

End-to-End Rate Enhancement in C-RAN Using Multi-Pair Two-Way Computation

Mahmoud A. Hasabelnaby and Anas Chaaban

School of Engineering, University of British Columbia, Kelowna, V1V1V7 BC, Canada

{mahmoud.hasabelnaby,anas.chaaban}@ubc.ca

Abstract—Cloud radio-access networks (C-RAN) have been proposed as an enabling technology for keeping up with the requirements of next-generation wireless networks. Most existing works on C-RAN consider the uplink or the downlink separately. However, designing the uplink and the downlink jointly may bring additional advantages, especially if message source-destination information is taken into account. In this paper, this idea is demonstrated by considering pairwise message exchange between users in a C-RAN. A multi-pair two-way transmission scheme is proposed which targets maximizing the end-to-end user data rates. The achievable rate of this scheme is derived, optimized, and evaluated numerically. Results reveal that significant end-to-end rate improvement can be achieved using the proposed scheme compared to existing schemes.

Index Terms—C-RAN; Compress-and-forward; Compute-and-forward; Reverse quantized-compute-and-forward.

I. INTRODUCTION

The cloud radio-access network (C-RAN) architecture is one of the methods that enhance communication capabilities towards meeting the critical requirements of next-generation wireless networks. However, the performance of a C-RAN is constrained by the limited capacity of fronthaul links that connect RRHs with the baseband processing unit (BBU) pool. This puts a constraint on the amount of information that can be exchanged between RRHs and the BBU pool. Therefore, advanced signal processing and relaying techniques are required in both uplink and downlink in order to make efficient use of the fronthaul links [1]. This has been the topic of many studies recently as discussed next.

In the uplink, multiple users transmit their codewords to the RRHs. Different relaying strategies can be used to relay information from RRHs to the BBU pool, such as decode-and-forward (DF) [2], compute-and-forward (CoF) [3], and compress-and-forward (CF) [4]. CoF is better than the CF under small fronthaul capacity values. However, the average performance of CF is better at moderate and high fronthaul capacities [2], [5], [6]. In the downlink, transmission strategies that enable cooperation between RRHs include data-sharing strategies [7], [8], compression based strategies [2], [9], reverse compute-and-forward (RCoF) [10], and reverse quantized-compute-and-forward (RQCoF) [11]. Since the exact characterization of the downlink C-RAN capacity is still an open problem, most works optimize the schemes using uplink-downlink duality to achieve downlink rates greater than or equal to the uplink rates [12], [13].

Note that all aforementioned works study either the uplink or the downlink, separately. This may incur performance loss, especially in scenarios where intra-cloud message exchange is desired. This paper demonstrates this idea via studying a C-RAN with intra-cloud pairwise communication. Such a scenario can occur in video conferencing or gaming applications for instance. A multi-pair two-way transmission scheme is proposed to maximize the end-to-end achievable rate. Using a lattice-based compression strategy, RRHs compress their observations and forward them to the BBU pool, which in turn computes integer linear combinations of codeword-pairs. This reduces the required number of computation steps at the BBU pool, thereby reducing the number of rate constraints. The BBU pool compresses the linear combinations and forwards them to the RRHs, which decompress the signals and transmit them to the users. Finally, users decode their desired message using their own messages as side information. The achievable rate of the scheme is derived, optimized, and evaluated numerically, showing superior performance to existing schemes in the literature.

In the sequel, the following notations will be used. Column vectors and matrices are denoted by boldface lowercase and uppercase letters, such as \mathbf{x} and \mathbf{X} , respectively. The transpose of matrix \mathbf{X} is denoted by \mathbf{X}^\top . All the logarithms are to the base 2, and $\log^+(x) = \max(0, \log(x))$.

II. SYSTEM MODEL

We consider a C-RAN consisting of M single-antenna user pairs ($K = 2M$ users), L single-antenna RRHs,¹ and a central processor (BBU pool). Each RRH $\ell \in \{1, \dots, L\}$, is connected to the BBU pool via a digital noiseless fronthaul link with a limited capacity C_ℓ . User pairs wish to communicate with each other using the C-RAN architecture. In other words, users $k, k' \in \{1, \dots, K\}$, $k \neq k'$, exchange messages with each other. As a result, a pairing matrix \mathbf{W} with dimensions of $M \times K$ is defined, where $w_{m,k} \in \{0, 1\}$ is a user-pair association indicator, i.e., $w_{m,k} = 1$ if user k belongs to pair m , and $w_{m,k} = 0$ otherwise. Note that $\sum_{m=1}^M w_{m,k} = 1$ for all k . The message exchange among the user-pairs is completed in two stages over n channel uses each, an uplink phase and a downlink phase.²

¹This work can be extended to deal with MIMO systems.

²We assume that the uplink and downlink occur over the same frequency band using a half-duplex fashion

In the uplink, user $k \in \{1, \dots, K\}$ encodes its message \mathbf{g}_k with rate R_k into a codeword $\mathbf{x}_k^{ul} \in \mathbb{R}^n$ and sends it to the RRHs. This signal is subject to a power constraint P_k^{ul} . Note that we consider a real-valued transmission model for the sake of simplicity, bearing in mind that a complex-valued model can be addressed using the real-valued vector representation. The received signal at the RRHs is given by

$$\mathbf{Y}^{ul} = \mathbf{H}^{ul} \mathbf{X}^{ul} + \mathbf{Z}^{ul} \quad (1)$$

where $\mathbf{Y}^{ul} = [\mathbf{y}_1^{ul}, \dots, \mathbf{y}_L^{ul}]^\top$, $\mathbf{y}_\ell^{ul} \in \mathbb{R}^n$ is the received signal at RRH ℓ , $\mathbf{H}^{ul} \in \mathbb{R}^{L \times K}$ is the uplink channel gain matrix between all users and RRHs, $\mathbf{X}^{ul} = [\mathbf{x}_1^{ul}, \dots, \mathbf{x}_K^{ul}]^\top$, and $\mathbf{Z}^{ul} \in \mathbb{R}^{L \times n}$ is additive white Gaussian noise with independent and identically distributed (i.i.d.) components with zero mean and unit variance $\mathcal{N}(0, 1)$. We assume that channels are Rayleigh fading and remain fixed across the transmission of a complete codeword (block fading). RRH $\ell \in \{1, \dots, L\}$ processes the signal \mathbf{y}_ℓ^{ul} into a message e_ℓ^{ul} , and sends this message to the BBU pool using the fronthaul link.

The downlink is described as follows. First, the BBU processes the received messages from all RRHs, then constructs messages e_ℓ^{dl} , $\ell \in \{1, \dots, L\}$, and sends e_ℓ^{dl} to RRH ℓ in the downlink using the fronthaul links. The RRH then processes e_ℓ^{dl} to construct a transmit signal $\mathbf{x}_\ell^{dl} \in \mathbb{R}^n$ with power constraint $p_\ell^{dl} \leq P_\ell^{dl}$ for transmission to the users. The received signals at users can be written in a matrix form as

$$\mathbf{Y}^{dl} = \mathbf{H}^{dl} \mathbf{X}^{dl} + \mathbf{Z}^{dl}, \quad (2)$$

where $\mathbf{Y}^{dl} = [\mathbf{y}_1^{dl}, \dots, \mathbf{y}_K^{dl}]^\top$, $\mathbf{H}^{dl} \in \mathbb{R}^{K \times L}$ is the downlink channel gain matrix between all RRHs and users, $\mathbf{X}^{dl} = [\mathbf{x}_1^{dl}, \dots, \mathbf{x}_L^{dl}]^\top$, and $\mathbf{Z}^{dl} \in \mathbb{R}^{K \times n}$ is additive white Gaussian noise with i.i.d $\mathcal{N}(0, 1)$ components. Note that we assume channel reciprocity, i.e., $\mathbf{H}^{dl} = \mathbf{H}^{ul\top}$. Finally, each user uses its received signal in combination with its own message to decode the message of the paired user.

The goal is to design an uplink/downlink transmission scheme which takes this user pairing into account, and to derive its achievable rate. This is discussed in the following sections which discuss the uplink phase and the downlink phase, respectively.

III. UPLINK TRANSMISSION

A. Encoding at the users

Using nested lattice coding [3], the lattice chain $\Lambda_c \subseteq \Lambda_{f,K} \subseteq \dots \subseteq \Lambda_{f,1}$ is generated, consisting of n -dimensional lattices. The coarse lattice Λ_c is chosen to be good for channel coding and quantization simultaneously, whereas the fine lattices $\Lambda_{f,k}$, $k \in \{1, 2, \dots, K\}$ are good for quantization only. User k generates its nested lattice codebook as $\mathcal{C}_k^{ul} = \{\Lambda_{f,k} \cap v_{\Lambda_c}\}$, where v_{Λ_c} is the fundamental Voronoi region of the coarse lattice. Using a one-to-one mapping, it maps its message \mathbf{g}_k into a lattice point $\mathbf{s}_k^{ul} \in \mathcal{C}_k^{ul}$. To make the transmitted signal independent on the lattice point, a random dither \mathbf{t}_k^{ul} , uniformly distributed over v_{Λ_c} and known to all

nodes in the network, is added. The result is then reduced using a modulo-lattice operation with respect to Λ_c . This leads to the signal

$$\mathbf{u}_k^{ul} = (\mathbf{s}_k^{ul} + \mathbf{t}_k^{ul}) \bmod \Lambda_c \quad (3)$$

which is then scaled by b_k^{ul} , and transmitted (i.e., $\mathbf{x}_k^{ul} = b_k^{ul} \mathbf{u}_k^{ul}$) to the RRHs. The transmitted signals from all users can be written as $\mathbf{X}^{ul} = \mathbf{B}^{ul} \mathbf{U}^{ul}$, where $\mathbf{B}^{ul} = \text{diag}(b_1^{ul}, \dots, b_K^{ul})$ is a $K \times K$ scaling matrix and $\mathbf{U}^{ul} = [\mathbf{u}_1^{ul}, \dots, \mathbf{u}_K^{ul}]^\top$ is the dithered codewords matrix. The uplink power constraint per user can be met by selecting the coarse lattice Λ_c with second moment $\sigma^2(\Lambda_c) = p^{ul}$ and assigning proper scaling value b_k^{ul} , so that $\frac{1}{n} \mathbb{E}[\|\mathbf{x}_k^{ul}\|^2] = (b_k^{ul})^2 p^{ul} \leq P_k^{ul}$.

B. Compression at the RRHs

The received signal \mathbf{y}_ℓ^{ul} at RRH ℓ is processed as follows. Given a lattice chain $\tilde{\Lambda}_{c,1} \subseteq \dots \subseteq \tilde{\Lambda}_{c,L} \subseteq \tilde{\Lambda}_{f,L} \subseteq \dots \subseteq \tilde{\Lambda}_{f,1}$, RRH ℓ generates its codebook as $\kappa_\ell^{ul} = \{\tilde{\Lambda}_{f,\ell} \cap v_{\tilde{\Lambda}_{c,\ell}}\}$ with rate equal to its fronthaul capacity C_ℓ . The coarse lattice $\tilde{\Lambda}_{c,\ell}$ is good for channel coding and quantization simultaneously, thereby the probability of error can be neglected. In addition, the fine lattice $\tilde{\Lambda}_{f,\ell}$ must be good for quantization in order to be able to find a precise relationship between the quantization rates and distortion levels [14]. Then, the ℓ th RRH adds a random dither vector $\tilde{\mathbf{t}}_\ell^{ul}$, uniformly distributed over $v_{\tilde{\Lambda}_{f,\ell}}$, to its observation to make the quantization error independent of the received signal \mathbf{y}_ℓ^{ul} . Moreover, using its generated codebook κ_ℓ^{ul} , the ℓ th RRH compresses its dithered observed signal using a lattice-based vector quantization as follows

$$\tilde{\mathbf{y}}_\ell^{ul} = [Q_{\tilde{\Lambda}_{f,\ell}}(\mathbf{y}_\ell^{ul} + \tilde{\mathbf{t}}_\ell^{ul})] \bmod \tilde{\Lambda}_{c,\ell}. \quad (4)$$

Then, RRH ℓ maps $\tilde{\mathbf{y}}_\ell^{ul}$ to an index $e_\ell^{ul} \in \{1, \dots, 2^{n C_\ell}\}$, and forwards it to the BBU pool via its fronthaul link.

C. Decompression at the BBU pool

Once the BBU pool receives the indices $e_1^{ul}, \dots, e_L^{ul}$, it recovers $\tilde{\mathbf{y}}_1^{ul}, \dots, \tilde{\mathbf{y}}_L^{ul}$, then subtracts the dithers, and reduces the result using the modulo-lattice operation with respect to $\tilde{\Lambda}_{c,\ell}$ as

$$\begin{aligned} \tilde{\mathbf{y}}_\ell^{ul} &= [\tilde{\mathbf{y}}_\ell^{ul} - \tilde{\mathbf{t}}_\ell^{ul}] \bmod \tilde{\Lambda}_{c,\ell} \\ &\stackrel{(a)}{=} [\mathbf{y}_\ell^{ul} + \mathbf{q}_\ell^{ul}] \bmod \tilde{\Lambda}_{c,\ell} = [\hat{\mathbf{y}}_\ell^{ul}] \bmod \tilde{\Lambda}_{c,\ell} \end{aligned} \quad (5)$$

where (a) is obtained from the distributive law of the modulo-lattice operation, $\mathbf{q}_\ell^{ul} = -[\mathbf{y}_\ell^{ul} + \tilde{\mathbf{t}}_\ell^{ul}] \bmod \tilde{\Lambda}_{f,\ell}$ is the compression distortion which is independent of \mathbf{y}_ℓ^{ul} and uniformly distributed over $v_{\tilde{\Lambda}_{f,\ell}}$, and $\hat{\mathbf{y}}_\ell^{ul} = \mathbf{y}_\ell^{ul} + \mathbf{q}_\ell^{ul}$. After that, the BBU pool proceeds to decode L integer linear combinations $\mathbf{v}_{r,1}^{ul}, \dots, \mathbf{v}_{r,L}^{ul}$ as in [5], by exploiting the correlation between the received signals at all RRHs, where

$$\mathbf{v}_{r,i}^{ul} = \left[\sum_{\ell=1}^L a_{r,i,\ell}^{ul} \hat{\mathbf{y}}_\ell^{ul} \right] \bmod \tilde{\Lambda}_{c,i} \stackrel{(b)}{=} \sum_{\ell=1}^L a_{r,i,\ell}^{ul} \hat{\mathbf{y}}_\ell^{ul} \quad (6)$$

$i \in \{1, \dots, L\}$, $a_{r,i,\ell}^{ul} \in \mathbb{Z}$ is an integer coefficient, and (b) is obtained with high probability (w.h.p.) if $\tilde{\Lambda}_{c,i}$ is good for

channel coding and $\frac{1}{n} \mathbb{E}[\|\mathbf{v}_{r,i}^{ul}\|^2] < \sigma^2(\tilde{\Lambda}_{c,i})$ [14]. We write (6) in a matrix form as $\mathbf{V}_r^{ul} = \mathbf{A}_r^{ul} \hat{\mathbf{Y}}^{ul}$, where \mathbf{A}_r^{ul} is a $L \times L$ full-rank integer coefficient matrix with full rank sub-matrices $\mathbf{A}_{r,[1:i]}^{ul}$ for $i \in \{1, \dots, L\}$ and $\hat{\mathbf{Y}}^{ul} = [\hat{\mathbf{y}}_1^{ul}, \dots, \hat{\mathbf{y}}_L^{ul}]^\top$. The integer coefficients can be selected to maximize the computation rate constraint, which allows us to increase the compression rate and decrease the compression distortion.

The compression rate at the ℓ th RRH can be expressed as

$$R_{r,\ell}^{ul} = \frac{1}{2} \log^+ \left(\frac{\mathbf{a}_{r,\ell}^{ul\top} (\mathbf{H}^{ul} \mathbf{B}^{ul} \mathbf{P}^{ul} \mathbf{B}^{ul\top} \mathbf{H}^{ul\top} + \mathbf{I}_L + \mathbf{D}^{ul}) \mathbf{a}_{r,\ell}^{ul}}{d_\ell^{ul}} \right) \quad (7)$$

where $R_{r,\ell}^{ul} \leq C_\ell$, $\mathbf{a}_{r,\ell}^{ul\top}$ is the ℓ th row of \mathbf{A}_r^{ul} , $\mathbf{P}^{ul} = p^{ul} \mathbf{I}_K$ is a $K \times K$ diagonal matrix, \mathbf{I}_L is a $L \times L$ identity matrix, \mathbf{D}^{ul} is a $L \times L$ uplink compression distortion effective covariance matrix whose diagonal elements are equal to $\mathbf{d}^{ul} = [d_1^{ul}, \dots, d_L^{ul}]^\top$, and d_ℓ^{ul} is the distortion power level of the fine lattice $\hat{\Lambda}_{f,\ell}$ at RRH ℓ .

By multiplying \mathbf{V}_r^{ul} by the inverse of the integer coefficient matrix $\mathbf{A}_{r,inv}^{ul}$, the BBU pool can recover

$$\hat{\mathbf{Y}}^{ul} = \mathbf{H}^{ul} \mathbf{X}^{ul} + \mathbf{Z}^{ul} + \mathbf{Q}^{ul} \quad (8)$$

where $\mathbf{Q}^{ul} = [\mathbf{q}_1^{ul}, \dots, \mathbf{q}_L^{ul}]^\top$.

D. Multi-pair computation at the BBU pool

The BBU pool proceeds to decode M integer linear combinations of user-pairs' codewords (instead of decoding K integer linear equations of users' individual codewords as in [5]), to obtain

$$\mathbf{v}_{\psi,j}^{ul} = \left[\sum_{m=1}^M a_{\psi,j,m}^{ul} \left[\sum_{k=1}^K w_{m,k} s_k^{ul} \right] \right] \bmod \Lambda_c, \quad (9)$$

where $j \in \{1, \dots, M\}$ and $a_{\psi,j,m}^{ul} \in \mathbb{Z}$ is an integer coefficient. Computing M combinations instead of K decreases the number of constraints on the computation rate, which improves performance. Note that (9) can be rewritten in matrix form as $\mathbf{V}_\psi^{ul} = [\mathbf{A}_\psi^{ul} \mathbf{W} \mathbf{S}^{ul}] \bmod \Lambda_c$, where \mathbf{A}_ψ^{ul} is an $M \times M$ integer coefficient matrix with a rank of $M = K/2$, and $\mathbf{S}^{ul} = [s_1^{ul}, \dots, s_K^{ul}]^\top$. This decoding can be done by linearly processing $\hat{\mathbf{Y}}^{ul}$ with a scaling equalizer $\boldsymbol{\rho}_j^{ul}$, removing the dither \mathbf{t}_k^{ul} , and reducing the result modulo Λ_c as

$$\begin{aligned} \boldsymbol{\mu}_j^{ul} &= \left[\boldsymbol{\rho}_j^{ul\top} \hat{\mathbf{Y}}^{ul} - \sum_{k=1}^K \mathbf{t}_k^{ul} \right] \bmod \Lambda_c \\ &= \left[\boldsymbol{\rho}_j^{ul\top} \mathbf{H}^{ul} \mathbf{B}^{ul} \mathbf{S}^{ul} + \boldsymbol{\rho}_j^{ul\top} (\mathbf{Z}^{ul} + \mathbf{Q}^{ul}) \right] \bmod \Lambda_c \\ &= \underbrace{[\mathbf{a}_{\psi,j}^{ul\top} \mathbf{W} \mathbf{S}^{ul}]}_{\text{desired signal}} + \mathbf{z}_{\text{eff},j}^{ul} \bmod \Lambda_c \end{aligned} \quad (10)$$

from which the BBU pool computes $[\mathbf{a}_{\psi,j}^{ul\top} \mathbf{W} \mathbf{S}^{ul}] \bmod \Lambda_c$, where $\mathbf{z}_{\text{eff},j}^{ul} = (\boldsymbol{\rho}_j^{ul\top} \mathbf{H}^{ul} \mathbf{B}^{ul} - \mathbf{a}_{\psi,j}^{ul\top} \mathbf{W}) \mathbf{S}^{ul} + \boldsymbol{\rho}_j^{ul\top} (\mathbf{Z}^{ul} + \mathbf{Q}^{ul})$ is the effective noise with power $\sigma_j^{ul2} = \frac{1}{n} \mathbb{E}[\|\mathbf{z}_{\text{eff},j}^{ul}\|^2]$

$$\sigma_j^{ul2} = \left\| (\boldsymbol{\rho}_j^{ul\top} \mathbf{H}^{ul} \mathbf{B}^{ul} - \mathbf{a}_{\psi,j}^{ul\top} \mathbf{W}) (\mathbf{P}^{ul})^{\frac{1}{2}} \right\|^2 + \boldsymbol{\rho}_j^{ul\top} (\mathbf{I}_L + \mathbf{D}^{ul}) \boldsymbol{\rho}_j^{ul} \quad (11)$$

In order to minimize the effective variance in (11), $\boldsymbol{\rho}_j^{ul}$ is chosen as the MMSE scaling equalizer given by

$$\boldsymbol{\rho}_j^{ul\top} = \mathbf{a}_{\psi,j}^{ul\top} \mathbf{W} \mathbf{P}^{ul} \mathbf{H}^{ul\top} (\mathbf{H}^{ul} \mathbf{P}^{ul} \mathbf{H}^{ul\top} + \mathbf{I}_L + \mathbf{D}^{ul})^{-1} \quad (12)$$

By substituting the MMSE solution into (11) and applying the matrix inversion lemma, the effective noise power σ_j^{ul2} can be rewritten as

$$\sigma_j^{ul2} = \mathbf{a}_{\psi,j}^{ul\top} \mathbf{W} [\mathbf{F}_\psi^{ul} \mathbf{F}_\psi^{ul\top}] \mathbf{W}^\top \mathbf{a}_{\psi,j}^{ul} = \|\mathbf{F}_\psi^{ul} \mathbf{W}^\top \mathbf{a}_{\psi,j}^{ul}\|^2 \quad (13)$$

where \mathbf{F}_ψ^{ul} is the Cholesky decomposition satisfying $\mathbf{F}_\psi^{ul} \mathbf{F}_\psi^{ul\top} = (\mathbf{P}^{ul} + \mathbf{H}^{ul\top} (\mathbf{I}_L + \mathbf{D}^{ul})^{-1} \mathbf{H}^{ul})^{-1}$.

Let user k belong to pair m_k , i.e., $w_{m,k} = 1$. Then, the achievable uplink computation rate for user k can be expressed as

$$R_k^{ul} \leq \min_{\substack{j \in \{1, \dots, M\} \\ a_{\psi,j,m_k}^{ul} \neq 0}} \frac{1}{2} \log^+ \left(\frac{p^{ul}}{\sigma_j^{ul2}} \right), \quad (14)$$

where σ_j^{ul2} is given by (13). Instead of recovering the original messages as in [5], the BBU compresses the previously computed equations directly and forwards them to the RRHs through the fronthaul links as described next.

IV. DOWNLINK TRANSMISSION

The basic idea of the downlink is to employ a reverse-quantized-compute-and-forward scheme [11].

A. Compression at the BBU pool

At first, the BBU pool uses the beamforming matrix \mathbf{B}^{dl} with dimensions $L \times M$ to produce

$$\mathbf{S}^{dl} = \mathbf{B}^{dl} \mathbf{V}_\psi^{ul} = \mathbf{B}^{dl} [\mathbf{A}_\psi^{ul} \mathbf{W} \mathbf{S}^{ul}] \bmod \Lambda_c \quad (15)$$

where $\mathbf{S}^{dl} = [s_1^{dl}, \dots, s_L^{dl}]^\top$. In order to enable each RRH to extract its desired quantized signal, the BBU pool pre-inverts the \mathbf{S}^{dl} with $\mathbf{A}_{r,inv}^{dl}$ as follows

$$\mathbf{V}_r^{dl} = \mathbf{A}_{r,inv}^{dl} \mathbf{S}^{dl} \quad (16)$$

where $\mathbf{V}_r^{dl} = [\mathbf{v}_1^{dl}, \dots, \mathbf{v}_L^{dl}]^\top$ and $\mathbf{A}_{r,inv}^{dl}$ is the inverse of the $L \times L$ full rank integer coefficient matrix \mathbf{A}_r^{dl} . Then, the BBU pool uses a lattice chain $\hat{\Lambda}_{c,1} \subseteq \dots \subseteq \hat{\Lambda}_{c,L} \subseteq \hat{\Lambda}_f$, where the coarse lattices and the fine lattice have the same properties as mentioned in the user encoding step. Next, the BBU pool adds a random dither matrix $\hat{\mathbf{T}}^{dl} = [\hat{\mathbf{t}}_1^{dl}, \dots, \hat{\mathbf{t}}_L^{dl}]^\top$ to \mathbf{V}_r^{dl} which is uniformly distributed over $v_{\hat{\Lambda}_f}$. The dithered output is then quantized as

$$\hat{\mathbf{V}}_r^{dl} = Q_{\hat{\Lambda}_f} (\mathbf{V}_r^{dl} + \hat{\mathbf{T}}^{dl}) \quad (17)$$

where $Q_{\hat{\Lambda}_f}$ is applied to each row of the dithered matrix separately. The BBU pool proceeds to generate integer linear combinations $\tilde{\mathbf{V}}_r^{dl} = \mathbf{A}_r^{dl} \hat{\mathbf{V}}_r^{dl}$ and performs the modulo-lattice operation with respect to $\hat{\Lambda}_{c,i}$, $i \in \{1, \dots, L\}$ to each i th row in $\tilde{\mathbf{V}}_r^{dl}$ to obtain

$$\begin{aligned} \tilde{\mathbf{v}}_{r,i}^{dl} &= \left[\mathbf{a}_{r,i}^{dl\top} \hat{\mathbf{V}}_r^{dl} \right] \bmod \hat{\Lambda}_{c,i} \\ &= \left[\mathbf{a}_{r,i}^{dl\top} Q_{\hat{\Lambda}_f} (\mathbf{V}_r^{dl} + \hat{\mathbf{T}}^{dl}) \right] \bmod \hat{\Lambda}_{c,i} \end{aligned} \quad (18)$$

Finally, the BBU pool maps its compressed linear equation $\tilde{\mathbf{v}}_{r,i}^{dl}$ to an index $e_i^{dl} \in \{1, \dots, 2^{n_{C_i}}\}$, and forwards it to the i th RRH.

B. Decompression at the RRHs

Once the ℓ th RRH receives the index e_ℓ^{dl} , it recovers $\tilde{\mathbf{v}}_{r,\ell}^{dl}$, then subtracts the dither $\hat{\mathbf{T}}^{dl}$, and reduces the result using the modulo-lattice operation with respect to $\hat{\Lambda}_{c,\ell}$ to obtain

$$\begin{aligned} \mathbf{x}_\ell^{dl} &= \left[\tilde{\mathbf{v}}_{r,\ell}^{dl} - \mathbf{a}_{r,\ell}^{dl\top} \hat{\mathbf{T}}^{dl} \right] \bmod \hat{\Lambda}_{c,\ell} \\ &= \left[\mathbf{a}_{r,\ell}^{dl\top} (\mathbf{V}_r^{dl} + \hat{\mathbf{T}}^{dl} + \mathbf{Q}^{dl}) - \mathbf{a}_{r,\ell}^{dl\top} \hat{\mathbf{T}}^{dl} \right] \bmod \hat{\Lambda}_{c,\ell} \\ &\stackrel{(c)}{=} \mathbf{s}_\ell^{dl\top} + \mathbf{a}_{r,\ell}^{dl\top} \mathbf{Q}^{dl} \end{aligned} \quad (19)$$

where $\mathbf{Q}^{dl} = [\mathbf{q}_1^{dl}, \dots, \mathbf{q}_L^{dl}]^\top$ is the downlink compression distortion with a $L \times L$ effective covariance matrix \mathbf{D}^{dl} whose diagonal elements is equal to $\text{diag}(d_1^{dl}, \dots, d_L^{dl})$, and (c) is obtained w.h.p. if $\frac{1}{n} \mathbb{E}[\|\mathbf{x}_\ell^{dl}\|^2] < \sigma^2(\hat{\Lambda}_{c,\ell})$. The downlink compression rate at RRH ℓ is given by

$$R_{r,\ell}^{dl} = \frac{1}{2} \log^+ \left(\frac{\mathbf{b}_\ell^{dl\top} \mathbf{P}_\psi^{ul} \mathbf{b}_\ell^{dl} + \mathbf{a}_{r,\ell}^{dl\top} \mathbf{D}^{dl} \mathbf{a}_{r,\ell}^{dl}}{d_\ell^{dl}} \right) \quad (20)$$

where $R_{r,\ell}^{dl} \leq C_\ell$ and $\mathbf{P}_\psi^{ul} = p^{ul} \mathbf{I}_M$ is $M \times M$ diagonal power matrix whose diagonal elements are equal to p^{ul} . Finally, after the ℓ th RRH recovers its desired signal, it broadcasts $\mathbf{x}_\ell^{dl} \in \mathbb{R}^n$ to the users with power

$$\frac{1}{n} \mathbb{E}[\|\mathbf{x}_\ell^{dl}\|^2] = \mathbf{b}_\ell^{dl\top} \mathbf{P}_\psi^{ul} \mathbf{b}_\ell^{dl} + \mathbf{a}_{r,\ell}^{dl\top} \mathbf{D}^{dl} \mathbf{a}_{r,\ell}^{dl} \triangleq p_\ell^{dl} \leq P_\ell^{dl}. \quad (21)$$

C. Decoding at the users

The received signals at all the users can be written in a matrix form as

$$\mathbf{Y}^{dl} = \mathbf{H}^{dl} (\mathbf{B}^{dl} \mathbf{V}_\psi^{ul} + \mathbf{A}_r^{dl} \mathbf{Q}^{dl}) + \mathbf{Z}^{dl}. \quad (22)$$

The k th user scales its received signal \mathbf{y}_k^{dl} by a linear scaling ρ_k^{dl} and reduces the result modulo Λ_c as follows

$$\begin{aligned} \boldsymbol{\mu}_k^{dl} &= [\rho_k^{dl} \mathbf{y}_k^{dl}] \bmod \Lambda_c \\ &= [\rho_k^{dl} \mathbf{h}_k^{dl} (\mathbf{B}^{dl} \mathbf{V}_\psi^{ul} + \mathbf{A}_r^{dl} \mathbf{Q}^{dl}) + \rho_k^{dl} \mathbf{z}_k^{dl}] \bmod \Lambda_c \\ &= \underbrace{[\mathbf{a}_{\psi,k}^{dl} \mathbf{V}_\psi^{ul}]}_{\text{intended signal}} + \mathbf{z}_{\text{eff},k}^{dl} \bmod \Lambda_c \end{aligned} \quad (23)$$

where $\mathbf{z}_{\text{eff},k}^{dl}$ is the effective noise given by $(\rho_k^{dl} \mathbf{h}_k^{dl} \mathbf{B}^{dl} - \mathbf{a}_{\psi,k}^{dl}) \mathbf{V}_\psi^{ul} + \rho_k^{dl} (\mathbf{h}_k^{dl} \mathbf{A}_r^{dl} \mathbf{Q}^{dl} + \mathbf{z}_k^{dl})$, $\mathbf{a}_{\psi,k}^{dl}$ is the k th row of \mathbf{A}_ψ^{dl} , a matrix with dimensions of $K \times M$ and rank of M , $\mathbf{A}_\psi^{dl} = \mathbf{W}^\top \mathbf{A}_\psi^{ul}$, \mathbf{A}_ψ^{ul} is the inverse of \mathbf{A}_ψ^{ul} matrix, \mathbf{h}_k^{dl} is the k th row of \mathbf{H}^{dl} , $\mathbf{v}_{\psi,k}^{dl} = [\mathbf{a}_{\psi,k}^{dl} \mathbf{V}_\psi^{ul}] \bmod \Lambda_c$ is the k th user's intended signal that includes the sum of the codewords of the user-pair $k, k' \in \{1, \dots, K\}$, $k \neq k'$. The power of the effective noise $\mathbf{z}_{\text{eff},k}^{dl}$ is given by

$$\begin{aligned} \sigma_k^{dl2} &= \frac{1}{n} \mathbb{E}[\|\mathbf{z}_{\text{eff},k}^{dl}\|^2] \\ &= \|(\rho_k^{dl} \mathbf{h}_k^{dl} \mathbf{B}^{dl} - \mathbf{a}_{\psi,k}^{dl}) (\mathbf{P}_\psi^{ul})^{\frac{1}{2}}\|^2 \\ &\quad + \rho_k^{dl2} (\mathbf{h}_k^{dl} \mathbf{A}_r^{dl} \mathbf{D}^{dl} \mathbf{A}_r^{dl\top} \mathbf{h}_k^{dl\top} + 1) \end{aligned} \quad (24)$$

This effective variance can be minimized by obtaining the MMSE coefficient for the linear scaling equalizer ρ_k^{dl} as

$$\rho_k^{dl} = \frac{\mathbf{a}_{\psi,k}^{dl} \mathbf{P}_\psi^{ul} \mathbf{B}^{dl\top} \mathbf{h}_k^{dl\top}}{\mathbf{h}_k^{dl} (\mathbf{A}_r^{dl} \mathbf{D}^{dl} \mathbf{A}_r^{dl\top} + \mathbf{B}^{dl} \mathbf{P}_\psi^{ul} \mathbf{B}^{dl\top}) \mathbf{h}_k^{dl\top} + 1} \quad (25)$$

Finally, user k decodes $[\mathbf{a}_{\psi,k}^{dl} \mathbf{V}_\psi^{ul}] \bmod \Lambda_c$, and uses its own codeword $[\mathbf{s}_k^{ul}] \bmod \Lambda_c$ as side information to recover its desired codeword $[\mathbf{s}_{k'}^{ul}] \bmod \Lambda_c$ as follows

$$\begin{aligned} \mathbf{s}_{\text{desired},k} &= [[\mathbf{a}_{\psi,k}^{dl} \mathbf{V}_\psi^{ul}] \bmod \Lambda_c - \mathbf{s}_k^{ul}] \bmod \Lambda_c \\ &= [\mathbf{a}_{\psi,k}^{dl} \mathbf{A}_\psi^{ul} \mathbf{W} \mathbf{S}^{ul} - \mathbf{s}_k^{ul}] \bmod \Lambda_c \\ &= \left[\sum_{m=1}^M w_{m,k} \left(\sum_{u=1}^K w_{m,u} \mathbf{s}_u^{ul} \right) - \mathbf{s}_k^{ul} \right] \bmod \Lambda_c \\ &= [\mathbf{s}_k^{ul} + \mathbf{s}_{k'}^{ul} - \mathbf{s}_k^{ul}] \bmod \Lambda_c \\ &= [\mathbf{s}_{k'}^{ul}] \bmod \Lambda_c \end{aligned} \quad (26)$$

Using this procedure, user k downlink rate is given by

$$R_{\psi,k}^{dl} = \frac{1}{2} \log^+ (p^{ul} (\sigma_k^{dl2})^{-1}) \quad (27)$$

At this point, we can summarize the end-to-end achievable rate of the proposed scheme as given next.

Theorem 1: The end-to-end data rate of user k achieved by the proposed scheme is given by

$$R_k = \min\{R_{\psi,k}^{ul}, R_{\psi,k}^{dl}\}, \quad (28)$$

where $R_{\psi,k}^{ul}$ and $R_{\psi,k}^{dl}$ are given in (14) and (27), respectively. *Proof:* This statement follows since the achievable end-to-end rate is bound by the smallest between the uplink rate and the downlink rate.

V. END-TO-END USER-RATE OPTIMIZATION

In this section, we propose an iterative multi-pair two-way rate optimization (MPTWR) algorithm to optimize the end-to-end rate in (28). The algorithm is carried in two steps, where the uplink and downlink user-rates are optimized iteratively.

A. Uplink Rate Optimization

Given \mathbf{H}^{ul} , \mathbf{P}^{ul} and P_k^{ul} , $k \in \{1, \dots, K\}$, the achievable uplink rate $R_{\psi,k}^{ul}$ can be optimized by assigning appropriate scaling matrix \mathbf{B}^{ul} , selecting proper full rank integer coefficient matrices \mathbf{A}_r^{ul} and \mathbf{A}_ψ^{ul} , and selecting the uplink compression distortion covariance matrix \mathbf{D}^{ul} to satisfy the fronthaul capacity constraint. The uplink optimization problem can be formulated as follows

$$\begin{aligned} \max_{\mathbf{A}_r^{ul}, \mathbf{D}^{ul}, \mathbf{A}_\psi^{ul}} \quad & \sum_{k=1}^K R_{\psi,k}^{ul} \\ \text{subject to} \quad & (b_k^{ul})^2 p^{ul} \leq P_k^{ul} \quad \forall k \in \{1, 2, \dots, K\} \\ & \text{rank}(\mathbf{A}_r^{ul}) = L, \quad \text{rank}(\mathbf{A}_\psi^{ul}) = M \\ & R_{r,\ell}^{ul} \leq C_\ell \quad \forall \ell \in \{1, 2, \dots, L\} \end{aligned} \quad (29)$$

This optimization problem is a mixed-integer non-linear programming (MINLP) which is NP-hard problem. Therefore, we can solve (29) by decoupling it into two separate steps

Algorithm 1 Iterative uplink optimization (IUO)

1: **Initialization:** Set $d_{\min} = 0$ and $d_{\max} = d^{ul} = \delta$ (large) such that $R_{r,\ell}^{ul} < C_\ell \forall \ell$.
2: **while** $\max_\ell (C_\ell - R_{r,\ell}^{ul}) > \epsilon$ **or** $\max_\ell (R_{r,\ell}^{ul} - C_\ell) > 0$ **do**
3: **if** $\max_\ell (R_{r,\ell}^{ul} - C_\ell) > 0$ **then**
4: $d_{\min} = d^{ul}$
5: **else**
6: $d_{\max} = d^{ul}$
7: **end if**
8: $d^{ul} = (d_{\max} + d_{\min})/2$
9: $\mathbf{F}_r^{ul} = \text{Chol}(\frac{1}{d^{ul}} \mathbf{H}^{ul} \mathbf{P}^{ul} \mathbf{H}^{ul\top} + \mathbf{I}_L(\frac{1}{d^{ul}} + 1))$
10: $\mathbf{A}_r^{ul} = \text{LLL}(\mathbf{F}_r^{ul})$
11: $R_{r,\ell}^{ul} = \frac{1}{2} \log^+(\|\mathbf{F}_r^{ul} \mathbf{a}_{r,\ell}^{ul}\|^2)$
12: **end while**
13: $\mathbf{F}_\psi^{ul} = \text{Chol}((\mathbf{P}^{ul^{-1}} + \mathbf{H}^{ul\top} (\mathbf{I}_L + \mathbf{D}^{ul})^{-1} \mathbf{H}^{ul})^{-1})$
14: $\mathbf{A}_\psi^{ul} = \text{LLL}(\mathbf{F}_\psi^{ul} \mathbf{W}^\top)$
15: Calculate $\sigma_j^{ul^2}$ using (13) $\forall m \in \{1, \dots, M\}$.
16: Calculate $\mathbf{r}_\psi^{ul} = [R_{\psi,1}^{ul}, \dots, R_{\psi,K}^{ul}]^\top$ using (14).
17: **return** $(\mathbf{D}^{ul}, \mathbf{A}_r^{ul}, \mathbf{A}_\psi^{ul}, \mathbf{r}_\psi^{ul})$

and iterating over them. At first, for a fixed scaling matrix \mathbf{B}^{ul} , we start by selecting the proper integer coefficient matrices \mathbf{A}_r^{ul} and \mathbf{A}_ψ^{ul} . This process is related to the Shortest Independent Vector Problem (SIVP) which is NP-hard. However, sub-optimal solutions can be obtained using the LLL algorithm [15]. For the sake of simplicity, it is assumed that all RRHs choose equal distortion levels, i.e., $d_\ell^{ul} = d^{ul}$, $\ell \in \{1, 2, \dots, L\}$. To maximize the uplink rates, we start by initializing d^{ul} to two extreme values and then calculate the corresponding \mathbf{A}_r^{ul} using the LLL algorithm on \mathbf{F}_r^{ul} which is the Cholesky decomposition satisfying $\mathbf{F}_r^{ul} \mathbf{F}_r^{ul\top} = \frac{1}{d^{ul}} \mathbf{H}^{ul} \mathbf{B}^{ul} \mathbf{P}^{ul} \mathbf{B}^{ul\top} \mathbf{H}^{ul\top} + \mathbf{I}_L(\frac{1}{d^{ul}} + 1)$. Next, we update d^{ul} using bisection until (29) is satisfied with equality. Finally, we use the obtained d^{ul} to calculate \mathbf{A}_ψ^{ul} using the LLL algorithm on \mathbf{F}_ψ^{ul} defined after (13). This is explained in detail in Algorithm 1.

Second, under the assignment of \mathbf{D}^{ul} , \mathbf{A}_r^{ul} , and \mathbf{A}_ψ^{ul} matrices, the algorithm solves to find the appropriate precoding matrix \mathbf{B}^{ul} using the barrier method which is a standard method to solve optimization problems with inequality constraints [16] as shown in Algorithm 2. Finally, the algorithm iterate over the two sub-steps until convergence occurs. The results of this algorithm will be used as inputs to optimize the achievable downlink user-rate.

B. Downlink Rate Optimization

The aim of this step is to obtain an achievable downlink user-rate greater or equal to the achievable one in the uplink, i.e., $R_{\psi,k}^{dl} \geq R_{\psi,k}^{ul}$. Given the downlink channel matrix $\mathbf{H}^{dl} = \mathbf{H}^{ul\top}$, the downlink powers P_ℓ^{dl} , $\ell = 1, \dots, L$, and the outputs of the uplink optimization algorithm, the achievable downlink user-rate $R_{\psi,k}^{dl}$ should be tuned by selecting proper integer coefficient matrix \mathbf{A}_r^{dl} , beamforming matrix \mathbf{B}^{dl} , and

Algorithm 2 Updating the Precoding Matrix \mathbf{B}^{ul} Algorithm

1: Set $\theta = \theta^{(0)} > 0$, and solve the barrier problem using newton method to get $\mathbf{B}^{ul^{(0)}} = \mathbf{B}^{ul*}(\theta)$.
2: For barrier parameter $\eta > 1$ and $\gamma = 1, 2, 3, \dots$, repeat
3: Solve barrier problem at $\theta = \theta^{(\gamma)}$ using newton method initialized at $\theta^{(\gamma-1)}$ to produce $\mathbf{B}^{ul^{(\gamma)}} = \mathbf{B}^{ul*}(\theta)$.
4: Stopping criteria at $K + L/\theta < \epsilon$
5: Else, increase $\theta := \eta\theta$
6: **return** (\mathbf{B}^{ul})

Algorithm 3 Iterative downlink optimization (IDO)

1: **Initialization:** set $\mathbf{H}^{dl} = \mathbf{H}^{ul\top}$, $\mathbf{A}_r^{dl} = \mathbf{A}_r^{ul\top}$, $\mathbf{C}^{dl} = \mathbf{C}^{ul\top}$, and $\mathbf{r}_\psi^{dl} = [R_{\psi,1}^{dl}, \dots, R_{\psi,K}^{dl}]^\top = \mathbf{0}_{K \times 1}$.
2: **while** $\sum_{k=1}^K |R_{\psi,k}^{dl} - R_{\psi,k}^{ul}| > \epsilon$ **or** $\max_\ell (p_\ell^{dl} - P_\ell^{dl}) > 0$ **do**
3: Set $\delta = 0$, initialize $\mathbf{B}_0^{dl} = L \times M$ matrix of i.i.d. $\mathcal{N}(0, 1)$, and $\mathbf{\Gamma}_0 = \mathbf{I}$.
4: **while** termination condition for BFGS method **do**
5: Compute line search $\Theta_\delta = -\mathbf{\Gamma}_\delta \nabla f_{un}(\mathbf{B}_\delta^{dl})$, and step length $\gamma_\delta > 0$.
6: Calculate $\mathbf{B}_{\delta+1}^{dl} = \mathbf{B}_\delta^{dl} + \gamma_\delta \Theta_\delta$ and $\mathbf{d}^{dl} = \mathbf{C}^{dl} \boldsymbol{\xi}^{dl}$
7: Calculate p_ℓ^{dl} , ρ_k^{dl} , and $\mathbf{r}_\psi^{dl} = [R_{\psi,1}^{dl}, \dots, R_{\psi,K}^{dl}]^\top$ using (21), (25), and (27), respectively
8: Calculate $\beta_\delta = \mathbf{B}_{\delta+1}^{dl} - \mathbf{B}_\delta^{dl}$.
9: Calculate $\mathbf{\Omega}_\delta = \nabla f_{un}(\mathbf{B}_{\delta+1}^{dl}) - \nabla f_{un}(\mathbf{B}_\delta^{dl})$.
10: Update $\mathbf{\Gamma}_\delta$ using β_δ and $\mathbf{\Omega}_\delta$ as in [17].
11: Set $\delta = \delta + 1$.
12: **end while**
13: **end while**
14: **return** $(\mathbf{r}_\psi^{ul}, \mathbf{r}_\psi^{dl})$

downlink compression distortion covariance matrix \mathbf{D}^{dl} . One way to formulate this problem is as follows

$$\begin{aligned} \min_{\mathbf{A}_r^{dl}, \mathbf{B}^{dl}, \mathbf{D}^{dl}} \sum_{k=1}^K (R_{\psi,k}^{dl} - R_{\psi,k}^{ul})^2 \\ \text{subject to } p_\ell^{dl} \leq P_\ell^{dl}, \text{ and } R_{r,\ell}^{dl} \leq C_\ell \forall \ell \end{aligned} \quad (30)$$

To simplify this problem we choose $\mathbf{A}_r^{dl} = \mathbf{A}_r^{ul\top}$. Then, we obtain \mathbf{B}^{dl} that minimize (30) using the BFGS Quasi-Newton algorithm with a cubic line search procedure [17] and calculate the downlink distortion levels $\mathbf{d}^{dl} = [d_1^{dl}, \dots, d_L^{dl}]^\top$ using (20). The line search in the BFGS algorithm must satisfy the Wolfe conditions in order to ensure sufficient step length taken in each search direction. Finally, the optimal \mathbf{B}^{dl} matrix is obtained when the partial derivatives of \mathbf{B}^{dl} are sufficiently too small [17]. If the constraints in (30) are not satisfied, we update the initial value of \mathbf{B}^{dl} , and repeat until the constraints are satisfied. The details of this procedure are given in Algorithm 2.

VI. NUMERICAL RESULTS

In this section, the proposed MPTWR optimization scheme is evaluated and compared to the conventional integer-forcing

VII. CONCLUSION

We proposed a multi-pair two-way user-rate optimization scheme for intra C-RAN communications, where users inside the network are grouped into communicating pairs. We used a multi-pair lattice-based computation strategy, where the BBU pool decodes integer linear combinations of paired users' codewords instead of decoding linear combinations of individual codewords. This reduces the required number of computation steps at the BBU pool, thereby reducing the number of rate constraints. In addition, instead of recovering the original messages as common in the BBU pool, the previously computed equations are compressed directly and forwarded to the RRHs through the fronthaul links. The scheme achieves significant improvement in the end-to-end rate compared to existing schemes.

REFERENCES

- [1] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: insights and challenges," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 152–160, 2015.
- [2] *Cloud Radio Access Networks: Principles, Technologies, and Applications*. Cambridge University Press, 2017.
- [3] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, 2011.
- [4] S. 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, 2013.
- [5] I. El Bakoury and B. Nazer, "Integer-forcing architectures for uplink cloud radio access networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2336–2351, 2020.
- [6] Y. Zhou and W. Yu, "Fronthaul compression and transmit beamforming optimization for multi-antenna uplink c-ran," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4138–4151, 2016.
- [7] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.
- [8] R. Zakhour and D. Gesbert, "Optimized data sharing in mimo with finite backhaul capacity," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6102–6111, 2011.
- [9] S. 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, no. 22, pp. 5646–5658, 2013.
- [10] S. Hong and G. Caire, "Reverse compute and forward: A low-complexity architecture for downlink distributed antenna systems," in *2012 Proc. IEEE Int. Symp. Info. Theory*, 2012, pp. 1147–1151.
- [11] —, "Compute-and-forward strategies for cooperative distributed antenna systems," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5227–5243, 2013.
- [12] L. Liu, P. Patil, and W. Yu, "An uplink-downlink duality for cloud radio access network," in *2016 IEEE Int. Symp. Info. Theory (ISIT)*, 2016, pp. 1606–1610.
- [13] I. E. Bakoury and B. Nazer, "Uplink-downlink duality for integer-forcing in cloud radio access networks," in *2018 56th Annual Allerton Conf. Commun., Control, and Comput. (Allerton)*, 2018, pp. 39–47.
- [14] O. Ordentlich and U. Erez, "Integer-forcing source coding," *IEEE Trans. Inf. Theory*, vol. 63, pp. 1253–1269, 2017.
- [15] A. K. Lenstra, H. W. Lenstra, and L. Lovasz, "Factoring polynomials with rational coefficients," *MATH. ANN.*, vol. 261, pp. 515–534, 1982.
- [16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [17] D. F. Shanno, "Conditioning of quasi-newton methods for function minimization," *Math. Comput.*, vol. 24, no. 111, pp. 647–656, 1970.
- [18] J. Guo, A. A. I. Ibrahim, A. Agaskar, D. J. Love, and N. Yazdani, "Implementation of rate-adaptive integer forcing compression in distributed wireless relay networking," in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, 2018, pp. 1–6.

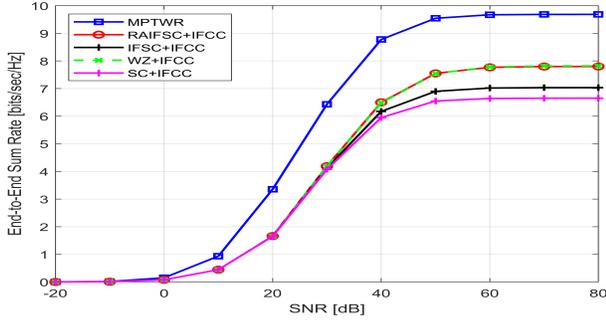


Fig. 1. End-to-End sum rate versus SNR in dB at $C_\ell = 4$ bits/transmission.

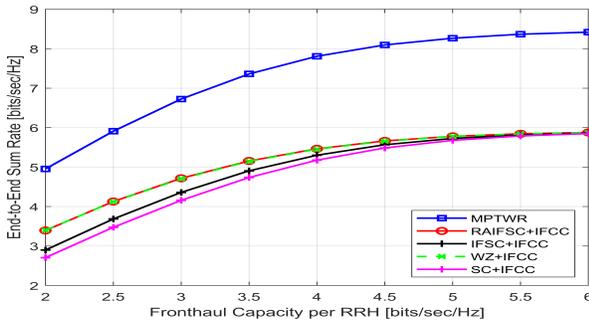


Fig. 2. End-to-End sum rate versus C_ℓ in bits/transmission at SNR = 35 dB.

source and channel coding (IFSC+IFCC) scheme [13] and the rate adaptive integer-forcing source and channel coding scheme (RAIFSC+IFCC) [18] under the fronthaul capacity constraint per each RRH. We also compare with optimized Wyner-Ziv (WZ+IFCC) and single-user (SU+IFCC) compression schemes with integer-forcing channel coding. We use 5000 realizations of the $L \times K$ channel matrix $\mathbf{H} = \mathbf{H}^{ul} = \mathbf{H}^{dlT}$, where each element $h_{\ell,k}$ is i.i.d $\mathcal{N}(0, 1)$. We set $L = 2$ RRHs, $K = 4$ users ($M = 2$ user-pairs). It is assumed that the coding power of the coarse lattice Λ_c is equal to $p^{ul} = \frac{10^{\text{SNR}/10}}{K}$, where SNR is the signal-to-noise ratio in dB. Also, the capacities of the fronthaul links are assumed to be equal, i.e., $C_1 = C_2$.

Fig. 1 and 2 show the total achievable end-to-end rate of our proposed scheme and conventional ones in bits/sec/Hz versus SNR at $C_\ell = 4$ bits/transmission and different fronthaul link capacity values at SNR of 35 dB, respectively. These figures demonstrate that our proposed scheme has a superior performance over other conventional approaches. This is due to the exploitation of the multi-pair lattice-based computation strategy that reduces the number of decoded linear combinations at the BBU pool to 2 equations instead of the 4 equations required by other IF schemes. In addition, the performance of the RAIFSC+IFCC scheme is nearly the same as that in the optimized WZ scheme, while the performance of IFSC+IFCC scheme is close to them. Further, the optimized SU has the poorest performance as usual because the correlation between the received signals at all RRHs is not exploited.