

# Two-Layer Decoding in Cellular Massive MIMO Systems with Spatial Channel Correlation

Trinh Van Chien, Christopher Mollén, Emil Björnson

Department of Electrical Engineering (ISY), Linköping University, SE-581 83 Linköping, Sweden

{trinh.van.chien, emil.bjornson}@liu.se, chris.mollen@gmail.com

**Abstract**—This paper studies a two-layer decoding method that mitigates inter-cell interference in multi-cell Massive MIMO systems. In layer one, each base station (BS) estimates the channels to intra-cell users and uses the estimates 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, while a closed-form expression is obtained for correlated Rayleigh fading channels, maximum-ratio combining (MRC), and large-scale fading decoding (LSFD) in the second layer. We formulate a non-convex sum SE maximization problem with both the data power and LSFD vectors as optimization variables and develop an algorithm based on the weighted MMSE (minimum mean square error) approach to obtain a stationary point with low computational complexity.

## I. INTRODUCTION

Massive MIMO is an emerging technology to handle the growing demand for wireless data traffic in the next generation cellular networks [1]. A Massive MIMO BS is equipped with hundreds of antennas to spatially multiplex a large number of users on the same time–frequency resource [2]. In a single-cell scenario, there is no need for computationally heavy decoding (or precoding methods) in Massive MIMO as both the thermal noise and mutual interference are effectively suppressed by linear processing, e.g., maximum-ratio combining (MRC) or regularized zero-forcing (RZF) combining, with a large number of BS antennas [3]. 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, so-called *pilot contamination* [4], unless high-complexity signal processing schemes are used to suppress it [5]. Pilot contamination reduces the benefit of having many antennas and the SEs achieved by low-complexity MRC or RZF saturate as the number of antennas grows.

Much research has been dedicated to mitigating the effects of pilot contamination; for example by increasing the length of the pilots [6] or assigning the pilots in a way that reduces the contamination [7]. In practical networks, however, it is not possible to make all pilots orthogonal due to the limited channel coherence block [6]. Besides, the pilot assignment is a combinatorial problem. Even though heuristic algorithms

with relatively low complexity can be developed, this approach still suffers from the asymptotic SE saturation since we only change one contaminating user for a less contaminating user. Instead of combating pilot contamination, one can utilize decoding schemes where the BSs are cooperating [8]. In the two-layer LSFD (*large-scale fading decoding*) framework, each BS applies an arbitrary local linear decoding method in the first layer. The results are 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 other inter-cell interference. The LSFD vectors are selected only based on the channel statistics (large-scale fading) and, therefore, there is no need for the BSs to share their local channel estimates. The new decoding design attains high SE even with a limited number of BS antennas [8]. Previous works on LSFD has either considered uncorrelated Rayleigh fading channels [8], [9] or special correlated Rayleigh fading based on the one-ring model [10]. The latter paper optimizes the system with respect to network-wide max-min fairness, which is a criterion that gives all the users the same SE, but usually a very low such SE [11].

In this paper, we generalize the LSFD method from [8], [10] to a scenario with arbitrary spatial correlation and also develop a method for joint power control and LSFD vector optimization in the system using the sum SE as the utility. We first quantify the SE in a system with arbitrary processing in the two layers and then derive a closed-form expression for the case when MRC is used in the first layer. The LSFD vector that maximizes the SE follows in closed form. Additionally, an uplink sum SE optimization problem with power constraints is formulated. Because it is a hard non-convex problem, we are not searching for the global optimum but develop an alternating low-complexity optimization algorithm that converges to a stationary point. Numerical results demonstrate the effectiveness of the optimized system for Massive MIMO systems with correlated Rayleigh fading.

*Notation:* Lower and upper case bold letters are used for vectors and matrices. The expectation of a random variable  $X$  is denoted by  $\mathbb{E}\{X\}$  and  $\|\cdot\|$  is the Euclidean norm. The transpose and Hermitian transpose of a matrix are written as  $(\cdot)^T$  and  $(\cdot)^H$ , respectively. The  $L \times L$ -dimensional diagonal matrix with the diagonal elements  $d_1, d_2, \dots, d_L$  is denoted  $\text{diag}(d_1, d_2, \dots, d_L)$ . Finally,  $\mathcal{CN}(\cdot, \cdot)$  is circularly symmetric complex Gaussian distribution.

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.

## II. SYSTEM MODEL

We consider a cellular network with  $L$  cells. Each cell consists of a BS equipped with  $M$  antennas that serves  $K$  single-antenna users. The channel vector in the uplink between user  $k$  in cell  $l$  and BS  $l'$  is denoted by  $\mathbf{h}_{l',k}^l \in \mathbb{C}^M$ . We consider the standard block-fading model [11], where the channels are static within a coherence block of size  $\tau_c$  channel uses and take an independent realization in each blocks, according to a stationary ergodic random process. Since practical channels are spatially correlated, we assume that each channel follows a correlated Rayleigh fading model:

$$\mathbf{h}_{l',k}^l \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{l',k}^l), \quad (1)$$

where  $\mathbf{R}_{l',k}^l \in \mathbb{C}^{M \times M}$  is the *spatial correlation matrix*. 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 [5], the channels are estimated by letting the users transmit  $K$ -symbol long *pilots* in a dedicated part of the coherence block, called the *pilot phase*. All the cells share a common set of  $K$  mutually orthogonal pilots  $\{\phi_1, \dots, \phi_K\}$  with  $\|\phi_k\|^2 = K$ . Without loss of generality, we assume that the users in different cells that have the same index use the same pilot and thereby cause pilot contamination to each other [2]. During the pilot phase, at BS  $l$  the received signal in the pilot phase is denoted  $\mathbf{Y}_l \in \mathbb{C}^{M \times K}$  and it is given by

$$\mathbf{Y}_l = \sum_{l'=1}^L \sum_{k=1}^K \sqrt{\hat{p}_{l',k}} \mathbf{h}_{l',k}^l \phi_k^H + \mathbf{N}_l, \quad (2)$$

where  $\hat{p}_{l',k}$  is the power of the pilot transmitted by 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 observation of the channel from user  $k$  to BS  $l$  is obtained by using standard MMSE estimation [11]. The channel estimates are used at BS  $l$  to compute decoding vectors for detecting the signals from the  $K$  intra-cell users.

### B. Uplink Data Transmission

In the *data phase*, it is assumed that user  $k$  in cell  $l'$  sends a zero-mean information symbol  $s_{l',k}$  with power  $\mathbb{E}\{|s_{l',k}|^2\} = 1$ . The received signal  $\mathbf{y}_l \in \mathbb{C}^M$  at BS  $l$  is then

$$\mathbf{y}_l = \sum_{l'=1}^L \sum_{k=1}^K \sqrt{p_{l',k}} \mathbf{h}_{l',k}^l s_{l',k} + \mathbf{n}_l, \quad (3)$$

where  $p_{l',k}$  denotes the transmit power of user  $k$  in cell  $l'$ . Based on the signals in (3), the BSs decode the symbols using the two-layers-decoding technique that is illustrated in Fig. 1. In the first layer, an estimate of the symbol from user  $k$  in cell  $l$  is obtained at BS  $l$  by local linear decoding as

$$\tilde{s}_{l,k} = \mathbf{v}_{l,k}^H \mathbf{y}_l = \sum_{l'=1}^L \sum_{k'=1}^K \sqrt{p_{l',k'}} \mathbf{v}_{l,k}^H \mathbf{h}_{l',k'}^l s_{l',k'} + \mathbf{v}_{l,k}^H \mathbf{n}_l, \quad (4)$$

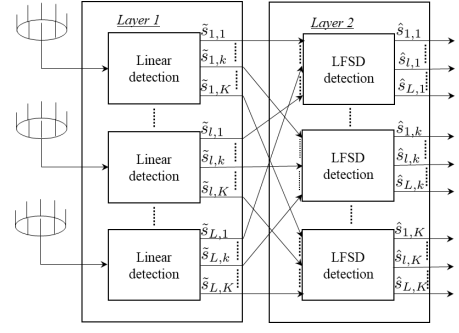


Fig. 1. Desired signals are detected by the two-layer decoding technique.

where  $\mathbf{v}_{l,k}$  is the *linear decoding vector*. The symbol estimate  $\tilde{s}_{l,k}$  contains interference and noise. In particular, the coherent interference caused by pilot contamination from pilot-sharing users in other cells is large in Massive MIMO. To mitigate the inter-cell interference, all the symbol estimates of the pilot-sharing users are collected in a vector

$$\tilde{\mathbf{s}}_k \triangleq [\tilde{s}_{1,k}, \tilde{s}_{2,k}, \dots, \tilde{s}_{L,k}]^T \in \mathbb{C}^L. \quad (5)$$

After the local decoding, a second layer of centralized decoding is performed. The final estimate of the data symbol from user  $k$  in cell  $l$  is obtained as

$$\hat{s}_{l,k} = \mathbf{a}_{l,k}^H \tilde{\mathbf{s}}_k = \sum_{l'=1}^L (a_{l',k}^{l'})^* \tilde{s}_{l',k} \quad (6)$$

where  $\mathbf{a}_{l,k} \triangleq [a_{l,k}^1, a_{l,k}^2, \dots, a_{l,k}^L]^T \in \mathbb{C}^L$  is called the *LSFD vector* and  $a_{l',k}^{l'}$  is the *LSFD weight*. Unlike previous works, in our framework, arbitrary linear combining methods can be used in the first layer and the LSFD vectors can still be optimized.

In the next section, we use the decoded signals  $\hat{s}_{l,k}$  together with the asymptotic channel properties [11, Section 2.5] to derive a closed-form expression for achievable uplink SE.

## III. UPLINK PERFORMANCE ANALYSIS

This section first derives a SE expression that can be used for any decoding vector and then a closed-form expression when using MRC. These expressions are then used to obtain the LSFD vectors that maximize the SE. The use-and-then-forget capacity bounding technique [3, Chapter 2.3.4], [5, Section 4.3] allows us to compute a lower bound on the uplink ergodic capacity (i.e., an achievable SE) of user  $k$  in cell  $l$  as

$$R_{l,k} = \max_{\{a_{l',k}^{l'}\}} \left(1 - \frac{K}{\tau_c}\right) \log_2(1 + \text{SINR}_{l,k}), \quad (7)$$

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

$$\frac{\mathbb{E}\{|\text{DS}_{l,k}|^2\}}{\mathbb{E}\{|\text{PC}_{l,k}|^2\} + \mathbb{E}\{|\text{BU}_{l,k}|^2\} + \mathbb{E}\{|\text{NI}_{l,k}|^2\} + \mathbb{E}\{|\text{AN}_{l,k}|^2\}}, \quad (8)$$

where  $\text{DS}_{l,k}$ ,  $\text{PC}_{l,k}$ ,  $\text{BU}_{l,k}$ ,  $\text{NI}_{l,k}$ , and  $\text{AN}_{l,k}$  stand for the desired signal, the pilot contamination, the beamforming gain uncer-

tainty, the non-coherent interference, and the additive noise, respectively, whose expectations are defined as

$$\mathbb{E}\{|\text{DS}_{l,k}|^2\} \triangleq p_{l,k} \left| \sum_{l'=1}^L (a'_{l',k})^* \mathbb{E}\{\mathbf{v}_{l',k}^H \mathbf{h}_{l',k}^1\} \right|^2, \quad (9)$$

$$\mathbb{E}\{|\text{PC}_{l,k}|^2\} \triangleq \sum_{\substack{l'=1 \\ l' \neq l}}^L p_{l',k} \left| \sum_{l'=1}^L (a'_{l',k})^* \mathbb{E}\{\mathbf{v}_{l',k}^H \mathbf{h}_{l',k}^1\} \right|^2, \quad (10)$$

$$\mathbb{E}\{|\text{BU}_{l,k}|^2\} \triangleq \sum_{l'=1}^L p_{l',k} \mathbb{E} \left\{ \left| \sum_{l''=1}^L (a'_{l'',k})^* \left( \mathbf{v}_{l',k}^H \mathbf{h}_{l'',k}^1 - \mathbb{E}\{\mathbf{v}_{l',k}^H \mathbf{h}_{l'',k}^1\} \right) \right|^2 \right\}, \quad (11)$$

$$\mathbb{E}\{|\text{NI}_{l',k'}|^2\} \triangleq \sum_{l'=1}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K p_{l',k'} \times \mathbb{E} \left\{ \left| \sum_{l''=1}^L (a'_{l'',k'})^* \mathbf{v}_{l',k'}^H \mathbf{h}_{l'',k'}^1 \right|^2 \right\}, \quad (12)$$

$$\mathbb{E}\{|\text{AN}_{l,k}|^2\} \triangleq \mathbb{E} \left\{ \left| \sum_{l'=1}^L (a'_{l',k})^* (\hat{\mathbf{h}}_{l',k}^1)^H \mathbf{n}_{l'} \right|^2 \right\}. \quad (13)$$

We notice that the SE expression in (7) can be applied together with any linear decoding method and any LSF vector, but the expectations have the evaluated numerically.

Maximizing the SE of user  $k$  in cell  $l$  is equivalent to selecting the LSF vector that maximizes a *Rayleigh quotient*.

**Theorem 1.** *If MRC, ZF or RZF is used, for a given set of pilot and data power coefficients, the SE of user  $k$  in cell  $l$  is*

$$R_{l,k} = \left(1 - \frac{K}{\tau_c}\right) \log_2 \left(1 + p_{l,k} \mathbf{b}_{l,k}^H \left( \sum_{i=1}^4 \mathbf{C}_{l,k}^{(i)} \right)^{-1} \mathbf{b}_{l,k}\right), \quad (14)$$

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

$$\mathbf{C}_{l,k}^{(1)} \triangleq \sum_{\substack{l'=1 \\ l' \neq l}}^L p_{l',k} \mathbf{b}_{l',k} \mathbf{b}_{l',k}^H, \quad (15)$$

$$\mathbf{C}_{l,k}^{(2)} \triangleq \sum_{l'=1}^L p_{l',k} \mathbb{E} \left\{ \tilde{\mathbf{b}}_{l',k} \tilde{\mathbf{b}}_{l',k}^H \right\}, \quad (16)$$

$$\mathbf{C}_{l,k}^{(3)} \triangleq \text{diag} \left( \sum_{\substack{l'=1 \\ k' \neq k}}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K p_{l',k'} \mathbb{E} \left\{ |\mathbf{v}_{1,k'}^H \mathbf{h}_{l',k'}^1|^2 \right\}, \dots, \sum_{\substack{l'=1 \\ k'=1 \\ k' \neq k}}^L \sum_{\substack{k'=1 \\ k' \neq k}}^K p_{l',k'} \mathbb{E} \left\{ |\mathbf{v}_{L,k'}^H \mathbf{h}_{l',k'}^L|^2 \right\} \right) \quad (17)$$

$$\mathbf{C}_{l,k}^{(4)} \triangleq \text{diag} \left( \sigma^2 \mathbb{E} \left\{ \|\mathbf{v}_{1,k}\|^2 \right\}, \dots, \sigma^2 \mathbb{E} \left\{ \|\mathbf{v}_{L,k}\|^2 \right\} \right), \quad (18)$$

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

$$\mathbf{b}_{l',k} \triangleq [\mathbb{E}\{\mathbf{v}_{1,k}^H \mathbf{h}_{l',k}^1\}, \dots, \mathbb{E}\{\mathbf{v}_{L,k}^H \mathbf{h}_{l',k}^L\}]^T, \quad (19)$$

$$\tilde{\mathbf{b}}_{l',k} \triangleq [\mathbf{v}_{1,k}^H \mathbf{h}_{l',k}^1, \dots, \mathbf{v}_{L,k}^H \mathbf{h}_{l',k}^L]^T - \mathbf{b}_{l',k}. \quad (20)$$

In order to attain this SE, the LSF vector is selected as

$$\mathbf{a}_{l,k} = \left( \sum_{i=1}^4 \mathbf{C}_{l,k}^{(i)} \right)^{-1} \mathbf{b}_{l,k}, \quad \forall l, k. \quad (21)$$

*Proof:* The proof relies on rewriting the SE as a generalized Rayleigh quotient and solving it. The details are available in the journal version of this paper [12]. ■

We stress that the LSF vector in (21) is designed to maximize the SE in (14) for every user in the network for a given data and pilot power and a given first-layer decoding method. This is a non-trivial generalization of the previous works [8]–[10], which only considered specific first-layer decoding methods that could provide closed-form expressions.<sup>1</sup> The following theorem states a closed-form expression of the SE for the case of MRC in arbitrary spatial correlation, which makes the results more practical than in [10].

**Theorem 2.** *When MRC is used, the SE in (7) of user  $k$  in cell  $l$  is given by*

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

where the SINR value is given in (23) on the top of the next page. The values  $b_{l',k}''', c_{l',k'}''',$  and  $d_{l',k}$  are given as

$$b_{l',k}''' = \sqrt{K \hat{p}_{l',k} \hat{p}_{l',k}'} \text{tr} \left( \Psi_{l',k}^{-1} \mathbf{R}_{l',k}'' \mathbf{R}_{l',k}''' \right), \quad (24)$$

$$c_{l',k'}''' = \hat{p}_{l',k} \text{tr} \left( \mathbf{R}_{l',k}'' \Psi_{l',k}^{-1} \mathbf{R}_{l',k}'' \mathbf{R}_{l',k'}''' \right), \quad (25)$$

$$d_{l',k} = \sigma^2 \hat{p}_{l',k} \text{tr} \left( \Psi_{l',k}^{-1} \mathbf{R}_{l',k}'' \mathbf{R}_{l',k}''' \right), \quad (26)$$

where  $\Psi_{l',k} = K \sum_{l=1}^K \hat{p}_{l,k} \mathbf{R}_{l,k}'' + \sigma^2 \mathbf{I}_M$  and  $\Psi_{l',k}$  is defined in the same manner.

*Proof.* The proof encompasses of computing the moments of complex Gaussian distributions and the detail is available in the journal version of this paper [12]. □

Theorem 2 describes the exact impact that the spatial correlation has on the system performance through the coefficients  $b_{l',k}''', c_{l',k'}''',$  and  $d_{l',k}$ . It is seen that the numerator of (23) grows as the square of the number of antennas,  $M^2$ , since the trace in (24) 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 (6) 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

<sup>1</sup>We stress that Theorem 1 also holds in other cases, if we replace  $\mathbf{C}_{l,k}^{(3)}$  as  $\mathbf{C}_{l,k}^{(3)} = \sum_{l'=1}^L \sum_{k'=1, k' \neq k}^K p_{l',k'} \mathbb{E} \left\{ \mathbf{z}_{k,l',k'} \mathbf{z}_{k,l',k'}^H \right\}$ , where  $\mathbf{z}_{k,l',k'} = [\mathbf{v}_{1,k}^H \mathbf{h}_{l',k'}^1, \dots, \mathbf{v}_{L,k}^H \mathbf{h}_{l',k'}^L] \in \mathbb{C}^L$ .

$$\text{SINR}_{l,k} = \frac{p_{l,k} \left| \sum_{l'=1}^L (a'_{l,k})^* b'_{l,k} \right|^2}{\sum_{l'=1, l' \neq l}^L p_{l',k} \left| \sum_{l''=1}^L (a'_{l',k})^* b'_{l'',k} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K \sum_{l''=1}^L p_{l',k'} |a'_{l',k}|^2 c_{l',k'}^{l'',k'} + \sum_{l'=1}^L |a'_{l,k}|^2 d_{l',k}} \quad (23)$$

only grow as  $M$ . Since the interference terms contain products of correlation matrices of different users, the interference is smaller between users that have different spatial correlation characteristics [11].

The following corollary gives the optimal LSFD vector  $\mathbf{a}_{l,k}$  that maximizes the SE of every user for a given set of pilot and data powers.

**Corollary 1.** *For a given set of data and pilot powers, by using MRC and optimal LSFD, the SE in Theorem 2 is given in closed form as*

$$R_{l,k} = \left(1 - \frac{K}{\tau_c}\right) \log_2 \left(1 + p_{l,k} \mathbf{b}_{l,k}^H \mathbf{C}_{l,k}^{-1} \mathbf{b}_{l,k}\right) \quad (27)$$

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

$$\mathbf{C}_{l,k} \triangleq \sum_{\substack{l'=1 \\ l' \neq l}}^L p_{l',k} \mathbf{b}_{l',k} \mathbf{b}_{l',k}^H + \text{diag} \left( \sum_{l'=1}^L \sum_{k'=1}^K p_{l',k'} c_{l',k'}^{l',k'} + d_{1,k}, \dots, \sum_{l'=1}^L \sum_{k'=1}^K p_{l',k'} c_{l',k'}^{l',k'} + d_{L,k} \right), \quad (28)$$

$$\mathbf{b}_{l,k} \triangleq [b_{l',k}^1, \dots, b_{l',k}^L]^T. \quad (29)$$

The SE in (27) is obtained by using LSFD vector

$$\mathbf{a}_{l,k} = \mathbf{C}_{l,k}^{-1} \mathbf{b}_{l,k}. \quad (30)$$

Although Corollary 1 is a special case of Theorem 1 when MRC is used, the LSFD vector  $\mathbf{a}_{l,k}$  is obtained in closed form.

#### IV. OPTIMIZING THE SUM SE

In this section, the sum SE maximization problem is formulated where the optimization variables are the data powers and LSFD vectors. Since this problem is NP-hard, an iterative algorithm is proposed to find a stationary point with low computational complexity.

##### A. Problem Formulation

We consider sum SE maximization

$$\begin{aligned} & \underset{\{p_{l,k} \geq 0\}, \{\mathbf{a}_{l,k}\}}{\text{maximize}} && \sum_{l=1}^L \sum_{k=1}^K R_{l,k} \\ & \text{subject to} && p_{l,k} \leq P_{\max,l,k} \quad \forall l, k. \end{aligned} \quad (31)$$

Inserting the SE expression (22) into (31), and removing the constant pre-log factor, yields the equivalent formulation

$$\begin{aligned} & \underset{\{p_{l,k} \geq 0\}, \{\mathbf{a}_{l,k}\}}{\text{maximize}} && \sum_{l=1}^L \sum_{k=1}^K \log_2 (1 + \text{SINR}_{l,k}) \\ & \text{subject to} && p_{l,k} \leq P_{\max,l,k} \quad \forall l, k. \end{aligned} \quad (32)$$

Sum SE maximization with imperfect CSI is known to be a non-convex and NP-hard problem [13] and this applies also to

(32), even if the optimal LSFD vectors are given in Corollary 1. Therefore, the global optimum is overly difficult to compute. Nevertheless, solving the ergodic sum SE maximization in (32) for a Massive MIMO system is more practical than maximizing the instantaneous SEs for a given small-scale fading realization, as is normally done in small-scale MIMO systems [14]. Since the sum SE maximization in (32) only depends on the large-scale fading coefficients, the solution can be used for as much time as the channel statistics are constant. Another key difference from prior work is that we jointly optimize the data powers and LSFD vectors.

Instead of seeking the global optimum to (32), we will obtain a stationary point to (32) by following the weighted MMSE approach from [14] and adapt it to the problem at hand. To this end, we first formulate the weighted MMSE problem that is equivalent to (32).

**Theorem 3.** *The optimization problem*

$$\begin{aligned} & \underset{\{p_{l,k} \geq 0\}, \{\mathbf{a}_{l,k}\}, \{w_{l,k} \geq 0\}}{\text{minimize}} && \sum_{l=1}^L \sum_{k=1}^K w_{l,k} e_{l,k} - \ln(w_{l,k}) \\ & \text{subject to} && p_{l,k} \leq P_{\max,l,k}, \quad \forall l, k, \end{aligned} \quad (33)$$

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

$$\begin{aligned} e_{l,k} \triangleq & |u_{l,k}|^2 \left( \sum_{l'=1}^L p_{l',k} \left| \sum_{l''=1}^L (a'_{l',k})^* b'_{l'',k} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K \sum_{l''=1}^L p_{l',k'} |a'_{l',k}|^2 c_{l',k'}^{l'',k'} \right. \\ & \left. + \sum_{l'=1}^L |a'_{l,k}|^2 d_{l',k} \right) - 2\sqrt{p_{l,k}} \times \\ & \Re \left( u_{l,k} \left( \sum_{l'=1}^L (a'_{l',k})^* b'_{l',k} \right) \right) + 1, \end{aligned} \quad (34)$$

is equivalent to the sum SE optimization problem (32) in the sense that (32) and (33) have the same optimal transmit powers  $\{p_{l,k}\}, \forall l, k$ , and LSFD vectors  $\{\mathbf{a}_{l,k}\}, \forall l, k, l'$ .

*Proof.* The proof is based on the signal detection process of a SISO system having the same SE as (27). The detail proof is available in our journal version [12].  $\square$

##### B. Iterative Algorithm

We will obtain a stationary point to (33) by decomposing it into a sequence of subproblems, each having a closed-form solution. To this end, the power variable  $p_{l,k}$  is substituted with  $\rho_{l,k} = \sqrt{p_{l,k}}$ . By alternating between solving the subproblems we obtain the following result.

**Theorem 4.** *A stationary point to (33) is obtained by iteratively updating  $\{\mathbf{a}_{l,k}, w_{l,k}, \rho_{l,k}\}$ . 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*

$$\tilde{u}_{l,k}^{(n-1)} = \sum_{l'=1}^L (\rho_{l',k}^{(n-1)})^2 \left| \sum_{l''=1}^L (a_{l,k}^{l'',(n-1)})^* b_{l',k}^{l''} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K \sum_{l''=1}^L (\rho_{l',k'}^{(n-1)})^2 |a_{l,k}^{l'',(n-1)}|^2 c_{l'',k'}^{l',k'} + \sum_{l'=1}^L |a_{l,k}^{l',(n-1)}|^2 d_{l',k} \quad (38)$$

$$\tilde{\mathbf{C}}_{l,k}^{(n-1)} = \sum_{l'=1}^L (\rho_{l',k}^{(n-1)})^2 \mathbf{b}_{l',k} \mathbf{b}_{l',k}^H + \text{diag} \left( \sum_{l'=1}^L \sum_{k'=1}^K (\rho_{l',k'}^{(n-1)})^2 c_{1,k}^{l',k'} + d_{1,k}, \dots, \sum_{l'=1}^L \sum_{k'=1}^K (\rho_{l',k'}^{(n-1)})^2 c_{L,k}^{l',k'} + d_{L,k} \right) \quad (42)$$

$$\rho_{l,k}^{(n)} = \min \left( \frac{w_{l,k}^{(n)} \Re \left( u_{l,k}^{(n)} \sum_{l'=1}^L (a_{l,k}^{l',(n)})^* b_{l',k}^{l'} \right)}{\sum_{l'=1}^L w_{l',k}^{(n)} |u_{l',k}^{(n)}|^2 \left| \sum_{l''=1}^L (a_{l',k}^{l'',(n)})^* b_{l',k}^{l''} \right|^2 + \sum_{l'=1}^L \sum_{k'=1}^K w_{l',k'}^{(n)} |u_{l',k'}^{(n)}|^2 \sum_{l''=1}^L |a_{l',k'}^{l'',(n)}|^2 c_{l'',k'}^{l',k'}} \right), \sqrt{P_{\max,l,k}}} \quad (43)$$

iteration  $n$ , the optimization parameters are updated in the following way:

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

$$u_{l,k}^{(n)} = \left( \rho_{l,k}^{(n-1)} \sum_{l'=1}^L a_{l,k}^{l',(n-1)} (b_{l',k}^{l'})^* \right) / \tilde{u}_{l,k}^{(n-1)}, \quad (35)$$

where the value  $\tilde{u}_{l,k}^{(n-1)}$  is defined in (38) on the top of this page.

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

$$w_{l,k}^{(n)} = 1/e_{l,k}^{(n)}, \quad (39)$$

where  $e_{l,k}^{(n)}$  is given by

$$e_{l,k}^{(n)} = |u_{l,k}^{(n)}|^2 \tilde{u}_{l,k}^{(n-1)} - 2\rho_{l,k}^{(n-1)} \times \Re \left( u_{l,k}^{(n)} \left( \sum_{l'=1}^L (a_{l,k}^{l',(n-1)})^* b_{l',k}^{l'} \right) \right) + 1. \quad (40)$$

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

$$\mathbf{a}_{l,k}^{(n)} = \tilde{u}_{l,k}^{*(n)} \left( \tilde{\mathbf{C}}_{l,k}^{(n-1)} \right)^{-1} \mathbf{b}_{l,k} / \sum_{l'=1}^L (a_{l,k}^{l',(n-1)})^* b_{l',k}^{l'}, \quad (41)$$

where  $\tilde{\mathbf{C}}_{l,k}^{(n-1)}$  is computed as in (42) on the top of this page.

- $\rho_{l,k}$  is updated as in (43) on the top of this page.

If we denote the stationary point to (33) 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  $(\rho_{l,k}^{\text{opt}})^2$ , 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 (32).

*Proof.* The closed-form expression to each optimization variable is obtained by taking the first derivative of the Lagrangian function and equating it to zero, while the same stationary point of problems (32) and (33) is based on the chain rule. The detail proof is available in the journal version [12].  $\square$

If the initial data power values are uniformly distributed over the range  $[0, \sqrt{P_{\max,l,k}}]$ , the initial LSFD vectors can be computed using Corollary 1. The iterative algorithm in Theorem 4 is then used to obtain a stationary point to problem (31). This algorithm is terminated when the variation between two consecutive iterations is sufficiently small.

## V. NUMERICAL RESULTS

We consider a wrapped-around cellular network with four cells. 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 carried out over 300 random sets of user locations. We model the system parameters and large-scale fading similar to the 3GPP LTE specifications [15]. 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  $\beta_{l,k}^{l'}$  is computed in decibel scale as  $\beta_{l,k}^{l'} = -148.1 - 37.6 \log_{10}(d_{l',k}^l/1 \text{ km}) + z_{l,k}^{l'}$ , where the decibel value of the shadow fading,  $z_{l,k}$ , has a Gaussian distribution with zero mean and standard derivation 7. The spatial correlation matrix of the channel from user  $k$  in cell  $l$  to BS  $l'$  is described by the exponential correlation model, that models a uniform linear array with the correlation magnitude  $\varsigma \in [0, 1]$  [16]. The correlation magnitude is multiplied with a unique phase-shift in every correlation matrix, selected as the user's incidence angle to the array. 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. The following methods are compared in the simulation:

- (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 a modified version of Theorem 4.
- (iii) *Two-layer decoding system with fixed data power and LSFD vectors:* The network deploys the two-layer decoding as shown in Fig. 1, using MRC and LSFD. The data symbols are transmitted at the maximum power 200 mW and the LSFD vectors are computed using Corollary 1.
- (iv) *Two-layer decoding system with optimized data power and LSFD vectors:* This is the proposed method, where the data powers and LSFD vectors are computed using the weighted MMSE algorithm as in Theorem 4.

Fig. 2 shows the convergence of the proposed method for sum SE optimization in Theorem 4. From the initial random data powers, uniformly distributed in the feasible set, updating the optimization variables gives improved sum SE in every

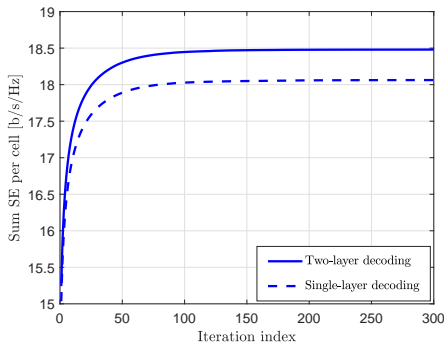


Fig. 2. Convergence of the proposed sum SE optimization with  $M = 200$ ,  $K = 5$ ,  $\varsigma = 0.8$ .

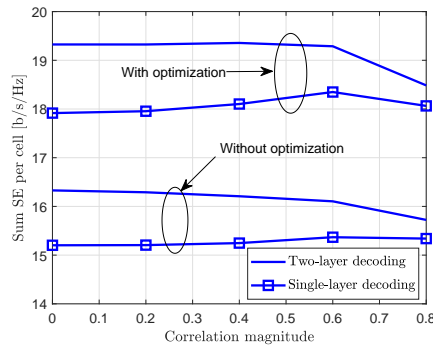


Fig. 3. Sum SE per cell versus different correlation magnitudes with  $M = 200$ ,  $K = 5$ .

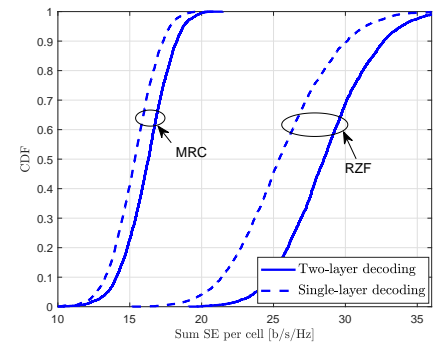


Fig. 4. CDF of sum SE per cell for MRC and RZF with  $M = 200$ ,  $K = 5$ ,  $\varsigma = 0.5$ .

iteration. For the two layer case (iv), the sum SE per cell is about 22.2% better at the stationary point than at the initial point. At convergence, (iv) gives 2.4% better sum SE than (ii). The proposed optimization methods need around 100 iterations to converge, but the complexity is low since every iteration in the algorithm consists of evaluating a closed-form expression.

Fig. 3 shows the sum SE per cell as a function of the channel correlation magnitude  $\varsigma$  for a multi-cell Massive MIMO system. First, we observe the substantial gains in sum SE attained by using LSFD. 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 optimized data powers and LSFD vectors. Moreover, this figure shows that the performance is greatly improved when the data powers are optimized. The gain varies from 17.9% to 20.3%. The gap becomes larger as the channel correlation magnitude increases. This shows the importance of doing joint data power control and LSFD optimization in Massive MIMO systems with spatially correlated channels.

Fig. 4 compares the cumulative distribution function (CDF) of the sum SE per cell with either MRC or RZF in the first layer, where the latter requires the use of the new general SE expression in Theorem 1. An equal power 200 mW is allocated to each transmitted symbol. Because RZF mitigates non-coherent interference effectively in the first layer, the second layer can increase the average SE by 11.80%. If MRC is used in the first layer, the SE gain from using LSFD using is only 5.84%. At the 95%-likely point, the two layer decoding system outperforms the single layer counterpart by 4.71% and 17.35% when using MRC or RZF, respectively.

## VI. CONCLUSION

This paper has investigated the ability of LSFD to mitigate inter-cell interference in multi-cell Massive MIMO systems with spatially correlated Rayleigh fading. LSFD is a two-layer decoding method, where a second decoding layer to mitigate inter-cell interference is applied after the classical decoding. We derived new SE expressions support arbitrary spatial correlation and first-layer decoding. We used these expressions to optimize the data powers and LSFD vectors, to maximize 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 demonstrate the effectiveness of LSFD in reducing pilot contamination with the improvement of sum SE for each cell about 17% in the tested scenarios, while the optimized data power control and LSFD design can improve the sum SE with more than 20%. The gains are larger when using RZF in the first layer than when using MRC.

## REFERENCES

- [1] 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.
- [2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [3] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.
- [4] 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.
- [5] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, 2018.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] E. Nayebe, A. Ashikhmin, T. L. Marzetta, and B. D. Rao, "Performance of cell-free Massive MIMO systems with MMSE and LSFD receivers," in *Proc. ASILOMAR*, 2016, pp. 203–207.
- [10] A. Adhikary and A. Ashikhmin, "Uplink Massive MIMO for channels with spatial correlation," 2018. [Online]. Available: <https://arxiv.org/abs/1807.04473>
- [11] 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.
- [12] T. Van Chien, C. Mollén, and E. Björnson, "Large-scale-fading decoding in cellular Massive MIMO systems with spatially correlated channels," *IEEE Trans. Commun.*, 2019, accepted for publication.
- [13] 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.
- [14] 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.

- [15] *Further advancements for E-UTRA physical layer aspects (Release 9)*. 3GPP TS 36.814, Mar. 2010.
- [16] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, 2001.