

Article

Energy Efficiency Optimization for Downlink Cloud RAN with Limited Fronthaul Capacity

Yong Wang, Lin Ma and Yubin Xu *

School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150080, China; dr_ywang@hotmail.com (Y.W.); malin@hit.edu.cn (L.M.)

* Correspondence: ybxu@hit.edu.cn; Tel.: +86-451-8641-2052 (ext. 8202)

Received: 22 March 2017; Accepted: 24 June 2017; Published: 26 June 2017

Abstract: In the downlink cloud radio access network (C-RAN), fronthaul compression has been developed to combat the performance bottleneck caused by the capacity-limited fronthaul links. Nevertheless, the state-of-arts focusing on fronthaul compression for spectral efficiency improvement become questionable for energy efficiency (EE) maximization, especially for meeting its requirements of large-scale implementation. Therefore, this paper aims to develop a low-complexity algorithm with closed-form solution for the EE maximization problem in a downlink C-RAN with limited fronthaul capacity. To solve such a non-trivial problem, we first derive an optimal solution using branch-and-bound approach to provide a performance benchmark. Then, by transforming the original problem into a parametric subtractive form, we propose a low-complexity two-layer decentralized (TLD) algorithm. Specifically, a bisection search is involved in the outer layer, while in the inner layer we propose an alternating direction method of multipliers algorithm to find a closed-form solution in a parallel manner with convergence guaranteed. Simulations results demonstrate that the TLD algorithm can achieve near optimal solution, and its EE is much higher than the spectral efficiency maximization one. Furthermore, the optimal and TLD algorithms are also extended to counter the channel error. The results show that the robust algorithms can provide robust performance in the case of lacking perfect channel state information.

Keywords: C-RAN; fronthaul compression; energy efficiency; alternating direction method of multipliers (ADMM); imperfect channel state information (CSI)

1. Introduction

To maintain the requirements of the expected scale of the increasing data traffic and mobile terminals, the fifth generation (5G) wireless network [1] faces some challenges in terms of system capacity, energy consumption, and so on. The cloud radio access network (C-RAN) [2,3], which has emerged as a promising solution in reducing both the capital and operating expenditures, is expected to be an effective approach to fulfil these requirements. In C-RAN, a central unit (CU) or baseband unit (BBU) pool connects all the deployed low-power base stations (BSs) using the finite-capacity fronthaul links that allows joint signal processing and transmission. Despite various attractive advantages brought by C-RANs [4–6], such as joint beamforming, and centralized encoding and decoding, the performance bottleneck for large-scale implementation comes with the high capacity requirements of the fronthaul links. Therefore, in practical C-RAN, the data sharing [7–13] and fronthaul compression designs [14–22] are recognized as two promising approaches to overcome the significant impact of the constrained fronthaul on spectral efficiency (SE) and energy efficiency (EE) (bit-per-Joule) [23–25]. The data sharing strategy [8–13] reduces the fronthaul consumption through limiting the data transfer among BSs (one BS serves a small number of the total MUs). For the latter one, the CU computes the precoded signals intended to be transmitted to each BS, and then the signals are quantized and sent

to the BSs through the capacity limited fronthaul links. The compression process is usually modeled as a test channel [14–22], where the uncompressed and compressed signals respectively represent the input and output, and the quantization noise are modeled as an independent additive Gaussian random variable [13,20,21]. The authors in [13] proved that the fronthaul compression achieves a better performance than the data sharing strategy, so that this paper considers the fronthaul compression.

In general, the fronthaul compression has been extensively studied in both the uplink [16–18] and downlink C-RANs [13,19–22]. The majority research of the existing literature on fronthaul compression in C-RAN is focused on the design of beamforming and quantization noises to maximize the achievable sum rate from the information-theoretic perspective [11,13,16–21,26]. A joint adaptive decompression and detection algorithm was proposed in [16] to improve the information-theoretic capacity for the uplink C-RAN. In [18], the authors developed a distributed compression for the sum rate maximization (SEmax) problem in the uplink C-RAN accounting for both perfect and imperfect channel state information (CSI). For the downlink C-RAN, a joint precoding and multivariate compression scheme has been studied in [19], where an iterative majorization-minimization approach was proved to achieve a stationary point solution of the SEmax problem. In [13], a comparison between the data sharing and fronthaul compression strategy was investigated for the power minimization problem with finite-capacity fronthaul links in downlink C-RAN. Then, a hybrid compression and data sharing strategy was designed in [11] for optimizing the achievable sum rate. However, maximizing the EE in accordance with fronthaul compression [22] is technically far more challenging compared to SRmax, and the closed-form solution remains unexplored.

This paper investigates the EEmax problem in downlink C-RANs with the consideration of the fronthaul compression. This optimization problem is formulated as a non-convex fractional programming problem with respect to the beamforming and quantization noises under a limited power budget and capacity-finite fronthaul links, which is NP-hard and is difficult to solve. To deal with the fraction objective function of the original problem, some work has been done to firstly transform the fractional objective function into an equivalent subtractive-form optimization problem via exploiting the fractional programming [27–30]. Although the power allocation schemes for orthogonal frequency division multiple access (OFDMA) system [28,29] were obtained using the dual decomposition approach, such approaches cannot be applied to solve our problem due to the joint consideration of fronthaul compression and beamforming in our problem. Moreover, the Lagrangian based decomposition algorithm [30] for the multicell system is still not applicable for decentralized implementation because the considered problem is more complex and the beamformings among BSs are coupled. The first-order Taylor expansion is adopted to linearize the non-convex data rate and fronthaul capacity constraint [20], however, the optimal solution cannot be obtained. By exploiting the relationship between the achievable data rate and the mean square error (MSE) in [31], Ref. [11,12] compared the sum rate performance of the data sharing and compression strategies, but they are limited to the SRmax problem. The authors in [22] considered a joint design of beamforming, multivariate compression and BS-MU link selection to maximize the EE. By using the epigraph form of the original EEmax problem, the authors [22] proposed a difference of convex (DC) function based algorithm. However, the optimal solution of [22] is still unknown, and the computational complexity is typically high since a series of semi-definite programming (SDP) or second-order cone programming (SOCP) problems are solved. More importantly, since the beamforming and quantization noises are computed centrally at the CU, the algorithms proposed in the literature [11–13,19,20,22] can be computation intensive for large-scale C-RANs. Therefore, in this paper, we first propose an optimal algorithm for the EEmax problem to provide a performance benchmark, and then develop a low-complexity decentralized algorithm with closed-form solution. For practical usage in downlink C-RANs, the impact of CSI errors on the EE performance is further investigated. This paper not only offers an optimal solution for the EEmax problem with limited fronthaul capacity, but also lays a sound foundation for decentralized implementation with closed-form solutions in this field. Thus, the methods derived in this paper are significant in advancing this field.

The main contributions of this paper are summarized as follows.

Firstly, we formulate the EEmax problem of joint beamforming and quantization noises design under a limited power budget and capacity-limited fronthaul links as a non-convex fractional programming problem. We first derive an optimal solution method based on branch-and-bound (BnB) technique [32,33] to solve the EEmax problem globally. Specifically, the BnB algorithm computes the upper and lower bounds, and deletes the regions that do not contain the optimal solution. The algorithm terminates when the difference between the upper and lower bounds is smaller than a predefined accuracy.

Secondly, to reduce the computational complexity of the optimal algorithm and facilitate decentralized implementation, we propose to transform the problem into a parametric subtractive form, and further proposed a two-layer decentralized (TLD) algorithm to solve the equivalent subtractive problem. Specifically, an one-dimension search approach is used to find the EE in the outer layer, and a decentralized algorithm based on alternating direction method of multipliers (ADMM) is proposed to solve a subproblem in the inner layer. The proposed algorithm achieves closed-form solution in parallel manner with convergence guaranteed.

Thirdly, considering the imperfection of the obtained CSI in practical C-RANs, the robust optimal and TLD algorithms for the considered EEmax problem are also proposed to characterize the performance degradation of the CSI errors. In particular, the robust optimal can also achieve a performance benchmark, and the robust TLD algorithm also has closed-form solution in a parallel manner.

Finally, we validate the effectiveness of the proposed algorithms through extensive simulations. The results demonstrate that both the optimal and TLD algorithms are convergent, and the TLD algorithm can achieve near optimal solution which is much higher than the SEmax one. Numerical analysis also show that the EE performance is susceptible to the channel errors, and a smaller channel error reaches a higher EE.

The remainder of the paper is organized as follows. In Section 2, the system model and problem formulation are presented. Section 3 describes the proposed optimal algorithm with perfect CSI. Section 4 presents TLD algorithm with perfect CSI. For imperfect CSI case, the robust optimal and TLD algorithms are also presented in Section 5. The simulation results are given in Section 6. Finally, we conclude this paper in Section 7.

Notations: We use \mathbb{C} to denote the set of complex numbers, and $\mathbb{C}^{M \times N}$ to denote the set of all $M \times N$ matrices with complex entries. We use boldface capital and lower case letters are respectively used to denote matrices and vectors. $(\mathbf{X})^{-1}$, \mathbf{X}^H and $\text{Tr}(\mathbf{X})$ represent the matrix inverse, Hermitian transport and the trace, respectively. $|\mathbf{x}|$ represents the Euclidean norm. $\mathbb{E}[\cdot]$ is the expectation operator, and $\text{diag}(x_1, \dots, x_L)$ represents a diagonal matrix with diagonal elements given by $\{x_1, \dots, x_L\}$. For a complex number x , $|x|$ is the mode of x . "s.t." stands for "subject to".

2. System Model and Problem Formulation

2.1. System Model

We consider a downlink C-RAN with L single-antenna BSs and K single-antenna MUs. The CU connects all the BSs via fronthaul links, and each link is finitely constrained by $C_l, l = 1, \dots, L$. Assume that the CU has access the global CSI. The data symbol for each MU (denoted by s_k for the k -th MU) is distributed as complex Gaussian with zero mean and unit variance. Denote by $x_l = \sum_{k=1}^K w_{kl}s_k$ the beamformed complex signal at the CU for BS l , where w_{kl} is the beamforming from BS l to MU k . To reduce the capacity requirements on the fronthaul network, the signals are compressed before being forwarded to the corresponding BSs via the finite-capacity fronthaul links. According to [18,20], the compression procedure is modeled as a test channel and the procedure can be expressed as

$\tilde{x}_l = x_l + e_l, \forall l$, where e_l is the quantization noise, and x_l and \tilde{x}_l are the input and out of the test channel, respectively. The received signal at MU k is given by

$$y_k = \sum_{l=1}^L h_{kl}^H \tilde{x}_l + n_k, k = 1, \dots, K, \quad (1)$$

where $h_{kl} \in \mathbb{C}$ is the channel from BS l to MU k , n_k is the additive Gaussian noise at MU k , with zero mean and σ^2 variance.

When employing the single user detection at each MU, the received signal-to-interference-plus-noise-ratio (SINR) at MU k is

$$\text{SINR}_k = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{w}_i|^2 + \sigma^2 + |\mathbf{h}_k \mathbf{Q} \mathbf{h}_k^H|}, \quad (2)$$

where $\mathbf{h}_k = [h_{k1}^T, \dots, h_{kL}^T]^T$ and $\mathbf{w}_k = [w_{k1}, \dots, w_{kL}]^T$ are the aggregated channel and beamforming from all BSs to MU k , respectively, and

$$\mathbf{Q} = \begin{bmatrix} q_{11} & \cdots & q_{1L} \\ \vdots & \ddots & \vdots \\ q_{L1} & \cdots & q_{LL} \end{bmatrix}, \quad (3)$$

is the covariance matrix of $\mathbf{e} = [e_1, \dots, e_L]^T$, i.e., $\mathbf{Q} = \mathbf{e} \mathbf{e}^H$. It is noted that multivariate compression is also possible and has been studied in [20], where $e_l e_j^H \neq 0, \forall l = 1, \dots, L, e_l e_j^H = 0$ when $l \neq j$. In this paper, we consider point to point fronthaul compression, and let $q_l = q_{ll}, l = 1, \dots, L$.

Considering an ideal vector quantizer, the quantization noise level q_l and the fronthaul capacity C_l for the l -th fronthaul link satisfy the following constraint [12,15]

$$\log_2 \left(1 + \frac{\sum_{k=1}^K |w_{kl}|^2}{q_l} \right) \leq C_l, \forall l. \quad (4)$$

The transmission power consumed at BS l is constrained by $\mathbb{E}[|\tilde{x}_l|^2] \leq P_l^{\max}$, where P_l^{\max} is the maximum transmit power of BS l . The transmit power of BS l , denoted by p_l , consists of quantization noise q_l and data transmission power (denoted by p_l^t), i.e., $p_l = p_l^t + q_l = \sum_{k=1}^K |w_{kl}|^2 + q_l$. It is obviously that

$$p_l \leq P_l^{\max}, \forall l. \quad (5)$$

The network power of C-RAN consists of the BS transmit power and relative fronthaul network power. In this paper, we adopt the power consumption model of C-RAN as [4]

$$P_{tot} = \sum_{l=1}^L \frac{1}{\eta_l} p_l + P^c, \quad (6)$$

where $P^c = \sum_{l=1}^L P_l^c$ is the total relative fronthaul link power consumption [4], $P_l^c \geq 0$ is the relative fronthaul link power consumption when switch off both the fronthaul link and the corresponding BSs. η_l ($\eta_l > 1$) is the drain efficiency of power amplifier of BS l . In this paper, we assume that all the BSs have the same drain efficiencies, i.e., $\eta = \eta_l, \forall l$. Since we do not consider the BS switch on/off scheme in this paper, P^c is a nonnegative constant and we call it by static power for brevity for the rest of the paper. We point out that based on the results obtained by the proposed algorithms in this paper, it is easily extended to add BSs selection (determine the BSs to be switch off or not) into consideration through ordering the BSs in accordance with the bisection search [4] to further improve EE. However, this is outside the scope of this paper.

2.2. Problem Formulation

To balance the sum rate and total power consumption, the EEmax problem is optimized in this paper. This problem over the beamforming in the presence of fronthaul compression can be formulated as

$$\mathbf{P}_0 : \max_{\mathbf{w}, \mathbf{q}} \frac{\sum_{k=1}^K R_k}{\sum_{l=1}^L \frac{1}{\eta_l} p_l + P_c} \quad (7a)$$

$$\text{s.t.} \quad \sum_{k=1}^K |w_{kl}|^2 + q_l \leq P_l^{\max}, \forall l \quad (7b)$$

$$\beta_l \sum_{k=1}^K |w_{kl}|^2 \leq q_l, \forall l, \quad (7c)$$

where $\mathbf{w} = [\mathbf{w}_1, \dots, \mathbf{w}_K]$ and $\mathbf{q} = [q_1, \dots, q_K]$ are the collection of beamforming vectors and quantization noises, respectively. $R_k = \log_2(1 + \text{SINR}_k)$ is the achievable data rate of MU k . (7b) is the BS transmit power constraint, and (7c) is the reformulation of the fronthaul capacity constraint (4) and $\beta_l = \frac{1}{2^{c_l-1}}$. Due to the non-convex R_k and the fractional objective function in (7a), \mathbf{P}_0 is a NP-hard problem, and it is challenging to find its global optimum. In the following, we first present an optimal approach and then propose a TLD framework solution.

3. Optimal Algorithm with Perfect CSI

In this section, we will propose a global optimal algorithm, which based on Branch-and-Bound method [32], to solve problem \mathbf{P}_0 . The essential idea of the proposed algorithm is based on the following equivalent transformation

$$\mathbf{P}_1 : \max_{w_{kl}, \mathbf{t}} f(\mathbf{t}) = t_{K+1} \sum_{k=1}^K t_k \quad (8a)$$

$$\text{s.t.} \quad \text{SINR}_k \geq 2^{t_k} - 1, \forall k \quad (8b)$$

$$\frac{1}{\frac{1}{\eta} \sum_{l=1}^L (\sum_{k=1}^K |w_{kl}|^2 + q_l) + P_c} \geq t_{K+1} \quad (8c)$$

(7b), (7c),

where $\mathbf{t} = [t_1, \dots, t_{K+1}]^T$ are the introduced variables, (8b) is the transformation of $\log_2(1 + \text{SINR}_k) \geq t_k$. The equivalence between problems \mathbf{P}_0 and \mathbf{P}_1 is that the constraints (8b) and (8c) hold with equality at optimum.

Although \mathbf{P}_1 is more tractable compared to \mathbf{P}_0 , it is still hard to solve due to the coupled variables of \mathbf{w}_k and \mathbf{t} . It is observed that if one increase each t_k in the feasible set of \mathbf{P}_1 , a better objective value can be obtained. This motivates us to use the monotonic optimization in [32], i.e., the optimal BnB algorithm, to solve problem \mathbf{P}_1 .

BnB Algorithm

To solve \mathbf{P}_1 , we first denote the feasible set for variables \mathbf{t} by Ξ , i.e., $\Xi = \{\mathbf{t} | \text{constraints of } \mathbf{P}_1\}$. Denote $\underline{\mathbf{t}} = [\underline{t}_1, \dots, \underline{t}_{K+1}]^T$ and $\bar{\mathbf{t}} = [\bar{t}_1, \dots, \bar{t}_{K+1}]^T$ by the aggregated lower and upper bound of t_k . The interval $\underline{\mathbf{t}} \leq \mathbf{t} \leq \bar{\mathbf{t}}$ indicates that each element of \mathbf{t} is bounded by its lower and upper bounds. The objective function $f(\mathbf{t})$ in \mathbf{P}_1 is monotonically increasing in the interval $\underline{\mathbf{t}} \leq \mathbf{t} \leq \bar{\mathbf{t}}$. In particular, t_k is upper bounded by ignoring the interferences, i.e., $t_k \leq \log(1 + \frac{1}{\sigma_k^2} \sum_{l=1}^L P_l^m |\bar{\mathbf{h}}_k|^2) = \bar{t}_k$, and the lower bound of t_k is $\underline{t}_k = 0 \leq t_k$ for $k = 1, \dots, K$. Similarly, we can constrain t_{K+1} by $\underline{t}_{K+1} \leq t_{K+1} \leq \bar{t}_{K+1}$, where $\bar{t}_{K+1} = 1/P_c$ and $\underline{t}_{K+1} = \frac{1}{\sum_{l=1}^L P_l^m / \eta + P_c}$. It is obvious that the feasible set \mathbf{t} in Ξ must be contained by $\Phi = [\underline{\mathbf{t}}, \bar{\mathbf{t}}]$.

For a given $\mathbf{t} \in \Phi$, problem \mathbf{P}_1 reduces to the feasibility problem given by

$$\begin{aligned} \mathbf{P}_2 : \text{find } & \mathbf{w}_1, \dots, \mathbf{w}_K, q_l \\ \text{s.t. } & (7b), (7c), (8b), (8c). \end{aligned} \tag{9}$$

Obviously, when $\mathbf{t} = [0, 0, \dots, \frac{1}{\sum_{l=1}^L P_l^m / \eta + P_c}]^T$, problem \mathbf{P}_2 is infeasible. That is because the sum rate should not equal to zero with maximum transmit power. In this paper, we will customize the BnB algorithm to solve problem \mathbf{P}_1 globally. The BnB algorithm divides the box Φ into smaller ones, and cuts off boxes that do not contain an optimal solution. The algorithm will converge to the global optimal solution after finite iterations.

Since \mathbf{P}_2 is non-convex due to the SINR constraint, in the following, we first recast them as convex ones. Let $\tilde{\gamma}_k = 2^{t_k} - 1$, (8b) is equivalently rewritten as [7,33]

$$\frac{1}{\tilde{\gamma}_k} \mathbf{h}_k \mathbf{w}_k \geq \left(\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{w}_i|^2 + \sigma^2 + \sum_{l=1}^L q_l |h_{kl}|^2 \right)^{1/2} \tag{10a}$$

$$\text{Im}(\mathbf{h}_k \mathbf{w}_k) = 0, \forall k \in \mathcal{K}. \tag{10b}$$

In the above formulation, we note that (10a) is a second-order cone (SOC) constraint. The constraint (10b) is without loss of generality due to the fact that a phase rotation of the beamformers does not effect the objective of the problem [25,33].

Moreover, (8c) is easily rewritten as

$$\frac{1}{\eta} \sum_{l=1}^L \left(\sum_{k=1}^K |w_{kl}|^2 + q_l \right) \leq \frac{1}{t_{K+1}} - P_c. \tag{11}$$

Then, problem \mathbf{P}_2 becomes a SOCP feasibility problem which can solved efficiently. In the proposed BnB algorithm, the bounding function can be formally expressed as

$$\phi_{\text{ub}}(\Phi) = \begin{cases} f(\mathbf{t}_{\text{max}}), & \mathbf{t}_{\text{min}} \in \Xi \\ 0, & \text{otherwise,} \end{cases} \tag{12}$$

$$\phi_{\text{lb}}(\Phi) = \begin{cases} f(\mathbf{t}_{\text{min}}), & \mathbf{t}_{\text{min}} \in \Xi \\ 0, & \text{otherwise,} \end{cases} \tag{13}$$

where $\phi_{\text{ub}}(\Phi)$ and $\phi_{\text{lb}}(\Phi)$ are the upper and lower bound respectively, Φ is defined as $\Phi \triangleq \{\mathbf{t} | t_{k,\text{min}} \leq t_k \leq t_{k,\text{max}}, \forall k\}$ where $t_{k,\text{min}}$ and $t_{k,\text{max}}$ denote the end points of the k th edge of Φ , $\mathbf{t}_{k,\text{min}} = [t_{1,\text{min}}, \dots, t_{K+1,\text{min}}]^T$ and $\mathbf{t}_{k,\text{max}} = [t_{1,\text{max}}, \dots, t_{K+1,\text{max}}]^T$.

We denote by V_i the collection of all created boxes at iteration i . Then, the work flow of the BnB algorithm to obtain the global optimal solution is presented in Algorithm 1.

Algorithm 1 Proposed Optimal Algorithm.

-
- 0: Initialization: given tolerance $\tau > 0$. Set $i = 1$, $\mathcal{B}_1 = \{\Phi\}$, $U_1 = \phi_{\text{ub}}(\Phi)$, $L_1 = \phi_{\text{lb}}(\Phi)$.
 - 1: Check the feasibility problem \mathbf{P}_2 with given \mathbf{t} . If it is infeasible, exit; Otherwise go to step 2.
 - 2: **repeat**
 - 3: Set $\Phi_i = \Phi$ where Φ satisfies $U_i = \phi_{\text{ub}}(\Phi)$.
 - 4: Branch Φ_i into two smaller boxes Φ_I and Φ_{II} using the bisection subdivision along the longest edge of Φ_i .
 - 5: Let $\mathcal{B}_{i+1} = (\mathcal{B}_i \setminus \{\Phi_i\}) \cup \{\Phi_I, \Phi_{II}\}$.
 - 6: Update $L_{i+1} = \max_{\Phi \in \mathcal{B}_{i+1}} \{\phi_{\text{lb}}(\Phi)\}$.
 - 7: Delete boxes that do not contain an optimal solution, $V_{i+1} = V_i \setminus \{\Phi_k | L_{i+1} > \phi_{\text{ub}}(\Phi_k)\}$ where $\Phi_k \in \mathcal{B}_{i+1}$.
 - 8: Update $U_{i+1} = \max_{\Phi \in V_{i+1}} \{\phi_{\text{ub}}(\Phi)\}$.
 - 9: Set $i = i + 1$.
 - 10: **Until** $U_{i+1} - L_{i+1} \leq \tau$.
-

Remark 1 (H). According to [34], the convergence of Algorithm 1 is guaranteed due to the monotonic property of $f(\mathbf{t})$. The main step of Algorithm 1 is to delete the boxes that do not contain the optimal solution. This step is referred to as pruning, and a smaller box that contains the optimal solution is obtained. Therefore, step 7 confirms the convergence of Algorithm 1. The corresponding optimal EE is $f(\mathbf{t})^* = U_i$, the optimal achieved data rate is \mathbf{t}^* , $\forall k \in \mathcal{K}$, and the network power consumption is $1/t_{K+1}^*$. This algorithm gives an optimal solution to problem \mathbf{P}_1 (equivalently to problem \mathbf{P}_0) when the tolerance τ is small enough. Algorithm 1 provides a performance benchmark for any other suboptimal algorithms. However, the computational complexity of Algorithm 1 is very high in general. Therefore, an improved box reduction approach was proposed in [33,34] to reduce the searching time, but we use the basic BnB approach in this paper for simplicity.

4. Decentralized Algorithm with Perfect CSI

In this section, we first transform the original problem into an equivalent subtractive-form using the Dinkelbach's method. By exploiting the equivalence between the achievable data rate and its MSE, an ADMM algorithm is proposed to solve a QCQP subproblem with closed-form solution in a parallel manner.

4.1. Equivalent Optimization Problem

It is noted that \mathbf{P}_0 is a nonlinear fractional programming problem and can be transformed using the Dinkelbach's method [27]. Defining the optimal EE of problem \mathbf{P}_0 by α^{opt} , we have

$$\alpha^{\text{opt}} = R^*/P_{\text{tot}}^{\text{opt}} = \max_{\{\mathbf{w}, \mathbf{q}\}} R/P_{\text{tot}}, \quad (14)$$

where $P_{\text{tot}}^{\text{opt}}$ is the optimal total power consumption, R^{opt} is the optimum of R , and $R^{\text{opt}} = \sum_{k=1}^K R_k^{\text{opt}}$ is the sum rate.

According to [29], the optimal EE α^{opt} can be achieved if and only if

$$\max_{\mathbf{w}, \mathbf{q}} \sum_{k=1}^K R_k - \alpha^{\text{opt}} P_{\text{tot}} = R^{\text{opt}} - \alpha^{\text{opt}} P_{\text{tot}}^{\text{opt}} = 0, \quad (15)$$

where $(\mathbf{w}, \mathbf{q}) \in \mathcal{D}$ and $\mathcal{D} = \{(\mathbf{w}, \mathbf{q}) | (7b), (7c)\}$ is the feasible region of problem α^{opt} .

Thus, based on the theoretical results in [25,30,33], problem \mathbf{P}_0 is transformed as the following parametric programming problem

$$\mathbf{P}_3 : \quad G(\alpha) = \max_{\mathbf{w}, \mathbf{q}} \sum_{k=1}^K R_k - \alpha P_{tot} \quad (16)$$

s.t. (7b), (7c),

It is noted that if problem \mathbf{P}_3 is optimally solved with $G(\alpha) = 0$, problem \mathbf{P}_0 can be solved optimally. However, if problem \mathbf{P}_3 cannot be optimally solved, we can still solve problem \mathbf{P}_0 through solving a sequence of problem \mathbf{P}_3 . Unfortunately, the optimal solution of problem \mathbf{P}_0 is not guaranteed in this case. We will provide detailed analysis in the next subsection.

The function of $G(\alpha)$ is a monotonically decreasing function over α . Therefore, a bisection method, which is demonstrated in Algorithm 2, should perform well enough to find α [25,30].

Algorithm 2 Outer Layer Algorithm.

- 1: Initialize the minimum and maximum α as α_{\min} and α_{\max} respectively, and a small threshold value ϵ .
 - 2: **repeat**
 - 3: Set $\alpha = (\alpha_{\max} + \alpha_{\min})/2$, solve problem \mathbf{P}_3 with α .
 - 4: If $G(\alpha) \geq 0$, $\alpha_{\min} = \alpha$. Otherwise, $\alpha_{\max} = \alpha$.
 - 5: **Until** $|\alpha_{\max} - \alpha_{\min}| \leq \epsilon$ or the maximum iteration number is reached.
-

It is important to initialize α in reducing the search time of Algorithm 2. Here, we initialize the interval $\alpha_{\min} \leq \alpha \leq \alpha_{\max}$ that α is bounded by its lower and upper bounds. Intuitively, α is lower bounded by $\alpha_{\min} = 0$ when the the sum rate equals to zero. For α_{\max} , it is upper bounded by ignoring the interference and using maximum transmit power in R_k , and ignoring the transmit power and the quantization noises in P_{tot} . Specifically, $R_k \leq \log_2(1 + \frac{1}{\sigma^2} \sum_{l=1}^L P_l^{\max} |\mathbf{h}_k|^2) = R_{k,\max}$, and $P_{t,\min} = P^c$. Therefore, $\alpha_{\max} = R_{k,\max}/P_{t,\min}$, and $\alpha = [\alpha_{\min}, \alpha_{\max}] = [0, \sum_{k=1}^K R_{k,\max}/P^c]$.

4.2. Decentralized Algorithm for Subproblem \mathbf{P}_3

The key step for finding the quantized noises and the beamformings in Algorithm 1 lies in solving the subproblem \mathbf{P}_3 . The main difficulty arises from the non-convex R_k in the objective function (16a). Fortunately, by extending the equivalence between the SRmax problem and MMSE problem [31,35], R_k in problem \mathbf{P}_3 can be reformulated into a tractable form.

$$R_k = \max_{\{\rho_k, u_k\}} \frac{1}{\ln 2} (\ln \rho_k - \rho_k e_k), \quad (17)$$

where $\rho_k \in \mathbb{R}$ is a scalar variable associated with MU k , $e_k \in \mathbb{R}$ is the MSE for MU k , given by

$$\begin{aligned} e_k &= \mathbb{E} \left[|u_k^H y_k - s_k|^2 \right] \\ &= |u_k|^2 \left(\sum_{j=1}^K |\mathbf{h}_k \mathbf{w}_j|^2 + \sigma^2 + \sum_{l=1}^L q_l |h_{kl}|^2 \right) \\ &\quad - 2\text{Re}\{u_k \mathbf{h}_k^H \mathbf{w}_k\} + 1. \end{aligned} \quad (18)$$

The proof of the equality in (17) is based on the first-order optimality condition [31], which is omitted here for brevity. Then, problem \mathbf{P}_3 can be recast as

$$\mathbf{P}_4 : \min_{\mathbf{w}, \mathbf{q}} \min_{\rho, \mathbf{u}} \frac{1}{\ln 2} \sum_{k=1}^K (\rho_k e_k - \ln \rho_k) + \alpha P_{tot} \quad (19)$$

s.t. (7b), (7c),

where $\mathbf{u} = [u_1, \dots, u_K]$ and $\rho = [\rho_1, \dots, \rho_L]$.

It is worth noting that problem \mathbf{P}_4 is not jointly convex in $\{\mathbf{w}, \mathbf{u}, \mathbf{q}, \rho\}$, but it is convex with respect to $\{\mathbf{w}, \mathbf{q}\}$ or $\{\mathbf{u}, \rho\}$ by given $\{\mathbf{u}, \rho\}$ or $\{\mathbf{w}, \mathbf{q}\}$, respectively. Thus, with fixed $\{\mathbf{w}_k\}$ and $\{q_l\}$, the optimal weight ρ_k is $\rho_k = 1/e_k^*$ where e_k^* is the optimal MSE for MU k . Then, the optimal receive beamforming coefficient $\{u_k\}$ under fixed $\{\mathbf{w}_k\}$, $\{q_l\}$ and $\{\rho_k\}$ is a MMSE receiver [31]

$$u_k = \left(\sum_{j=1}^K |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2 + \sum_{l=1}^L q_l |h_{kl}|^2 \right)^{-1} \mathbf{h}_k^H \mathbf{w}_k. \quad (20)$$

With fixed $\{u_k\}$ and $\{\rho_k\}$, the optimal $\{\mathbf{w}_k\}$ and $\{q_l\}$ can be obtained by solving the following quadratic constraint quadratic programming (QCQP) problem.

$$\mathbf{P}_5 : \min_{\mathbf{w}, \mathbf{q}} f(\mathbf{w}_k) + \sum_{k=1}^K \rho_k |u_k|^2 \sum_{l=1}^L q_l |h_{kl}|^2 + \frac{\alpha}{\tilde{\eta}} \sum_{l=1}^L p_l \quad (21)$$

s.t. (7b), (7c),

where $f(\mathbf{w}_k) = \sum_{k=1}^K (\mathbf{w}_k^H \mathbf{A} \mathbf{w}_k - \text{Re}\{\mathbf{b}_k^H \mathbf{w}_k\})$ is a quadratic objective function, and $\mathbf{b}_k = 2\rho_k u_k \mathbf{h}_k$ and $\mathbf{A} = \sum_{j=1}^K \rho_j |u_j|^2 \mathbf{h}_j \mathbf{h}_j^H$, and $\tilde{\eta} = \eta \ln 2$. It is observed that this problem is a convex optimization problem with respect to \mathbf{w} and \mathbf{q} , which can be solved centrally by standard mathematical tools, i.e., CVX [36]. It is noted that by simply replacing the constraints of the SINR and the maximum transmit power per BS as in [20], such an alternative optimization can also be adopted to solve \mathbf{P}_0 with multivariate compression. That because the replacement does not affect the convexity of the subproblems. However, such an interior-point method solves \mathbf{P}_5 with high computational complexity and it does not reveal the structure of the solution. Meanwhile, it is implementation intensive for large-scale C-RANs due to the centralized computation of the beamformings and quantization noises at the CU. Unlike the beamforming design problem in multicell system [30], the beamformings are coupled among BSs in our problem, making the Lagrangian based decomposition algorithm invalid in solving \mathbf{P}_5 . Towards this end, we propose a novel approach using ADMM method to solve \mathbf{P}_5 with closed-form solution optimally in a parallel manner.

In particular, in \mathbf{P}_5 , the two constraints (7b) and (7c), respectively, provide an upper and lower bound on \mathbf{q} . Then, the constraint $\beta_l p_l^t \leq P_l^{\max} - p_l^t$ should be satisfied. By rearranging this constraint, we have

$$(1 + \beta_l) p_l^t \leq P_l^{\max}, \forall l. \quad (22)$$

Since the objective function of \mathbf{P}_5 is monotonically decreasing over \mathbf{q} , we can replace the inequality constraint (7c) with equality, i.e., $q_l = \beta_l \sum_{k=1}^K |w_{kl}|^2$. We denote a new beamforming vector for BS l as $\tilde{\mathbf{w}}_l = [w_{1l}, \dots, w_{Kl}]^T$, and then we have $q_l = \beta_l |\tilde{\mathbf{w}}_l|^2$. As a result, problem \mathbf{P}_5 is equivalent to the following problem in only a single set of variables \mathbf{w} .

$$\mathbf{P}_6 : \min_{\mathbf{w}} f(\mathbf{w}_k) + \sum_{l=1}^L \mu_l |\tilde{\mathbf{w}}_l|^2 \quad (23a)$$

$$\text{s.t. } (1 + \beta_l) |\tilde{\mathbf{w}}_l|^2 \leq P_l^{\max}, \forall l, \quad (23b)$$

where $\mu_l = \sum_{k=1}^K \rho_k |u_k|^2 \beta_l |h_{kl}|^2 + \frac{\alpha}{\beta} (1 + \beta_l)$.

The objective function in (23a) contains two parts, and they are functions of different variables, i.e., \mathbf{w}_k and $\tilde{\mathbf{w}}_l$, rather than the same variable. Therefore, problem \mathbf{P}_6 is not a standard group lasso absolute shrinkage and selection operator (LASSO) problem [37]. Hence, the existing algorithms for the group LASSO problems are not directly applicable. This fact motivates us to find new algorithm to solve problem \mathbf{P}_6 . Fortunately, it is observed that problem \mathbf{P}_6 has a special structure that can be solved by developing the famous ADMM algorithm. To account for the difference between \mathbf{w}_k and $\tilde{\mathbf{w}}_l$ in problem \mathbf{P}_6 , we first introduce a copy $\tilde{\mathbf{z}}_l$ for $\tilde{\mathbf{w}}_l$, and define $\mathbf{z} = [\tilde{\mathbf{z}}_1^T, \dots, \tilde{\mathbf{z}}_L^T]^T$. Problem \mathbf{P}_6 can be equivalently expressed as

$$\mathbf{P}_7 : \min_{\mathbf{w}, \mathbf{z}} f(\mathbf{w}_k) + \sum_{l=1}^L \mu_l |\tilde{\mathbf{z}}_l|^2 \quad (24a)$$

$$\text{s.t. } (1 + \beta_l) |\tilde{\mathbf{z}}_l|^2 \leq P_l^{\max}, \forall l \quad (24b)$$

$$\tilde{\mathbf{z}}_l = \tilde{\mathbf{w}}_l, \forall l. \quad (24c)$$

The partial augmented Lagrangian function of problem \mathbf{P}_7 is

$$\begin{aligned} L(\mathbf{w}, \mathbf{z}, \mathbf{y}) &= f(\mathbf{w}_k) + \sum_{l=1}^L \mu_l |\tilde{\mathbf{z}}_l|^2 \\ &+ \sum_{l=1}^L \text{Re} \left(\tilde{\mathbf{y}}_l^H (\tilde{\mathbf{z}}_l - \tilde{\mathbf{w}}_l) \right) + \frac{c}{2} \sum_{l=1}^L |\tilde{\mathbf{z}}_l - \tilde{\mathbf{w}}_l|^2, \end{aligned} \quad (25)$$

where $\mathbf{y} = [\tilde{\mathbf{y}}_1^T, \dots, \tilde{\mathbf{y}}_L^T]^T$ with $\tilde{\mathbf{y}}_k = [\tilde{\mathbf{y}}_{k1}^T, \dots, \tilde{\mathbf{y}}_{kL}^T]^T$ is the vector of Lagrangian dual variables for the equality constraint (24c), and $c > 0$ is some constant.

The idea of the ADMM is to update the local variables when fixing the other variables. Specifically, the variables updating procedure of the ADMM algorithm is detailed as follows.

By fixing $\mathbf{w}^{(m)}$ and $\mathbf{y}^{(m)}$ at the (m) -th iteration, $\mathbf{z}^{(m+1)}$ at the $(m+1)$ -th iteration is updated by solving the following convex problem

$$\min_{\mathbf{z}} L(\mathbf{w}^{(m)}, \mathbf{z}, \mathbf{y}^{(m)}) \quad \text{s.t. } (24b). \quad (26)$$

We show that problem (26) can be solved in a parallel manner. Specifically, we first decompose problem (26) into L subproblems

$$\min_{\tilde{\mathbf{z}}_l} \mu_l |\tilde{\mathbf{z}}_l|^2 + \text{Re} \left(\tilde{\mathbf{y}}_l^H (\tilde{\mathbf{z}}_l - \tilde{\mathbf{w}}_l^{(m)}) \right) + \frac{c}{2} |\tilde{\mathbf{z}}_l - \tilde{\mathbf{w}}_l^{(m)}|^2 \quad \text{s.t. } (24b), \quad (27)$$

that can be solved independently at the CU.

The Karush-Kuhn-Tucker (KKT) conditions of problem (27) are

$$(2\mu_l + 2\theta_l^* + c)\tilde{\mathbf{z}}_l^* - c\mathbf{s}_l = \mathbf{0}, \tag{28}$$

$$(1 + \beta_l)|\tilde{\mathbf{z}}_l^*|^2 \leq P_l^{\max}, \theta_l^* \geq 0, \tag{29}$$

$$\left((1 + \beta_l)|\tilde{\mathbf{z}}_l^*|^2 - P_l^{\max} \right) \theta_l^* = 0, \tag{30}$$

where θ_l^* is the optimal Lagrangian multiplier associated with the power constraint, $\tilde{\mathbf{z}}_l^*$ is the optimum of $\tilde{\mathbf{z}}_l$, and \mathbf{s}_l is defined as $\mathbf{s}_l = \tilde{\mathbf{w}}_l^{(m)} - \tilde{\mathbf{y}}_l^{(m)}$.

If $\theta_l^* = 0$, we have $\tilde{\mathbf{z}}_l^* = \frac{c\mathbf{s}_l}{2\mu_l+c}$ under the condition of $(1 + \beta_l)|\tilde{\mathbf{z}}_l^*|^2 \leq P_l^{\max}$ (equivalent to $|\mathbf{s}_l| \leq \sqrt{\frac{P_l^{\max}}{1+\beta_l}}(2\mu_l/c + 1)$). If $\theta_l^* > 0$, $(1 + \beta_l)|\tilde{\mathbf{z}}_l^*|^2 = P_l^{\max}$. Then, $\tilde{\mathbf{z}}_l^* = \sqrt{\frac{P_l^{\max}}{1+\beta_l}}$ when $|\mathbf{s}_l| > \sqrt{\frac{P_l^{\max}}{1+\beta_l}}(2\mu_l/c + 1)$. Specially, when $|\mathbf{s}_l| = 0$, $\tilde{\mathbf{z}}_l^* = \mathbf{0}$. Therefore, $\mathbf{z}_l^{(m+1)} = \tilde{\mathbf{z}}_l^*$ is updated as

$$\mathbf{z}_l^{(m+1)} = \begin{cases} \mathbf{0}, & |\mathbf{s}_l| = 0 \\ \frac{c\mathbf{s}_l}{2\mu_l+c}, & |\mathbf{s}_l| \leq \sqrt{\frac{P_l^{\max}}{1+\beta_l}}(2\mu_l/c + 1) \\ \sqrt{\frac{P_l^{\max}}{1+\beta_l}}, & |\mathbf{s}_l| > \sqrt{\frac{P_l^{\max}}{1+\beta_l}}(2\mu_l/c + 1). \end{cases} \tag{31}$$

To update $\mathbf{w}^{(m+1)}$ with fixed $\{\mathbf{y}_k^{(m)}, \mathbf{z}_k^{(m+1)}\}$, the following problem is solved

$$\min_{\mathbf{w}} L(\mathbf{w}, \mathbf{z}^{(m+1)}, \mathbf{y}^{(m)}). \tag{32}$$

Due to the relationship between $\mathbf{w}_k = [w_{kl}, \dots, w_{kL}]^T$ and $\tilde{\mathbf{w}}_l = [w_{1l}, \dots, w_{kl}]^T, \forall l, k$, we have $\sum_{l=1}^L |\tilde{\mathbf{w}}_l - \tilde{\mathbf{z}}_l^{(m+1)} - \tilde{\mathbf{y}}_l^{(m)} / c|^2 = \sum_{k=1}^K |\mathbf{w}_k - \mathbf{z}_k^{(m+1)} - \mathbf{y}_k^{(m)} / c|^2$. Problem (32) is equivalent to

$$\min_{\mathbf{w}} f(\mathbf{w}_k) + \frac{c}{2} \sum_{k=1}^K |\mathbf{w}_k - \mathbf{z}_k^{(m+1)} - \mathbf{y}_k^{(m)} / c|^2. \tag{33}$$

Then, problem (33) can be decomposed into the following K subproblems, and can be solved in a parallel manner with each MU.

$$\min_{\mathbf{w}_k} \mathbf{w}_k^H \mathbf{A} \mathbf{w}_k - \text{Re}\{\mathbf{b}_k^H \mathbf{w}_k\} + \frac{c}{2} |\mathbf{w}_k - \mathbf{z}_k^{(m+1)} - \mathbf{y}_k^{(m)} / c|^2. \tag{34}$$

By differentiating (34) with respect to $\{\mathbf{w}_k\}$ and set to zero, we obtain the optimal $\{\mathbf{w}_k^*\}$ with closed-form expression in the $(m + 1)$ -th iteration, given by

$$\mathbf{w}_k^{(m+1)} = \mathbf{w}_k^* = (2\mathbf{A}_k + c\mathbf{I})^{-1}(2\mathbf{b}_k + c\mathbf{z}_k^{(m+1)} + \mathbf{y}_k^{(m)}). \tag{35}$$

Using the relationship between \mathbf{w}_k and $\tilde{\mathbf{w}}_l$, $\tilde{\mathbf{w}}_l^{(m+1)}$ is obtained. Then, with obtained $\tilde{\mathbf{w}}_l^{(m+1)}$ and $\tilde{\mathbf{z}}_l^{(m+1)}$, the multipliers $\tilde{\mathbf{y}}_l^{(m+1)}$ are updated

$$\tilde{\mathbf{y}}_l^{(m+1)} = \tilde{\mathbf{y}}_l^{(m)} + c(\tilde{\mathbf{z}}_l^{(m+1)} - \tilde{\mathbf{w}}_l^{(m+1)}); \tag{36}$$

Therefore, the decentralized algorithm for solving \mathbf{P}_3 is summarized in Algorithm 3. The convergence of Algorithm 3 is guaranteed by Theorem 1.

Algorithm 3 Decentralized Algorithm for Subproblem \mathbf{P}_3 .

-
- 1: Initialization: choose initial value for $\mathbf{w}^{(0)}$ and $\mathbf{q}^{(0)}$, set iteration index $n = 0$, and choose the initial value for $\mathbf{z}^{(0)}$ and $\mathbf{y}^{(0)}$.
 - 2: **Repeat**
 - 3: Update $\{\mathbf{u}\}$ by (20), and obtain $\{\mathbf{u}^{(n)}\}$.
 - 4: Calculate $\{e_k\}$ according to (18) with $\{\mathbf{w}^{(n)}, \mathbf{q}^{(n)}\}$.
 - 5: Update $\{\rho_k\}$ with $\rho_k = 1/e_k$, and obtain $\rho^{(n)}$.
 - 6: Let iteration index $m = 0$ and $\mathbf{w}^{(m)} = \mathbf{w}^{(n)}$.
 - 7: **while** $\epsilon_{ADMM} \geq 10^{-5}$ **do**
 - 8: Update $\mathbf{z}^{(m+1)}$ by (31);
 - 9: Update $\mathbf{w}^{(m+1)}$ by (35);
 - 10: Update the multipliers $\mathbf{y}^{(m+1)}$ by (36).
 - 11: $m = m + 1$.
 - 12: **end while**
 - 13: $n = n + 1$.
 - 14: Update $\mathbf{w}^{(n)} = \mathbf{w}^{(m+1)}$.
 - 15: Update $\{\mathbf{q}^{(n)}\}$ by $q_l = \beta_l \sum_{k=1}^K |w_{kl}|^2, \forall l$ with $\mathbf{w}^{(n)}$.
 - 16: **Until** convergence or the maximum iteration number is reached.
-

Theorem 1. Algorithm 3 generates a sequence $\{\mathbf{w}^{(n)}, \mathbf{q}^{(n)}\}$ that converge to a stationary point of problem \mathbf{P}_3 .

Proof. The proof is based on the convergence of the alternative optimization method and ADMM algorithm. With initialized $\{\mathbf{w}^{(0)}, \mathbf{z}^{(0)}, \mathbf{y}^{(0)}\}$, the inner loop of Algorithm 3 from steps 7 to 12 converges to an optimal solution of \mathbf{P}_5 due to the convergence of ADMM algorithm, and the proof can be found in [38]. On the other hand, the outer loop of Algorithm 3 converges to a stationary point of subproblem \mathbf{P}_3 due to the convergence property of block coordinate decent algorithm [31,35]. According to [25,30], for an arbitrary α , the objective (16a) in problem \mathbf{P}_3 is shown to be non-decreasing during each iteration of the outer loop of Algorithm 3. Therefore, Algorithm 3 is guaranteed to converge to a stationary point of problem \mathbf{P}_3 , and the proof is completed. \square

Denote by α^{opt} the actually optimal solution of problem \mathbf{P}_0 , and α^* the obtained suboptimal solution returned by the Algorithm 2 when using Algorithm 3 to solve problem \mathbf{P}_3 . Since Algorithm 3 converges to a stationary point of problem \mathbf{P}_3 , the suboptimal objective of $G(\alpha)$ equals to zero which equivalently equals to $G(\alpha^*) > 0$. Moreover, considering the fact that $G(\alpha)$ is monotonically decreasing, the actually optimal solution α^{opt} that satisfies $G(\alpha^{\text{opt}}) = 0$ must be no smaller than α^* , i.e., $\alpha^{\text{opt}} \geq \alpha^*$. Therefore, the obtained solution returned by Algorithm 2 when using Algorithm 3 is no larger than the optimal solution of problem \mathbf{P}_0 . Simulation results will demonstrate that the obtained solution is very close to the optimal one that verifies the effectiveness of the proposed algorithm numerically.

Combining Algorithms 2 and 3, it is concluded that problem \mathbf{P}_0 can be solved efficiently by the proposed TLD algorithm. Since the limits of the upper and lower bounds are updated iteratively, the bisection procedure will stop any way. However, we may not have $G(\alpha) = 0$.

4.2.1. Parallelized Implementation

Since problem (26) is decomposed into L subproblems, $\mathbf{z}^{(m+1)}$ at step 8 in Algorithm 3 is updated in a parallel manner with closed-form solution. Similarly, the beamformings $\mathbf{w}^{(m+1)}$ and multipliers $\mathbf{y}^{(m+1)}$ are also updated using (35) and (36) in a parallel manner with closed-form solution. Therefore, we derive closed-form expressions for the optimal beamformings, the optimal receiver filters and the auxiliary variables in Algorithm 3, that provide some insights on the EEmax problem. For example, if $\mathbf{w}_l = \mathbf{0}$ (no transmission power consumed by BS l), BS l can be switched off to save static power and improve EE.

4.2.2. Complexity Analysis

According to the algorithm flow, the computational complexity of Algorithm 3 contains two parts. In Algorithms 3, the computational complexity of computing (20) arises from the matrix inversion of the receive beamforming, i.e., $\mathcal{O}(L^3)$. When the interior-point method is adopted to solve \mathbf{P}_5 , the computational complexity is $\mathcal{O}((LK)^{3.5})$ [39]. In this case, to serve K MUs, the overall computational complexity is in the order of $\mathcal{O}(KL^3 + (LK)^{3.5})$. On the other hand, since the ADMM algorithm is applied to solve \mathbf{P}_5 , the main computational complexity is the matrix inversion in step 9 of Algorithm 3 where the transmit beamformings are computed by (35). Thus, the overall computational complexity of Algorithm 3 is in the order of $\mathcal{O}(KL^3 + KL^3)$. For the outer layer algorithm, simulation results will show that it converges rapidly (about 5 iterations). Therefore, we can deduce that the proposed TLD algorithm is computation efficient.

4.2.3. Generalization to the Multi-Antenna System

Although a single-antenna is equipped at each BS and each MU in the above discussion, we claim that proposed TLD algorithm can be generalized to the multi-antennas BSs and the single-antenna MUs scenarios. This is because deployed multiple antennas (each with N antennas) at BSs, one only needs to replace the corresponding channel coefficient h_{kl} from BS l to MU k with $\mathbf{h}_{kl} \in \mathbb{C}^{N \times 1}$. Similarly, the transmit beamforming coefficient w_{kl} and receive beamforming coefficient u_k from BS l to MU k are replaced with vectors $\mathbf{w}_{kl} \in \mathbb{C}^{N \times 1}$ and $\mathbf{u}_k \in \mathbb{C}^{N \times 1}$, respectively. While e_l is replaced by $\mathbf{e}_l \in \mathbb{C}^{N \times 1}$, and the corresponding covariance matrix is $\mathbb{E}[\mathbf{e}_l \mathbf{e}_l^H] = \text{diag}(q_{11}, \dots, q_{1N})$. If $q_l = q_{11} = \dots = q_{1N}$, one only need to replace q_l in (7b) and (7c) with Nq_l . Therefore, the proposed TLD algorithm can be extended to solve the EEmax problem with multi-antenna BSs and single-antenna MUs but requires additional efforts. Moreover, for the multi-antenna BSs and multi-antenna MUs C-RAN, we point out that the weighted minimum MSE algorithm in [31] might provide some insights on how to apply the proposed algorithm.

5. Robust Algorithms with Imperfect CSI

In practical C-RAN, due to the limited feedback [40], partial CSI [41] or estimated error, the obtained channel is not perfect. According to [42–44], the imperfection in CSI has significantly impact on the system performance. Therefore, we will extend the proposed algorithms in the previous two sections to solve the robust EEmax problem in the presence of imperfect CSI.

The same system model is considered as in Section 2. Since the path-loss fading and the log-normal fading can be estimated accurately, the imperfection usually comes from the uncertain small-scale fading. Thus, different from the worst-case design [42,43], the Gaussian distribution in [44] is adopted to model the channel imperfection. The real channel from BS l to MU k is expressed as

$$h_{kl} = g_{kl}(\tilde{h}_{kl} + \tilde{\Delta}_{kl}), \forall l, k, \quad (37)$$

where $g_{kl} = GL(d_{kl})\varphi_{kl}$ is the channel gain consisting of the antenna gain G , the path-loss fading $L(d_{kl})$ at distance d_{kl} in km and the log-normal random fading φ_{kl} . \tilde{h}_{kl} and $\tilde{\Delta}_{kl}$ are respectively the estimated channel and channel uncertainty from BS l to MU k . $\tilde{\Delta}_{kl}$ is assumed to be independent identically distributed (i.i.d) zero mean circularly symmetrical complex Gaussian (ZMCSCG) random variables with variance σ_e^2 . The channel from all the BSs to MU k is $\mathbf{h}_k = \mathbf{g}_k(\tilde{\mathbf{h}}_k + \tilde{\Delta}_k)$, $\forall k$, where $\mathbf{g}_k = \text{diag}(g_{k1}, \dots, g_{kL})$, and $\tilde{\mathbf{h}}_k$ and $\tilde{\Delta}_k$ are the aggregated collection of \tilde{h}_{kl} and $\tilde{\Delta}_{kl}$, respectively. Let $\tilde{\mathbf{h}}_k = \mathbf{g}_k \tilde{\mathbf{h}}_k$, $\tilde{h}_{kl} = g_{kl} \tilde{h}_{kl}$ and $\mathbf{D}_k = \text{diag}(|g_{k1}|^2, \dots, |g_{kL}|^2)$. The received SINR at MU k is expressed as

$$\text{SINR}_k = \frac{|\bar{\mathbf{h}}_k^H \mathbf{w}_k|^2}{\sum_{i \neq k} |\bar{\mathbf{h}}_k^H \mathbf{w}_i|^2 + \sigma^2 + \sum_{l=1}^L q_l (|\bar{h}_{kl}|^2) + Y}, \quad (38)$$

where $Y = \sigma_e^2 \text{Tr}(\mathbf{D}_k) \sum_{i=1}^K |\mathbf{w}_i|^2 + \sum_{l=1}^L q_l \sigma_e^2$ denotes the interference caused by uncertainty part of CSI. The first term of Y contains the interference caused by the CSI error of the intended signal and the signals of other MUs.

The robust EEmax problem has the same form as problem \mathbf{P}_0 by simply using the imperfect channel, i.e., replacing h_{kl} by (28). The details of the extended algorithms to solve the robust problem with the consideration of imperfect CSI are given in the following two subsections, respectively.

5.1. Robust Optimal Algorithm

To apply the optimal algorithm under imperfect CSI, we only need to reformulate (8b) because the CSI is only involved in (8b). In particular, (8b) is recast as

$$\frac{1}{\tilde{\gamma}_k} \bar{\mathbf{h}}_k \mathbf{w}_k \geq \left(\sum_{i \neq k} |\bar{\mathbf{h}}_k^H \mathbf{w}_i|^2 + \sigma^2 + \sum_{l=1}^L q_l (|\bar{h}_{kl}|^2) + Y \right)^{1/2}, \quad (39a)$$

$$\text{Im}(\mathbf{h}_k \mathbf{w}_k) = 0, \forall k \in \mathcal{K}. \quad (39b)$$

Follow the same procedure as demonstrated in Algorithm 1, one only need to replace constraint (10a) in problem \mathbf{P}_2 with (39a). In this case, the computational complexity is the same as Algorithm 1 (both of them are very high but provide optimal solutions).

5.2. Robust TLD Algorithm

Since the CSI is involved only in R_k , the TLD algorithm performs well to solve the robust EEmax problem but requires some transformations. In particular, the out layer iteration of the robust TLD algorithm follows the procedure as Algorithm 2. To solve the robust subproblem \mathbf{P}_3 in the inner layer, Algorithm 3 cannot be applied directly. Fortunately, with the same transformations as Section 4.2, the corresponding MMSE receiver can be expressed as

$$\tilde{u}_k = \frac{\bar{\mathbf{h}}_k^H \mathbf{w}_k}{\sum_{j=1}^K |\bar{\mathbf{h}}_k^H \mathbf{w}_j|^2 + \sigma^2 + \sum_{l=1}^L q_l (|\bar{h}_{kl}|^2) + Y}, \quad (40)$$

and the MSE of the k -th MU is

$$\begin{aligned} \tilde{e}_k &= \mathbb{E} \left[|\tilde{u}_k^H y_k - s_k|^2 \right] \\ &= |\tilde{u}_k|^2 \sum_{j=1}^K \text{Tr} \left(\mathbf{w}_j (\bar{\mathbf{h}}_k \bar{\mathbf{h}}_k^H) \mathbf{w}_j^H \right) + |\tilde{u}_k|^2 \sigma^2 + 1 \\ &\quad - 2 \text{Re} \{ \tilde{u}_k \bar{\mathbf{h}}_k^H \mathbf{w}_k \} + |\tilde{u}_k|^2 \left(\sum_{l=1}^L q_l |\bar{h}_{kl}|^2 + Y \right). \end{aligned} \quad (41)$$

Then, with fixed $\{\tilde{u}_k\}$ and $\{\rho_k\}$, the robust \mathbf{P}_5 with respect to $\{\mathbf{w}_k\}$ and $\{q_l\}$ becomes

$$\begin{aligned} \mathbf{P}_8 : \min_{\mathbf{w}, \mathbf{q}} \quad & \sum_{k=1}^K \rho_k \tilde{e}_k + \frac{\alpha}{\tilde{\eta}} \sum_{l=1}^L p_l \\ \text{s.t.} \quad & (6b), (6c). \end{aligned} \quad (42)$$

We further reformulate problem \mathbf{P}_8 as

$$\mathbf{P}_9 : \min_{\mathbf{w}} \tilde{f}(\mathbf{w}_k) + \sum_{l=1}^L \tilde{\mu}_l |\tilde{\mathbf{w}}_l|^2 \quad (43a)$$

$$\text{s.t. } (1 + \beta_l) |\tilde{\mathbf{w}}_l|^2 \leq P_l^{\max}, \forall l, \quad (43b)$$

where $\tilde{f}(\mathbf{w}_k) = \sum_{k=1}^K (\mathbf{w}_k^H \tilde{\mathbf{A}} \mathbf{w}_k - \text{Re}\{\tilde{\mathbf{b}}_k^H \mathbf{w}_k\})$, $\tilde{\mu}_l = \sum_{k=1}^K \rho_k |\tilde{u}_k|^2 \beta_l (|\tilde{h}_{kl}|^2 + \sigma_e^2) + \frac{\alpha}{\beta_l} (1 + \beta_l)$, and

$$\tilde{\mathbf{A}} = \sum_{j=1}^K \rho_j |\tilde{u}_j|^2 (\tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^H + \sigma_e^2 \mathbf{D}_j), \quad (44)$$

$$\tilde{\mathbf{b}}_k = 2\rho_k \tilde{u}_k \tilde{\mathbf{h}}_k. \quad (45)$$

Since problems \mathbf{P}_6 and \mathbf{P}_9 have the same form, the ADMM algorithm presented in the inner layer of Algorithm 3 performs well in solving \mathbf{P}_9 . Particularly, \mathbf{z} is updated by (31), and $\{\mathbf{w}_k\}$ is updated using the following closed-form expression

$$\mathbf{w}_k = (2\tilde{\mathbf{A}}_k + c\mathbf{I})^{-1} (2\tilde{\mathbf{b}}_k + c\mathbf{z}_k + \mathbf{y}_k), \quad (46)$$

where $\tilde{\mathbf{A}}_k$ and $\tilde{\mathbf{b}}_k$ are shown in (44) and (45), respectively.

The corresponding multipliers \mathbf{y} are updated by (36). Similar to the TLD algorithm, the robust TLD algorithm procedure is omitted here for brevity. It is easily verified that the robust TLD algorithm is convergent, and it has the same computational complexity as the non-robust TLD algorithm, and it achieves a suboptimal solution with closed-form expression in a parallel manner as well.

6. Simulation Results and Discussions

In this section, we evaluate the performance of the proposed algorithms via Monte-Carlo simulation. We consider a downlink C-RAN with $L = 5$ single-antenna BSs and $K = 4$ single-antenna MUs, where one BS locates at the circle centre, and the other four BSs are located in a circle region at equal distances apart with radius 0.5 km, as shown in Figure 1. The four single-antenna MUs are randomly deployed in the circle with uniform distribution. Each BS and MU are equipped with a single antenna, and we set the maximum transmit power $P^m = P_l^{\max}$ and the fronthaul capacity $C = C_l$ for all BSs. The convergence errors are set as ϵ for all the proposed algorithms. Unless specified, other simulation parameters are listed in Table 1, and all the simulation results are averaged over 50 times independent MU locations, each with a single random channel realization. In order to verify the effectiveness of the proposed algorithms, we consider two baseline algorithms for comparison:

SRmax algorithm: The proposed TLD algorithm is adopted to solve the SRmax problem by simply setting $\alpha = 0$.

DC algorithm: We also modify the algorithm proposed in [22] for comparison. Specifically, to arrive at a tractable formulation, one can use the epigraph form of the original problem. Based on the transformation, one can transform the constraints into convex ones by using the first-order Taylor expansion. Then, the approximated problem is iteratively solved, and the solution converges to a KKT point [22]. The computational complexity of the DC algorithm using the interior-point method is $O(I_{\max} L^{6.5} K^{3.5} N^{6.5})$ (I_{\max} is the maximum of iteration number), which is much higher than the TLD algorithm.

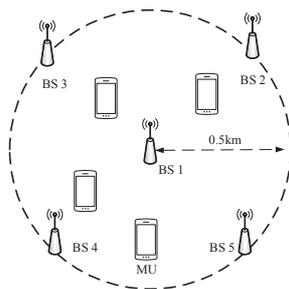


Figure 1. Simulation setup in a downlink C-RAN.

Table 1. Simulation parameters.

Parameter	Value
Maximum transmit power P^m	1 W
Power amplifier efficiency η [4]	1/4
Transmit antenna power gain G	9 dBi
Background noise σ^2	-104 dBm
Path-loss fading from BS l to MU k [4]	$148.1 + 37.6\log_{10} d_{kl}$
Log-normal shadowing φ_{kl}	8 dB
Static power P^c	2 W
Convergence errors $\epsilon = \epsilon_{ADMM}$	10^{-5}
Maximum iteration number	50

6.1. Non-Robust Performance

We first investigate the convergence behavior of the proposed algorithms over a typical random channel realization for $P^c = 2$ W, $P^m = 30$ dBm, and $C = 5$ bits/s/Hz, which are presented in Figure 2.

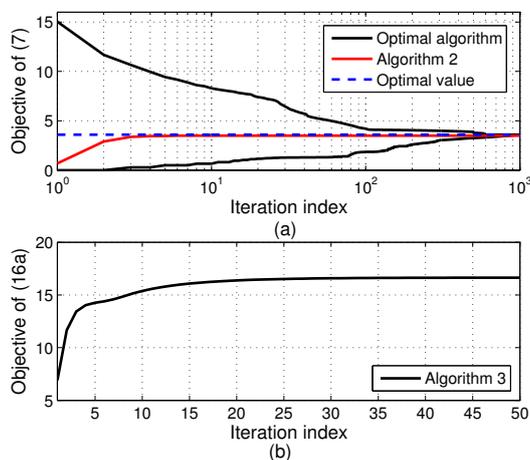


Figure 2. Convergence of the proposed algorithms over a typical random channel realization. (a) EE of Algorithms 1 and 2; (b) Objective of problem P_3 with $\alpha = 2.90$ (at the 2nd iteration of Algorithm 2 in Figure 2a).

It is observed that the outer layer (Algorithm 2) of TLD in Figure 2a converges to a near optimum very fast (about 5 iterations). Meanwhile, Figure 2b illustrates that the objective function of problem (16) of Algorithm 3 at the 2nd iteration of Algorithm 2 in Figure 2a increases and converges to a stationary point in less than 30 iterations. Thus, the proposed TLD algorithm converges rapidly. It is also found that the convergence rate of the optimal algorithm is much slower than the TLD algorithm. Due to a large number of infeasible boxes being removed, the gap between upper and lower bounds of the optimal algorithm reduces rapidly during the first iterations.

Then, we explore the effect of fronthaul capacity on the performance in terms of EE, sum rate and transmit power, shown in Figures 3–5, respectively.

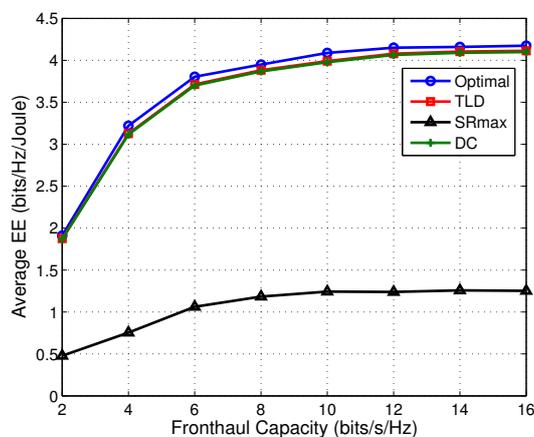


Figure 3. Average EE at various fronthaul capacities.

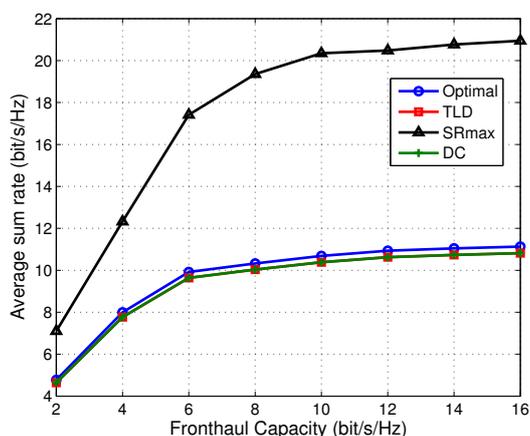


Figure 4. Average sum rate at various fronthaul capacities.

The TLD algorithm achieves a comparable performance compared to the DC one in terms of EE. Figure 3 shows the EE comparison among the four algorithms under different fronthaul capacities. As can be seen from Figure 3 that the TLD and DC algorithms achieve an approximate EE performance as the optimal one, and both of them outperform the SRmax one by about three times in the middle-high fronthaul capacity region. Since the network power is penalized in (7a), and the sum rate and power are jointly optimized, the EE is improved compared to SRmax. This can be explained from Figures 4 and 5 that the saved ratio of the transmit power of TLD (save for about 95% compared with SRmax) is much higher than its sum rate reduction ratio (decrease by about 50% compared with SRmax), resulting in a higher EE. It is also observed from Figure 3 that the EE performance increases with the growing fronthaul capacity in the low fronthaul capacity region, and then gradually converge to a constant in the middle to high fronthaul capacity region, and a similar trend can be found in Figure 4. According to the expression of quantization noises (QN), i.e., $q_l = \frac{1}{2^{c-1}} \sum_{k=1}^K |w_{kl}|^2$, it reduces exponentially with the increasing fronthaul capacity (shown in Figure 5). As a result, at low fronthaul capacity region, the SINR is limited by the high QN, leading to a low sum rate in this region. Moreover, the corresponding data transmit power (DTP) in Figure 5 increases very slowly from low to high fronthaul capacity region. That because with the increase of the fronthaul capacity, more powers of DTP and QN are consumed to improve the sum rate. However, more network power is consumed which makes the increment of EE gradually. It should be noted that the optimal algorithm cannot provide DTP and QN but only the total network power, as a result only the transmit power (sum of DTP and QN) of the

optimal algorithm is given in Figure 5. In summary, Figure 3–5 illustrate that the TLD and optimal algorithms can achieve the balance between the sum rate and power consumption, implying a much higher EE than the SRmax one.

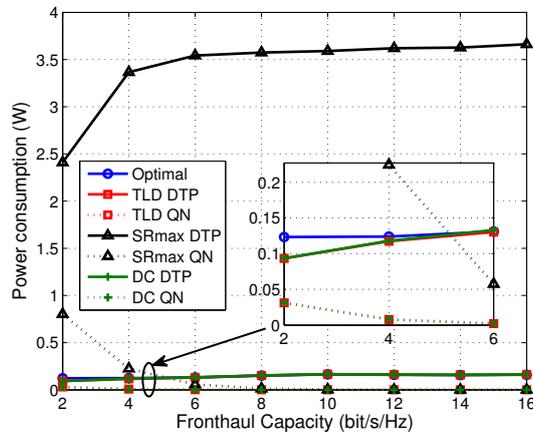


Figure 5. Average power consumption at various fronthaul capacities.

Figure 6 compares the EE performance with respect to the maximum transmit power P^m for a given fronthaul capacity $C = 5$ bits/s/Hz. It is observed that TLD and DC significantly outperform the SRmax in terms of EE in the middle to high transmit power region (≥ 18 dBm). This is because by jointly optimizing the sum rate and network power, the increased sum rate is slightly larger than the increased power consumption ratio, resulting in a gradually increase of EE. While for the SRmax algorithm, because of the interference between MUs, its sum rate gain cannot compensate for the negative impact of the network power, making its EE worse than the TLD and DC algorithms in the event that $P^m \geq 18$ dBm. It is also observed that the optimal, and TLD and DC algorithms have comparable EE performance as the SRmax one at low maximum transmit power (≤ 18 dBm), which suggests that at this region, transmitting with the maximum transmit power provides comparable sum rate gain for the optimal, EEmax and SRmax algorithms.

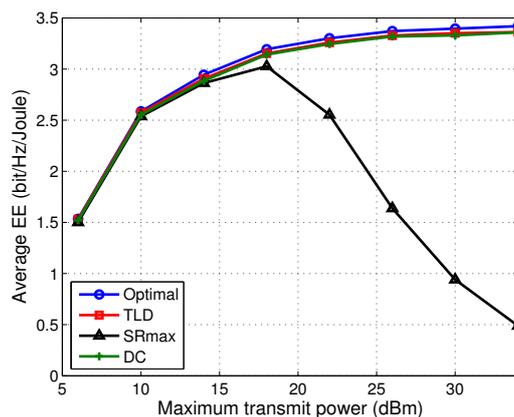


Figure 6. Average EE versus maximum transmit power with $P^c = 2$ W and $C = 5$ bits/s/Hz.

To investigate further, we also test the impact of static power consumption on the EE performance of the proposed algorithms, shown in Figure 7. It depicts that, at low to middle static power ($P^c \leq 6$ W), the EE of the optimal, DC and TLD algorithms decreases significantly with the growing static power P^c . This can be explained that the static power dominates the network power at this region and the transmit power is optimized as well, leading to a rapid decrease of EE. Whereas, since the maximum transmit power is usually adopted by the SRmax algorithm for transmission and it accounts for a large amount of the network power for P^c ranging from 2 to 10 W, the decrease EE of the SRmax algorithm

becomes gentle. These facts bring a higher EE performance for the TLD algorithm than the SRmax at different static powers.

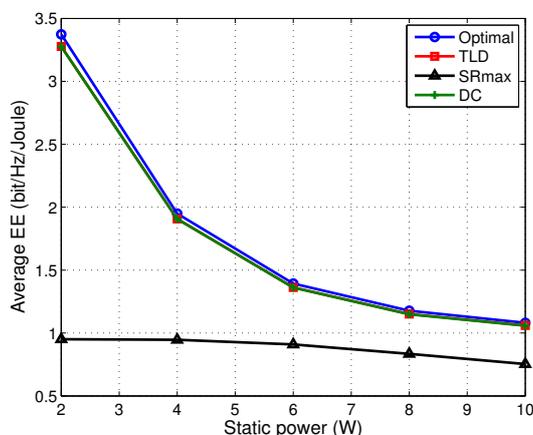


Figure 7. Average EE versus static power with $P^m = 30$ dBm and $C = 5$ bits/s/Hz.

The effect of MU number on the EE performance is also investigated with $C = 5$ bits/s/Hz, $P^m = 30$ dBm and $P^c = 2$ W, which is shown in Figure 8. We learn from Figure 8 that, with more MUs to be served, the EE of the TLD and DC algorithms increase significantly and exhibits obvious advantage over the SRmax one, e.g., higher than SRmax by more than 300%. Intuitively, the data rate of each MU decreases with the growing number of served MUs due to the increased interference between MUs. The obtained sum rate gain is very large when the serving MU number is small, while the sum rate increases much more slowly if the served MU number is large. It is attributed to the rapidly increase of interference among MUs by serving large number of MUs, implying a data rate reduction for each MU and a slightly increase on the sum rate of all MUs. Besides, more transmit power is needed to achieve the sum rate gain when serving more MUs, which also perform negative for boosting EE performance. It should also be noted that by further increasing the number of MUs, the EE of all the four algorithms will firstly become stable and then show a decrease trend. This is because, due to the increased interference caused by increasing the MU number, the decreased sum data rate of MUs is much larger than the increased data rate brought by the newly added MUs.

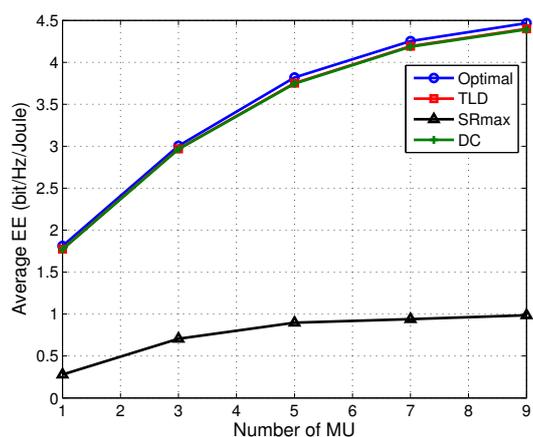


Figure 8. Average EE versus MU number with $P^m = 30$ dBm, $C = 5$ bits/s/Hz and $P^c = 2$ W.

6.2. Robust Performance

The EE performance of the proposed robust TLD algorithm is investigated over the channels with different channel errors, and the results are shown in Figure 9. Similar to the non-robust design, each point in Figure 9 is averaged over 100 channel realizations. In the simulation, $\sigma_e^2 = 0$ means

perfect CSI is known at the CU. It was observed in the previous subsection that the SRmax algorithm has a worse EE performance than that of the TLD (TLD has near optimal performance), and thus we only plot the EE curve of robust TLD algorithm under different CSI errors. Due to the high complexity imposed by the DC algorithm, we do not compare it with the TLD in this section.

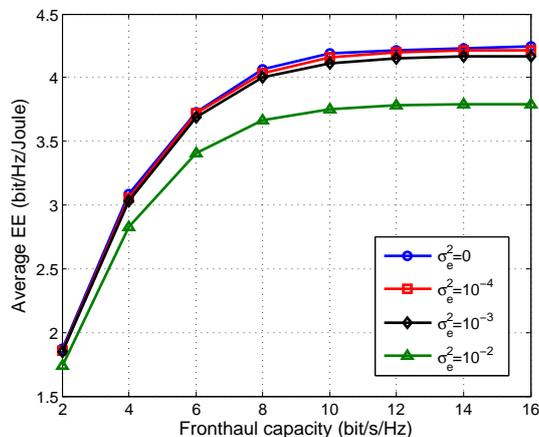


Figure 9. Average EE versus fronthaul capacity with different channel errors, $P^m = 30$ dBm, $P^c = 2$ W and $C = 5$ bits/s/Hz.

We can see from Figure 9 that by increasing the channel errors from $\sigma_e^2 = 0$ to $\sigma_e^2 = 10^{-2}$, the EE performance decreases significantly especially in the middle to high fronthaul capacity. In the low fronthaul capacity region, according to (38), the received SINR is slightly influenced by the channel errors since the increased interference is very small. While a worse sum rate is obtained in the middle to high fronthaul capacity due to the impairment of channel error on the received SINR, and more amount of transmit power is consumed in order to achieve the same sum rate for a larger channel errors. As a result, the sum rate gain cannot compensate the network power reduction that caused by the channel error, resulting in lower EE. In summary, the EE performance is susceptible to the channel error especially for a larger channel error.

7. Conclusions

In this paper, we have studied the EEmax problem with fronthaul compression in a downlink C-RAN. The optimization problem is formulated as a non-convex fractional programming problem. We have proposed an optimal algorithm based on BnB method to provide a performance benchmark. Further, a near optimal TLD algorithm has been proposed via a bisection search procedure in conjunction with an ADMM method. The proposed algorithms were guaranteed to be convergent, and the solution of TLD was achieved with closed-form in a parallel manner. Simulation results illustrated that the TLD algorithm converged to a near optimal solution very fast, and it achieved a much higher EE than the SRmax algorithm. The results further indicated that EE could be improved by increasing the fronthaul link capacity or optimizing the network power. Numerical analysis also demonstrated that the robust TLD algorithm can provide robust performance in the case of lacking perfect CSI, and its performance is susceptible to the channel error.

Acknowledgments: This paper is supported by National Natural Science Foundation of China (61571162), Heilongjiang Province Natural Science Foundation (F2016019), and National Science and Technology Major Specific Projects of China (2015ZX03004002-004).

Author Contributions: Yong Wang conceived the original idea and wrote this paper; Lin Ma and Yubin Xu analyzed the data and revised the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Andrews, J.; Buzzi, S.; Choi, W.; Hanly, S.; Lozano, A.; Soong, A.; Zhang, J. What Will 5G Be? *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1065–1082.
- China Mobile Research Institute. *C-RAN: The Road Towards Green RAN*; White Paper, Version 2.5; China Mobile Research Institute: Beijing, China, 2011.
- Peng, M.; Li, Y.; Jiang, J.; Li, J.; Wang, C. Heterogeneous Cloud Radio Access Networks: A New Perspective for Enhancing Spectral and Energy Efficiencies. *IEEE Wirel. Commun.* **2014**, *21*, 126–135.
- Shi, Y.; Zhang, J.; Letaief, K.B. Group Sparse Beamforming for Green Cloud-RAN. *IEEE Trans. Wirel. Commun.* **2014**, *5*, 2809–2823.
- Tang, J.; Tay, W.; Quek, T. Cross-Layer Resource Allocation with Elastic Service Scaling in Cloud Radio Access Network. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 5068–5081.
- Luo, S.; Zhang, R.; Lim, T. Downlink and Uplink Energy Minimization Through User Association and Beamforming in Cloud RAN. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 494–508.
- Cheng, Y.; Pesavento, M.; Philipp, A. Joint Network Optimization and Downlink Beamforming for COMP Transmissions Using Mixed Integer Conic Programming. *IEEE Trans. Signal Process.* **2013**, *61*, 3972–3987.
- Zhuang, F.; Lau, V. Backhaul Limited Asymmetric Cooperation for MIMO Cellular Networks via Semidefinite Relaxation. *IEEE Trans. Signal Process.* **2014**, *62*, 982–997.
- Zhao, J.; Quek, T.; Lei, Z. Coordinated Multipoint Transmission with Limited Backhaul Data Transfer. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 2762–2775.
- Wang, Y.; Liang, B.; Xu, Y. A Two-Stage Rank Selection Scheme in Downlink COMP Transmission Networks. In Proceedings of the IEEE International Conference of Communication (ICC), Kuala Lumpur, Malaysia, 23–27 May 2016; pp. 1–6.
- Patil, P.; Yu, W. Hybrid Compression and Message-Sharing Strategy for The Downlink Cloud Radio Access Network. In Proceedings of the 2014 IEEE Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 9–14 February 2014; pp. 1–6.
- Patil, P.; Dai, B.; Yu, W. Performance Comparison of Data-Sharing and Compression Strategies for Cloud Radio Access Networks. In Proceedings of the 2015 23rd European IEEE Signal Processing Conference (EUSIPCO), Nice, France, 31 August–4 September 2015; pp. 2456–2460.
- Dai, B.; Yu, W. Energy Efficiency of Downlink Transmission Strategies for Cloud Radio Access Networks. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 1037–1050.
- Integrated Device Technology, Inc. *Front-Haul Compression for Emerging C-RAN and Small Cell Networks*; IDT Integrated Device Technology, Inc.: San Jose, CA, USA, 2013.
- Peng, M.; Wang, C.; Lau, V.; Poor, H. Fronthaul-Constrained Cloud Radio Access Networks: Insights and Challenges. *IEEE Wirel. Commun.* **2015**, *22*, 152–160.
- Vu, T.; Nguyen, H.; Quek, T. Adaptive Compression and Joint Detection for Fronthaul Uplinks in Cloud Radio Access Networks. *IEEE Trans. Commun.* **2015**, *63*, 4565–4575.
- Zhou, Y.; Yu, W. Optimized Backhaul Compression for Uplink Cloud Radio Access Network. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1295–1307.
- Park, S.; Simeone, O.; Sahin, O.; Shamai, S. Robust and Efficient Distributed Compression for Cloud Radio Access Networks. *IEEE Trans. Signal Process.* **2013**, *62*, 692–703.
- Park, S.; Simeone, O.; Sahin, O.; Shamai, S. Fronthaul Compression for Cloud Radio Access Networks: Signal Processing Advances Inspired by Network Information Theory. *IEEE Signal Process. Mag.* **2014**, *31*, 69–79.
- Park, S.; Simeone, O.; Sahin, O.; Shamai, S. Joint Precoding and Multivariate Backhaul Compression for the Downlink of Cloud Radio Access Networks. *IEEE Trans. Signal Process.* **2013**, *61*, 5646–5658.
- Gamal, A.; Kim, Y. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.
- Nguyen, K.-G.; Vu, Q.-D.; Juntti, M.; Tran, L.-N. Energy Efficient Precoding C-RAN Downlink with Compression at Fronthaul. *arXiv* **2017**, arXiv:1703.05996v1.
- Feng, D.; Jiang, C.; Lim, G.; Cimini, L.; Feng, G.; Li, G. A Survey of Energy-Efficient Wireless Communications. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 167–178.
- Fehske, A.; Marsch, P.; Fettweis, G. Bit Per Joule Efficiency of Cooperating Base Station in Cellular Networks. In Proceedings of the IEEE Globecom Workshops, Miami, FL, USA, 5–10 December 2010; pp. 1406–1411.

25. Xu, W.; Cui, Y.; Zhang, H.; Li, G.Y.; You, X. Robust Beamforming With Partial Channel State Information for Energy Efficient Networks. *IEEE J. Sel. Areas Commun.* **2015**, *33*, 2920–2935.
26. Park, S.; Simeone, O.; Sahin, O.; Shamai, S. Inter-Cluster Design of Precoding and Fronthaul Compression for Cloud Radio Access Networks. *IEEE Wirel. Commun. Lett.* **2014**, *3*, 369–372.
27. Dinkelbach, W. On Nonlinear Fractional Programming. *Manag. Sci.* **1967**, *13*, 492–498.
28. Peng, M.; Zhang, K.; Jiang, J.; Wang, J.; Wang, W. Energy-Efficient Resource Assignment and Power Allocation in Heterogeneous Cloud Radio Access Networks. *IEEE Trans. Veh. Technol.* **2015**, *64*, 5275–5287.
29. Ng, D.; Lo, S.; Schober, R. Energy-Efficient Resource Allocation in OFDMA Systems with Large Numbers of Base Station Antennas. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 3292–3304.
30. He, S.; Huang, Y.; Jin, S.; Yang, L. Coordinated Beamforming for Energy Efficient Transmission in Multicell Multiuser Systems. *IEEE Trans. Commun.* **2013**, *61*, 4961–4971.
31. Shi, Q.; Razaviyayn, M.; Luo, Z.; He, C. An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for A MIMO Interfering Broadcast Channel. *IEEE Trans. Signal Process.* **2011**, *59*, 4331–4340.
32. Joshi, S.; Weeraddana, P.; Codreanu, M.; Latva-aho, M. Weighted Sum-Rate Maximization for MISO Downlink Cellular Networks via Branch and Bound. *IEEE Trans. Signal Process.* **2012**, *60*, 2090–2095.
33. Tervo, O.; Tran, L.-N.; Juntti, M. Optimal Energy-Efficient Transmit Beamforming for Multi-User MISO Downlink. *IEEE Trans. Signal Process.* **2015**, *63*, 5574–5587.
34. Tuy, H.; Al-Khayyal, F.; Thach, P. Monotonic Optimization: Branch and Cut Methods. In *Essays and Surveys in Global Optimization*; Audet, C., Hansen, P., Savard, G., Eds.; Springer: New York, NY, USA, 2005; pp. 39–78.
35. Christensen, S.; Agarwal, R.; De Carvalho, E.; Cioffi, J. Weighted Sum-Rate Maximization Using Weighted MMSE for MIMO-BC Beamforming Design. *IEEE Trans. Wirel. Commun.* **2008**, *7*, 4792–4799.
36. Grant, M.; Boyd, S.; Ye, Y. CVX: Matlab Software for Disciplined Convex Programming, Version 3.0 Beta. Available online: <http://cvxr.com/cvx> (accessed on 31 March 2017).
37. Deng, W.; Yin, W.; Zhang, Y. *Group Sparse Optimization by Alternating Direction Method*; Technical Report; Rice University: Houston, TX, USA, 2011.
38. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122.
39. Pólik, I.; Terlaky, T. Interior Point Methods for Nonlinear Optimization. In *Nonlinear Optimization*, 1st ed.; Pillo, G., Schoen, F., Eds.; Springer: Berlin, Germany, 2010.
40. Love, D.; Heath, R., Jr.; Lau, V.; Gesbert, D.; Rao, B.; Andrews, M. An Overview of Limited Feedback in Wireless Communication Systems. *IEEE J. Sel. Areas Commun.* **2008**, *26*, 1341–1365.
41. Li, J.; Björnson, E.; Svensson, T.; Eriksson, T.; Debbah, M. Joint Precoding and Load Balancing Optimization for Energy-Efficient Heterogeneous Networks. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 5810–5822.
42. Shi, Y.; Zhang, J.; Letaief, K. Robust Group Sparse Beamforming for Multicast Green Cloud-RAN with Imperfect CSI. *IEEE Trans. Signal Process.* **2015**, *63*, 4647–4659.
43. Shen, C.; Chang, T.; Wang, K.; Qiu, Z.; Chi, C. Distributed Robust Multicell Coordinated Beamforming with Imperfect CSI: An ADMM Approach. *IEEE Trans. Signal Process.* **2012**, *60*, 2988–3003.
44. Wang, J.; Bengtsson, M.; Ottersten, B.; Palomar, D. Robust MIMO Precoding for Several Classes of Channel Uncertainty. *IEEE Trans. Signal Process.* **2013**, *61*, 3056–3070.

