

Large-scale-fading decoding in cellular Massive MIMO systems with spatially correlated channels

Trinh Van Chien, Christopher Mollén and Emil Björnson

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-156657>

N.B.: When citing this work, cite the original publication.

Van Chien, T., Mollén, C., Björnson, E., (2019), Large-scale-fading decoding in cellular Massive MIMO systems with spatially correlated channels, *IEEE Transactions on Communications*, 67(4), 2746-2762. <https://doi.org/10.1109/TCOMM.2018.2889090>

Original publication available at:

<https://doi.org/10.1109/TCOMM.2018.2889090>

Copyright: IEEE Communications Society

<http://www.ieee.org/>

©2019 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.



Large-Scale-Fading Decoding in Cellular Massive MIMO Systems with Spatially Correlated Channels

Trinh Van Chien, Student Member, IEEE, Christopher Mollén, Emil Björnson, Senior Member, IEEE

Abstract—Massive multiple-input–multiple-output (MIMO) systems can suffer from coherent intercell interference due to the phenomenon of pilot contamination. This paper investigates a two-layer decoding method that mitigates both coherent and non-coherent interference in multi-cell Massive MIMO. To this end, each base station (BS) first estimates the channels to intra-cell users using either minimum mean-squared error (MMSE) or element-wise MMSE (EW-MMSE) estimation based on uplink pilots. The estimates are used for local decoding on each BS followed by a second decoding layer where the BSs cooperate to mitigate inter-cell interference. An uplink achievable spectral efficiency (SE) expression is computed for arbitrary two-layer decoding schemes. A closed-form expression is then obtained for correlated Rayleigh fading, maximum-ratio combining, and the proposed large-scale fading decoding (LSFD) in the second layer. We also formulate a sum SE maximization problem with both the data power and LSFD vectors as optimization variables. Since this is an NP-hard problem, we develop a low-complexity algorithm based on the weighted MMSE approach to obtain a local optimum. The numerical results show that both data power control and LSFD improve the sum SE performance over single-layer decoding multi-cell Massive MIMO systems.

Index Terms—Massive MIMO, Large-Scale Fading Decoding, Sum Spectral Efficiency Optimization, Channel Estimation.

I. Introduction

MASSIVE MIMO BSs, which are equipped with hundreds of antennas, exploit channel reciprocity to estimate the channel based on uplink pilots and spatially multiplex a large number of users on the same time–frequency resource [2], [3]. It is a promising technique to meet the growing demand for wireless data traffic of tomorrow [4], [5]. In a single-cell scenario, there is no need for computationally heavy decoding or precoding methods in Massive MIMO, such as successive interference cancellation or dirty paper coding. Linear processing schemes (e.g., zero-forcing combining) can effectively suppress interference and noise if the BS is equipped with a large number of antennas [6]. In a multi-cell scenario, however, pilot-based channel estimation is contaminated by the non-orthogonal transmission in other cells. This results in coherent intercell interference in the data transmission,

This paper was supported by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 641985 (5Gwireless). It was also supported by ELLIIT and CENIIT. Parts of this paper were submitted to IEEE International Conference on Communications 2019 [1].

T. V. Chien and E. Björnson are with the Department of Electrical Engineering (ISY), Linköping University, 581 83 Linköping, Sweden (e-mail: trinh.van.chien@liu.se, emil.bjornson@liu.se).

C. Mollén was with the Department of Electrical Engineering (ISY), Linköping University, 581 83 Linköping, Sweden, when this work was done (e-mail: chris.mollen@gmail.com).

so-called pilot contamination [7], unless some advanced processing schemes are used to suppress it [8]. Pilot contamination causes the gain of having more antennas to decrease and the SE of linear decoding methods, such as maximum-ratio combining (MRC) or zero-forcing, to saturate as the number of antennas grows.

Much work has been done to mitigate the effects of pilot contamination. The first and intuitive approach to mitigate pilot contamination is to increase the length of the pilots. In practical networks, however, it is not possible to make all pilots orthogonal due to the limited channel coherence block [9]. Hence, there is a trade-off between having longer pilots and low pilot overhead. Another method to mitigate pilot contamination is to assign the pilots in a way that reduces the contamination [10], since only a few users from other cells cause substantial contamination. The pilot assignment is a combinatorial problem and heuristic algorithms with low computational complexity can be developed to mitigate the pilot contamination. In [11], a greedy pilot assignment method is developed that exploits the statistical channel information and mutual interference between users. Pilot assignment approaches still suffer from asymptotic SE saturation since we only change one contaminating user for a less contaminating user. A third method is to utilize the spatial correlation to mitigate the coherent interference using multi-cell minimum-mean-square error (M-MMSE) decoding [8], but this method has high computational complexity.

Instead of combating pilot contamination, one can utilize it using more advanced decoding schemes [12]–[14]. This approach was initially called pilot contamination decoding since the BSs cooperate in suppressing the pilot contamination [12]. The original form of this technique used simplistic MRC, which does not suppress interference very well, thus it required a huge number of antennas to be effective [13]. The latest version of this decoding design, called large-scale fading decoding (LSFD) [14], was designed to be useful also with a practical number of antennas. In the two-layer LSFD framework, each BS applies an arbitrary local linear decoding method in the first layer, preferably one that suppresses intra-cell interference. The result is then gathered at a common central station that applies so-called LSFD vectors in a second-layer to combine the signals from multiple BSs to suppress pilot contamination and inter-cell interference. This new decoding design overcomes the aforementioned limitations in [12] and attains high SE even with a limited number of BS antennas.

To explain why LFSF vectors are necessary to mitigate pilot contamination, we consider a toy example comprising of two BSs, each serving one user with the same index as their BS. There are four different channels and $\mathbf{h}_{i,j} \sim \mathcal{CN}(\mathbf{0}, \beta_{i,j} \mathbf{I}_M)$ denotes the channel between BS i and user j for $i, j \in \{1, 2\}$. Let s_i denote the desired signal from the user in cell i . When using single-layer decoding with MRC, the noise vanishes as $M \rightarrow \infty$, but pilot contamination remains [3]. The resulting detected signals \hat{s}_1, \hat{s}_2 at the two BSs are then given by

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \begin{bmatrix} \beta_{1,1}s_1 + \beta_{1,2}s_2 \\ \beta_{2,1}s_1 + \beta_{2,2}s_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{bmatrix}}_{\triangleq \mathbf{B}} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}. \quad (1)$$

Since each BS observes a linear combination of the two signals, the asymptotic SE achieved with single-layer decoding is limited due to interference. However, in a two-layer decoding system, a central station can process \hat{s}_1 and \hat{s}_2 to remove the interference as follows:

$$\mathbf{B}^{-1} \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \mathbf{B}^{-1} \mathbf{B} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}. \quad (2)$$

The rows of the inverse matrix \mathbf{B}^{-1} are called the LFSF vectors and only depends on the statistical parameters $\beta_{1,1}, \beta_{1,2}, \beta_{2,1}, \beta_{2,2}$, so the central station does not need to know the instantaneous channels. Since the resulting signals in (2) are free from noise and interference, the network can achieve an unbounded SE as $M \rightarrow \infty$.

This motivating example, adapted from [12], exploits the fact that the channels are spatially uncorrelated and requires an infinite number of antennas. The prior works [12], [14] are only considering uncorrelated Rayleigh fading channels and rely on the particular asymptotic properties of that channel model. The generalization to more practical correlated channels is non-trivial and has not been considered until now.¹ In this paper, we consider spatially correlated channels with a finite number of antennas. We stress that these generalizations are practically important: if two-layer decoding will ever be implemented in practice, the channels will be subject to spatial correlation and the BSs will have a limited number of antennas.

A. Main Contributions

In this paper, we generalize the LFSF method from [12], [14] to a scenario with correlated Rayleigh fading and arbitrary first-layer decoders, and also develop a method for data power control in the system. We evaluate the performance by deriving an SE expression for the system. Our main contributions are summarized as follows:

¹The concurrent work [15] appeared just as we were submitting this paper. It contains the uplink SE for correlated Rayleigh fading described by the one-ring model and MMSE estimation, while we consider arbitrary spatial correlation and uses two types of channel estimators. Moreover, they consider joint power control and LFSF for max-min fairness, while we consider sum SE optimization, making the papers complementary.

- An uplink per-user SE is derived as a function of the second-layer decoding weights. A closed-form expression is then obtained for correlated Rayleigh fading and a system that uses MRC in the first decoding layer and an arbitrary choice of LFSF in the second layer. The second-layer decoding weights that maximize the SE follows in closed form.
- An uplink sum SE optimization problem with power constraints is formulated. Because it is non-convex and NP-hard, we propose an alternating optimization approach that converges to a local optimum with polynomial complexity.
- Numerical results demonstrate the effectiveness of two-layer decoding for Massive MIMO communication systems with correlated Rayleigh fading.

The rest of this paper is organized as follows: Multi-cell Massive MIMO with two-layer decoding is presented in Section II. An SE for the uplink together with the optimal LFSF design is presented in Section III. A maximization problem for the sum SE is formulated and a solution is proposed in Section IV. Numerical results in Section V demonstrate the performance of the proposed system. Section VI states the major conclusions of the paper.

Reproducible research: All the simulation results can be reproduced using the Matlab code and data files available at: <https://github.com/emilbjornson/large-scale-fading-decoding>

Notation: Lower and upper case bold letters are used for vectors and matrices. The expectation of the random variable X is denoted by $\mathbb{E}\{X\}$ and the Euclidean norm of the vector \mathbf{x} by $\|\mathbf{x}\|$. The transpose and Hermitian transpose of a matrix \mathbf{M} are written as \mathbf{M}^T and \mathbf{M}^H , respectively. The $L \times L$ -dimensional diagonal matrix with the diagonal elements d_1, d_2, \dots, d_L is denoted $\text{diag}(d_1, d_2, \dots, d_L)$. $\Re(\cdot)$ and $\Im(\cdot)$ are the real and imaginary parts of a complex number. $\nabla g(\mathbf{x})|_{\mathbf{x}_0}$ denotes the gradient of a multivariate function g at $\mathbf{x} = \mathbf{x}_0$. Finally, $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ is a vector of circularly symmetric, complex, jointly Gaussian distributed random variables with zero mean and correlation matrix \mathbf{R} .

II. System Model

We consider a network with L cells. Each cell consists of a BS equipped with M antennas that serves K single-antenna users.² The M -dimensional channel vector in the uplink between user k in cell l and BS l' is denoted by $\mathbf{h}'_{l,k} \in \mathbb{C}^M$. We consider the standard block-fading model, where the channels are static within a coherence block of size τ_c channel uses and assume one independent realization in each block, according to a stationary ergodic random process. Each channel follows a correlated Rayleigh fading model:

$$\mathbf{h}'_{l,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}'_{l,k}), \quad (3)$$

²In the uplink, the considered network consists of multiple interfering single-input multiple-output (SIMO) channels. Such a setup has been referred to as multiuser MIMO in the information theoretic literature for decades, which is why we adopt this terminology in the paper.

where $\mathbf{R}_{l,k}^{l'} \in \mathbb{C}^{M \times M}$ is the spatial correlation matrix of the channel. The BSs know the channel statistics, but have no prior knowledge of the channel realizations, which need to be estimated in every coherence block.

A. Channel Estimation

As in conventional Massive MIMO [8], the channels are estimated by letting the users transmit τ_p -symbol long pilots in a dedicated part of the coherence block, called the pilot phase. All the cells share a common set of $\tau_p = K$ mutually orthogonal pilots $\{\phi_1, \dots, \phi_K\}$, where the pilot $\phi_k \in \mathbb{C}^{\tau_p}$ spans τ_p symbols. Such orthogonal pilots are disjointly distributed among the K users in each cell:

$$\phi_k^H \phi_{k'} = \begin{cases} \tau_p & k = k', \\ 0 & k \neq k'. \end{cases} \quad (4)$$

Without loss of generality, we assume that all the users in different cells, which share the same index, use the same pilot and thereby cause pilot contamination to each other [3].

During the pilot phase, at BS l , the signals received in the pilot phase are collectively denoted by the $M \times \tau_p$ -dimensional matrix \mathbf{Y}_l and it is given by

$$\mathbf{Y}_l = \sum_{l'=1}^L \sum_{k=1}^K \sqrt{\hat{p}_{l',k}} \mathbf{h}_{l',k}^l \phi_k^H + \mathbf{N}_l, \quad (5)$$

where $\hat{p}_{l',k}$ is the power of the pilot of user k in cell l' and \mathbf{N}_l is a matrix of independent and identically distributed noise terms, each distributed as $\mathcal{CN}(0, \sigma^2)$.

An intermediate observation of the channel from user k to BS l is obtained through correlation with the pilot of user k in the following way:

$$\tilde{\mathbf{y}}_{l,k} = \mathbf{Y}_l \phi_k = \tau_p \sqrt{\hat{p}_{l,k}} \mathbf{h}_{l,k}^l + \sum_{\substack{l'=1 \\ l' \neq l}}^L \tau_p \sqrt{\hat{p}_{l',k}} \mathbf{h}_{l',k}^l + \tilde{\mathbf{n}}_{l,k}, \quad (6)$$

where $\tilde{\mathbf{n}}_{l,k} \triangleq \mathbf{N}_l \phi_k \sim \mathcal{CN}(\mathbf{0}, \tau_p \sigma^2 \mathbf{I}_M)$ are independent over l and k . The channel estimate and estimation error of the MMSE estimation of $\mathbf{h}_{l,k}^l$ is presented in the following lemma.

Lemma 1. If BS l uses MMSE estimation based on the observation in (6), the estimate of the channel between user k in cell l and BS l is

$$\hat{\mathbf{h}}_{l,k}^l = \sqrt{\hat{p}_{l,k}} \mathbf{R}_{l,k}^l \Psi_{l,k}^{-1} \tilde{\mathbf{y}}_{l,k}, \quad (7)$$

where $\Psi_{l,k} = \mathbb{E}\{\tilde{\mathbf{y}}_{l,k} (\tilde{\mathbf{y}}_{l,k})^H\} / \tau_p$ is given by

$$\Psi_{l,k} \triangleq \sum_{l'=1}^L \tau_p \hat{p}_{l',k} \mathbf{R}_{l',k}^l + \sigma^2 \mathbf{I}_M. \quad (8)$$

The channel estimate is distributed as

$$\hat{\mathbf{h}}_{l,k}^l \sim \mathcal{CN}(\mathbf{0}, \tau_p \hat{p}_{l,k} \mathbf{R}_{l,k}^l \Psi_{l,k}^{-1} \mathbf{R}_{l,k}^l), \quad (9)$$

and the channel estimation error, $\mathbf{e}_{l,k}^l \triangleq \mathbf{h}_{l,k}^l - \hat{\mathbf{h}}_{l,k}^l$, is independently distributed as

$$\mathbf{e}_{l,k}^l \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{l,k}^l - \tau_p \hat{p}_{l,k} \mathbf{R}_{l,k}^l \Psi_{l,k}^{-1} \mathbf{R}_{l,k}^l). \quad (10)$$

Proof. This lemma follows from adopting standard MMSE estimation results from [16], [17, Section 3] to our system model and notation. \square

Lemma 1 provides statistical information for the BS to construct the decoding and precoding vectors for the up- and downlink data transmission. However, to compute the MMSE estimate, the inverse matrix of $\Psi_{l,k}$ has to be computed for every user, which can lead to a computational complexity that might be infeasible when there are many antennas. This motivates us to use the simpler estimation technique called element-wise MMSE (EW-MMSE) [17].

To simplify the presentation, we make the standard assumption that the correlation matrix $\mathbf{R}_{l,k}^{l'}$ has equal diagonal elements, denoted by $\beta_{l,k}^{l'}$. This assumption is well motivated for elevated macro BSs that only observe far-field scattering effects from every cell. However, EW-MMSE estimation of the channel can also be done when the diagonal elements are different. The generalization to this case is straightforward. EW-MMSE estimation is given in Lemma 2 together with the statistics of the estimates.

Lemma 2. If BS l uses EW-MMSE estimation and the diagonal elements of the spatial correlation matrix of the channel are equal, the channel estimate between user k in cell l and BS l is

$$\hat{\mathbf{h}}_{l,k}^l = \varrho_{l,k}^l \tilde{\mathbf{y}}_{l,k}, \quad (11)$$

where

$$\varrho_{l,k}^l = \frac{\sqrt{\hat{p}_{l,k}} \beta_{l,k}^l}{\sum_{l'=1}^L \tau_p \hat{p}_{l',k} \beta_{l',k}^l + \sigma^2}, \quad (12)$$

and the channel estimate and estimation error of $\mathbf{h}_{l,k}^l$ are distributed as

$$\hat{\mathbf{h}}_{l,k}^l \sim \mathcal{CN}(\mathbf{0}, (\varrho_{l,k}^l)^2 \tau_p \Psi_{l,k}), \quad (13)$$

$$\mathbf{e}_{l,k}^l \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{l,k}^l - (\varrho_{l,k}^l)^2 \tau_p \Psi_{l,k}) \quad (14)$$

and are not independent.

Proof. The statistics of the estimate and estimation error follow from straightforward computation of the correlation matrices and the derivation is therefore omitted. \square

As compared to MMSE estimation, EW-MMSE estimation simplifies the computations, since no inverse matrix computation is involved. Moreover, each BS only needs to know the diagonal of the spatial correlation matrices, which are easier to acquire in practice than the full matrices. We can also observe the relationship between two users utilizing nonorthogonal pilots by a simple expression as shown in Corollary 1.

Corollary 1. When the diagonal elements of the spatial correlation matrix of the channel are equal, the two EW-MMSE estimates $\hat{\mathbf{h}}_{l,k}^l$ and $\hat{\mathbf{h}}_{l',k}^{l'}$ of the channels of users k in cells l and l' that are computed at BS l are related as:

$$\frac{\hat{\mathbf{h}}_{l,k}^l}{\sqrt{\hat{p}_{l,k} \beta_{l,k}^l}} = \frac{\hat{\mathbf{h}}_{l',k}^{l'}}{\sqrt{\hat{p}_{l',k} \beta_{l',k}^l}}, \quad (15)$$

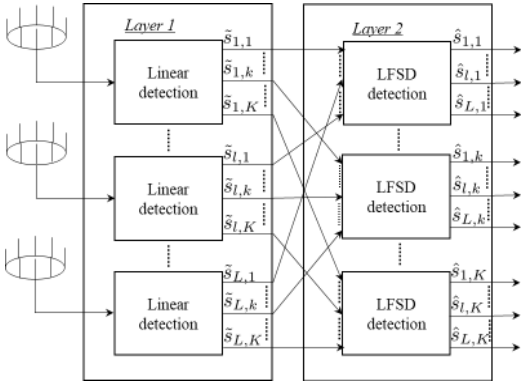


Fig. 1. Desired signals are detected by the two-layer decoding technique.

where $\hat{\mathbf{h}}_{l',k}^l = \varrho_{l',k}^l \tilde{\mathbf{y}}_{l',k}$ with

$$\varrho_{l',k}^l = \sqrt{\hat{p}_{l',k} \beta_{l',k}^l} / \left(\sum_{l''=1}^L \tau_p \hat{p}_{l'',k} \beta_{l'',k}^l + \sigma^2 \right). \quad (16)$$

Corollary 1 mathematically shows that the channel estimates of two users with the same pilot signal only differ from each other by a scaling factor. Using EW-MMSE estimation leads to severe pilot contamination that cannot be mitigated by linear processing of the data signal only, at least not with the approach in [8].

B. Uplink Data Transmission

During the data phase, it is assumed that user k in cell l' sends a zero-mean symbol $s_{l',k}$ with variance $\mathbb{E}\{|s_{l',k}|^2\} = 1$. The received signal at BS l is then

$$\mathbf{y}_l = \sum_{l'=1}^L \sum_{k=1}^K \sqrt{p_{l',k}} \mathbf{h}_{l',k}^l s_{l',k} + \mathbf{n}_l, \quad (17)$$

where $p_{l',k}$ denotes the transmit power of user k in cell l' . Based on the signals in (17), the BSs decode the symbols with the two-layers-decoding technique that is illustrated in Fig. 1. The general idea of a two-layer decoding system is that each BS decodes the desired signals from its coverage area in the first layer. A central station is then collecting the decoded signals of all users that used the same pilot and jointly processes these signals in the second layer to suppress inter-cell interference using LSFSD vectors. In detail, an estimate of the symbol from user k in cell l is obtained by local linear decoding in the first layer as

$$\tilde{s}_{l,k} = \mathbf{v}_{l,k}^H \mathbf{y}_l = \sum_{l'=1}^L \sum_{k'=1}^K \sqrt{p_{l',k'}} \mathbf{v}_{l,k}^H \mathbf{h}_{l',k'}^l s_{l',k'} + \mathbf{v}_{l,k}^H \mathbf{n}_l, \quad (18)$$

where $\mathbf{v}_{l,k}$ is the linear decoding vector. The symbol estimate $\tilde{s}_{l,k}$ generally contains interference and, in Massive MIMO, the pilot contamination from all the users with the same pilot sequence is particularly large. To mitigate the pilot contamination, all the symbol estimates of the contaminating users are collected in a vector

$$\tilde{\mathbf{s}}_k \triangleq [\tilde{s}_{1,k}, \tilde{s}_{2,k}, \dots, \tilde{s}_{L,k}]^T \in \mathbb{C}^L. \quad (19)$$

After the local decoding, a second layer of centralized decoding is performed on this vector using the LSFSD vector $\mathbf{a}_{l,k} \triangleq [a_{l,k}^1, a_{l,k}^2, \dots, a_{l,k}^L]^T \in \mathbb{C}^L$, where $a_{l,k}^{l'}$ is the LSFSD weight. The final estimate of the data symbol from user k in cell l is then given by

$$\hat{s}_{l,k} = \mathbf{a}_{l,k}^H \tilde{\mathbf{s}}_k = \sum_{l'=1}^L (a_{l,k}^{l'})^* \tilde{s}_{l',k}. \quad (20)$$

In the next section, we use the decoded signals $\hat{s}_{l,k}$ together with the asymptotic channel properties [17, Section 2.5] to derive a closed-form expression of an uplink SE.

III. Uplink Performance Analysis

This section first derives a general SE expression for each user k in each cell l and a closed-form expression when using MRC. These expressions are then used to obtain the LSFSD vectors that maximize the SE. The use-and-then-forget capacity bounding technique [6, Chapter 2.3.4], [8, Section 4.3] allows us to compute a lower bound on the uplink ergodic capacity (i.e., an achievable SE). We first rewrite (20) as

$$\begin{aligned} \hat{s}_{l,k} &= \sum_{l'=1}^L (a_{l,k}^{l'})^* \mathbb{E}\{\mathbf{v}_{l',k}^H \mathbf{h}_{l',k}^{l'}\} \sqrt{p_{l',k}} s_{l',k} \\ &+ \sum_{l'=1}^L (a_{l,k}^{l'})^* \sum_{\substack{l''=1 \\ l'' \neq l}}^L \mathbb{E}\{\mathbf{v}_{l',k}^H \mathbf{h}_{l'',k}^{l''}\} \sqrt{p_{l'',k}} s_{l'',k} \\ &+ \sum_{l'=1}^L (a_{l,k}^{l'})^* \sum_{l''=1}^L \left(\mathbf{v}_{l',k}^H \mathbf{h}_{l'',k}^{l''} - \mathbb{E}\{\mathbf{v}_{l',k}^H \mathbf{h}_{l'',k}^{l''}\} \right) \sqrt{p_{l'',k}} s_{l'',k} \\ &+ \sum_{l'=1}^L (a_{l,k}^{l'})^* \sum_{l''=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K \sqrt{p_{l'',k'}} \mathbf{v}_{l',k}^H \mathbf{h}_{l'',k'}^{l''} s_{l'',k'} \\ &+ \sum_{l'=1}^L (a_{l,k}^{l'})^* \mathbf{v}_{l',k}^H \mathbf{n}_{l'}, \end{aligned} \quad (21)$$

then by considering the first part of (21) as the desired signal from user k in cell l while the remaining is effective Gaussian noise, a lower bound on the uplink ergodic capacity is shown in Lemma 3.

Lemma 3. A lower bound on the uplink ergodic capacity is

$$R_{l,k} = \max_{\{a_{l,k}^{l'}\}} \left(1 - \frac{\tau_p}{\tau_c} \right) \log_2 (1 + \text{SINR}_{l,k}), \quad (22)$$

where the effective SINR, denoted by $\text{SINR}_{l,k}$, is

$$\text{SINR}_{l,k} = \mathbb{E}\{|\text{DS}_{l,k}|^2\} / D_{l,k}, \quad (23)$$

where $D_{l,k}$ is given by

$$\begin{aligned} D_{l,k} &= \mathbb{E}\{|\text{PC}_{l,k}|^2\} + \mathbb{E}\{|\text{BU}_{l,k}|^2\} \\ &+ \sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K \mathbb{E}\{|\text{NI}_{l',k'}|^2\} + \mathbb{E}\{|\text{AN}_{l,k}|^2\}. \end{aligned} \quad (24)$$

Here $\text{DS}_{l,k}$, $\text{PC}_{l,k}$, $\text{BU}_{l,k}$, $\text{NI}_{l',k'}$, and $\text{AN}_{l,k}$ stand for the desired signal, the pilot contamination, the beamforming gain uncertainty, the non-coherent interference, and the additive noise, respectively, whose expectations are defined as

$$\mathbb{E}\{|\text{DS}_{l,k}|^2\} \triangleq p_{l,k} \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* \mathbb{E}\{\mathbf{v}_{l',k}^H \mathbf{h}_{l',k}^1\} \right|^2, \quad (25)$$

$$\mathbb{E}\{|\text{PC}_{l,k}|^2\} \triangleq \sum_{\substack{l''=1 \\ l'' \neq l}}^L p_{l'',k} \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* \mathbb{E}\{\mathbf{v}_{l',k}^H \mathbf{h}_{l',k}^{l''}\} \right|^2, \quad (26)$$

$$\mathbb{E}\{|\text{BU}_{l,k}|^2\} \triangleq \sum_{l''=1}^L p_{l'',k} \mathbb{E} \left\{ \left| \sum_{l'=1}^L (a_{l',k}^{l''})^* \left(\mathbf{v}_{l',k}^H \mathbf{h}_{l',k}^{l''} - \mathbb{E}\{\mathbf{v}_{l',k}^H \mathbf{h}_{l',k}^{l''}\} \right) \right|^2 \right\}, \quad (27)$$

$$\mathbb{E}\{|\text{NI}_{l',k'}|^2\} \triangleq p_{l',k'} \mathbb{E} \left\{ \left| \sum_{l''=1}^L (a_{l'',k'}^{l''})^* \mathbf{v}_{l'',k'}^H \mathbf{h}_{l'',k'}^{l''} \right|^2 \right\}, \quad (28)$$

$$\mathbb{E}\{|\text{AN}_{l,k}|^2\} \triangleq \mathbb{E} \left\{ \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* (\hat{\mathbf{h}}_{l',k}^{l'})^H \mathbf{n}_{l'} \right|^2 \right\}. \quad (29)$$

Note that the lower bound on the uplink ergodic capacity in Lemma 3 can be applied to any linear decoding method and any LSFD design.

To maximize the SE of user k in cell l is equivalent to selecting the LSFD vector that maximizes a Rayleigh quotient as shown in the proof of the following theorem. This is the first main contribution of this paper.

Theorem 1. For a given set of pilot and data power coefficients, the SE of user k in cell l is

$$R_{l,k} = \left(1 - \frac{\tau_p}{\tau_c} \right) \log_2 \left(1 + p_{l,k} \mathbf{b}_{l,k}^H \left(\sum_{i=1}^4 \mathbf{C}_{l,k}^{(i)} \right)^{-1} \mathbf{b}_{l,k} \right), \quad (30)$$

where the matrices $\mathbf{C}_{l,k}^{(1)}$, $\mathbf{C}_{l,k}^{(2)}$, $\mathbf{C}_{l,k}^{(3)}$, $\mathbf{C}_{l,k}^{(4)} \in \mathbb{C}^{L \times L}$ and the vector $\mathbf{b}_{l,k}$ are defined as

$$\mathbf{C}_{l,k}^{(1)} \triangleq \sum_{\substack{l'=1 \\ l' \neq l}}^L p_{l',k} \mathbf{b}_{l',k} \mathbf{b}_{l',k}^H, \quad (31)$$

$$\mathbf{C}_{l,k}^{(2)} \triangleq \sum_{l'=1}^L p_{l',k} \mathbb{E} \left\{ \tilde{\mathbf{b}}_{l',k} \tilde{\mathbf{b}}_{l',k}^H \right\}, \quad (32)$$

$$\mathbf{C}_{l,k}^{(3)} \triangleq \text{diag} \left(\sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K p_{l',k'} \mathbb{E} \left\{ |\mathbf{v}_{1,k}^H \mathbf{h}_{l',k'}^1|^2 \right\}, \dots, \sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K p_{l',k'} \mathbb{E} \left\{ |\mathbf{v}_{L,k}^H \mathbf{h}_{l',k'}^L|^2 \right\} \right), \quad (33)$$

$$\mathbf{C}_{l,k}^{(4)} \triangleq \text{diag} \left(\sigma^2 \mathbb{E} \left\{ \|\mathbf{v}_{1,k}\|^2 \right\}, \dots, \sigma^2 \mathbb{E} \left\{ \|\mathbf{v}_{L,k}\|^2 \right\} \right), \quad (34)$$

and the vectors $\mathbf{b}_{l',k}, \tilde{\mathbf{b}}_{l',k} \in \mathbb{C}^L, \forall l' = 1, \dots, L$, are defined as

$$\mathbf{b}_{l',k} \triangleq [\mathbb{E}\{\mathbf{v}_{1,k}^H \mathbf{h}_{l',k}^1\}, \dots, \mathbb{E}\{\mathbf{v}_{L,k}^H \mathbf{h}_{l',k}^L\}]^T, \quad (35)$$

$$\tilde{\mathbf{b}}_{l',k} \triangleq [\mathbf{v}_{1,k}^H \mathbf{h}_{l',k}^1, \dots, \mathbf{v}_{L,k}^H \mathbf{h}_{l',k}^L]^T - \mathbf{b}_{l',k}. \quad (36)$$

In order to attain this SE, the LSFD vector is formulated as

$$\mathbf{a}_{l,k} = \left(\sum_{i=1}^4 \mathbf{C}_{l,k}^{(i)} \right)^{-1} \mathbf{b}_{l,k}, \quad \forall l, k. \quad (37)$$

Proof. The proof is available in Appendix B. \square

We stress that the LSFD vector in (37) is designed to maximize the SE in (30) for every user in the network for a given data and pilot power and a given first-layer decoder. Note that Theorem 1 can be applied to practical correlated Rayleigh fading channels with either MMSE or EW-MMSE estimation and any conceivable choice of first-layer decoder. This stands in contrast to the previous work [14], [18] that only considered uncorrelated Rayleigh fading channels, which are unlikely to occur in practice, and particular linear combining methods that were selected to obtain closed-form expressions. Theorem 1 explicitly reveals the influence that mutual interference and noise have on the SE when utilizing the optimal LSFD vector given in (37): $\mathbf{C}_{l,k}^{(1)}$ determines the amount of remaining pilot contamination from the $(L-1)$ users using the same pilot sequence as user k in cell l . The beamforming gain uncertainty is represented by $\mathbf{C}_{l,k}^{(2)}$, while $\mathbf{C}_{l,k}^{(3)}$ is the noncoherent mutual interference from the remaining users and $\mathbf{C}_{l,k}^{(4)}$ represent the additive noise.

The following theorem states a closed-form expression of the SE for the case of MRC, i.e., $\mathbf{v}_{l,k} = \hat{\mathbf{h}}_{l,k}^l$. This is the second main contribution of this paper.

Theorem 2. When MRC is used, the SE in (22) of user k in cell l is given by

$$R_{l,k} = \left(1 - \frac{\tau_p}{\tau_c} \right) \log_2 (1 + \text{SINR}_{l,k}), \quad (38)$$

where the SINR value is given in (39). The values $b_{l',k}^{l''}$, $c_{l',k}^{l',k'}$, and $d_{l',k}$ are different depending on the channel estimation technique. MMSE estimation results in

$$b_{l',k}^{l''} = \sqrt{\tau_p \hat{p}_{l',k} \hat{p}_{l'',k}} \text{tr} \left(\mathbf{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l''} \mathbf{R}_{l',k}^{l''} \right), \quad (40)$$

$$c_{l',k}^{l',k'} = \hat{p}_{l',k} \text{tr} \left(\mathbf{R}_{l',k}^{l''} \mathbf{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l''} \mathbf{R}_{l',k}^{l',k'} \right), \quad (41)$$

$$d_{l',k} = \sigma^2 \hat{p}_{l',k} \text{tr} \left(\mathbf{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l'} \mathbf{R}_{l',k}^{l'} \right), \quad (42)$$

whereas EW-MMSE results in

$$b_{l',k}^{l''} = \sqrt{\tau_p} \varrho_{l',k}^{l''} \varrho_{l',k}^{l''} \text{tr} \left(\mathbf{\Psi}_{l',k} \right), \quad (43)$$

$$c_{l',k}^{l',k'} = (\varrho_{l',k}^{l''})^2 \text{tr} \left(\mathbf{R}_{l',k}^{l''} \mathbf{\Psi}_{l',k} \right), \quad (44)$$

$$d_{l',k} = (\varrho_{l',k}^{l'})^2 \sigma^2 \text{tr} \left(\mathbf{\Psi}_{l',k} \right). \quad (45)$$

Proof. The proofs consist of computing the moments of complex Gaussian distributions. They are available in

$$\text{SINR}_{l,k} = \frac{p_{l,k} \left| \sum_{l'=1}^L (a'_{l,k})^* b'_{l,k} \right|^2}{\sum_{\substack{l'=1 \\ l' \neq l}}^L p_{l',k} \left| \sum_{l''=1}^L (a'_{l',k})^* b'_{l',k} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K \sum_{l''=1}^L p_{l',k'} |a'_{l',k}|^2 c'_{l',k'} + \sum_{l'=1}^L |a'_{l,k}|^2 d_{l',k}}. \quad (39)$$

Appendix C and Appendix D for MMSE and EW-MMSE estimation, respectively. \square

Theorem 2 describes the exact impact of the spatial correlation of the channel on the system performance through the coefficients $b'_{l',k}$, $c'_{l',k'}$, and $d_{l',k}$. It is seen that the numerator of (39) grows as the square of the number of antennas, M^2 , since the trace in (40) is the sum of M terms. This gain comes from the coherent combination of the signals from the M antennas. It can also be seen from Theorem 2 that the pilot contamination in (20) combines coherently, i.e., its variance—the first term in the denominator that contains $b'_{l',k}$ —grows as M^2 . The other terms in the denominator represent the impact of non-coherent interference and Gaussian noise, respectively. These two terms only grow as M . Since the interference terms contain products of correlation matrices of different users, the interference is smaller between users that have very different spatial correlation characteristics [17].

The following corollary gives the optimal LSF vector $\mathbf{a}_{l,k}$ that maximizes the SE of every user in the network for a given set of pilot and data powers, which is expected to work well when each BS is equipped with a practical number of antennas.

Corollary 2. For a given set of data and pilot powers, by using MRC and LSF, the SE in Theorem 2 is given in the closed form as

$$R_{l,k} = \left(1 - \frac{\tau_p}{\tau_c}\right) \log_2 \left(1 + p_{l,k} \mathbf{b}_{l,k}^H \mathbf{C}_{l,k}^{-1} \mathbf{b}_{l,k}\right) \quad (46)$$

where $\mathbf{C}_{l,k} \in \mathbb{C}^{L \times L}$ and $\mathbf{b}_{l,k} \in \mathbb{C}^L$ are defined as

$$\mathbf{C}_{l,k} \triangleq \sum_{\substack{l'=1 \\ l' \neq l}}^L p_{l',k} \mathbf{b}_{l',k} \mathbf{b}_{l',k}^H + \text{diag} \left(\sum_{l'=1}^L \sum_{k'=1}^K p_{l',k'} c'_{l',k'} + d_{1,k}, \dots, \sum_{l'=1}^L \sum_{k'=1}^K p_{l',k'} c'_{L,k} + d_{L,k} \right), \quad (47)$$

$$\mathbf{b}_{l',k} \triangleq [b_{l',k}^1, \dots, b_{l',k}^L]^T. \quad (48)$$

The SE in (46) is obtained by using LSF vector defined as

$$\mathbf{a}_{l,k} = \mathbf{C}_{l,k}^{-1} \mathbf{b}_{l,k}. \quad (49)$$

Even though Corollary 2 is a special case of Theorem 1 when MRC is used, its contributions are two-fold: The LSF vector $\mathbf{a}_{l,k}$ is computed in the closed form which is independent of the small-scale fading, so it is easy to compute and store. Moreover, this LSF vector is the generalization of the vector given in [14] to the larger class of correlated Rayleigh fading channels.

IV. Data Power Control and LFS Design for Sum SE Optimization

In this section, how to choose the powers $\{p_{l,k}\}$ (power control) and the LSF vector to maximize the sum SE is investigated. The sum SE maximization problem for a multi-cell Massive MIMO system is first formulated based on results from previous sections. Next, an iterative algorithm based on solving a series of convex optimization problems is proposed to efficiently find a stationary point.

A. Problem Formulation

We consider sum SE maximization:

$$\begin{aligned} & \underset{\{p_{l,k} \geq 0\}, \{\mathbf{a}_{l,k}\}}{\text{maximize}} && \sum_{l=1}^L \sum_{k=1}^K R_{l,k} \\ & \text{subject to} && p_{l,k} \leq P_{\max,l,k} \quad \forall l, k. \end{aligned} \quad (50)$$

Using the rate (38) in (50), and removing the constant pre-log factor, we obtain the equivalent formulation

$$\begin{aligned} & \underset{\{p_{l,k} \geq 0\}, \{\mathbf{a}_{l,k}\}}{\text{maximize}} && \sum_{l=1}^L \sum_{k=1}^K \log_2 (1 + \text{SINR}_{l,k}) \\ & \text{subject to} && p_{l,k} \leq P_{\max,l,k} \quad \forall l, k. \end{aligned} \quad (51)$$

This can be shown to be a non-convex and NP-hard problem using the same methodology as in [19], even if the fine details will be different since that paper considers small-scale multi-user MIMO systems with perfect channel knowledge. Therefore, the global optimum is difficult to find in general. Nevertheless, solving the ergodic sum SE maximization (51) for a Massive MIMO system is more practical than maximizing the instantaneous SEs for a small-scale MIMO network and a given realization of the small-scale fading [20], [21]. In contrast, the sum SE maximization in (51) only depends on the large-scale fading coefficients, which simplifies matters and allows the solution to be used for a long period of time. Another key difference from prior work is that we jointly optimize the data powers and LSF vectors.

Instead of seeking the global optimum to (51), which has an exponential computational complexity, we will use the weighted MMSE method [22], [23] to obtain a stationary point to (51) in polynomial time. This is a standard method to break down a sum SE problem into subproblems that can be solved sequentially. We stress that the resulting subproblems and algorithms are different for every problem that the method is applied to, thus our solution is a novel contribution. To this end, we first formulate the weighted MMSE problem from (51) as shown in Theorem 3.

Theorem 3. The optimization problem

$$\begin{aligned} & \underset{\substack{\{\rho_{l,k} \geq 0\}, \{\mathbf{a}_{l,k}\}, \\ \{w_{l,k} \geq 0\}, \{u_{l,k}\}}}{\text{minimize}} & & \sum_{l=1}^L \sum_{k=1}^K w_{l,k} e_{l,k} - \ln(w_{l,k}) \\ & \text{subject to} & & \rho_{l,k} \leq P_{\max,l,k}, \forall l, k, \end{aligned} \quad (52)$$

where $e_{l,k}$ is defined as

$$\begin{aligned} e_{l,k} \triangleq & |u_{l,k}|^2 \left(\sum_{l'=1}^L \rho_{l',k} \left| \sum_{l''=1}^L (a_{l',k}^{l''})^* b_{l',k}^{l''} \right|^2 \right. \\ & \left. + \sum_{l'=1}^L \sum_{k'=1}^K \sum_{l''=1}^L \rho_{l',k'} |a_{l',k'}^{l''}|^2 c_{l'',k'}^{l',k'} + \sum_{l''=1}^L |a_{l',k}^{l''}|^2 d_{l',k} \right) \\ & - 2\sqrt{\rho_{l,k}} \Re \left(u_{l,k} \left(\sum_{l'=1}^L (a_{l',k}^{l'})^* b_{l',k}^{l'} \right) \right) + 1, \end{aligned} \quad (53)$$

is equivalent to the sum SE optimization problem (51) in the sense that (51) and (52) have the same global optimal power solution $\{\rho_{l,k}\}, \forall l, k$, and the same LSF elements $\{a_{l',k}^{l'}\}, \forall l, k, l'$.

Proof. The proof is available in Appendix E. \square

B. Iterative Algorithm

We now find a stationary point to (52) by decomposing it into a sequence of subproblems, each having a closed-form solution. By changing variable as $\rho_{l,k} = \sqrt{\rho_{l,k}}$, the optimization problem (52) is equivalent to

$$\begin{aligned} & \underset{\substack{\{\rho_{l,k} \geq 0\}, \{\mathbf{a}_{l,k}\}, \\ \{w_{l,k} \geq 0\}, \{u_{l,k}\}}}{\text{minimize}} & & \sum_{l=1}^L \sum_{k=1}^K w_{l,k} e_{l,k} - \ln(w_{l,k}) \\ & \text{subject to} & & \rho_{l,k}^2 \leq P_{\max,l,k}, \forall l, k, \end{aligned} \quad (54)$$

where $e_{l,k}$ is

$$\begin{aligned} e_{l,k} = & |u_{l,k}|^2 \left(\sum_{l'=1}^L \rho_{l',k}^2 \left| \sum_{l''=1}^L (a_{l',k}^{l''})^* b_{l',k}^{l''} \right|^2 \right. \\ & \left. + \sum_{l'=1}^L \sum_{k'=1}^K \sum_{l''=1}^L \rho_{l',k'}^2 |a_{l',k'}^{l''}|^2 c_{l'',k'}^{l',k'} + \sum_{l''=1}^L |a_{l',k}^{l''}|^2 d_{l',k} \right) \\ & - 2\rho_{l,k} \Re \left(u_{l,k} \left(\sum_{l'=1}^L (a_{l',k}^{l'})^* b_{l',k}^{l'} \right) \right) + 1. \end{aligned} \quad (55)$$

As a third main contribution of this paper, the following theorem provides an algorithm that relies on alternating optimization to find a stationary point to (54).

Theorem 4. A stationary point to (54) is obtained by iteratively updating $\{\mathbf{a}_{l,k}, u_{l,k}, w_{l,k}, \rho_{l,k}\}$. Let $\mathbf{a}_{l,k}^{n-1}, u_{l,k}^{n-1}, w_{l,k}^{n-1}, \rho_{l,k}^{n-1}$ the values after iteration $n-1$. At iteration n , the optimization parameters are updated in the following way:

- $u_{l,k}$ is updated as

$$u_{l,k}^{(n)} = \frac{\rho_{l,k}^{(n-1)} \sum_{l'=1}^L a_{l',k}^{l',(n-1)} (b_{l',k}^{l'})^*}{\tilde{u}_{l,k}^{(n-1)}}, \quad (56)$$

where the value $\tilde{u}_{l,k}^{(n-1)}$ is defined in (57).

- $w_{l,k}$ is updated as

$$w_{l,k}^{(n)} = \left(e_{l,k}^{(n)} \right)^{-1}, \quad (58)$$

where $e_{l,k}^{(n)}$ is defined as

$$\begin{aligned} e_{l,k}^{(n)} = & |u_{l,k}^{(n)}|^2 \tilde{u}_{l,k}^{(n-1)} \\ & - 2\rho_{l,k}^{(n-1)} \Re \left(u_{l,k}^{(n)} \left(\sum_{l'=1}^L (a_{l',k}^{l',(n-1)})^* b_{l',k}^{l'} \right) \right) + 1. \end{aligned} \quad (59)$$

- $\mathbf{a}_{l,k}$ is updated as

$$\mathbf{a}_{l,k}^{(n)} = \frac{\tilde{u}_{l,k}^{*(n)}}{\sum_{l'=1}^L (a_{l',k}^{l',(n-1)})^* b_{l',k}^{l'}} \left(\tilde{\mathbf{C}}_{l,k}^{(n-1)} \right)^{-1} \mathbf{b}_{l,k}, \quad (60)$$

where $\tilde{\mathbf{C}}_{l,k}^{(n-1)}$ is computed as in (61).

- $\rho_{l,k}$ is updated as in (62).

If we denote the stationary point to (54) that is attained by the above iterative algorithm as $n \rightarrow \infty$ by $u_{l,k}^{\text{opt}}, w_{l,k}^{\text{opt}}, \mathbf{a}_{l,k}^{\text{opt}}$, and $\rho_{l,k}^{\text{opt}}$, for all l, k , then the solution $\{\mathbf{a}_{l,k}^{\text{opt}}\}, \{(\rho_{l,k}^{\text{opt}})^2\}$, is also a stationary point to the problem (51).

Proof. The proof is available in Appendix F. \square

The iterative algorithm is summarized in Algorithm 1. With initial data power values in the feasible set, the related LSF vectors are computed by using Corollary 2.³ After that, the iterative algorithm in Theorem 4 is used to obtain a stationary point to the sum SE optimization problem (50). Algorithm 1 can be terminated when the variation between two consecutive iterations are small. In particular, for a given $\epsilon \geq 0$, the stopping criterion can, for instance, be defined as

$$\left| \sum_{l=1}^L \sum_{k=1}^K R_{l,k}^{(n)} - \sum_{l=1}^L \sum_{k=1}^K R_{l,k}^{(n-1)} \right| \leq \epsilon. \quad (63)$$

Because all the update states in Algorithm 1 are in closed form, for an initial point in the feasible set we can compute the exact number of arithmetic operations needed to obtain a given ϵ -accuracy. For simplicity, let us only count complex multiplications, divisions, and logarithms, which are the main operations. Then, the number of arithmetic operations that Algorithm 1 requires is

$$\begin{aligned} N \left(11L^3K^2 + 6L^3K + \frac{L^4K + 53L^2K}{3} \right. \\ \left. + 3L^2K^2 + 16LK + 2 \right), \end{aligned} \quad (64)$$

where N is the number of iterations to reach the stationary point. We obtain (64) by assuming that a Cholesky decomposition is used to compute matrix inversion in (60).

³We observe faster convergence with a hierarchical initialization of $\rho_{l,k}^{(0)}, \forall l, k$, than with an all-equal initialization. In simulations, we initialize $\rho_{l,k}^{(0)}, \forall l, k$, as uniformly distributed over the range $[0, \sqrt{P_{\max,l,k}}]$.

$$\tilde{u}_{l,k}^{(n-1)} = \sum_{l'=1}^L (\rho_{l',k}^{(n-1)})^2 \left| \sum_{l''=1}^L (a_{l,k}^{l'',(n-1)})^* b_{l',k}^{l''} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K \sum_{l''=1}^L (\rho_{l',k'}^{(n-1)})^2 |a_{l,k}^{l'',(n-1)}|^2 c_{l',k}^{l'',k'} + \sum_{l'=1}^L |a_{l,k}^{l',(n-1)}|^2 d_{l',k}. \quad (57)$$

$$\tilde{\mathbf{C}}_{l,k}^{(n-1)} = \sum_{l'=1}^L (\rho_{l',k}^{(n-1)})^2 \mathbf{b}_{l',k} \mathbf{b}_{l',k}^H + \text{diag} \left(\sum_{l'=1}^L \sum_{k'=1}^K (\rho_{l',k'}^{(n-1)})^2 c_{l,k}^{l',k'} + d_{1,k}, \dots, \sum_{l'=1}^L \sum_{k'=1}^K (\rho_{l',k'}^{(n-1)})^2 c_{l,k}^{l',k'} + d_{L,k} \right). \quad (61)$$

$$\rho_{l,k}^{(n)} = \min \left(\frac{w_{l,k}^{(n)} \Re \left(u_{l,k}^{(n)} \sum_{l'=1}^L (a_{l,k}^{l',(n)})^* b_{l',k}^{l'} \right)}{\sum_{l'=1}^L w_{l',k}^{(n)} |u_{l',k}^{(n)}|^2 \left| \sum_{l''=1}^L (a_{l',k}^{l'',(n)})^* b_{l',k}^{l''} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K w_{l',k'}^{(n)} |u_{l',k'}^{(n)}|^2 \sum_{l''=1}^L |a_{l',k'}^{l'',(n)}|^2 c_{l',k}^{l'',k'}}, \sqrt{P_{\max,l,k}} \right). \quad (62)$$

Algorithm 1 Alternating optimization approach for (54)

Input: Maximum data powers $P_{\max,l,k}, \forall l, k$; Large-scale fading coefficients $\beta_{l,k}^{l'}, \forall l, k, l'$. Initial coefficients $\rho_{l,k}^{(0)}, \forall l, k$. Set up $n = 1$ and compute $\mathbf{a}_{l,k}^{(0)}, \forall l, k$, using Corollary 2.

1. Iteration n :

- 1.1. Update $u_{l,k}^{(n)}$ using (56) where $\tilde{u}_{l,k}^{(n-1)}$ is computed as in (57).
- 1.2. Update $w_{l,k}^{(n)}$ using (58) where $e_{l,k}^{(n)}$ is computed as in (59).
- 1.3. Update $\mathbf{a}_{l,k}^{(n)}$ by using (60) where $\mathbf{C}_{l,k}^{(n-1)}$ is computed as in (61) and $\mathbf{b}_{l,k}$ as in (48).
- 1.4. Update $\rho_{l,k}^{(n)}$ by using (62).

2. If Stopping criterion (63) is satisfied \rightarrow Stop. Otherwise, go to Step 3.

3. Store the currently best solution: $\rho_{l,k}^{(n)}$ and $\mathbf{a}_{l,k}^{(n)}, \forall l, k$. Set $n = n + 1$, then go to Step 1.

Output: The optimal solutions: $\rho_{l,k}^{\text{opt}} = \rho_{l,k}^{(n)}, \mathbf{a}_{l,k}^{\text{opt}} = \mathbf{a}_{l,k}^{(n)}, \forall l, k$.

C. Sum SE Optimization Without Using LFSF

For completeness, we also study a multi-cell Massive MIMO system that only uses one-layer decoding. This scenario is considered as a benchmark to investigate the improvements of our proposed joint data power control and LSF design in the previous section. Mathematically, this is a special case of the above analysis, in which the elements of the LSF vector $\mathbf{a}_{l,k}, \forall l, k$ are defined as

$$a_{l,k}^{l'} = \begin{cases} 1, & \text{for } l' = l, \\ 0, & \text{for } l' \neq l. \end{cases} \quad (65)$$

With the LSF vector fixed, the SE for each user in the network is a function only of the data power coefficients and it saturates when the number of BS antennas is increased without bound. The SE can, thus, only be improved through data power control. For this communication scenario, the

problem (54) becomes

$$\begin{aligned} & \underset{\{\rho_{l,k} \geq 0\}, \{w_{l,k} \geq 0\}, \{u_{l,k}\}}{\text{minimize}} && \sum_{l=1}^L \sum_{k=1}^K w_{l,k} e_{l,k} - \ln(w_{l,k}) \\ & \text{subject to} && \rho_{l,k}^2 \leq P_{\max,l,k}, \forall l, k, \end{aligned} \quad (66)$$

where $e_{l,k}$ is defined as

$$\begin{aligned} e_{l,k} \triangleq & |u_{l,k}|^2 \left(\sum_{l'=1}^L \rho_{l',k}^2 |b_{l',k}^l|^2 + \sum_{l'=1}^L \sum_{k'=1}^K \rho_{l',k'}^2 c_{l,k}^{l',k'} + d_{l,k} \right) \\ & - 2\rho_{l,k} \Re \left(u_{l,k} b_{l,k}^l \right) + 1. \end{aligned} \quad (67)$$

The alternating optimization approach in Algorithm 1 can also be applied to the problem in (66) to obtain a stationary point as shown in the following corollary.

Corollary 3. A stationary point to (66) is obtained by iteratively updating $\{u_{l,k}, w_{l,k}, \rho_{l,k}\}$. At iteration n , these optimization parameters are updated as

- $u_{l,k}$ is updated as

$$u_{l,k}^{(n)} = \frac{\rho_{l,k}^{(n-1)} (b_{l,k}^l)^*}{\tilde{u}_{l,k}^{(n-1)}}, \quad (68)$$

where $\tilde{u}_{l,k}^{(n-1)}$ is computed as

$$\begin{aligned} \tilde{u}_{l,k}^{(n-1)} = & \sum_{l'=1}^L (\rho_{l',k}^{(n-1)})^2 |b_{l',k}^l|^2 \\ & + \sum_{l'=1}^L \sum_{k'=1}^K (\rho_{l',k'}^{(n-1)})^2 c_{l,k}^{l',k'} + d_{l,k}. \end{aligned} \quad (69)$$

- $w_{l,k}$ is updated as:

$$w_{l,k}^{(n)} = \left(e_{l,k}^{(n)} \right)^{-1}, \quad (70)$$

where $e_{l,k}^{(n)}$ is computed as

$$e_{l,k}^{(n)} = |u_{l,k}^{(n)}|^2 \tilde{u}_{l,k}^{(n-1)} - 2\rho_{l,k}^{(n-1)} \Re \left(u_{l,k}^{(n)} b_{l,k}^l \right) + 1. \quad (71)$$

- $\rho_{l,k}$ is updated as

$$\rho_{l,k}^{(n)} = \min \left(\tilde{\rho}_{l,k}^{(n)}, \sqrt{P_{\max,l,k}} \right), \quad (72)$$

where

$$\tilde{\rho}_{l,k}^{(n)} \triangleq \frac{w_{l,k}^{(n)} \Re \left(u_{l,k}^{(n)} b_{l,k}^l \right)}{\sum_{l'=1}^L w_{l',k}^{(n)} |u_{l',k}^{(n)}|^2 \left| b_{l',k}^{l'} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K w_{l',k'}^{(n)} |u_{l',k'}^{(n)}|^2 c_{l',k'}^{l,k}}. \quad (73)$$

After initializing the data power coefficients to a point in the feasible set, Corollary 3 provides closed-form expressions to update each variable in the optimization (66) iteratively. This benchmark only treats the data powers as optimization variables, so it is a simplification of Algorithm 1.

V. Numerical Results

To demonstrate the effectiveness of the proposed algorithms, we consider a wrapped-around cellular network with four cells as illustrated in Fig. 2. The distance between user k in cell l' and BS l is denoted by $d_{l',k}^l$. The users in each cell are uniformly distributed over the cell area that is at least 35 m away from the BS, i.e. $d_{l',k}^l \geq 35$ m. Monte-Carlo simulations are done over 300 random sets of user locations, for almost figures, but Fig. 10 is obtained by 3000 random sets of user locations which the moments of complex Gaussian distributions are computed by 1000 random realizations of small-scale fading,

We model the system parameters and large-scale fading similar to the 3GPP LTE specifications [24]. The system uses 20 MHz of bandwidth, the noise variance is -96 dBm, and the noise figure is 5 dB. The large-scale fading coefficient $\beta_{l,k}^{l'}$ in decibel is computed as

$$\left[\beta_{l,k}^{l'} \right]_{\text{dB}} = -148.1 - 37.6 \log_{10} \left(d_{l',k}^l / 1 \text{ km} \right) + z_{l,k}^{l'}, \quad (74)$$

where the decibel value of the shadow fading, $z_{l,k}^{l'}$, has a Gaussian distribution with zero mean and standard deviation 7. The spatial correlation matrix of the channel from user k in cell l to BS l' is described by the exponential correlation model, which models a uniform linear array [25]:

$$\mathbf{R}_{l,k}^{l'} = \beta_{l,k}^{l'} \begin{bmatrix} 1 & r_{l,k}^{l',*} & \dots & (r_{l,k}^{l',*})^{M-1} \\ r_{l,k}^{l'} & 1 & \dots & (r_{l,k}^{l',*})^{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ (r_{l,k}^{l'})^{M-1} & (r_{l,k}^{l'})^{M-2} & \dots & 1 \end{bmatrix}, \quad (75)$$

where the correlation coefficient $r_{l,k}^{l'} = \zeta e^{j\theta_{l,k}^{l'}}$, the correlation magnitude ζ is in the range $[0, 1]$ and the user incidence angle to the array boresight is $\theta_{l,k}^{l'}$.

We assume that the power is fixed to 200 mW for each pilot symbol and it is also the maximum power that each user can allocate to a data symbol, i.e., $P_{\max,l,k} = 200$ mW. Extensive numerical results will be presented from the following methods with either MMSE or EW-MMSE estimation:

- (i) Single-layer decoding system with fixed data power: Each BS uses MRC for data decoding for the users

in the own cell, and all users transmit data symbols with the same power 200 mW.

- (ii) Single-layer decoding system with data power control: This benchmark is similar to (i), but the data powers are optimized using the weighted MMSE algorithm in Corollary 3.
- (iii) Two-layer decoding system with fixed data power and LSF vectors: The network deploys the two-layer decoding as shown in Fig. 1, using MRC and LSF. The data symbols have fixed power 200 mW and the LSF vectors are computed using Corollary 2.
- (iv) Two-layer decoding system with fixed data power and approximate LSF vectors: This benchmark is similar to (iii), but the LSF vectors are computed using only the diagonal elements of the channel correlation matrices. This allows us to study how inaccurate LSF vectors degrade sum SE.
- (v) Two-layer decoding system with optimized data power and LSF vectors: This benchmark is similar to (iii), but the data powers and LSF vectors are computed using the weighted MMSE algorithm as in Theorem 4.
- (vi) Two-layer decoding system with optimized data power and approximate LSF vectors: This benchmark is similar to (v), but the LSF vectors are computed by Corollary 2 based on only the diagonal coefficients of the channel correlation matrices.

A. Convergence

Fig. 3 shows the convergence of the proposed methods for sum SE optimization in Theorem 4 and Corollary 3 for both MMSE and EW-MMSE estimation. From the initial data powers, in the feasible set, updating the optimization variables gives improved sum SE in every iteration. For a system that uses MMSE estimation and LSF, the sum SE per cell is about 22.2% better at the stationary point than at the initial point. The corresponding improvement for the system that uses EW-MMSE estimation is about 24.7%. By using MMSE estimation, the two-layer decoding system gives 2.4% better sum SE than a system with single-layer decoding. The corresponding gain for EW-MMSE estimation is up to 7.5%. Besides, MMSE estimation gives an SE that is up to 12.1% higher than EW-MMSE. The proposed optimization methods need around 100 iterations to converge, but the complexity is low since every update in the algorithm consists of evaluating a closed-form expression.

The approximation in (vi) of the channel correlation matrix as diagonal breaks the convergence statement in Theorem 4, so it is not included in Fig. 3. Hereafter, when we consider (vi) for comparison, we select the highest sum SE among 500 iterations.

B. Impact of Spatial Correlation

Figs. 4 and 5 show the sum SE per cell as a function of the channel correlation magnitude ζ for a multi-cell Massive MIMO system using either MMSE or EW-MMSE estimation. First, we observe the large gains in sum SE

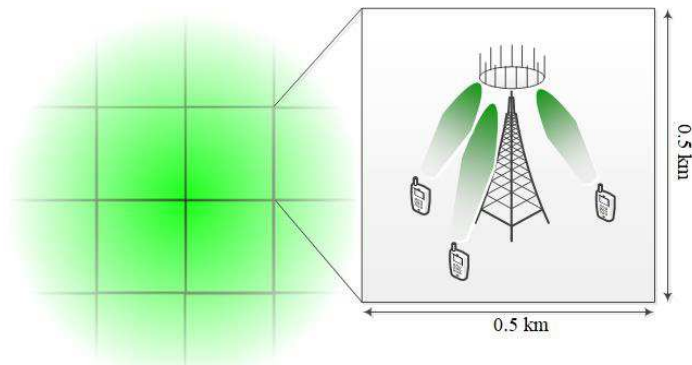


Fig. 2. A wrapped-around cellular network used for simulation.

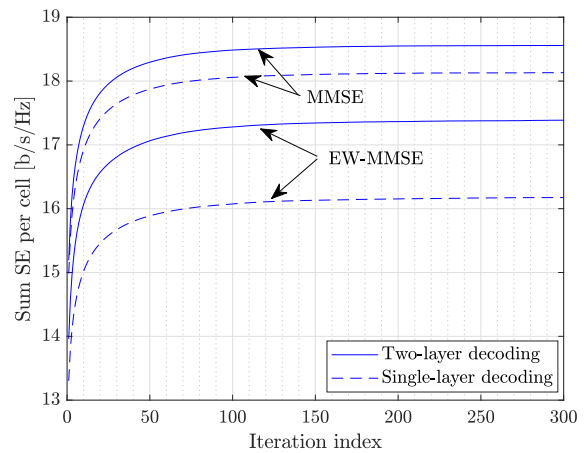


Fig. 3. Convergence of the proposed sum SE optimization with $M = 200$, $K = 5$, and $\zeta = 0.8$.

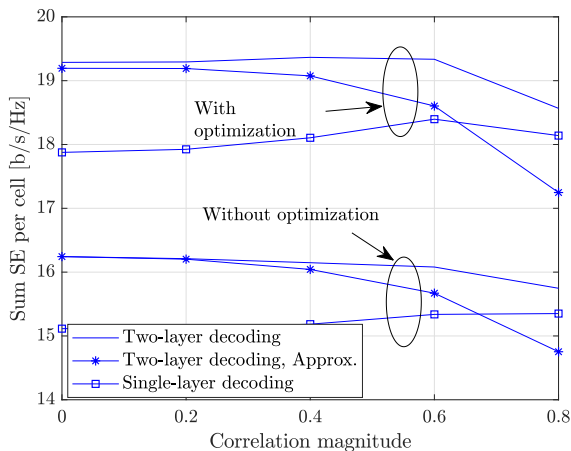


Fig. 4. Sum SE per cell [b/s/Hz] versus different correlation magnitudes. The network uses MMSE estimation, $M = 200$, and $K = 5$.

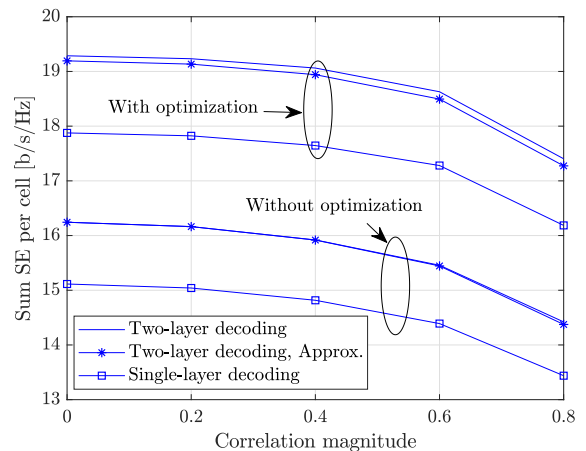


Fig. 5. Sum SE per cell [b/s/Hz] versus different correlation magnitudes. The network uses EW-MMSE estimation, $M = 200$, and $K = 5$.

attained by using LSFV detection. With MMSE estimation, the sum SE increases with up to 7.5% in the case of equally fixed data powers, while that gain is about 7.9% for jointly optimizing data powers and LSFV vectors. The same maximum gains are observed when using EW-MMSE estimation, since these gains occur when $\zeta = 0$ (in which case MMSE and EW-MMSE coincide). The performance of EW-MMSE estimation is worse than that of MMSE when the correlation magnitude is increased, because EW-MMSE does not use the knowledge of the spatial correlation to improve the estimation quality. For example, MMSE estimation obtains 6.68% and 9.91% higher SE than EW-MMSE with and without data power control, respectively. The advantage of EW-MMSE is the reduced computational complexity.

Interestingly, Figs. 4 and 5 indicate that it is sufficient to use only the large-scale fading coefficients when constructing the LSFV vectors in many scenarios. Specifically, in the system with EW-MMSE estimation, the approximate LSFV vectors yield almost the same sum SE as the optimal ones. Meanwhile, in the case of MMSE estimation, the loss from the approximation of

LSFV vectors, which are only based on the diagonal values of channel correlation matrices, grows up to 6.7% when having a correlation magnitude of 0.8. In comparison to MMSE, increasing the spatial correlation does not improve the performance of the approximate LSFV vectors when using EW-MMSE estimation, since it does not utilize the spatial correlation in the estimation phase. Consequently, MMSE and EW-MMSE perform almost equally with the maximum difference 0.71%.

Moreover, the performance is greatly improved when the data powers are optimized. The gain varies from 17.9% to 20.7%. The gap becomes larger as the channel correlation magnitude increases. This shows the importance of doing joint data power control and LSFV optimization in Massive MIMO systems with spatially correlated channels.

C. Impact of Number of Antennas and Users

Figs. 6 and 7 show the sum SE per cell as a function of the number of BS antennas with MMSE and EW-MMSE estimation, respectively. Two-layer decoding gives improvements in all the cases. In case of MMSE estimation, by increasing the number of BS antennas from 100 to 300,

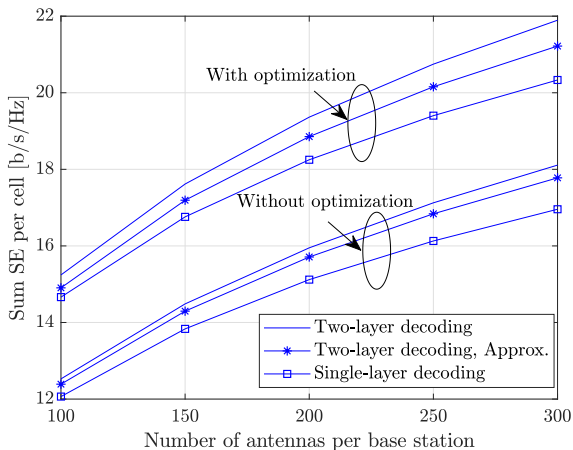


Fig. 6. Sum SE per cell [b/s/Hz] versus different number of BS antennas. The network uses MMSE estimation, $K = 5$, and $\varsigma = 0.5$.

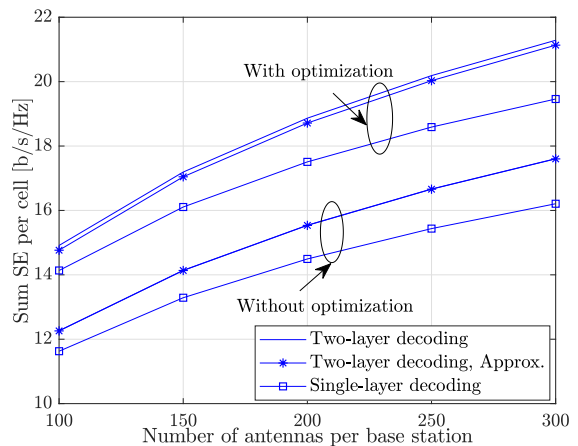


Fig. 7. Sum SE per cell [b/s/Hz] versus different number of BS antennas. The network uses EW-MMSE estimation, $K = 5$, and $\varsigma = 0.5$.

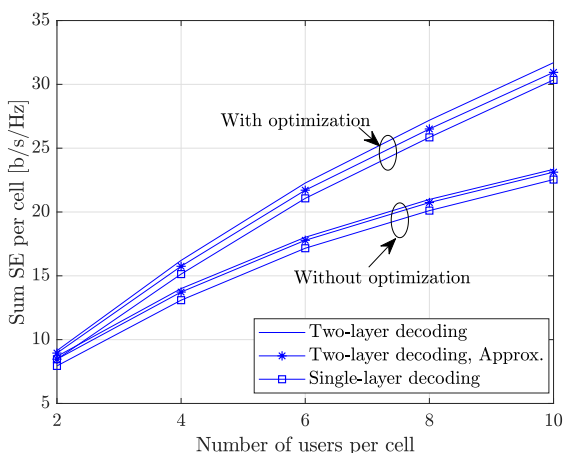


Fig. 8. Sum SE per cell [b/s/Hz] versus different number of users per cell. The network uses MMSE estimation, $M = 200$, and $\varsigma = 0.5$.

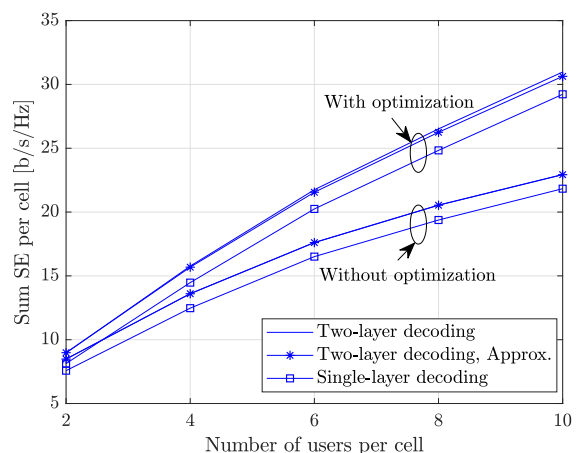


Fig. 9. Sum SE per cell [b/s/Hz] versus different number of users per cell. The network uses EW-MMSE estimation, $M = 200$, and $\varsigma = 0.5$.

the gain of using LSF D increases from 4.0% to 7.7% with optimized data power, and from 3.8% to 6.8% with equal data power. In case of EW-MMSE estimation and fixed transmitted power level, LSF D increase the sum SE by 5.5% to 8.6% compared to using only MRC. Besides, by optimizing the data powers, the gain from using LSF D is between 5.5% and 9.4%. Among all considered scenarios, MMSE estimation provides up to 4.6% higher sum SE than EW-MMSE.

Figs. 8 and 9 show the sum SE per cell as a function of the number of users per cell when using MMSE and EW-MMSE estimation, respectively. The figures demonstrate how the gain from power control increases with the number of users. The gain grows from 5.2% for two users to 35.8% for for ten users. The approximated version of LSF D detection works properly in all tested scenarios, in the sense that the maximum loss in SE is only up to 2.9%. In these figures, MMSE provides up to 5% higher SE than EW-MMSE.

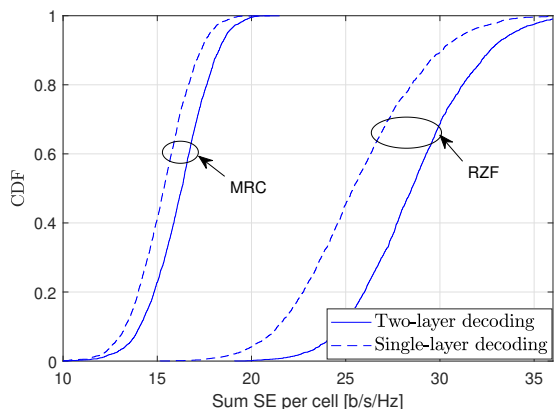


Fig. 10. CDF of sum SE per cell [b/s/Hz] for MRC and RZF. The network uses $M = 200$, $K = 5$, $\varsigma = 0.5$, and MMSE estimation.

D. Performance of Regularized Zero-Forcing

Fig. 10 compares the cumulative distribution function (CDF) of the sum SE per cell when using either MRC or RZF in the first layer. MMSE estimation is used for channel estimation. An equal pilot and data power of 200

mW is allocated to each transmitted symbol. We first observe that RZF achieves much higher SE than MRC. The performance gain is 67.2% and 76.6% with single-layer and two-layer decoding, respectively. Because RZF cancels non-coherent interference effectively at the first layer, the second layer can improve the sum SE by 11.80%. Meanwhile, the improvement is only 5.84% if MRC is used.

VI. Conclusion

This paper has investigated the performance of the LSF design in mitigating mutual interference for multi-cell Massive MIMO systems with correlated Rayleigh fading. This decoding design is deployed as a second decoding layer to mitigate the interference that remains after classical linear decoding. Numerical results demonstrate the effectiveness of the LSF design in reducing pilot contamination with the improvement of sum SE for each cell up to about 10% in the tested scenarios. We have also investigated joint data power control and LSF design, which efficiently improves the sum SE of the network. Even though the sum SE optimization is a non-convex and NP-hard problem, we proposed an iterative approach to obtain a stationary point with low computational complexity. Numerical results showed improvements of sum SE for each cell up to more than 20% with using the limited number of BS antennas.

Appendix A

Useful Lemma and Definition

Lemma 4 (Lemma 2 in [26]). Let a random vector be distributed as $\mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$ and consider an arbitrary, deterministic matrix \mathbf{M} . It holds that

$$\mathbb{E}\{|\mathbf{u}^H \mathbf{M} \mathbf{u}|^2\} = |\text{tr}(\mathbf{\Lambda} \mathbf{M})|^2 + \text{tr}(\mathbf{\Lambda} \mathbf{M} \mathbf{\Lambda} \mathbf{M}^H). \quad (76)$$

Definition 1 (Stationary point, [27]).⁴ Consider the optimization problem

$$\underset{\mathbf{x} \in \mathcal{X}}{\text{minimize}} \quad g(\mathbf{x}), \quad (77)$$

where the feasible set \mathcal{X} is convex and $g(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ is differentiable. A point $\mathbf{y} \in \mathcal{X}$ is a stationary point to the optimization problem (77) if the following property is true for all $\mathbf{x} \in \mathcal{X}$:

$$(\mathbf{x} - \mathbf{y})^T \nabla g(\mathbf{x})|_{\mathbf{y}} \geq 0. \quad (78)$$

Note that a stationary point \mathbf{y} of $g(\mathbf{x})$ can be obtained by solving the equation $\nabla g(\mathbf{x}) = \mathbf{0}$.

Appendix B

Proof of Theorem 1

The numerator of (23) is reformulated into

$$\mathbb{E}\{|\text{DS}_{l,k}|^2\} = p_{l,k} |\mathbf{a}_{l,k}^H \mathbf{b}_{l,k}|^2. \quad (79)$$

⁴Definition 1 guarantees the existence of at least one stationary point for any non-convex problem as long as the feasible set is convex. The stationary point is even the global optimum if (77) is a convex problem.

Meanwhile, the pilot contamination term in the denominator of (23) is rewritten as

$$\begin{aligned} \mathbb{E}\{|\text{PC}_{l,k}|^2\} &= \sum_{\substack{l''=1 \\ l'' \neq l}}^L p_{l'',k} \left| \sum_{l'=1}^L (a_{l',k}^{l''})^* \mathbb{E}\{\mathbf{v}_{l',k}^H \mathbf{h}_{l'',k}^{l''}\} \right|^2 \\ &= \sum_{\substack{l''=1 \\ l'' \neq l}}^L p_{l'',k} \mathbf{a}_{l',k}^H \mathbf{b}_{l'',k} \mathbf{b}_{l'',k}^H \mathbf{a}_{l',k} \\ &= \mathbf{a}_{l,k}^H \mathbf{C}_{l,k}^{(1)} \mathbf{a}_{l,k}. \end{aligned} \quad (80)$$

The beamforming gain uncertainty term in the denominator of (23) is rewritten as

$$\mathbb{E}\{|\text{BU}_{l,k}|^2\} = \mathbf{a}_{l,k}^H \mathbf{C}_{l,k}^{(2)} \mathbf{a}_{l,k}. \quad (81)$$

Similarly, the non-coherent interference term in the denominator is computed as

$$\begin{aligned} &\sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K \mathbb{E}\{|\text{NI}_{l',k'}|^2\} \\ &= \sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K \left| \sum_{l''=1}^L (a_{l',k'}^{l''})^* \sqrt{p_{l'',k'}} \mathbf{v}_{l'',k'}^H \mathbf{h}_{l',k'}^{l''} \right|^2 \\ &= \mathbf{a}_{l',k'}^H \mathbf{C}_{l',k'}^{(3)} \mathbf{a}_{l',k'} \end{aligned} \quad (82)$$

and the additive noise term is computed as

$$\begin{aligned} \mathbb{E}\{|\text{AN}_{l,k}|^2\} &= \mathbb{E}\left\{ \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* \mathbf{v}_{l',k}^H \mathbf{n}_{l'} \right|^2 \right\} \\ &= \mathbf{a}_{l,k}^H \mathbf{C}_{l,k}^{(4)} \mathbf{a}_{l,k}. \end{aligned} \quad (83)$$

The lower-bound on the uplink capacity given in Lemma 3 is written as

$$R_{l,k} = \left(1 - \frac{\tau_p}{\tau_c} \right) \log_2 \left(1 + \frac{p_{l,k} |\mathbf{a}_{l,k}^H \mathbf{b}_{l,k}|^2}{\mathbf{a}_{l,k}^H \left(\sum_{i=1}^4 \mathbf{C}_{l,k}^{(i)} \right) \mathbf{a}_{l,k}} \right). \quad (84)$$

Since the SINR expression in (84) is a generalized Rayleigh quotient with respect to $\mathbf{a}_{l,k}$, we can apply [17, Lemma B.10] to obtain the maximizing vector $\mathbf{a}_{l,k}$ as in (37). Hence, using (37) in (84), maximizes the SE for both MMSE and EW-MMSE estimation. The maximum SE is given by (30) in the theorem.

Appendix C

Proof of Theorem 2 in case of MMSE

Here the expectations in (23) are computed. The numerator of (23), $\mathbb{E}\{|\text{DS}_{l,k}|^2\}$, becomes

$$\begin{aligned}
& p_{l,k} \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* \mathbb{E} \left\{ \sqrt{\hat{p}_{l',k}} (\mathbf{R}_{l',k}^{l'} \boldsymbol{\Psi}_{l',k}^{-1} \mathbf{y}_{l',k})^H \mathbf{h}_{l',k}^{l'} \right\} \right|^2 \\
&= p_{l,k} \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* \sqrt{\hat{p}_{l',k}} \text{tr} \left(\boldsymbol{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l'} \mathbb{E} \{ \mathbf{h}_{l',k}^{l'} \mathbf{y}_{l',k}^H \} \right) \right|^2 \\
&= p_{l,k} \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* \tau_p \sqrt{\hat{p}_{l',k} \hat{p}_{l',k}} \text{tr} \left(\boldsymbol{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l'} \mathbf{R}_{l',k}^{l'} \right) \right|^2 \\
&= \tau_p p_{l,k} \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* b_{l',k}^{l'} \right|^2, \tag{85}
\end{aligned}$$

The variance of the pilot contamination in the denominator of (23) is computed as

$$\begin{aligned}
& \sum_{\substack{l''=1 \\ l'' \neq l}}^L p_{l'',k} \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* \tau_p \sqrt{\hat{p}_{l',k} \hat{p}_{l'',k}} \text{tr} \left(\boldsymbol{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l'} \mathbf{R}_{l'',k}^{l''} \right) \right|^2 \\
&= \tau_p \sum_{\substack{l''=1 \\ l'' \neq l}}^L p_{l'',k} \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* b_{l',k}^{l''} \right|^2. \tag{86}
\end{aligned}$$

The variance of the beamforming gain uncertainty, $\mathbb{E}\{|\text{BU}_{l,k}|^2\}$, is evaluated as

$$\begin{aligned}
& \sum_{l'=1}^L p_{l',k} \sum_{l''=1}^L |a_{l',k}^{l''}|^2 \left(\underbrace{\mathbb{E} \left\{ |(\hat{\mathbf{h}}_{l',k}^{l''})^H \mathbf{h}_{l',k}^{l''}|^2 \right\}}_{\mathcal{I}_1} \right. \\
& \quad \left. - \underbrace{\left| \mathbb{E} \{ (\hat{\mathbf{h}}_{l',k}^{l''})^H \mathbf{h}_{l',k}^{l''} \} \right|^2}_{\mathcal{I}_2} \right), \tag{87}
\end{aligned}$$

where the expectation \mathcal{I}_1 is computed by applying the property in Lemma 4 as

$$\begin{aligned}
\mathcal{I}_1 &= \mathbb{E} \left\{ |(\hat{\mathbf{h}}_{l',k}^{l''})^H \mathbf{h}_{l',k}^{l''}|^2 \right\} + \mathbb{E} \left\{ |(\hat{\mathbf{h}}_{l',k}^{l''})^H \mathbf{e}_{l',k}^{l''}|^2 \right\} \\
&= \hat{p}_{l',k} \hat{p}_{l'',k} \mathbb{E} \left\{ |\mathbf{y}_{l',k}^H \boldsymbol{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l''} \mathbf{R}_{l'',k}^{l''} \boldsymbol{\Psi}_{l'',k}^{-1} \mathbf{y}_{l'',k}|^2 \right. \\
& \quad \left. + \text{tr} \left(\mathbb{E} \{ \mathbf{e}_{l',k}^{l''} (\mathbf{e}_{l',k}^{l''})^H \hat{\mathbf{h}}_{l',k}^{l''} (\hat{\mathbf{h}}_{l',k}^{l''})^H \} \right) \right\} \\
&= \tau_p^2 \hat{p}_{l',k} \hat{p}_{l'',k} |\text{tr}(\boldsymbol{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l''} \mathbf{R}_{l'',k}^{l''})|^2 \\
& \quad + \tau_p \hat{p}_{l'',k} \text{tr}(\mathbf{R}_{l',k}^{l''} \mathbf{R}_{l'',k}^{l''} \boldsymbol{\Psi}_{l',k}^{-1} \mathbf{R}_{l'',k}^{l''}) \\
&= \tau_p \left((b_{l',k}^{l''})^2 + c_{l',k}^{l''} \right) \tag{88}
\end{aligned}$$

and the expectation \mathcal{I}_2 is computed as

$$\begin{aligned}
\mathcal{I}_2 &= \left| \mathbb{E} \{ (\hat{\mathbf{h}}_{l',k}^{l''})^H \hat{\mathbf{h}}_{l',k}^{l''} \} \right|^2 \\
&= \hat{p}_{l',k} \hat{p}_{l'',k} \left| \mathbb{E} \{ \mathbf{y}_{l',k}^H \boldsymbol{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l''} \mathbf{R}_{l'',k}^{l''} \boldsymbol{\Psi}_{l'',k}^{-1} \mathbf{y}_{l'',k} \} \right|^2 \\
&= \tau_p^2 \hat{p}_{l',k} \hat{p}_{l'',k} |\text{tr}(\boldsymbol{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l''} \mathbf{R}_{l'',k}^{l''})|^2 \\
&= \tau_p (b_{l',k}^{l''})^2. \tag{89}
\end{aligned}$$

Combining (87), (88), and (89), we obtain the variance of the beamforming gain uncertainty as

$$\begin{aligned}
& \sum_{l''=1}^L \sum_{l'=1}^L p_{l'',k} \tau_p \hat{p}_{l',k} |a_{l',k}^{l''}|^2 \text{tr}(\mathbf{R}_{l'',k}^{l''} \mathbf{R}_{l',k}^{l'} \boldsymbol{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l'}) \\
&= \sum_{l''=1}^L \sum_{l'=1}^L \tau_p p_{l'',k} |a_{l',k}^{l''}|^2 c_{l',k}^{l''}. \tag{90}
\end{aligned}$$

The variance of the non-coherent interference term, $\sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K \mathbb{E}\{|\text{NI}_{l',k'}|^2\}$, is computed based on the independent channel properties

$$\begin{aligned}
& \sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K p_{l',k'} \sum_{l''=1}^L |a_{l',k}^{l''}|^2 \mathbb{E} \left\{ \left| (\hat{\mathbf{h}}_{l',k}^{l''})^H \mathbf{h}_{l',k'}^{l''} \right|^2 \right\} \\
&= \sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K \sum_{l''=1}^L p_{l',k'} \tau_p \hat{p}_{l'',k} |a_{l',k}^{l''}|^2 \text{tr} \left(\mathbf{R}_{l',k'}^{l''} \mathbf{R}_{l'',k}^{l''} \boldsymbol{\Psi}_{l'',k}^{-1} \mathbf{R}_{l'',k}^{l''} \right) \\
&= \sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K \sum_{l''=1}^L \tau_p p_{l',k'} |a_{l',k}^{l''}|^2 c_{l',k'}^{l'',k}. \tag{91}
\end{aligned}$$

The last expectation in the denominator is computed as

$$\begin{aligned}
\mathbb{E}\{|\text{AN}_{l,k}|^2\} &= \sum_{l'=1}^L |a_{l',k}^{l'}|^2 \mathbb{E}\{ |(\hat{\mathbf{h}}_{l',k}^{l'})^H \mathbf{n}_{l'}|^2 \} \\
&= \sum_{l'=1}^L |a_{l',k}^{l'}|^2 \sigma^2 \tau_p \hat{p}_{l',k} \text{tr} \left(\mathbf{R}_{l',k}^{l'} \boldsymbol{\Psi}_{l',k}^{-1} \mathbf{R}_{l',k}^{l'} \right) \\
&= \sum_{l'=1}^L \tau_p |a_{l',k}^{l'}|^2 d_{l',k}. \tag{92}
\end{aligned}$$

Using (85), (86), (90), (91), and (92) in (22), we obtain the closed-form expression for the SE as shown in Theorem 2.

Appendix D

Proof of Theorem 2 in Case of EW-MMSE

The main steps to prove the results for the case of EW-MMSE are similar to that of MMSE, but the distributions of the estimate and estimation errors are different (and not independent). However, we can use the relationship in Corollary 1 between the channels of the users sending non-orthogonal pilot signals to perform the derivation. The main steps of the proof are summarized as follows: The numerator of (23) is computed based on the relationship between the estimates of the channels to BS l and users k in cells l' and l as stated in Corollary 1:

$$\begin{aligned}
& p_{l,k} \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* \frac{\sqrt{\hat{p}_{l',k}} \beta_{l',k}^{l'}}{\sqrt{\hat{p}_{l',k} \beta_{l',k}^{l'}}} \mathbb{E} \{ \|\hat{\mathbf{h}}_{l',k}^{l'}\|^2 \} \right|^2 \\
&= p_{l,k} \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* \sqrt{\tau_p} \varrho_{l',k}^{l'} \text{tr}(\boldsymbol{\Psi}_{l',k}) \right|^2 \\
&= p_{l,k} \left| \sum_{l'=1}^L (a_{l',k}^{l'})^* b_{l',k}^{l'} \right|^2. \tag{93}
\end{aligned}$$

Similarly, we use the relationship between the channel estimates of users k in cells l' and l'' to compute the variance of the pilot contamination term in the denominator of (23) as

$$\begin{aligned} & \sum_{\substack{l''=1 \\ l'' \neq l}}^L p_{l'',k} \left| \sum_{l'=1}^L (a'_{l',k})^* \frac{\sqrt{\hat{p}_{l'',k} \beta'_{l'',k}}}{\sqrt{\hat{p}_{l',k} \beta'_{l',k}}} \mathbb{E}\{\|\hat{\mathbf{h}}_{l',k}\|^2\} \right|^2 \\ &= \sum_{\substack{l''=1 \\ l'' \neq l}}^L p_{l'',k} \left| \sum_{l'=1}^L (a'_{l',k})^* \sqrt{\tau_P} \varrho'_{l',k} \varrho'_{l'',k} \text{tr}(\Psi_{l',k}) \right|^2 \quad (94) \\ &= \sum_{\substack{l''=1 \\ l'' \neq l}}^L p_{l'',k} \left| \sum_{l'=1}^L (a'_{l',k})^* b'_{l'',k} \right|^2. \end{aligned}$$

The variance of the beam uncertainty term in the denominator of (23) is computed as

$$\begin{aligned} & \sum_{l'=1}^L p_{l',k} \sum_{l''=1}^L |a'_{l',k}|^2 \left(\mathbb{E}\left\{ |(\hat{\mathbf{h}}_{l'',k})^H \hat{\mathbf{h}}_{l'',k}|^2 \right\} \right. \\ & \quad \left. - \left(\mathbb{E}\{(\hat{\mathbf{h}}_{l'',k})^H \hat{\mathbf{h}}_{l'',k}\} \right)^2 \right) \\ & \quad + \sum_{l'=1}^L p_{l',k} \mathbb{E} \left\{ \left| \sum_{l''=1}^L (a'_{l',k})^* (\hat{\mathbf{h}}_{l'',k})^H \mathbf{e}'_{l'',k} \right|^2 \right\}. \quad (95) \end{aligned}$$

By using the relationship between the two channel estimates $\hat{\mathbf{h}}_{l'',k}$ and $\hat{\mathbf{h}}_{l',k}$ in Corollary 1, (95) is equal to

$$\begin{aligned} & \sum_{l'=1}^L p_{l',k} \sum_{l''=1}^L |a'_{l',k}|^2 \frac{\hat{p}_{l',k} (\beta'_{l',k})^2}{\hat{p}_{l'',k} (\beta'_{l'',k})^2} \left(\mathbb{E}\{\|\hat{\mathbf{h}}_{l'',k}\|^4\} \right. \\ & \quad \left. - \left(\mathbb{E}\{\|\hat{\mathbf{h}}_{l'',k}\|^2\} \right)^2 \right) \\ & \quad + \sum_{l'=1}^L \sum_{l''=1}^L p_{l',k} |a'_{l',k}|^2 \mathbb{E} \left\{ |(\hat{\mathbf{h}}_{l'',k})^H \mathbf{e}'_{l'',k}|^2 \right\} \\ &= \sum_{l'=1}^L \sum_{l''=1}^L p_{l',k} \tau_P |a'_{l',k}|^2 (\varrho'_{l'',k})^2 (\varrho'_{l',k})^2 \text{tr}(\Psi_{l'',k}^2) \\ & \quad + \sum_{l'=1}^L \sum_{l''=1}^L p_{l',k} |a'_{l',k}|^2 (\varrho'_{l'',k})^2 \\ & \quad \times \text{tr} \left(\left(\mathbf{R}_{l'',k}^{l''} - \tau_P (\varrho'_{l'',k})^2 \Psi_{l'',k} \right) \Psi_{l'',k} \right) \\ &= \sum_{l'=1}^L \sum_{l''=1}^L p_{l',k} |a'_{l',k}|^2 (\varrho'_{l'',k})^2 \text{tr}(\mathbf{R}_{l'',k}^{l''} \Psi_{l'',k}) \\ &= \sum_{l'=1}^L \sum_{l''=1}^L p_{l',k} |a'_{l',k}|^2 c_{l'',k}^{l',k}. \quad (96) \end{aligned}$$

By performing MMSE estimation separately for every element of a channel vector, it is straightforward to prove that $\hat{\mathbf{h}}_{l'',k}$ and $\hat{\mathbf{h}}_{l',k}$ are independent since the joint density function is the product of their respective marginal densities. Consequently, the variance of the

non-coherent interference in the denominator of (23), $\sum_{l'=1}^L \sum_{k'=1, k' \neq k}^K \mathbb{E}\{|\mathbf{N}_{l',k'}|^2\}$, is computed as

$$\begin{aligned} & \sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K p_{l',k'} \mathbb{E} \left\{ \left| \sum_{l''=1}^L (a'_{l',k'})^* (\hat{\mathbf{h}}_{l'',k'})^H \mathbf{h}_{l'',k'} \right|^2 \right\} \\ &= \sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K p_{l',k'} \sum_{l''=1}^L |a'_{l',k'}|^2 \mathbb{E} \left\{ |(\hat{\mathbf{h}}_{l'',k'})^H \mathbf{h}_{l'',k'}|^2 \right\} \quad (97) \\ &= \sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K \sum_{l''=1}^L p_{l',k'} |a'_{l',k'}|^2 (\varrho'_{l'',k'})^2 \text{tr}(\mathbf{R}_{l'',k'}^{l''} \Psi_{l'',k'}) \\ &= \sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K \sum_{l''=1}^L p_{l',k'} |a'_{l',k'}|^2 c_{l'',k'}^{l',k'}. \end{aligned}$$

The last expectation in the denominator of (22) is computed by using the fact that the noise and the channel estimate are independent, leading to

$$\begin{aligned} \mathbb{E}\{|\mathbf{A}\mathbf{N}_{l,k}|^2\} &= \sum_{l'=1}^L |a'_{l',k}|^2 (\varrho'_{l',k})^2 \sigma^2 \text{tr}(\Psi_{l',k}) \\ &= \sum_{l'=1}^L |a'_{l',k}|^2 d_{l',k}. \quad (98) \end{aligned}$$

Applying (93), (94), (96), (97), and (98) in (22), we obtain the closed-form expression of the SE as shown in Theorem 2.

Appendix E Proof of Theorem 3

The whole system with the aggregate effect of channel and decoding can be viewed as a SISO channel with deterministic channel gain, whose SE is the equivalent of (38), namely:

$$\begin{aligned} y_{l,k} &= \sqrt{p_{l,k}} \left(\sum_{l'=1}^L (a'_{l',k})^* b'_{l',k} \right) s_{l,k} \\ & \quad + \sum_{\substack{l'=1 \\ l' \neq l}}^L \sqrt{p_{l',k}} \left(\sum_{l''=1}^L (a'_{l',k})^* b'_{l'',k} \right) s_{l',k} \\ & \quad + \sum_{l'=1}^L \sum_{k'=1}^K \sqrt{\sum_{l''=1}^L p_{l',k'} |a'_{l',k'}|^2 c_{l'',k'}^{l',k'}} z_{l',k'} \\ & \quad + \sqrt{\sum_{l'=1}^L |a'_{l',k}|^2 d_{l',k} n_{l,k}}, \quad (99) \end{aligned}$$

where $z_{l',k'} \sim \mathcal{CN}(0,1)$ and $n_{l,k} \sim \mathcal{CN}(0,1)$. The desired signal $s_{l,k}$ is decoded by using a beamforming coefficient

$u_{l,k} \in \mathbb{C}$ as

$$\begin{aligned} \hat{s}_{l,k} &= u_{l,k} y_{l,k} = \sqrt{p_{l,k}} u_{l,k} \left(\sum_{l'=1}^L (a'_{l,k})^* b'_{l,k} \right) s_{l,k} \\ &+ \sum_{\substack{l'=1 \\ l' \neq l}}^L \sqrt{p_{l',k}} u_{l,k} \left(\sum_{l''=1}^L (a'_{l',k})^* b'_{l',k} \right) s_{l',k} \\ &+ \sum_{l'=1}^L \sum_{k'=1}^K u_{l,k} \sqrt{\sum_{l''=1}^L p_{l',k'} |a'_{l',k'}|^2 c_{l',k'} z_{l',k'}} \\ &+ u_{l,k} \sqrt{\sum_{l'=1}^L |a'_{l',k}|^2 d_{l',k}} n_{l,k}. \quad (100) \end{aligned}$$

We now compute the mean-square error as

$$\begin{aligned} e_{l,k} &= \mathbb{E}\{|\hat{s}_{l,k} - s_{l,k}|^2\} = \mathbb{E}\{|u_{l,k} y_{l,k} - s_{l,k}|^2\} \\ &= \left| \sqrt{p_{l,k}} u_{l,k} \left(\sum_{l'=1}^L (a'_{l,k})^* b'_{l,k} \right) - 1 \right|^2 + \sum_{\substack{l'=1 \\ l' \neq l}}^L p_{l',k} |u_{l,k}|^2 \times \\ &\left| \sum_{l''=1}^L (a'_{l',k})^* b'_{l',k} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K \sum_{l''=1}^L |u_{l,k}|^2 p_{l',k'} |a'_{l',k'}|^2 c_{l',k'} \\ &+ |u_{l,k}|^2 \sum_{l'=1}^L |a'_{l',k}|^2 d_{l',k}. \quad (101) \end{aligned}$$

After some algebra, we obtain $e_{l,k}$ as shown in (53). The optimal solution of $u_{l,k}$ is computed by equating the first derivative of $e_{l,k}$ with respect to $u_{l,k}$ to zero, leading to

$$\begin{aligned} u_{l,k}^* &\left(\sum_{l'=1}^L p_{l',k} \left| \sum_{l''=1}^L (a'_{l',k})^* b'_{l',k} \right|^2 \right. \\ &+ \sum_{l'=1}^L \sum_{k'=1}^K \sum_{l''=1}^L p_{l',k'} |a'_{l',k'}|^2 c_{l',k'} \\ &\left. + \sum_{l'=1}^L |a'_{l',k}|^2 d_{l',k} \right) - \sqrt{p_{l,k}} \sum_{l'=1}^L (a'_{l',k})^* b'_{l',k} = 0. \quad (102) \end{aligned}$$

Therefore, the optimal solution $u_{l,k}^{\text{opt}}$ for a given set $\{\mathbf{a}_{l,k}, w_{l,k}, \rho_{l,k}\}$ is given in (103).

The optimal solution $w_{l,k}^{\text{opt}}$ is obtained by taking the first derivative of the objective function of the optimization problem (52) with respect to $w_{l,k}$ and equating to zero:

$$w_{l,k}^{\text{opt}} = e_{l,k}^{-1}. \quad (104)$$

Using (103) and (104) in (52), we obtain the optimization problem

$$\begin{aligned} &\underset{\{p_{l,k} \geq 0\}, \{\mathbf{a}_{l,k}\}}{\text{minimize}} && KL - \sum_{l=1}^L \sum_{k=1}^K \ln(1 + \text{SINR}_{l,k}) \\ &\text{subject to} && p_{l,k} \leq P_{\max,l,k}, \forall l, k. \end{aligned} \quad (105)$$

Since (105) is easily converted to (51), we have completed the proof.

Appendix F Proof of Theorem 4

For sake of simplicity, we omit the iteration index in the proof. The optimal solution of $u_{l,k}$ and $w_{l,k}$ are easily computed by (103) and (104) by noting that $\rho_{l,k} = \sqrt{p_{l,k}}$. We can find the optimal solution to $\mathbf{a}_{l,k}$ for a given set of $\{u_{l,k}, w_{l,k}, \rho_{l,k}\}$ from the optimization problem

$$\underset{\{\mathbf{a}_{l,k}\}}{\text{minimize}} \quad \sum_{l=1}^L \sum_{k=1}^K w_{l,k} \tilde{e}_{l,k} \quad (106)$$

where $\tilde{e}_{l,k}$ in (106) depends on $\{\mathbf{a}_{l,k}\}$ and is defined as

$$\begin{aligned} \tilde{e}_{l,k} &= |u_{l,k}|^2 \left(\sum_{l'=1}^L \rho_{l',k}^2 \left| \sum_{l''=1}^L (a'_{l',k})^* b'_{l',k} \right|^2 \right. \\ &+ \sum_{l'=1}^L \sum_{k'=1}^K \sum_{l''=1}^L \rho_{l',k'}^2 |a'_{l',k'}|^2 c_{l',k'} + \sum_{l'=1}^L |a'_{l',k}|^2 d_{l',k} \left. \right) \\ &- \rho_{l,k} u_{l,k} \left(\sum_{l'=1}^L (a'_{l',k})^* b'_{l',k} \right) - \rho_{l,k} u_{l,k}^* \left(\sum_{l'=1}^L a'_{l',k} (b'_{l',k})^* \right). \quad (107) \end{aligned}$$

By denoting $f(\mathbf{a}_{l,k}) = \sum_{l=1}^L \sum_{k=1}^K w_{l,k} \tilde{e}_{l,k}$ and using the expression of $\tilde{e}_{l,k}$ in (107), we can write $f(\mathbf{a}_{l,k})$ as

$$\begin{aligned} f(\mathbf{a}_{l,k}) &= \sum_{l=1}^L \sum_{k=1}^K w_{l,k} \left(|u_{l,k}|^2 \mathbf{a}_{l,k}^H \tilde{\mathbf{C}}_{l,k} \mathbf{a}_{l,k} \right. \\ &\left. - u_{l,k} \rho_{l,k} \mathbf{a}_{l,k}^H \mathbf{b}_{l,k} - u_{l,k}^* \rho_{l,k} \mathbf{b}_{l,k}^H \mathbf{a}_{l,k} \right). \quad (108) \end{aligned}$$

Taking the first derivative of $f(\mathbf{a}_{l,k})$ with respect to $\mathbf{a}_{l,k}$, we obtain

$$\nabla f = 2w_{l,k} |u_{l,k}|^2 \tilde{\mathbf{C}}_{l,k} \mathbf{a}_{l,k} - 2w_{l,k} u_{l,k} \rho_{l,k} \mathbf{b}_{l,k}. \quad (109)$$

Therefore, the solution is

$$\mathbf{a}_{l,k}^{\text{opt}} = \frac{\rho_{l,k}}{u_{l,k}^*} \tilde{\mathbf{C}}_{l,k}^{-1} \mathbf{b}_{l,k}. \quad (110)$$

After removing $\rho_{l,k}$ in both the numerator and denominator of the fraction in (110) and doing some algebra, the optimal solution to $\mathbf{a}_{l,k}$ is expressed as in (60).

We now compute the optimal solution for $\rho_{l,k}$ for a given set of optimization variables $\{\mathbf{a}_{l,k}, w_{l,k}, u_{l,k}\}$. In this case, (54) simplifies to

$$\begin{aligned} &\underset{\{\rho_{l,k} \geq 0\}}{\text{minimize}} && \sum_{l=1}^L \sum_{k=1}^K w_{l,k} e_{l,k} \\ &\text{subject to} && \rho_{l,k}^2 \leq P_{\max,l,k}, \forall l, k. \end{aligned} \quad (111)$$

The Lagrangian function of the optimization (111) is

$$\begin{aligned} \mathcal{L}(\{\rho_{l,k}\}, \{\lambda_{l,k}\}) &= \sum_{l=1}^L \sum_{k=1}^K w_{l,k} e_{l,k} + \\ &\sum_{l=1}^L \sum_{k=1}^K \lambda_{l,k} (\rho_{l,k}^2 - P_{\max,l,k}), \quad (112) \end{aligned}$$

$$u_{l,k}^{\text{opt}} = \frac{\sqrt{p_{l,k}} \sum_{l'=1}^L a_{l',k}^{l'} (b_{l',k}^{l'})^*}{\sum_{l'=1}^L p_{l',k} \left| \sum_{l''=1}^L (a_{l'',k}^{l''})^* b_{l'',k}^{l''} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K \sum_{l''=1}^L p_{l',k'} |a_{l',k}^{l''}|^2 c_{l'',k'}^{l',k'} + \sum_{l'=1}^L |a_{l',k}^{l'}|^2 d_{l',k}}. \quad (103)$$

where $\lambda_{l,k} \geq 0$ is the Lagrange multiplier associated with $\rho_{l,k}^2 \leq P_{\max,l,k}$. Taking the first derivative of $\mathcal{L}(\{\rho_{l,k}\}, \{\lambda_{l,k}\})$ with respect to $\rho_{l,k}$, we obtain

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \rho_{l,k}} &= 2\rho_{l,k} \sum_{l'=1}^L w_{l',k} |u_{l',k}|^2 \left| \sum_{l''=1}^L (a_{l'',k}^{l''})^* b_{l'',k}^{l''} \right|^2 \\ &\quad + 2\rho_{l,k} \sum_{l'=1}^L \sum_{k'=1}^K w_{l',k'} |u_{l',k'}|^2 \sum_{l''=1}^L |a_{l',k'}^{l''}|^2 c_{l'',k'}^{l',k} \\ &\quad - w_{l,k} u_{l,k} \left(\sum_{l'=1}^L (a_{l',k}^{l'})^* b_{l',k}^{l'} \right) - w_{l,k} u_{l,k}^* \left(\sum_{l'=1}^L a_{l',k}^{l'} (b_{l',k}^{l'})^* \right) \\ &\quad + 2\lambda_{l,k} \rho_{l,k}. \end{aligned} \quad (113)$$

By equating the above derivative to zero, the stationary point is obtained as shown in (114). The Lagrangian multiplier $\lambda_{l,k}$ must satisfy the complementary slackness condition [28]

$$\lambda_{l,k} (\rho_{l,k}^2 - P_{\max,l,k}) = 0. \quad (115)$$

Therefore, we obtain the solution to $\rho_{l,k}$ as

$$\rho_{l,k} = \begin{cases} \min(\tilde{\rho}_{l,k}, \sqrt{P_{\max,l,k}}), & \text{if } \lambda_{l,k} = 0, \\ \sqrt{P_{\max,l,k}}, & \text{if } \lambda_{l,k} \neq 0, \end{cases} \quad (116)$$

where $\tilde{\rho}_{l,k}$ is defined as in (117), which is obtained from (114) by setting $\lambda_{l,k} = 0$. From (116), the optimal solution to $\rho_{l,k}$ is derived as shown in (62).

We now prove that Algorithm 1 converges to a stationary point, as defined in Definition 1. The optimization problem (54) is first converted to the following equivalent unconstrained problem:

$$\begin{aligned} &\underset{\substack{\{\rho_{l,k} \geq 0\}, \{\mathbf{a}_{l,k}\}, \\ \{w_{l,k} \geq 0\}, \{u_{l,k}\}}} \text{minimize} && g(\{u_{l,k}\}, \{w_{l,k}\}, \{\mathbf{a}_{l,k}\}, \{\rho_{l,k}\}) \end{aligned} \quad (118)$$

where the objective function g is defined as

$$g(\{u_{l,k}\}, \{w_{l,k}\}, \{\mathbf{a}_{l,k}\}, \{\rho_{l,k}\}) = \sum_{l=1}^L \sum_{k=1}^K w_{l,k} e_{l,k} - \ln(w_{l,k}) + \lambda_{l,k} (\rho_{l,k}^2 - P_{\max,l,k}). \quad (119)$$

Since in every iteration, each subproblem is convex and has a unique optimal solution, the objective function of the optimization problem (118) is monotonically decreasing after iterations. Additionally, this function is lower bounded by zero, so Algorithm 1 must converge to a limit point, attained by a solution that we call $\{(u_{l,k}^{\text{opt}}, w_{l,k}^{\text{opt}}, \mathbf{a}_{l,k}^{\text{opt}}, \rho_{l,k}^{\text{opt}})\}$. Note that $g(\{u_{l,k}\}, \{w_{l,k}\}, \{\mathbf{a}_{l,k}\}, \{\rho_{l,k}\})$ is convex in each optimization variable, when the others variables are fixed, and the optimal solution to each sub-problem is computed from the first derivative of the cost function. By applying the standard trick in [29, Remark 2.2] to decompose a

complex number into the real and imaginary parts, the following properties are obtained:

$$\left(\Re(u_{l,k}) - \Re(u_{l,k}^{\text{opt}}) \right) \frac{\partial g}{\partial \Re(u_{l,k})} \Big|_{\Re(u_{l,k}^{\text{opt}})} \geq 0, \quad (120)$$

$$\left(\Im(u_{l,k}) - \Im(u_{l,k}^{\text{opt}}) \right) \frac{\partial g}{\partial \Im(u_{l,k})} \Big|_{\Im(u_{l,k}^{\text{opt}})} \geq 0, \quad (121)$$

$$(w_{l,k} - w_{l,k}^{\text{opt}}) \frac{\partial g}{\partial w_{l,k}} \Big|_{w_{l,k}^{\text{opt}}} \geq 0, \quad (122)$$

$$\left(\Re(\mathbf{a}_{l,k}) - \Re(\mathbf{a}_{l,k}^{\text{opt}}) \right)^T \nabla g \Big|_{\Re(\mathbf{a}_{l,k}^{\text{opt}})} \geq 0, \quad (123)$$

$$\left(\Im(\mathbf{a}_{l,k}) - \Im(\mathbf{a}_{l,k}^{\text{opt}}) \right)^T \nabla g \Big|_{\Im(\mathbf{a}_{l,k}^{\text{opt}})} \geq 0, \quad (124)$$

$$(\rho_{l,k} - \rho_{l,k}^{\text{opt}}) \frac{\partial g}{\partial \rho_{l,k}} \Big|_{\rho_{l,k}^{\text{opt}}} \geq 0. \quad (125)$$

These properties mean that the limit point is a stationary point to (54).

We now prove that the optimal solution $\{\mathbf{a}_{l,k}^{\text{opt}}\}, \{(\rho_{l,k}^{\text{opt}})^2\}$ forms a stationary point of (51). In fact, the optimization problem (51) is equivalent to

$$\begin{aligned} &\underset{\{\rho_{l,k} \geq 0\}, \{\mathbf{a}_{l,k}\}} \text{maximize} && h(\{\mathbf{a}_{l,k}\}, \{\rho_{l,k}\}) \\ &\text{subject to} && \rho_{l,k}^2 \leq P_{\max,l,k}, \forall l, k. \end{aligned} \quad (126)$$

where the objective function is

$$h(\{\mathbf{a}_{l,k}\}, \{\rho_{l,k}\}) \triangleq \sum_{l=1}^L \sum_{k=1}^K \log_2(1 + \text{SINR}_{l,k}). \quad (127)$$

Here, the SINR value has a similar expression as in (39), but with $p_{l,k} = \rho_{l,k}^2$. For given $w_{l,k} = w_{l,k}^{\text{opt}}$ and $u_{l,k} = u_{l,k}^{\text{opt}}$, for all l, k , it is sufficient to prove the following equalities

$$\frac{\partial g}{\partial \rho_{l',k'}} = \frac{1}{\log_2(e)} \frac{\partial h}{\partial \rho_{l',k'}} + 2\lambda_{l',k'} \rho_{l',k'}, \forall l', k', \quad (128)$$

$$\nabla g(\mathbf{a}_{l',k'}) = \frac{1}{\log_2(e)} \nabla h(\mathbf{a}_{l',k'}), \forall l', k'. \quad (129)$$

By using (103), (104), and the chain rule, (128) is proved as

$$\begin{aligned} \frac{\partial g}{\partial \rho_{l',k'}} &= \sum_{l=1}^L \sum_{k=1}^K w_{l,k}^{\text{opt}} \frac{\partial e_{l,k}}{\partial \rho_{l',k'}} + 2\lambda_{l',k'} \rho_{l',k'} \\ &= \sum_{l=1}^L \sum_{k=1}^K (e_{l,k}^{\text{opt}})^{-1} \frac{\partial e_{l,k}^{\text{opt}}}{\partial \rho_{l',k'}} + 2\lambda_{l',k'} \rho_{l',k'}, \end{aligned} \quad (130)$$

$$\rho_{l,k} = \frac{w_{l,k} \Re \left(u_{l,k} \sum_{l'=1}^L (a_{l',k}^{l'})^* b_{l',k}^{l'} \right)}{\sum_{l'=1}^L w_{l',k} |u_{l',k}|^2 \left| \sum_{l''=1}^L (a_{l'',k}^{l''})^* b_{l'',k}^{l''} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K w_{l',k'} |u_{l',k'}|^2 \sum_{l''=1}^L |a_{l'',k'}^{l''}|^2 c_{l'',k'}^{l,k} + \lambda_{l,k}}. \quad (114)$$

$$\tilde{\rho}_{l,k} \triangleq \frac{w_{l,k} \Re \left(u_{l,k} \sum_{l''=1}^L (a_{l'',k}^{l''})^* b_{l'',k}^{l''} \right)}{\sum_{l'=1}^L w_{l',k} |u_{l',k}|^2 \left| \sum_{l''=1}^L (a_{l'',k}^{l''})^* b_{l'',k}^{l''} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K w_{l',k'} |u_{l',k'}|^2 \sum_{l''=1}^L |a_{l'',k'}^{l''}|^2 c_{l'',k'}^{l,k}}. \quad (117)$$

where $e_{l,k}^{\text{opt}} = (1 + \text{SINR}_{l,k})^{-1}$ is derived by using (103) in (67) and some algebra. It leads to

$$\begin{aligned} \frac{\partial g}{\partial \rho_{l',k'}} &= \sum_{l=1}^L \sum_{k=1}^K (1 + \text{SINR}_{l,k}) \frac{\partial (1 + \text{SINR}_{l,k})^{-1}}{\partial \rho_{l',k'}} \\ &\quad + 2\lambda_{l',k'} \rho_{l',k'} \\ &= \sum_{l=1}^L \sum_{k=1}^K (1 + \text{SINR}_{l,k})^{-1} \frac{\partial \text{SINR}_{l,k}}{\partial \rho_{l',k'}} + 2\lambda_{l',k'} \rho_{l',k'} \\ &= \frac{1}{\log_2(e)} \frac{\partial h}{\partial \rho_{l',k'}} + 2\lambda_{l',k'} \rho_{l',k'}. \end{aligned} \quad (131)$$

The proof of (129) is similar to how (128) just was proved.

References

- [1] T. V. Chien, C. Mollén, and E. Björnson, “Two-layer decoding in cellular Massive MIMO systems with spatial channel correlation,” in Proc. ICC, 2019, submitted.
- [2] E. Björnson, E. G. Larsson, and T. L. Marzetta, “Massive MIMO: 10 myths and one critical question,” *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, 2016.
- [3] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.
- [4] A. Gupta and R. K. Jha, “A survey of 5G network: Architecture and emerging technologies,” *IEEE Access*, vol. 3, pp. 1206–1232, 2015.
- [5] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What will 5G be?” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [6] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.
- [7] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, “Pilot contamination and precoding in multi-cell TDD systems,” *IEEE Trans. Commun.*, vol. 10, no. 8, pp. 2640–2651, 2011.
- [8] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO has unlimited capacity,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, 2018.
- [9] E. Björnson, E. G. Larsson, and M. Debbah, “Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?” *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, 2016.
- [10] S. Jin, M. Li, Y. Huang, Y. Du, and X. Gao, “Pilot scheduling schemes for multi-cell massive multiple-input-multiple-output transmission,” *IET Communications*, vol. 9, no. 5, pp. 689–700, 2015.
- [11] X. Zhu, Z. Wang, L. Dai, and C. Qian, “Smart pilot assignment for Massive MIMO,” *IEEE Commun. Letters*, vol. 19, no. 9, pp. 1644–1647, 2015.
- [12] A. Ashikhmin and T. Marzetta, “Pilot contamination precoding in multi-cell large scale antenna systems,” in Proc. IEEE ISIT, 2012.
- [13] A. Ashikhmin, L. Li, and T. L. Marzetta, “Interference reduction in multi-cell Massive MIMO systems with large-scale fading precoding,” *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6340–6361, 2018.
- [14] A. Adhikary, A. Ashikhmin, and T. L. Marzetta, “Uplink interference reduction in large-scale antenna systems,” *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2194–2206, 2017.
- [15] A. Adhikary and A. Ashikhmin, “Uplink Massive MIMO for channels with spatial correlation,” 2018. [Online]. Available: <https://arxiv.org/abs/1807.04473>
- [16] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.
- [17] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO networks: Spectral, energy, and hardware efficiency,” *Foundations and Trends in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.
- [18] E. Nayebe, A. Ashikhmin, T. L. Marzetta, and B. D. Rao, “Performance of cell-free Massive MIMO systems with MMSE and LSFDR receivers,” in Proc. ASILOMAR, 2016, pp. 203–207.
- [19] V. Annapureddy and V. Veeravalli, “Sum capacity of MIMO interference channels in the low interference regime,” *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2565–2581, 2011.
- [20] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, “An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, 2011.
- [21] P. Weeraddana, M. Codreanu, M. Latva-aho, A. Ephremides, and C. Fischione, “Weighted sum-rate maximization in wireless networks: A review,” *Foundations and Trends in Networking*, vol. 6, no. 1-2, pp. 1–163, 2012.
- [22] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, “Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, 2008.
- [23] R. Brandt, E. Björnson, and M. Bengtsson, “Weighted sum rate optimization for multicell MIMO systems with hardware-impaired transceivers,” in Proc. ICASSP, 2014, pp. 479–483.
- [24] Further advancements for E-UTRA physical layer aspects (Release 9). 3GPP TS 36.814, Mar. 2010.
- [25] S. Loyka, “Channel capacity of MIMO architecture using the exponential correlation matrix,” *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, 2001.
- [26] E. Björnson, M. Matthaiou, and M. Debbah, “Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, 2015.
- [27] Y. Yang and M. Pesavento, “A unified successive pseudoconvex approximation framework,” *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3313–3328, 2017.
- [28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [29] E. Björnson and E. Jorswieck, “Optimal resource allocation in coordinated multi-cell systems,” *Foundations and Trends in Communications and Information Theory*, vol. 9, no. 2-3, pp. 113–381, 2013.

PLACE
PHOTO
HERE

Trinh Van Chien (S'16) received the B.S. degree in electronics and telecommunications from the Hanoi University of Science and Technology, Vietnam, in 2012, and the M.S. degree in electrical and computer engineering from Sungkyunkwan University, South Korea, in 2014. He is currently pursuing the Ph.D. degree in communication systems from Linköping University, Sweden and being a Marie Skłodowska-Curie research fellow at the H2020 5Gwireless ITN project. His interest lies in convex optimization problems for wireless communications and image and video processing. He was an IEEE Wireless Communications Letters Exemplary Reviewer for 2016 and 2017. He was a recipient of the Award of Scientific Excellence in the first year of the 5Gwireless project funded by European Union Horizon's 2020.

PLACE
PHOTO
HERE

Christopher Mollén received the M.Sc. and Ph.D. degrees from Linköping University, Sweden, in 2013 and 2018, respectively. His Ph.D. thesis "High-End Performance with Low-End Hardware: Analysis of Massive MIMO Base Stations Transceivers" explored low-complexity hardware implementations of massive MIMO base stations, including low-PAR precoding, low-resolution ADCs, and nonlinear amplifiers. Previously, he was an intern at Ericsson in Kista, Sweden, and in Shanghai, China. From 2011 to 2012, he was an exchange student in the Erasmus Programme at the Eidgenössische Technische Hochschule Zürich, Switzerland. From 2015 to 2016, he visited the University of Texas at Austin as a Fulbright Scholar. The IEEE Transactions on Communications named him exemplary reviewer in 2017. This work was done while at Linköping University. He is currently with the 3D Vision Team at Apple.

PLACE
PHOTO
HERE

Emil Björnson (S'07-M'12-SM'17) received the M.S. degree in Engineering Mathematics from Lund University, Sweden, in 2007. He received the Ph.D. degree in Telecommunications from KTH Royal Institute of Technology, Sweden, in 2011. From 2012 to mid 2014, he was a joint postdoc at the Alcatel-Lucent Chair on Flexible Radio, SUPELEC, France, and at KTH. He joined Linköping University, Sweden, in 2014 and is currently Associate Professor and Docent at the Division of Communication Systems.

He performs research on multi-antenna communications, Massive MIMO, radio resource allocation, energy-efficient communications, and network design. He is on the editorial board of the IEEE Transactions on Communications (since 2017) and the IEEE Transactions on Green Communications and Networking (since 2016). He is the first author of the textbooks "Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency" (2017) and "Optimal Resource Allocation in Coordinated Multi-Cell Systems" from 2013. He is dedicated to reproducible research and has made a large amount of simulation code publicly available.

Dr. Björnson has performed MIMO research for more than ten years and has filed more than ten related patent applications. He received the 2018 Marconi Prize Paper Award in Wireless Communications, the 2016 Best PhD Award from EURASIP, the 2015 Ingvar Carlsson Award, and the 2014 Outstanding Young Researcher Award from IEEE ComSoc EMEA. He also co-authored papers that received best paper awards at the conferences WCSP 2017, IEEE ICC 2015, IEEE WCNC 2014, IEEE SAM 2014, IEEE CAMSAP 2011, and WCSP 2009.