

Deep Learning-Aided Finite-Capacity Fronthaul Cell-Free Massive MIMO with Zero Forcing

Manijeh Bashar^{*†}, Ali Akbari^{*}, Kanapathippillai Cumanan[†], Hien Quoc Ngo[‡], Alister G. Burr[†], Pei Xiao^{*},
and Mérouane Debbah[§]

^{*}University of Surrey, UK., [†]University of York, UK, [‡]Queen’s University Belfast, UK, [§]CentraleSupélec,
Université Paris-Saclay, France. E-mail: {m.bashar,ali.akbari,p.xiao}@surrey.ac.uk,
{kanapathippillai.cumanan,alister.burr}@york.ac.uk, hien.ngo@qub.ac.uk, m.debbah@centralesupelec.fr

Abstract—We consider a cell-free massive multiple-input multiple-output (MIMO) system where the channel estimates and the received signals are quantized at the access points (APs) and forwarded to a central processing unit (CPU). Zero-forcing technique is used at the CPU to detect the signals transmitted from all users. To solve the non-convex sum rate maximization problem, a heuristic sub-optimal scheme is proposed to convert the problem into a geometric programme (GP). Exploiting a deep convolutional neural network (DCNN) allows us to determine both a *mapping* from the large-scale fading (LSF) coefficients and the optimal power by solving the optimization problem using the quantized channel. Depending on how the optimization problem is solved, different power control schemes are investigated; i) small-scale fading (SSF)-based power control; ii) LSF use-and-then-forget (UatF)-based power control; and iii) LSF deep learning (DL)-based power control. The SSF-based power control scheme needs to be solved for each coherence interval of the SSF, which is practically impossible in real time systems. Numerical results reveal that the proposed LSF-DL-based scheme significantly increases the performance compared to the practical and well-known LSF-UatF-based power control, thanks to the mapping obtained using DCNN.

I. INTRODUCTION

Cell-free massive multiple-input multiple-output (MIMO) is a promising technology, where large number of distributed access points (APs) serve much smaller number of users [1], [2]. A limited fronthaul cell-free massive MIMO is considered, where the access points (APs) send the quantized versions of the received signals and the channel estimates to a central processing unit (CPU) through limited fronthaul links [3], [4].

Zero forcing (ZF) receiver is investigated which can remove the inter-user interference [2]. In this paper, we study the sum rate optimization problem which needs to be solved for each coherence interval of the small-scale fading (SSF), which is infeasible in real time systems. This practical restriction is a more challenging in cell-free massive MIMO, as the optimization problem should be centrally solved at the CPU,

which introduces excessive delay. Therefore, it is essential to design a low complexity and practically feasible solution. Unfortunately, however no closed form expression for the sum rate is available in terms of the large-scale-fading (LSF) coefficients, and hence we formulate an optimization based on the quantized channel estimates (which is a function of the SSF coefficients), and provide a new heuristic approach to its solution. Finally we use the results from this optimization to train a deep convolutional neural network (DCNN) to determine the power control coefficients based on the LSF. Based on the results in [5], [6], the LSF changes much more slowly with time, some 120 times slower compared to the SSF.

The contributions of the paper are summarized as follows: (i) We propose a sum rate maximization problem taking into account the per-user power constraints and throughput requirement constraints, as well as the quantized channel estimates. Next, a heuristic and sub-optimal approach is proposed to reformulate the original optimization problem as a standard geometric programme (GP); (ii) We propose a LSF-deep learning (DL)-based power control scheme to allocate the power control coefficients using only the LSF coefficients as input. The main idea of the current paper is to train a neural network so that it will derive “optimum” transmit powers for the users using the LSF coefficients as inputs, based on the results of the optimization over a large number of randomly chosen SSF coefficients; (iii) Finally, the computational complexity of the proposed scheme is presented.

There are three important differences between the proposed DCNN-based algorithm in this paper and the scheme presented in [7], [8], which are: (i) In [7], [8], the authors exploit deep learning to solve an optimization problem which could be solved through the standard convex optimization software. However, without the machine learning tool, it is impossible to find a mapping between the LSF coefficients and the optimal power elements which are obtained by solving the optimization problem with knowledge of the quantized channel (which is a function of the SSF coefficients). This is because, given that only the quantized version of the estimated channel is available at the CPU as side information, the sum rate is a function of the SSF. So, it is *not* possible to explicitly find a mapping between the sum rate and LSF coefficients. This is the main difference between the current work and the work in [7],

The work of K. Cumanan and A. G. Burr was supported by H2020-MSCA-RISE-2015 under grant number 690750.

The work of H. Q. Ngo was supported by the UK Research and Innovation Future Leaders Fellowships under Grant MR/S017666/1.

The work of P. Xiao was supported by the U.K. Engineering and Physical Sciences Research Council under Grant EP/P008402/2 and EP/R001588/1. The authors also would like to acknowledge the support of the 5G Innovation Centre, University of Surrey, U.K. (<http://www.surrey.ac.uk/5gic>) for this work.

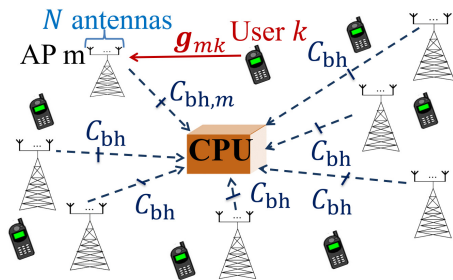


Figure 1. The uplink of a cell-free massive MIMO system with K single-antenna users and M APs.

[8]; **(ii)** The authors in [7] consider a cellular massive MIMO system, while here we consider a cell-free massive MIMO system. Note that unlike [7], having pure LSF components (i.e., the coefficients β_{mk} defined in Section II in (1)) as a raw input of the DCNN does not work in cell-free massive MIMO, and the network cannot learn the power elements obtained through the convex optimization approach. Hence, we generate a novel and unique input matrix to feed as the input to the DCNN for ZF receiver; **(iii)** Finally, in [7], [8], the authors use 2,000,000 training samples to train the neural network whereas only 60,000 training examples are enough for our proposed network.

II. SYSTEM MODEL

We investigate uplink transmission of a cell-free massive MIMO system with M APs (each equipped with N antennas) and K randomly distributed single-antenna users. The channel coefficients between the k th user and the m th AP, $\mathbf{g}_{mk} \in \mathbb{C}^{N \times 1}$, is modeled as

$$\mathbf{g}_{mk} = \sqrt{\beta_{mk}} \mathbf{h}_{mk}, \quad (1)$$

where β_{mk} denotes the LSF and the elements of \mathbf{h}_{mk} are independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ random variables, representing the SSF [1]. For each coherence interval, the transmission occurs into 2 main phases: channel estimation and uplink data transmission. All pilot sequences transmitted by all the K users in the channel estimation phase are collected in a matrix $\Phi \in \mathbb{C}^{\tau_p \times K}$, where τ_p is the length of the pilot sequence (in symbols) for each user and the k th column of Φ , ϕ_k , represents the pilot sequence used for the k th user. After performing a de-spreading operation, the minimum mean square error (MMSE) estimate of the channel vector between the k th user and the m th AP is given by [1]

$$\hat{\mathbf{g}}_{mk} = c_{mk} \left(\sqrt{\tau_p p_p} \mathbf{g}_{mk} + \sqrt{\tau_p p_p} \sum_{k' \neq k} \mathbf{g}_{mk'} \phi_k^H \phi_{k'} + \mathbf{W}_{p,m} \phi_k \right), \quad (2)$$

where $\mathbf{W}_{p,m} \in \mathbb{C}^{N \times \tau_p}$ denotes the noise at the m th AP whose elements are i.i.d. $\mathcal{CN}(0, 1)$, p_p represents the normalized signal-to-noise ratio (SNR) of each pilot symbol, and c_{mk} is given by $c_{mk} = \frac{\sqrt{\tau_p p_p} \beta_{mk}}{\tau_p p_p \sum_{k'=1}^K \beta_{mk'} |\phi_k^H \phi_{k'}|^2 + 1}$. Let the transmitted signal from the k th user be $x_k = \sqrt{q_k} s_k$, where s_k ($\mathbb{E}\{|s_k|^2\} = 1$) and q_k denote the transmitted symbol and the

transmit power of the k th user, respectively. Then the signal received at the m th AP is given by

$$\mathbf{y}_m = \sqrt{\rho} \sum_{k=1}^K \mathbf{g}_{mk} \sqrt{q_k} s_k + \mathbf{n}_m, \quad (3)$$

where $\mathbf{n}_m \in \mathbb{C}^{N \times 1}$, whose elements are i.i.d. $\mathcal{CN}(0, 1)$, is the noise at the m th AP, and ρq_k is the normalized uplink SNR corresponding to the k th user.

A. Quantization

First the m th AP quantizes the estimated channels, $\hat{\mathbf{g}}_{mk}$, $\forall k$ and the received signal, \mathbf{y}_m , using the optimal uniform quantization. Then it sends the quantized versions to the CPU. Using the Bussgang decomposition [9]–[14], the quantized signal can be expressed as

$$[\check{\mathbf{y}}_m]_n = \tilde{a} [\mathbf{y}_m]_n + [\mathbf{e}_m^{y,B}]_n \quad \forall m, n, \quad (4)$$

where $[\mathbf{e}_m^{y,B}]_n$ refers to the quantization distortion. Moreover, note that \tilde{a} is given in [3, Table I] and variance of the quantization distortion is given by $\sigma_{[\mathbf{e}_m^{y,B}]_n}^2 = \sigma_{\tilde{e},B}^2 \left(\rho \sum_{k'=1}^K \beta_{mk'} q_{k'} + 1 \right)$, $\forall m, n$, [3] and it is assumed that the same number of bits is used at all APs and all antennas to quantize the received signal. The optimal values of $\sigma_{\tilde{e},B}^2$ for different numbers of quantization bits are given in [3, Table I]. Moreover, we let α denote the number of quantization bits. Next, using the analysis in [3], the linear quantization is modeled as $\mathcal{Q}(z) = h(z) = z + \tilde{n}_d$, $\forall k$, where the output of the quantizer and the distortion are uncorrelated [15], [16]. Hence, similar to the scheme in [3], we quantize the estimated channel with the optimal quantizer obtained using the Max algorithm [15] as follows:

$$[\check{\mathbf{g}}_{mk}]_n = [\hat{\mathbf{g}}_{mk}]_n + [\mathbf{e}_{mk}^g]_n, \quad \forall k, m, n, \quad (5)$$

where $[\mathbf{e}_{mk}^g]_n$ is the quantization distortion. Moreover, the variance of the quantization distortion is $\sigma_{[\mathbf{e}_{mk}^g]_n}^2 = \sigma_{\tilde{e},g}^2 \gamma_{mk}$, $\forall m, k, n$, where $\sigma_{\tilde{e},g}^2 = \sigma_{\tilde{e}}^2$, which is given in [3, Table I].

B. Data Detection

Let $\check{\mathbf{V}} \in \mathbb{C}^{MN \times K}$ be the linear detector matrix depending on the side information at the receiver $\check{\mathbf{g}}_{mk}$, $\forall m, k$. We assume $\check{\mathbf{v}}_k = [\check{\mathbf{v}}_{1k}^T, \dots, \check{\mathbf{v}}_{Mk}^T]^T$ refers to the k th column of the detector matrix $\check{\mathbf{V}}$, and $\check{\mathbf{v}}_{mk} \in \mathbb{C}^N$. The received signal after using the linear detector at the CPU is given by

$$\check{s}_k = \check{\mathbf{v}}_k^H [\check{\mathbf{y}}_1^T, \dots, \check{\mathbf{y}}_M^T]^T, \quad (6)$$

where $\check{\mathbf{y}}_m$ is defined in (4). Then the transmitted signals from all K users will be detected from \check{s}_k .

$$\text{SINR}_k = \frac{\mathbb{E} \left\{ |DS_k s_k| \check{\mathbf{G}}|^2 \right\}}{\sum_{k'=1}^K \mathbb{E} \left\{ |IUI_{kk'}| \check{\mathbf{G}}|^2 \right\} + \mathbb{E} \left\{ |TN_k| \check{\mathbf{G}}|^2 \right\} + \frac{1}{\check{\alpha}^2} \mathbb{E} \left\{ |TQY_k| \check{\mathbf{G}}|^2 \right\} + \sum_{k'=1}^K \mathbb{E} \left\{ |TQG_{kk'}| \check{\mathbf{G}}|^2 \right\} + \sum_{k'=1}^K \mathbb{E} \left\{ |TEE_{kk'}| \check{\mathbf{G}}|^2 \right\}}. \quad (9)$$

$$\text{SINR}_k = \frac{\rho q_k}{\rho \sum_{k'=1}^K q_{k'} \sum_{m=1}^M \left[\beta_{mk'} \left(1 + \frac{\sigma_{\check{e}}^2}{\check{\alpha}^2} \right) - \gamma_{mk'} (1 - \sigma_{\check{e}^y}^2) \right] \|\check{\mathbf{v}}_{mk}\|^2 + \left(1 + \frac{\sigma_{\check{e}^y, B}^2}{\check{\alpha}^2} \right) \sum_{m=1}^M \|\check{\mathbf{v}}_{mk}\|^2}. \quad (10)$$

III. ACHIEVABLE RATE ANALYSIS

From (4) and (6), the received signal after using the linear detector is

$$\begin{aligned} \check{s}_k &= \sum_{m=1}^M \check{\mathbf{v}}_{mk}^H \check{\mathbf{y}}_m = \sum_{m=1}^M \check{\mathbf{v}}_{mk}^H (\tilde{\alpha} \mathbf{y}_m + \mathbf{e}_m^{y, B}) \\ &= \sum_{m=1}^M \check{\mathbf{v}}_{mk}^H \left(\tilde{\alpha} \sqrt{\rho} \sum_{k'=1}^K \mathbf{g}_{mk} \sqrt{q_{k'}} s_{k'} + \tilde{\alpha} \mathbf{n}_m + \mathbf{e}_m^{y, B} \right) \\ &= \underbrace{\tilde{\alpha} \sqrt{\rho q_k} \sum_{m=1}^M \check{\mathbf{v}}_{mk}^H \check{\mathbf{g}}_{mk} s_k}_{DS_k} + \underbrace{\tilde{\alpha} \sum_{k' \neq k} \sqrt{\rho q_{k'}} \sum_{m=1}^M \check{\mathbf{v}}_{mk}^H \check{\mathbf{g}}_{mk'} s_{k'}}_{IUI_{kk'}} \\ &\quad + \underbrace{\tilde{\alpha} \sum_{m=1}^M \check{\mathbf{v}}_{mk}^H \mathbf{n}_m}_{TN_k} - \underbrace{\tilde{\alpha} \sum_{k'=1}^K \sqrt{\rho} \sum_{m=1}^M \check{\mathbf{v}}_{mk}^H \sqrt{q_{k'}} \mathbf{e}_{mk'}^g s_{k'}}_{TQG_{kk'}} \\ &\quad + \underbrace{\sum_{m=1}^M \check{\mathbf{v}}_{mk}^H \mathbf{e}_m^{y, B}}_{TQY_k} + \underbrace{\tilde{\alpha} \sum_{k'=1}^K \sqrt{\rho} \sum_{m=1}^M \check{\mathbf{v}}_{mk}^H \sqrt{q_{k'}} \check{\mathbf{g}}_{mk'} s_{k'}}_{TEE_{kk'}}, \quad (7) \end{aligned}$$

where DS_k , $IUI_{kk'}$, and $TEE_{kk'}$ represent the desired signal, interuser interference, and total estimation error, respectively. Moreover, TN_k accounts for the total noise, and finally TQY_k and $TQG_{kk'}$ are total quantization errors due to quantizing the received signal \mathbf{y}_m and the estimated channel \mathbf{g}_{mk} , respectively. By using the capacity bound with side information provided in [17], we obtain the following achievable rate

$$R_k \approx \mathbb{E}_{\text{SSF}} \{ \log_2 (1 + \text{SINR}_k) \}, \quad (8)$$

where SINR_k is given by (9) shown at the top of the next page. Moreover, \mathbb{E}_{SSF} indicates that the expectation is taken over the SSF coefficients. With ZF, the decoder matrix is $\check{\mathbf{V}} = (\check{\mathbf{G}} \check{\mathbf{G}}^H)^{-1} \check{\mathbf{G}}$, where $\check{\mathbf{G}} = [\check{\mathbf{g}}_1, \dots, \check{\mathbf{g}}_K]$ which yields to $\sum_{m=1}^M \check{\mathbf{v}}_{mk}^H \check{\mathbf{g}}_{mk} = \sqrt{\rho q_k}$, and $\sum_{m=1}^M \check{\mathbf{v}}_{mk}^H \check{\mathbf{g}}_{mk'} = 0$, for $k \neq k'$. Therefore, SINR_k in (9) can be simplified as (10) (at the top of this page), where $\gamma_{mk} = \sqrt{\tau_p p_p} \beta_{mk} c_{mk}$.

IV. SUM RATE MAXIMIZATION

We aim to choose the transmit power $q_k, \forall k$, to maximize the sum rate as follows:

$$P_1 : \max_{q_k} \sum_{k=1}^K \mathbb{E}_{\text{SSF}} \{ \log_2 (1 + \text{SINR}_k) \} \quad (11a)$$

$$\text{s.t.} \quad 0 \leq q_k \leq p_{\max}^{(k)}, \forall k, \quad (11b)$$

$$\text{SINR}_k \geq \text{SINR}_k^{\text{Req}}, \forall k. \quad (11c)$$

where $p_{\max}^{(k)}$ is the maximum transmit power available at the k th user, and the constraints in (11c) refer to the throughput requirement constraints. Without loss of generality, Problem P_1 is equivalent to the following problem:

$$P_2 : \max_{q_k} \mathbb{E}_{\text{SSF}} \left\{ \prod_{k=1}^K (1 + \text{SINR}_k) \right\} \quad (12a)$$

$$\text{s.t.} \quad 0 \leq q_k \leq p_{\max}^{(k)}, \forall k \quad (12b)$$

$$\text{SINR}_k \geq \text{SINR}_k^{\text{Req}}, \forall k. \quad (12c)$$

A. Small-Scale-Fading-Based Power Control

To achieve the best performance, $q_k, \forall k$ should be optimally chosen for each realization inside the expectation. Thus we need to solve the following optimization problem:

$$P_3 : \max_{q_k} \prod_{k=1}^K (1 + \text{SINR}_k) \quad (13a)$$

$$\text{s.t.} \quad 0 \leq q_k \leq p_{\max}^{(k)}, \forall k \quad (13b)$$

$$\text{SINR}_k \geq \text{SINR}_k^{\text{Req}}, \forall k. \quad (13c)$$

Problem P_3 can be reformulated as follows:

$$P_4 : \min_{q_k} \prod_{k=1}^K (1 + \text{SINR}_k)^{-1} \quad (14a)$$

$$\text{s.t.} \quad 0 \leq q_k \leq p_{\max}^{(k)}, \forall k \quad (14b)$$

$$\text{SINR}_k \geq \text{SINR}_k^{\text{Req}}, \forall k. \quad (14c)$$

Problem P_4 is not convex, but it can be reformulated as a standard GP [18]. We re-write Problem P_4 as follows:

$$P_5 : \min_{q_k, t_k} \prod_{k=1}^K (1 + t_k)^{-1} \quad (15a)$$

$$\text{s.t.} \quad 0 \leq q_k \leq p_{\max}^{(k)}, \forall k, \quad (15b)$$

$$\text{SINR}_k \geq t_k, \forall k \quad (15c)$$

$$\text{SINR}_k \geq \text{SINR}_k^{\text{Req}}, \forall k, \quad (15d)$$

where $t_k, \forall k$ are the slack variables. Problem P_5 is a non-convex signomial problem. Moreover, all constraints in (15c) and (15d) can be reformulated into posynomial functions. As a result, if the objective function in (15a) is reformulated into a posynomial function, problem P_5 is a standard GP. Therefore, we present a heuristic solution to tackle the non-convexity

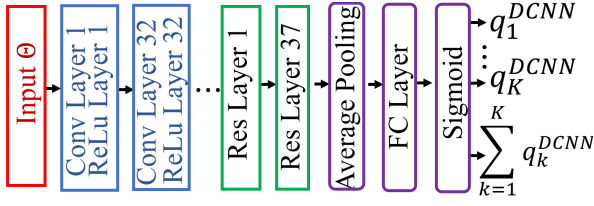


Figure 2. The proposed DL-based power scheme.

issue of Problem P_5 . To end this, we propose to reformulate Problem P_6 as follows:

$$P_6 : \min_{q_k, t_k} \prod_{k=1}^K t_k^{-1} \quad (16a)$$

$$\text{s.t. } 0 \leq q_k \leq p_{\max}^{(k)}, \forall k, \quad (16b)$$

$$\text{SINR}_k \geq t_k, \forall k \quad (16c)$$

$$\text{SINR}_k \geq \text{SINR}_k^{\text{Req}}, \forall k. \quad (16d)$$

Proposition 1. *Problem P_6 can be casted as a standard GP.*

Remark 1. *Since SINR_k is a function of the SSF coefficients, we refer the solution obtained by solving Problem P_6 as “the SSF-based power control”.*

B. Proposed Deep-Learning-Based Power Control

1) *Why Are Small-Scale-Fading-Based Power Control Schemes NOT Practical?:* In the practical systems, some users move very quickly, and hence, the channel coherence time may be only a few milliseconds [7]. Thus, it is not practical to design the power coefficients based on the SSF. For the SSF-based power control scheme in Section IV-A, the optimal transmit powers have to be recomputed on the SSF time scale. It is not practical to re-run the sum rate optimization problem every channel coherence time. The complexity of the sum rate optimization problem makes this approach infeasible. Therefore, we propose to use a deep learning scheme to control the power which needs to be re-run only after many coherence times. The authors in [6] define the spatial wide-sense stationary (WSS) property which is given by

$$Q_{\text{WSS}} = \frac{T_{\text{LT}}}{T_c}, \quad (17)$$

where T_{LT} refers to the long-term (LT) time, where the statistics of the channel may be considered constant within this interval, whereas T_c is the channel coherence time. The measurement results for an outdoor scenario at a center frequency of 2 GHz shows that $Q_{\text{WSS}} = 120$. As a result, the proposed DL-based power control needs to be run every $120T_c$, while the optimization problem in the SSF-based scheme needs to be solved at the beginning of each coherence time.

2) *Deep-Learning-Based Power Control Scheme:* We aim to determine a mapping from the LSF components and the optimal power obtained through solving Problem P_5 , i.e., \mathbf{q}^* . The proposed DL-based power control scheme is provided in Figure 2. The authors in [7] investigated a multi-cell massive MIMO. However, since the APs in cell-free massive MIMO are distributed, the neural network cannot learn a map between

the coefficients $\beta_{mk}, \forall m, k$ and the power elements obtained by the convex programming software CVX (obtained using the quantized version of the estimated channel). As depicted in Fig. 2, the architecture of the proposed neural network consists of five parts: convolution (Conv), residual (Res), average pooling, fully connected (FC) and sigmoid parts. The input of the network Θ_{IN} is a matrix of fixed size $K \times (K+1)$, which is defined in (18) (at the top of the next page). The input matrix is first passed through the convolution part which consists of a stack of 32 convolution layers. Each convolution layer is followed by the rectified linear unit (ReLU) layer. Each filter in the convolution layer has small receptive field of size 3×3 and its stride (i.e., step size of each filter) is fixed to 1 pixel. Furthermore, 1-pixel zero-padding is also carried out in each layer to preserve the spatial resolution after the convolution. Each convolution layer is followed by the ReLU activation layer. The ReLU function introduces non-linearity to the network which helps a variety of complex functions to be learned by training the CNN on a set of training data. For all convolution filters in other residual layers, the stride is set to 1 pixel and 1-pixel zero-padding is also carried. Finally, the output of the last FC layer is passed through the sigmoid part to bring the output values in the range $[0, 1]$. The output of the network \mathbf{q}_{DCNN} is a vector of size $K+1$.

For training the above CNN network, first a set containing a large number of training pairs $(\Theta_{\text{IN}}, \mathbf{q}^*)$ are collected, where \mathbf{q}^* is the solution of Problem P_6 . All channel matrices are then converted to dB which becomes the input of the network. The above CNN network is trained to minimize the following loss function:

$$L = \|\mathbf{q}^* - \mathbf{q}^{\text{DCNN}}\|^2. \quad (19)$$

This loss is averaged over the training data set and the aim of training is to minimize this loss.

V. USE-AND-THEN-FORGET BOUNDING TECHNIQUE

In this section, for the completeness, we summarize the UatF bounding technique [17], where the power control can be performed on the LSF time scale. From (7) and by using the UatF bounding technique, we can obtain the following achievable rate

$$R_k^{\text{UatF}} = \log_2(1 + \text{SINR}_k^{\text{UatF}}), \quad (20)$$

where $\text{SINR}_k^{\text{UatF}}$ is given in (21) (at the top of the next page). With ZF, we have $\text{DS}_k = \sqrt{\rho q_k}$, and $\text{Var}\{\text{DS}_k\} = 0$. Furthermore we have,

$$\text{IUI}_{kk'} = \rho q_{k'} \sum_{m=1}^M \check{\mathbf{v}}_{mk}^H \check{\mathbf{g}}_{mk'} = 0, \quad (23)$$

$$\mathbb{E}\{|\text{TEE}_{kk'}|^2\} = \rho \sum_{k'=1}^K q_{k'} \sum_{m=1}^M (\beta_{mk'} - \gamma_{mk'}) \mathbb{E}\{\|\check{\mathbf{v}}_{mk}\|^2\}, \quad (24)$$

Table I. Computational Complexity

$$\Theta_{\text{IN}} = \begin{bmatrix} \sum_{m=1}^M \left[\beta_{m1} \left(1 + \frac{\sigma_{\tilde{e},B}^2}{\tilde{a}^2} \right) - \gamma_{m1} (1 - \sigma_{\tilde{e}y}^2) \right] \mathbb{E} \{ \|\tilde{\mathbf{v}}_{m1}\|^2 \} & \dots & \sum_{m=1}^M \left[\beta_{mK} \left(1 + \frac{\sigma_{\tilde{e},B}^2}{\tilde{a}^2} \right) - \gamma_{mK} (1 - \sigma_{\tilde{e}y}^2) \right] \mathbb{E} \{ \|\tilde{\mathbf{v}}_{m1}\|^2 \} & \frac{1 + \frac{\sigma_{\tilde{e}y,B}^2}{\tilde{a}^2}}{\rho} \sum_{m=1}^M \mathbb{E} \{ \|\tilde{\mathbf{v}}_{m1}\|^2 \} \\ \sum_{m=1}^M \left[\beta_{m1} \left(1 + \frac{\sigma_{\tilde{e},B}^2}{\tilde{a}^2} \right) - \gamma_{m1} (1 - \sigma_{\tilde{e}y}^2) \right] \mathbb{E} \{ \|\tilde{\mathbf{v}}_{m2}\|^2 \} & \dots & \sum_{m=1}^M \left[\beta_{mK} \left(1 + \frac{\sigma_{\tilde{e},B}^2}{\tilde{a}^2} \right) - \gamma_{mK} (1 - \sigma_{\tilde{e}y}^2) \right] \mathbb{E} \{ \|\tilde{\mathbf{v}}_{m2}\|^2 \} & \frac{1 + \frac{\sigma_{\tilde{e}y,B}^2}{\tilde{a}^2}}{\rho} \sum_{m=1}^M \mathbb{E} \{ \|\tilde{\mathbf{v}}_{m2}\|^2 \} \\ \vdots & \ddots & \vdots & \vdots \\ \sum_{m=1}^M \left[\beta_{m1} \left(1 + \frac{\sigma_{\tilde{e},B}^2}{\tilde{a}^2} \right) - \gamma_{m1} (1 - \sigma_{\tilde{e}y}^2) \right] \mathbb{E} \{ \|\tilde{\mathbf{v}}_{mK}\|^2 \} & \dots & \sum_{m=1}^M \left[\beta_{mK} \left(1 + \frac{\sigma_{\tilde{e},B}^2}{\tilde{a}^2} \right) - \gamma_{mK} (1 - \sigma_{\tilde{e}y}^2) \right] \mathbb{E} \{ \|\tilde{\mathbf{v}}_{mK}\|^2 \} & \frac{1 + \frac{\sigma_{\tilde{e}y,B}^2}{\tilde{a}^2}}{\rho} \sum_{m=1}^M \mathbb{E} \{ \|\tilde{\mathbf{v}}_{mK}\|^2 \} \end{bmatrix}. \quad (18)$$

$$\text{SINR}_k^{\text{UatF}} = \frac{|\mathbb{E} \{ \text{DS}_k \}|^2}{\text{Var} \{ \text{DS}_k \} + \sum_{k' \neq k}^K \mathbb{E} \{ |\text{IUI}_{kk'}|^2 \} + \sum_{k'=1}^K \mathbb{E} \{ |\text{TEE}_{kk'}|^2 \} + \sum_{k'=1}^K \mathbb{E} \{ |\text{TQG}_{kk'}|^2 \} + \frac{1}{\tilde{a}^2} \mathbb{E} \{ |\text{TQY}_k|^2 \} + \mathbb{E} \{ |\text{TN}_k|^2 \}}. \quad (21)$$

$$\text{SINR}_k^{\text{UatF}} = \frac{\rho q_k}{\rho \sum_{k'=1}^K q_{k'} \sum_{m=1}^M \left[\beta_{mk'} \left(1 + \frac{\sigma_{\tilde{e},B}^2}{\tilde{a}^2} \right) - \gamma_{mk'} (1 - \sigma_{\tilde{e}y}^2) \right] \mathbb{E} \{ \|\tilde{\mathbf{v}}_{mk}\|^2 \} + \left(1 + \frac{\sigma_{\tilde{e}y,B}^2}{\tilde{a}^2} \right) \sum_{m=1}^M \mathbb{E} \{ \|\tilde{\mathbf{v}}_{mk}\|^2 \}}. \quad (22)$$

Power control schemes Beamforming Optimization		
SSF-based scheme	$\mathcal{O}((M+N)^3)$	$Q_{\text{WSS}} \times \mathcal{O}(2K^3)$
LSF-DL-based scheme	$\mathcal{O}((M+N)^3)$	$\mathcal{O}(2K^3)$
LSF-UatF-based scheme	$\mathcal{O}((M+N)^3)$	$\mathcal{O}(2K^3)$

$$\mathbb{E} \{ |\text{TN}_k|^2 \} = \mathbb{E} \left\{ \left| \sum_{m=1}^M \tilde{\mathbf{v}}_{mk}^H \mathbf{n}_m \right|^2 \right\} = \sum_{m=1}^M \mathbb{E} \{ \|\tilde{\mathbf{v}}_{mk}\|^2 \}, \quad (25)$$

$$\begin{aligned} \mathbb{E} \{ |\text{TQY}_k|^2 \} &= \mathbb{E} \left\{ \left| \sum_{m=1}^M \tilde{\mathbf{v}}_{mk}^H \mathbf{e}_m^{y,B} \right|^2 \right\} \\ &= \sum_{m=1}^M \mathbb{E} \{ \|\tilde{\mathbf{v}}_{mk}\|^2 \} \sigma_{\tilde{e}y,B}^2 \left(\rho \sum_{k'=1}^K \beta_{mk'} q_{k'} + 1 \right), \quad (26) \end{aligned}$$

and

$$\mathbb{E} \{ |\text{TQG}_{kk'}|^2 \} = \rho \sum_{k'=1}^K q_{k'} \sum_{m=1}^M \sigma_{\tilde{e}g}^2 \gamma_{mk'} \mathbb{E} \{ \|\tilde{\mathbf{v}}_{mk}\|^2 \}. \quad (27)$$

Therefore, from (20), the achievable rate of the k th user for ZF can be simplified as

$$R_k^{\text{UatF}} = \log_2 (1 + \text{SINR}_k^{\text{UatF}}), \quad (28)$$

where $\text{SINR}_k^{\text{UatF}}$ is given in (22) (defined at the top of the next page), where $\mathbb{E} \{ \|\tilde{\mathbf{v}}_{mk}\|^2 \}$ can be numerically calculated. Next, let us assume the length of the frame (which represents the length of the uplink data) is $\tau_f = \tau_c - \tau_p$, where τ_c denotes the number of samples for each coherence interval. Defining the number of quantization bits as α_m , where the index m denotes the m th AP. The required number of bits for each AP to quantize the estimated channel and the uplink data during each coherence interval is given by $R_{\text{th},m} = \frac{2N(K + \tau_f)\alpha_m}{T_c}$, where T_c (in sec.) refers to the coherence time.

Remark 2. We refer the schemes using “only the LSF coefficients” to solve the optimization problem as “the LSF-based power control scheme”.

VI. COMPLEXITY ANALYSIS

The ZF receiver is designed with the complexity of $\mathcal{O}((M+N)^3)$. In addition, a standard GP in Problem P_6 can be solved with complexity equivalent to $\mathcal{O}(2K^3)$ [18]. Note that the DL-based power control needs to be solved every $Q_{\text{WSS}} T_c$, while the optimization problem in the SSF-based scheme must run at the beginning of each coherence time. Hence, the complexity of solving the optimization problem is Q_{WSS} times more than the complexity of the proposed LSF-DL-based power control scheme. The number of arithmetic operations are provided in Table I.

VII. NUMERICAL RESULTS AND DISCUSSION

A. Simulation Parameters

We assume that M APs and K users are uniformly distributed in a $D \times D$ area. The channel coefficients between users and APs and the noise power are modeled in [1]. It is assumed that \bar{p}_p and $\bar{\rho}$ denote the power of the pilot sequence and the uplink data powers, respectively, where $p_p = \frac{\bar{p}_p}{p_n}$ and $\rho = \frac{\bar{\rho}}{p_n}$ are normalized transmit SNRs. Note that p_n refers to the noise power and is defined in [1]. In simulations, we set $\bar{p}_p = 100$ mW and $\bar{\rho} = 1$ W. To make sure all users can achieve a certain level of throughput, we have K SINR constraints as indicated in (11c). We set $\text{SINR}_k^{\text{Req}} = \text{SINR}_k^{\text{UatF}} (q_k = 1), \forall k$.

B. How to Generate the Training Set?

We generate 60,000 training sets. To run this simulation setup, we used a PC with Core(TM)i7 CPU @ 3.41 GHz with 64 GB Installed memory (RAM) for 4 days. Moreover, note that training a DCNN takes around 14 hours with this PC.

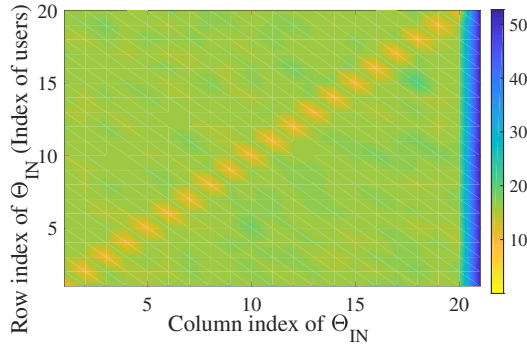


Figure 3. The pattern obtained by taking the average over the LSF coefficients (on a linear scale) for each element of Θ_{IN} for $M = 15$, $N = 6$, K

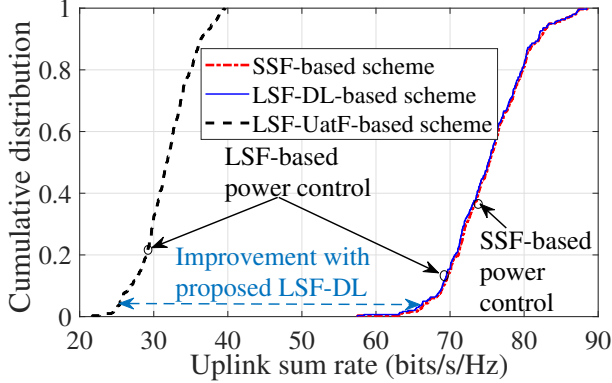


Figure 4. Cumulative distribution of sum rate performance of cell-free massive MIMO with the ZF receiver with $M = 15$, $K = 20$, $N = 6$, $\tau_p = 20$, $\tau_c = 200$, $\tau_f = \tau_c - \tau_p = 180$ and $T_c = 1$ ms. Note that we set $\alpha = 3$ which results in $R_{th,m} = \frac{2N(K+\tau_f)\alpha_m}{T_c} = 7.2$ Mbits/s.

C. Numerical Results

1) *Pattern in the Input of the Network*: We aim to have a closer look at the input of the network and investigate “what is the DCNN really doing?”. We investigate *The pattern* obtained by taking the average over the LSF coefficients for each element of the input matrices Θ_{IN} and Θ_{IN} . Fig. 3 reveals the pattern obtained by taking the average over the LSF coefficients for each element of the input matrix Θ_{IN} for for $\{M = 15, N = 6, K = 20\}$. The size of Θ_{IN} in both cases is 20×21 (obtained from (18)). As a result, the DCNN learns an unknown map between the input matrix and the optimal power elements obtained by CVX.

2) *CDF of the Achievable Sum Rate*: A cell-free massive MIMO system is considered with 15 APs ($M = 15$) where each AP is equipped with $N = 6$ antennas. Moreover, 20 users ($K = 20$) are uniformly distributed at random points over the simulation area of size 1×1 km². We also set $\alpha = 3$ quantization bits resulting in $R_{th,m} = \frac{2N(K+\tau_f)\alpha_m}{T_c} = 7.2$ Mbits/s. Fig. 4 presents the cumulative distribution of the achievable uplink sum rates for the proposed LSF-DL-based power scheme, the UatF bounding technique and the power elements are designed based on the quantized version of the estimated channel. As shown in Fig. 4, the performance of the proposed LSF-DL-based power scheme is significantly improved compared to the performance of the UatF bounding scheme. Note that in Fig. 4, the power elements in “SSF-

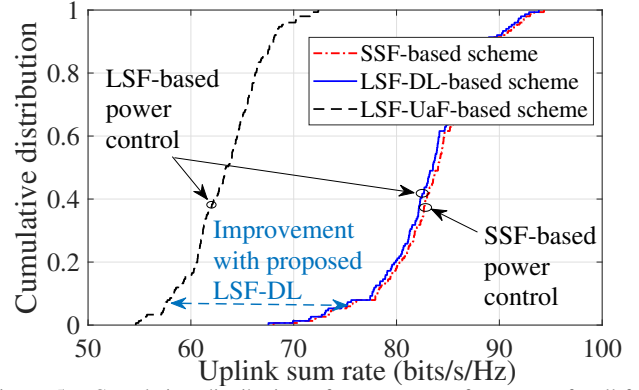


Figure 5. Cumulative distribution of sum rate performance of cell-free massive MIMO with the ZF receiver with $M = 50$, $K = 20$, $N = 2$, $\tau_p = 20$, $\tau_c = 200$, $\tau_f = \tau_c - \tau_p = 180$ and $T_c = 1$ ms. Note that we set $\alpha = 3$ which results in $R_{th,m} = 1.8$ Mbits/s.

based scheme” are designed based on the quantized version of the channel whereas in “LSF-DL-based scheme”, we need only the statistics of the channel to solve the optimization problems. Moreover, in “LSF-UatF-based scheme”, the CPU has access to the quantized version of the estimated channel to detect the data, however, it exploits only LSF coefficients to design the power elements. It is practically impossible to design the power elements based on the quantized version of the channel due to its high complexity. Interestingly, the sum rate performance of cell-free massive MIMO with the power elements obtained from the quantized version of the channel is almost as good as the performance of cell-free massive MIMO with the power elements obtained from the quantized version of the estimated channel -which reveals the beauty of DCNN.

Next, we investigate the performance of cell-free massive MIMO for $M = 50$ APs, each equipped with $N = 2$ antennas, and $K = 20$ users. Fig. 5 presents the cumulative distribution of sum rate performance of the cell-free massive MIMO system with three schemes, i.e., the UatF bounding technique, the proposed LSF-DL power control scheme and the scheme in which quantized version of the estimated channel are exploited to solve the sum rate maximization problem.

For further investigation, we evaluate the performance of the system for $M = 50$ APs, each equipped with $N = 2$ antennas, and $K_{serv} = 5$ users (only K_{serv} ($K_{serv} < K$) users are active and served in the area) while using the DCNN trained for $K = 20$ users, as assumed in Fig. 6. As the figure shows, the proposed scheme works very well even if we have fewer active users in the area.

3) *Training Curve*: Fig. 7 demonstrates the loss function for both training and test sets, which shows less than 0.02% loss (see (19)), confirming the accuracy of the proposed training scheme. Note that it is impossible to achieve the exact performance of the scheme with only the statistics of the channel (as the CPU exploits knowledge of the quantized version of the estimated channel to design the power elements).

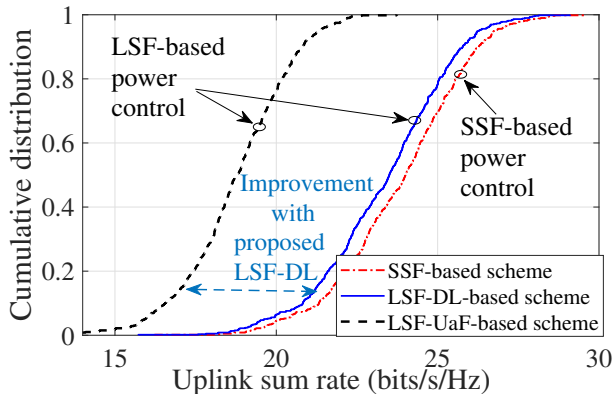


Figure 6. Cumulative distribution of sum rate performance of cell-free massive MIMO with the ZF receiver with $M = 50$, $K_{\text{serv}} = 5$, $N = 2$ and $\tau_p = 5$. We use the DCNN trained for the case of $K = 20$ users when we have only $K = 5$ active users in the area.

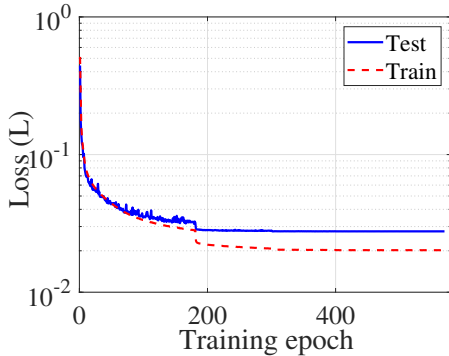


Figure 7. Loss function for both training and test sets.

VIII. CONCLUSIONS

We have developed a deep learning algorithm using a neural network for a cell-free massive MIMO to find a mapping between the LSF components and the power elements obtained from the SSF coefficients. We have proposed a sum rate optimization scheme with LSF-DL-based power control, which is practically feasible in cell-free massive MIMO due to its low complexity. The results showed less than 0.02% CNN loss function. We have studied the pattern in the input of the deep learning network. The numerical results showed that the proposed LSF-DL-based power control scheme significantly increases the achievable uplink sum rate of the cell-free massive MIMO system compared to the existing and practical LSF-UaF-based scheme.

REFERENCES

- [1] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.
- [2] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and H. Yang, "Cell-free massive MIMO systems," in *Proc. IEEE Asilomar*, Nov. 2015, pp. 695–699.
- [3] M. Bashar, H. Q. Ngo, A. Burr, D. Maryopi, K. Cumanan, and E. G. Larsson, "On the performance of backhaul constrained cell-free massive MIMO with linear receivers," in *Proc. IEEE Asilomar*, Nov. 2018, pp. 1–7.
- [4] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, E. G. Larsson, and P. Xiao, "Energy efficiency of the cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. Green Commun. and Net.*, pp. 1–18, Sep. 2019.
- [5] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

- [6] I. Viering, H. Hofstetter, and W. Utschick, "Spatial long-term variations in urban, rural and indoor environments," in *Proc. the 5th Meeting of COST273, Lisbon, Portugal*, Sep. 2002.
- [7] T. V. Chien, T. N. Canh, E. Björnson, and E. G. Larsson, "Power control in cellular massive MIMO with varying user activity: A deep learning solution," [online]. Available: <https://arxiv.org/pdf/1901.03620.pdf>, pp. 1–30, 2019.
- [8] C. D'Andrea, A. Zappone, S. Buzzi, and M. Debbah, "Uplink power control in cell-free massive MIMO via deep learning," [online]. Available: <https://arxiv.org/pdf/1908.11121.pdf>, pp. 1–5, 2019.
- [9] P. Zillmann, "Relationship between two distortion measures for memoryless nonlinear systems," *IEEE Signal Process. Lett.*, vol. 17, no. 11, pp. 917–920, Feb. 2010.
- [10] J. Xu, W. Xu, J. Zhu, D. W. K. Ng, and A. Lee Swindlehurst, "Secure massive MIMO communication with low-resolution DACs," *IEEE Trans. Wireless Commun.*, vol. 67, no. 5, pp. 3265–3278, May 2019.
- [11] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, "Cell-free massive MIMO with limited backhaul," in *Proc. IEEE ICC*, May 2018, pp. 1–7.
- [12] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, "Max-min SINR of cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. Commun.*, To appear.
- [13] A. G. Burr, M. Bashar, and D. Maryopi, "Cooperative access networks: Optimum fronthaul quantization in distributed massive MIMO and cloud RAN," in *Proc. IEEE VTC*, Jun. 2018, pp. 1–7.
- [14] D. Maryopi, M. Bashar, and A. Burr, "On the uplink throughput of zero-forcing in cell-free massive MIMO with coarse quantization," *IEEE Trans. Veh. Technol.*, pp. 1–5, Accepted.
- [15] J. Max, "Quantizing for minimum distortion," *IEEE Trans. Inf. Theory*, vol. 6, no. 1, pp. 7–12, Mar. 1960.
- [16] A. Kakkavas, J. Munir, A. Mezghani, H. Brunner, and J. A. Nossek, "Weighted sum rate maximization for multiuser MISO systems with low resolution digital to analog converters," in *Proc. IEEE WSA*, Mar. 2016.
- [17] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.
- [18] S. P. Boyd, S. J. Kim, A. Hassibi, and L. Vandenbarghe, "A tutorial on geometric programming," *Optim. Eng.*, vol. 8, no. 1, pp. 67–128, 2007.