

“© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

A New Class of Structured Beamforming for Content-Centric Fog Radio Access Networks

W. Zhu^{1,2}, H. D. Tuan², E. Dutkiewicz², Y. Fang¹, and L. Hanzo³

Abstract—A multi-user fog radio access network (F-RAN) is designed for supporting content-centric services. The requested contents are partitioned into sub-contents, which are then ‘beamformed’ by the remote radio heads (RRHs) for transmission to the users. Since a large number of beamformers must be designed, this poses a computational challenge. We tackle this challenge by proposing a new class of regularized zero forcing beamforming (RZFB) for directly mitigating the inter-content interferences, while the ‘intra-content interference’ is mitigated by successive interference cancellation at the user end. Thus each beamformer is decided by a single real variable (for proper Gaussian signaling) or by a pair of complex variables (for improper Gaussian signaling). Hence the total number of decision variables is substantially reduced to facilitate tractable computation. To address the problem of energy efficiency optimization subject to multiple constraints, such as individual user-rate requirement and the fronthauling constraint of the links between the RRHs and the centralized baseband signal processing unit, as well as the total transmit power budget, we develop low-complexity path-following algorithms. Finally, we actualize their performance by simulations.

Index Terms—Fog radio access network (F-RAN), soft-transfer fronthauling, regularized zero-forcing beamforming, content service, Gaussian signaling.

I. INTRODUCTION

Fog radio access networks (F-RANs) [1]–[5], which place processing units at the network edge for reducing the backhaul’s latency and traffic load, have emerged for meeting the radical requirements of mission-critical cloud radio access networks (C-RANs), as exemplified by augmented reality/virtual reality (AR/VR), device-to-device (D2D) communications, smart living and smart cities, etc. (see e.g. [6]–[9] and references therein). On one hand, the uniform quality of service (QoS) provision for cell-edge and cell-center users is satisfied by exploiting the strategic spatial distribution of the remote radio heads (RRHs) over the network [10]–[13]. On the other hand, employing RRHs for proactively caching popular contents enables flawless low-latency service provision relying on a high network throughput at a high energy efficiency,

This work was supported in part by the Australian Research Council’s Discovery Projects under Grant DP190102501, in part by the Engineering and Physical Sciences Research Council projects EP/P034284/1 and EP/P003990/1 (COALESCE), and in part by the European Research Council’s Advanced Fellow Grant QuantCom (Grant No. 789028)

¹School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China; Email: wenbozhu@shu.edu.cn, yfang@staff.shu.edu.cn

²School of Electrical and Data Engineering, University of Technology Sydney, Broadway, NSW 2007, Australia; Email: wenbo.zhu@student.uts.edu.au, tuan.hoang@uts.edu.au, eryk.dutkiewicz@uts.edu.au

³School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K; Email: lh@ecs.soton.ac.uk

because only a small fraction of the requested video clips must be fetched through the limited-capacity fronthaul links [14]–[17].

Beamforming aided RRHs have been considered in [18]–[22] and in the references therein. However, this design problem is high-dimensional, as it involves many beamforming vectors calculated for covering specific segments of contents. The resultant problem is nonconvex because the throughput function is neither convex nor concave, hence making the objection function constructed for maximizing the sub-contents’ throughput nonconcave, while the associated fronthaul constraints are nonconvex. The challenge of dimension is even further escalated in [18], [19], which treat beamforming vectors of dimension n_t (n_t is the number of RRH antennas), hence ultimately arriving at rank-one matrices of dimension $n_t \times n_t$. The resultant multiple rank-one constraints are then dropped for the sake of facilitating so-called difference of two convex functions based iterations [23]. Despite dropping these rank-one constraints, the problem still remains computationally complex, since it relies on logarithmic determinant function optimization, which has unknown computational complexity. In [18], no complexity analysis was provided, while the numerical examples in [19] are limited to the simplest and lowest-dimensional case of three single-antenna RRHs ($n_t = 1$) serving three users. The path-following algorithms proposed in [21] handled more practical scenarios of five four-antenna RRHs serving five users, which became possible as a benefit of the structural exploitation of approximate zero-forcing beamforming. In contrast to the iterations used in [18], [19], the path-following algorithms of [21] invoke a convex fractional solver of polynomial complexity at each iteration. However, it still remained an open challenge to reduce the dimension of the approximate zero forcing beamformers, while maintaining the throughput.

Against the above background, this paper offers the following new contributions to the design of RRHs beamformers.

- We propose a new class of regularized zero forcing beamformers (RZFB) operating in the presence of both ‘intra-content’ and ‘inter-content’ interferences, which allows us to recast the beamforming design into an optimization problem of moderate dimension.
- We conceive a new convex quadratic approximation for the fronthaul constraints, which allows us to design new path-following algorithms for determining the beamforming vectors by relying on a low-complexity convex quadratic solver at each iteration for generating an improved feasible point.

The paper is organized as follows. Section II is devoted to the modeling of the constrained-backhaul FRAN. Sections III and IV constitute the main technical contribution of the paper, which are respectively devoted to RZFB-based proper Gaussian signaling (PGS) and improper Gaussian signaling (IGS). Our simulations are discussed in Section V, while our conclusions offered in Section VI. The Appendix provides fundamental inequalities, which support the derivation of the technical results.

Notation and most frequently used mathematics. Only the optimization variables are boldfaced to emphasize their appearance in nonlinear functions; $|\mathcal{A}|$ is the cardinality of the set \mathcal{A} ; $[X]^2 \triangleq XX^H$ and $\langle X \rangle$ is the trace of the matrix X ; $\mathcal{CN}(0, a)$ is the set of circular Gaussian random variables with zero mean and variance a (each $s \in \mathcal{CN}(0, a)$ is also called proper because $\mathbb{E}(s^2) = 0$), while $\mathcal{C}(0, a)$ is the set of non-circular Gaussian random variables with zero means and variance a ($\mathbb{E}(s^2) \neq 0$ for $s \in \mathcal{C}(0, a)$ so is also called improper); $\text{diag}[x]$ for vector x is a diagonal matrix with the entries of x as its diagonal entries; $\mathcal{I}(x, y)$ is the mutual information between the random variable x and random variable y ; The dot product of the matrices X and Y of appropriate size is defined as $\langle X, Y \rangle \triangleq \langle X^H Y \rangle$; $X \succ 0$ ($X \succeq 0$, resp.) means X is a Hermitian symmetric ($X = X^H$) and positive definite (positive semi-definite) matrix. Accordingly, $X \succeq Y$ merely means $X - Y \succeq 0$. One of the most fundamental properties of positive definite matrices is $X \succeq Y \succ 0 \Leftrightarrow Y^{-1} \succeq X^{-1} \succ 0$ [24]. $\|\cdot\|$ is the Frobenius norm, i.e. $\|X\| = \sqrt{\langle [X]^2 \rangle}$, which also implies $\langle A, [X]^2 \rangle = \|A^{1/2}X\|^2$ whenever $A \succeq 0$. Therefore, the function $f(X) = \langle A, [X]^2 \rangle + \langle B, X \rangle$ for $A \succeq 0$ is termed as convex quadratic, while $-f(X)$ is concave quadratic.

II. FRAN MODELING AND SIGNALING

As illustrated by Fig. 1, we consider a typical F-RAN consisting of a centralized baseband signal processing unit (BBU) and N_R uniformly distributed RRHs indexed by $i \in \mathcal{N}_R \triangleq \{1, \dots, N_R\}$, which provides content delivery for N_u randomly localized users (UEs) indexed by $k \in \mathcal{N}_u \triangleq \{1, \dots, N_u\}$. Each RRH is equipped by an n_t -antenna array, while each user equipment (UE) has a single antenna. The RRHs are connected to the BBU through fronthaul links, each of which is of capacity C .

A. Edge caching

The file library consists of F files $f \in \mathcal{F} \triangleq \{1, \dots, F\}$ labelled in order of their popularity, which is distributed according to Zipf's distribution obeying [25] $P(f) = f^{-\gamma_z} / \sum_{j \in \mathcal{F}} j^{-\gamma_z}$ with the popularity exponent being $\gamma_z > 0$. A larger γ_z results in a reduced number of extremely popular contents. Under the uncoded strategy, each file f is split into L subfiles (f, ℓ) , $\ell \in \mathcal{L} \triangleq \{1, \dots, L\}$. Each RRH can store a fraction $\mu = N_c/F_c$ of each file in the pre-fetching phase [26], where F_c is the library capacity. We define the binary-

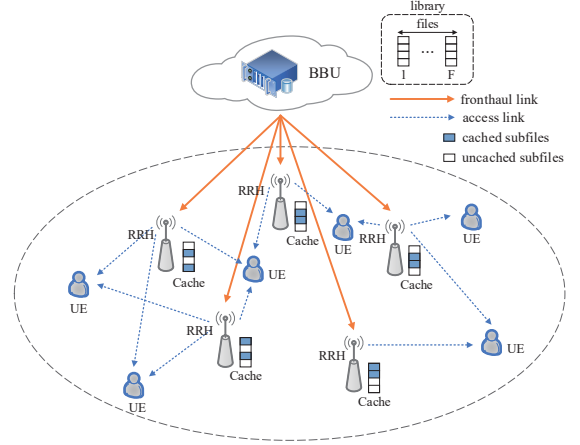


Fig. 1. F-RAN

indicator function associated with $c_{f,\ell}^i$ such that $c_{f,\ell}^i = 1$ if and only if (f, ℓ) is cached by RRH i .¹ Then we have:

$$\sum_{f=1}^F \sum_{\ell=1}^L c_{f,\ell}^i / L \leq \mu F, \quad i \in \mathcal{N}_R.$$

B. Content transmission and reception

Let us denote the set of requested files by \mathcal{F}_{req} :

$$\mathcal{F}_{req} \triangleq \{f_\nu \in \mathcal{F} : \nu = 1, \dots, N_{req}\},$$

where N_{req} ($1 \leq N_{req} \leq N_u$) is the number of requested files. For each ν , we define \mathcal{K}_ν as the set of UEs requesting file f_ν :

$$\mathcal{K}_\nu \triangleq \{1_\nu, \dots, K^{(\nu)}\}, \quad (1)$$

where it is plausible that $K^{(\nu)}$ is equal to the cardinality $|\mathcal{K}_\nu|$ of \mathcal{K}_ν .

In what follows, for convenience of presentation, we use the notation $c_{\nu,\ell}^i$ to represent the above $c_{f_\nu,\ell}^i$ and $\mathcal{N}_{req} \triangleq \{1, \dots, N_{req}\}$.

Furthermore, the subfile (f_ν, ℓ) requested by UE k_ν is transferred by the fronthaul links to those N_F RRHs, which have the highest channel gains from them to UE k_ν among the RRHs, but do not have (f_ν, ℓ) stored in their cache. Given (f_ν, ℓ) , let us define $d_{\nu,\ell}^i$ as:

$$d_{\nu,\ell}^i = \begin{cases} 1 & \text{if subfile } (f_\nu, \ell) \text{ is transferred to RRH } i, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

For $i = 1, \dots, N_R$, we denote by \mathcal{N}_i the specific set of subfiles that are either in the cache of RRH i or are received by RRH i from the BBU:

$$\mathcal{N}_i \triangleq \{(\nu, \ell) : \nu \in \mathcal{N}_{req}, \text{ and } c_{\nu,\ell}^i = 1 \text{ or } d_{\nu,\ell}^i = 1\}. \quad (3)$$

Let $s_{\nu,\ell} \in \mathcal{CN}(0, 1)$ be the proper Gaussian information source that encodes the subfile (f_ν, ℓ) . Each $s_{\nu,\ell}$ for $(\nu, \ell) \in \mathcal{N}_i$ is 'beamformed' into the signal $\lambda_{\nu,\ell}^i(s_{\nu,\ell})$ for transmission.

¹ $c_{f,\ell}^i$ is thus pre-determined by the caching strategy RRH used

TABLE I
NOMENCLATURE

| Notation | Description |
|--|---|
| f_ν | a requested file |
| (f_ν, ℓ) | the ℓ -th subfile of file f_ν |
| $s_{\nu, \ell}$ | the information source encoding (f_ν, ℓ) |
| \mathcal{K}_ν | the set of users requesting file f_ν (defined by (1)) |
| $ \mathcal{K}_\nu $ | the number of users requesting file f_ν |
| k_ν | user k requests file f_ν |
| $G_{k_\nu, \ell}^i$ | transmit beamformer for $s_{\nu, \ell}$ at RRH i by the user k_ν 's request (defined by (11)) |
| $\mathbf{p}_{k_\nu, \ell}^i$ | optimization variable for allocating the power to $G_{k_\nu, \ell}^i$ in PGS (13) |
| $(\mathbf{p}_{k_\nu, \ell}^{i,1}, \mathbf{p}_{k_\nu, \ell}^{i,1})$ | optimization variable for allocating the power to $G_{k_\nu, \ell}^i$ in IGS (49) |
| v_i | quantization noise in BBU transmission to RRH i |
| \mathbf{x}^i | the introduced variable to control v (defined by (54)) |

Let $\Xi^i \triangleq \mathcal{N}_i \setminus \{(\nu, \ell) : c_{\nu, \ell}^i = 1\}$. Under the soft-transfer fronthauling (STF) regime of [27], the BBU transfers the following quantization-contaminated version of the ‘beamformed’ subfiles to RRH i :

$$\xi_i = \sum_{(\nu, \ell) \in \Xi^i} \lambda_{\nu, \ell}^i(s_{\nu, \ell}) + v_i, \quad (4)$$

with the independent quantization noise given by $v_i \in \mathcal{C}(0, \Delta^2(\tilde{\mathbf{x}}^i) + \epsilon I_{n_t})$ and

$$\Delta(\mathbf{x}^i) \triangleq \text{diag}[\tilde{\mathbf{x}}^i] \in \mathbb{R}^{n_t \times n_t}, \tilde{\mathbf{x}}^i \triangleq (\tilde{x}_1^i, \dots, \tilde{x}_{n_t}^i)^T \in \mathbb{R}_+^{n_t}. \quad (5)$$

The fronthaul-rate constraint is constituted by the following reliably recovered rate-constraint of SFT [27]

$$\mathcal{I} \left(\xi_i, \sum_{(\nu, \ell) \in \Xi^i} \lambda_{\nu, \ell}^i(s_{\nu, \ell}) \right) \leq C, i \in \mathcal{N}_R. \quad (6)$$

It is plausible that a finer quantization results in a reduced error covariance $\Delta(\mathbf{x}^i)$, hence leading to a higher throughput, but the soft-transfer of the quantized signals is limited by the capacity of STF according to (6). As analysed in detail in [21], SFT is much more efficient than hard-transfer fronthauling (HFT) in the face of limited fronthauling capacity.

Furthermore, the signal transmitted by RRH i is

$$\chi_i \triangleq \sum_{(\nu', \ell) \in \mathcal{N}_i} \lambda_{\nu', \ell}^i(s_{\nu', \ell}) + v_i, \quad (7)$$

which contains the quantization noise v_i . This noise cannot be completely eliminated due to the limited fronthaul-rate constraint (6). The signal received by UE k_ν is

$$\begin{aligned} y_{k_\nu} &= \sum_{i=1}^{N_R} h_{k_\nu, i}^H \chi_i + z_{k_\nu} \\ &= \sum_{i=1}^{N_R} h_{k_\nu, i}^H \left(\sum_{(\nu', \ell) \in \mathcal{N}_i} \lambda_{\nu', \ell}^i(s_{\nu', \ell}) \right) + z_{k_\nu}, \end{aligned} \quad (8)$$

where $h_{k_\nu, i} \in \mathbb{C}^{n_t}$ is the channel spanning from RRH i to UE k_ν , and $z_{k_\nu} \in \mathbb{C}$ is the background noise of covariance σ .

III. REGULARIZED ZERO-FORCING BEAMFORMING FOR PROPER GAUSSIAN SIGNALING

To assist the reader, Table I provides a summary of the key notations at a glance.

A. PGS Problem statement

It follows from (8) that the interference of $\lambda_{\nu, \ell}^i(s_{\nu, \ell})$ to UE $k_{\nu'}$ ($\nu' \neq \nu$) is given by $h_{k_{\nu'}, i}^H \lambda_{\nu, \ell}^i(s_{\nu, \ell})$. Thus, for $n_\nu \triangleq \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} |\mathcal{K}_{\nu'}|$, we define the matrix of interfering channels by

$$\Phi_{k_\nu, \ell}^i \triangleq \text{Row}(h_{k_{\nu'}, i}^H)_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}, k=1, \dots, |\mathcal{K}_{\nu'}|} \in \mathbb{C}^{n_\nu \times n_t}, \quad (9)$$

and then the matrix of signal plus interfering channels as

$$\tilde{\Phi}_{k_\nu, \ell}^i \triangleq \begin{bmatrix} h_{k_\nu, i}^H \\ \Phi_{k_\nu, \ell}^i \end{bmatrix} \in \mathbb{C}^{(n_\nu+1) \times n_t}. \quad (10)$$

Here the operation Row in (9) arranges n_ν row-vectors $h_{k_{\nu'}, i}^H \in \mathbb{C}^{1 \times n_t}$, $\nu' \in \mathcal{N}_{req} \setminus \{\nu\}$, $k = 1, \dots, |\mathcal{K}_{\nu'}|$ in a matrix of n_ν rows.

For $(\nu, \ell) \in \mathcal{N}_i$ and

$$G_{k_\nu, \ell}^i \triangleq \left((\tilde{\Phi}_{k_\nu, \ell}^i)^H \tilde{\Phi}_{k_\nu, \ell}^i + \alpha I_{n_t} \right)^{-1} h_{k_\nu, i}, k = 1, \dots, |\mathcal{K}_\nu|, \quad (11)$$

associated with²

$$\alpha = \begin{cases} 0 & \text{if } \text{rank}(\tilde{\Phi}_{k_\nu, \ell}^i) = n_\nu + 1 \\ N_R \sigma / P & \text{otherwise,} \end{cases} \quad (12)$$

we propose the following class of RZFB

$$\begin{aligned} \lambda_{\nu, \ell}^i(s_{\nu, \ell}) &= \sum_{k=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k_\nu, \ell}^i G_{k_\nu, \ell}^i s_{\nu, \ell} \\ &= \left(\sum_{k=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k_\nu, \ell}^i G_{k_\nu, \ell}^i \right) s_{\nu, \ell}, \mathbf{p}_{k_\nu, \ell}^i \in \mathbb{R}. \end{aligned} \quad (13)$$

Let $e_1 \in \mathbb{R}^{n_\nu+1}$ having all-zero entries except for the first entry, which is 1. Then we have

$$\begin{aligned} \tilde{\Phi}_{k_\nu, \ell}^i G_{k_\nu, \ell}^i &= \tilde{\Phi}_{k_\nu, \ell}^i \left((\tilde{\Phi}_{k_\nu, \ell}^i)^H \tilde{\Phi}_{k_\nu, \ell}^i + \alpha I_{n_t} \right)^{-1} (\tilde{\Phi}_{k_\nu, \ell}^i)^H e_1 \\ &= \tilde{\Phi}_{k_\nu, \ell}^i (\tilde{\Phi}_{k_\nu, \ell}^i)^H \left(\tilde{\Phi}_{k_\nu, \ell}^i (\tilde{\Phi}_{k_\nu, \ell}^i)^H + \alpha I_{n_\nu+1} \right)^{-1} e_1 \\ &\approx I_{n_\nu+1} e_1 \end{aligned} \quad (15)$$

$$= e_1, \quad (16)$$

implying that the power of $\lambda_{\nu, \ell}^i(s_{\nu, \ell})$ can be still amplified, while its interference inflicted upon UE $k_{\nu'}$ ($\nu' \neq \nu$) is

² P is the power budget

approximately forced to zero. When the rank of $\tilde{\Phi}_{k\nu,\ell}^i$ is $(n_\nu + 1)$, which only occurs for $n_t > n_\nu + 1$, the perfect zero-interference condition of $\lambda_{\nu,\ell}^i(s_{\nu,\ell})$ is achieved.

It should be noted that (14) represents a brand new class of RZFB specifically tailored for mitigating the inter-content interference only and as such it has quite a different structure compared to the traditional RZFB (see e.g. [28]), which aims for suppressing the multi-user interference.

Under the RZFB of (13), the signal ξ_i defined in (4) and transferred from the BBU to RRH i through the limited-rate backhaul link is reformulated as:

$$\xi = \sum_{(\nu,\ell) \in \Xi^i} \left(\sum_{k=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k\nu,\ell}^i G_{k\nu,\ell}^i \right) s_{\nu,\ell} + v_i. \quad (17)$$

Then, the mutual information (MI) in the right-hand side (RHS) of (6) may be expressed as:

$$\begin{aligned} \mathcal{I} \left(\xi_i, \sum_{(\nu,\ell) \in \Xi^i} \left(\sum_{k=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k\nu,\ell}^i G_{k\nu,\ell}^i \right) s_{\nu,\ell} \right) \\ = \ln \left| I_{n_t} + L_i(\mathbf{p}) \left([\Delta(\tilde{\mathbf{x}}^i)]^2 + \epsilon I_{n_t} \right)^{-1} \right| \\ = \ln \left| I_{n_t} + \epsilon^{-1} L_i(\mathbf{p}) \left([\Delta(\mathbf{x}^i)]^2 + I_{n_t} \right)^{-1} \right| \\ \triangleq m_i(\mathbf{p}, \mathbf{x}^i), \end{aligned} \quad (18)$$

for $\mathbf{p} \triangleq \{\mathbf{p}_{k\nu,\ell}^i\} : (\nu,\ell) \in \mathcal{N}_i, k_\nu = 1, \dots, |\mathcal{K}_\nu|, i = 1, \dots, N_R\}$, and

$$L_i(\mathbf{p}) \triangleq \sum_{(\nu,\ell) \in \Xi^i} \left[\sum_{k=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k\nu,\ell}^i G_{k\nu,\ell}^i \right]^2 \in \mathbb{C}^{n_t \times n_t},$$

and

$$\mathbf{x}^i = \tilde{\mathbf{x}}^i / \sqrt{\epsilon}. \quad (19)$$

Then the fronthaul-rate constraint (6) becomes:

$$m_i(\mathbf{p}, \mathbf{x}^i) \leq \log_2(e)C, i \in \mathcal{N}_R. \quad (20)$$

Observe that the unit in (18) is nat, while the unit of C is bit, so the factor $\log_2(e)$ in (20) converts its right-hand side (RHS) to nats for consistency with its left-hand side (LHS).

The transmit signal at RRH i is specialized to

$$\chi_i = \sum_{(\nu,\ell) \in \mathcal{N}_i} \left(\sum_{k=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k\nu,\ell}^i G_{k\nu,\ell}^i \right) s_{\nu,\ell} + v_i. \quad (21)$$

For each $\nu \in \mathcal{N}_{req}$ let \mathcal{L}_ν be the set of RRHs that do not have the entire file ν in their cache, i.e.

$$\mathcal{L}_\nu \triangleq \{i \in \mathcal{N}_R : c_{\nu,\ell}^i \neq 1\}. \quad (22)$$

The power consumption of transmitting the signal carrying f_ν from RRH i is expressed as:

$$\begin{aligned} \sum_{\ell: (\nu,\ell) \in \mathcal{N}_i} \left\| \sum_{k=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k\nu,\ell}^i G_{k\nu,\ell}^i \right\|^2 + \sum_{i \in \mathcal{L}_\nu} \sum_{j=1}^{n_t} [(\tilde{\mathbf{x}}_j^i)^2 + \epsilon] = \\ \sum_{\ell: (\nu,\ell) \in \mathcal{N}_i} \left\| \sum_{k=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k\nu,\ell}^i G_{k\nu,\ell}^i \right\|^2 + \epsilon \sum_{i \in \mathcal{L}_\nu} \sum_{j=1}^{n_t} [(\mathbf{x}_j^i)^2 + 1] \triangleq \end{aligned}$$

$$\pi_\nu^i(\mathbf{p}, \mathbf{x}), \quad (23)$$

for $\mathbf{x} \triangleq \{\mathbf{x}^i, i = 1, \dots, N_R\}$. The total power dissipated by delivering f_ν is given by the following convex quadratic function

$$\pi_\nu(\mathbf{p}, \mathbf{x}) \triangleq \sum_{i=1}^{N_R} \pi_\nu^i(\mathbf{p}, \mathbf{x}) + P_{non}, \quad (24)$$

where P_{non} is the total 'non-transmission' power dissipation at the RRHs formulated as $N_R n_t P_a$, where the antenna circuit power is P_a .

Let us now rewrite equation (8) of the signal received by user k_ν as

$$\begin{aligned} y_{k_\nu} = \sum_{\ell=1}^L \sum_{i: (\nu,\ell) \in \mathcal{N}_i} h_{k_\nu,i}^H \lambda_{\nu,\ell}^i(s_{\nu,\ell}) \\ + \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} \sum_{\ell=1}^L \sum_{i: (\nu',\ell) \in \mathcal{N}_i} \lambda_{\nu',\ell}^i(s_{\nu',\ell}) \\ + \sum_{i=1}^{N_R} h_{k_\nu,i}^H v_i + z_{k_\nu} \\ = \sum_{\ell=1}^L a_{k_\nu,\ell}(\mathbf{p}) s_{\nu,\ell} + \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} \sum_{\ell=1}^L b_{k_\nu,\ell,\nu'}(\mathbf{p}) s_{\nu',\ell} \\ + \sum_{i=1}^{N_R} h_{k_\nu,i}^H v_i + z_{k_\nu}, \end{aligned} \quad (25)$$

for

$$a_{k_\nu,\ell}(\mathbf{p}) \triangleq \sum_{i: (\nu,\ell) \in \mathcal{N}_i} \sum_{k'=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k'\nu,\ell}^i h_{k_\nu,i}^H G_{k'\nu,\ell}^i, \quad (26)$$

and

$$b_{k_\nu,\ell,\nu'}(\mathbf{p}) \triangleq \sum_{i: (\nu',\ell) \in \mathcal{N}_i} \sum_{k'=1}^{|\mathcal{K}_{\nu'}|} \mathbf{p}_{k'\nu',\ell}^i h_{k_\nu,i}^H G_{k'\nu',\ell}^i, \quad (27)$$

with

$$\mathbb{E} \{ |h_{k_\nu,i}^H v_i|^2 \} = \epsilon h_{k_\nu,i}^H \left([\Delta(\mathbf{x}^i)]^2 + I_{n_t} \right) h_{k_\nu,i}.$$

Upon employing successive interference cancelation (SIC), user k_ν subtracts the detected and remodulated signal of

$$\sum_{\ell'=1}^{\ell-1} a_{k_\nu,\ell'}(\mathbf{p}) s_{\nu,\ell'} \quad (28)$$

from the RHS of (25), yielding

$$\begin{aligned} \sum_{\ell'=\ell}^L a_{k_\nu,\ell'}(\mathbf{p}) s_{\nu,\ell'} + \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} \sum_{\ell=1}^L b_{k_\nu,\ell,\nu'}(\mathbf{p}) s_{\nu',\ell} \\ + \sum_{i=1}^{N_R} h_{k_\nu,i}^H v_i + z_{k_\nu} \end{aligned} \quad (29)$$

for detecting $s_{\nu,\ell}$ by considering

$$a_{k_\nu,\ell}(\mathbf{p}) s_{\nu,\ell} \quad (30)$$

in (29) as the signal of interest, and

$$\sum_{\ell'=\ell+1}^L a_{k_\nu, \ell'}(\mathbf{p}) s_{\nu, \ell'} + \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} \sum_{\ell=1}^L b_{k_\nu, \ell, \nu'}(\mathbf{p}) s_{\nu', \ell} + \sum_{i=1}^{N_R} h_{k_\nu, i}^H v_i + z_{k_\nu} \quad (31)$$

as the interference-plus-noise term. As such, the throughput of $s_{\nu, \ell}$ at user k_ν is formulated as

$$r_{k_\nu, \ell}(\mathbf{p}, \mathbf{x}) \triangleq \ln \left(1 + \frac{|a_{k_\nu, \ell}(\mathbf{p})|^2}{\psi_{k_\nu, \ell}(\mathbf{p}, \mathbf{x})} \right), \quad (32)$$

for

$$\begin{aligned} \psi_{k_\nu, \ell}(\mathbf{p}, \mathbf{x}) &\triangleq \sum_{\ell'=\ell+1}^L |a_{k_\nu, \ell'}(\mathbf{p})|^2 + \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} \sum_{\ell=1}^L |b_{k_\nu, \ell, \nu'}(\mathbf{p})|^2 \\ &+ \epsilon \sum_{i=1}^{N_R} h_{k_\nu, i}^H \left([\Delta(\mathbf{x}^i)]^2 + I_{n_t} \right) h_{k_\nu, i} + \sigma, \end{aligned} \quad (33)$$

which is a convex quadratic function.

The throughput of $s_{\nu, \ell}$ after SIC becomes:

$$r_{\nu, \ell}(\mathbf{p}, \mathbf{x}) \triangleq \min_{k=1, \dots, |\mathcal{K}_\nu|} r_{k_\nu, \ell}(\mathbf{p}, \mathbf{x}). \quad (34)$$

The energy efficiency optimization related problem may then be formulated as³:

$$\begin{aligned} \max_{\mathbf{p}, \mathbf{x}} \theta(\mathbf{p}, \mathbf{x}) &\triangleq \min_{\nu=1, \dots, N_{req}} \frac{\sum_{\ell=1}^L r_{\nu, \ell}(\mathbf{p}, \mathbf{x})}{\pi_\nu(\mathbf{p}, \mathbf{x})} \\ &\text{s.t. (20),} \end{aligned} \quad (35a)$$

$$r_{\min} \leq r_{\nu, \ell}(\mathbf{p}, \mathbf{x}), (\nu, \ell) \in \mathcal{N}_{req} \times \mathcal{L}, \quad (35b)$$

$$\sum_{\nu=1}^{N_{req}} \pi_\nu^i(\mathbf{p}, \mathbf{x}) \leq P, i \in \mathcal{N}_R, \quad (35c)$$

where P is the maximum power budget of each RRH, and r_{\min} is the minimum rate required for guaranteeing the delivery of all files requested within a certain delivery deadline [29]. As such (35b) constitutes the rate constraint to ensure a high-probability successful delivery according to Shannonian information theory.

B. PGS Computation

Note that this problem is non-convex because the objective function (OF) in (35a) is not only non-smooth but also non-concave, while both the SFT constraint (20) and the rate-constraint (35b) are non-convex. We now develop a path-following algorithm, which iterates for finding ever better feasible points for (35) with the aid of convex solvers. To this end, we have to derive (i) an inner convex approximation for the SFT constraint (20) and the rate-constraint (35b); and (ii) a lower-bounding concave approximation for the OF in (35a).

Let $(p^{(\kappa)}, x^{(\kappa)})$ be the specific feasible point for (35) that is found by the κ th iteration.

³The final result should be divided by $\log_2(e)$ to express the energy efficiency in terms of bps/Hz/W

1) *Inner convex approximation for the SFT constraint (20):* Using the inequality (86) in the appendix yields

$$\begin{aligned} m_i(\mathbf{p}, \mathbf{x}^i) &\leq m_i(p^{(\kappa)}, x^{i, (\kappa)}) + \epsilon \langle \left(\epsilon I_{n_t} + L_i(p^{(\kappa)}) + \epsilon [\Delta(x^{i, (\kappa)})]^2 \right)^{-1} \rangle \\ &+ \langle [\Delta(x^{i, (\kappa)})]^2 \rangle - \langle \left([\Delta(x^{i, (\kappa)})]^2 + I_{n_t} \right)^{-1} \rangle \\ &+ \langle \left(\epsilon I_{n_t} + L_i(p^{(\kappa)}) + \epsilon [\Delta(x^{i, (\kappa)})]^2 \right)^{-1}, \\ &L_i(\mathbf{p}) + \epsilon [\Delta(\mathbf{x}^i)]^2 - 2\langle \Delta(x^{i, (\kappa)}), \Delta(\mathbf{x}^i) \rangle \rangle \\ &+ \langle I_{n_t} - \left([\Delta(x^{i, (\kappa)})]^2 + I_{n_t} \right)^{-1}, [\Delta(\mathbf{x}^i)]^2 \rangle \end{aligned} \quad (36)$$

$$\begin{aligned} &= m_i^{(\kappa)} + \sum_{(\nu, \ell) \in \Xi^i} \left\| \left(\epsilon I_{n_t} + L_i(p^{(\kappa)}) + \epsilon [\Delta(x^{i, (\kappa)})]^2 \right)^{-1/2} \right. \\ &\quad \times \sum_{k=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k_\nu, \ell}^i G_{k_\nu, \ell}^i \left. \right\|^2 \\ &+ \epsilon \left\| \left(\epsilon I_{n_t} + L_i(p^{(\kappa)}) + \epsilon [\Delta(x^{i, (\kappa)})]^2 \right)^{-1/2} \Delta(\mathbf{x}^i) \right\|^2 \\ &- 2\langle \Delta(x^{i, (\kappa)}), \Delta(\mathbf{x}^i) \rangle \\ &+ \left\| \left(I_{n_t} - \left([\Delta(x^{i, (\kappa)})]^2 + I_{n_t} \right)^{-1} \right)^{-1/2} \Delta(\mathbf{x}^i) \right\|^2 \end{aligned} \quad (37)$$

$$\triangleq m_i^{(\kappa)}(\mathbf{p}, \mathbf{x}^i), \quad (38)$$

for

$$\begin{aligned} m_i^{(\kappa)} &\triangleq m_i(p^{(\kappa)}, x^{i, (\kappa)}) \\ &+ \epsilon \langle \left(\epsilon I_{n_t} + L_i(p^{(\kappa)}) + \epsilon [\Delta(x^{i, (\kappa)})]^2 \right)^{-1} \rangle \\ &+ \langle [\Delta(x^{i, (\kappa)})]^2 \rangle - \langle \left([\Delta(x^{i, (\kappa)})]^2 + I_{n_t} \right)^{-1} \rangle. \end{aligned}$$

The function $m_i^{(\kappa)}(\mathbf{p}, \mathbf{x}^i)$ is convex quadratic. The non-convex constraint (20) seen in (35) is then innerly approximated by the following convex constraint:

$$m_i^{(\kappa)}(\mathbf{p}, \mathbf{x}^i) \leq \log_2(e)C, i \in \mathcal{N}_R. \quad (39)$$

2) *Inner convex quadratic approximation for the rate-constraint (35b):* Furthermore, applying the inequality (85) of the Appendix yields

$$\begin{aligned} r_{k_\nu, \ell}(\mathbf{p}, \mathbf{x}) &\geq a_{k_\nu, \ell}^{(\kappa)} + 2\Re\{b_{k_\nu, \ell}^{(\kappa)} a_{k_\nu, \ell}(\mathbf{p})\} \\ &- c_{k_\nu, \ell}^{(\kappa)} (|a_{k_\nu, \ell}(\mathbf{p})|^2 + \psi_{k_\nu, \ell}(\mathbf{p}, \mathbf{x})) \\ &\triangleq r_{k_\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) \end{aligned} \quad (40)$$

$$\triangleq r_{k_\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) \quad (41)$$

in conjunction with

$$a_{k_\nu, \ell}^{(\kappa)} \triangleq r_{k_\nu, \ell}(p^{(\kappa)}, x^{(\kappa)}) - \frac{|a_{k_\nu, \ell}(p^{(\kappa)})|^2}{\psi_{k_\nu, \ell}(p^{(\kappa)}, x^{(\kappa)})},$$

and

$$b_{k_\nu, \ell}^{(\kappa)} \triangleq \frac{a_{k_\nu, \ell}^H(p^{(\kappa)})}{\psi_{k_\nu, \ell}(p^{(\kappa)}, x^{(\kappa)})},$$

as well as

$$0 < c_{k\nu,\ell}^{(\kappa)} \triangleq \frac{|a_{k\nu,\ell}(p^{(\kappa)})|^2}{\psi_{k\nu,\ell}(p^{(\kappa)}, x^{(\kappa)})} \times \frac{1}{(|a_{k\nu,\ell}(p^{(\kappa)})|^2 + \psi_{k\nu,\ell}(p^{(\kappa)}, x^{(\kappa)})}.$$

Since the function $r_{k\nu,\ell}^{(\kappa)}(\mathbf{p}, \mathbf{x})$ is concave quadratic, the following function provides a concave lower-bounding approximation for $r_{\nu,\ell}(\mathbf{p}, \Delta)$:

$$r_{\nu,\ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) \triangleq \min_{k=1,\dots,|\mathcal{K}_\nu|} r_{k\nu,\ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}),$$

leading to the following inner convex quadratic approximation for the non-convex constraint (35b) in (35)

$$r_{\nu,\ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) \geq r_{\min}(\nu, \ell) \in \mathcal{N}_{req} \times \mathcal{L}. \quad (42)$$

3) *PGS Algorithm*: We now solve the following convex optimization problem at the κ -th iteration for generating the next feasible point $(p^{(\kappa+1)}, x^{(\kappa+1)})$ for (35)

$$\max_{\mathbf{p}, \mathbf{x}} \min_{\nu=1,\dots,N_{req}} \left[\sum_{\ell=1}^L r_{\nu,\ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) - \theta(p^{(\kappa)}, x^{(\kappa)}) \pi_\nu(\mathbf{p}, \mathbf{x}) \right] \quad \text{s.t.} \quad (35c), (39), (42), \quad (43)$$

which is equivalent to the following convex quadratic problem:

$$\max_{\mathbf{p}, \mathbf{x}, \gamma} \gamma \quad \text{s.t.} \quad (35c), (39), (42), \quad \sum_{\ell=1}^L r_{\nu,\ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) - \theta(p^{(\kappa)}, x^{(\kappa)}) \pi_\nu(\mathbf{p}, \mathbf{x}) \geq \gamma, \nu = 1, \dots, N_{req}.$$

The computational complexity of (43), which involves $n_v = \sum_{i=1}^{N_R} |\mathcal{N}_i| + N_R n_t + 1$ decision variables and $n_c = 2N_R + N_{req}(L+1)$ constraints is on the order of [30, p.4]

$$\mathcal{O} \{ (n_c)^{2.5} [(n_v)^2 + n_c] \}. \quad (44)$$

As $(p^{(\kappa+1)}, x^{(\kappa+1)})$ and $(p^{(\kappa)}, x^{(\kappa)})$ are the optimal solution and a feasible point for (43), we have

$$\min_{\nu=1,\dots,N_{req}} \left[\sum_{\ell=1}^L r_{\nu,\ell}^{(\kappa)}(p^{(\kappa+1)}, x^{(\kappa+1)}) - \theta(p^{(\kappa)}, x^{(\kappa)}) \pi_\nu(p^{(\kappa+1)}, x^{(\kappa+1)}) \right] > \min_{\nu=1,\dots,N_{req}} \left[\sum_{\ell=1}^L r_{\nu,\ell}^{(\kappa)}(p^{(\kappa)}, x^{(\kappa)}) - \theta(p^{(\kappa)}, x^{(\kappa)}) \pi_\nu(p^{(\kappa)}, x^{(\kappa)}) \right] = 0.$$

Therefore, we have:

$$\sum_{\ell=1}^L r_{\nu,\ell}^{(\kappa)}(p^{(\kappa+1)}, x^{(\kappa+1)}) - \theta(p^{(\kappa)}, x^{(\kappa)}) \pi_\nu(p^{(\kappa+1)}, x^{(\kappa+1)}) > 0 \quad \forall \nu = 1, \dots, N_{req}$$

that yields

$$\theta(p^{(\kappa+1)}, x^{(\kappa+1)}) > \theta(p^{(\kappa)}, x^{(\kappa)}), \quad (45)$$

i.e. $(p^{(\kappa+1)}, x^{(\kappa+1)})$ provides a better feasible point for the nonconvex problem (35) than $(p^{(\kappa)}, x^{(\kappa)})$, provided that $(p^{(\kappa+1)}, x^{(\kappa+1)}) \neq (p^{(\kappa)}, x^{(\kappa)})$. As such, the sequence $\{(p^{(\kappa)}, x^{(\kappa)})\}$ converges to a locally optimal solution of (35)

Algorithm 1 RZFB PGS Algorithm

- 1: **Set** $\kappa = 0$. Iterate (46) for a feasible point $(p^{(0)}, x^{(0)})$ for (35).
 - 2: **Repeat until convergence**. Solve the convex optimization problem (43) to generate the next feasible point $(p^{(\kappa+1)}, x^{(\kappa+1)})$ for (35). Set $\kappa := \kappa + 1$.
 - 3: **Output** $(p^{(\kappa)}, x^{(\kappa)})$ as the optimal solution of (35).
-

[31]. However, in the scenario of [32], this locally optimal solution turned out to be the globally optimal one. Algorithm 1 provides the pseudo-code of this computational procedure.

To find a feasible initial point for (35), initialized by any feasible point for the convex constraints (35c), we iterate

$$\max_{\mathbf{p}, \mathbf{x}} \min \left\{ \min_{(\nu,\ell) \in \mathcal{N}_{req} \times \mathcal{L}} \left[\frac{r_{\nu,\ell}^{(\kappa)}(\mathbf{p}, \mathbf{x})}{r_{\min}/\log_2(e)} - 1 \right], \min_{i \in \mathcal{N}_R} \left[1 - \frac{1}{\log_2(e)C} m_i^{(\kappa)}(\mathbf{p}, \mathbf{x}^i) \right] \right\} \quad \text{s.t.} \quad (35c), \quad (46)$$

until the value of the objective in (46) becomes more than 0 or equal to 0. The problem (46) is equivalent to the following convex quadratic problem:

$$\max_{\mathbf{p}, \mathbf{x}, \gamma} \gamma \quad \text{s.t.} \quad (35c), \quad r_{\nu,\ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) \geq r_{\min}(1 + \gamma)/\log_2(e), (\nu, \ell) \in \mathcal{N}_{req} \times \mathcal{L}, \quad m_i^{(\kappa)}(\mathbf{p}, \mathbf{x}^i) \leq \log_2(e)C(1 - \gamma), i \in \mathcal{N}_R.$$

Remark 1. Instead of the energy efficiency optimization problem (35) one can consider the following problem of max-min rate optimization

$$\max_{\mathbf{p}, \mathbf{x}} \min_{(\nu,\ell) \in \mathcal{N}_{req} \times \mathcal{L}} r_{\nu,\ell}(\mathbf{p}, \mathbf{x}) \quad \text{s.t.} \quad (20), (35c). \quad (47)$$

Then, instead of (43), the following problem is solved at the κ -th iteration to generate a better feasible point $(p^{(\kappa+1)}, x^{(\kappa+1)})$:

$$\max_{\mathbf{p}, \mathbf{x}} \min_{(\nu,\ell) \in \mathcal{N}_{req} \times \mathcal{L}} r_{\nu,\ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) \quad \text{s.t.} \quad (35c), (39), \quad (48)$$

while for an initial feasible point, we choose a reasonable r_{\min} and then iterate (46).

IV. RZFB BASED IMPROPER GAUSSIAN SIGNALING

A. IGS Problem statement

Observe that the ‘beamformed’ signal $\lambda_{\nu,\ell}^i(s_{\nu,\ell})$ defined by (14) is proper Gaussian ($\mathbb{E}((\lambda_{\nu,\ell}^i(s_{\nu,\ell}))^2) = 0$), because so is the information source $s_{\nu,\ell} \in \mathcal{CN}(0, 1)$. As such, the RZFB technique of the previous section represents proper Gaussian signaling (PGS).

In this section, we propose the following new class of RZFB

$$\lambda_{\nu,\ell}^i(s_{\nu,\ell}) = \sum_{k=1}^{|\mathcal{K}_\nu|} \left(\mathbf{p}_{k\nu,\ell}^{i,1} G_{k\nu,\ell}^i s_{\nu,\ell} + \mathbf{p}_{k\nu,\ell}^{i,2} G_{k\nu,\ell}^i s_{\nu,\ell}^* \right) \quad (49) = \left(\sum_{k=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k\nu,\ell}^{i,1} G_{k\nu,\ell}^i \right) s_{\nu,\ell} + \left(\sum_{k=1}^{|\mathcal{K}_\nu|} \mathbf{p}_{k\nu,\ell}^{i,2} G_{k\nu,\ell}^i \right) s_{\nu,\ell}^*,$$

$$\mathbf{p}_{k_\nu, \ell}^i \triangleq (\mathbf{p}_{k_\nu, \ell}^{i,1}, \mathbf{p}_{k_\nu, \ell}^{i,2}) \in \mathbb{C}^2, \quad (50)$$

where $\lambda_{\nu, \ell}^i(s_{\nu, \ell})$ is an improper Gaussian signal ($\mathbb{E}((\lambda_{\nu, \ell}^i(s_{\nu, \ell}))^2) \neq 0$) capable of offering additional degree of freedom for improving the throughput of the proper Gaussian information source $s_{\nu, \ell}$ of section III [33]. Explicitly, this is regarded as improper Gaussian signaling (IGS) in contrast to the proper Gaussian signaling (PGS) scheme of (14), which has been shown to enhance the throughput of severely interference-limited networks [34]–[36]. The reader is also referred to [37] for another instantiation of IGS based (traditional) RZFB.

Based on the IGS philosophy of (50), we rewrite (4) in an form as:

$$\xi_i^A \triangleq \begin{bmatrix} \xi_i \\ \xi_i^* \end{bmatrix} = \sum_{(\nu, \ell) \in \Xi^i} \left(\sum_{k=1}^{|\mathcal{K}_\nu|} \mathcal{G}_{k_\nu, \ell}^i(\mathbf{p}_{k_\nu, \ell}^i) \right) s_{\nu, \ell}^A + v_i^A, \quad (51)$$

for

$$\begin{aligned} s_{\nu, \ell}^A &= \begin{bmatrix} s_{\nu, \ell} \\ s_{\nu, \ell}^* \end{bmatrix}, v_i^A = \begin{bmatrix} v_i \\ v_i^* \end{bmatrix}, \\ \mathcal{G}_{k_\nu, \ell}^i(\mathbf{p}_{k_\nu, \ell}^i) &\triangleq \begin{bmatrix} \mathbf{p}_{k_\nu, \ell}^{i,1} G_{k_\nu, \ell}^i & \mathbf{p}_{k_\nu, \ell}^{i,2} G_{k_\nu, \ell}^i \\ (\mathbf{p}_{k_\nu, \ell}^{i,2})^* (G_{k_\nu, \ell}^i)^* & (\mathbf{p}_{k_\nu, \ell}^{i,1})^* (G_{k_\nu, \ell}^i)^* \end{bmatrix} \in \mathbb{C}^{(2n_t) \times 2}. \end{aligned} \quad (52)$$

For $\Delta^A(\tilde{\mathbf{x}}^i) \triangleq \text{diag}[\Delta(\tilde{\mathbf{x}}^i), \Delta(\tilde{\mathbf{x}}^i)] \in \mathbb{R}^{(2n_t) \times (2n_t)}$, $\mathbf{p}^i \triangleq \{\mathbf{p}_{k_\nu, \ell}^i, (\nu, \ell) \in \mathcal{N}_i, i = 1, \dots, N_R, k_\nu = 1, \dots, |\mathcal{K}_\nu|\}$, and

$$\Lambda_i(\mathbf{p}^i) \triangleq \sum_{(\nu, \ell) \in \Xi^i} \left[\sum_{k=1}^{|\mathcal{K}_\nu|} \mathcal{G}_{k_\nu, \ell}^i(\mathbf{p}_{k_\nu, \ell}^i) \right]^2 \in \mathbb{C}^{(2n_t) \times (2n_t)},$$

the mutual information between $\sum_{(\nu, \ell) \in \Xi^i} \lambda_{\nu, \ell}^i(s_{\nu, \ell})$ and quantized version ξ is [38]

$$\begin{aligned} &\frac{1}{2} \mathcal{I} \left(\xi_i^A, \sum_{(\nu, \ell) \in \Xi^i} \left(\sum_{k=1}^{|\mathcal{K}_\nu|} \mathcal{G}_{k_\nu, \ell}^i(\mathbf{p}_{k_\nu, \ell}^i) \right) s_{\nu, \ell}^A \right) \\ &= \frac{1}{2} \ln \left| I_{2n_t} + \Lambda_i(\mathbf{p}^i) \left([\Delta^A(\tilde{\mathbf{x}}^i)]^2 + \epsilon I_{2n_t} \right)^{-1} \right| \\ &= \frac{1}{2} \ln \left| I_{2n_t} + \epsilon^{-1} \Lambda_i(\mathbf{p}^i) \left([\Delta^A(\mathbf{x}^i)]^2 + I_{2n_t} \right)^{-1} \right| \\ &\triangleq \mu_i(\mathbf{p}, \mathbf{x}^i), \end{aligned} \quad (53)$$

for

$$\mathbf{x}^i = \tilde{\mathbf{x}}^i / \sqrt{\epsilon}. \quad (54)$$

Similarly to the PGS expression of (20), the fronthaul-rate constraint (6) is reformulated as:

$$\mu_i(\mathbf{p}, \mathbf{x}^i) \leq \log_2(e)C, i \in \mathcal{N}_R. \quad (55)$$

Instead of (21), the signal transmitted by RRH i now becomes

$$\chi_i = \sum_{(\nu, \ell) \in \mathcal{N}_i} \left(\sum_{k=1}^{|\mathcal{K}_\nu|} \left[\mathbf{p}_{k_\nu, \ell}^{i,1} G_{k_\nu, \ell}^i \quad \mathbf{p}_{k_\nu, \ell}^{i,2} G_{k_\nu, \ell}^i \right] \right) s_{\nu, \ell}^A + v_i. \quad (56)$$

Instead of (23), the power consumption of transmitting the signal carrying f_ν from RRH i is

$$\begin{aligned} &\frac{1}{2} \sum_{i=1}^{N_R} \sum_{\ell: (\nu, \ell) \in \mathcal{N}_i} \left\| \sum_{k=1}^{|\mathcal{K}_\nu|} \mathcal{G}_{k_\nu, \ell}^i(\mathbf{p}_{k_\nu, \ell}^i) \right\|^2 + \sum_{i \in \nu} \sum_{j=1}^{n_t} [(\tilde{\mathbf{x}}_j^i)^2 + \epsilon] = \\ &\frac{1}{2} \sum_{i=1}^{N_R} \sum_{\ell: (\nu, \ell) \in \mathcal{N}_i} \left\| \sum_{k=1}^{|\mathcal{K}_\nu|} \mathcal{G}_{k_\nu, \ell}^i(\mathbf{p}_{k_\nu, \ell}^i) \right\|^2 + \epsilon \sum_{i \in \nu} \sum_{j=1}^{n_t} [(x_j^i)^2 + 1] \triangleq \\ &\tilde{\pi}_\nu^i(\mathbf{p}, \mathbf{x}), \end{aligned} \quad (57)$$

for $\mathbf{x} \triangleq \{\mathbf{x}^i, i = 1, \dots, N_R\}$, where ν is defined from (22) as the set of RRHs that do not have the entire file ν in their cache.

Instead of (24), the total power consumption of delivering f_ν becomes

$$\tilde{\pi}_\nu(\mathbf{p}, \mathbf{x}) \triangleq \sum_{i=1}^{N_R} \tilde{\pi}_\nu^i(\mathbf{p}, \mathbf{x}) + P_{non}. \quad (58)$$

Under the RZFB (50), the augmented form of equation (8) formulating the signal received by user k_ν is

$$\begin{aligned} y_{k_\nu}^A &\triangleq \begin{bmatrix} y_{k_\nu} \\ y_{k_\nu}^* \end{bmatrix} \\ &= \sum_{\ell=1}^L \mathcal{A}_{k_\nu, \ell}(\mathbf{p}) s_{\nu, \ell}^A + \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} \sum_{\ell=1}^L \mathcal{B}_{k_\nu, \ell, \nu'}(\mathbf{p}) s_{\nu', \ell}^A \\ &\quad + \sum_{i=1}^{N_R} \mathcal{H}_{k_\nu, i} v_i^A + z_{k_\nu}^A, \end{aligned} \quad (59)$$

for

$$\mathcal{H}_{k_\nu, i} \triangleq \begin{bmatrix} h_{k_\nu, i}^H & 0_{1 \times n_t} \\ 0_{1 \times n_t} & h_{k_\nu, i}^T \end{bmatrix} \in \mathbb{C}^{2 \times (2n_t)}, z_{k_\nu}^A \triangleq \begin{bmatrix} z_{k_\nu} \\ z_{k_\nu}^* \end{bmatrix} \quad (61)$$

and

$$\mathcal{A}_{k_\nu, \ell}(\mathbf{p}) \triangleq \sum_{i: (\nu, \ell) \in \mathcal{N}_i} \mathcal{H}_{k_\nu, i} \sum_{k'=1}^{|\mathcal{K}_\nu|} \mathcal{G}_{k_\nu, \ell}^i(\mathbf{p}_{k_\nu, \ell}^i) \in \mathbb{C}^{2 \times 2}, \quad (62)$$

as well as

$$\mathcal{B}_{k_\nu, \ell, \nu'}(\mathbf{p}) \triangleq \sum_{i: (\nu', \ell) \in \mathcal{N}_i} \mathcal{H}_{k_\nu, i} \sum_{k'=1}^{|\mathcal{K}_{\nu'}|} \mathcal{G}_{k_\nu, \ell}^i(\mathbf{p}_{k_\nu, \ell}^i) \in \mathbb{C}^{2 \times 2}. \quad (63)$$

It is readily shown that

$$\mathbb{E} \{ [\mathcal{H}_{k_\nu, i}^H v_i^A]^2 \} = \epsilon \mathcal{H}_{k_\nu, \ell} \left([\Delta^A(\mathbf{x}^i)]^2 + I_{2n_t} \right) \mathcal{H}_{k_\nu, \ell}^H.$$

In SIC, user k_ν subtracts the detected and remodulated signal

$$\sum_{\ell'=1}^{\ell-1} \mathcal{A}_{k_\nu, \ell'}(\mathbf{p}) s_{\nu, \ell'}^A \quad (64)$$

from the RHS of (60), yielding

$$\begin{aligned} &\sum_{\ell'=1}^L \mathcal{A}_{k_\nu, \ell'}(\mathbf{p}) s_{\nu, \ell'}^A + \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} \sum_{\ell=1}^L \mathcal{B}_{k_\nu, \ell, \nu'}(\mathbf{p}) s_{\nu', \ell}^A \\ &\quad + \sum_{i=1}^{N_R} \mathcal{H}_{k_\nu, i} v_i^A + z_{k_\nu}^A \end{aligned} \quad (65)$$

for detecting $s_{\nu,\ell}$ by considering

$$\mathcal{A}_{k_{\nu},\ell}(\mathbf{p})s_{\nu,\ell}^A \quad (66)$$

in (65) as the signal of interest, and

$$\begin{aligned} \sum_{\ell'=\ell+1}^L \mathcal{A}_{k_{\nu},\ell'}(\mathbf{p})s_{\nu,\ell'}^A + \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} \sum_{\ell=1}^L \mathcal{B}_{k_{\nu},\ell,\nu'}(\mathbf{p})s_{\nu',\ell}^A \\ + \sum_{i=1}^{N_R} \mathcal{H}_{k_{\nu},i}v_i^A + z_{k_{\nu}}^A \end{aligned} \quad (67)$$

as the interference-plus-noise. As such the throughput $s_{f_{\nu},\ell}$ of user k_{ν} is calculated as [39]

$$\begin{aligned} \frac{1}{2} \mathcal{I} \left(\mathcal{A}_{k_{\nu},\ell}(\mathbf{p})s_{\nu,\ell}^A, \sum_{\ell'=\ell+1}^L \mathcal{A}_{k_{\nu},\ell'}(\mathbf{p})s_{\nu,\ell'}^A \right. \\ \left. + \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} \sum_{\ell=1}^L \mathcal{B}_{k_{\nu},\ell,\nu'}(\mathbf{p})s_{\nu',\ell}^A + \sum_{i=1}^{N_R} \mathcal{H}_{k_{\nu},i}v_i + z_{k_{\nu}} \right). \end{aligned} \quad (68)$$

Then, based on [33] we have:

$$\rho_{k_{\nu},\ell}(\mathbf{p}, \mathbf{x}) \triangleq \ln \left| I_2 + [\mathcal{A}_{k_{\nu},\ell}(\mathbf{p})]^2 \Psi_{k_{\nu},\ell}^{-1}(\mathbf{p}, \mathbf{x}) \right|, \quad (69)$$

where

$$\begin{aligned} \Psi_{k_{\nu},\ell}(\mathbf{p}, \mathbf{x}) \\ \triangleq \sum_{\ell'=\ell+1}^L [\mathcal{A}_{k_{\nu},\ell'}(\mathbf{p})]^2 + \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} \sum_{\ell=1}^L [\mathcal{B}_{k_{\nu},\ell,\nu'}(\mathbf{p})]^2 \\ + \epsilon \sum_{i=1}^{N_R} \mathcal{H}_{k_{\nu},i} \left([\Delta^A(\mathbf{x}^i)]^2 + I_{2n_t} \right) \mathcal{H}_{k_{\nu},i}^H + \sigma I_2. \end{aligned} \quad (70)$$

The throughput of $s_{\nu,\ell}$ attained by SIC thus becomes:

$$\frac{1}{2} \rho_{\nu,\ell}(\mathbf{p}, \mathbf{x}), \quad (71)$$

where we have:

$$\rho_{\nu,\ell}(\mathbf{p}, \mathbf{x}) = \min_{k=1, \dots, |\mathcal{K}_{\nu}|} \rho_{k_{\nu},\ell}(\mathbf{p}, \mathbf{x}). \quad (72)$$

The problem of IGS energy efficiency optimization is reminiscent of its PGS counterpart in (35), which is formulated as⁴:

$$\begin{aligned} \max_{\mathbf{p}, \mathbf{x}} \Theta(\mathbf{p}, \mathbf{x}) \triangleq \min_{\nu=1, \dots, N_{req}} \frac{\sum_{\ell=1}^L \rho_{\nu,\ell}(\mathbf{p}, \mathbf{x})}{2\tilde{\pi}_{\nu}(\mathbf{p}, \mathbf{x})} \\ \text{s.t. (5), (55),} \end{aligned} \quad (73a)$$

$$2r_{\min} \leq \rho_{\nu,\ell}(\mathbf{p}, \mathbf{x}), (\nu, \ell) \in \mathcal{N}_{req} \times \mathcal{L}, \quad (73b)$$

$$\sum_{\nu=1}^{N_{req}} \tilde{\pi}_{\nu}^i(\mathbf{p}, \mathbf{x}) \leq P, i \in \mathcal{N}_R. \quad (73c)$$

This is non-convex, because the OF of (73a) is both non-smooth and non-concave, while both the fronthaul-rate constraint (55) and (73b) are non-convex. In contrast to the PGS throughput defined by (32), which is a logarithmic function,

⁴Again the final result should be divided by $\log_2(e)$ to convert the energy-efficiency to bps/Hz/W

the IGS throughput defined by (69) is a log-determinant function. As a result, the computational algorithms developed in the previous section are not applicable to address the problem (73).

B. IGS Computation

Similarly to PGS, we now derive an inner convex approximation for both (55) as well as (73b) and a concave lower-bounding approximation for the OF in (73a) that are used for a path-following computational procedure to generate a sequence of gradually improved feasible points for (73).

Let $(p^{(\kappa)}, x^{(\kappa)})$ be the specific feasible point for (73) that is found by the κ th iteration.

1) *Inner convex quadratic approximation for the SFT constraint (55)*: Similarly to (38), we use the inequality (86) of the Appendix for obtaining the following convex quadratic upper-bounding approximation for the function defined in (53):

$$\begin{aligned} \mu_i(\mathbf{p}, \mathbf{x}^i) \\ \leq \mu_i^{(\kappa)} + \sum_{(\nu,\ell) \in \Xi^i} \left\| \left(\epsilon I_{n_t} + \Lambda_i(p^{(\kappa)}) + \epsilon \left[\Delta(x^{i,(\kappa)}) \right]^2 \right)^{-1/2} \right. \\ \left. \times \sum_{k=1}^{|\mathcal{K}_{\nu}|} \mathcal{G}_{k_{\nu},\ell}^i(\mathbf{p}_{k_{\nu},\ell}^i) \right\|^2 \\ + \epsilon \left\| \left(\epsilon I_{2n_t} + \Lambda_i(p^{(\kappa)}) + \epsilon \left[\Delta^A(x^{i,(\kappa)}) \right]^2 \right)^{-1/2} \Delta^A(\mathbf{x}^i) \right\|^2 \\ - 2 \langle \Delta^A(x^{i,(\kappa)}), \Delta^A(\mathbf{x}^i) \rangle \\ + \left\| \left(I_{2n_t} - \left(\left[\Delta^A(x^{i,(\kappa)}) \right]^2 + I_{2n_t} \right)^{-1} \right)^{-1/2} \Delta^A(\mathbf{x}^i) \right\|^2 \end{aligned} \quad (74)$$

$$\triangleq \mu_i^{(\kappa)}(\mathbf{p}, \mathbf{x}^i), \quad (75)$$

for

$$\begin{aligned} \mu_i^{(\kappa)} \triangleq \mu_i(p^{(\kappa)}, x^{i,(\kappa)}) \\ + \frac{1}{2} \left[\epsilon \left\langle \left(\epsilon I_{2n_t} + \Lambda_i(p^{i,(\kappa)}) + \epsilon \left[\Delta^A(x^{i,(\kappa)}) \right]^2 \right)^{-1} \right. \right. \\ \left. \left. + \left\langle \left[\Delta^A(x^{i,(\kappa)}) \right]^2 \right\rangle - \left\langle \left(\left[\Delta^A(x^{i,(\kappa)}) \right]^2 + I_{2n_t} \right)^{-1} \right\rangle \right]. \end{aligned}$$

The non-convex constraint (55) in (73) is then innerly approximated by the following convex quadratic constraint:

$$\mu_i^{(\kappa)}(\mathbf{p}, \mathbf{x}^i) \leq \log_2(e)C, i \in \mathcal{N}_R. \quad (76)$$

2) *Inner convex quadratic approximation for the rate-constraint (73b)*: Furthermore, applying the inequality (84) of the Appendix yields

$$\begin{aligned} \rho_{k_{\nu},\ell}(\mathbf{p}, \mathbf{x}) \\ \geq a_{k_{\nu},\ell}^{(\kappa)} + 2\Re\{\langle B_{k_{\nu},\ell}^{(\kappa)} \mathcal{A}_{k_{\nu},\ell}(\mathbf{p}) \rangle\} \\ - \langle C_{k_{\nu},\ell}^{(\kappa)}, [\mathcal{A}_{k_{\nu},\ell}(\mathbf{p})]^2 + \Psi_{k_{\nu},\ell}(\mathbf{p}, \mathbf{x}) \rangle \\ = a_{k_{\nu},\ell}^{(\kappa)} + 2\Re\{\langle B_{k_{\nu},\ell}^{(\kappa)} \mathcal{A}_{k_{\nu},\ell}(\mathbf{p}) \rangle\} - \|(C_{k_{\nu},\ell}^{(\kappa)})^{1/2} \mathcal{A}_{k_{\nu},\ell}(\mathbf{p})\|^2 \end{aligned} \quad (77)$$

$$\begin{aligned}
& - \sum_{\ell'=1}^L \|(C_{k_\nu, \ell'}^{(\kappa)})^{1/2} \mathcal{A}_{k_\nu, \ell'}(\mathbf{p})\|^2 \\
& - \sum_{\nu' \in \mathcal{N}_{req} \setminus \{\nu\}} \sum_{\ell=1}^L \|(C_{k_\nu, \ell}^{(\kappa)})^{1/2} \mathcal{B}_{k_\nu, \ell, \nu'}(\mathbf{p})\|^2 \\
& - \epsilon \sum_{i=1}^{N_R} \|(C_{k_\nu, \ell}^{(\kappa)})^{1/2} \mathcal{H}_{k_\nu, i} \Delta^A(\mathbf{x}^i)\|^2 \\
& - \epsilon \sum_{i=1}^{N_R} \|(C_{k_\nu, \ell}^{(\kappa)})^{1/2} \mathcal{H}_{k_\nu, i}\|^2 - \sigma \langle C_{k_\nu, \ell}^{(\kappa)} \rangle, \\
& \triangleq \rho_{k_\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}), \tag{78}
\end{aligned}$$

with

$$a_{k_\nu, \ell}^{(\kappa)} \triangleq \rho_{k_\nu, \ell}(\mathbf{p}^{(\kappa)}, \mathbf{x}^{(\kappa)}) - \langle [\mathcal{A}_{k_\nu, \ell}(\mathbf{p}^{(\kappa)})]^2 \Psi_{k_\nu, \ell}^{-1}(\mathbf{p}^{(\kappa)}, \mathbf{x}^{(\kappa)}) \rangle,$$

and

$$B_{k_\nu, \ell}^{(\kappa)} \triangleq (\mathcal{A}_{k_\nu, \ell}(\mathbf{p}^{(\kappa)}))^H \Psi_{k_\nu, \ell}^{-1}(\mathbf{p}^{(\kappa)}, \mathbf{x}^{(\kappa)}),$$

as well as

$$\begin{aligned}
0 \preceq C_{k_\nu, \ell}^{(\kappa)} & \triangleq \Psi_{k_\nu, \ell}^{-1}(\mathbf{p}^{(\kappa)}, \mathbf{x}^{(\kappa)}) \\
& - \left([\mathcal{A}_{k_\nu, \ell}(\mathbf{p}^{(\kappa)})]^2 + \Psi_{k_\nu, \ell}(\mathbf{p}^{(\kappa)}, \mathbf{x}^{(\kappa)}) \right)^{-1}.
\end{aligned}$$

Since $\rho_{k_\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x})$ is a concave quadratic function, the following function provides a lower-bounding concave approximation for $\rho_{\nu, \ell}(\mathbf{p}, \Delta)$:

$$\rho_{\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) \triangleq \min_{k=1, \dots, |\mathcal{K}_\nu|} \rho_{k_\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}),$$

leading to the following inner convex quadratic approximation for the non-convex constraint (73b) in (73)

$$\rho_{\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) \geq 2r_{\min}, (\nu, \ell) \in \mathcal{N}_{req} \times \mathcal{L}. \tag{79}$$

3) *IGS Algorithm*: We now solve the following convex problem at the κ -th iteration for generating the next feasible point $(\mathbf{p}^{(\kappa+1)}, \mathbf{x}^{(\kappa+1)})$ for (73)

$$\begin{aligned}
\max_{\mathbf{p}, \mathbf{x}} \min_{\nu=1, \dots, N_{req}} & \left[\sum_{\ell=1}^L \rho_{\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) - 2\Theta(\mathbf{p}^{(\kappa)}, \mathbf{x}^{(\kappa)}) \tilde{\pi}_\nu(\mathbf{p}, \mathbf{x}) \right] \\
\text{s.t.} & \quad (5), (73c), (76), (79), \tag{80}
\end{aligned}$$

which is equivalent to the convex quadratic problem:

$$\begin{aligned}
& \max_{\mathbf{p}, \mathbf{x}, \gamma} \gamma \quad \text{s.t.} \quad (5), (73c), (76), (79), \\
& \sum_{\ell=1}^L \rho_{\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) - 2\Theta(\mathbf{p}^{(\kappa)}, \mathbf{x}^{(\kappa)}) \tilde{\pi}_\nu(\mathbf{p}, \mathbf{x}) \geq \gamma, \nu = 1, \dots, N_{req}.
\end{aligned}$$

The computational complexity of (80) is given by (44) with $n_\nu = 2 \sum_{i=1}^{N_R} |\mathcal{N}_i| + N_R n_t + 1$ and $n_c = 2N_R + N_{req}(L+1)$.

Similarly to (45), we can show that

$$\Theta(\mathbf{p}^{(\kappa+1)}, \mathbf{x}^{(\kappa+1)}) > \Theta(\mathbf{p}^{(\kappa)}, \mathbf{x}^{(\kappa)}),$$

provided that $(\mathbf{p}^{(\kappa+1)}, \mathbf{x}^{(\kappa+1)}) \neq (\mathbf{p}^{(\kappa)}, \mathbf{x}^{(\kappa)})$, so similarly to Algorithm 1, Algorithm 2 also generates a sequence $\{(\mathbf{p}^{(\kappa)}, \mathbf{x}^{(\kappa)})\}$ of improved feasible points for (73) that converges to a locally optimal solution of (73).

Algorithm 2 RZFB IGS Algorithm

- 1: **Set** $\kappa = 0$. Iterate (81) for a feasible point $(\mathbf{p}^{(0)}, \mathbf{x}^{(0)})$ for (73).
 - 2: **Repeat until convergence**. Solve the convex optimization problem (80) to generate the next feasible point $(\mathbf{p}^{(\kappa+1)}, \mathbf{x}^{(\kappa+1)})$ for (73). Set $\kappa := \kappa + 1$.
 - 3: **Output** $(\mathbf{p}^{(\kappa)}, \mathbf{x}^{(\kappa)})$ as the optimal solution of (73).
-

TABLE II
PARAMETER SETTINGS

| Parameter | Value |
|--|------------------------|
| Radius of cell | 300 m |
| Carrier frequency/ Bandwidth(Bw) | 2 GHz / 10 MHz |
| RRH maximum transmit power | 30 dBm |
| Shadowing standard deviation | 8 dB |
| Noise power density | -174 dBm/Hz |
| The circuit power per antenna (P_a) | 5.6×10^{-3} W |
| The number of transmit antennas at RRH (n_t) | 4 |
| The number of RRHs (N_R) | 5 |
| The number of UEs (N_u) | 10 |
| The number of subfiles per file (L) | 4 |
| The number of files in the library (F) | 100 |
| Caching capacity (μ) | 1/2 |
| Popularity exponent for Zipf's distribution (γ_z) | 1 |

To find an initial feasible point for (73), initialized by any feasible point for the convex constraints (5), (73c) we iterate the following optimization procedure

$$\begin{aligned}
\max_{\mathbf{p}, \mathbf{x}} \min & \left\{ \min_{(\nu, \ell) \in \mathcal{N}_{req} \times \mathcal{L}} \left[\frac{\rho_{\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x})}{2r_{\min}/\log_2(e)} - 1 \right], \right. \\
& \left. \min_{i \in \mathcal{N}_R} \left[1 - \frac{1}{\log_2(e)C} \mu_i^{(\kappa)}(\mathbf{p}^i, \mathbf{x}^i) \right] \right\} \\
& \text{s.t.} \quad (5), (73c), \tag{81}
\end{aligned}$$

until the value of the objective in (81) becomes more than 0 or equal to 0. The problem (81) is equivalent to

$$\begin{aligned}
& \max_{\mathbf{p}, \mathbf{x}, \gamma} \gamma \quad \text{s.t.} \quad (5), (73c), \\
& \rho_{\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) \geq 2r_{\min}(1 + \gamma)/\log_2(e), (\nu, \ell) \in \mathcal{N}_{req} \times \mathcal{L}, \\
& \mu_i^{(\kappa)}(\mathbf{p}^i, \mathbf{x}^i) \leq \log_2(e)C(1 - \gamma), i \in \mathcal{N}_R.
\end{aligned}$$

Before closing this section, let us point out that Algorithm 2 can also be readily adjusted to solve the following max-min rate optimization problem:

$$\max_{\mathbf{p}, \mathbf{x}} \min_{(\nu, \ell) \in \mathcal{N}_{req} \times \mathcal{L}} \frac{1}{2} \rho_{\nu, \ell}(\mathbf{p}, \mathbf{x}) \quad \text{s.t.} \quad (5), (55), (73c). \tag{82}$$

Then, instead of (80) the following problem is solved at the κ -th iteration to generate $(\mathbf{p}^{(\kappa+1)}, \mathbf{x}^{(\kappa+1)})$:

$$\max_{\mathbf{p}, \mathbf{x}} \min_{(\nu, \ell) \in \mathcal{N}_{req} \times \mathcal{L}} \frac{1}{2} \rho_{\nu, \ell}^{(\kappa)}(\mathbf{p}, \mathbf{x}) \quad \text{s.t.} \quad (5), (73c), (76), \tag{83}$$

while for an initial feasible, we choose a reasonable r_{\min} and then iterate (81).

V. SIMULATION RESULTS

The efficiency of the proposed algorithms along with their convergence is demonstrated through the numerical results

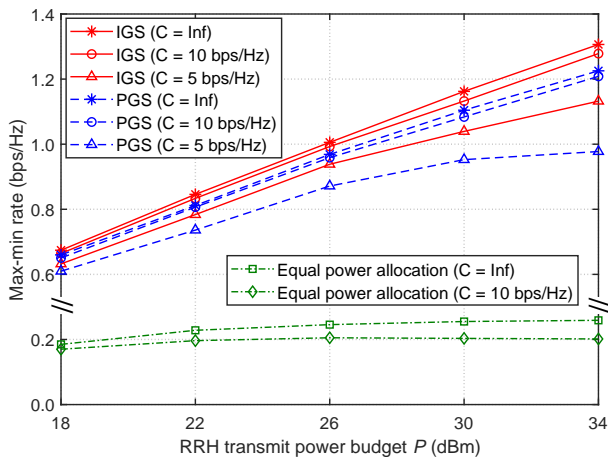


Fig. 2. The max-min rate versus RRH transmit power budget P using the parameters of Table II

of this section. The channel vector $h_{k,i}$ between the RRH i and UE k at the distance $d_{k,i}$ in km is modeled as $h_{k,i} = \sqrt{10^{-\rho_{k,i}/10}} \tilde{h}_{k,i}$, where $\rho_{k,i} = 148.1 + 37.6 \log_{10}(d_{k,i})$ (dB) is the pathloss and $\tilde{h}_{k,i}$ having independent and identically distributed complex entries is the small-scale fading [40]. The error tolerance for declaring convergence is set to $\varepsilon = 1e-3$. Unless stated otherwise, the main parameters follow Table II. We simulate three of the most popular caching strategies:

- Caching the most popular files (CMP): Each RRH stores \mathcal{F}_i number of the most popular files, so we have $|\mathcal{F}_i| = \lfloor \mu F \rfloor$ and $c_{f,\ell}^i = 1$ if and only if $f \leq |\mathcal{F}_i|$.
- Caching fractions of distinct files (CFD): Each RRH stores up to $\lfloor L/2 \rfloor$ fragments of each file that are randomly chosen, so $c_{f,\ell}^i = 1$ if and only if $\ell \in \mathcal{L}_f^i$, where \mathcal{L}_f^i is a set of $\lfloor L/2 \rfloor$ numbers picked randomly from $\mathcal{L} \triangleq \{1, \dots, L\}$, which are independent across the file f and the RRH index i . This strategy promotes collaboration between RRHs, when their cache capacity C is small.
- Random caching (RanC): Each RRH stores a set \mathcal{F}_i number of distinct files, which are arbitrarily selected from F files, so $|\mathcal{F}_i| = \lfloor \mu F \rfloor$ and $c_{f,\ell}^i = 1$ if and only if $f \in \mathcal{F}_i$.

Each UE is served by a single file at each time. The UE's requests are generated by Zipf's distribution associated with $\gamma_z = 1$.

Fig. 2 plots the max-min user rate versus the transmit power budget P at the RRH. Observe that as expected, the achievable max-min rate is improved upon increasing the transmit power budget. Furthermore, the IGS algorithm outperforms the PGS algorithm. Additionally, the max-min rate generated without any fronthaul constraint and that generated by observing the fronthaul constraint of either $C = 10$ bps/Hz or $C = 5$ bps/Hz are also illustrated in Fig. 2. To expound further, observe that increasing the fronthaul capacity C results in increased r_{min} for both the IGS and PGS algorithms. The fixed power allocation policy is also characterized in Fig. 2, in which the max-min rate optimization is carried out under the equal power allocation and uniform quantization. We can observe that the

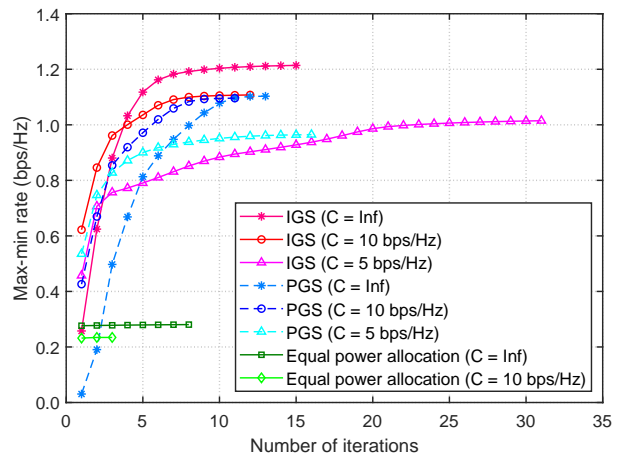


Fig. 3. The convergence behaviors of the max-min rate optimization schemes

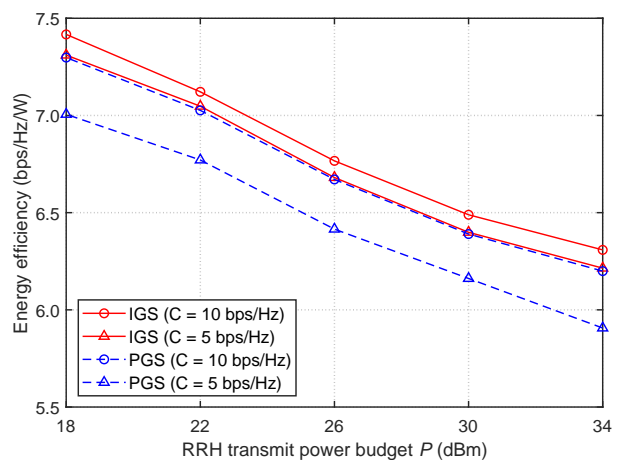


Fig. 4. The energy efficiency versus RRH transmit power budget P using the parameters of Table II

achievable max-min rate under this policy is much worse than that of the proposed RZFB. Table III provides the average number of iterations required for declaring convergence in Fig. 2.

Fig. 3 plots the convergence behaviors of the max-min rate optimization schemes in simulating Fig. 2 with $P = 30$ dBm. The max-min rate improves after each iteration by the proposed RZFB schemes, and the PGS algorithm converges faster than the IGS algorithm as the latter needs to handle more decision variables.

Fig. 4 plots the achievable energy efficiency (EE) versus the transmit power budget P at the RRH with r_{min} fixed at 0.4 bps/Hz, where the fronthaul capacity C set to either 10 bps/Hz or 5 bps/Hz. The proposed IGS algorithm outperforms the PGS algorithm in terms of its EE under different power budgets. As expected, a reduced EE is attained at an increased power budget. At the fronthaul capacity of $C = 10$ bps/Hz, which allows the BBU to transfer the uncached contents to RRHs at higher rates, the EE improves for both the IGS and PGS algorithms. As the fronthaul capacity C is reduced from 10 bps/Hz to 5 bps/Hz, the achievable EE of the IGS algorithm exhibits a modest but non-negligible performance

TABLE III
THE AVERAGE NUMBER OF ITERATIONS FOR IMPLEMENTING IGS, PGS AND EQUAL POWER ALLOCATION ALGORITHMS IN OBTAINING FIG. 2

| | $P = 18$ dBm | $P = 22$ dBm | $P = 26$ dBm | $P = 30$ dBm | $P = 34$ dBm |
|---|--------------|--------------|--------------|--------------|--------------|
| IGS ($C = \text{Inf}$) | 9.6 | 8.9 | 12.0 | 16.3 | 20.3 |
| IGS ($C = 10$ bps/Hz) | 10.3 | 9.0 | 14.1 | 13.6 | 17.4 |
| IGS ($C = 5$ bps/Hz) | 11.6 | 14.3 | 19.8 | 27.0 | 26.4 |
| PGS ($C = \text{Inf}$) | 6.9 | 11.0 | 13.6 | 13.6 | 19.3 |
| PGS ($C = 10$ bps/Hz) | 6.1 | 6.1 | 7.5 | 10.9 | 13.3 |
| PGS ($C = 5$ bps/Hz) | 10.4 | 17.4 | 14.0 | 13.6 | 18.8 |
| Equal power allocation ($C = \text{Inf}$) | 2.0 | 2.6 | 3.5 | 6.3 | 4.9 |
| Equal power allocation ($C = 10$ bps/Hz) | 4.0 | 5.6 | 5.3 | 5.3 | 5.1 |

TABLE IV
THE AVERAGE NUMBER OF ITERATIONS FOR IMPLEMENTING IGS AND PGS ALGORITHMS IN OBTAINING FIG. 4

| | $P = 18$ dBm | $P = 22$ dBm | $P = 26$ dBm | $P = 30$ dBm | $P = 34$ dBm |
|------------------------|--------------|--------------|--------------|--------------|--------------|
| IGS ($C = 10$ bps/Hz) | 5.5 | 6.4 | 7.6 | 15.9 | 16.3 |
| IGS ($C = 5$ bps/Hz) | 9.0 | 9.8 | 11.9 | 14.1 | 15.9 |
| PGS ($C = 10$ bps/Hz) | 4.6 | 6.3 | 6.8 | 8.1 | 12.0 |
| PGS ($C = 5$ bps/Hz) | 7.4 | 9.4 | 11.1 | 11.9 | 15.6 |

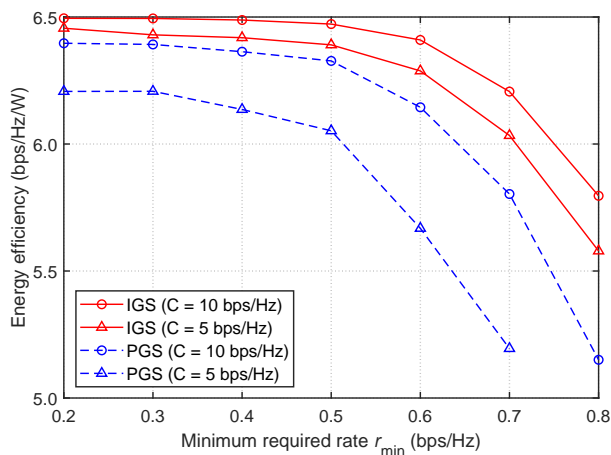


Fig. 5. The energy efficiency versus the minimum required rate r_{min} using the parameters of Table II

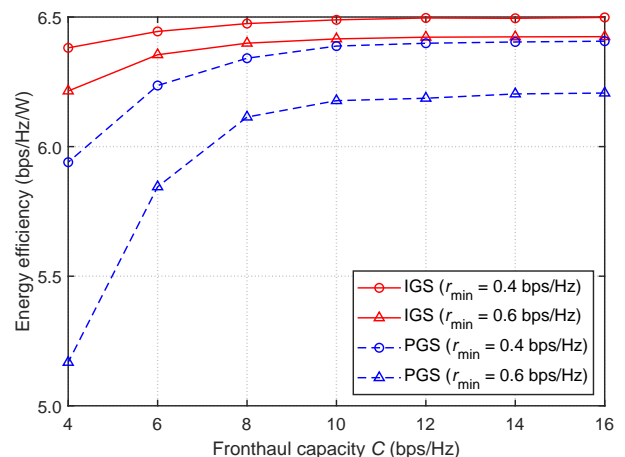


Fig. 6. The energy efficiency versus the fronthaul capacity C using the parameters of Table II

erosion compared to that of PGS. The average number of iterations required for convergence in Fig. 4 is shown in Table IV.

Fig. 5 plots the achievable EE versus the minimum required rate r_{min} with the fronthaul capacity C either set to 10 bps/Hz or to 5 bps/Hz. The transmit power budget P is fixed at 30 dBm in the following simulations. Observe that the EE of IGS algorithm is higher than that of PGS. When the fronthaul capacity C is sufficient ($C = 10$ bps/Hz), the EE decreases for both IGS and PGS upon increasing r_{min} . However, for the limited fronthaul capacity of $C = 5$ bps/Hz, the EE of PGS degrades more dramatically than that of IGS. Additionally, $r_{min} > 0.7$ cannot be attained by PGS. By contrast, IGS attains better fronthaul utilization and thus has higher r_{min} .

Fig. 6 plots the achievable EE versus the fronthaul capacity C parameterized by the minimum required rate r_{min} set to 0.4 bps/Hz and 0.6 bps/Hz, respectively. The EE exhibits an upward trend upon increasing C . In order to reach a higher r_{min} , more transmit power is needed. The EE at $r_{min} = 0.6$ bps/Hz is lower than that at $r_{min} = 0.4$ bps/Hz. Furthermore, IGS outperforms PGS, especially at low fronthaul capacity.

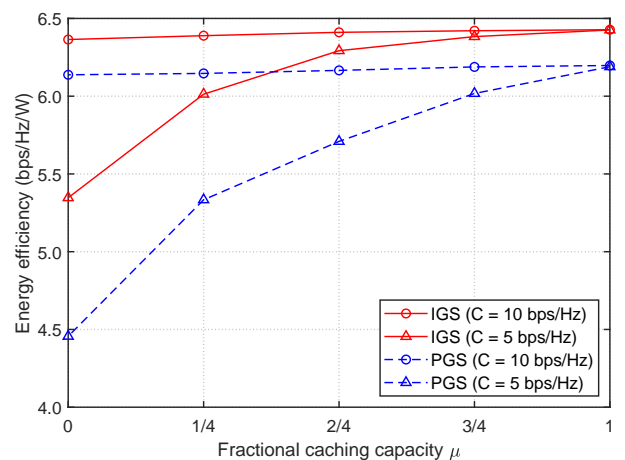


Fig. 7. The energy efficiency versus the fractional caching capacity μ using the parameters of Table II

The beneficial effect of the fractional caching capacity μ on the EE is revealed in Fig. 7. We set r_{min} to 0.6 bps/Hz, and C either to 10 bps/Hz or to 5 bps/Hz. The reduction of the

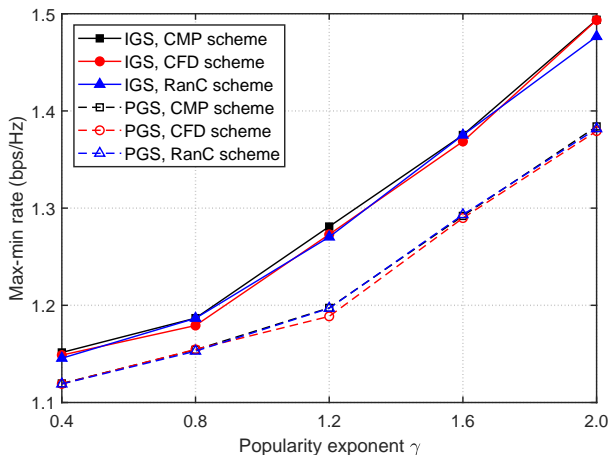


Fig. 8. The max-min rate versus the popularity exponent γ without fronthaul constraint using the parameters of Table II

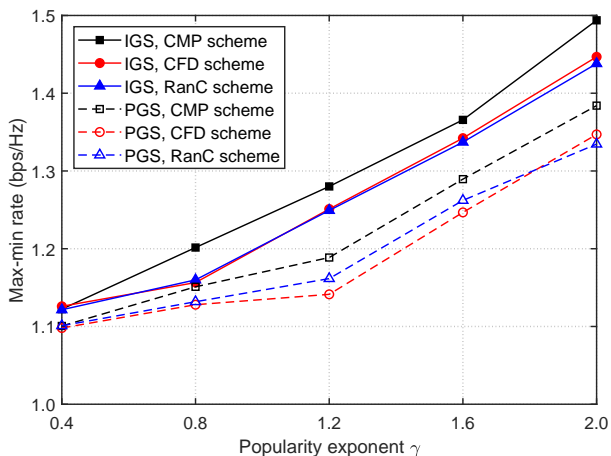


Fig. 9. The max-min rate versus the popularity exponent γ with fronthaul constraint using the parameters of Table II

fractional caching capacity μ leads to increasing the number of content downloads via the fronthaul links, inevitably degrading the EE. Observe that the EE of IGS is higher than that of PGS, especially for the lower fronthaul capacity of $C = 5$ bps/Hz. This further illustrates the benefits of IGS. The average number of iterations required for declaring convergence in Fig. 7 is shown in Table V.

In Fig. 8 we characterize the impact of the popularity exponent γ on the CMP, CFD, and RanC caching schemes. Fig. 8 plots the max-min users rate versus the popularity exponent γ without any fronthaul constraint. We fix the fractional caching capacity to $\mu = 1/4$. The three caching schemes provide similar performances under the condition that the fronthaul capacity is sufficiently large. The IGS algorithm outperforms the PGS based algorithm for all the three caching schemes. Fig. 9 plots the max-min user rate versus the popularity exponent γ at the fronthaul constraint of $C = 10$ bps/Hz, where IGS still exhibits better performance. The achievable rate tends to increase with the popularity exponent γ , because the UE requests are generated by the Zipf distribution, where typically the most popular files are requested at higher γ , and

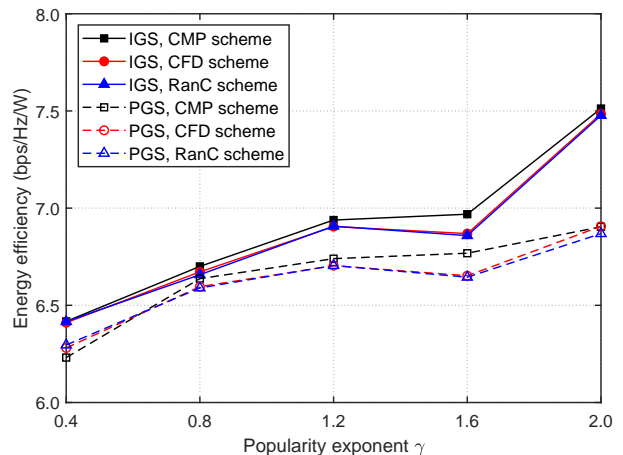


Fig. 10. The energy efficiency versus the popularity exponent γ using the parameters of Table II

TABLE V
THE AVERAGE NUMBER OF ITERATIONS FOR IMPLEMENTING IGS AND PGS ALGORITHMS IN OBTAINING FIG. 7

| | $\mu = 0$ | $\mu = 1/4$ | $\mu = 1/2$ | $\mu = 3/4$ | $\mu = 1$ |
|------------------------|-----------|-------------|-------------|-------------|-----------|
| IGS ($C = 10$ bps/Hz) | 10.3 | 14.4 | 12.5 | 13.1 | 13.6 |
| IGS ($C = 5$ bps/Hz) | 18.8 | 17.6 | 15.9 | 18.9 | 13.6 |
| PGS ($C = 10$ bps/Hz) | 8.3 | 8.4 | 8.8 | 9.0 | 8.8 |
| PGS ($C = 5$ bps/Hz) | 15.0 | 15.1 | 12.0 | 11.1 | 8.8 |

TABLE VI
THE AVERAGE NUMBER OF ITERATIONS FOR IMPLEMENTING IGS AND PGS ALGORITHMS IN OBTAINING FIG. 10

| | $\gamma = 0.4$ | $\gamma = 0.8$ | $\gamma = 1.2$ | $\gamma = 1.6$ | $\gamma = 2.0$ |
|------------------|----------------|----------------|----------------|----------------|----------------|
| IGS, CMP scheme | 10.6 | 11.1 | 12.0 | 16.1 | 18.9 |
| IGS, CFD scheme | 15.0 | 13.9 | 10.9 | 16.3 | 14.9 |
| IGS, RanC scheme | 9.6 | 10.3 | 11.1 | 16.5 | 14.8 |
| PGS, CMP scheme | 9.8 | 9.6 | 14.4 | 16.1 | 14.0 |
| PGS, CFD scheme | 8.3 | 8.3 | 10.3 | 15.6 | 13.8 |
| PGS, RanC scheme | 8.0 | 8.3 | 11.1 | 10.1 | 12.9 |

thus the burden on fronthauling is alleviated.

Fig. 10 plots the achievable EE versus the popularity exponent γ . We fix $r_{min} = 0.6$ bps/Hz, $C = 10$ bps/Hz, and $\mu = 1/4$. As the popularity exponent γ is increased, the distribution of requested files is dominated by a small number of the most popular files. Hence less files have to be pre-coded and transmitted over the fronthaul, hence increasing the EE. Since the users' requests are randomly generated, the distribution of the randomly generated requested files becomes unpredictable, especially when γ varies from 0.8 to 1.6. The EE of all three caching schemes fluctuates within this range. As illustrated in Fig. 8 and Fig. 9, IGS has the highest rate. It also provides high EE as illustrated in Fig. 10. Additionally, the CMP scheme outperforms the other two schemes for high γ . This is because when the UEs tend to request the most popular files, caching these popular files in the RRHs will put less burden on fronthauling, hence improving the EE. Table VI provides the average number of iterations required for convergence in Fig. 10.

VI. CONCLUSIONS

The performance of content-driven F-RANs is limited by both inter-content and intra-content interferences. We mitigated this problem by a new class of RZFB specifically tailored for managing the inter-content interference in the face of limited-capacity backhaul links and minimum user-rate requirements. We also derived expressions having moderate numbers of decision variables in the optimized solutions. The main computational challenges have been resolved by tractable convex quadratic function optimization at a low complexity. Explicitly, we have developed a pair of path-following algorithms for solving our convex quadratic problem by generating an improved feasible point at each iteration. The resultant procedures converged rapidly to locally optimal solutions. An extension enhancing the performance of our data driven multilayered network is under current study.

APPENDIX: FUNDAMENTAL INEQUALITIES FOR CONVEX QUADRATIC APPROXIMATIONS

The following inequality is valid for all positive definite matrices \mathbf{Y} and $\bar{\mathbf{Y}}$ of size $n \times n$ [41]:

$$\begin{aligned} \ln |I_n + [\mathbf{V}]^2 \mathbf{Y}^{-1}| \\ \geq \ln |I_n + [\bar{\mathbf{V}}]^2 \bar{\mathbf{Y}}^{-1}| - \langle [\bar{\mathbf{V}}]^2 \bar{\mathbf{Y}}^{-1} \rangle + 2\Re\{ \langle \bar{\mathbf{V}}^H \bar{\mathbf{Y}}^{-1} \mathbf{V} \rangle \} \\ - \langle \bar{\mathbf{Y}}^{-1} - ([\bar{\mathbf{V}}]^2 + \bar{\mathbf{Y}})^{-1}, [\mathbf{V}]^2 + \mathbf{Y} \rangle. \end{aligned} \quad (84)$$

The particular case of $n = 1$ in (84) is

$$\begin{aligned} \ln \left(1 + \frac{|\mathbf{v}|^2}{\mathbf{y}} \right) \geq \ln \left(1 + \frac{|\bar{v}|^2}{\bar{y}} \right) - \frac{|\bar{v}|^2}{\bar{y}} + 2 \frac{\Re\{\bar{v}^* \mathbf{v}\}}{\bar{y}} \\ - \frac{|\bar{v}|^2}{\bar{y}(|\bar{v}|^2 + \bar{y})} (|\mathbf{v}|^2 + \mathbf{y}) \\ \forall \mathbf{v} \in \mathbb{C}, \mathbf{y} > 0 \quad \& \quad \bar{v} \in \mathbb{C}, \bar{y} > 0. \end{aligned} \quad (85)$$

Let us define a function

$$f(\mathbf{X}, \mathbf{Y}) = \ln |I_m + [\mathbf{X}]^2 ([\mathbf{Y}]^2 + I_m)^{-1}|$$

where we have $\mathbf{X} \in \mathbb{C}^{m \times n}$ and $\mathbf{Y} \in \mathbb{C}^{m \times q}$. Then for any $\bar{\mathbf{X}} \in \mathbb{C}^{m \times n}$ and $\bar{\mathbf{Y}} \in \mathbb{C}^{m \times q}$, the following inequality holds

$$\begin{aligned} f(\mathbf{X}, \mathbf{Y}) \leq f(\bar{\mathbf{X}}, \bar{\mathbf{Y}}) + \langle (I_m + [\bar{\mathbf{X}}]^2 + [\bar{\mathbf{Y}}]^2)^{-1} \rangle + \langle [\bar{\mathbf{Y}}]^2 \rangle \\ - \langle ([\bar{\mathbf{Y}}]^2 + I_m)^{-1} \rangle + \langle (I_m + [\bar{\mathbf{X}}]^2 + [\bar{\mathbf{Y}}]^2)^{-1} \rangle, \\ [\mathbf{X}]^2 + [\mathbf{Y}]^2 - 2\Re\{ \langle \bar{\mathbf{Y}}^H \mathbf{Y} \rangle \} \\ + \langle I_m - ([\bar{\mathbf{Y}}]^2 + I_m)^{-1}, [\mathbf{Y}]^2 \rangle. \end{aligned} \quad (86)$$

REFERENCES

- [1] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wirel. Commun.*, vol. 22, pp. 152–160, Apr. 2015.
- [2] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," *IEEE Internet Things J.*, vol. 3, pp. 854–864, Jun. 2016.
- [3] M. Peng, S. Yan, K. Zhang, and C. Wang, "Fog-computing-based radio access networks: issues and challenges," *IEEE Network*, vol. 30, pp. 46–53, Apr. 2016.
- [4] Y.-J. Ku, D.-Y. Lin, C.-F. Lee, P.-J. Hsieh, H.-Y. Wei, C.-T. Chou, and A.-C. Pang, "5G radio access network design with the fog paradigm: Confluence of communications and computing," *IEEE Commun. Mag.*, vol. 55, pp. 46–52, Apr. 2017.
- [5] L. F. Bittencourt, J. Diaz-Montes, R. Buyya, O. F. Rana, and M. Parashar, "Mobility-aware application scheduling in fog computing," *IEEE Cloud Comput.*, vol. 4, p. 26035, Mar./Apr. 2017.
- [6] X. Zhu et al., "Improving video performance with edge servers in the fog computing architecture," *J. Intel Technol.*, vol. 19, pp. 202–224, Apr. 2015.
- [7] C. Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow, and P. A. Polakos, "A comprehensive survey on fog computing: State-of-the-art and research challenges," *IEEE Commun. Surv. Tut.*, vol. 20, no. 1, pp. 416–464, 2018.
- [8] T. Dang and M. Peng, "Joint radio communication caching and computing design for mobile virtual reality delivery in fog radio access networks," *IEEE J. Select. Areas Commun.*, vol. 37, pp. 1594–1607, Jul. 2019.
- [9] R. Karasik, O. Simeone, and S. S. Shitz, "How much can D2D communication reduce content delivery latency in fog networks with edge caching?," *IEEE Trans. Commun.*, vol. 68, pp. 2308–2323, Apr. 2020.
- [10] G. S. Paschos, G. Iosifidis, M. Tao, D. Towsley, and G. Caire, "The role of caching in future communication systems and networks," *IEEE J. Select. Areas Commun.*, vol. 36, pp. 1111–1125, Jun. 2018.
- [11] S. M. Azimi, O. Simeone, A. Sengupta, and R. Tandon, "Online edge caching and wireless delivery in fog-aided networks with dynamic content popularity," *IEEE J. Select. Areas Commun.*, vol. 36, pp. 1189–1202, Jun. 2018.
- [12] S. He, J. Ren, J. Wang, Y. Huang, Y. Zhang, W. Zhuang, and S. Shen, "Cloud-edge coordinated processing: Low-latency multicasting transmission," *IEEE J. Select. Areas Commun.*, vol. 37, pp. 1144–1158, May 2019.
- [13] B. Yin, M. Peng, S. Yan, and C. Hu, "Tradeoff between ergodic rate and delivery latency in fog radio access networks," *IEEE Trans. Wirel. Commun.*, vol. 19, pp. 2240–2251, Apr. 2020.
- [14] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, pp. 131–139, Feb. 2014.
- [15] D. Liu and C. Yang, "Energy efficiency of downlink networks with caching at base stations," *IEEE J. Sel. Areas Commun.*, vol. 34, pp. 907–922, Apr. 2016.
- [16] C. Yang, Y. Yao, Z. Chen, and B. Xia, "Analysis on cache-enabled wireless heterogeneous networks," *IEEE Trans. Wirel. Commun.*, vol. 15, pp. 131–145, Jan. 2016.
- [17] H. Freeman and T. Zhang, "The emerging era of fog computing and networking [the president's page]," *IEEE Commun. Mag.*, vol. 54, pp. 4–5, Jun. 2016.
- [18] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud RAN," *IEEE Trans. Wirel. Commun.*, vol. 15, pp. 6118–6131, Sept. 2016.
- [19] S.-H. Park, O. Simeone, and S. Shamai, "Joint optimization of cloud and edge processing for fog radio access networks," *IEEE Trans. Wirel. Commun.*, vol. 15, pp. 7621–7632, Nov. 2016.
- [20] M. Tao, D. Gunduz, F. Xu, and J. S. P. Roig, "Content caching and delivery in wireless radio access networks," *IEEE Trans. Commun.*, vol. 67, pp. 4724–4749, Jul. 2019.
- [21] H. T. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and W.-J. Hwang, "Nonsmooth optimization algorithms for multicast beamforming in content-centric fog radio access networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 1455–1469, 2020.
- [22] K. Wang, J. Li, Y. Yang, W. Chen, and L. Hanzo, "Content-centric heterogeneous fog networks relying on energy efficiency optimization," *IEEE Trans. Vehic. Techn.*, vol. 69, pp. 13579–13592, Nov. 2020.
- [23] H. H. Kha, H. D. Tuan, and H. H. Nguyen, "Fast global optimal power allocation in wireless networks by local d.c. programming," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 510–515, Feb. 2012.
- [24] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 1985.
- [25] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: Evidence and implications," in *Proc. IEEE INFOCOM*, vol. 1, pp. 126–134, Mar. 1999.
- [26] A. Sengupta, R. Tandon, and O. Simeone, "Fog-aided wireless networks for content delivery: Fundamental latency tradeoffs," *IEEE Trans. Info. Theory*, vol. 63, pp. 6650–6678, Oct. 2017.
- [27] S. H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Signal Process.*, vol. 61, pp. 5646–5658, Nov. 2013.

- [28] 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.
- [29] H. Ahleghagh and S. Dey, "Video-aware scheduling and caching in the radio access network," *IEEE/ACM Trans. Network.*, vol. 22, pp. 1444–1462, May 2014.
- [30] D. Peaucelle, D. Henrion, Y. Labit, and K. Taitz, "User's guide for SeDuMi interface 1.04," 2002.
- [31] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, pp. 681–683, Jul. 1978.
- [32] 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.
- [33] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge Univ. Press, 2010.
- [34] 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. Sel. Topics Signal Process.*, vol. 13, pp. 496–507, Jun. 2019.
- [35] H. Yu, H. D. Tuan, T. Q. Duong, Y. Fang, and L. Hanzo, "Improper Gaussian signaling for integrated data and energy networking," *IEEE Trans. Commun.*, vol. 68, pp. 3922–3934, Jun. 2020.
- [36] A. A. Nasir, H. D. Tuan, H. H. Nguyen, T. Q. Duong, and H. V. Poor, "Signal superposition in NOMA with proper and improper Gaussian signaling," *IEEE Trans. Commun.*, vol. 68, pp. 6537–6551, Oct. 2020.
- [37] H. Yu, H. D. Tuan, A. A. Nasir, T. Q. Duong, and L. Hanzo, "Improper Gaussian signaling for computationally tractable energy and information beamforming," *IEEE Trans. Veh. Techn.*, vol. 69, pp. 13990–13995, Nov. 2020.
- [38] T. M. Cover and J. A. Thomas, *Elements of Information Theory (second edition)*. John Wileys & Sons, 2006.
- [39] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov./Dec. 1999.
- [40] E. Bjornson, M. Kountouris, and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," in *Proc. IEEE ICT*, pp. 1–5, May 2013.
- [41] 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, Jun. 2016.



Wenbo Zhu received the B.S. degree in Electrical Engineering and Automation from Shanghai University, Shanghai, China, in 2014, and the M.S. degree in Control Science and Engineering from Anhui Polytechnic University, in 2018. He is currently pursuing the Ph.D. degree with the School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW, Australia, and with Shanghai University, Shanghai, China. His current research interests include optimization methods for wireless communication and signal processing.



Hoang Duong Tuan received the Diploma (Hons.) and Ph.D. degrees in applied mathematics from Odessa State University, Ukraine, in 1987 and 1991, respectively. He spent nine academic years in Japan as an Assistant Professor in the Department of Electronic-Mechanical Engineering, Nagoya University, from 1994 to 1999, and then as an Associate Professor in the Department of Electrical and Computer Engineering, Toyota Technological Institute, Nagoya, from 1999 to 2003. He was a Professor with the School of Electrical Engineering and Telecommunications, University of New South Wales, from 2003 to 2011. He is currently a Professor with the School of Electrical and Data Engineering, University of Technology Sydney. He has been involved in research with the areas of optimization, control, signal processing, wireless communication, and biomedical engineering for more than 20 years.

communications, University of New South Wales, from 2003 to 2011. He is currently a Professor with the School of Electrical and Data Engineering, University of Technology Sydney. He has been involved in research with the areas of optimization, control, signal processing, wireless communication, and biomedical engineering for more than 20 years.



at Hokkaido University, Japan. His current research interests include 5G/6G and the IoT networks.

Eryk Dutkiewicz (SM'17) received the B.E. degree in electrical and electronic engineering and the M.Sc. degree in applied mathematics from The University of Adelaide in 1988 and 1992, respectively, and the Ph.D. degree in telecommunications from the University of Wollongong in 1996. His industry experience includes management of the Wireless Research Laboratory at Motorola in early 2000's. He is currently the Head of the School of Electrical and Data Engineering, University of Technology Sydney, Australia. He also holds a professorial appointment



Yong Fang received the Ph.D. degree in electronic engineering from the City University of Hong Kong, Hong Kong, in 1999. He is currently a Professor with the School of Communication and Information Engineering, Shanghai University, Shanghai, China. His research interests include communication signal processing, blind signal processing, and adaptive information systems.



Lajos Hanzo (<http://www-mobile.ecs.soton.ac.uk>, https://en.wikipedia.org/wiki/Lajos_Hanzo) (FIEEE'04) received his Master degree and Doctorate in 1976 and 1983, respectively from the Technical University (TU) of Budapest. He was also awarded the Doctor of Sciences (DSc) degree by the University of Southampton (2004) and Honorary Doctorates by the TU of Budapest (2009) and by the University of Edinburgh (2015). He is a Foreign Member of the Hungarian Academy of Sciences and a former Editor-in-Chief of the IEEE

Press. He has served several terms as Governor of both IEEE ComSoc and of VTS. He has published 2000+ contributions at IEEE Xplore, 19 Wiley-IEEE Press books and has helped the fast-track career of 123 PhD students. Over 40 of them are Professors at various stages of their careers in academia and many of them are leading scientists in the wireless industry. He is also a Fellow of the Royal Academy of Engineering (FREng), of the IET and of EURASIP.