



**QUEEN'S
UNIVERSITY
BELFAST**

Cell-Free Massive MIMO in the Short Blocklength Regime for URLLC

Nasir, A. A., Tuan, H. D., Ngo, H. Q., Duong, T. Q., & Poor, H. V. (2021). Cell-Free Massive MIMO in the Short Blocklength Regime for URLLC. *IEEE Transactions on Wireless Communications*. Advance online publication. <https://doi.org/10.1109/TWC.2021.3070836>

Published in:

IEEE Transactions on Wireless Communications

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright 2021, IEEE.

This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

Cell-Free Massive MIMO in the Short Blocklength Regime for URLLC

A. A. Nasir, H. D. Tuan, H. Q. Ngo, T. Q. Duong, and H. V. Poor

Abstract—This paper considers cell-free massive MIMO (cfm-MIMO) for the downlink ultra reliable and low-latency communication (URLLC). At the time of writing, cfm-MIMO has only been considered for communication in the long block regime (LBR), whose throughput is determined by the Shannon function rate with the interference treated as Gaussian noise. Conjugate beamforming (CB) is often used as it requires only local channel state information (CSI) for implementation but its design is based on a large-scale nonconvex problem, which is computationally intractable. The rate function in URLLC is much more complex than the Shannon function rate. The paper proposes a special class of CB, which admits a low-scale optimization formulation for computational tractability. Accordingly, a new path-following algorithm, which generates a sequence of better feasible points and converges at least to a locally optimal solution, is developed for optimizing URLLC rates and cfm-MIMO energy efficiency. Furthermore, the paper also develops improper Gaussian signaling to improve both Shannon function rate and URLLC rate.

Index Terms—Cell-free massive MIMO (cfm-MIMO), ultra reliable and low-latency communication (URLLC), conjugate beamforming (CB), energy efficiency, nonconvex optimization.

I. INTRODUCTION

Ultra reliable and low-latency communication (URLLC) has attracted a lot of recent research attention (see e.g. [1], [2] and references therein) due to its promise in enabling innovative Internet-of-Thing (IoT) applications such as tactile internet, haptic feedback, vehicle network, autonomous systems, etc (see e.g. [2], [3] and references therein). Information theoretic analysis in URLLC is based on the URLLC rate function, which is a complicated function of the transmission power, the blocklength and the decoding error probability [4]. As such, the design of resource allocation to optimize URLLC rate performance poses very computationally challenging problems

The work was supported in part by the KFUPM Research Project #SB191006, in part by the Australian Research Council's Discovery Projects under Grant DP190102501, in part by the U.K. Research and Innovation Future Leaders Fellowships under Grant MR/S017666/1, in part by the Royal Academy of Engineering (RAEng) under the RAEng Research Chair and Senior Research Fellowship scheme Grant RCSRF2021\11\41, and in part by the U.S. National Science Foundation under Grant CCF-1908308

Ali A. Nasir is with the Department of Electrical Engineering, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia (email: anasir@kfupm.edu.sa).

Hoang D. Tuan is with the School of Electrical and Data Engineering, University of Technology Sydney, Broadway, NSW 2007, Australia (email: Tuan.Hoang@uts.edu.au).

Hien Q. Ngo and Trung Q. Duong are with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast BT3 9DT, U.K. (e-mail: hien.ngo@qub.ac.uk; trung.q.duong@qub.ac.uk).

H. Vincent Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA (email: poor@princeton.edu).

[5]–[9]. For massive access beyond 5G (B5G), cell-free massive MIMO (cfm-MIMO) [10]–[12], under which a massive number of single-antenna access points (APs) are distributed across the area to serve multiple single-antenna users, has attracted a great attention.

Zero-forcing beamforming (ZFB) and conjugate beamforming (CB) are the most popular classes of massive MIMO information delivery [13]. The authors of [14] employed ZFB for cfm-MIMO while only considering the scenario of no pilot contamination that requires the orthogonality of pilot sequences. Its optimal power allocation is easy but its implementation may not be practical, requiring that all APs have global channel state information (CSI) of the overall system. Besides, there is no analytical expression for the mathematical expectation of nonlinear terms involving inverse of random matrices [14, eq. (15)]. On the other hand, the authors of [10] used CB, which requires only local CSI for each AP for implementation and works under a general scenario of pilot contamination. Besides, its effective signal-to-interference-plus-noise (SINR) is also asymptotically better than that by ZFB [13, Table 3.1]. However, its computational complexity is very high because there is a large number of decision variables that makes it impractical for massive connectivity. Unlike ZFB, CB cannot suppress the so called multi-user interference (MUI), so proper Gaussian signaling (PGS), which relies on circularly symmetric Gaussian signals, is not expected as optimal in these works. Recent studies such as [15]–[20] and references therein have showed that improper Gaussian signaling (IGS) [21], which releases the signal circular symmetry, can improve information throughput essentially for severely interfered networks such as those with the number of transmit antennas less than the number of served users.

While proper Gaussian signal is fully characterized by its covariance, improper Gaussian signal in contrast is fully characterized by its so called augmented covariance, which involves not only the covariance but also the pseudo-covariance. The Shannon function rate under IGS is a log-determinant $\log \det(\cdot)$ of signal augmented covariance matrices instead of logarithm function of signal power in multi-input single-output (MISO) such as cfm-MIMO, so the design of beamforming under IGS is much more computationally challenging than that under PGS. The merits of IGS over PGS have been shown in the literature for improving Shannon function rate [15]–[20] but has not been known for URLLC. The URLLC rate expression is a very complicated function of the transmission power coefficients [4], making the design of CB under the assumption of IGS for URLLC is much more computationally

challenging.

Against the above background, in this paper we opt for a particular class of CB, which takes a similarly low computational complexity of ZFB and the same simple access requirement of CB to optimize the downlink URLLC rates of a cfm-MIMO and its energy efficiency under both PGS and IGS. The paper's contribution is three-fold:

- A new class of CB with low computational complexity, which enables massive connectivity by using local CSI.
- The paper is the first work to consider cfm-MIMO for multi-user URLLC. We formulate two important design problems of (i) max-min rate optimization and (ii) energy efficiency optimization under transmit power constraint. Due to the complexity of rate expression in the short blocklength regime, the aforementioned problems are quite challenging to solve. Accordingly, new path-following algorithms, which generate a sequence of better feasible points and converge at least to a locally optimal solution, are developed for optimizing URLLC rates and cfm-MIMO energy efficiency.
- This is the first work to propose IGS to improve both the URLLC rate and energy efficiency of cfm-MIMO. The performance of proposed algorithms is analyzed through rigorous simulations under varying values of APs, number of users, transmit power budget, URLLC transmission duration, and the length of pilot sequences. Particularly, IGS is shown to outperform PGS in low-power regime that is very important for B5G.

The paper is organized as follows. Section II is devoted to PGS for cfm-MIMO in serving URLLC, where the basic modeling and communication protocol are recalled, a new class of CB is introduced, the problem of max-min users' URLLC rate is accordingly formulated and then new path-following algorithm is developed for its computation. The problem of maximizing cfm-MIMO efficiency under the quality of service (QoS) constraints in terms of users' URLLC rate thresholds is also addressed. Section III introduces IGS for the downlink cfm-MIMO to address the problem of max-min users' URLLC rate optimization and of maximizing cfm-MIMO energy-efficiency under IGS. Section IV provides simulations to confirm the information theoretic results of Sections II and III. Section VI concludes the paper. The appendix provides ingredient inequalities used in the derivations of the theoretical results.

Notation: Bold-faced lower-case letters, e.g., \mathbf{x} , are used for vectors, while bold-face upper letters, e.g., \mathbf{X} , are used for matrices. \mathbf{x}^H , \mathbf{x}^T , and \mathbf{x}^* denote Hermitian transpose, normal transpose, and conjugate of the vector \mathbf{x} , respectively. $\|\cdot\|$ stands for the vector's Euclidean norm and $|\cdot|$ stands for the absolute value of a complex/real scalar number. $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^H \mathbf{y}$ for the vectors \mathbf{x} and \mathbf{y} and $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{trace}(\mathbf{X}^H \mathbf{Y})$ for the matrices \mathbf{X} and \mathbf{Y} . We also use $\langle \mathbf{X} \rangle = \text{trace}(\mathbf{X})$ for the matrix \mathbf{X} . Also $\mathbf{X} \succeq 0$ ($\mathbf{X} \succ 0$, resp.) means the Hermitian symmetric matrix \mathbf{X} is positive semi-definite (positive definite, resp.). \mathbb{C} and \mathbb{R} , and \mathbb{R}_+ denote the set of all complex, real numbers, and positive numbers, respectively. $\Re\{x\}$ denotes the real part of a complex number x , and \mathbf{I}_n is the identity matrix of size

$n \times n$.

II. SYSTEM MODEL AND PROPER GAUSSIAN SIGNALING

Consider a cfm-MIMO system, which consists of a massive number M of single-antenna access points (APs) indexed by $m \in \mathcal{M} \triangleq \{1, \dots, M\}$ serving N single-antenna users (UEs) indexed by $n \in \mathcal{N} \triangleq \{1, \dots, N\}$. Like remote radio heads in a cloud radio access network, these APs, which are linked to a central processing unit (CPU) through backhalls, are distributed across the area to make them closer to UEs. The channel h_{mn} between AP m and UE n is modelled as

$$h_{mn} = \sqrt{\beta_{mn}} g_{mn}, \quad (1)$$

where β_{mn} represents the large-scaling fading and $g_{mn} \in \mathcal{CN}(0, 1)$ represents the small scale fading. The standard block fading model is used, hence the channel h_{mn} is constant in time-frequency blocks of t channel uses, where t is the length of the coherence interval in samples defined by the product of the coherence time and bandwidth. In this work, we assume perfect channel knowledge hence our results will depict the upper bound on the achievable throughput under LBR or URLLC. In practice, the channel knowledge at the APs can be acquired via uplink training [10]. While each user can estimate the channels through a simple and low-overhead beamforming training scheme [22]. In this work, by assuming perfect channel knowledge, we keep our focus in considering cfm-MIMO for multi-user URLLC and proposing a new class of CB with low computational complexity.

The authors in [10] proposed the following CB

$$x_m = \sum_{n' \in \mathcal{N}} \sqrt{p_{mn'}} h_{mn'}^* s_{n'},$$

where $p_{mn'}$ expresses the power allocation at the AP m for the symbol $s_{n'} \in \mathcal{CN}(0, 1)$ intended for UE n' , which is normalized as $\mathbb{E}(|s_{n'}|^2) = 1$. The design of CB thus involves MN decision variables $p_{mn'}$, which means a large-scale nonconvex problems, which is computationally intractable. For instance, for $M = 128$ and $N = 32$, it involves already 4096 decision variables. As the signal fades with the distance due to path-loss, the computational complexity can be reduced by choosing only those APs that are closer to the UE n , as serving APs for the UE n . However, the complexity will still be enormous because if, say, 20 nearby APs are selected to serve each UE, there will still be 640 decision variables.

To circumvent this computational issue, we propose the following CB at AP m ,

$$x_m = \sum_{n' \in \mathcal{N}} \sqrt{p_{n'}} h_{mn'}^* s_{n'}, \quad (2)$$

where $p_{n'}$ expresses the power allocation across all the APs for the symbol $s_{n'}$ intended for UE n' . The design of this CB involves only N decision variables instead of MN of those. Nevertheless, to avoid unnecessary interference to the users, each AP is chosen to serve users in its close vicinity only. Thus, by following [23], only the APs that are closer to the UE n (within certain radius), are selected as the serving APs for the UE n .

The total transmit power by the APs is given by

$$\begin{aligned}\mathbb{E}(|x_m|^2) &= \sum_{n \in \mathcal{N}} \|h_{mn}\|^2 p_n \\ &\triangleq \pi_m(\mathbf{p}),\end{aligned}\quad (3)$$

which is a linear function in $\mathbf{p} = (p_1, p_2, \dots, p_N)^T \in \mathbb{R}_+^N$.

The signal received at UE n is

$$\begin{aligned}y_n &= \sum_{m \in \mathcal{M}} h_{mn} \sum_{n' \in \mathcal{N}} \sqrt{p_{n'}} h_{mn'}^* s_{n'} + \nu_n \\ &= \sqrt{p_n} \lambda_{nn} s_n + \sum_{n' \in \mathcal{N} \setminus \{n\}} \sqrt{p_{n'}} \lambda_{nn'} s_{n'} + \nu_n,\end{aligned}\quad (4)$$

where $\nu_n \in \mathcal{CN}(0, \sigma^2)$ is the noise background at UE n , and

$$\lambda_{nn'} \triangleq \sum_{m \in \mathcal{M}} h_{mn} h_{mn'}^*, \quad (n, n') \in \mathcal{N} \times \mathcal{N}.\quad (5)$$

Given $\lambda_{nn'}$, the interference-plus-noise term in (4), $\sum_{n' \in \mathcal{N} \setminus \{n\}} \sqrt{p_{n'}} \lambda_{nn'} s_{n'} + \nu_n$, is Gaussian. In addition, it can be easily shown that the above interference-plus-noise term is uncorrelated with the desired signal part, $\sqrt{p_n} \lambda_{nn} s_n$. Therefore, the corresponding effective SINR at UE n is

$$\gamma_n(\mathbf{p}) = \frac{\lambda_n p_n}{\varphi_n(\mathbf{p}) + \sigma^2}\quad (6)$$

with $\lambda_n \triangleq \lambda_{nn}^2$ and $\varphi_n(\mathbf{p}) \triangleq \sum_{n' \in \mathcal{N} \setminus \{n\}} p_{n'} |\lambda_{nn'}|^2$, which is a positive linear function.

In the long block regime (LBR), the achievable rate in nats/sec/Hz for decoding the signal $\sqrt{p_n} \lambda_{nn} s_n$ in (4) by treating other terms there as Gaussian noise is the following Shannon rate function

$$\tilde{r}_n(\mathbf{p}) = \ln(1 + \gamma_n(\mathbf{p})).\quad (7)$$

Let \mathcal{B} is the communication bandwidth. According to [24], the achievable URLLC rate in nats/sec/Hz of decoding the signal $\sqrt{p_n} \lambda_{nn} s_n$ in (4) by treating other terms there as Gaussian noise is approximated by

$$r_n(\mathbf{p}) \triangleq \tilde{r}_n(\mathbf{p}) - \frac{1}{\sqrt{\mathcal{B}t_t}} Q_G^{-1}(\epsilon^c) \delta_n(\mathbf{p}),\quad (8)$$

with

$$\delta_n(\mathbf{p}) \triangleq \left[2 \frac{\gamma_n(\mathbf{p})}{1 + \gamma_n(\mathbf{p})} \right]^{1/2},\quad (9)$$

where the factor

$$V_n \triangleq 2 \frac{\gamma_n(\mathbf{p})}{1 + \gamma_n(\mathbf{p})}\quad (10)$$

is the channel dispersion [24, eq. (27)] under the SINR $\gamma_n(\mathbf{p})$, t_t is the URLLC transmission duration, $Q_G^{-1}(\cdot)$ is the inverse of the Gaussian Q-function $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt$, and ϵ^c is defined as an acceptable decoding error probability, which implies that under the considered block fading channel model, one out of $1/\epsilon^c$ short blocklength packets (URLLC transmissions) may experience outage [1].

It is noteworthy that the channel dispersion of spherical codebook is close to the that of the Gaussian codebook, when the number of interferers is large [24, Fig. 3]. This justifies our assumption of using Gaussian codebook instead of achievable

spherical codebook as our system model considers many UEs (and hence, the number of interferers is large).

Here we would like to make a note that many URLLC transmissions are expected to be implemented during data transmission because the duration of each URLLC transmission is usually much shorter than the coherence interval t_d for payload data transmission.

A. Max-min rate based efficient payload data transmission

Having the expression (8) for the user URLLC rate, the problem of max-min users' URLLC rate optimization subject to transmission power constraint can be formulated as follows:

$$\max_{\mathbf{p} \in \mathbb{R}_+^N} f(\mathbf{p}) \triangleq \min_{n \in \mathcal{N}} r_n(\mathbf{p})\quad (11a)$$

$$\text{s.t. } \pi_m(\mathbf{p}) \leq P, m \in \mathcal{M},\quad (11b)$$

where P is a given power budget. This optimization problem is non-convex as each URLLC rate function $r_n(\mathbf{p})$ is nonconcave so their minimum, which is the objective function in (11a) is non-concave. In fact, $r_n(\mathbf{p})$ defined by (8) is a difference of two functions, which are non-concave and non-convex. To propose a path-following algorithm, which generates a sequence of better feasible points for (11) and converges at least to its locally optimal solution, we need to develop a lower bounding concave function approximation for $r_n(\mathbf{p})$.

Let $\mathbf{p}^{(\kappa)} \triangleq (p_1^{(\kappa)}, p_2^{(\kappa)}, \dots, p_N^{(\kappa)})^T$ be the feasible point for (11) that is found from the $(\kappa - 1)$ -th iteration.

By applying the inequality (59) in the Appendix for $(x, y) = (\lambda_n p_n, \varphi_n(\mathbf{p}) + \sigma^2)$ and $(\bar{x}, \bar{y}) = (\lambda_n p_n^{(\kappa)}, \varphi_n(\mathbf{p}^{(\kappa)}) + \sigma^2)$ we obtain

$$\begin{aligned}\tilde{r}_n(\mathbf{p}) &\geq \tilde{r}_n(\mathbf{p}^{(\kappa)}) + \frac{\lambda_n p_n^{(\kappa)}}{\varphi_n(\mathbf{p}^{(\kappa)}) + \sigma^2} \\ &\quad \times \left(2 \frac{\sqrt{p_n}}{\sqrt{p_n^{(\kappa)}}} - \frac{\lambda_n p_n + \varphi_n(\mathbf{p}) + \sigma^2}{\lambda_n p_n^{(\kappa)} + \varphi_n(\mathbf{p}^{(\kappa)}) + \sigma^2} - 1 \right) \\ &\triangleq \tilde{r}_n^{(\kappa)}(\mathbf{p}).\end{aligned}\quad (12)$$

The function $\tilde{r}_n^{(\kappa)}(\mathbf{p})$ is concave.

Regarding the function $\delta_n(\mathbf{p})$ in (8), by applying the inequality (61) in the Appendix for $x = 2\gamma_n(\mathbf{p})/(1 + \gamma_n(\mathbf{p}))$ and $\bar{x} = 2\gamma_n(\mathbf{p}^{(\kappa)})/(1 + \gamma_n(\mathbf{p}^{(\kappa)}))$, we obtain

$$\begin{aligned}\delta_n(\mathbf{p}) &\leq \frac{1}{2} \left(\delta_n(\mathbf{p}^{(\kappa)}) + \frac{2}{\delta_n(\mathbf{p}^{(\kappa)})} \frac{\gamma_n(\mathbf{p})}{1 + \gamma_n(\mathbf{p})} \right) \\ &= \frac{\delta_n(\mathbf{p}^{(\kappa)})}{2} + \frac{1}{\delta_n(\mathbf{p}^{(\kappa)})} \frac{\lambda_n p_n}{\lambda_n p_n + \varphi_n(\mathbf{p}) + \sigma^2}\end{aligned}\quad (13)$$

Since the RHS of (13) is still non-convex, we approximate the function $\frac{\lambda_n p_n}{\lambda_n p_n + \varphi_n(\mathbf{p}) + \sigma^2}$ as follows:

$$\begin{aligned}\frac{\lambda_n p_n}{\lambda_n p_n + \varphi_n(\mathbf{p}) + \sigma^2} &\leq \frac{0.5 \lambda_n p_n^2 / p_n^{(\kappa)} + 0.5 \lambda_n p_n^{(\kappa)}}{\lambda_n p_n + \varphi_n(\mathbf{p}) + \sigma^2} \\ &= \frac{\lambda_n p_n^2}{2 p_n^{(\kappa)} (\lambda_n p_n + \varphi_n(\mathbf{p}) + \sigma^2)} \\ &\quad + \frac{0.5 \lambda_n p_n^{(\kappa)}}{\lambda_n p_n + \varphi_n(\mathbf{p}) + \sigma^2}\end{aligned}\quad (15)$$

where (14) is obtained by applying the inequality (62) in the Appendix to the numerator of the second factor in (13) for $x = \lambda_n p_n$ and $\bar{x} = \lambda_n p_n^{(\kappa)}$.

Define the function

$$\delta_n^{(\kappa)}(\mathbf{p}) \triangleq \frac{\delta_n(\mathbf{p}^{(\kappa)})}{2} + \frac{1}{\delta_n(\mathbf{p}^{(\kappa)})} \times \text{RHS of (15)}, \quad (16)$$

which is an upper bounding convex function approximation of the function $\delta_n(\mathbf{p})$:

$$\delta_n(\mathbf{p}) \leq \delta_n^{(\kappa)}(\mathbf{p}). \quad (17)$$

It then follows from (12) and (17) that the function

$$r_n^{(\kappa)}(\mathbf{p}) \triangleq \tilde{r}_n^{(\kappa)}(\mathbf{p}) - \frac{1}{\sqrt{\beta t_t}} Q_G^{-1}(\epsilon^c) \delta_n^{(\kappa)}(\mathbf{p}) \quad (18)$$

is a lower bound concave function approximation of the URLLC rate function $r_n(\mathbf{p})$ defined from (8):

$$r_n(\mathbf{p}) \geq r_n^{(\kappa)}(\mathbf{p}). \quad (19)$$

Also $r_n^{(\kappa)}(\mathbf{p})$ matches with $r_n(\mathbf{p})$ at $p^{(\kappa)}$: $r_n(\mathbf{p}^{(\kappa)}) = r_n^{(\kappa)}(\mathbf{p}^{(\kappa)})$.

Therefore, the function

$$f^{(\kappa)}(\mathbf{p}) \triangleq \min_{n \in \mathcal{N}} r_n^{(\kappa)}(\mathbf{p}) \quad (20)$$

is concave as the minimum of concave functions $r_n^{(\kappa)}(\mathbf{p})$, $n \in \mathcal{N}$ [25] and is a lower bounding concave function approximation of the objective function $f(\mathbf{p})$ defined from (11a):

$$f^{(\kappa)}(\mathbf{p}) \leq f(\mathbf{p}). \quad (21)$$

At the κ -th iteration, we solve the following convex problem of the computational complexity $\mathcal{O}(N^3 M)$ [26, p. 4], to generate the next iterative point $\mathbf{p}^{(\kappa+1)}$:

$$\max_{\mathbf{p} \in \mathbb{R}_+^N} f^{(\kappa)}(\mathbf{p}) \quad \text{s.t.} \quad (11b). \quad (22)$$

Note that $f^{(\kappa)}(\mathbf{p}^{(\kappa+1)}) > f^{(\kappa)}(\mathbf{p}^{(\kappa)})$ as far as $\mathbf{p}^{(\kappa+1)} \neq \mathbf{p}^{(\kappa)}$ because the former is the optimal solution of (22) while the latter is a feasible point for (22). Therefore

$$f(\mathbf{p}^{(\kappa+1)}) \geq f^{(\kappa)}(\mathbf{p}^{(\kappa+1)}) > f^{(\kappa)}(\mathbf{p}^{(\kappa)}) = f(\mathbf{p}^{(\kappa)}), \quad (23)$$

i.e. $\mathbf{p}^{(\kappa+1)}$ is a better feasible point than $p^{(\kappa)}$ for (11). Such sequence $\{\mathbf{p}^{(\kappa)}\}$ of improved feasible points for (11) converges at least to its locally optimal solution [27].

It is important to have a good initial point $p^{(0)}$ with the positive URLLC rate, so we take it as the optimal solution of the following problem of max-min users' Shannon function rate optimization

$$\max_{\mathbf{p} \in \mathbb{R}_+^N} \min_{n \in \mathcal{N}} \tilde{r}_n(\mathbf{p}) \quad \text{s.t.} \quad (11b), \quad (24)$$

which is computed by iterating the following convex problem

$$\max_{\mathbf{p} \in \mathbb{R}_+^N} \min_{n \in \mathcal{N}} \tilde{r}_n^{(\kappa)}(\mathbf{p}) \quad \text{s.t.} \quad (11b) \quad (25)$$

until the convergence of the objective function in (24). The initial point of this iterative process is any feasible point for the convex constraint (11b).

Algorithm 1 provides the pseudo-code for the proposed path-following procedure.

Algorithm 1 Path-following algorithm for solving the problem (11)

- 1: **Initialization:** Iterate the convex problem (25) until the convergence to obtain an initial point $\mathbf{p}^{(0)}$. Set $\kappa = 0$.
 - 2: **Repeat until convergence of the objective function in (11):** Solve the convex problem (22) to generate $\mathbf{p}^{(\kappa+1)}$. Reset $\kappa \leftarrow \kappa + 1$.
-

B. Energy efficient payload data transmission

The problem of maximizing quality-of-service (QoS)-aware energy efficiency [28], [29] can be formulated as

$$\max_{\mathbf{p} \in \mathbb{R}_+^N} \frac{\sum_{n \in \mathcal{N}} r_n(\mathbf{p})}{\pi_{te}(\mathbf{p})} \quad \text{s.t.} \quad (11b), \quad (26a)$$

$$r_n(\mathbf{p}) \geq \bar{r}, \quad (26b)$$

where \bar{r} is a given threshold for URLLC users' rates to set the QoS, and [30], [31]

$$\pi_{te}(\mathbf{p}) = \alpha \sum_{m \in \mathcal{M}} \pi_m(\mathbf{p}) + P_{sc} + M \left(P_0 + P_{bt} \sum_{n \in \mathcal{N}} r_n(\mathbf{p}) \right) \quad (27)$$

which is the total power consumption. Here, α is the reciprocal of drain efficiency of the the power amplifier at APs, P_{sc} is the internal power for running the circuit components, P_0 is the power consumption of each backhaul, and P_{bt} is the traffic-dependent power (in Watt per bits/s). The first and second terms in (27) recap the transmission power consumption, while the third term recaps the power consumption of the backhauls [32].

Like [30], observe that

$$\begin{aligned} \frac{\sum_{n \in \mathcal{N}} r_n(\mathbf{p})}{\pi_{te}(\mathbf{p})} &= \left(\frac{\alpha \sum_{m \in \mathcal{M}} \pi_m(\mathbf{p}) + P_{sc} + MP_0}{\sum_{n \in \mathcal{N}} r_n(\mathbf{p})} + MP_{bt} \right)^{-1} \\ &= \left(\frac{\pi(\mathbf{p})}{\sum_{n \in \mathcal{N}} r_n(\mathbf{p})} + MP_{bt} \right)^{-1} \end{aligned} \quad (28)$$

for

$$\pi(\mathbf{p}) \triangleq \alpha \sum_{m \in \mathcal{M}} \pi_m(\mathbf{p}) + \nu, \quad (29)$$

which is an affine function, and

$$\nu \triangleq P_{sc} + MP_0, \quad (30)$$

which is a positive constant.

It can be observed from the RHS of (28) that the factor MP_{bt} in $\left(\frac{\pi(\mathbf{p})}{\sum_{n \in \mathcal{N}} r_n(\mathbf{p})} + MP_{bt} \right)$ is an independent additive factor and is not a function of the optimization variable \mathbf{p} . Therefore, maximizing the objective function in (26) is equivalent to the maximization of the objective function $\frac{\sum_{n \in \mathcal{N}} r_n(\mathbf{p})}{\pi(\mathbf{p})}$. The problem (26) is thus equivalent to the following optimization problem:

$$\max_{\mathbf{p} \in \mathbb{R}_+^N} \frac{\sum_{n \in \mathcal{N}} r_n(\mathbf{p})}{\pi(\mathbf{p})} \quad \text{s.t.} \quad (11b), (26b). \quad (31)$$

Let $\mathbf{p}^{(\kappa)} \triangleq (p_1^{(\kappa)}, p_2^{(\kappa)}, \dots, p_N^{(\kappa)})^T$ be the feasible point for (31) that is found from the $(\kappa - 1)$ -th iteration and

$$\gamma^{(\kappa)} = \frac{\sum_{n \in \mathcal{N}} r_n(\mathbf{p}^{(\kappa)})}{\pi(\mathbf{p}^{(\kappa)})}. \quad (32)$$

Algorithm 2 Path-following algorithm for solving the problem (26)

- 1: **Initialization:** Iterate the convex problem (22) until its optimal value is larger or equal to \bar{r} for obtaining an initial point $\mathbf{p}^{(0)}$. Set $\kappa = 0$.
- 2: **Repeat until convergence of $\gamma^{(\kappa)}$ defined by (32):** Solve the convex problem (33) to generate $\mathbf{p}^{(\kappa+1)}$ and $\gamma^{(\kappa+1)}$ (according to (35)). Reset $\kappa \leftarrow \kappa + 1$.
- 3: **Output $\mathbf{p}^{(\kappa)}$** as the optimal solution of (26) and

$$\left(\frac{1}{\gamma^{(\kappa)}} + MP_{bt} \right)^{-1} = \frac{\gamma^{(\kappa)}}{1 + MP_{bt}\gamma^{(\kappa)}}$$

as the optimal value of (26).

Recalling the function $r_n^{(\kappa)}(\mathbf{p})$ from (18), at the κ th iteration we solve the following convex problem of the computational complexity $\mathcal{O}(N^3(M+N))$ [26, p. 4], to generate the next iterative point $\mathbf{p}^{(\kappa+1)}$:

$$\max_{\mathbf{p} \in \mathbb{R}_+^N} \left[\sum_{n \in \mathcal{N}} r_n^{(\kappa)}(\mathbf{p}) - \gamma^{(\kappa)} \pi(\mathbf{p}) \right] \quad \text{s.t.} \quad (11b), r_n^{(\kappa)}(\mathbf{p}) \geq \bar{r}. \quad (33)$$

Note that

$$\begin{aligned} \sum_{n \in \mathcal{N}} r_n^{(\kappa)}(\mathbf{p}^{(\kappa+1)}) - \gamma^{(\kappa)} \pi(\mathbf{p}^{(\kappa+1)}) \\ > \sum_{n \in \mathcal{N}} r_n^{(\kappa)}(\mathbf{p}^{(\kappa)}) - \gamma^{(\kappa)} \pi(\mathbf{p}^{(\kappa)}) \\ = 0 \end{aligned} \quad (34)$$

as far as $\mathbf{p}^{(\kappa+1)} \neq \mathbf{p}^{(\kappa)}$ because the former is the optimal solution of (33) while the latter is its feasible point. Therefore

$$\gamma^{(\kappa+1)} \triangleq \frac{\sum_{n \in \mathcal{N}} r_n(\mathbf{p}^{(\kappa+1)})}{\pi(\mathbf{p}^{(\kappa+1)})} \quad (35)$$

$$> \gamma^{(\kappa)}, \quad (36)$$

i.e. $\mathbf{p}^{(\kappa+1)}$ is a better feasible point than $\mathbf{p}^{(\kappa)}$ for (31). Consequently, the sequence $\{\mathbf{p}^{(\kappa)}\}$ of improved feasible points for (31) converges at least to its locally optimal solution [33].

To find a feasible point $\mathbf{p}^{(0)}$ for (31) we iterate the convex problem (22) until its optimal value is large or equal to \bar{r} .

Algorithm 2 provides the pseudo-code for the proposed path-following procedure.

III. IMPROPER GAUSSIAN SIGNALING

Until now, the information source $s_{n'}$ in (2) is assumed to be proper Gaussian in $\mathcal{CN}(0,1)$. It has been recently shown that (see e.g. [15]–[19] and references therein) the release of Gaussian properness improves the signal degree of freedom, which helps to improve the Shannon function rate performances of interfering networks. This section aims to use improper Gaussian sources to improve the max-min users's URLLC rate as well as the energy efficiency of the considered

cfm-MIMO system. A normalized improper Gaussian source $s_{n'}$ is fully characterized by its augmented covariance [21]

$$C_{n'} \triangleq \begin{bmatrix} \mathbb{E}(|s_{n'}|^2) & \mathbb{E}(s_{n'}^2) \\ (\mathbb{E}(s_{n'}^2))^* & \mathbb{E}(|s_{n'}|^2) \end{bmatrix} \quad (37)$$

$$= \begin{bmatrix} 1 & \tilde{q}_{n'} \\ (\tilde{q}_{n'})^* & 1 \end{bmatrix}, \quad (38)$$

where $\tilde{q}_{n'} \triangleq \mathbb{E}(s_{n'}^2)$ is called its pseudo covariance. Thus, a Gaussian source is proper if and only if its pseudo covariance is zero. Note that $|\tilde{q}_{n'}| = |\mathbb{E}(s_{n'}^2)| \leq \mathbb{E}(|s_{n'}|^2) = 1$, so

$$C_{n'} \succeq 0. \quad (39)$$

Inversely, if

$$|\tilde{q}_{n'}| \leq 1 \quad (40)$$

then $\begin{bmatrix} 1 & \tilde{q}_{n'} \\ (\tilde{q}_{n'})^* & 1 \end{bmatrix}$ is qualified as a pseudo covariance of a normalized improper Gaussian source.

Obviously, $\sqrt{p_{n'}}s_{n'}$ is also improper Gaussian with the augmented covariance

$$\mathbf{Q}_{n'} = p_{n'} \begin{bmatrix} 1 & \tilde{q}_{n'} \\ (\tilde{q}_{n'})^* & 1 \end{bmatrix} \quad (41)$$

$$= \begin{bmatrix} p_{n'} & p_{n'}\tilde{q}_{n'} \\ p_{n'}\tilde{q}_{n'}^* & p_{n'} \end{bmatrix} \quad (42)$$

$$= \begin{bmatrix} p_{n'} & q_{n'} \\ q_{n'}^* & p_{n'} \end{bmatrix} \quad (43)$$

for

$$q_{n'} = p_{n'}\tilde{q}_{n'} \in \mathbb{C}, \quad (44)$$

which satisfies the constraint

$$|q_{n'}| < p_{n'}. \quad (45)$$

Inversely, if $q_{n'}$ satisfies (45), then $\tilde{q}_{n'} \triangleq q_{n'}/p_{n'}$ also satisfies (40) so it is a pseudo-covariance of a normalized improper Gaussian source. Therefore, in what follows, we will design $\mathbf{Q}_{n'}$ defined by (43) with $p_{n'}$ and $q_{n'}$ satisfying the constraint (45), which is convex.

Using (4), the achievable rate by decoding the signal $\sqrt{p_n}\lambda_{nn}s_n$ in (4) under IGS in LBR is approximated by $(1/2)\tilde{\rho}_n(\mathbf{p}, \mathbf{q})$ with [34]

$$\tilde{\rho}_n(\mathbf{p}, \mathbf{q}) = \ln |\mathbf{I}_2 + \lambda_{nn}^2 \mathbf{Q}_n \mathcal{L}_n^{-1}(\mathbf{p}, \mathbf{q})| \quad (46)$$

$$= 4 \ln \lambda_{nn} + \ln |\mathbf{Q}_n| \\ + \ln \left| (\lambda_{nn}^2 \mathbf{Q}_n)^{-1} + \mathcal{L}_n^{-1}(\mathbf{p}, \mathbf{q}) \right| \quad (47)$$

with

$$\mathcal{L}_n(\mathbf{p}, \mathbf{q}) \triangleq \sum_{n' \in \mathcal{N} \setminus \{n\}} \begin{bmatrix} \lambda_{nn'} & 0 \\ 0 & \lambda_{nn'}^* \end{bmatrix} \mathbf{Q}_{n'} \begin{bmatrix} \lambda_{nn'}^* & 0 \\ 0 & \lambda_{nn'} \end{bmatrix} \\ + \sigma^2 \mathbf{I}_2, \quad (48)$$

which is the augmented covariance of the interference plus noise in (4). In what follows, we shall call $\tilde{\rho}_n(\mathbf{p}, \mathbf{q})$ defined by (47) the IGS Shannon function rate.

Following (8), the achievable URLLC rate by decoding $\sqrt{p_n}\lambda_{nn}s_n$ under IGS can be approximated by

$$\rho_n(\mathbf{p}, \mathbf{q}) \triangleq \frac{1}{2} \tilde{\rho}_n(\mathbf{p}, \mathbf{q}) - \frac{1}{\sqrt{Bt_t}} Q_G^{-1}(\epsilon^c) \delta_n(\mathbf{p}), \quad (49)$$

with $\delta_n(\mathbf{p})$ is defined in (9) and depends on the SINR $\gamma_n(\mathbf{p})$ (see (6)).

The problem of max-min users' URLLC rate optimization is formulated accordingly by

$$\max_{\substack{\mathbf{p} \in \mathbb{R}_+^N \\ \mathbf{q} \in \mathbb{C}^N}} \min_{n \in \mathcal{N}} \rho_n(\mathbf{p}, \mathbf{q}) \quad \text{s.t.} \quad (45) \quad (50a)$$

$$\pi_m(\mathbf{p}) \leq P, m \in \mathcal{M}. \quad (50b)$$

Let $(\mathbf{p}^{(\kappa)}, \mathbf{q}^{(\kappa)})$ be the feasible point for (50) that is found from the $(\kappa - 1)$ -th iteration.

We need to find a lower bounding concave function approximation of the function $\tilde{\rho}_n(\mathbf{p}, \mathbf{q})$, defined from (47). Applying the inequality (63) in the Appendix to the last term in the RHS of (47) for $(\mathbf{X}, \mathbf{Y}) = (\lambda_{nn}^2 \mathbf{Q}_n, \mathcal{L}_n(\mathbf{p}, \mathbf{q}))$ and $(\bar{\mathbf{X}}, \bar{\mathbf{Y}}) = (\lambda_{nn}^2 \mathbf{Q}_n^{(\kappa)}, \mathcal{L}_n(\mathbf{p}^{(\kappa)}, \mathbf{q}^{(\kappa)}))$ yields

$$\begin{aligned} & \ln \left| (\lambda_{nn}^2 \mathbf{Q}_n)^{-1} + \mathcal{L}_n^{-1}(\mathbf{p}, \mathbf{q}) \right| \\ & \geq \hat{a}_n^{(\kappa)} - \lambda_{nn}^2 \langle \hat{\mathbf{B}}^{(\kappa)}, \mathbf{Q}_n \rangle \\ & \quad - \sum_{n' \in \mathcal{N} \setminus \{n\}} \left\langle \hat{\mathbf{C}}^{(\kappa)}, \begin{bmatrix} \lambda_{nn'} & 0 \\ 0 & \lambda_{nn'}^* \end{bmatrix} \mathbf{Q}_{n'} \begin{bmatrix} \lambda_{nn'}^* & 0 \\ 0 & \lambda_{nn'} \end{bmatrix} \right\rangle \end{aligned} \quad (51)$$

for

$$\begin{aligned} \hat{a}_n^{(\kappa)} & \triangleq \ln \left| (\lambda_{nn}^2 \mathbf{Q}_n^{(\kappa)})^{-1} + \mathcal{L}_n^{-1}(\mathbf{p}^{(\kappa)}, \mathbf{q}^{(\kappa)}) \right| + 2 - \sigma^2 \langle \hat{\mathbf{C}}^{(\kappa)}, \mathbf{Q}_n \rangle, \\ 0 & \leq \hat{\mathbf{B}}^{(\kappa)} \triangleq \lambda_{nn}^{-2} (\mathbf{Q}_n^{(\kappa)})^{-1} - (\lambda_{nn}^2 \mathbf{Q}_n^{(\kappa)} + \mathcal{L}_n(\mathbf{p}^{(\kappa)}, \mathbf{q}^{(\kappa)}))^{-1}, \\ 0 & \leq \hat{\mathbf{C}}^{(\kappa)} \triangleq \mathcal{L}_n^{-1}(\mathbf{p}^{(\kappa)}, \mathbf{q}^{(\kappa)}) - (\lambda_{nn}^2 \mathbf{Q}_n^{(\kappa)} + \mathcal{L}_n(\mathbf{p}^{(\kappa)}, \mathbf{q}^{(\kappa)}))^{-1}. \end{aligned}$$

Next, by applying the inequality (64) in the Appendix to the second term in the RHS of (47) for $\mathbf{X} = \mathbf{Q}_n$ and $\bar{\mathbf{X}} = \mathbf{Q}_n^{(\kappa)}$, we obtain

$$\ln |\mathbf{Q}_n| \geq \ln |\mathbf{Q}_n^{(\kappa)}| + 2 - \langle \mathbf{Q}_n^{(\kappa)}, \mathbf{Q}_n^{-1} \rangle, \quad (52)$$

It follows from (52), (51) and the presentation (47) that the function

$$\tilde{\rho}_n^{(\kappa)}(\mathbf{p}, \mathbf{q}) \triangleq 4 \ln \lambda_{nn} + \text{RHS of (52)} + \text{RHS of (51)}, \quad (53)$$

is a lower bounding concave function approximation of $\rho_n(\mathbf{p}, \mathbf{q})$ and the former matches with the latter at $(\mathbf{p}^{(\kappa)}, \mathbf{q}^{(\kappa)})$.

Therefore, the function

$$\rho_n^{(\kappa)}(\mathbf{p}, \mathbf{q}) \triangleq \frac{1}{2} \tilde{\rho}_n^{(\kappa)}(\mathbf{p}, \mathbf{q}) - \frac{1}{\sqrt{\mathcal{B}t_t}} Q_G^{-1}(\epsilon^c) \delta_n^{(\kappa)}(\mathbf{p}) \quad (54)$$

provides a lower bounding concave function approximation of $\rho_n(\mathbf{p}, \mathbf{q})$ and the former matches with the latter at $(\mathbf{p}^{(\kappa)}, \mathbf{q}^{(\kappa)})$, where $\delta_n^{(\kappa)}(\mathbf{p})$ is derived in (16). We solve the following convex problem of the computational complexity $\mathcal{O}((2N)^3(M+N))$ [26, p. 4], at the κ -th iteration, to generate the next iterative point $(\mathbf{p}^{(\kappa+1)}, \mathbf{q}^{(\kappa+1)})$

$$\max_{\substack{\mathbf{p} \in \mathbb{R}_+^N \\ \mathbf{q} \in \mathbb{C}^N}} \min_{n \in \mathcal{N}} \rho_n^{(\kappa)}(\mathbf{p}, \mathbf{q}) \quad \text{s.t.} \quad (45), (50b) \quad (55)$$

Like Algorithm 1, Algorithm 3 converges at least to a local optimal solution of the problem (50). Similarly, like Algorithm

Algorithm 3 Path-following algorithm for solving the problem (50)

- 1: **Initialization:** Iterate the convex problem $\max_{\mathbf{p}, \mathbf{q}} \min_{n \in \mathcal{N}} \tilde{\rho}_n^{(\kappa)}(\mathbf{p}, \mathbf{q})$ s.t. (45), (50b) until the convergence to obtain an initial point $(\mathbf{p}^{(0)}, \mathbf{q}^{(0)})$. Set $\kappa = 0$.
- 2: **Repeat until convergence of the objective function in (50):** Solve the convex problem (55) to generate $(\mathbf{p}^{(\kappa+1)}, \mathbf{q}^{(\kappa+1)})$. Reset $\kappa \leftarrow \kappa + 1$.

Algorithm 4 Path-following algorithm for solving the problem (56)

- 1: **Initialization:** Iterate the convex problem (55) until its optimal value is larger or equal to \bar{r} for obtaining an initial point $(\mathbf{p}^{(0)}, \mathbf{q}^{(0)})$. Set $\kappa = 0$.
- 2: **Repeat until convergence of $\gamma^{(\kappa)} \triangleq \sum_{n \in \mathcal{N}} \rho_n(\mathbf{p}^{(\kappa)}, \mathbf{q}^{(\kappa)}) / \pi(\mathbf{p}^{(\kappa)})$:** Solve the convex problem $\max_{(\mathbf{p}, \mathbf{q})} \sum_{n \in \mathcal{N}} \rho_n^{(\kappa)}(\mathbf{p}, \mathbf{q}) - \gamma^{(\kappa)} \pi(\mathbf{p})$ s.t. (11b), (45), $\rho_n^{(\kappa)}(\mathbf{p}, \mathbf{q}) \geq \bar{r}, n \in \mathcal{N}$ to generate $(\mathbf{p}^{(\kappa+1)}, \mathbf{q}^{(\kappa+1)})$ and $\gamma^{(\kappa+1)}$. Reset $\kappa \leftarrow \kappa + 1$.
- 3: **Output $(\mathbf{p}^{(\kappa)}, \mathbf{q}^{(\kappa)})$** as the optimal solution of (56) and

$$\left(\frac{1}{\gamma^{(\kappa)}} + MP_{bt} \right)^{-1} = \frac{\gamma^{(\kappa)}}{1 + MP_{bt} \gamma^{(\kappa)}}$$

as the optimal value of (56).

2, Algorithm 3 converges at least to a locally optimal solution of the problem of maximizing QoS aware energy efficiency

$$\max_{(\mathbf{p}, \mathbf{q})} \frac{\sum_{n \in \mathcal{N}} \rho_n(\mathbf{p}, \mathbf{q})}{\pi_{te}(\mathbf{p}, \mathbf{q})} \quad \text{s.t.} \quad (11b), (45), \rho_n(\mathbf{p}, \mathbf{q}) \geq \bar{r}, n \in \mathcal{N}. \quad (56)$$

with $\pi(\mathbf{p}) \triangleq \alpha \sum_{m \in \mathcal{M}} \pi_m(\mathbf{p}) + \nu$ with ν defined from (30) and $\pi_{te}(\mathbf{p}, \mathbf{q}) \triangleq \pi(\mathbf{p}) + P_{sc} + M(P_0 + P_{bt} \sum_{n \in \mathcal{N}} \rho_n(\mathbf{p}, \mathbf{q}))$.

IV. SIMULATIONS

In this section, we analyze the performance of our proposed algorithms. We start by first describing our simulation parameters. The large-scale fading coefficient in (1) is given by

$$\beta_{mn} = 10^{(-\text{PL}_{mn} + X_{mn})/10}, \quad (57)$$

where PL_{mn} is the path-loss (in dB) and $10^{X_{mn}/10}$ represents the shadowing effect with $X_{mn} \sim \mathcal{N}(0, \sigma_{\text{sh}}^2)$. We use a three-slope model for path-loss [35], which is given by (in dB)

$$\text{PL}_{mn} = \begin{cases} L + 35 \log_{10}(d_{mn}), & d_{mn} > d_1 \\ L + 15 \log_{10}(d_1) + 20 \log_{10}(d_{mn}), & d_o < d_{mn} \leq d_1 \\ L + 15 \log_{10}(d_1) + 20 \log_{10}(d_o), & d_{mn} \leq d_o \end{cases} \quad (58)$$

where the constant factor L depends on the carrier frequency and the heights of the users and APs and d_{mn} is the distance between the AP m and UE n . In our simulations, we set $\sigma_{\text{sh}} = 8$ dB, $d_o = 10$ m, $d_1 = 50$ m, and $L = 140.7$ dB. These parameters resemble those in [10], [31].

We have randomly deployed the APs and the UEs within a square of 1×1 km². The square is wrapped around at the

edges to avoid boundary effects. Corresponding to a coherence bandwidth of 320 KHz and coherence time of 1 ms, the length of coherence interval is $t = 320$ samples. We set the system bandwidth to 10 MHz, noise power density is set to -174 dBm/Hz and noise figure is equal to 9 dB.

To simulate the energy efficiency performance of a network, the threshold rate is set to $\bar{r} = 0.5$ bps/Hz, the drain efficiency of power amplifier is set to 0.4, i.e., $\alpha = 1/0.4$ in (27), internal power for running the circuit components is set to $P_{sc} = 0.2$ W, fixed power consumption for each backhaul is $P_0 = 0.825$ W, and traffic-dependent backhaul power is $P_{bt} = 0.25$ W/(Gbits/s). These values are taken from [31].

Unless otherwise specified, we use $M = 128$ APs, $N = 32$ UEs, transmission duration $= t_t = 0.1$ ms (defined below (8))¹, and transmit power budget $P = 10$ mW. The decoding error probability ϵ_c (defined below (8)) is set to 10^{-5} . The power allocated to the pilots, p_u , is set to P . Note that following [23], only the APs that are closer to the UE n (within the radius of around 150 meters), are selected as the serving APs for the UE n .

In what follows, we refer to our proposed algorithms as follows for convenience:

- *PGS (Alg. 1)* refers to the PGS based max-min rate Algorithm 1.
- *IGS (Alg. 3)* refers to the IGS based max-min rate Algorithm 3.
- *PGS (Alg. 2)* refers to the PGS based energy efficiency optimization Algorithm 2.
- *IGS (Alg. 4)* refers to the IGS based energy efficiency optimization Algorithm 4.

We simulate both max-min rate and max-min URLLC rate for comparison purpose. Similarly, to analyze the energy efficiency performance, we simulate it using both Shannon and URLLC rates for comparison purpose. Note that in all the figures, the rates are calculated in bps/Hz, which are obtained by dividing the rates (defined in nats/sec/Hz) by $\ln 2$. Following the simulations setup in [10], the results are averaged over random realizations of the AP/user locations, shadow fading, and small-scale fading.

In the following subsections, we will separately present the max-min rate performance and the energy efficiency performance of our proposed algorithms.

A. Max-min rate performance

Fig. 1 plots the optimized max-min user rate versus the number of APs M . As expected, the max-min user rate increases with the increase in the number of APs due to the availability of more resources (recall that each AP has its own independent power budget P). Fig. 1 shows that the IGS based Alg. 3 clearly outperforms the PGS based Alg. 1 and the performance gap increases with the increase in M . As expected, Shannon function rate performance is better than the URLLC rate performance.

¹there could be several short blocklength communications within coherence time because the coherence time is usually longer than the URLLC transmission duration.

Fig. 2 plots the optimized max-min user rate versus the number of users N . As expected, the max-min user rate decreases with the increase in the the number of users N . This is because when more users will be competing for the resources, the minimum user-rate will be more affected. Fig. 2 shows that the IGS based Alg. 3 clearly outperforms the PGS based Alg. 1 by almost 0.5 bps/Hz.

Fig. 3 plots the optimized max-min user rate versus the transmit power budget P . The max-min user rate increases with the increase in the power budget P , however, there is almost no further improvement in the max-min rate as P increases beyond 0 dBm. Fig. 3 also shows the supremacy of IGS based Alg. 3 over the PGS based Alg. 1.

Fig. 4 plots the optimized max-min user rate versus the transmission duration t_t . Thus, Fig. 4 shows the effect of short blocklength transmission for varying values of blocklength. Fig. 4 shows that the URLLC rate increases with the increase in the transmission duration and gets closer to the Shannon function rate. The Shannon function rate is fixed as it does not depend on the transmission duration and assumes infinite blocklength transmission. Fig. 4 shows that the performance gap between the IGS based Alg. 3 and the PGS based Alg. 1 does not change over different considered values of t_t .

B. Energy efficiency performance

Fig. 5 plots the optimized energy efficiency versus the number of APs M . The energy efficiency decreases with the increase in the number of APs. This is because energy efficiency, as defined in the objective of (26a), is the ratio of sum-rate to the total power consumption (defined in (27)). We know that sum-rate increases with M due to the availability of more resources, however, the power consumption also increases with M (see (27)). The increase in the power consumption dominates the increase in the sum-rate, which results in the decrease in the energy efficiency performance. Fig. 5 shows that the IGS based Alg. 4 clearly outperforms the PGS based Alg. 2.

Fig. 6 plots the optimized energy efficiency versus the number of users N . The energy efficiency increases with the increase in the number of users. This is because the sum-rate increases with the increase in the number of users. Fig. 6 shows that the performance gain of the IGS based Alg. 4 over the PGS based Alg. 2 increases with the increase in the number of users. It can be observed from Figs. 5-6 that the performance gap between IGS and PGS based on the URLLC rate is more compared to that based on the Shannon function rate.

Fig. 7 plots the optimized energy efficiency versus the transmit power budget P . The energy efficiency increases with the increase in the power budget P due to the availability of more resources. Fig. 7 also shows the supremacy of IGS over the PGS. Fig. 8 plots the optimized energy efficiency versus the transmission duration t_t . Fig. 8 shows that the energy efficiency based on the URLLC rate increases with the increase in the transmission duration and gets closer to that based on the Shannon function rate. The energy efficiency based on Shannon function rate is fixed as it does not depend

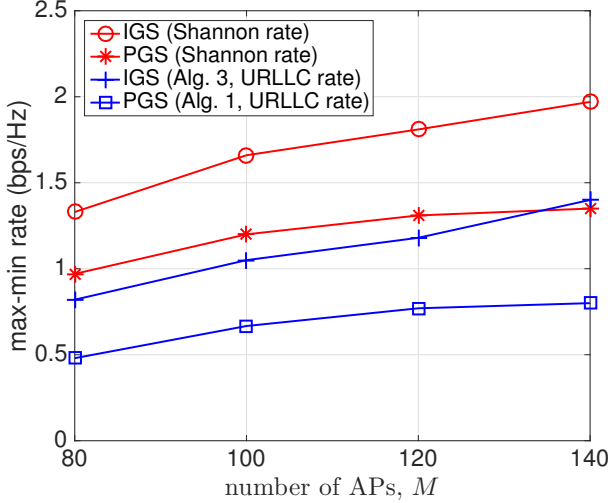


Fig. 1: Optimized max-min user rate versus the number of APs M .

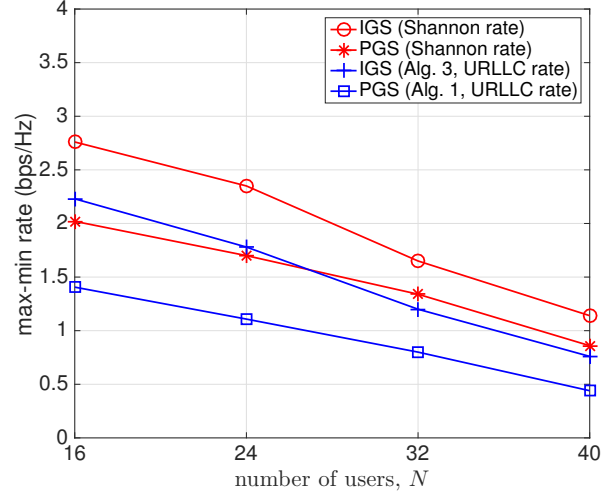


Fig. 2: Optimized max-min user rate versus the number of users N .

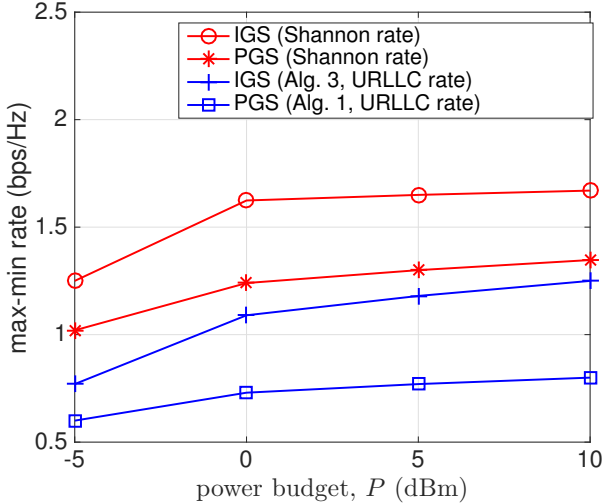


Fig. 3: Optimized max-min user rate versus the transmit power budget P .

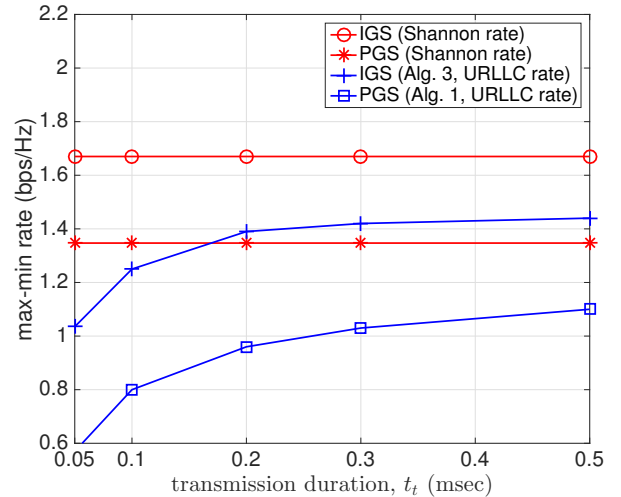


Fig. 4: Optimized max-min user rate versus the transmission duration t_t .

on the transmission duration and assumes infinite blocklength transmission. Fig. 8 shows that the performance gap between the IGS based Alg. 4 and the PGS based Alg. 2 decreases with the increase in the transmission duration t_t .

Finally, Fig. 9 plots the optimized energy efficiency versus the threshold rate \bar{r} . It can be observed from Fig. 9 that the performance gain by IGS over PGS increases by increasing the threshold rate. Particularly, the proposed IGS based Alg. 4 shows around 22% gain in terms of energy efficiency compared to the proposed PGS based Alg. 2, under $\bar{r} = 0.5$ bps/Hz and short blocklength regime. One interesting observation is that the performance gap between IGS and PGS based on the URLLC rate is more compared to that based on the Shannon function rate. This shows the advantage of IGS over PGS under practical short blocklength communication.

V. CONCLUSIONS

We have introduced a particular class of conjugate beamforming (CB) for a cfm-MIMO to maintain the low computational complexity for its design while requiring only local CSI for its transmit implementation. This CB has been used to optimize both the users' Shannon rate (in the long blocklength) and URLLC rate (in the short blocklength). IGS has been also employed to further improve the max-min user rate performance. New path-following algorithms have been developed for their computations. The energy-efficiency of cfm-MIMO has been also addressed. The obtained theoretical results have been backed up by intensive simulations.

One possible future research option is to consider URLLC in cfm-MIMO under imperfect CSI assumption. One has to consider that short pilot sequences in short blocklength communication may impact the quality of estimated CSI,

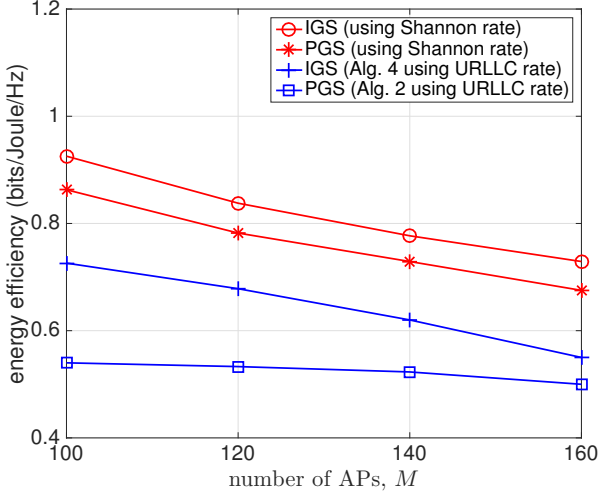


Fig. 5: Optimized energy efficiency versus the number of APs M .

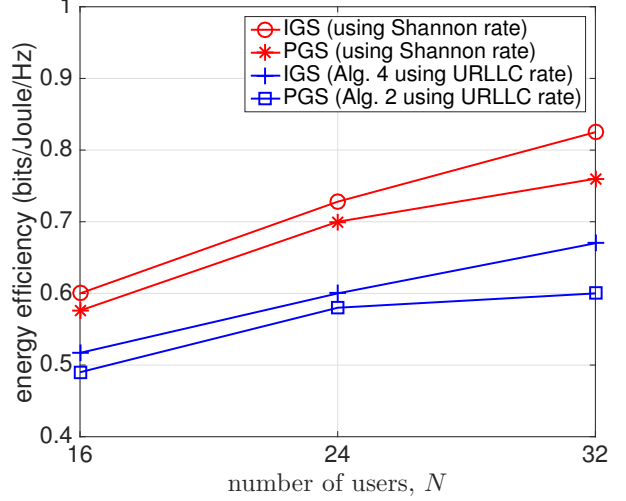


Fig. 6: Optimized energy efficiency versus the number of users N .

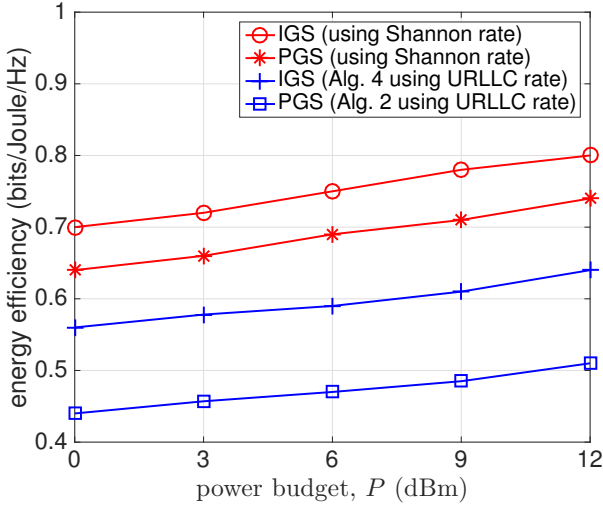


Fig. 7: Optimized energy efficiency versus the transmit power budget P .

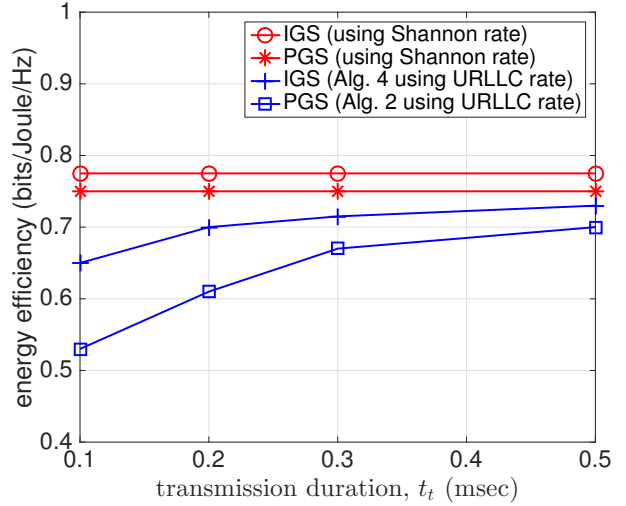


Fig. 8: Optimized energy efficiency versus the transmission duration t_t .

which in turn can affect the reliability. Moreover, since many URLLC transmissions can take place during payload data transmission as the duration of each URLLC transmission is usually much shorter than the coherence interval for payload data transmission, the channel estimate obtained from the first short blocklength packet (first URLLC transmission) can be used for the remaining short blocklength packets (subsequent URLLC transmissions) during the coherence interval.

APPENDIX: FUNDAMENTAL INEQUALITIES

The following inequalities for all $x > 0$, $y > 0$, and $\bar{x} > 0$, $\bar{y} > 0$, were proved in [36], [37]:

$$\ln\left(1 + \frac{x}{y}\right) \geq \ln\left(1 + \frac{\bar{x}}{\bar{y}}\right) + \frac{\bar{x}}{\bar{y}} \left(2\frac{\sqrt{x}}{\sqrt{\bar{x}}} - \frac{x+y}{\bar{x}+\bar{y}} - 1\right), \quad (59)$$

and

$$\frac{x^2}{y} \geq \frac{\bar{x}}{\bar{y}} \left(2x - \frac{\bar{x}}{\bar{y}}y\right). \quad (60)$$

and

$$\sqrt{x} \leq \frac{1}{2} \left(\sqrt{\bar{x}} + \frac{x}{\sqrt{\bar{x}}}\right). \quad (61)$$

The following inequality is based on least-square

$$x \leq 0.5(x^2/\bar{x} + \bar{x}) \quad \forall x > 0, \bar{x} > 0. \quad (62)$$

We also recall the following inequality for all $\mathbf{X} \succ 0$, $\mathbf{Y} \succ 0$ and $\bar{\mathbf{X}} \succ 0$, $\bar{\mathbf{Y}} \succ 0$ of dimension 2×2 [33]:

$$\begin{aligned} \ln|\mathbf{X}^{-1} + \mathbf{Y}^{-1}| &\geq \ln|\bar{\mathbf{X}}^{-1} + \bar{\mathbf{Y}}^{-1}| + 2 \\ &\quad - \langle \bar{\mathbf{X}}^{-1} - (\bar{\mathbf{X}} + \bar{\mathbf{Y}})^{-1}, \mathbf{X} \rangle \\ &\quad - \langle \bar{\mathbf{Y}}^{-1} - (\bar{\mathbf{X}} + \bar{\mathbf{Y}})^{-1}, \mathbf{Y} \rangle, \end{aligned} \quad (63)$$

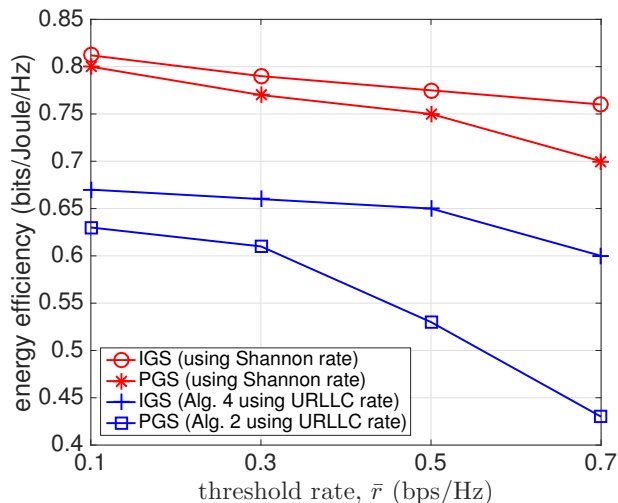


Fig. 9: Optimized energy efficiency versus threshold rate \bar{r} .

and

$$\log |\mathbf{X}| \geq \log |\bar{\mathbf{X}}| + 2 - \langle \bar{\mathbf{X}}, (\mathbf{X})^{-1} \rangle. \quad (64)$$

REFERENCES

- [1] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultra-reliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, pp. 1711–1726, Aug. 2016.
- [2] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proc. IEEE*, vol. 106, pp. 1834–1853, Oct. 2018.
- [3] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-enabled tactile internet," *IEEE J. Sel. Areas Commun.*, vol. 34, pp. 460–473, Mar. 2016.
- [4] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2307–2359, May 2010.
- [5] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, pp. 4232–4264, Jul. 2014.
- [6] C. She, C. Yang, and T. Q. S. Quek, "Cross-layer optimization for ultra-reliable and low-latency radio access networks," *IEEE Trans. Wirel. Commun.*, vol. 17, pp. 127–141, Jan. 2018.
- [7] C. She, C. Yang, and T. Q. S. Quek, "Joint uplink and downlink resource configuration for ultra-reliable and low-latency communications," *IEEE Trans. Commun.*, vol. 66, pp. 2266–2280, May 2018.
- [8] C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications," *IEEE Trans. Wirel. Commun.*, vol. 18, pp. 402–415, Jan. 2019.
- [9] A. Azari, M. Ozger, and C. Cavdar, "Risk-aware resource allocation for URLLC: Challenges and strategies with machine learning," *ArXiv Technical Report*, 2019.
- [10] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wirel. Commun.*, vol. 16, pp. 1834–1849, Mar. 2017.
- [11] T. M. Hoang, H. Q. Ngo, T. Q. Duong, H. D. Tuan, and A. Marshall, "Cell-free massive MIMO networks: Optimal power control against active eavesdropping," *IEEE Trans. Commun.*, vol. 26, pp. 4724–4737, Oct. 2018.
- [12] M. Bashar, K. Cumanan, A. G. Burr, M. Debbah, and H. Q. Ngo, "On the uplink maxmin SINR of cell-free massive MIMO systems," *IEEE Trans. Wirel. Commun.*, vol. 18, pp. 2021–2036, Apr. 2019.
- [13] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. UK.: Cambridge Univ. Press, 2016.
- [14] E. Nayeibi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Pre-coding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wirel. Commun.*, vol. 16, pp. 4445–4459, Jul. 2017.
- [15] C. Hellings, M. Joham, and W. Utschick, "QoS feasibility in MIMO broadcast channels with widely linear transceivers," *IEEE Signal Process. Lett.*, vol. 20, pp. 1134–1137, Nov. 2013.
- [16] Y. Zeng, C. M. Yetis, E. Gunawan, Y. L. Guan, and R. Zhang, "Transmit optimization with improper Gaussian signaling for interference channels," *IEEE Trans. Signal Process.*, vol. 61, pp. 2899–2913, Jun. 2013.
- [17] S. Lagen, A. Agustin, and J. Vidal, "On the superiority of improper Gaussian signaling in wireless interference MIMO scenarios," *IEEE Trans. Commun.*, vol. 64, pp. 3350–3368, Aug. 2016.
- [18] A. A. Nasir, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Improper Gaussian signaling for broadcast interference networks," *IEEE Signal Process. Lett.*, vol. 26, pp. 808–812, Jun. 2019.
- [19] H. D. Tuan, A. A. Nasir, H. H. Nguyen, T. Q. Duong, and H. V. Poor, "Non-orthogonal multiple access with improper Gaussian signaling," *IEEE J. Select. Topics Signal Process.*, vol. 13, pp. 496–507, Mar. 2019.
- [20] M. Soleymani, C. Lameiro, I. Santamaria, and P. J. Schreier, "Energy-efficient improper signaling for K-user interference channels," in *Proc. European Signal Processing Conference (EUSIPCO)*, pp. 1–5, 2019.
- [21] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.
- [22] G. Interdonato, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Downlink training in cell-free massive mimo: A blessing in disguise," *IEEE Trans. Wirel. Commun.*, vol. 18, pp. 5153–5169, Nov. 2019.
- [23] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1–4, May 2019.
- [24] J. Scarlett, V. Y. F. Tan, and G. Durisi, "The dispersion of nearest-neighbor decoding for additive non-Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 63, pp. 81–92, Jan. 2017.
- [25] H. Tuy, *Convex Analysis and Global Optimization (second edition)*. Springer International, 2017.
- [26] D. Peaucelle, D. Henrion, and Y. Labit, "Users guide for SeDuMi interface 1.03," 2002.
- [27] A. A. Nasir, H. D. Tuan, D. T. Ngo, T. Q. Duong, and H. V. Poor, "Beamforming design for wireless information and power transfer systems: Receive power-splitting versus transmit time-switching," *IEEE Trans. Commun.*, vol. 65, pp. 876–889, Feb. 2017.
- [28] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Select. Areas Commun.*, vol. 34, pp. 697–709, Apr. 2016.
- [29] A. Zappone, L. Sanguinetti, G. Bacci, E. A. Jorswieck, and M. Debbah, "Energy-efficient power control: A look at 5G wirel. technologies," *IEEE Trans. Signal Process.*, vol. 64, pp. 1668–1683, Apr. 2016.
- [30] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, pp. 1871–1874, Aug. 2017.
- [31] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. Network.*, vol. 2, pp. 25–39, Jan. 2018.
- [32] E. Bjornson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?," *IEEE Trans. Wirel. Commun.*, vol. 14, pp. 3059–3075, Jun. 2015.
- [33] L. D. Nguyen, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Multi-user regularized zero-forcing beamforming," *IEEE Trans. Signal Process.*, vol. 67, pp. 2839–2853, Jun. 2019.
- [34] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov./Dec. 1999.
- [35] A. Tang, J. Sun, and K. Gong, "Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area," in *Proc. IEEE Vehicular Technology Conference (VTC)*, Spring, vol. 1, pp. 333–336, May 2001.
- [36] H. H. M. Tam, H. D. Tuan, and D. T. Ngo, "Successive convex quadratic programming for quality-of-service management in full-duplex MU-MIMO multicell networks," *IEEE Trans. Commun.*, vol. 64, pp. 2340–2353, June 2016.
- [37] Z. Sheng, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "Power allocation for energy efficiency and secrecy of wireless interference networks," *IEEE Trans. Wirel. Commun.*, vol. 17, pp. 3737–3751, Jun. 2018.