

# Optimized Uplink Transmission in Multi-Antenna C-RAN With Spatial Compression and Forward

Liang Liu, *Member, IEEE*, and Rui Zhang, *Senior Member, IEEE*

**Abstract**—MIMO and cloud radio access network (C-RAN) are promising techniques for implementing future wireless communication systems, where a large number of antennas are deployed either being co-located at the base station or totally distributed at separate sites called remote radio heads (RRHs), both to achieve enormous spectrum efficiency and energy efficiency gains. Here, we consider a general antenna deployment method for wireless networks, termed multi-antenna C-RAN, where a flexible number of antennas can be equipped at each RRH to more effectively balance the performance and fronthaul complexity tradeoff beyond the conventional massive MIMO and single-antenna C-RAN. To coordinate and control the fronthaul traffic over multi-antenna RRHs, under the uplink communication setup, we propose a new “spatial-compression-and-forward (SCF)” scheme, where each RRH first performs a linear spatial filtering to denoise and maximally compress its received signals from multiple users to a reduced number of dimensions, then conducts uniform scalar quantization over each of the resulting dimensions in parallel, and finally sends the total quantized bits via a finite-rate fronthaul link to the baseband unit (BBU) for joint information decoding. Under this scheme, we maximize the minimum SINR of all users at the BBU by a joint resource allocation over the wireless transmission and fronthaul links. Specifically, each RRH determines its own spatial filtering solution in a distributed manner to reduce the signaling overhead with the BBU, while the BBU jointly optimizes the users’ transmit power, the RRHs’ fronthaul bits allocation, and the BBU’s receive beamforming with fixed spatial filters at individual RRHs. Numerical results show that, given a total number of antennas to be deployed, multi-antenna C-RAN with the proposed SCF and joint optimization significantly outperforms both massive MIMO and single-antenna C-RAN under practical fronthaul capacity constraints.

**Index Terms**—Beamforming, fronthaul constraint, massive multiple-input multiple-output (MIMO), multi-antenna cloud radio access network (C-RAN), power control, signal-to-interference-plus-noise ratio (SINR) maximization, spatial-compression-and-forward (SCF).

Manuscript received January 20, 2015; revised April 28, 2015 and May 29, 2015; accepted June 15, 2015. Date of publication June 25, 2015; date of current version August 24, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tongtong Li. The work was supported in part by the National University of Singapore under Research Grant R-263-000-B46-112.

L. Liu is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 138601, Singapore (e-mail: liu\_liang@u.nus.edu).

R. Zhang is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 138601, Singapore, and also with the Institute for Infocomm Research, A\*STAR, Singapore (e-mail: elezhang@nus.edu.sg).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2015.2450199

## I. INTRODUCTION

It is anticipated that the mobile data traffic will grow 1000 times higher in the next decade with a rate of roughly a factor of two per year [1]. As a result, the fifth-generation or 5G wireless system on the roadmap has to make a major paradigm shift to accommodate this dramatic explosion in future demands for wireless data communications. Both as highly promising candidate techniques for the 5G wireless network, massive multiple-input multiple-output (MIMO) [2], [3] and cloud radio access network (C-RAN) [4] have recently drawn significant attentions. Both techniques advocate the use of very large number of antennas and centralized signal processing to achieve enormously improved efficiency in spectrum and energy usage. However, they are practically implemented with different performance and complexity trade-off considerations.

In a massive MIMO system, as depicted in Fig. 1(a), a large number of antennas are deployed at the base station (BS) to reap all the benefits of the conventional multi-user MIMO system but with a much greater scale. It is shown in [2] that in the asymptotic regime where the number of antennas at the BS is much larger than that of the users, the channels of different users are orthogonal with each other, and thus a simple matched filter, i.e., maximal ratio combining (MRC) in the uplink and maximal ratio transmission (MRT) in the downlink, is capacity-achieving. Moreover, the required transmit energy per bit vanishes as the number of antennas goes to infinity [5]. Despite the above appealing benefits, there are many issues that potentially limit the performance of massive MIMO in practice. For instance, due to the limited space for installation, the number of antennas at the BS is finite and their channels are generally correlated; as a result, the theoretical performance limit in the asymptotic regime cannot be fully achieved in practice with simple MRC/MRT. Moreover, similar to conventional MIMO system, the performance of the cell-edge users in the massive MIMO system is still a bottleneck since their channel conditions may be still weak due to more significant path attenuation.

On the other hand, in a single-antenna C-RAN as shown in Fig. 1(b), each antenna is deployed at one separate site called remote radio head (RRH), which is connected to the central unit (CU) via a high-speed fronthaul link (fiber or wireless) where joint information processing is performed. Unlike the BS in the massive MIMO system which decodes or encodes the user messages locally, in C-RAN each RRH merely forwards the signals from/to the CU via its fronthaul link, while leaving the joint encoding/decoding complexity to a baseband unit (BBU) in the CU. As a result, if the data can be perfectly transmitted between the BBU and RRHs over the fronthaul links without any error, the single-antenna C-RAN can be viewed as another form

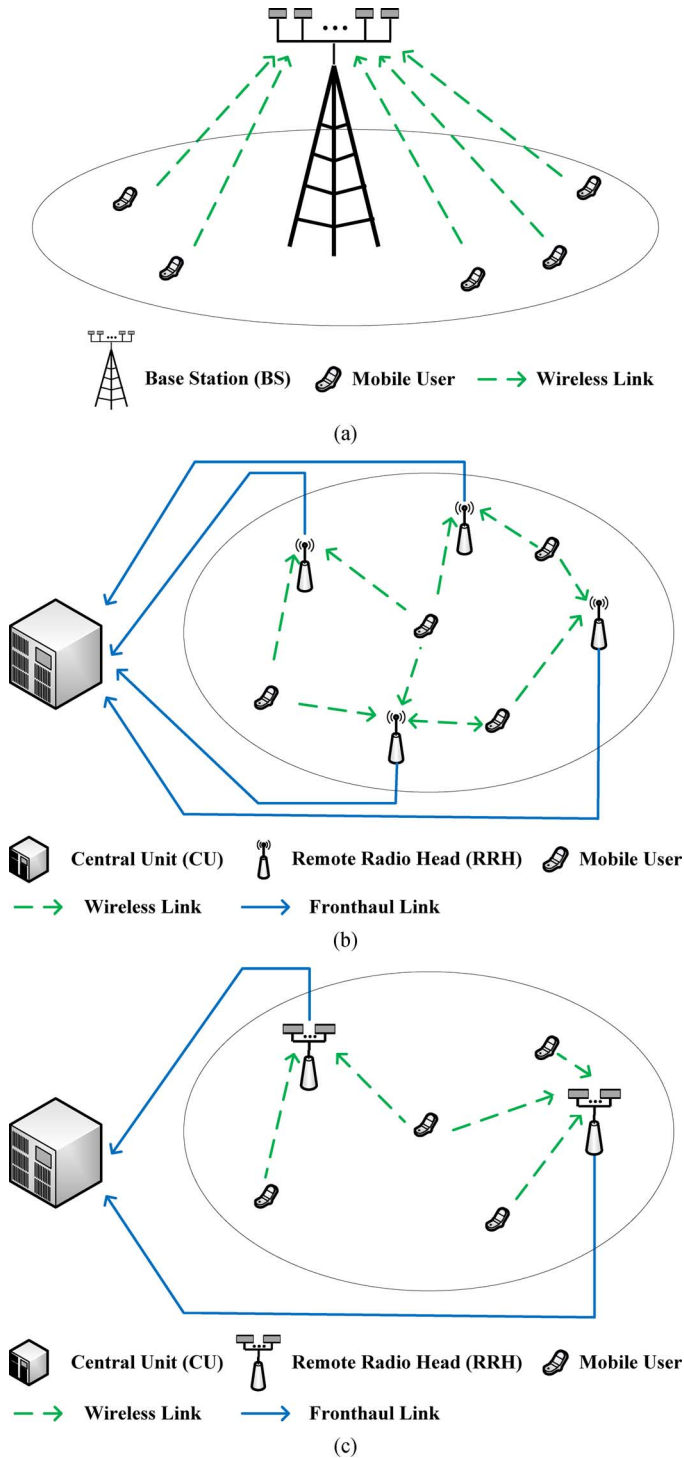


Fig. 1. Comparison of massive MIMO, single-antenna C-RAN, and multi-antenna C-RAN in the uplink transmission. (a) Massive MIMO. (b) Single-antenna C-RAN. (c) Multi-antenna C-RAN.

of massive MIMO but with the antennas densely distributed over the whole served region such that all the mobile users can be served by some adjacent antennas with strong channel conditions. However, one obstacle constraining the practically achievable throughput of C-RAN is the limited-capacity fronthaul link between each RRH and BBU. According to [4], the capacity of the current commercial fibers, which is about several Gbps (gigabits per second), can be easily overwhelmed in

C-RAN even under moderate data traffic; whereas wireless fronthaul techniques such as those exploiting the millimeter-wave bands in general operate with even lower data rate per link and shorter distance as compared to the fiber based solution. In the literature, a considerable amount of effort has been dedicated to study fronthaul quantization/compression techniques in the uplink communication [6]–[12]. Specifically, the “quantize-and-forward (QF)” scheme [13]–[15] is adopted to reduce the communication rates between the BBU and RRHs, where each single-antenna RRH samples, quantizes and forwards its received signals to the BBU over its fronthaul link. Moreover, the QF scheme is usually studied under an information-theoretical Gaussian test channel model, and the quantization noise levels of all the RRHs are jointly optimized to maximize the end-to-end throughput subject to the capacity constraints of individual fronthaul links. Different from the complex quantization schemes considered in [9]–[12], the simplest and probably most widely used quantization technique in practice is the so-called uniform scalar quantization, where the step size between adjacent quantization levels is constant. Recently, [16] reveals that subject to the capacity of the current commercial fibers, the above theoretical performance upper bound can be approached by applying a simple uniform scalar quantization at each RRH for orthogonal frequency division multiplexing (OFDM) based single-antenna C-RAN. Moreover, another line of research in C-RAN studies the control of the overall fronthaul traffic by selectively activating only a small subset of all the RRHs to meet the user demands [17].

In the aforementioned massive MIMO system and single-antenna C-RAN, the antennas are deployed in a totally co-located and separated manner, respectively. In addition to these two extreme cases, a more general solution is to distribute the antennas at a reduced number of sites as compared to single-antenna C-RAN, while the RRH at each site is equipped with multiple antennas. This so-called multi-antenna C-RAN design can provide both improved channel conditions to the users as in single-antenna C-RAN as well as local spatial multiplexing gain as in massive MIMO; furthermore, by adjusting the number of antennas at each RRH and the number of RRHs, these benefits can be flexibly traded-off with a given total number of antennas deployed. However, the fronthaul issue becomes more severe in the multi-antenna C-RAN as compared to conventional single-antenna counterpart, since the limited fronthaul capacity at each RRH needs to be allocated over the outputs of more antennas, thus leading to more quantization errors in general. To tackle this challenge and address the above design trade-off, in this paper we focus our study on the uplink communication in C-RAN with multi-antenna RRHs each subject to a finite fronthaul capacity constraint. The main contributions of this paper are summarized as follows.

- We propose a novel “spatial-compression-and-forward (SCF)” scheme for multi-antenna RRHs to balance between the information conveyed to the BBU and the data traffic over the fronthaul links. Specifically, each RRH first performs spatial filtering to its received signals at different antennas from the users such that the signals are denoised and maximally compressed into a reduced number of dimensions. Then, each RRH applies the simple uniform scalar quantization [16] over each of these dimensions

in parallel by appropriately allocating the number of quantization bits over them under a total rate constraint. Last, the quantized bits by all RRHs are sent to the BBU for joint information decoding via their fronthaul links. It is worth noting that spatial filtering by means of e.g., Karhunen-Loeve transform, has been widely adopted in the literature for compressing and denoising spatially correlated signals with additive noise/interference [18], [19]. It is also worth noting that at each RRH, the compression of the correlated signals received by all its antennas is fulfilled by a simple linear filter, rather than the complicated distributed source coding, e.g., Wyner-Ziv coding, as in single-antenna C-RAN where all antennas are physically separated [11], [12].

- With the proposed SCF scheme, we formulate the optimization problem of joint users' power allocation, RRHs' spatial filter design and quantization bits allocation, as well as BBU's receive beamforming to maximize the minimum signal-to-interference-plus-noise ratio (SINR) of all the users in the uplink. We propose an efficient solution to this complicated design problem for ease of practical implementation. First, each RRH computes its own spatial filter in a distributed manner based on its locally received signal covariance matrix. This distributed solution helps save significant fronthaul resources for exchanging control signals with the BBU. Then, the BBU jointly optimizes the other parameters to maximize the minimum SINR of all users based on the alternating optimization technique. Specifically, with given quantization bits allocation at each RRH, we extend the well-known fixed-point method [20] to obtain the optimal transmit power levels for the users and receive beamforming vectors at the BBU. On the other hand, given the above optimized parameters, the corresponding quantization bits allocation at each RRH is efficiently solved.
- Last, we investigate the following interesting question: Given a total number of antennas for a target area, should all of them be deployed at one BS, i.e., massive MIMO, or one at each RRH, i.e., single-antenna C-RAN, or optimally divided over a certain number of RRHs, i.e., multi-antenna C-RAN, given a practical total fronthaul capacity constraint for C-RAN? Through numerical examples, it is shown that with the proposed SCF scheme and joint wireless-fronthaul-BBU optimization, multi-antenna C-RAN generally performs noticeably better than both massive MIMO and single-antenna C-RAN under practical fronthaul rate constraints.

The rest of this paper is organized as follows. Section II presents the system model for C-RAN with the SCF scheme. Section III formulates the minimum-SINR maximization problem. Section IV presents the proposed solution for this problem. Section V provides numerical results to verify the effectiveness of the proposed multi-antenna C-RAN with SCF. Finally, Section VI concludes the paper.

*Notation:* Scalars are denoted by lower-case letters, vectors by bold-face lower-case letters, and matrices by bold-face upper-case letters.  $\mathbf{I}$  and  $\mathbf{0}$  denote an identity matrix and an all-zero matrix, respectively, with appropriate dimensions. For a square matrix  $\mathbf{S}$ ,  $\mathbf{S} \succeq \mathbf{0}$  ( $\mathbf{S} \preceq \mathbf{0}$ ) means that  $\mathbf{S}$  is positive

(negative) semi-definite. For a matrix  $\mathbf{M}$  of arbitrary size,  $\mathbf{M}^H$  and  $\text{rank}(\mathbf{M})$  denote the conjugate transpose and rank of  $\mathbf{M}$ , respectively.  $\text{diag}(\mathbf{a})$  denotes a diagonal matrix with the main diagonal given by the vector  $\mathbf{a}$ , while  $\text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_K)$  denotes a block diagonal matrix with the diagonal given by square matrices  $\mathbf{A}_1, \dots, \mathbf{A}_K$ .  $E[\cdot]$  denotes the statistical expectation. The distribution of a circularly symmetric complex Gaussian (CSCG) random vector with mean  $\mathbf{x}$  and covariance matrix  $\mathbf{\Sigma}$  is denoted by  $\mathcal{CN}(\mathbf{x}, \mathbf{\Sigma})$ ; and  $\sim$  stands for "distributed as".  $\mathbb{C}^{x \times y}$  denotes the space of  $x \times y$  complex matrices.  $\|\mathbf{x}\|$  denotes the Euclidean norm of a complex vector  $\mathbf{x}$ . For two real vectors  $\mathbf{x}$  and  $\mathbf{y}$ ,  $\mathbf{x} \geq \mathbf{y}$  means that  $\mathbf{x}$  is greater than or equal to  $\mathbf{y}$  in a component-wise manner.

## II. SYSTEM MODEL

This paper considers the uplink communication in a multi-antenna C-RAN. As shown in Fig. 1(c), the system consists of one BBU,  $N$  RRHs, denoted by the set  $\mathcal{N} = \{1, \dots, N\}$ , and  $K$  users, denoted by the set  $\mathcal{K} = \{1, \dots, K\}$ . It is assumed that each RRH is equipped with  $M \geq 1$  antennas, while each user is equipped with one single antenna. It is further assumed that each RRH  $n$  is connected to the BBU via a digital error-free fronthaul link with a capacity of  $\bar{T}_n$  bits per second (bps). In the uplink, each RRH processes the signals received from all the users into digital bits and forwards them to the BBU via its fronthaul link. Then, the BBU jointly decodes the users' messages based on the signals from all the RRHs. The details of our studied uplink multi-antenna C-RAN are given as follows.

It is assumed that all the  $K$  users transmit over quasi-static flat-fading channels over a given bandwidth of  $B$  Hz.<sup>1</sup> The equivalent baseband complex symbol received at RRH  $n$  is then expressed as

$$\mathbf{y}_n = \sum_{k=1}^K \mathbf{h}_{n,k} \sqrt{p_k} s_k + \mathbf{z}_n = \mathbf{H}_n \mathbf{P}^{\frac{1}{2}} \mathbf{s} + \mathbf{z}_n, \quad \forall n, \quad (1)$$

where  $\mathbf{s} = [s_1, \dots, s_K]^T$  with  $s_k \sim \mathcal{CN}(0, 1)$  denoting the transmit symbol of user  $k$  which is modelled as a CSCG random variable with zero-mean and unit-variance,  $\mathbf{P} = \text{diag}(p_1, \dots, p_K)$  with  $p_k$  denoting the transmit power of user  $k$ ,  $\mathbf{H}_n = [\mathbf{h}_{n,1}, \dots, \mathbf{h}_{n,K}] \in \mathbb{C}^{M \times K}$  with  $\mathbf{h}_{n,k} \in \mathbb{C}^{M \times 1}$  denoting the channel vector from user  $k$  to RRH  $n$ , and  $\mathbf{z}_n \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$  denotes the additive white Gaussian noise (AWGN) at RRH  $n$ . It is assumed that  $\mathbf{z}_n$ 's are independent over  $n$ .

At each multi-antenna RRH, we consider the use of the SCF scheme to forward the received baseband symbol  $\mathbf{y}_n$  to the BBU via the finite-capacity digital fronthaul link. An illustration of the SCF scheme is shown in Fig. 2, where each RRH first converts the signal received at each antenna to the baseband and then into digital bits by analog-to-digital converter (ADC), which is assumed to be perfect with no conversion error. After that, each RRH processes the digital signals from all of its antennas via spatial filtering, then conducts a scalar

<sup>1</sup>It is worth noting that this paper studies the narrowband wireless channel, while the results obtained can be readily extended to broadband based C-RAN by viewing each frequency subchannel as one narrowband channel considered here and accordingly scaling down the fronthaul capacity  $\bar{T}_n$  by the number of frequency subchannels in use.

quantization over each of the output dimensions in parallel, and finally forwards the total quantized bits from all dimensions to the BBU via the fronthaul link. Specifically, RRH  $n$  first implements spatial filtering to its received signal  $\mathbf{y}_n$  to obtain

$$\begin{aligned}\bar{\mathbf{y}}_n &= \mathbf{V}_n \mathbf{y}_n = \mathbf{V}_n \sum_{k=1}^K \mathbf{h}_{n,k} \sqrt{p_k} s_k + \mathbf{V}_n \mathbf{z}_n \\ &= \mathbf{V}_n \mathbf{H}_n \mathbf{P}^{\frac{1}{2}} \mathbf{s} + \mathbf{V}_n \mathbf{z}_n, \quad n = 1, \dots, N,\end{aligned}\quad (2)$$

where  $\mathbf{V}_n = [\mathbf{v}_{n,1}, \dots, \mathbf{v}_{n,L}]^T \in \mathbb{C}^{L \times M}$  denotes the filtering matrix at RRH  $n$  with  $L \leq M$  denoting the (reduced) dimension of the output signal  $\bar{\mathbf{y}}_n$  after filtering. Specifically, in this paper we set  $L = \min(M, K)$  such that there is no information loss due to spatial filtering at each RRH [18], [19]. Then, a simple uniform scalar quantization is applied to each element of  $\bar{\mathbf{y}}_n = [\bar{y}_{n,1}, \dots, \bar{y}_{n,L}]^T$  at RRH  $n$ . Note that each complex symbol  $\bar{y}_{n,l}$  can be represented by its in-phase (I) and quadrature (Q) parts as

$$\bar{y}_{n,l} = \bar{y}_{n,l}^I + j\bar{y}_{n,l}^Q, \quad \forall n, l, \quad (3)$$

where  $j^2 = -1$ , and the I-branch symbol  $\bar{y}_{n,l}^I$  and Q-branch symbol  $\bar{y}_{n,l}^Q$  are both real Gaussian random variables with zero mean and variance  $(\sum_{k=1}^K p_k |\mathbf{v}_{n,l}^T \mathbf{h}_{n,k}|^2 + \sigma^2 \|\mathbf{v}_{n,l}\|^2)/2$ . A typical method to implement the uniform quantization is via separate I/Q scalar quantization, the details of which can be found in [16]. After uniform scalar quantization, the baseband quantized symbol of  $\bar{\mathbf{y}}_n$  is then given by

$$\tilde{\mathbf{y}}_n = \bar{\mathbf{y}}_n + \mathbf{e}_n = \mathbf{V}_n \sum_{k=1}^K \mathbf{h}_{n,k} \sqrt{p_k} s_k + \mathbf{V}_n \mathbf{z}_n + \mathbf{e}_n, \quad \forall n, \quad (4)$$

where  $\mathbf{e}_n = [e_{n,1}, \dots, e_{n,L}]^T$  with  $e_{n,l}$ ,  $1 \leq l \leq L$ , denoting the quantization error for  $\bar{y}_{n,l}$  with zero mean and variance  $q_{n,l}$ . Let  $D_{n,l}$  denote the number of bits that RRH  $n$  uses to quantize the I-branch or Q-branch of  $\bar{y}_{n,l}$ , i.e.,  $\bar{y}_{n,l}^I$  or  $\bar{y}_{n,l}^Q$ . The quantization noise level  $q_{n,l}$  due to  $e_{n,l}$  for uniform quantization is then given by [16]

$$q_{n,l} = \begin{cases} 3 \left( \sum_{k=1}^K p_k |\mathbf{v}_{n,l}^T \mathbf{h}_{n,k}|^2 + \sigma^2 \|\mathbf{v}_{n,l}\|^2 \right) 2^{-2D_{n,l}}, & \text{if } D_{n,l} > 0, \\ \infty, & \text{if } D_{n,l} = 0. \end{cases} \quad (5)$$

Note that  $e_{n,l}$ 's are independent over  $l$  due to independent scalar quantization for each element of  $\bar{\mathbf{y}}_n$ , and also over  $n$  due to independent processing at different RRHs. As a result, the covariance matrix of  $\mathbf{e}_n$  is a function of  $\mathbf{p} = [p_1, \dots, p_K]^T$ ,  $\mathbf{V}_n$  as well as  $\mathbf{D}_n = [D_{n,1}, \dots, D_{n,L}]^T$ , which is given by

$$\mathbf{Q}_n(\mathbf{p}, \mathbf{V}_n, \mathbf{D}_n) = E[\mathbf{e}_n \mathbf{e}_n^H] = \text{diag}(q_{n,1}, \dots, q_{n,L}). \quad (6)$$

Then, each RRH forwards the quantized bits to the BBU via the fronthaul link. The transmission rate in RRH  $n$ 's fronthaul link is expressed as [16]

$$T_n = 2B \sum_{l=1}^L D_{n,l}, \quad n = 1, \dots, N. \quad (7)$$

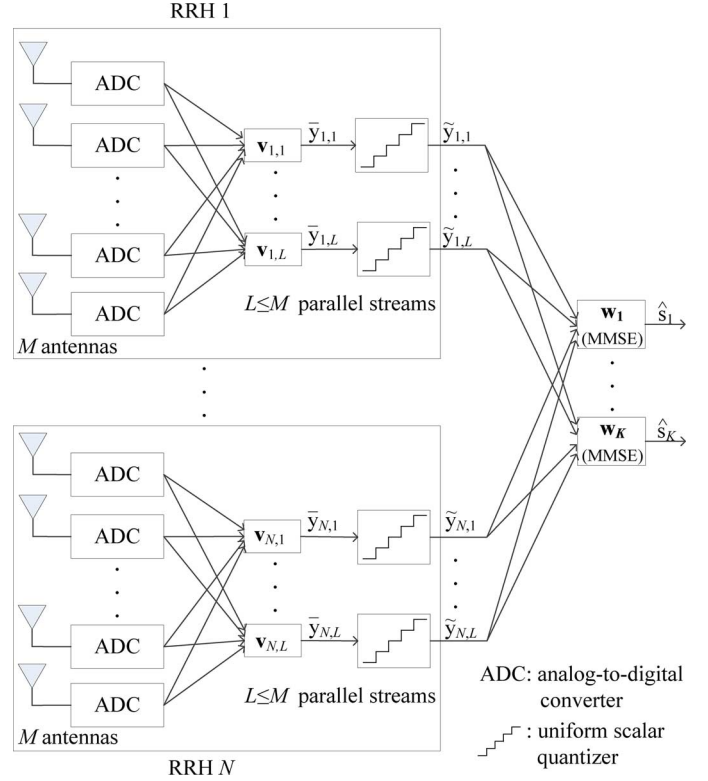


Fig. 2. Illustration of C-RAN with SCF applied at each multi-antenna RRH.

It is worth noting that if  $D_{n,l} = 0$ , then the symbol  $\bar{y}_{n,l}$  at the spatial filter output dimension  $l$  is not quantized at RRH  $n$  and thus is not forwarded to the BBU; as a result,  $q_{n,l} = \infty$  according to (5). To summarize, our proposed SCF scheme is a two-stage compression method, where the output signal dimension at each RRH is first reduced from  $M$  to  $L$  (if  $M > L$ ) by spatial filtering, and then further reduced by allocating the limited fronthaul quantization bits to only  $L^* \leq L$  out of the  $L$  output dimensions with  $D_{n,l} > 0$ .

The received signal at the BBU from all RRHs is expressed as

$$\tilde{\mathbf{y}} = [\tilde{\mathbf{y}}_1^T, \dots, \tilde{\mathbf{y}}_N^T]^T = \mathbf{V} \sum_{k=1}^K \mathbf{h}_{(k)} \sqrt{p_k} s_k + \mathbf{V} \mathbf{z} + \mathbf{e}, \quad (8)$$

where  $\mathbf{V} = \text{diag}(\mathbf{V}_1, \dots, \mathbf{V}_N)$ ,  $\mathbf{h}_{(k)} = [\mathbf{h}_{1,k}^T, \dots, \mathbf{h}_{N,k}^T]^T$ ,  $\mathbf{z} = [\mathbf{z}_1^T, \dots, \mathbf{z}_N^T]^T$ , and  $\mathbf{e} = [\mathbf{e}_1^T, \dots, \mathbf{e}_N^T]^T$ . To decode  $s_k$ , we consider that a linear beamforming<sup>2</sup> is applied to  $\tilde{\mathbf{y}}$ , i.e.,

$$\begin{aligned}\hat{s}_k &= \mathbf{w}_k^H \tilde{\mathbf{y}} \\ &= \mathbf{w}_k^H \mathbf{V} \mathbf{h}_{(k)} \sqrt{p_k} s_k + \sum_{j \neq k} \mathbf{w}_k^H \mathbf{V} \mathbf{h}_{(j)} \sqrt{p_j} s_j \\ &\quad + \mathbf{w}_k^H \mathbf{V} \mathbf{z} + \mathbf{w}_k^H \mathbf{e}, \quad \forall k,\end{aligned}\quad (9)$$

where  $\mathbf{w}_k = [w_{k,1}, \dots, w_{k,NL}]^T \in \mathbb{C}^{NL \times 1}$  is the receive beamformer for  $s_k$ . Note that  $w_{k,(n-1)L+l} = 0$  if the  $l$ th signal

<sup>2</sup>It is worth noting that the subsequent results also hold similarly for the case with the non-linear successive interference cancellation scheme with any fixed decoding order, which is not considered in this paper due to its high complexity for implementation.

dimension at RRH  $n$  is not quantized, i.e.,  $D_{n,l} = 0$ . According to (9), the SINR for decoding  $s_k$  is expressed as

$$\gamma_k = \frac{p_k |\mathbf{w}_k^H \mathbf{V} \mathbf{h}_{(k)}|^2}{\sum_{j \neq k} p_j |\mathbf{w}_k^H \mathbf{V} \mathbf{h}_{(j)}|^2 + \sigma^2 \|\mathbf{w}_k^H \mathbf{V}\|^2 + \mathbf{w}_k^H \mathbf{Q}(\mathbf{p}, \{\mathbf{V}_n, \mathbf{D}_n\}) \mathbf{w}_k}, \quad (10)$$

where

$$\begin{aligned} \mathbf{Q}(\mathbf{p}, \{\mathbf{V}_n, \mathbf{D}_n\}) &= E[\mathbf{e}\mathbf{e}^H] \\ &= \text{diag}(\mathbf{Q}_1(\mathbf{p}, \mathbf{V}_1, \mathbf{D}_1), \dots, \mathbf{Q}_N(\mathbf{p}, \mathbf{V}_N, \mathbf{D}_N)). \end{aligned}$$

### III. PROBLEM FORMULATION

In this paper, we aim to maximize the minimum SINR of all the users by optimizing the users' power allocation, i.e.,  $\mathbf{p}$ , the spatial filter and quantization bits allocation at each RRH, i.e.,  $\mathbf{V}_n$  and  $\mathbf{D}_n$ ,  $\forall n$ , and the receive beamforming vectors at the BBU, i.e.,  $\mathbf{w}_k$ ,  $\forall k$ . Specifically, we aim to solve the following problem:

$$\begin{aligned} &\text{Maximize} && \gamma \\ &\gamma, \mathbf{p}, \{\mathbf{V}_n, \mathbf{D}_n\}, \{\mathbf{w}_k\} \\ &\text{Subject to} \end{aligned} \quad (11)$$

$$\frac{p_k |\mathbf{w}_k^H \mathbf{V} \mathbf{h}_{(k)}|^2}{\sum_{j \neq k} p_j |\mathbf{w}_k^H \mathbf{V} \mathbf{h}_{(j)}|^2 + \sigma^2 \|\mathbf{w}_k^H \mathbf{V}\|^2 + \mathbf{w}_k^H \mathbf{Q}(\mathbf{p}, \{\mathbf{V}_n, \mathbf{D}_n\}) \mathbf{w}_k} \geq \gamma, \quad \forall k, \quad (12)$$

$$p_k \leq \bar{P}_k, \quad \forall k, \quad (13)$$

$$2B \sum_{l=1}^L D_{n,l} \leq \bar{T}_n, \quad \forall n, \quad (14)$$

$$D_{n,l} \text{ is an integer}, \quad \forall n, l, \quad (15)$$

where  $\gamma$  denotes the common SINR target for all the  $K$  users and  $\bar{P}_k$  denotes the transmit power constraint of user  $k$ .

It can be observed that problem (11) is a non-convex optimization problem since the design variables are complicatedly coupled in its constraint (12). Note that with given  $\mathbf{V}_n$ 's and by setting  $D_{n,l} = \infty$ ,  $\forall n, l$ , problem (11) reduces to the well-known SINR balancing problem via transmit power control and receive beamforming only. This problem has been efficiently solved by [21], [22] based on the non-negative matrix theory [23]. However, due to the new quantization noise term in the SINR expression, i.e.,  $\mathbf{w}_k^H \mathbf{Q}(\mathbf{p}, \{\mathbf{V}_n, \mathbf{D}_n\}) \mathbf{w}_k$ , it can be shown that the algorithm proposed in [21], [22] cannot be directly applied to obtain the optimal power control and beamforming solution to problem (11) even with given  $\mathbf{V}_n$ 's and finite values of  $D_{n,l}$ 's. As a result, with the additional variables  $\mathbf{V}_n$ 's and integer variables  $\mathbf{D}_n$ 's, it is generally difficult to globally solve problem (11).

In this paper, we propose an efficient algorithm to solve problem (11) suboptimally, for which the main procedures are illustrated in Fig. 3. Specifically, in the channel training phase, all the users send orthogonal pilot signals to the RRHs with equal power  $\bar{P} \leq \min_{1 \leq k \leq K} \bar{P}_k$ . Let  $\mathbf{y}_{\text{pilot},n} \in \mathbb{C}^{M \times 1}$  denote the received signal at RRH  $n$  during the channel training

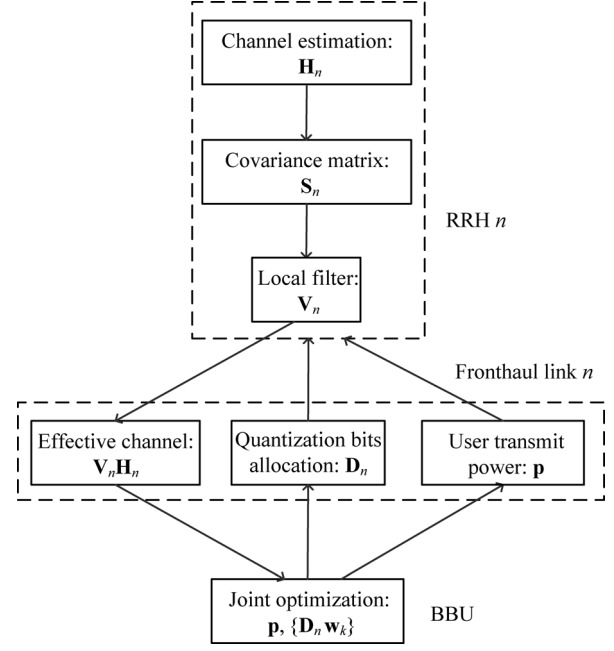


Fig. 3. A diagram for illustrating the proposed algorithm for Problem (11).

phase. Then, each RRH  $n$  estimates the MIMO channel  $\mathbf{H}_n$  (assumed to be perfect) and computes the covariance matrix of its received signal (with the presumed knowledge of  $\bar{P}$  and  $\sigma^2$ ), i.e.,

$$\begin{aligned} \mathbf{S}_n &= E[\mathbf{y}_{\text{pilot},n} \mathbf{y}_{\text{pilot},n}^H] = \mathbf{H}_n \mathbf{P} \mathbf{H}_n^H + \sigma^2 \mathbf{I} \\ &= \bar{P} \mathbf{H}_n \mathbf{H}_n^H + \sigma^2 \mathbf{I}, \quad n = 1, \dots, N. \end{aligned} \quad (16)$$

Then each RRH independently determines its spatial filter  $\mathbf{V}_n$  based on  $\mathbf{S}_n$ . Next, each RRH  $n$  sends the effective channel  $\mathbf{V}_n \mathbf{H}_n \triangleq \tilde{\mathbf{H}}_n = [\tilde{\mathbf{h}}_{n,1}, \dots, \tilde{\mathbf{h}}_{n,K}]$  to the BBU. Then, based on its global channel state information (CSI), i.e.,  $\tilde{\mathbf{H}}_n$ 's,  $n = 1, \dots, N$ , the BBU computes  $\mathbf{p}$ ,  $\mathbf{D}_n$ 's and  $\mathbf{w}_k$ 's by solving the following reduced problem of problem (11):

$$\begin{aligned} &\text{Maximize} && \gamma \\ &\gamma, \mathbf{p}, \{\mathbf{D}_n\}, \{\mathbf{w}_k\} \\ &\text{Subject to} \end{aligned} \quad (17)$$

$$\frac{p_k |\mathbf{w}_k^H \tilde{\mathbf{h}}_{(k)}|^2}{\sum_{j \neq k} p_j |\mathbf{w}_k^H \tilde{\mathbf{h}}_{(j)}|^2 + \sigma^2 \|\mathbf{w}_k^H \mathbf{V}\|^2 + \mathbf{w}_k^H \mathbf{Q}(\mathbf{p}, \{\mathbf{V}_n, \mathbf{D}_n\}) \mathbf{w}_k} \geq \gamma, \quad \forall k, \quad (18)$$

$$p_k \leq \bar{P}_k, \quad \forall k, \quad (19)$$

$$2B \sum_{l=1}^L D_{n,l} \leq \bar{T}_n, \quad \forall n, \quad (20)$$

$$D_{n,l} \text{ is an integer}, \quad \forall n, l, \quad (21)$$

where  $\tilde{\mathbf{h}}_{(k)} = [\tilde{\mathbf{h}}_{1,k}^T, \dots, \tilde{\mathbf{h}}_{N,k}^T]^T$ . Finally, the BBU sends  $\mathbf{D}_n$  to RRH  $n$ ,  $n = 1, \dots, N$ , via its fronthaul link for the implementation of scalar quantization, and  $\mathbf{p}$  to one RRH (say RRH  $n$ ), which then broadcasts  $\mathbf{p}$  to all the users in the downlink. In the following section, we first show how each RRH computes its spatial filter in a distributed manner, and then discuss how the BBU solves problem (17).

#### IV. PROPOSED ALGORITHM FOR PROBLEM (11)

In this section, we present details of the two-stage algorithm for solving problem (11) based on the diagram shown in Fig. 3.

##### A. Distributed Spatial Filter Design at RRHs

First, we show how to obtain the spatial filter at each RRH. It is observed from problem (11) that each user's SINR at the BBU depends on  $\mathbf{V}_n$ 's,  $n = 1, \dots, N$ , at all RRHs, which are also coupled with other design variables. As a result, it is difficult to obtain the jointly optimal spatial filters at all RRHs. Therefore, in the following we propose to design  $\mathbf{V}_n$ 's in a distributed manner, where each RRH  $n$  computes its own spatial filter only based on its local information, i.e.,  $\mathbf{S}_n$  in (16). First, each RRH  $n$  obtains the eigenvalue decomposition (EVD) of  $\mathbf{S}_n$  as

$$\mathbf{S}_n = \mathbf{U}_n^H \mathbf{\Lambda}_n \mathbf{U}_n, \quad n = 1, \dots, N, \quad (22)$$

where  $\mathbf{\Lambda}_n = \text{diag}(\lambda_{n,1}, \dots, \lambda_{n,M}) \in \mathbb{C}^{M \times M}$  with  $\lambda_{n,1} \geq \dots \geq \lambda_{n,M} \geq 0$  denoting the eigenvalues of  $\mathbf{S}_n$ , and  $\mathbf{U}_n = [\mathbf{u}_{n,1}, \dots, \mathbf{u}_{n,M}] \in \mathbb{C}^{M \times M}$  is a unitary matrix with  $\mathbf{u}_{n,m}$  denoting the eigenvector of  $\mathbf{S}_n$  associated with the eigenvalue  $\lambda_{n,m}$ ,  $m = 1, \dots, M$ . Notice that  $\mathbf{U}_n$  can also be obtained as the singular vectors of  $\mathbf{H}_n$  from its singular value decomposition (SVD). Note that since the rank of  $\mathbf{S}_n$  is  $\text{rank}(\mathbf{S}_n) = \min(M, K)$ ,  $\lambda_{n,m} = 0$  if  $\min(M, K) < m \leq M$ , and thus each RRH only needs to extract the first  $L = \min(M, K)$  user signal subspace by setting its spatial filter as

$$\mathbf{V}_n = \bar{\mathbf{U}}_n \triangleq [\mathbf{u}_{n,1}, \dots, \mathbf{u}_{n,L}], \quad n = 1, \dots, N. \quad (23)$$

As a result, the equivalent channel from all users to RRH  $n$  after filtering is given by

$$\tilde{\mathbf{H}}_n = \bar{\mathbf{U}}_n^H \mathbf{H}_n, \quad n = 1, \dots, N, \quad (24)$$

which is sent to the BBU in channel training as discussed in Section III.

The reasons to set the spatial filter at each RRH as in (23) are as follows. First, if  $M > K$ ,  $\bar{\mathbf{U}}_n$  denoises the received signal before quantization by extracting the user signal subspace from the received signal plus noise without loss of information [18], [19]. Second, if  $M > K$ , the signal dimension is reduced from  $M$  to  $L = K$  at each RRH  $n$ , which simplifies the subsequent uniform scalar quantization at each dimension. Last but not least, with this spatial filter applied at RRH  $n$  in data transmission, the filter output  $\tilde{\mathbf{y}}_n$  given in (2) has the following covariance matrix,

$$E[\tilde{\mathbf{y}}_n \tilde{\mathbf{y}}_n^H] = \bar{\mathbf{U}}_n^H \mathbf{H}_n \mathbf{P} \mathbf{H}_n^H \bar{\mathbf{U}}_n + \sigma^2 \mathbf{I}, \quad (25)$$

which is (approximately) diagonal if the transmit power of users, i.e.,  $p_k$ 's, are (closely) equal, i.e., the elements in  $\tilde{\mathbf{y}}_n$  are approximately uncorrelated. As a result, independent scalar quantization at each output dimension of  $\tilde{\mathbf{y}}_n$  as shown in (4) becomes more effective.

##### B. Joint Optimization at BBU

Next, given the spatial filter at each RRH  $\mathbf{V}_n = \bar{\mathbf{U}}_n$ ,  $\forall n$ , as in (23), we jointly optimize users' power allocation, RRHs' quantization bits allocation, and BBU's receive beamforming by solving problem (17).

We propose to apply the alternating optimization technique to solve problem (17). Specifically, first we fix the quantization bits allocation  $D_{n,l} = \bar{D}_{n,l}$ ,  $\forall n, l$ , in problem (17) and optimize the users' power allocation and BBU's receive beamforming solution by solving the following problem.

$$\text{Maximize}_{\gamma, \mathbf{p}, \{\mathbf{w}_k\}} \gamma \quad (26)$$

Subject to

$$\frac{p_k |\mathbf{w}_k^H \tilde{\mathbf{h}}_{(k)}|^2}{\sum_{j \neq k} p_j |\mathbf{w}_k^H \tilde{\mathbf{h}}_{(j)}|^2 + \sigma^2 \|\mathbf{w}_k\|^2 + \mathbf{w}_k^H \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \mathbf{w}_k} \geq \gamma, \quad \forall k, \quad (27)$$

$$p_k \leq \bar{P}_k, \quad \forall k. \quad (28)$$

Let  $\bar{\mathbf{p}} = [\bar{p}_1, \dots, \bar{p}_K]^T$  and  $\bar{\mathbf{w}}_k$ 's denote the solution to problem (26). Then, we fix the users' power allocation and BBU's receive beamforming solution as  $\mathbf{p} = \bar{\mathbf{p}}$ ,  $\mathbf{w}_k = \bar{\mathbf{w}}_k$ 's, respectively, and optimize the RRHs' quantization bits allocation by solving the following problem.

$$\text{Maximize}_{\gamma, \{\bar{\mathbf{D}}_n\}} \gamma \quad (29)$$

Subject to

$$\frac{\bar{p}_k |\bar{\mathbf{w}}_k^H \tilde{\mathbf{h}}_{(k)}|^2}{\sum_{j \neq k} \bar{p}_j |\bar{\mathbf{w}}_k^H \tilde{\mathbf{h}}_{(j)}|^2 + \sigma^2 \|\bar{\mathbf{w}}_k\|^2 + \bar{\mathbf{w}}_k^H \mathbf{Q}(\bar{\mathbf{p}}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \bar{\mathbf{w}}_k} \geq \gamma, \quad \forall k, \quad (30)$$

$$2B \sum_{l=1}^L D_{n,l} \leq \bar{T}_n, \quad \forall n, \quad (31)$$

$$D_{n,l} \text{ is an integer}, \quad \forall n, l. \quad (32)$$

Let  $\bar{\mathbf{D}}_n$ 's denote the solution to problem (29). The above procedure is iterated until convergence. In the following, we show how to solve problems (26) and (29), respectively.

1) *Solution to Problem (26)*: First, we solve problem (26). As mentioned in Section III, due to the additional term associated with the quantization noise, i.e.,  $\mathbf{w}_k^H \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \mathbf{w}_k$ 's, in the SINR expression, which is coupled with  $\mathbf{p}$  and  $\mathbf{w}_k$ 's, the algorithm proposed in [21] and [22] cannot be applied to obtain the optimal solution to problem (26). In the following, we propose to apply the bisection method [24] together with the celebrated fixed-point method [20] to globally solve problem (26). Specifically, given any common SINR target  $\bar{\gamma}$  for all the users, we solve the following feasibility problem.

$$\text{Find } \mathbf{p}, \{\mathbf{w}_k\} \quad (33)$$

Subject to

$$\frac{p_k |\mathbf{w}_k^H \tilde{\mathbf{h}}_{(k)}|^2}{\sum_{j \neq k} p_j |\mathbf{w}_k^H \tilde{\mathbf{h}}_{(j)}|^2 + \sigma^2 \|\mathbf{w}_k\|^2 + \mathbf{w}_k^H \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \mathbf{w}_k} \geq \bar{\gamma}, \quad \forall k, \quad (34)$$

$$p_k \leq \bar{P}_k, \quad \forall k. \quad (35)$$

The feasibility problem (33) is still a non-convex problem. In the following, we show that the feasibility of problem (33) can be efficiently checked by solving a sum-power minimization

problem without the individual power constraints in (35). The sum-power minimization problem is formulated as

$$\begin{aligned} & \underset{\mathbf{p}, \{\mathbf{w}_k\}}{\text{Minimize}} \quad \sum_{k=1}^K p_k \\ & \text{Subject to} \end{aligned} \quad (36)$$

$$\frac{p_k |\mathbf{w}_k^H \tilde{\mathbf{h}}_{(k)}|^2}{\sum_{j \neq k} p_j |\mathbf{w}_k^H \tilde{\mathbf{h}}_{(j)}|^2 + \sigma^2 \|\mathbf{w}_k\|^2 + \mathbf{w}_k^H \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \mathbf{w}_k} \geq \bar{\gamma}, \quad \forall k. \quad (37)$$

Note that given any power allocation  $\mathbf{p}$ , the optimal linear receive beamforming solution to problem (36) is the minimum-mean-square-error (MMSE) based receiver given by

$$\begin{aligned} & \mathbf{w}_k^{\text{MMSE}} \\ &= \left( \sum_{j \neq k} p_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \right)^{-1} \tilde{\mathbf{h}}_{(k)}, \quad \forall k. \end{aligned} \quad (38)$$

With the above MMSE receivers, the SINR of user  $k$  given in (10) reduces to

$$\begin{aligned} \gamma_k &= p_k \tilde{\mathbf{h}}_{(k)}^H \\ & \cdot \left( \sum_{j \neq k} p_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \right)^{-1} \tilde{\mathbf{h}}_{(k)}, \quad \forall k. \end{aligned} \quad (39)$$

As a result, problem (36) reduces to the following power control problem.

$$\underset{\mathbf{p}}{\text{Minimize}} \quad \sum_{k=1}^K p_k \quad (40)$$

Subject to

$$\begin{aligned} & p_k \tilde{\mathbf{h}}_{(k)}^H \left( \sum_{j \neq k} p_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \right)^{-1} \tilde{\mathbf{h}}_{(k)} \\ & \geq \bar{\gamma}, \quad \forall k. \end{aligned} \quad (41)$$

Define  $\mathbf{I}(\mathbf{p}) = [\mathbf{I}_1(\mathbf{p}), \dots, \mathbf{I}_K(\mathbf{p})]^T \in \mathbb{R}^{K \times 1}$  with the  $k$ th element denoted by

$$\mathbf{I}_k(\mathbf{p}) = \frac{\bar{\gamma}}{\tilde{\mathbf{h}}_{(k)}^H \left( \sum_{j \neq k} p_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \right)^{-1} \tilde{\mathbf{h}}_{(k)}}, \quad \forall k. \quad (42)$$

Then, problem (40) reduces to the following problem.

$$\underset{\mathbf{p}}{\text{Minimize}} \quad \sum_{k=1}^K p_k \quad (43)$$

$$\text{Subject to} \quad \mathbf{p} \geq \mathbf{I}(\mathbf{p}). \quad (44)$$

*Lemma 4.1:*  $\mathbf{I}(\mathbf{p})$  given in (42) is a standard interference function [20]. In other words, it satisfies 1.  $\mathbf{I}(\mathbf{p}) \geq \mathbf{0}$  if  $\mathbf{p} \geq \mathbf{0}$ ;

TABLE I  
ALGORITHM I: ALGORITHM FOR PROBLEM (26)

1. **Initialize**  $\gamma_{\min} = 0, \gamma_{\max} \geq \gamma^*$ ;
2. **Repeat**
  - a) Set  $\bar{\gamma} = \frac{\gamma_{\max} + \gamma_{\min}}{2}$ ;
  - b) Solve problem (43) by the fixed-point method as in Corollary 4.1 with initial point  $\mathbf{p}^{(0)} = \mathbf{0}$ ;
  - c) If the obtained solution  $\mathbf{p}^*$  is infinity or does not satisfy all the given individual power constraints in (28), set  $\gamma_{\max} = \bar{\gamma}$ ; otherwise, set  $\gamma_{\min} = \bar{\gamma}$ ;
3. **Until**  $\gamma_{\max} - \gamma_{\min} \leq \epsilon$ , where  $\epsilon > 0$  is a small constraint to control the algorithm accuracy;
4. Calculate  $\mathbf{w}_k$ 's by substituting  $\mathbf{p}^*$  into (38).

2. if  $\mathbf{p} \geq \mathbf{p}'$ , then  $\mathbf{I}(\mathbf{p}) \geq \mathbf{I}(\mathbf{p}')$ ; and 3.  $\forall a > 1$ , it follows that  $a\mathbf{I}(\mathbf{p}) > \mathbf{I}(a\mathbf{p})$ .

*Proof:* Please refer to Appendix A. ■

Based on Lemma 4.1, we have the following corollaries.

*Corollary 4.1:* If  $\bar{\gamma}$  is not achievable for all the  $K$  users in problem (40), the iterative fixed-point method  $\mathbf{p}^{(i+1)} = \mathbf{I}(\mathbf{p}^{(i)})$  with the initial point  $\mathbf{p}^{(0)} = \mathbf{0}$ , where  $\mathbf{p}^{(i)}$  denotes the power allocation solution obtained in the  $i$ th iteration of the above fixed-point method, will converge to  $\mathbf{p} \rightarrow \infty$ .

*Proof:* Please refer to Appendix B. ■

*Corollary 4.2:* If  $\bar{\gamma}$  is achievable for all the  $K$  users in problem (40), the iterative fixed-point method  $\mathbf{p}^{(i+1)} = \mathbf{I}(\mathbf{p}^{(i)})$  will converge to the optimal solution to problem (43) with any initial point  $\mathbf{p}^{(0)} \geq \mathbf{0}$ .

*Proof:* Please refer to [20, Theorem 2]. ■

*Corollary 4.3:* If  $\bar{\gamma}$  is achievable for all the  $K$  users in problem (40), the optimal power solution  $\mathbf{p}^*$  to problem (43) is component-wise minimum in the sense that any other feasible power solution  $\mathbf{p}'$  that satisfies (44) must satisfy  $\mathbf{p}' \geq \mathbf{p}^*$ .

*Proof:* Please refer to Appendix C. ■

According to Corollaries 4.1–4.3, the feasibility of problem (33) can be efficiently checked as follows. First, we apply the fixed-point method in Corollary 4.1 to solve the sum-power minimization problem (43) with the initial point  $\mathbf{p}^{(0)} = \mathbf{0}$ . If the obtained solution, denoted by  $\mathbf{p}^*$ , is infinity (unbounded), then  $\bar{\gamma}$  cannot be achieved by all the  $K$  users simultaneously even without the individual power constraints. Therefore, problem (33) is not feasible. Otherwise, if  $\mathbf{p}^*$  is of finite value,  $\bar{\gamma}$  can be achieved by all the users without the individual power constraints. In this case, we check whether  $\mathbf{p}^*$  satisfies all the given individual power constraints. If  $\mathbf{p}^*$  does not satisfy all the individual power constraints, according to Corollary 4.3, all the power solutions that satisfy the SINR constraints cannot satisfy the given individual power constraints. As a result, problem (33) is not feasible. Otherwise, if  $\mathbf{p}^*$  satisfies all the individual power constraints, it is a feasible solution to problem (33), i.e., problem (33) is feasible.

Let  $\gamma^*$  denote the optimal value of problem (26). The algorithm to solve problem (26) based on the bisection method is summarized in Table I. Note that in Step 1,  $\gamma_{\max}$  can be set as  $\gamma_{\max} = \max_{1 \leq k \leq K} \bar{\gamma}_k$ , where  $\bar{\gamma}_k$  is obtained by setting  $p_k = \bar{P}_k$  and  $p_j = 0, \forall j \neq k$ , in (39).

2) *Solution to Problem (29):* Next, we consider problem (29). Note that the quantization noise power given in (5), i.e.,  $q_{n,l}$ , is a continuous function over quantization bits allocation



$D_{n,l}$  (assuming it is continuous), except when  $D_{n,l} = 0$ . To deal with this issue, we approximate the quantization noise power with the following continuous function of  $D_{n,l}$ , i.e.,

$$q_{n,l} = 3 \left( \sum_{k=1}^K p_k |\mathbf{v}_{n,l}^T \mathbf{h}_{n,k}|^2 + \sigma^2 \|\mathbf{v}_{n,l}\|^2 \right) 2^{-2D_{n,l}}, \quad D_{n,l} \geq 0, \quad \forall n, l. \quad (45)$$

The difference of (45) from (5) is that when  $D_{n,l} = 0$ ,  $q_{n,l}$  equals  $3(\sum_{k=1}^K p_k |\mathbf{v}_{n,l}^T \mathbf{h}_{n,k}|^2 + \sigma^2 \|\mathbf{v}_{n,l}\|^2)$  instead of infinity. This indicates that even with  $D_{n,l} = 0$ , some quantized information at the  $l$ th signal dimension of RRH  $n$  is forwarded to the BBU. However, since  $3(\sum_{k=1}^K p_k |\mathbf{v}_{n,l}^T \mathbf{h}_{n,k}|^2 + \sigma^2 \|\mathbf{v}_{n,l}\|^2)$  is generally much larger than the signal power of any user  $k$ , i.e.,  $p_k |\mathbf{v}_{n,l}^T \mathbf{h}_{n,k}|^2$ 's, its effect on the user SINR given in (10) is negligible. Therefore, we use (45) to approximate the quantization noise power for solving problem (29). Note that when the actual SINR is computed using (10), the quantization noise power is set to be infinity if the obtained  $D_{n,l}$  is zero.

With the approximation (45), problem (29) is still challenging to solve due to the integer constraints for  $D_{n,l}$ 's. Let problem (29)-NoInt denote the relaxation of problem (29) without the integer constraints in (32). In the following, we first solve problem (29)-NoInt, the solution to which may not satisfy all the integer constraints. Then, we propose an efficient algorithm to obtain a set of integer solutions for all  $D_{n,l}$ 's based on the solution of relaxed problem (29)-NoInt.

Similar to problem (26), problem (29)-NoInt can be globally solved by the bisection method. Given any common SINR target  $\bar{\gamma}$  for all the users, we need to solve the following feasibility problem.

$$\text{Find } \{\mathbf{D}_n\} \quad (46)$$

Subject to

$$\frac{\bar{p}_k |\bar{\mathbf{w}}_k^H \tilde{\mathbf{h}}_{(k)}|^2}{\sum_{j \neq k} \bar{p}_j |\bar{\mathbf{w}}_k^H \tilde{\mathbf{h}}_{(j)}|^2 + \sigma^2 \|\bar{\mathbf{w}}_k\|^2 + \bar{\mathbf{w}}_k^H \mathbf{Q}(\bar{\mathbf{p}}, \{\bar{\mathbf{U}}_n, \mathbf{D}_n\}) \bar{\mathbf{w}}_k} \geq \bar{\gamma}, \quad \forall k, \quad (47)$$

$$2B \sum_{l=1}^L D_{n,l} \leq \bar{T}_n, \quad \forall n. \quad (48)$$

Note that we have

$$\begin{aligned} \bar{\mathbf{w}}_k^H \mathbf{Q}(\bar{\mathbf{p}}, \{\bar{\mathbf{U}}_n, \mathbf{D}_n\}) \bar{\mathbf{w}}_k &= \sum_{n=1}^N \sum_{l=1}^L q_{n,l} |\bar{w}_{k,(n-1)L+m}|^2 \\ &= \sum_{n=1}^N \sum_{l=1}^L \theta_{n,l,k} 2^{-2D_{n,l}}, \quad \forall k, \end{aligned} \quad (49)$$

where  $\bar{w}_{i,j}$  denotes the  $j$ th element of  $\bar{\mathbf{w}}_i$ ,  $1 \leq i \leq K$ ,  $1 \leq j \leq NL$ , and

$$\theta_{n,l,k} = 3 |\bar{w}_{k,(n-1)L+l}|^2 \left( \sum_{i=1}^K p_i |\mathbf{h}_{n,i}|^2 + \sigma^2 \right), \quad \forall n, l. \quad (50)$$

Note that  $\theta_{n,l,k}$  can be interpreted as the effective quantization noise power due to the  $l$ th quantized dimension at RRH  $n$  in

TABLE II  
ALGORITHM II: ALGORITHM FOR PROBLEM (29)-NOINT.

- 
1. **Initialize**  $\gamma_{\min} = 0, \gamma_{\max} \geq \gamma^*$ ;
  2. **Repeat**
    - a) Set  $\bar{\gamma} = \frac{\gamma_{\max} + \gamma_{\min}}{2}$ ;
    - b) Solve problem (51) by the interior-point method. If problem (51) is feasible, set  $\gamma_{\min} = \bar{\gamma}$ ; otherwise, set  $\gamma_{\max} = \bar{\gamma}$ ;
  3. **Until**  $\gamma_{\max} - \gamma_{\min} \leq \epsilon$ , where  $\epsilon > 0$  is a small constraint to control the algorithm accuracy.
- 

decoding  $s_k$  at the BBU. Then, problem (46) is equivalent to the following problem.

$$\text{Find } \{\mathbf{D}_n\} \quad (51)$$

$$\text{Subject to } \sum_{n=1}^N \sum_{l=1}^L \theta_{n,l,k} 2^{-2D_{n,l}} \leq \frac{\bar{p}_k |\bar{\mathbf{w}}_k^H \tilde{\mathbf{h}}_{(k)}|^2}{\bar{\gamma}} - \sum_{j \neq k} \bar{p}_j |\bar{\mathbf{w}}_k^H \tilde{\mathbf{h}}_{(j)}|^2 - \sigma^2 \|\bar{\mathbf{w}}_k\|^2, \quad \forall k, \quad (52)$$

$$2B \sum_{l=1}^L D_{n,l} \leq \bar{T}_n, \quad \forall n. \quad (53)$$

Problem (51) can be shown to be a convex feasibility problem, and thus can be efficiently solved via the interior-point method [24]. Let  $\gamma^*$  denote the optimal value of problem (29)-NoInt. The algorithm for problem (29)-NoInt is then summarized in Table II. Note that in Step 1,  $\gamma_{\max}$  can be similarly set as in Algorithm I.

It is worth noting that the solution obtained by Algorithm II may not satisfy all the integer constraints given by (32) in problem (29). In the following, we show how to obtain a set of integer solutions  $\{\mathbf{D}_n\}$  for problem (29) based on the solution of relaxed problem (29)-NoInt. Similar to [16], we propose to round each  $D_{n,l}$  to its nearby integer as follows.

$$\hat{D}_{n,l} = \begin{cases} \lfloor D_{n,l} \rfloor, & \text{if } D_{n,l} - \lfloor D_{n,l} \rfloor \leq \alpha_n, \\ \lceil D_{n,l} \rceil, & \text{otherwise,} \end{cases} \quad \forall n, l, \quad (54)$$

where  $0 \leq \alpha_n \leq 1, \forall n$ , and  $\lfloor x \rfloor, \lceil x \rceil$  denote the maximum integer that is no larger than  $x$  and the minimum integer that is no smaller than  $x$ , respectively. Note that we can always find a feasible solution of  $\hat{D}_{n,l}$ 's by simply setting  $\alpha_n = 1, \forall n$ , in (54) since in this case we have  $\sum_{l=1}^L \hat{D}_{n,l} \leq \sum_{l=1}^L D_{n,l} \leq \bar{T}_n/2B$ . In the following, we show how to find a generally better feasible solution by optimizing  $\alpha_n$ 's. Since  $\hat{D}_{n,l}$ 's decrease with  $\alpha_n$ 's, as  $\alpha_n$ 's become smaller, the resulting  $\hat{D}_{n,l}$ 's from (54) achieve higher SINRs for all the users, but the fronthaul constraints for the RRHs become more difficult to be satisfied. Hence, we propose to apply a simple bisection method to find the optimal values of  $\alpha_n$ 's, and then substitute  $\alpha_n$ 's into (54) to obtain  $\hat{D}_{n,l}$ 's. The algorithm is summarized in Table III.

3) *Overall Algorithm for Problem (17)*: After problems (26) and (29) are solved by Algorithm I and Algorithms II and III, respectively, we are ready to propose the overall algorithm for problem (17) with given  $\mathbf{V}_n = \bar{\mathbf{U}}_n, \forall n$ , based on the alternating optimization technique, which is summarized in Table IV. The algorithm starts with equal quantization bits allocation for all the spatial filter output dimensions at each RRH, i.e.,  $D_{n,l} =$



TABLE III  
ALGORITHM III: ALGORITHM TO FIND FEASIBLE INTEGER  
SOLUTION  $\hat{D}_{n,l}$ 'S FOR PROBLEM (29)

1. Initialize  $\alpha_n^{\min} = 0$ ,  $\alpha_n^{\max} = 1$ ,  $\forall n$ ;
2. For  $n = 1 : N$ , repeat
  - a. Set  $\alpha_n = \frac{\alpha_n^{\min} + \alpha_n^{\max}}{2}$ ;
  - b. Substitute  $\alpha_n$  into (54). If  $\hat{D}_{n,l}$ 's,  $\forall l$ , satisfy  $2B \sum_{l=1}^L \hat{D}_{n,l} \leq \bar{T}_n$ , set  $\alpha_n^{\max} = \alpha_n$ ; otherwise, set  $\alpha_n^{\min} = \alpha_n$ ;
3. Until  $\alpha_n^{\max} - \alpha_n^{\min} < \varepsilon$ ,  $\forall n$ , where  $\varepsilon > 0$  is a small constraint to control the accuracy of the algorithm.

TABLE IV  
ALGORITHM IV: ALGORITHM FOR PROBLEM (17) WITH GIVEN  $\mathbf{V}_n = \bar{\mathbf{U}}_n$

1. Initialize: Set  $D_{n,l}^{(0)} = \lfloor \bar{T}_n / (2BL) \rfloor$ ,  $\forall n, l$ ,  $\gamma^{(0)} = 0$ , and  $i = 0$ ;
2. Repeat
  - a.  $i = i + 1$ ;
  - b. Obtain  $\mathbf{p}^{(i)}$  and  $\{\mathbf{w}_k^{(i)}\}$  by solving problem (26) with  $\bar{D}_{n,l} = D_{n,l}^{(i-1)}$ ,  $\forall n, l$ , using Algorithm I;
  - c. Obtain  $\{D_n^{(i)}\}$  by solving problem (29) with  $\bar{\mathbf{p}} = \mathbf{p}^{(i)}$  and  $\bar{\mathbf{w}}_k = \mathbf{w}_k^{(i)}$ ,  $\forall k$ , using Algorithms II and III;
3. Until  $\gamma^{(i)} - \gamma^{(i-1)} \leq \varepsilon$ , where  $\gamma^{(i)}$  denotes the objective value of problem (17) achieved by  $\mathbf{p}^{(i)}$ ,  $\{\mathbf{w}_k^{(i)}\}$  and  $\{D_n^{(i)}\}$ , and  $\varepsilon > 0$  is a small constraint to control the accuracy of the algorithm.

$\lfloor \bar{T}_n / (2BL) \rfloor$ ,  $\forall n, l$ . Note that the proposed algorithm terminates in either of the following two cases: 1.  $0 \leq \gamma^{(i)} - \gamma^{(i-1)} \leq \varepsilon$ , i.e., the minimum SINR of users cannot be improved above the positive threshold; and 2.  $\gamma^{(i)} - \gamma^{(i-1)} < 0$ , i.e., the minimum SINR obtained at the current iteration is even reduced compared with that at the previous iteration. The second case may occur since problem (29) is generally not globally solved by Algorithms II and III due to the integer constraints in (32).

## V. NUMERICAL RESULTS

In this section, we provide numerical results to verify our results. In the following numerical examples, the bandwidth of the wireless link is  $B = 10$  MHz, while the path loss model of the wireless channel is given as  $30.6 + 36.7 \log_{10}(d)$  dB [25], where  $d$  (in meter) denotes the distance between the user and the RRH. The transmit power constraint for each user is  $\bar{P}_k = 23$  dBm,  $\forall k$ . The power spectral density of the background noise at each RRH is assumed to be  $-169$  dBm/Hz, and the noise figure due to the receiver processing is 7 dB. Furthermore, it is worth noting that current wireless systems are mostly broadband to support high-rate multimedia traffic. As mentioned in Section II, our studied narrowband system can be viewed as one frequency subchannel in an OFDM-based broadband C-RAN. To make our simulation results more practically relevant to broadband communications, in this section we consider  $\bar{T}_n$  as the normalized fronthaul capacity of RRH  $n$  for each frequency subchannel, which is obtained by dividing the practical fronthaul capacity (typically tens of Gbps) over the number of subchannels used (32, 64, etc.). Last, it is assumed that all the RRHs possess the identical normalized fronthaul capacity, i.e.,  $\bar{T}_n = T$ ,  $\forall n$ .

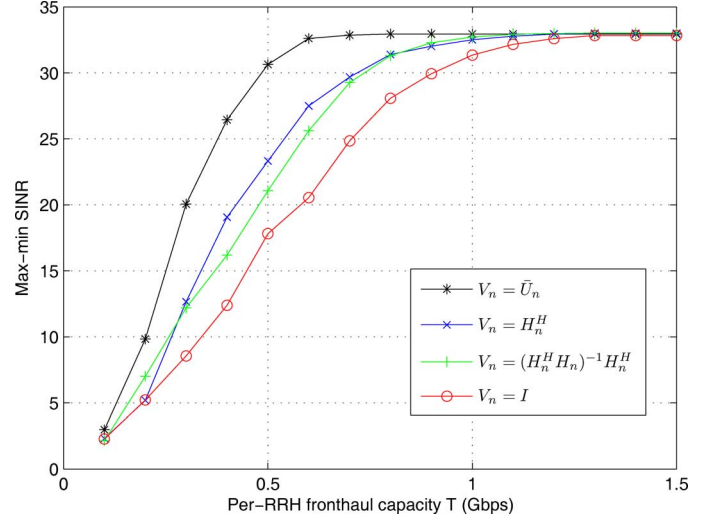


Fig. 4. Performance comparison of different distributed spatial filtering designs at RRH versus per-RRH fronthaul capacity.

### A. Effectiveness of Distributed Spatial Filtering at RRHs

First, we illustrate the effectiveness of the proposed RRHs' distributed spatial filtering via SCF. In this example, there are  $N = 4$  RRHs and  $K = 8$  users randomly distributed in a circle area of radius 500 m. Besides the proposed spatial filter design given in (23) at each RRH, we also consider the following schemes for performance benchmark: 1. Matched filtering, i.e.,  $\mathbf{V}_n = \mathbf{H}_n^H$ ,  $\forall n$ ; 2. Zero-forcing (ZF) filtering, i.e.,  $\mathbf{V}_n = (\mathbf{H}_n^H \mathbf{H}_n)^{-1} \mathbf{H}_n^H$ ,  $\forall n$ ; and 3. Without spatial filtering, i.e.,  $\mathbf{V}_n = \mathbf{I}$ ,  $\forall n$ . Note that for the matched and ZF filtering, like the proposed design, the signal dimension is reduced from  $M$  to  $K$  at each RRH, since in this example we set  $K < M$ .

Fig. 4 shows the maximized minimum (max-min) SINR of all the users achieved by Algorithm IV with different spatial filtering designs at the RRHs, where each RRH is equipped with  $M = 10$  antennas and the per-RRH fronthaul capacity  $T$  varies from  $T = 0.1$  Gbps to  $T = 1.5$  Gbps. It is observed that the max-min SINR achieved with the proposed spatial filtering in (23) based on the EVD of the covariance matrix  $\mathbf{S}_n$  (or SVD of  $\mathbf{H}_n$ ) is higher than that achieved by the other benchmark schemes for all values of fronthaul capacities. This observation also holds in other simulation setups, which are omitted due to the space limitation. The performance gain is explained as follows. For matched filtering, the dimension of the spatial filter output  $\bar{\mathbf{y}}_n$  given in (2) is reduced from  $M$  to  $L = K$ , but its elements are still correlated at each RRH  $n$ , which makes the subsequent scalar quantization ineffective. On the other hand, with ZF filtering, each output dimension of RRH  $n$  only contains one user's message, while some users' signals with low signal-to-noise ratio (SNR) at RRH  $n$  are not quantized and forwarded to the BBU if  $D_{n,l} = 0$  for some  $l$  after performing Algorithm IV. With this user selection based scheme at each RRH, although the signal dimension is reduced and the elements in  $\bar{\mathbf{y}}_n$  are decorrelated in general, the noise effect becomes more severe due to ZF filtering (and is further amplified by subsequent quantization), which degrades the performance of joint decoding at the BBU. Furthermore, with  $\mathbf{V}_n = \mathbf{I}$ ,  $\forall n$ , our proposed SCF scheme reduces to the antenna selection based quan-

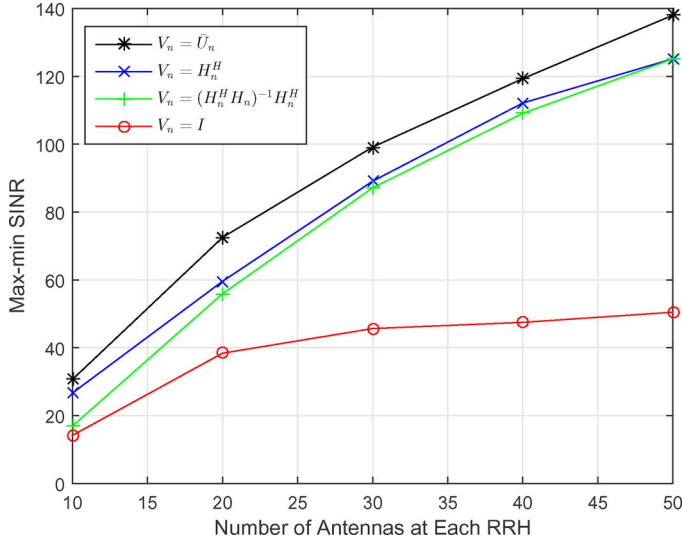


Fig. 5. Performance comparison of different spatial filter designs at RRH versus per-RRH number of antennas.

tization where the received signals from some antennas are not forwarded to the BBU if the quantization bits allocation obtained from Algorithm IV yields  $D_{n,m} = 0$  for antenna  $m$  at RRH  $n$ . With limited per-RRH fronthaul capacity, this scheme is not effective as substantial user signal information is discarded at each RRH due to antenna selection, even with optimized quantization bits allocation. Last, it is observed that as the per-RRH fronthaul capacity increases to 1.5 Gbps, the max-min SINRs achieved with all spatial filter solutions converge to the same value. This is because when the fronthaul capacity is large enough, the quantization error becomes negligible as compared to the receiver noise at each RRH.

Fig. 5 shows the max-min SINR of all the users achieved by Algorithm IV with different spatial filtering designs at the RRHs versus the number of antennas at each RRH, where the per-RRH fronthaul capacity is fixed as  $T = 0.5$  Gbps. It is observed that the proposed filter outperforms both matched filter and ZF filter for all values of  $M$ . Moreover, it is also observed that as  $M$  increases, the performance gap between the schemes with versus without filtering is enlarged. This is because when  $M$  is large, we need to compress the signals received at each RRH more significantly by spatial filtering before scalar quantization, i.e., with  $L \ll M$ .

#### B. Performance Gain of Joint Wireless and Fronthaul Resource Allocation

Next, we show the performance gain of joint optimization for wireless and fronthaul resource allocation. In addition to the proposed Algorithm IV, we consider the following three benchmark schemes as well as one performance upper bound for performance comparison.

- **Benchmark Scheme 1: Optimized quantization bits allocation without power control.** In this scheme, all the users transmit at their maximum power, i.e.,  $p_k = \bar{P}_k, \forall k$ , while the BBU iteratively updates its MMSE based receive beamforming as given in (38) and RRHs' quantization bits allocation via Algorithms II and III.

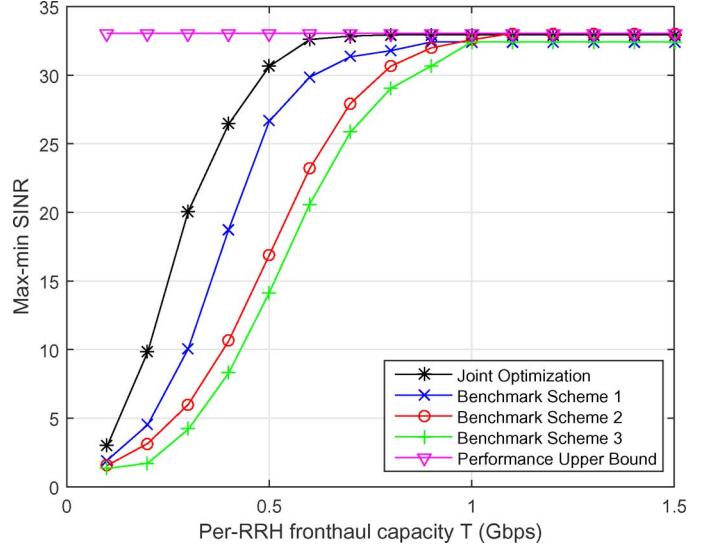


Fig. 6. Performance comparison of different resource allocation schemes in the case of small number of users.

- **Benchmark Scheme 2: Optimized power control with equal quantization bits allocation.** In this scheme, each RRH equally allocates its fronthaul capacity to all the  $L$  dimensions of the output signal after spatial filtering, i.e.,  $D_{n,l} = \bar{T}_n / (2BL), \forall l, n$ . Then, the BBU computes the users' power allocation  $\mathbf{p}$  as well as its beamforming solution  $\mathbf{w}_k$ 's using Algorithm I.
- **Benchmark Scheme 3: Equal quantization bits allocation without power control.** In this scheme, all the users transmit at their maximum power, and each RRH equally allocates its fronthaul capacity to all the  $L$  dimensions of the filter output signal, while the BBU computes the resulting MMSE receive beamforming.
- **Performance Upper Bound.** For this performance upper bound, it is assumed that each RRH  $n$  has an infinite fronthaul capacity, i.e.,  $\bar{T}_n \rightarrow \infty, \forall n$ , in problem (11). In this case, there is no quantization error at each RRH, thus spatial compression is not necessary, i.e., we can set  $\mathbf{V}_n = \mathbf{I} \in \mathbb{C}^{M \times M}, \forall n$ , in problem (11). Moreover, the optimal users' power allocation and BBU's receive beamforming solutions to problem (17) can be efficiently obtained by Algorithm II with  $D_{n,l} = \infty, \forall n, l$ .

Fig. 6 shows the performance comparison of different schemes versus per-RRH fronthaul capacity  $T$  under the same setup as for Fig. 4. It is observed that the proposed joint optimization of wireless and fronthaul resource allocation, i.e., Algorithm IV, achieves higher max-min SINR over Benchmark Schemes 1–3 with separate resource allocation. Moreover, it is observed that under small and moderate per-RRH fronthaul capacity, Benchmark Scheme 1 performs closer to Algorithm IV, as compared to Benchmark Schemes 2 and 3, which equally quantize all the signal dimensions at each RRH. In other words, in this example, most of the joint resource allocation gain is due to joint quantization bits allocation at RRHs. Note that in this example, the total number of antennas from all RRHs is  $MN = 40$ , which is much larger than that of the users  $K = 8$ . Since there are sufficient spatial dimensions to separate the users' signals, power control becomes less effective than

TABLE V  
NUMBER OF QUANTIZED DIMENSIONS AT EACH RRH.

RRH Index	1	2	3	4
No. of Quantized Dimensions ( $T=0.1\text{Gbps}$ )	3	3	3	3
No. of Quantized Dimensions ( $T=0.5\text{Gbps}$ )	6	7	5	7
No. of Quantized Dimensions ( $T=1.0\text{Gbps}$ )	8	8	8	8

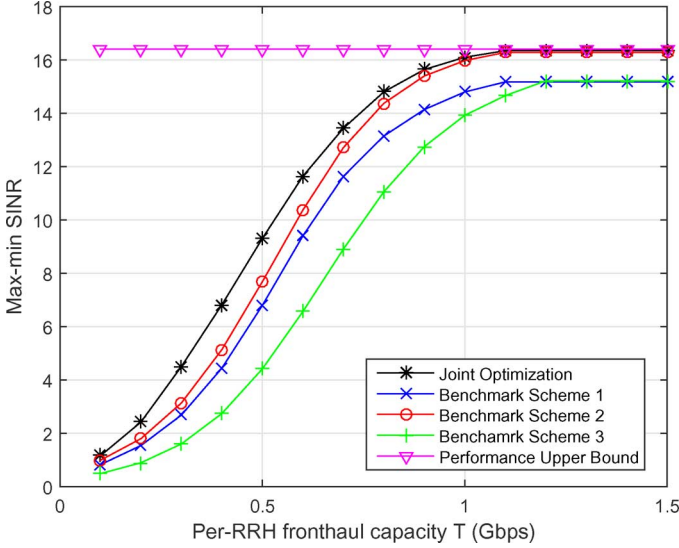


Fig. 7. Performance comparison of different resource allocation schemes in the case of large number of users.

quantization bits allocation. Table V shows the number of quantized dimensions with  $D_{n,l} > 0$ ,  $l = 1, \dots, L$ , at each RRH  $n$  when the per-RRH fronthaul capacity is  $T = 0.1, 0.5, 1$  Gbps, respectively. It is observed that when  $T = 0.1$  Gbps, each RRH should quantize 3 out of  $L = \min(M, K) = 8$  signal dimensions. As the fronthaul capacity increases, more signal dimensions can be quantized. Particularly, when the fronthaul capacity is sufficiently large, i.e.,  $T = 1$  Gbps, all the 8 signal dimensions after spatial filtering should be quantized at all RRHs. Last but not least, it is observed that when  $T > 1$  Gbps, the gap between the max-min SINR achieved by our proposed joint optimization solution or Benchmark Scheme 2 against the performance upper bound vanishes. The reason is as follows. When the common fronthaul capacity  $T$  is large enough, the impact of quantization error on the max-min SINR performance is negligible, and thus the scheme with optimal users' power control can achieve the performance upper bound closely.

Fig. 7 shows the performance comparison for the case of large number of users with  $K = 20$ . It is observed that Algorithm IV still outperforms Benchmark Schemes 1–3. However, different from the case with smaller number of users ( $K = 8$ ) shown in Fig. 6, it is observed that with  $K = 20$ , Benchmark Scheme 2 performs better than Benchmark Schemes 1 and 3. This is because with more users but fixed total number of spatial degrees of freedom  $MN = 40$ , the inter-user interference is more severe and thus power control becomes more effective.

### C. Multi-Antenna C-RAN Versus Massive MIMO

As discussed in Section I, besides multi-antenna C-RAN considered in this paper, massive MIMO and single-antenna

C-RAN are two promising techniques proposed for 5G wireless networks. Therefore, an interesting as well as important question we seek to address in this subsection is as follows: *given a total amount of antennas to be deployed, should we equip them in one single BS, i.e., massive MIMO, or distribute them over a given area by connecting to the BBU via finite-rate fronthaul links, i.e., C-RAN? Moreover, if C-RAN is preferred, what is the optimal antenna deployment solution, i.e., single-antenna C-RAN versus multi-antenna C-RAN?* Intuitively, if more single-antenna RRHs are deployed in the network, with higher probability each user can be served by one or more nearby RRHs with strong channel conditions. However, with multi-antenna RRHs, we can efficiently perform SCF at each RRH to better utilize the limited fronthaul capacity given for each RRH. In the following, we provide a case of study to show the advantage of multi-antenna C-RAN over its two extreme counterparts: massive MIMO with all the antennas deployed at one BS and single-antenna C-RAN with only one antenna at each RRH, i.e.,  $M = 1$ .

To make a fair comparison, we assume that there are in total  $\bar{M}$  antennas to serve  $K$  users in a given area. Specifically, for the massive MIMO system, we assume that there is only one BS which is equipped with all  $\bar{M}$  antennas, while for the C-RAN, we assume that there are  $N$ ,  $1 \leq N \leq \bar{M}$ , RRHs each equipped with  $\bar{M}/N$  antennas.<sup>3</sup> Moreover, we assume there is a total fronthaul capacity for all RRHs in C-RAN, denoted by  $\bar{T}$ , which is a constant regardless of the number of RRHs  $N$ . Note that this is a valid assumption since in practice the signals from different RRHs in the same area are usually first multiplexed at a local hub, which then forwards the signals to the BBU via a fronthaul link with capacity  $\bar{T}$ . For simplicity, we set  $\bar{T}_n = \bar{T}/N$  for all the  $N$  RRHs in the case of C-RAN.

Fig. 8 shows the max-min SINR performance comparison between massive MIMO versus C-RAN with different number of RRHs and fronthaul sum-capacity. The total number of antennas is  $\bar{M} = 50$ . Moreover, there are  $K = 20$  users randomly located in a circle area of radius 700 m. For the massive MIMO, the BS is located in the center of the circle, while for the C-RAN, the RRHs are randomly located in the circle. Note that since the BS in massive MIMO system directly decodes the user messages, the optimal power allocation for users and decoding beamforming vectors at the BS can be obtained by Algorithm I with  $\bar{D}_{n,l} = \infty$ ,  $\forall n, l$ . It is observed from Fig. 8 that if the fronthaul sum-capacity is sufficiently large, i.e.,  $\bar{T} = 3$  Gbps or 5 Gbps, the max-min SINR in C-RAN is larger than that in massive MIMO when the number of RRHs  $N > 4$ . This is because the densification gain due to larger number of RRHs dominates the performance of C-RAN. However, if the fronthaul sum-capacity is limited, i.e.,  $\bar{T} = 1$  Gbps, the performance of C-RAN with any  $N$  is inferior to that of massive MIMO, since quantization errors dominate its performance. Moreover, for the case of single-antenna C-RAN with  $N = 50$  single-antenna RRHs (which is not shown in Fig. 8 for brevity), the achieved max-min SINRs for the cases of  $\bar{T} = 1, 3, 5$  Gbps are 2.25, 10.23 and 22.08, respectively, which are significantly lower than the maximum SINRs achieved in multi-antenna C-RAN with the corre-

<sup>3</sup>In the case that  $\bar{M}/N$  is not an integer, we can simply let some RRHs possess  $\lfloor \bar{M}/N \rfloor$  antennas, while the others possess  $\lceil \bar{M}/N \rceil$  antennas such that the total number of antennas is equal to  $\bar{M}$ .

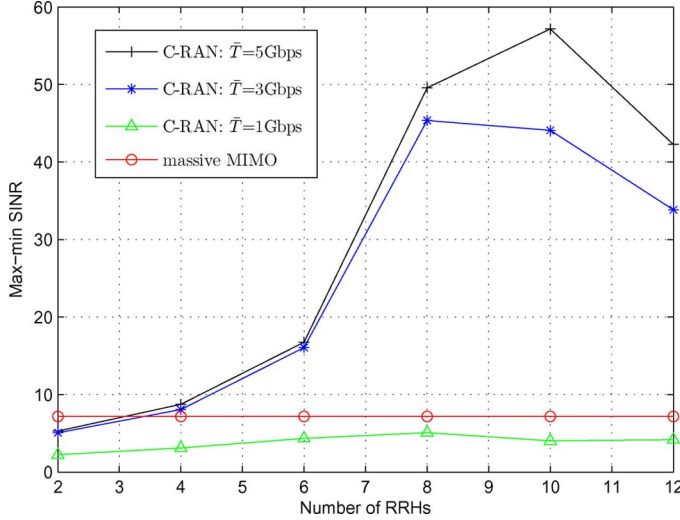


Fig. 8. Performance comparison between multi-antenna C-RAN and massive MIMO.

sponding optimal number of RRHs shown in Fig. 8. This is because in the case of single-antenna C-RAN, the received signals at nearby RRHs are highly correlated, and as a result independent scalar quantization at these RRHs without spatial filtering is inefficient in utilization of the given fronthaul sum-capacity. Last, it is observed that the achieved max-min SINR in multi-antenna C-RAN first increases with the number of RRHs, but then decreases after a certain optimal number of RRHs for all three cases of  $\bar{T} = 1, 3, 5$  Gbps. This validates the effectiveness of multi-antenna C-RAN by more flexibly balancing between the performance and fronthaul trade-off. As  $\bar{T}$  increases, it is observed that more RRHs should be deployed in C-RAN to exploit better channels from all users.

## VI. CONCLUSIONS

This paper considers a flexible antenna deployment design for C-RAN termed multi-antenna C-RAN and proposes a new “spatial-compression-and-forward (SCF)” scheme for efficient and low-complexity processing at each RRH in the uplink multiuser communication. With the proposed distributed spatial filters at multi-antenna RRHs, a joint optimization across the wireless transmission, the fronthaul quantization and the decoding at the BBU is performed to maximize the minimum SINR of all the users. Our results show that the proposed SCF scheme with joint resource allocation achieves significant performance gains over the conventional “quantize-and-forward” based single-antenna C-RAN as well as massive MIMO, thanks to the more flexibly optimized antenna deployment in multi-antenna C-RAN.

## APPENDIX

### Proof of Lemma 4.1:

First, since

$$\left( \sum_{j \neq k} p_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \right)^{-1} \succ \mathbf{0}, \forall k,$$

we have  $\mathbf{I}(\mathbf{p}) > \mathbf{0}$  if  $\mathbf{p} > \mathbf{0}$ . The first property in Lemma 4.1 is proved.

Next, if  $\mathbf{p} \geq \mathbf{p}'$ , then we have

$$\begin{aligned} \sum_{j \neq k} p_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \\ \succeq \sum_{j \neq k} p'_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(\mathbf{p}', \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}), \quad \forall k. \end{aligned} \quad (55)$$

**Lemma A.1:** [23, Corollary 7.7.4] If  $\mathbf{A} \succ \mathbf{0}$ ,  $\mathbf{B} \succ \mathbf{0}$ , and  $\mathbf{A} - \mathbf{B} \succeq \mathbf{0}$ , then  $\mathbf{A}^{-1} - \mathbf{B}^{-1} \preceq \mathbf{0}$ .

According to Lemma A.1, it follows that

$$\begin{aligned} \left( \sum_{j \neq k} p_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \right)^{-1} \\ \preceq \left( \sum_{j \neq k} p'_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(\mathbf{p}', \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \right)^{-1}, \quad \forall k. \end{aligned} \quad (56)$$

As a result, we have  $\mathbf{I}(\mathbf{p}) \geq \mathbf{I}(\mathbf{p}')$  if  $\mathbf{p} \geq \mathbf{p}'$ . The second property in Lemma 4.1 is thus proved.

Last, if  $a > 1$ , then we have

$$\begin{aligned} a \left( \sum_{j \neq k} p_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \right) \\ \succ \sum_{j \neq k} ap_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(a\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}), \quad \forall k. \end{aligned} \quad (57)$$

According to Lemma 4.1, we have

$$\begin{aligned} \frac{1}{a} \left( \sum_{j \neq k} p_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \right)^{-1} \\ \prec \left( \sum_{j \neq k} ap_j \tilde{\mathbf{h}}_{(j)} \tilde{\mathbf{h}}_{(j)}^H + \sigma^2 \mathbf{I} + \mathbf{Q}(a\mathbf{p}, \{\bar{\mathbf{U}}_n, \bar{\mathbf{D}}_n\}) \right)^{-1}, \quad \forall k. \end{aligned} \quad (58)$$

As a result, it follows that  $\forall a > 1$ ,  $a\mathbf{I}(\mathbf{p}) > \mathbf{I}(a\mathbf{p})$ . The third property of Lemma 4.1 is thus proved.

To summarize,  $\mathbf{I}(\mathbf{p})$  given in (42) is a standard interference function. Lemma 4.1 is thus proved.

### Proof of Corollary 4.2:

If the initial point of the fixed-point method is  $\mathbf{p}^{(0)} = \mathbf{0}$ , then we have  $\mathbf{p}^{(1)} = \mathbf{I}(\mathbf{p}^{(0)}) > \mathbf{0} = \mathbf{p}^{(0)}$ . Moreover, according to the second property of Lemma 4.1, if  $\mathbf{p}^{(i)} \geq \mathbf{p}^{(i-1)}$ , then  $\mathbf{p}^{(i+1)} = \mathbf{I}(\mathbf{p}^{(i)}) \geq \mathbf{I}(\mathbf{p}^{(i-1)}) = \mathbf{p}^{(i)}$ . As a result, with the initial point  $\mathbf{p}^{(0)} = \mathbf{0}$ , the resulted power solution increases after each iteration, i.e.,  $\mathbf{p}^{(0)} < \mathbf{p}^{(1)} \leq \mathbf{p}^{(2)} \leq \dots$ . Suppose that at the  $i$ th iteration, the above procedure converges to a finite power solution  $\mathbf{p}^{(i)}$ . Then we have  $\mathbf{p}^{(i+1)} = \mathbf{p}^{(i)} = \mathbf{I}(\mathbf{p}^{(i)})$ . As a result,  $\mathbf{p}^{(i)}$  can satisfy the SINR constraints (44), which contradicts to the fact that  $\bar{\gamma}$  cannot be achieved by all the users. To summarize, the power solution by the fixed-point method with  $\mathbf{p}^{(0)} = \mathbf{0}$  will increase to an infinity power allocation. Corollary 4.2 is thus proved.

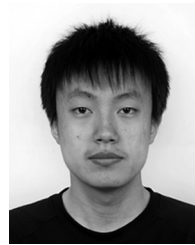


### Proof of Corollary 4.3:

Let  $\mathbf{p}^{(0)}$  denote a feasible solution to problem (43). Since it satisfies (44), we have  $\mathbf{p}^{(0)} \geq \mathbf{I}(\mathbf{p}^{(0)}) = \mathbf{p}^{(1)}$ . In the following, we show that  $\mathbf{p}^{(i+1)} \leq \mathbf{p}^{(i)}$  holds  $\forall i \geq 1$  by induction. Suppose  $\mathbf{p}^{(i)} \leq \mathbf{p}^{(i-1)}$ . Then we have  $\mathbf{p}^{(i+1)} = \mathbf{I}(\mathbf{p}^{(i)}) \stackrel{(a)}{\leq} \mathbf{I}(\mathbf{p}^{(i-1)}) = \mathbf{p}^{(i)}$ , where (a) is due to the second property in Lemma 4.1. Since we have  $\mathbf{p}^{(1)} \leq \mathbf{p}^{(0)}$ , we thus have  $\mathbf{p}^{(i+1)} \leq \mathbf{p}^{(i)}$  holds  $\forall i \geq 1$ . In other words, given any feasible power solution  $\mathbf{p}'$  as the initial point, a monotonic convergence can be guaranteed by the fixed-point method, i.e.,  $\mathbf{p}' \geq \mathbf{p}^*$ . Corollary 4.3 is thus proved.

### REFERENCES

- [1] "Report of 3GPP RAN Workshop on Release 12 and Onwards," 3GPP, RWS- 120052, 2012, 3GPP Workshop on Release 12 Onward.
- [2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.
- [4] China Mobile Research Inst., "C-RAN: The Road Towards Green RAN," White Paper, Ver. 2.5 2011.
- [5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [6] A. Sanderovich, S. Shamai, and Y. Steinberg, "Distributed MIMO receiver—Achievable rates and upper bounds," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4419–4438, Oct. 2009.
- [7] A. Sanderovich, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Uplink macro diversity of limited backhaul cellular network," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3457–3478, Aug. 2009.
- [8] S. H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint decompression and decoding for cloud radio access networks," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 503–506, May 2013.
- [9] A. D. Coso and S. Simoes, "Distributed compression for MIMO coordinated networks with a backhaul constraint," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4698–4709, Sep. 2009.
- [10] S. H. Park, O. Simeone, O. Sahin, and S. Shamai, "Robust and efficient distributed compression for cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 692–703, Feb. 2013.
- [11] Y. Zhou and W. Yu, "Optimized backhaul compression for uplink cloud radio access network," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1295–1307, Jun. 2014.
- [12] L. Zhou and W. Yu, "Uplink multicell processing with limited backhaul via per-base-station successive interference cancellation," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 10, pp. 1981–1993, Oct. 2013.
- [13] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.
- [14] S. H. Lim, Y. H. Kim, A. El Gamal, and S. Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.
- [15] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, "Communication via decentralized processing," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3008–3023, Jul. 2008.
- [16] L. Liu, S. Bi, and R. Zhang, "Joint power control and fronthaul rate allocation for throughput maximization in OFDMA-based cloud radio access network, submitted to," *IEEE Trans. Commun.*
- [17] S. Luo, R. Zhang, and T. J. Lim, "Downlink and uplink energy minimization Through user association and beamforming in cloud RAN," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 494–508, Jan. 2015.
- [18] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loeve transform," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5171–5196, Dec. 2006.
- [19] I. D. Schizas, G. B. Giannakis, and Z. Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4284–4299, Aug. 2007.
- [20] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1341–1347, Sep. 1995.
- [21] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.
- [22] L. Zhang, Y. C. Liang, and Y. Xin, "Joint beamforming and power control for multiple access channels in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 38–51, Jan. 2008.
- [23] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [25] "Further advancements for E-UTRA physical layer aspects," 3GPP, TR 36.814 V9.0.0, 2010.



**Liang Liu** (S'14–M'15) received the B.Eng degree from the Tianjin University, China, in 2010, and the Ph.D. degree from the National University of Singapore in 2014. He is currently a Research Fellow in the Department of Electrical and Computer Engineering at the National University of Singapore. His research interests include convex optimization, resource allocation in interference channel, energy harvesting, and 5G. He was the recipient of a Best Paper Award from IEEE WCSP 2011.



**Rui Zhang** (S'00–M'07–SM'15) received the B.Eng. (First-Class Hons.) and M.Eng. degrees from the National University of Singapore in 2000 and 2001, respectively, and the Ph.D. degree from the Stanford University, Stanford, CA USA, in 2007, all in electrical engineering.

From 2007 to 2010, he worked at the Institute for Infocomm Research, A-STAR, Singapore, where he now holds a Senior Research Scientist joint appointment. Since 2010, he has joined the Department of Electrical and Computer Engineering of the National

University of Singapore as an Assistant Professor. His current research interests include energy-efficient and energy-harvesting-enabled wireless communications, wireless information and power transfer, multiuser MIMO, cognitive radio, smart grids, and optimization methods.

Dr. Zhang has published over 200 papers. He was the co-recipient of the Best Paper Award from the IEEE PIMRC in 2005, and the IEEE Marconi Prize Paper Award in Wireless Communications in 2015. He was the recipient of the 6th IEEE Communications Society Asia-Pacific Best Young Researcher Award in 2011, the Young Investigator Award and the Young Researcher Award of the National University of Singapore in 2011 and 2015, respectively. He has served for over 30 IEEE conferences as TPC members and Organizing Committee members. He is now an elected member of IEEE Signal Processing Society SPCOM and SAM Technical Committees, and serves as the Vice Chair of the IEEE ComSoc Asia-Pacific Board Technical Affairs Committee. He is an editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the Green Communications and Networking Series of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.