mathbb{C}^{M}$ is the corresponding precoding vector normalized such that the average transmitted power is λ, i.e., $\mathbb{E}\{\|\mathbf{b}_{lk}\|^2\} = 1$, and $\mathbf{e} \in \mathbb{C}^{C_d}$ is the AWGN noise distributed as $\mathcal{CN}(\mathbf{0},\mathbf{I})$.

A. Channel Estimation

A dedicated set of P ($\geq K$) symbols is allocated to UL pilots for channel estimation in each coherence block, as shown in Figs. 1(a) and 2. In other words, let $\mathbf{p}_k \in \mathbb{C}^P$ denote the ChEst pilot sequence used by the k^{th} user in any of the L cells. Then, for another user m in the same cell, we have $\mathbf{p}_k^H\mathbf{p}_m = P\delta_{km}$. We assume that the same P pilots are reused in each cell and each user is randomly allocated one of these pilots for channel estimation.

The pilot transmissions in all cells are assumed to be synchronized. Then, the received signal at BS j during pilot transmissions in the n^{th} coherence block (denoted as $\mathbf{Y}_j^{(p)}[n]$) can be written as

$$\mathbf{Y}_j^{(p)}[n] = \sum_{l=1}^{L}\sum_{k=1}^{K} \sqrt{\mu}\mathbf{h}_{jlk}\mathbf{p}_k^{\mathsf{T}} + \mathbf{N}_j^{(p)}[n] \tag{2}$$

where $\mathbf{N}_j^{(p)}[n] \in \mathbb{C}^{M\times P}$ is the noise during pilot transmission.

We consider LMMSE and element-wise LMMSE techniques for estimating $\mathbf{h}_{jlk}$ from the observed signal $\mathbf{Y}_j^{(p)}$ given in (2). In what follows, we first review these estimation techniques.

1) LMMSE Channel Estimation: From (2), the least-squares (LS) channel estimate of user (j,u) at BS j in the n^{th} coherent block (denoted as $\hat{\mathbf{h}}_{jju}^{LS}[n]$) can be obtained as follows[3]

$$\begin{aligned}\hat{\mathbf{h}}_{jju}^{LS}[n] &= \arg\min_{\mathbf{g}} \ \|\mathbf{Y}_j^{(p)}[n] - \sqrt{\mu}\mathbf{g}\mathbf{p}_u^{\mathsf{T}}\|^2 \\ &= \mathbf{h}_{jju} + \sum_{l\neq j}\mathbf{h}_{jlu} + \frac{1}{P\sqrt{\mu}}\mathbf{N}_j^{(p)}[n]\mathbf{p}_u^*. \end{aligned} \quad (3)$$

Using this LS channel estimate that serves as a sufficient statistic for $\mathbf{h}_{jju}$, the resultant LMMSE estimate can be easily derived to be [15]

$$\begin{aligned}\hat{\mathbf{h}}_{jju}^{LMMSE}[n] &= \mathbf{R}_{jju}\mathbf{Q}_{ju}^{-1}\hat{\mathbf{h}}_{jju}^{LS}[n] \\ \mathbf{Q}_{ju} &\triangleq \mathbb{E}\{\hat{\mathbf{h}}_{jju}^{LS}[n](\hat{\mathbf{h}}_{jju}^{LS}[n])^H\} = \sum_{l=1}^{L}\mathbf{R}_{jlu} + \frac{1}{P\mu}\mathbf{I}. \end{aligned} \quad (4)$$

Although the channel estimates in the above equation assume that the covariance information is known, in practice it has to be estimated at the BS. Therefore, it is reasonable to replace these matrices with estimated covariance matrices ($\hat{\mathbf{R}}_{jju}$, and $\hat{\mathbf{Q}}_{ju}$) to get LMMSE-type channel estimate as follows

$$\hat{\mathbf{h}}_{jju}[n] = \hat{\mathbf{R}}_{jju}\hat{\mathbf{Q}}_{ju}^{-1}\hat{\mathbf{h}}_{jju}^{LS}[n] \quad (5)$$

For known covariance case, the computational complexity in evaluating (4) is $\mathcal{O}(M^3)$. Furthermore, the computational complexity of a sample covariance matrix of an $M\times 1$ channel vector is $\mathcal{O}(NM^2)$, where N is the number of samples. Therefore, the total computational complexity involved in evaluating (5) is $\mathcal{O}(M^3 + M^2\ N_R + M^2\ N_Q)$, where N_R and N_Q are the number of pilot sequences (samples) used for computing $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$, respectively.

2) Element-Wise LMMSE Channel Estimation: An alternative approach for LMMSE channel estimation is to use the element-wise LMMSE estimate; this technique requires a fewer number of samples/pilots for the covariance estimation that does not grow with M [10].

The element-wise LMMSE estimate of the channel can be obtained as

$$[\hat{\mathbf{h}}_{jju}^{\mathrm{el-LMMSE}}[n]]_p = \frac{[\mathbf{S}_{jju}]_{pp}}{[\mathbf{P}_{ju}]_{pp}}[\hat{\mathbf{h}}_{jju}^{LS}[n]]_p, \quad p\in\{1,\ldots,M\}$$

where $\mathbf{S}_{jju} \triangleq \mathrm{diag}(\mathbf{R}_{jju})$ and $\mathbf{P}_{ju} \triangleq \mathrm{diag}(\mathbf{Q}_{ju})$. The element-wise LMMSE-type estimate with estimated covariance matrices ($\hat{\mathbf{S}}_{jju}$, and $\hat{\mathbf{P}}_{ju}$) can be written as

$$[\hat{\mathbf{h}}_{jju}^{\mathrm{el}}[n]]_p = \frac{[\hat{\mathbf{S}}_{jju}]_{pp}}{[\hat{\mathbf{P}}_{ju}]_{pp}}[\hat{\mathbf{h}}_{jju}^{LS}[n]]_p, \quad p\in\{1,\ldots,M\} \quad (6)$$

[3]As the channel observations, in this case, are linear measurements in gaussian noise, one should note that this is also an MMSE estimator.

Here, each diagonal element of $\hat{\mathbf{S}}_{jju}$ ($\hat{\mathbf{P}}_{ju}$) is computed using a sample variance estimator of the corresponding element of the channel vector. If we use N_R (N_Q) number of channel samples for estimating $\hat{\mathbf{S}}_{jju}$ ($\hat{\mathbf{P}}_{ju}$), the computational complexity involved in evaluating each element of $\hat{\mathbf{h}}_{jju}^{\mathrm{el}}[n]$ is $\mathcal{O}(N_R+N_Q)$. Therefore, the total computational complexity involved in evaluating $\hat{\mathbf{h}}_{jju}^{\mathrm{el}}$ is $\mathcal{O}(MN_R+MN_Q)$. Element-wise LMMSE-type channel estimation substantially reduces the computational complexity at the cost of some performance degradation caused due to the fact that we ignore non-diagonal elements of $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$. Later in Section V, we compare the performance of these two channel estimation methods using simulations.

B. Covariance Matrix Estimation

Several methods to address the covariance matrix estimation problem have been proposed in literature [11], [14]–[16]. However, among these methods, only the estimators in [15] and [16] are in closed-form and consequently, lend themselves to mathematical analysis. Moreover, since [16] is seen to outperform [15], we select the estimator in [16] for performance analysis when the estimate is used for both LMMSE-type and element-wise LMMSE-type channel estimation. We assume that the BSs have knowledge of the random phase sequences.

In this subsection, we briefly review the pilot structure introduced in [16] and the corresponding spatial covariance estimation method. The objective is to compute a pair of $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$ (or $\hat{\mathbf{S}}_{jju}$ and $\hat{\mathbf{P}}_{ju}$) for each set of τ_s contiguous coherence blocks.

To obtain $\hat{\mathbf{Q}}_{ju}$, since the matrix $\mathbf{Q}_{ju}$ is defined as the covariance matrix of $\hat{\mathbf{h}}_{jju}^{LS}[n]$, we use these LS channel estimates from multiple coherence blocks in a sample covariance estimator. We use a set of N_Q ($\geq M$) number of LS estimates for computing $\hat{\mathbf{Q}}_{ju}$. Therefore, we have the following unbiased covariance estimator $\hat{\mathbf{Q}}_{ju} = \frac{1}{N_Q}\sum_{n=1}^{N_Q}\hat{\mathbf{h}}_{jju}^{LS}[n](\hat{\mathbf{h}}_{jju}^{LS}[n])^H$. Similarly, the unbiased estimate of $\mathbf{P}_{ju}$ is obtained using a sample covariance estimator as follows

$$[\hat{\mathbf{P}}_{ju}]_{pp} = \frac{1}{N_Q}\sum_{n=1}^{N_Q}|[\hat{\mathbf{h}}_{jju}^{LS}[n]]_p|^2, \quad \forall p\in 1\ldots M.$$

For estimating $\mathbf{R}_{jju}$ and $\mathbf{S}_{jju}$, as depicted by the red coherence blocks in Fig. 1(b), each user transmits an additional pilot sequence of length P symbols for N_R out of the τ_s coherence blocks. Specifically, the CovEst pilots, denoted as $\{\boldsymbol{\phi}_{lk}[n]\}_{n=1}^{N_R}$, are transmitted by the user (l,k), with the pilot sequence in n^{th} coherence block given as a phase-shifted version of the ChEst pilot, i.e., $\boldsymbol{\phi}_{lk}[n] = e^{j\theta_{ln}}\mathbf{p}_k$. The phase-shifts $\{\theta_{ln}\}_{n=1}^{N_R}$ are (pseudo-)random and are generated such that $\{\theta_{ln}\}_{n=1}^{N_R}$ is independent of the channel vectors and satisfies $\mathbb{E}\{e^{j\theta_{ln}}\} = 0$ [16]. A random sequence that satisfies both these properties is $\theta_{ln}\sim\mathcal{U}[0,2\pi)$. Furthermore, the random phase sequences are assumed to be i.i.d across cells.

Now, let $\mathbf{Y}_j^{(r)}[n]$ be the received signal when the users transmit the CovEst pilots $\boldsymbol{\phi}_{ju}[n]$. Then, $\mathbf{Y}_j^{(r)}[n]$ can be

written as

$$\mathbf{Y}_j^{(r)}[n] = \sum_{l=1}^{L}\sum_{k=1}^{K}\sqrt{\mu}\mathbf{h}_{jlk}\boldsymbol{\phi}_{lk}^{\mathsf{T}}[n] + \mathbf{N}_j^{(r)}[n] \tag{7}$$

where $\mathbf{N}_j^{(r)}[n]$ is the AWGN noise at the BS that has the same statistics as $\mathbf{N}_j^{(p)}[n]$.

We denote LS channel estimates obtained from the pilots $\mathbf{p}_u$ and $\boldsymbol{\phi}_{ju}$ as $\hat{\mathbf{h}}_{jju}^{(1)}[n]$ and $\hat{\mathbf{h}}_{jju}^{(2)}[n]$, respectively. Using (2) and (7) and by using the fact that $\boldsymbol{\phi}_{lk}[n] = e^{j\theta_{ln}}\mathbf{p}_k$, the LS estimates can be straightforwardly obtained as [16]

$$\hat{\mathbf{h}}_{jju}^{(1)}[n] = \mathbf{h}_{jju} + \sum_{l\neq j}\mathbf{h}_{jlu} + \frac{1}{P\sqrt{\mu}}\mathbf{N}_j^{(p)}[n]\mathbf{p}_u^* \tag{8}$$

$$\hat{\mathbf{h}}_{jju}^{(2)}[n] = \mathbf{h}_{jju} + \sum_{l\neq j}\mathbf{h}_{jlu}e^{j(\theta_{ln}-\theta_{jn})} + \frac{1}{P\sqrt{\mu}}\mathbf{N}_j^{(r)}[n]\mathbf{p}_u^* e^{-j\theta_{jn}} \tag{9}$$

In the following subsections, we describe both cases of complete and diagonal matrix estimation using the aforementioned LS channel estimates.

1) Estimation of $\hat{\mathbf{R}}_{jju}$: Note that the second and third terms in (8), corresponding to the interference and noise, respectively, are independent of the second and third terms in (9). Consequently, the cross-correlation of $\hat{\mathbf{h}}_{jju}^{(1)}[n]$ and $\hat{\mathbf{h}}_{jju}^{(2)}[n]$ can be easily shown to be same as the covariance matrix $\mathbf{R}_{jju}$ [16]. Therefore, we can use the following unbiased Hermitian-symmetric sample cross-covariance matrix as an estimate for $\mathbf{R}_{jju}$

$$\ddot{\mathbf{R}}_{jju} = \frac{1}{2N_R}\sum_{n=1}^{N_R}\hat{\mathbf{h}}_{jju}^{(1)}[n]\left(\hat{\mathbf{h}}_{jju}^{(2)}[n]\right)^H + \frac{1}{2N_R}\sum_{n=1}^{N_R}\hat{\mathbf{h}}_{jju}^{(2)}[n]\left(\hat{\mathbf{h}}_{jju}^{(1)}[n]\right)^H \tag{10}$$

As $N_R \to \infty$, one can show that the estimated covariance matrix converges in probability to the true covariance matrix, i.e., $\ddot{\mathbf{R}}_{jju} \xrightarrow[N_R\to\infty]{P} \mathbf{R}_{jju}$. However, this unbiased covariance estimator does not guarantee positive diagonal elements for finite N_R. Therefore, we consider a regularized estimate for the covariance matrix given by

$$\hat{\mathbf{R}}_{jju} = \alpha_R\ddot{\mathbf{R}}_{jju} + (1-\alpha_R)\mathbf{R}_b \tag{11}$$

where $\mathbf{R}_b$ is an arbitrary symmetric positive definite bias-matrix, and α_R is a design parameter. Additionally, it is useful to define $\bar{\mathbf{R}}_{jju}$ to denote the expected value of $\hat{\mathbf{R}}_{jju}$ as $\bar{\mathbf{R}}_{jju} \triangleq \mathbb{E}\{\hat{\mathbf{R}}_{jju}\} = \alpha_R\mathbf{R}_{jju} + (1-\alpha_R)\mathbf{R}_b$.

2) Estimation of $\mathbf{S}_{jju}$: For element-wise LMMSE-type estimation, it is sufficient to estimate the diagonal matrix $\mathbf{S}_{jju}$. Therefore, we use an unbiased Hermitian-symmetric covariance estimate $\ddot{\mathbf{S}}_{jju}$ (similar to $\ddot{\mathbf{R}}_{jju}$) as follows

$$[\ddot{\mathbf{S}}_{jju}]_{pp} = \frac{1}{2N_R}\sum_{n=1}^{N_R}[\hat{\mathbf{h}}_{jju}^{(1)}[n]]_p[\hat{\mathbf{h}}_{jju}^{(2)}[n]]_p^* + \frac{1}{2N_R}\sum_{n=1}^{N_R}[\hat{\mathbf{h}}_{jju}^{(2)}[n]]_p[\hat{\mathbf{h}}_{jju}^{(1)}[n]]_p^*, \quad \forall p \in 1\ldots M. \tag{12}$$

A regularized estimate for $\mathbf{S}_{jju}$ is given by

$$\hat{\mathbf{S}}_{jju} = \alpha_R\ddot{\mathbf{S}}_{jju} + (1-\alpha_R)\mathrm{diag}(\mathbf{R}_b). \tag{13}$$

We define $\bar{\mathbf{S}}_{jju}$ as the expected value of $\hat{\mathbf{S}}_{jju}$, $\bar{\mathbf{S}}_{jju} \triangleq \mathbb{E}\{\hat{\mathbf{S}}_{jju}\} = \alpha_R\mathbf{S}_{jju} + (1-\alpha_R)\mathrm{diag}(\mathbf{R}_b)$, for future use.

In summary, the BS needs to compute channel covariance matrices for each set of τ_s coherence blocks in order to obtain the LMMSE-type/element-wise LMMSE-type channel estimates in each coherence block within the set.

III. Average UL and DL Spectral Efficiencies as Performance Metric

In order to analyze the performance of the covariance estimation algorithm in a massive MIMO system, we derive a closed-form expression for a performance metric that captures the impact of pilot overhead. Clearly, the performance of a massive MIMO system directly depends on the quality of channel covariance estimates. In the literature [14]–[16], the achievable SE value is typically computed for a single set of τ_s coherence blocks as a function of the estimated covariance matrices corresponding to that set. Such an SE value corresponding to a single realization of covariance matrices is particularly important for designing practical receivers but does not clearly capture the impact of covariance estimation. Therefore, we consider average SE computed across different realizations of the covariance matrices (ensemble average) as the performance metric. We derive closed-form expressions for average SE for both UL and DL data for LMMSE-type and element-wise LMMSE-type channel estimation.[4] Note that the maximum ratio combining corresponds to a lower SE value when compared to LMMSE combining [10]. Moreover, the aim of the average SE-based performance metric derived in this paper is to capture the impact pilot overhead but not to present an achievable rate. Therefore, we use matched filter precoding and maximum ratio combining for deriving the SE expressions in the DL and UL communication, respectively.

A. Uplink Spectral Efficiency

In this section, the average SE for the UL channel of a target user (j,u) is derived when the channel estimates are used in a maximum ratio combiner at the BS. For the maximum ratio combiner, the combining vector $\mathbf{v}_{ju}[n]$ can be written as $\mathbf{v}_{ju}[n] = \hat{\mathbf{h}}_{jju}[n] = \hat{\mathbf{W}}_{ju}\hat{\mathbf{h}}_{jju}^{LS}[n]$, where

$$\hat{\mathbf{W}}_{ju} = \begin{cases} \hat{\mathbf{R}}_{jju}\hat{\mathbf{Q}}_{ju}^{-1}, & \text{LMMSE-type channel estimate} \\ \hat{\mathbf{S}}_{jju}\hat{\mathbf{P}}_{ju}^{-1}, & \text{element-wise LMMSE-type channel estimate.} \end{cases}$$

For the sake of mathematical tractability, we make the following assumptions

- $\hat{\mathbf{R}}_{jju}$ ($\hat{\mathbf{S}}_{jju}$) and $\hat{\mathbf{Q}}_{ju}$ ($\hat{\mathbf{P}}_{ju}$) are each computed from a different non-overlapping set of coherence blocks that does not include n^{th} block [15]. Consequently, the random variables $\hat{\mathbf{R}}_{jju}/\hat{\mathbf{S}}_{jju}$, $\hat{\mathbf{Q}}_{ju}/\hat{\mathbf{P}}_{ju}$, and $\hat{\mathbf{h}}_{jju}^{LS}[n]$ are mutually uncorrelated.

[4]Note that, [14]–[16] utilize numerically computed average rate as a performance measure for covariance matrix estimation method.

- For the LMMSE-type channel estimate, N_Q is assumed greater than M, so that the distribution of $\hat{\mathbf{Q}}_{ju}^{-1}$ is non-degenerate inverse Wishart.

The received combined signal is given by

$$\begin{aligned}\mathbf{v}_{ju}^H\mathbf{y}_j &= \sqrt{\mu}\mathbb{E}\{\mathbf{v}_{ju}^H\mathbf{h}_{jju}\}x_{ju} \\ &\quad + \sqrt{\mu}(\mathbf{v}_{ju}^H\mathbf{h}_{jju} - \mathbb{E}\{\mathbf{v}_{ju}^H\mathbf{h}_{jju}\})x_{ju} \\ &\quad + \sum_{k\neq u}\sqrt{\mu}\mathbf{v}_{ju}^H\mathbf{h}_{jjk}x_{jk} \\ &\quad + \sum_{l\neq j}\sum_{k=1}^{K}\sqrt{\mu}\mathbf{v}_{ju}^H\mathbf{h}_{jlk}x_{lk} + \mathbf{v}_{ju}^H\mathbf{n}_j \end{aligned} \tag{14}$$

In (14), the first term corresponds to the signal component, the second term is a result of the uncertainty in the array gain, the third term corresponds to the non-coherent intra-cell interference, the fourth term corresponds to the coherent interference from pilot contamination, and the last term corresponds to the additive noise component. Since the first term is uncorrelated with the subsequent terms, a lower bound on SE of the UL channel from user (j,u) to BS j can be obtained as [15]

$$SE_{ju}^{(\mathrm{ul})} = \left(1 - \frac{P}{C_u} - \frac{N_R P}{C_u\tau_s}\right)\log_2\left(1+\gamma_{ju}^{(ul)}\right), \quad \text{[bits/s/Hz]}$$

where $\gamma_{ju}^{(\mathrm{ul})}$ is given by

$$\gamma_{ju}^{(\mathrm{ul})} = \frac{|\mathbb{E}\{\mathbf{v}_{ju}^H\mathbf{h}_{jju}\}|^2}{\sum\limits_{l,k}\mathbb{E}\{|\mathbf{v}_{ju}^H\mathbf{h}_{jlk}|^2\} - |\mathbb{E}\{\mathbf{v}_{ju}^H\mathbf{h}_{jju}\}|^2 + \frac{1}{\mu}\mathbb{E}\{\mathbf{v}_{ju}^H\mathbf{v}_{ju}\}}$$

and the expectation $\mathbb{E}\{\cdot\}$ is over the channel realizations. In the pre-log factor, P/C_u accounts for ChEst pilots, and $N_RP/C_u\tau_s$ accounts for CovEst pilots. However, since we assume that $\hat{\mathbf{W}}_{ju}$ and $\hat{\mathbf{h}}_{jju}^{LS}[n]$ are mutually independent, we have $\mathbb{E}\{\cdot\} = \mathbb{E}_W\{\mathbb{E}_{h^{LS}}\{\cdot\}\}$, where $\mathbb{E}_W$ is the expectation over $\hat{\mathbf{W}}_{ju}$, and $\mathbb{E}_{h^{LS}}$ is the expectation over the LS estimate. Let $\mathbf{R}_s \triangleq \sum\limits_{l=1}^{L}\sum\limits_{k=1}^{K}\mathbf{R}_{jlk} + \frac{1}{\mu}\mathbf{I}$. Then, the signal to interference plus noise ratio (SINR) expression can be further simplified to (15) [15], as shown at the bottom of the next page.

B. Uplink Spectral Efficiency When $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju}\hat{\mathbf{Q}}_{ju}^{-1}$

In this subsection, expressions for all the terms given in (15) are derived for the case when $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju}\hat{\mathbf{Q}}_{ju}^{-1}$. In what follows, $\mathbb{E}_R\{\cdot\}$ represents the expectation over $\hat{\mathbf{R}}_{jju}$, $\mathbb{E}_Q\{\cdot\}$ represents the expectation over $\hat{\mathbf{Q}}_{ju}$, and $\mathbb{E}_W\{\cdot\}$ represents the expectation over both $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$. It should be noted that, as already mentioned, we have assumed that $\hat{\mathbf{R}}_{jju}$ and $\hat{\mathbf{Q}}_{ju}$ are estimated from different pilot resources (coherence blocks) such that the estimates are independent to each other. Therefore, $\mathbb{E}_R\{\cdot\}$ and $\mathbb{E}_Q\{\cdot\}$ can be evaluated independently.

Before analytically deriving the expectations for the terms in (15), we present some useful lemmas.

Lemma 1: Given an arbitrary matrix $\mathbf{A} \in \mathbb{C}^{M\times M}$*, and for any mutually independent M-dimensional random vector* $\mathbf{h}$ *distributed as* $\mathcal{CN}(\mathbf{0},\mathbf{R})$*, we have*

$$\mathbb{E}\{\mathbf{h}\mathbf{h}^H\mathbf{A}\mathbf{h}\mathbf{h}^H\} = \mathbf{R}\mathbf{A}\mathbf{R} + \mathbf{R}\mathrm{tr}(\mathbf{A}\mathbf{R}) \tag{17}$$

$$\mathbb{E}\{|\mathbf{h}^H\mathbf{A}\mathbf{h}|^2\} = |\mathrm{tr}(\mathbf{A}^H\mathbf{R})|^2 + \mathrm{tr}(\mathbf{A}\mathbf{R}\mathbf{A}^H\mathbf{R}). \tag{18}$$

Proof: The proof is available in Appendix A. ■

Lemma 2: Given a Hermitian matrix $\mathbf{C} \in \mathbb{C}^{M\times M}$*, an arbitrary matrix* $\mathbf{A} \in \mathbb{C}^{M\times M}$*, and a complex Wishart matrix,* $\mathbf{X} \in \mathbb{C}^{M\times M}$*, distributed as* $\mathcal{W}(N,\mathbf{I})$*, we have*

$$\mathbb{E}\{[\mathbf{X}^{-1}]_{ij}\} = \frac{[\mathbf{I}]_{ij}}{N-M} \tag{19}$$

$$\mathbb{E}\{[\mathbf{X}^{-1}]_{ij}[\mathbf{X}^{-1}]_{lk}\} = \frac{[\mathbf{I}]_{ij}[\mathbf{I}]_{lk} + \frac{1}{N-M}[\mathbf{I}]_{lj}[\mathbf{I}]_{ik}}{(N-M)^2-1} \tag{20}$$

$$\mathbb{E}\{\mathrm{tr}(\mathbf{X}^{-2}\mathbf{C})\} = \frac{N}{(N-M)^3-(N-M)}\mathrm{tr}(\mathbf{C}) \tag{21}$$

$$\mathbb{E}\{|\mathrm{tr}(\mathbf{X}^{-1}\mathbf{A})|^2\} = \frac{|\mathrm{tr}(\mathbf{A})|^2 + \frac{1}{N-M}\mathrm{tr}(\mathbf{A}\mathbf{A}^H)}{(N-M)^2-1}. \tag{22}$$

Proof: The proof is available in Appendix B. ■

Lemma 3: Given an arbitrary matrix $\mathbf{A} \in \mathbb{C}^{M\times M}$*, we have*

$$\begin{aligned}&\mathbb{E}\{\ddot{\mathbf{R}}_{jju}\mathbf{A}\ddot{\mathbf{R}}_{jju}\} \\ &\quad = \mathbf{R}_{jju}\mathbf{A}\mathbf{R}_{jju} \\ &\qquad + \frac{1}{2N_R}\mathbf{Q}_{ju}\mathrm{tr}(\mathbf{A}\mathbf{Q}_{ju}) + \frac{1}{2N_R}\mathbf{R}_{jju}\mathrm{tr}(\mathbf{A}\mathbf{R}_{jju})\end{aligned} \tag{23}$$

$$\begin{aligned}&\mathbb{E}\{|\mathrm{tr}(\ddot{\mathbf{R}}_{jju}\mathbf{A})|^2\} \\ &\quad = |\mathrm{tr}(\mathbf{R}_{jju}\mathbf{A})|^2 \\ &\qquad + \frac{1}{2N_R}\left\{\mathrm{tr}(\mathbf{A}\mathbf{Q}_{ju}\mathbf{A}^H\mathbf{Q}_{ju}) + \mathrm{tr}(\mathbf{A}\mathbf{R}_{jju}\mathbf{A}^H\mathbf{R}_{jju})\right\}\end{aligned} \tag{24}$$

Proof: The proof of this lemma uses Lemma 1 and is presented in Appendix C. ■

Now we are ready to formulate the key theorem of this subsection.

Theorem 1: The numerator term of (15) *when* $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju}\hat{\mathbf{Q}}_{ju}^{-1}$ *is given by*

$$\begin{aligned}&\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jju})\} \\ &\quad = \mathrm{tr}(\mathbf{W}_{ju}^H\mathbf{R}_{jju}) \\ &\qquad + \left\{\frac{N_Q}{N_Q-M}\mathrm{tr}(\bar{\mathbf{W}}_{ju}^H\mathbf{R}_{jju}) - \mathrm{tr}(\mathbf{W}_{ju}^H\mathbf{R}_{jju})\right\}\end{aligned} \tag{25}$$

The first and second terms of the denominator in (15) *are given by*

$$\begin{aligned}&\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\hat{\mathbf{W}}_{ju}^H\mathbf{R}_s)\} \\ &\quad = \mathrm{tr}(\mathbf{W}_{ju}\mathbf{Q}_{ju}\mathbf{W}_{ju}^H\mathbf{R}_s) \\ &\qquad + \Big\{\kappa_1\mathrm{tr}(\bar{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\bar{\mathbf{W}}_{ju}^H\mathbf{R}_s) - \mathrm{tr}(\mathbf{W}_{ju}\mathbf{Q}_{ju}\mathbf{W}_{ju}^H\mathbf{R}_s) \\ &\qquad + \frac{\alpha_R^2\kappa_1}{2N_R}M\mathrm{tr}(\mathbf{R}_s\mathbf{Q}_{ju}) + \frac{\alpha_R^2\kappa_1}{2N_R}\mathrm{tr}(\mathbf{W}_{ju})\mathrm{tr}(\mathbf{R}_s\mathbf{R}_{jju})\Big\}\end{aligned} \tag{26}$$

and (16)*, as shown at the bottom of the next page, respectively. Here,* $\kappa_1 \triangleq N_Q\kappa_2/(N_Q-M)$*,* $\kappa_2 \triangleq N_Q^2/((N_Q-M)^2-1)$*,* $\bar{\mathbf{W}}_{ju} \triangleq \bar{\mathbf{R}}_{jju}\mathbf{Q}_{ju}^{-1}$ *and* $\mathbf{W}_{lu} \triangleq \mathbf{R}_{jlu}\mathbf{Q}_{ju}^{-1}$ *for all* $l = 1$ *to* L*.*

Proof: We define a matrix $\tilde{\mathbf{Q}}_{ju}$ as

$$\tilde{\mathbf{Q}}_{ju} \triangleq N_Q(\mathbf{Q}_{ju}^{-\frac{1}{2}}\hat{\mathbf{Q}}_{ju}\mathbf{Q}_{ju}^{-\frac{1}{2}}). \tag{27}$$

It can be seen that $\tilde{\mathbf{Q}}_{ju}$ is Wishart distributed, i.e., $\mathcal{W}(N_Q, \mathbf{I})$.

Using (27) and the fact that $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju}\hat{\mathbf{Q}}_{ju}^{-1}$, the numerator term of (15) can be written as

$$\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})\} = N_Q \mathbb{E}_W\{\mathrm{tr}(\mathbf{Q}_{ju}^{-\frac{1}{2}}\tilde{\mathbf{Q}}_{ju}^{-1}\mathbf{Q}_{ju}^{-\frac{1}{2}}\hat{\mathbf{R}}_{jju}\mathbf{R}_{jju})\}. \quad (28)$$

Taking direct expectation over $\hat{\mathbf{R}}_{jju}$ in (28) and also using Lemma 2, (25) can be obtained.

Proof of (26) and (16) is as follows. Substituting $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju}\hat{\mathbf{Q}}_{ju}^{-1}$ into the first and second terms in the denominator of (15) and using Lemma 2, we get the following equations

$$\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\hat{\mathbf{W}}_{ju}^H\mathbf{R}_s)\} = \kappa_1 \mathbb{E}_R\{\mathrm{tr}(\mathbf{Q}_{ju}^{-1}\hat{\mathbf{R}}_{jju}\mathbf{R}_s\hat{\mathbf{R}}_{jju})\} \quad (29)$$

$$\mathbb{E}_W\{|\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jlu})|^2\} = \kappa_2\mathbb{E}_R\{|\mathrm{tr}(\mathbf{Q}_{ju}^{-1}\hat{\mathbf{R}}_{jju}\mathbf{R}_{jlu})|^2\} + \frac{\kappa_1}{N_Q}\mathbb{E}_R\{\mathrm{tr}(\mathbf{Q}_{ju}^{-1}\hat{\mathbf{R}}_{jju}\mathbf{R}_{jlu}^2\hat{\mathbf{R}}_{jju}\mathbf{Q}_{ju}^{-1})\}. \quad (30)$$

Then using Lemma 3, and substituting (11) into (29) and (30), we get (26) and (16), respectively. ∎

Note that the expectation terms given in Theorem 1 contain two components: (i) the component that corresponds to known covariance information (first term of the right-hand side of the equations) and (ii) a penalty component (all terms except the first term of the right-hand side of the equations) due to regularization of $\mathbf{R}_{jju}$ estimate and due to imperfect channel covariance information. Note that for $\alpha_R = 1$, and as N_R and N_Q tend to infinity, the penalty components of the expectation terms vanish.

C. Uplink Spectral Efficiency When $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju}\hat{\mathbf{P}}_{ju}^{-l}$

In this subsection, we present the derivations for all the terms given in (15) when $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju}\hat{\mathbf{P}}_{ju}^{-1}$. In what follows, $\mathbb{E}_S\{\cdot\}$ represents the expectation over $\hat{\mathbf{S}}_{jju}$, $\mathbb{E}_P\{\cdot\}$ represents the expectation over $\hat{\mathbf{P}}_{ju}$, and $\mathbb{E}_W\{\cdot\}$ represents the expectation over both $\hat{\mathbf{S}}_{jju}$ and $\hat{\mathbf{P}}_{ju}$.

Before analytically deriving the expectations for the terms in (15), we present some useful lemmas.

Lemma 4: Given a zero mean complex Gaussian 2×1 random vector $\mathbf{h} = [h_1, h_2]^\intercal$ with covariance matrix

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$$

we can state that $\mathbb{E}\{|h_1|^2|h_2|^2\} = r_{11}r_{22} + r_{12}r_{21}$.

Proof: The proof of this lemma is straight forward to obtain and we omit it due to lack of space. ∎

Lemma 5: Given arbitrary matrices $\mathbf{A}_1 \in \mathbb{C}^{M\times M}$, $\mathbf{A}_2 \in \mathbb{C}^{M\times M}$, $\mathbf{A} \in \mathbb{C}^{M\times M}$, and a matrix $\mathbf{Y} = \mathbf{Z}/2$, where $\mathbf{Z}$ is a diagonal matrix whose elements are i.i.d. χ^2 random variables with $2N$-degrees of freedom ($N > 2$), we have

$$\mathbb{E}\{\mathrm{tr}(\mathbf{Y}^{-1}\mathbf{A}_1\mathbf{Y}^{-1}\mathbf{A}_2)\} = \tau_1\mathrm{tr}(\mathbf{A}_1\mathbf{A}_2) + \tau_2\mathrm{tr}(\mathbf{A}_{1d}\mathbf{A}_{2d}) \quad (31)$$

$$\mathbb{E}\{|\mathrm{tr}(\mathbf{Y}^{-1}\mathbf{A})|^2\} = \tau_1|\mathrm{tr}(\mathbf{A})|^2 + \tau_2\mathrm{tr}(\mathbf{A}_d^H\mathbf{A}_d) \quad (32)$$

where $\tau_1 \triangleq 1/(N-1)^2$, $\tau_2 \triangleq \tau_1/(N-2)$, $\mathbf{A}_{1d} \triangleq \mathrm{diag}(\mathbf{A}_1)$, $\mathbf{A}_{2d} \triangleq \mathrm{diag}(\mathbf{A}_2)$, and $\mathbf{A}_d \triangleq \mathrm{diag}(\mathbf{A})$.

Proof: The proof is available in Appendix D. ∎

Lemma 6: Given an arbitrary matrix $\mathbf{A} \in \mathbb{C}^{M\times M}$ and an arbitrary diagonal matrix $\mathbf{D} \in \mathbb{R}^{M\times M}$, then

$$\mathbb{E}\{\ddot{\mathbf{S}}_{jju}\mathbf{A}\ddot{\mathbf{S}}_{jju}\} = \mathbf{S}_{jju}\mathbf{A}\mathbf{S}_{jju} + \frac{1}{2N_R}\mathbf{A}\circ\mathbf{Q}_{ju}\circ\mathbf{Q}_{ju} + \frac{1}{2N_R}\mathbf{A}\circ\mathbf{R}_{jju}\circ\mathbf{R}_{jju} \quad (33)$$

$$\mathbb{E}\{|\mathrm{tr}(\ddot{\mathbf{S}}_{jju}\mathbf{D})|^2\} = |\mathrm{tr}(\mathbf{S}_{jju}\mathbf{D})|^2 + \frac{1}{2N_R}\sum_{p=1}^{M}\sum_{q=1}^{M}[\mathbf{D}(\mathbf{Q}_{ju}\circ\mathbf{Q}_{ju})\mathbf{D}]_{pq} + \frac{1}{2N_R}\sum_{p=1}^{M}\sum_{q=1}^{M}[\mathbf{D}(\mathbf{R}_{jju}\circ\mathbf{R}_{jju})\mathbf{D}]_{pq} \quad (34)$$

Proof: The proof is available in Appendix E. ∎

Now we are ready to formulate the key theorem of this subsection.

Theorem 2: The numerator term of (15) when $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju}\hat{\mathbf{P}}_{ju}^{-1}$ is given by

$$\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jju})\} = \mathrm{tr}(\bar{\mathbf{W}}_{ju}^H\mathbf{R}_{jju}) + \left\{\frac{N_Q}{(N_Q-1)}\mathrm{tr}(\bar{\mathbf{W}}_{ju}^H\mathbf{R}_{jju}) - \mathrm{tr}(\mathbf{W}_{ju}^H\mathbf{R}_{jju})\right\} \quad (35)$$

$$\gamma_{ju}^{(\mathrm{ul})} = \frac{|\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jju})\}|^2}{\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\hat{\mathbf{W}}_{ju}^H\mathbf{R}_s)\} + \sum_{l=1}^{L}\mathbb{E}_W\{|\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jlu})|^2\} - |\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jju})\}|^2} \quad (15)$$

$$\mathbb{E}_W\{|\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jlu})|^2\} = |\mathrm{tr}(\mathbf{W}_{ju}^H\mathbf{R}_{jlu})|^2 + \left\{\kappa_2|\mathrm{tr}(\bar{\mathbf{W}}_{ju}^H\mathbf{R}_{jlu})|^2 - |\mathrm{tr}(\mathbf{W}_{ju}^H\mathbf{R}_{jlu})|^2 + \frac{\alpha_R^2\kappa_2}{2N_R}\mathrm{tr}(\mathbf{W}_{lu}\mathbf{Q}_{ju}\mathbf{W}_{lu}^H\mathbf{Q}_{ju}) + \frac{\alpha_R^2\kappa_2}{2N_R}\mathrm{tr}(\mathbf{W}_{lu}\mathbf{R}_{jju}\mathbf{W}_{lu}^H\mathbf{R}_{jju}) + \frac{\kappa_1}{N_Q}\mathrm{tr}(\bar{\mathbf{W}}_{ju}^H\bar{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\mathbf{W}_{lu}^H\mathbf{W}_{lu}\mathbf{Q}_{ju}) + \frac{\alpha_R^2\kappa_1}{2N_QN_R}M\mathrm{tr}(\mathbf{W}_{jlu}^2\mathbf{Q}_{ju}^2) + \frac{\alpha_R^2\kappa_1}{2N_QN_R}\mathrm{tr}(\mathbf{W}_{ju})\mathrm{tr}(\mathbf{W}_{jlu}^2\mathbf{Q}_{ju}\mathbf{R}_{jju})\right\} \quad (16)$$

The first and second terms of the denominator in (15) *are given by* (36) *and* (37)*, as shown at the bottom of the next page, respectively, where* $\kappa_3 = N_Q^2/(N_Q-1)^2$, $\kappa_4 = \kappa_3/(N_Q-2)$, $\mathbf{S}_s \triangleq \mathrm{diag}(\mathbf{R}_s)$, $\bar{\mathbf{W}}_{ju} \triangleq \bar{\mathbf{S}}_{jju}\mathbf{P}_{ju}^{-1}$ *and* $\mathbf{W}_{lu} \triangleq \mathbf{S}_{jlu}\mathbf{P}_{ju}^{-1}$ *for all* $l = 1$ *to* L.

Proof: We define the diagonal matrix $\tilde{\mathbf{P}}_{ju}$ as follows

$$\tilde{\mathbf{P}}_{ju} \triangleq N_Q(\mathbf{P}_{ju}^{-1}\hat{\mathbf{P}}_{ju}). \tag{39}$$

It can be seen that the elements of $2\tilde{\mathbf{P}}_{ju}$ are i.i.d. χ^2 random variables with $2N$-degrees of freedom. Using (39) and the fact that $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju}\hat{\mathbf{P}}_{ju}^{-1}$, the numerator term of (15) can be written as

$$\begin{aligned}&\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jju})\}\\ &\quad= N_Q\mathbb{E}_W\{\mathrm{tr}(\tilde{\mathbf{P}}_{ju}^{-1}\mathbf{P}_{ju}^{-1}\hat{\mathbf{S}}_{jju}\mathbf{R}_{jju})\}\\ &\quad= N_Q\sum_{p=1}^{M}\mathbb{E}_P\{[\tilde{\mathbf{P}}_{ju}^{-1}]_{pp}\}\mathbb{E}_S\{[\mathbf{P}_{ju}^{-1}\hat{\mathbf{S}}_{jju}\mathbf{R}_{jju}]_{pp}\}.\end{aligned} \tag{40}$$

Taking direct expectation over $\hat{\mathbf{S}}_{jju}$ in (40) and using the properties of inverse χ^2 distribution, (35) can be obtained.

Proof of (36) and (37) is as follows. Substituting $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju}\hat{\mathbf{P}}_{ju}^{-1}$ and (39) into the first and second denominator terms of (15) and using Lemma 5, we get the following equations

$$\begin{aligned}&\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\hat{\mathbf{W}}_{ju}^H\mathbf{R}_s)\}\\ &\quad= \kappa_3\mathbb{E}_S\{\mathrm{tr}(\mathbf{P}_{ju}^{-1}\mathbf{Q}_{ju}\mathbf{P}_{ju}^{-1}\hat{\mathbf{S}}_{jju}\mathbf{R}_s\hat{\mathbf{S}}_{jju})\}\\ &\qquad+\kappa_4\mathbb{E}_S\{\mathrm{tr}(\mathbf{P}_{ju}^{-1}\hat{\mathbf{S}}_{jju}\mathbf{S}_s\hat{\mathbf{S}}_{jju})\}\end{aligned} \tag{41}$$

$$\begin{aligned}&\mathbb{E}_W\{|\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jlu})|^2\}\\ &\quad= \kappa_3\mathbb{E}_S\{|\mathrm{tr}(\mathbf{P}_{ju}^{-1}\hat{\mathbf{S}}_{jju}\mathbf{S}_{jlu})|^2\}\\ &\qquad+\kappa_4\mathbb{E}_S\{\mathrm{tr}(\mathbf{P}_{ju}^{-2}\hat{\mathbf{S}}_{jju}^2\mathbf{S}_{jlu}^2)\}.\end{aligned} \tag{42}$$

Then using Lemma 6 and substituting (13) into (41) and (42), we get (36) and (37), respectively. ∎

Similar to Theorem 1, the penalty components of the expectation terms given in Theorem 2 also vanish if $\alpha_R = 1$, and as N_R and N_Q tend to infinity.

D. Downlink Spectral Efficiency

The DL spectral efficiency for user (j,u) is given in this section for a matched filter precoder, i.e., $\mathbf{b}_{ju} = \hat{\mathbf{h}}_{jju}[n]/\sqrt{\mathbb{E}\{\|\hat{\mathbf{h}}_{jju}[n]\|^2\}} = \hat{\mathbf{W}}_{ju}\hat{\mathbf{h}}_{jju}^{LS}/\sqrt{\mathbb{E}\{\|\hat{\mathbf{W}}_{ju}\hat{\mathbf{h}}_{jju}^{LS}[n]\|^2\}}$. Therefore, the received signal at user (j,u) can be written as

$$\begin{aligned}z_{ju} =\ &\sqrt{\lambda}\mathbb{E}\{\mathbf{b}_{ju}^H\mathbf{h}_{jju}\}d_{ju}\\ &+\sqrt{\lambda}(\mathbf{b}_{ju}^H\mathbf{h}_{jju}-\mathbb{E}\{\mathbf{b}_{ju}^H\mathbf{h}_{jju}\})d_{ju}\\ &+\sum_{k\neq u}\sqrt{\lambda}(\mathbf{b}_{ju}^H\mathbf{h}_{jjk})d_{jk}\\ &+\sum_{l\neq j}\sum_{k=1}^{K}\sqrt{\lambda}(\mathbf{b}_{ju}^H\mathbf{h}_{jlk})d_{lk}+e_{ju}.\end{aligned} \tag{43}$$

Here, we assume that the scalar in the denominator of the precoding vector, $\sqrt{\mathbb{E}\{\|\hat{\mathbf{h}}_{jju}[n]\|^2\}}$, is a known constant at the BS. The first term in (43) corresponds to the desired signal component, the second term corresponds to the uncertainty in the DL transmit array gain, the third term corresponds to the non-coherent intra-cell interference, the coherent interference from pilot contamination given by the fourth term, and the last term represents the additive noise component. The second term in (43) corresponds to the uncertainty in the DL transmit array gain. Then, due to the similarity between the UL received combined signal in (14) to the DL received signal, a lower bound on DL channel SE of the user (j,u) can be easily shown to be

$$\mathrm{SE}_{ju}^{(dl)} = \log_2\left(1+\gamma_{ju}^{(dl)}\right) \quad [\text{bits/s/Hz}],$$

where $\gamma_{ju}^{(dl)}$ is given by (38), as shown at the bottom of the next page, and $\mathbf{R}_s^{(dl)} \triangleq \sum_{l=1}^{L}\sum_{k=1}^{K}\mathbf{R}_{jlk}$. We utilize channel reciprocity and avoid the use of DL pilots. Consequently, there is no pre-log factor for the SE expression. The expectation taken in all the terms of (38) is over the random matrix $\hat{\mathbf{W}}_{ju}$. However, $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{R}}_{jju}\hat{\mathbf{Q}}_{ju}^{-1}$ for the LMMSE-type channel estimation and $\hat{\mathbf{W}}_{ju} = \hat{\mathbf{S}}_{jju}\hat{\mathbf{P}}_{ju}^{-1}$ for the element-wise LMMSE-type channel estimation. These expectation terms are already presented in Theorems 1, and 2 for the LMMSE-type, and element-wise LMMSE-type, respectively. Furthermore, $\mathbf{R}_s$ should be replaced by $\mathbf{R}_s^{(dl)}$ in computing the expectation terms.

IV. Main Discussion: Impact of Pilot Overhead and Channel Estimation

Based on the closed-form SE expressions derived in the previous section, we have established a direct relation between the average SE value and the parameters N_R and N_Q. Thus, we discuss here the impact of these parameters on the SE corresponding to the LMMSE-type and element-wise LMMSE-type channel estimation. We also address the question of how to choose between LMMSE-type and element-wise LMMSE-type channel estimations.

It can be noted from the expectation terms in Theorems 1 and 2 that the penalty components due to imperfect covariance information gradually vanish with an increase in N_R and N_Q, but the penalty due to the regularization remains finite. Furthermore, if $||\mathbf{W}_{ju}-\bar{\mathbf{W}}_{ju}||/||\mathbf{W}_{ju}|| \ll 1$ (i.e., if α_R is close to 1), one can state that these expectation terms converge to the values that correspond to the known covariance case. However, despite leading to an improvement in $\gamma_{ju}^{(ul)}$ (due to convergence of the expected values), an increase in N_R also causes a degradation in the pre-log factor of the derived UL SE expression. Therefore, the choice of N_R impacts UL SE in two ways: (i) smaller the value of N_R, higher the error in covariance estimation and hence lower the value of UL SE and (ii) larger the value of N_R, higher the consumption of UL resources and hence lower the value of UL SE. Whereas, due to the absence of DL pilots, the DL SE does not degrade with an increase in N_R; it gradually rises to the DL SE value that corresponds to the known covariance case. Larger N_Q makes both the UL and DL SE better due to the improved estimates of $\mathbf{Q}_{ju}$ (or $\mathbf{P}_{ju}$). Therefore, given an SE requirement, the aim

here is to choose minimum N_R and N_Q values that are sufficient to provide the desired SE.

Since estimating $\mathbf{Q}_{ju}$ (or $\mathbf{P}_{ju}$) does not involve additional pilot transmission, choosing N_Q is not as critical as choosing N_R. Therefore, if we consider N_Q as known, it is also important to derive N_R values that make the LMMSE-type channel estimation preferable to the element-wise LMMSE-type one, and vice-versa. By comparing the UL/DL SINR values (in (15) or (38)) for the two channel estimation techniques, we can compute a threshold, $\bar{N}_R$, such that the element-wise LMMSE-type estimator is preferable if $N_R < \bar{N}_R$, and the LMMSE-type estimator is preferable otherwise. Note that $\bar{N}_R$ is different for UL and DL covariance estimation. It can be obtained by equating the SINR expressions for the LMMSE-type and element-wise LMMSE-type channel estimation techniques (for UL and DL) and solving the corresponding equation for N_R. After some straight forward algebra, $\bar{N}_R$ can be obtained in the form

$$\bar{N}_R = \frac{fc - ah}{ag - fb} \tag{44}$$

where b, c, g and h are given at the bottom of the next page along with the following parameters

$$a = \left(\frac{N_Q}{N_Q - M}\mathrm{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})\right)^2$$

$$f = \left(\frac{N_Q}{(N_Q - 1)}\mathrm{tr}(\bar{\mathbf{W}}_{ju}^H \mathbf{R}_{jju})\right)^2$$

$$\bar{\mathbf{S}}_s = \mathrm{diag}(\bar{\mathbf{R}}_s); \qquad \bar{\mathbf{R}}_s = \begin{cases} \mathbf{R}_s, & \text{for UL} \\ \mathbf{R}_s^{(dl)}, & \text{for DL} \end{cases}$$

$$d = \begin{cases} 0, & \text{for UL} \\ \dfrac{1}{\lambda}, & \text{for DL.} \end{cases}$$

Note that $\bar{N}_R$ is a function of N_Q which can take any real value. Thus, if $\bar{N}_R$ is negative for some value of N_Q, it means, for that particular choice of N_Q, there is no valid N_R that makes the LMMSE-type channel estimation preferable. Consequently, using (44), we can also compute a threshold for N_Q below which element-wise LMMSE-type channel estimation is always preferred. However, deriving a theoretical expression for such a threshold is extremely difficult. It can be easily computed numerically.

Therefore, the closed-form expressions for average UL and DL SE, for the LMMSE-type and element-wise LMMSE-type channel estimation methods serve as tools for choosing different design parameters, and also as a tool for choosing a preferred channel estimation technique. In practice, with approximate models of the covariance matrix of an individual user in a massive MIMO system, the derived expressions for average SE enables us to choose these parameters for the desired UL and DL SE values.

In what follows, we validate the derived theoretical SE expressions with simulated SE obtained by averaging over multiple realizations of random covariance estimation matrices. Then, we compare the theoretical SE expressions with the SE expressions that correspond to known covariance case. Finally, we also depict the impact of N_R on the SE.

V. Simulation Results

We consider a massive MIMO system with $L = 7$ cells, each comprising a BS with $M = 100$ antennas and $K = 10$ users. The BSs are at a distance of 300 m apart from each other, and the users are uniformly spaced at a distance of 120 m from the BS in their cells. The users that reuse the same pilot in different cells are at the same position relative to the corresponding BSs. The angular spread of the channel cluster is assumed to be 20°, within which the received paths from a user are assumed to be uniformly distributed. We consider a 3GPP urban macro (UMa) [20] scenario with a non-line-of-sight (N-LOS) channel for simulating the path loss model. The mean path loss of the received signal from a

$$\begin{aligned}
&\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\hat{\mathbf{W}}_{ju}^H\mathbf{R}_s)\} \\
&\quad = \mathrm{tr}(\mathbf{W}_{ju}\mathbf{Q}_{ju}\mathbf{W}_{ju}^H\mathbf{R}_s) + \Big\{\kappa_3\mathrm{tr}(\bar{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\bar{\mathbf{W}}_{ju}^H\mathbf{R}_s) - \mathrm{tr}(\mathbf{W}_{ju}\mathbf{Q}_{ju}\mathbf{W}_{ju}^H\mathbf{R}_s) \\
&\qquad + \frac{\alpha_R^2\kappa_3}{2N_R}\mathrm{tr}\Big(\mathbf{P}_{ju}^{-1}\mathbf{Q}_{ju}\mathbf{P}_{ju}^{-1}\{\mathbf{R}_s \circ \mathbf{Q}_{ju} \circ \mathbf{Q}_{ju}\} + \mathbf{P}_{ju}^{-1}\mathbf{Q}_{ju}\mathbf{P}_{ju}^{-1}\{\mathbf{R}_s \circ \mathbf{R}_{jju} \circ \mathbf{R}_{jju}\}\Big) + \kappa_4\mathrm{tr}(\bar{\mathbf{W}}_{ju}\mathbf{P}_{ju}\bar{\mathbf{W}}_{ju}^H\mathbf{S}_s) \\
&\qquad + \frac{\alpha_R^2\kappa_4}{2N_R}\mathrm{tr}(\mathbf{S}_s\mathbf{P}_{ju}) + \frac{\alpha_R^2\kappa_4}{2N_R}\mathrm{tr}(\mathbf{W}_{ju}\mathbf{S}_s\mathbf{S}_{jju})\Big\}
\end{aligned} \tag{36}$$

$$\begin{aligned}
&\mathbb{E}_W\{|\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jlu})|^2\} \\
&\quad = |\mathrm{tr}(\mathbf{W}_{ju}^H\mathbf{S}_{jlu})|^2 + \Big\{\kappa_3|\mathrm{tr}(\bar{\mathbf{W}}_{ju}^H\mathbf{S}_{jlu})|^2 - |\mathrm{tr}(\mathbf{W}_{ju}^H\mathbf{S}_{jlu})|^2 + \frac{\alpha_R^2\kappa_3}{2N_R}\sum_{p=1}^{M}\sum_{q=1}^{M}[\mathbf{W}_{lu}(\mathbf{Q}_{ju} \circ \mathbf{Q}_{ju})\mathbf{W}_{lu}]_{pq} \\
&\qquad + \frac{\alpha_R^2\kappa_3}{2N_R}\sum_{p=1}^{M}\sum_{q=1}^{M}[\mathbf{W}_{lu}(\mathbf{R}_{jju} \circ \mathbf{R}_{jju})\mathbf{W}_{lu}]_{pq} + \kappa_4\mathrm{tr}(\bar{\mathbf{W}}_{ju}^2\mathbf{S}_{jlu}^2) + \frac{\alpha_R^2\kappa_4}{2N_R}\mathrm{tr}(\mathbf{W}_{lu}^2\mathbf{P}_{ju}^2) + \frac{\alpha_R^2\kappa_4}{2N_R}\mathrm{tr}(\mathbf{W}_{lu}^2\mathbf{S}_{jju}^2)\Big\}
\end{aligned} \tag{37}$$

$$\gamma_{ju}^{(dl)} = \frac{|\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jju})\}|^2}{\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\hat{\mathbf{W}}_{ju}^H\mathbf{R}_s^{(dl)})\} + \sum_{l=1}^{L}\mathbb{E}_W\{|\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jlu})|^2\} - |\mathbb{E}_W\{\mathrm{tr}(\hat{\mathbf{W}}_{ju}^H\mathbf{R}_{jju})\}|^2 + \frac{1}{\lambda}} \tag{38}$$

user that is at a distance d (in m) from the BS is given as $PL(f,d) = 32.4 + 20\log_{10}(f/1\text{ GHz}) + 30\log_{10}(d_{3D}/1\text{ m})$, where $d_{3D} = \sqrt{d^2 + (h_{BS} - h_{UT})^2}$ m, f is the carrier frequency, h_{BS} is the height of a BS in m, and h_{UT} is the height of a UE in m. Therefore, the mean received SNR, in dB, is given by $SNR = P_T - PL - 10\log_{10}(kT_oB) - NF$, where P_T is the transmit power, k is the Boltzmann constant, $T_0 = 290$ K is the nominal temperature, B is the signal bandwidth, and NF is the noise figure in dB. In this setup, we consider $f = 3.4$ GHz, $P_T = 6$ dBm, $B = 40$ MHz, $NF = 10$ dB, $h_{BS} = 25$ m, and $h_{UT} = 1.5$ m which results in the mean SNR of the received signal from a user that is at a distance d from the BS to be given by $46.93 - 30\log_{10} d_{3D}$.

The number of symbols that are dedicated for UL transmission within each coherence block is chosen to be $C_u = 100$ symbols. We choose the number of symbols used for the ChEst (and also the CovEst) pilot to be $P = 10$. Second-order statistics of the channel are assumed to be constant for $\tau_s = 25000$ coherence blocks, and the UL transmit power is $\mu = 1$, and the DL transmit power is $\lambda = 10$. Additionally, we choose $\alpha_R = 0.95$, and $\mathbf{R}_b = \mathbf{I}$. Sample averaging for all the expectation terms is computed using 2000 trials for different values of $N_R = \{125, 250, 500, 1000, 2000, 4000, 8000\}$.

A. Uplink Spectral Efficiency

For this simulation example, we consider the UL SE expressions that correspond to the two channel estimation techniques: LMMSE-type channel estimation and the element-wise LMMSE-type channel estimation, each in combination with two beamforming techniques, which are MRC and ZF combining.[5] In Fig. 3, we plot the SE as a function of N_R for the two aforementioned channel estimation techniques and beamforming techniques. Fig. 3(a) depicts the SE values for $N_Q = 125$ and Fig. 3(b) shows SE values for $N_Q = 4000$. In both the subplots, we present SE values corresponding to known covariance matrices and theoretical SE values (only for the MRC combining case) as well as simulated SE values corresponding to the two channel estimation techniques that use the estimated covariance matrices.

In Fig. 3, it can be noticed that the theoretical SE, corresponding to LMMSE-type channel estimation and MRC combining, initially rises with N_R to approach the SE that corresponds to LMMSE channel estimation, followed by a drop in the theoretical SE at $N_R = 8000$. In contrast, the theoretical SE, corresponding to element-wise LMMSE-type channel estimation and MRC combining, approaches the SE corresponding to element-wise LMMSE channel estimation for N_R value as low as 125 and reaches its maximum at $N_R = 500$. Then, the theoretical SE reduces linearly with a further increase in N_R as the pilot overhead increases. Moreover, the simulated SEs match the theoretical values for both the channel estimation techniques tested, thereby validating the derivations presented in the paper.

The initial rise of the theoretical SEs is due to the improvement in the covariance estimates caused by the increase in the number of samples for estimation. However, a further increase in N_R results in a drop in UL SEs due to the pre-log factor. Despite the improvement in estimation quality of the covariance matrices, the SEs drop because of the consumption of UL resources for the additional CovEst pilots. This validates the theoretical analysis done in Section IV. Moreover, it should be noted that the LMMSE should always perform better than element-wise LMMSE as the correlation between antenna elements' channels are ignored in element-wise LMMSE. However, due to imperfect covariance information, element-wise LMMSE-type is not necessarily better than LMMSE-type. Specifically, element-wise LMMSE-type is expected to outperform the LMMSE-type when $N_R \leq \bar{N}_R$.

It can be seen from Fig. 3(a) and Fig. 3(b) that using element-wise LMMSE channel estimation instead of LMMSE channel estimation leads to a drop in SE. However, it is evident that the element-wise LMMSE-type channel estimation completely outperforms the LMMSE-type channel estimation for all the N_R values and for $N_Q = 125$. It can also be noted that even for $N_Q = 4000$, the element-wise LMMSE-type channel estimation outperforms the LMMSE-type channel estimation

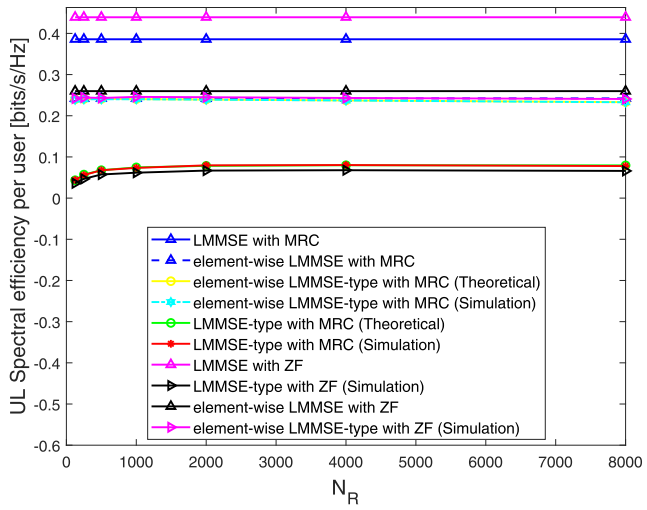
[5] In practice, ZF is a good choice for beamforming [21].

$$b = \kappa_1 \mathrm{tr}(\bar{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\bar{\mathbf{W}}_{ju}^H\bar{\mathbf{R}}_s) + \sum_{l=1}^{L}\left\{\kappa_2|\mathrm{tr}(\bar{\mathbf{W}}_{ju}^H\mathbf{R}_{jlu})|^2 + \frac{\kappa_1}{N_Q}\mathrm{tr}(\bar{\mathbf{W}}_{ju}^H\bar{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\mathbf{W}_{lu}^H\mathbf{W}_{lu}\mathbf{Q}_{ju})\right\} - a + d$$

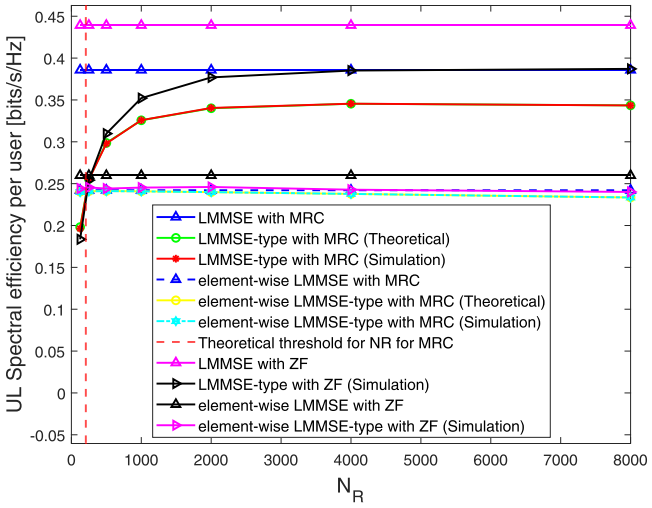
$$c = \frac{\alpha_R^2\kappa_1}{2}\left\{M\mathrm{tr}(\bar{\mathbf{R}}_s\mathbf{Q}_{ju}) + \mathrm{tr}(\mathbf{W}_{ju})\mathrm{tr}(\bar{\mathbf{R}}_s\mathbf{R}_{jju})\right\} + \frac{\alpha_R^2\kappa_2}{2}\sum_{l=1}^{L}\left\{\mathrm{tr}(\mathbf{W}_{lu}\mathbf{Q}_{ju}\mathbf{W}_{lu}^H\mathbf{Q}_{ju}) + \mathrm{tr}(\mathbf{W}_{lu}\mathbf{R}_{jju}\mathbf{W}_{lu}^H\mathbf{R}_{jju})\right\}$$
$$+ \frac{\alpha_R^2\kappa_1}{2N_Q}\sum_{l=1}^{L}\left\{M\mathrm{tr}(\mathbf{W}_{jlu}^2\mathbf{Q}_{ju}^2) + \mathrm{tr}(\mathbf{W}_{ju})\mathrm{tr}(\mathbf{W}_{jlu}^2\mathbf{Q}_{ju}\mathbf{R}_{jju})\right\}$$

$$g = \kappa_3\left\{\mathrm{tr}(\bar{\mathbf{W}}_{ju}\mathbf{Q}_{ju}\bar{\mathbf{W}}_{ju}^H\bar{\mathbf{R}}_s) + \sum_{l=1}^{L}|\mathrm{tr}(\bar{\mathbf{W}}_{ju}^H\mathbf{S}_{jlu})|^2\right\} + \kappa_4\left\{\mathrm{tr}(\bar{\mathbf{W}}_{ju}\mathbf{P}_{ju}\bar{\mathbf{W}}_{ju}^H\bar{\mathbf{S}}_s) + \sum_{l=1}^{L}\mathrm{tr}(\bar{\mathbf{W}}_{ju}^2\mathbf{S}_{jlu}^2)\right\} - f + d$$

$$h = \frac{\alpha_R^2\kappa_3}{2}\mathrm{tr}\Big(\mathbf{P}_{ju}^{-1}\mathbf{Q}_{ju}\mathbf{P}_{ju}^{-1}\{\bar{\mathbf{R}}_s \circ \mathbf{Q}_{ju} \circ \mathbf{Q}_{ju}\} + \mathbf{P}_{ju}^{-1}\mathbf{Q}_{ju}\mathbf{P}_{ju}^{-1}\{\bar{\mathbf{R}}_s \circ \mathbf{R}_{jju} \circ \mathbf{R}_{jju}\}\Big) + \frac{\alpha_R^2\kappa_4}{2}\Big\{\mathrm{tr}(\bar{\mathbf{S}}_s\mathbf{P}_{ju}) + \mathrm{tr}(\mathbf{W}_{ju}\bar{\mathbf{S}}_s\mathbf{S}_{jju})$$
$$+ \sum_{l=1}^{L}\mathrm{tr}(\mathbf{W}_{lu}^2\mathbf{P}_{ju}^2) + \sum_{l=1}^{L}\mathrm{tr}(\mathbf{W}_{lu}^2\mathbf{S}_{jju}^2)\Big\} + \frac{\alpha_R^2\kappa_3}{2}\sum_{p=1}^{M}\sum_{q=1}^{M}\Big\{[\mathbf{W}_{lu}(\mathbf{Q}_{ju} \circ \mathbf{Q}_{ju})\mathbf{W}_{lu}]_{pq} + [\mathbf{W}_{lu}(\mathbf{R}_{jju} \circ \mathbf{R}_{jju})\mathbf{W}_{lu}]_{pq}\Big\}$$



(a) SE vs N_R with $N_Q = 125$.



(b) SE vs N_R with $N_Q = 4000$.

Fig. 3. UL SE for different channel estimation techniques.

for $N_R = 125$. Moreover, for $N_Q = 4000$, $\bar{N}_R$ given in Section IV matches exactly with the N_R value for which the LMMSE-type and element-wise LMMSE-type channel estimations have the same performance. Therefore, the minimum SE guaranteed for a massive MIMO system with imperfect covariance information is the SE provided by the element-wise LMMSE channel estimator.[6] This SE can be achieved by using element-wise LMMSE-type channel estimation with very low values of N_R and N_Q, and with low computational complexity. From simulations, we also observe the threshold value for N_Q to be 263, such that for $N_Q < 263$, element-wise LMMSE-type channel estimation always outperforms LMMSE-type channel estimation.

In Fig. 3, it can also be seen that the SE simulation curves corresponding to ZF combining behave similarly to the case of MRC combining. ZF combining performs well only for the larger number of pilots ($N_Q = 4000$ and $N_R \geq 500$) and needs additional computational complexity, but it does not significantly improve the performance of element-wise LMMSE channel estimation and it gives only marginally better performance than that corresponding to MRC combining. Moreover, for $N_Q = 4000$, the SE curve for LMMSE-type channel estimation crosses the SE for element-wise LMMSE-type channel estimation closer to the $\bar{N}_R$ value theoretically computed for MRC combining (44). For large N_R and N_Q values, ZF combining outperforms MRC combining. This is due to better covariance estimates. Therefore, the SE expressions derived in this paper serve as conservative bounds for an achievable spectral efficiency of the system considered.

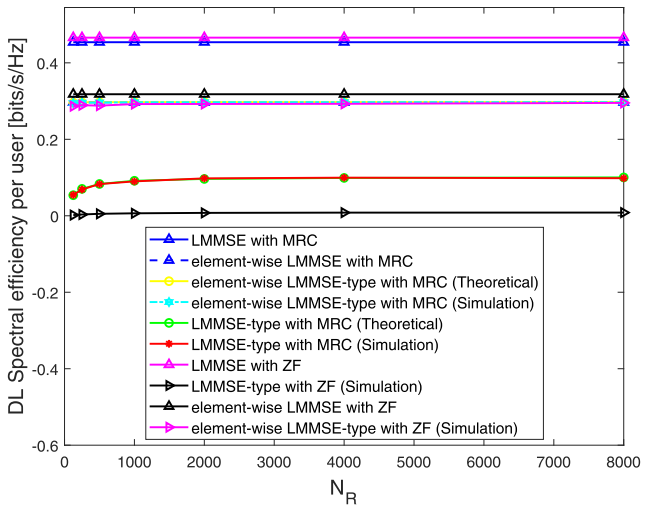
[6]Note that the objective is to have N_R and N_Q as low as possible for guaranteeing a desired SE.

B. Downlink Spectral Efficiency

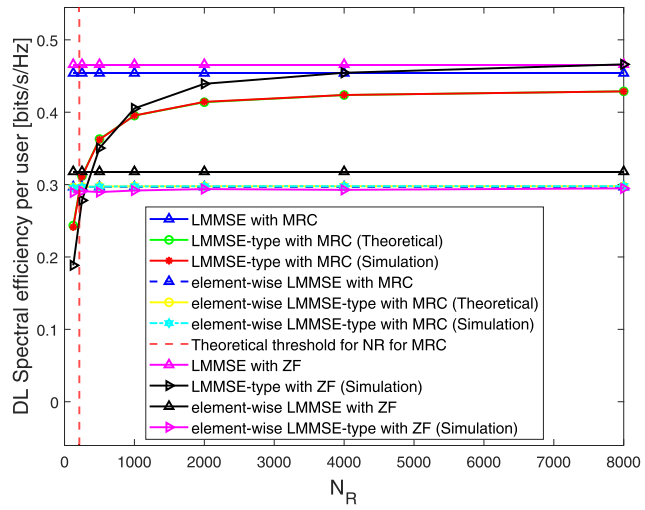
Similar to the UL example, in this simulation example, we consider the DL SE expressions that correspond to the two channel estimation techniques: LMMSE-type channel estimation and the element-wise LMMSE-type channel estimation, each in combination with two beamforming techniques, which are matched filter precoding and ZF precoding. In Fig. 4, we plot the SE as a function of N_R for the two aforementioned channel estimation techniques. Fig. 4(a) depicts the SE values for $N_Q = 125$, and Fig. 4(b) shows SE values for $N_Q = 4000$. We perform a study on these plots similar to the study done in Subsection V-A.

It can be observed from Fig. 4 that the DL SE plots are similar to the plots in Subsection V-A. However, unlike in UL SE, an increase in N_R does not result in a drop in SE as there is no pilot overhead in DL. The simulated SEs match the theoretical values for both the channel estimation techniques used, thereby validating the derivations presented in the paper. Moreover, for $N_Q = 4000$, $\bar{N}_R$ given in Section IV matches exactly with the N_R value for which LMMSE-type and element-wise LMMSE-type channel estimations have the same performance. From Fig. 4(a) and Fig. 4(b), the minimum DL SE guaranteed for a massive MIMO system with imperfect covariance information is the SE provided by element-wise LMMSE channel estimator. This SE can be achieved by using element-wise LMMSE-type channel estimation with very low values of N_R and N_Q, with low computational complexity. From simulations, we also compute the threshold value for N_Q to be 272, such that for $N_Q < 272$, element-wise LMMSE-type channel estimation always outperforms LMMSE-type channel estimation.

It can also be noticed from Fig. 4 that the SE simulation curves corresponding to ZF precoding behave similarly to the case of matched filter precoding. ZF precoding performs well only for the larger number of pilots ($N_Q = 4000$ and $N_R \geq 500$) and needs additional computational complexity, but it does not significantly improve the performance corresponding to element-wise LMMSE channel estimation, and it gives only marginally better performance than that corresponding to matched filter precoding. Moreover, for $N_Q = 4000$, the SE curve for LMMSE-type channel estimation crosses the SE for element-wise LMMSE-type channel estimation closer to the $\bar{N}_R$ value theoretically computed for matched filter precoding (44). For large N_R and N_Q values, ZF precoding



(a) SE vs N_R with $N_Q = 125$.



(b) SE vs N_R with $N_Q = 4000$.

Fig. 4. DL SE for different channel estimation techniques.

outperforms the matched filter precoding. This is due to better covariance estimates. Therefore, the SE expressions derived in this paper serve as conservative bounds for an achievable spectral efficiency of the system considered.

VI. Conclusion

We have derived closed-form expressions for average UL and DL SEs of a massive MIMO system that implements MRC and matched filter precoder, respectively, as a function of N_R and N_Q, which represent the UL pilot overhead. These combiners use channel estimates that utilize estimated covariance matrices in addition to channel observations. The LMMSE-type and element-wise LMMSE-type channel estimates have been considered. Using theoretical analysis of these closed-form expressions as well as simulation results, we have demonstrated the impact of different values of N_R and N_Q on SEs of a user in a massive MIMO system, thereby presenting the closed-form expressions as the tools for solving the problem of choosing these parameters optimally. Based on numerical study, we have demonstrated that the ZF beamforming does not significantly improve the SE for the case of element-wise LMMSE-type channel estimation. It is useful for the case of LMMSE-type channel estimation but at the cost of large pilot overhead and computational complexity. Furthermore, we have shown that the choice of pilot overhead made based on the LS beamforming serves as a more conservative result than for the ZF beamforming case, but accurate and very useful estimate of the pilot overhead. Finally, we have shown that the element-wise LMMSE-type channel estimator with very low N_R and N_Q and with simple LS combiner provides the minimum SE guarantee with low computational complexity.

Appendix A
Proof of Lemma 1

Let us start with a proof of (17). Let the rank of the covariance matrix of $\mathbf{h}$, $\mathbf{R}$, be K. Then, we denote $\mathbf{\Lambda} \in \mathbb{R}^{K\times K}$ is a diagonal matrix containing positive eigenvalues of $\mathbf{R}$ and $\mathbf{U} \in \mathbb{R}^{M\times K}$ is a matrix containing K eigenvectors corresponding to eigenvalues. Now, let us also define $\mathbf{B} \triangleq \mathbf{U}\mathbf{\Lambda}^{1/2} \in \mathbb{C}^{M\times K}$. Then, there exists a unique $\mathbf{g} \in \mathbb{C}^K$ such that $\mathbf{h} = \mathbf{B}\mathbf{g}$ and $\mathbb{E}\{\mathbf{g}\mathbf{g}^H\} = \mathbf{I}$. Therefore, we have $\mathbb{E}\{\mathbf{h}\mathbf{h}^H\mathbf{A}\mathbf{h}\mathbf{h}^H\} = \mathbf{B}\mathbb{E}\{\mathbf{g}\mathbf{g}^H\tilde{\mathbf{A}}\mathbf{g}\mathbf{g}^H\}\mathbf{B}^H$ where $\tilde{\mathbf{A}} \triangleq \mathbf{B}^H\mathbf{A}\mathbf{B}$. However, since $\mathbf{g}$ is distributed as $\mathcal{CN}(\mathbf{0},\mathbf{I})$, the term $\mathbb{E}\{\mathbf{g}\mathbf{g}^H\tilde{\mathbf{A}}\mathbf{g}\mathbf{g}^H\}$ can be evaluated as follows

$$\mathbb{E}\{[\mathbf{g}\mathbf{g}^H\tilde{\mathbf{A}}\mathbf{g}\mathbf{g}^H]_{ij}\} = \sum_{p=1}^{K}\sum_{q=1}^{K}\mathbb{E}\{[\mathbf{g}]_i[\mathbf{g}]_p^*[\mathbf{g}]_q[\mathbf{g}]_j^*\}[\tilde{\mathbf{A}}]_{pq} = \begin{cases}[\tilde{\mathbf{A}}]_{ij} & \text{if } i \neq j\\ [\tilde{\mathbf{A}}]_{ii} + \mathrm{tr}(\tilde{\mathbf{A}}) & \text{otherwise}\end{cases}$$

and $\mathbb{E}\{\mathbf{g}\mathbf{g}^H\tilde{\mathbf{A}}\mathbf{g}\mathbf{g}^H\} = \tilde{\mathbf{A}} + \mathbf{I}\mathrm{tr}(\tilde{\mathbf{A}})$. Therefore, $\mathbb{E}\{\mathbf{h}\mathbf{h}^H\mathbf{A}\mathbf{h}\mathbf{h}^H\} = \mathbf{R}\mathbf{A}\mathbf{R} + \mathbf{R}\mathrm{tr}(\mathbf{A}\mathbf{R})$.

Proof of (18) is as follows. We first compute that $\mathbb{E}\{|\mathbf{h}^H\mathbf{A}\mathbf{h}|^2\} = \mathbb{E}\{\mathbf{h}^H\mathbf{A}\mathbf{h}\mathbf{h}^H\mathbf{A}^H\mathbf{h}\} = \mathbb{E}\{\mathrm{tr}(\mathbf{A}\mathbf{h}\mathbf{h}^H\mathbf{A}^H\mathbf{h}\mathbf{h}^H)\}$. Using (17), we have $\mathbb{E}\{|\mathbf{h}^H\mathbf{A}\mathbf{h}|^2\} = |\mathrm{tr}(\mathbf{A}^H\mathbf{R})|^2 + \mathrm{tr}(\mathbf{A}\mathbf{R}\mathbf{A}^H\mathbf{R})$.

Appendix B
Proof of Lemma 2

Proof of (19) and (20) is given in [22].

Using the eigenvalue decomposition of $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, let us define $\tilde{\mathbf{X}} \triangleq \mathbf{U}^H\mathbf{X}\mathbf{U}$. It should be noted that $\tilde{\mathbf{X}}$ is distributed as $\mathcal{W}(N,\mathbf{I})$. Then, (21) can be proved as follows. First, we compute the following expectation term.

$$\mathbb{E}\{\mathrm{tr}(\mathbf{X}^{-2}\mathbf{C})\} = \mathbb{E}\{\mathrm{tr}(\tilde{\mathbf{X}}^{-2}\mathbf{\Lambda})\} = \sum_{i=1}^{M}[\mathbb{E}\{\tilde{\mathbf{X}}^{-2}\}]_{ii}[\mathbf{\Lambda}]_{ii}$$

But from (20), we have

$$\mathbb{E}\{\mathrm{tr}(\mathbf{X}^{-2}\mathbf{C})\} = \sum_{i=1}^{M}\frac{N}{(N-M)^3-(N-M)}[\mathbf{\Lambda}]_{ii} = \frac{N}{(N-M)^3-(N-M)}\mathrm{tr}(\mathbf{C})$$

For (21), we expand $\mathbb{E}\{|\mathrm{tr}(\mathbf{X}^{-1}\mathbf{A})|^2\}$ using (20) as follows.

$$\begin{aligned}
&\mathbb{E}\{|\mathrm{tr}(\mathbf{X}^{-1}\mathbf{A})|^2\}\\
&\quad=\sum_{p=1}^{M}\sum_{q=1}^{M}\sum_{r=1}^{M}\sum_{s=1}^{M}\mathbb{E}\{[\mathbf{X}^{-1}]_{pp}[\mathbf{X}^{-1}]_{ss}\}[\mathbf{A}]_{pp}[\mathbf{A}^H]_{ss}\\
&\quad=\sum_{p=1}^{M}\mathbb{E}\{[\mathbf{X}^{-1}]_{pp}[\mathbf{X}^{-1}]_{pp}\}[\mathbf{A}]_{pp}[\mathbf{A}^H]_{pp}\\
&\qquad+\sum_{p=1}^{M}\sum_{s=1,s\neq p}^{M}\mathbb{E}\{[\mathbf{X}^{-1}]_{pp}[\mathbf{X}^{-1}]_{ss}\}[\mathbf{A}]_{pp}[\mathbf{A}^H]_{ss}\\
&\qquad+\sum_{p=1}^{M}\sum_{s=1,s\neq p}^{M}\mathbb{E}\{[\mathbf{X}^{-1}]_{ps}[\mathbf{X}^{-1}]_{sp}\}[\mathbf{A}]_{sp}[\mathbf{A}^H]_{ps}
\end{aligned}$$

Using (20), the above equation can be further simplified to (22).

Appendix C
Proof of Lemma 3

Let us define a pair of mutually independent random vectors as follows.

$$\mathbf{g}^{(1)}_{jju}[n]\triangleq\hat{\mathbf{h}}^{(1)}_{jju}[n]-\mathbf{h}_{jju},\ \mathbf{g}^{(2)}_{jju}[n]\triangleq\hat{\mathbf{h}}^{(2)}_{jju}[n]-\mathbf{h}_{jju}$$

The covariance matrices for $\mathbf{g}^{(1)}[n]$ and $\mathbf{g}^{(2)}[n]$ are identically equal to $\mathbf{Q}_{ju}-\mathbf{R}_{jju}$. Additionally, we also define mutually independent set of matrices

$$\breve{\mathbf{R}}_{jju}[n]\triangleq\hat{\mathbf{h}}^{(1)}_{jju}[n](\hat{\mathbf{h}}^{(2)}_{jju}[n])^H+\hat{\mathbf{h}}^{(2)}_{jju}[n](\hat{\mathbf{h}}^{(1)}_{jju}[n])^H$$

for all $n\in\{1\dots N_R\}$ such that $\ddot{\mathbf{R}}_{jju}=\frac{1}{N_R}\sum_{n=1}^{N}\breve{\mathbf{R}}_{jju}[n]$ by definition (i.e., (10)).

Using the definition of $\mathbf{g}^{(1)}_{jju}[n]$ and $\mathbf{g}^{(2)}_{jju}[n]$, and also using Lemma 1, it can be shown that, for all $n=\{1\dots N_R\}$, we have

$$\begin{aligned}
\mathbb{E}\{\breve{\mathbf{R}}_{jju}[n]\mathbf{A}\breve{\mathbf{R}}_{jju}[n]\}&=\mathbf{R}_{jju}\mathbf{A}\mathbf{R}_{jju}+\frac{1}{2}\mathbf{Q}_{ju}\mathrm{tr}(\mathbf{A}\mathbf{Q}_{ju})\\
&\quad+\frac{1}{2}\mathbf{R}_{jju}\mathrm{tr}(\mathbf{A}\mathbf{R}_{jju})\qquad(45)\\
\mathbb{E}\{|\mathrm{tr}(\breve{\mathbf{R}}_{jju}[n]\mathbf{A})|^2\}&=|\mathrm{tr}(\mathbf{R}_{jju}\mathbf{A})|^2+\frac{1}{2}\mathrm{tr}(\mathbf{A}\mathbf{Q}_{ju}\mathbf{A}^H\mathbf{Q}_{ju})\\
&\quad+\frac{1}{2}\mathrm{tr}(\mathbf{A}\mathbf{R}_{jju}\mathbf{A}^H\mathbf{R}_{jju}).\qquad(46)
\end{aligned}$$

Finally, along with the equation $\ddot{\mathbf{R}}_{jju}=\frac{1}{N_R}\sum_{n=1}^{N}\breve{\mathbf{R}}_{jju}[n]$, (45) and (46) will result in (23) and (24), respectively.

Appendix D
Proof of Lemma 5

Since $\mathbf{Y}=\mathbf{Z}/2$, and the elements of the diagonal matrix $\mathbf{Z}$ are χ^2 distributed with $2N$ degrees of freedom, we have $\mathbb{E}\{[\mathbf{Y}^{-1}]_{pp}\}=2\mathbb{E}\{[\mathbf{Z}^{-1}]_{pp}\}=1/(N-1)$ and $\mathbb{E}\{[\mathbf{Y}^{-1}]^2_{pp}\}=4\mathbb{E}\{[\mathbf{Z}^{-1}]^2_{pp}\}=1/(N-1)(N-2)$.

Using the above results, (31) can be derived as follows

$$\begin{aligned}
&\mathbb{E}\{\mathrm{tr}(\mathbf{Y}^{-1}\mathbf{A}_1\mathbf{Y}^{-1}\mathbf{A}_2)\}\\
&\quad=\left(\frac{1}{N-1}\right)^2\sum_{p=1}^{M}\sum_{q\neq p}[\mathbf{A}_1]_{pq}[\mathbf{A}_2]_{qp}\\
&\qquad+\frac{1}{(N-1)(N-2)}\sum_{p=1}^{M}[\mathbf{A}_1]_{pp}[\mathbf{A}_2]_{pp}\\
&\quad=\tau_1\mathrm{tr}(\mathbf{A}_1\mathbf{A}_2)+\tau_2\mathrm{tr}(\mathbf{A}_{1d}\mathbf{A}_{2d})
\end{aligned}$$

where $\tau_1\triangleq 1/(N-1)^2$, $\tau_2\triangleq 1/((N-1)^2(N-2))$, $\mathbf{A}_{1d}\triangleq\mathrm{diag}(\mathbf{A}_1)$, and $\mathbf{A}_{2d}\triangleq\mathrm{diag}(\mathbf{A}_2)$.

In what follows, proof of (32) is presented

$$\begin{aligned}
\mathbb{E}\{|\mathrm{tr}(\mathbf{Y}^{-1}\mathbf{A})|^2\}&=\frac{1}{(N-1)^2}\sum_{p=1}^{M}\sum_{q\neq p}[\mathbf{A}]_{pp}[\mathbf{A}]^*_{qq}\\
&\quad+\frac{1}{(N-1)(N-2)}\sum_{p=1}^{M}|[\mathbf{A}]_{pp}|^2\\
&=\tau_1|\mathrm{tr}(\mathbf{A})|^2+\tau_2\mathrm{tr}(\mathbf{A}_d^H\mathbf{A}_d)
\end{aligned}$$

where $\mathbf{A}_d\triangleq\mathrm{diag}(\mathbf{A})$.

Appendix E
Proof of Lemma 6

Let us define a pair of mutually independent random vectors as follows.

$$\mathbf{g}^{(1)}_{jju}[n]\triangleq\hat{\mathbf{h}}^{(1)}_{jju}[n]-\mathbf{h}_{jju},\ \mathbf{g}^{(2)}_{jju}[n]\triangleq\hat{\mathbf{h}}^{(2)}_{jju}[n]-\mathbf{h}_{jju}$$

The covariance matrices for $\mathbf{g}^{(1)}_{jju}[n]$ and $\mathbf{g}^{(2)}_{jju}[n]$ are identically equal to $\mathbf{Q}_{ju}-\mathbf{R}_{jju}$. Additionally, we also define mutually independent set of matrices as $\breve{\mathbf{S}}_{jjk}[n]\triangleq\mathrm{diag}(\hat{\mathbf{h}}^{(1)}_{jju}[n](\hat{\mathbf{h}}^{(2)}_{jju}[n])^H+\hat{\mathbf{h}}^{(2)}_{jju}[n](\hat{\mathbf{h}}^{(1)}_{jju}[n])^H)$ for all $n\in\{1\dots N_R\}$ such that $\ddot{\mathbf{S}}_{jju}=\frac{1}{N}\sum_{n=1}^{N}\breve{\mathbf{S}}_{jju}[n]$ by definition (i.e., (12)).

Using the definitions of $\mathbf{g}^{(1)}_{jju}[n]$ and $\mathbf{g}^{(1)}_{jju}[n]$ together with Lemma 1 (for scalar case), and Lemma 4, it can be shown that

$$\begin{aligned}
&\mathbb{E}\{[\breve{\mathbf{S}}_{jju}]_{pp}[\breve{\mathbf{S}}_{jju}]_{qq}\}\\
&=\mathbb{E}\{|[\mathbf{h}_{jju}]_p|^2|[\mathbf{h}_{jju}]_q|^2\}\\
&\quad+\frac{1}{2}[\mathbf{R}_{jju}]_{pq}[\mathbf{Q}_{ju}-\mathbf{R}_{jju}]_{qp}+\frac{1}{2}[\mathbf{Q}_{ju}-\mathbf{R}_{jju}]_{pq}[\mathbf{R}_{jju}]_{qp}\\
&\quad+\frac{1}{2}[\mathbf{Q}_{ju}-\mathbf{R}_{jju}]_{pq}[\mathbf{Q}_{ju}-\mathbf{R}_{jju}]_{qp}\\
&=[\mathbf{S}_{jju}]_{pp}[\mathbf{S}_{jju}]_{qq}+\frac{1}{2}[\mathbf{Q}_{jju}]_{pq}[\mathbf{Q}_{jju}]_{pq}\\
&\quad+\frac{1}{2}[\mathbf{R}_{jju}]_{pq}[\mathbf{R}_{jju}]_{pq}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\mathbb{E}\{[\breve{\mathbf{S}}_{jju}\mathbf{A}\breve{\mathbf{S}}_{jju}]_{pq}\}\\
&\quad=[\mathbf{A}]_{pq}\big\{[\mathbf{S}_{jju}]_{pp}[\mathbf{S}_{jju}]_{qq}+\frac{1}{2}[\mathbf{Q}_{jju}]_{pq}[\mathbf{Q}_{jju}]_{pq}\\
&\qquad+\frac{1}{2}[\mathbf{R}_{jju}]_{pq}[\mathbf{R}_{jju}]_{pq}\big\}\qquad(47)\\
&\mathbb{E}\{|\mathrm{tr}(\breve{\mathbf{S}}_{jju}\mathbf{D})|^2\}\\
&\quad=\sum_{p=1}^{M}\sum_{q=1}^{M}\Big\{[\mathbf{S}_{jju}]_{pp}[\mathbf{S}_{jju}]_{qq}+\frac{1}{2}[\mathbf{Q}_{ju}]_{pq}[\mathbf{Q}_{ju}]_{pq}\\
&\qquad+\frac{1}{2}[\mathbf{R}_{jju}]_{pq}[\mathbf{R}_{jju}]_{pq}\Big\}[\mathbf{D}]_{pp}[\mathbf{D}]_{qq}
\end{aligned}$$

$$= |\mathrm{tr}(\mathbf{S}_{jju}\mathbf{D})|^2 + \frac{1}{2}\sum_{p=1}^{M}\sum_{q=1}^{M}[\mathbf{D}(\mathbf{Q}_{ju}\circ\mathbf{Q}_{ju})\mathbf{D}]_{pq} + \frac{1}{2}\sum_{p=1}^{M}\sum_{q=1}^{M}[\mathbf{D}(\mathbf{R}_{jju}\circ\mathbf{R}_{jju})\mathbf{D}]_{pq}. \quad (48)$$

Finally, along with the equation $\ddot{\mathbf{S}}_{jju} = \frac{1}{N}\sum_{n=1}^{N}\breve{\mathbf{S}}_{jju}[n]$, (47) and (48) will result in (33) and (34), respectively.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[5] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[6] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[7] 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.

[8] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[9] A. Pitarokoilis, E. Björnson, and E. G. Larsson, "On the effect of imperfect timing synchronization on pilot contamination," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[10] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.

[11] S. Haghighatshoar and G. Caire, "Massive MIMO pilot decontamination and channel interpolation via wideband sparse channel estimation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8316–8332, Dec. 2017.

[12] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse Bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, Mar. 2019.

[13] 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.

[14] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 863–867, Jun. 2018.

[15] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. 50th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 974–978.

[16] K. Upadhya and S. A. Vorobyov, "Covariance matrix estimation for massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 546–550, Apr. 2018.

[17] A. K. Kocharlakota, K. Upadhya, and S. A. Vorobyov, "On achievable rates for massive MIMO system with imperfect channel covariance information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4504–4508.

[18] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, Jul. 2015.

[19] I. Viering, H. Hofstetter, and W. Utschick, "Spatial long-term variations in urban, rural and indoor environments," in *Proc. 5th Meeting COST*, Lisbon, Portugal, vol. 273, Sep. 2002, pp. 1–10.

[20] *Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document 3GPP, (TR) 38.901, version 16.1.0, Dec. 2019. [Online]. Available: http://www.3gpp.org/specs/archive/38_series/38.901/38901-g10.zip

[21] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.

[22] D. Maiwald and D. Kraus, "On moments of complex Wishart and complex inverse Wishart distributed matrices," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Munich, Germany, vol. 5, Apr. 1997, pp. 3817–3820.

Atchutaram K. Kocharlakota (Student Member, IEEE) received the M.Tech. degree from IIT Madras, India, in 2011. He is currently pursuing the D.Sc.(Tech.) degree from Aalto University, Finland. Prior to his Ph.D., he was part of the Research and Development Team, Intel Mobile Communications, India, from 2011 to 2017. His research interests include massive MIMO, cell-free massive MIMO, and intelligent reflecting surface aided communications.

Karthik Upadhya (Member, IEEE) received the M.Tech. degree from IIT Madras, India, in 2011, and the D.Sc.(Tech.) degree from Aalto University, Finland, in 2018. Prior to his Ph.D., he was a Member of Technical Staff with Saankhya Labs, India, from 2011 to 2013, and a Research Assistant with Indian Institute of Science, India, from 2013 to 2014. He was a Visiting Researcher with the Wireless Networking and Communications Group, The University of Texas at Austin, in 2017. He is currently with Nokia Bell Labs, Espoo, Finland. His research interests include massive MIMO and physical layer security.

Sergiy A. Vorobyov (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in systems and control from Kharkiv National University of Radio Electronics, Kharkiv, Ukraine, in 1994 and 1997, respectively.

He is currently a Professor with the Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland. He has been previously with the University of Alberta, AB, Canada, as an Assistant Professor and an Associate Professor, and then as a Full Professor. Since his graduation, he has been holding various research and faculty positions with Kharkiv National University of Radio Electronics; the Institute of Physical and Chemical Research, Japan; McMaster University, Canada; the University of Duisburg-Essen, Germany; Darmstadt University of Technology, Germany; and the Joint Research Institute between Heriot-Watt University and The University of Edinburgh, U.K. His research interests include optimization and multi-liner algebra methods in signal processing and data analysis, statistical and array signal processing, sparse signal processing, estimation, detection, and learning theory and methods, and multi-antenna, very large, cooperative, and cognitive systems. He was a member of the Sensor Array and Multi-Channel Signal Processing Committee of the IEEE Signal Processing Society from 2007 to 2012 and the Signal Processing for Communications and Networking Technical Committee of the IEEE Signal Processing Society from 2010 to 2016. He was a recipient of the 2004 IEEE Signal Processing Society Best Paper Award, the 2007 Alberta Ingenuity New Faculty Award, the 2011 Carl Zeiss Award (Germany), the 2012 NSERC Discovery Accelerator Award, and other awards. He was the Track Chair of Asilomar 2011, Pacific Grove, CA, USA; and the Tutorial Chair of ISWCS 2013, Ilmenau, Germany. He was the Technical Co-Chair of IEEE CAMSAP 2011, Puerto Rico; IEEE SAM 2018, Sheffield, U.K.; and IEEE CAMSAP 2021, Costa-Rica. He was an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2006 to 2010 and IEEE SIGNAL PROCESSING LETTERS from 2007 to 2009. He has been a Senior Area Editor of IEEE SIGNAL PROCESSING LETTERS since 2016.