Uplink Achievable Rate Maximization for Reconfigurable Intelligent Surface Aided Millimeter Wave Systems With Resolution-Adaptive ADCs

Yue Xiu, Jun Zhao, Member, IEEE, Ertugrul Basar, Senior Member, IEEE, Marco Di Renzo, Fellow, IEEE, Wei Sun, Student Member, IEEE, Guan Gui, Member, IEEE, and Ning Wei, Member, IEEE

Abstract—In this letter, we investigate the uplink of a reconfigurable intelligent surface (RIS)-aided millimeter-wave (mmWave) multi-user system. In the considered system, however, problems with hardware cost and power consumption arise when massive antenna arrays coupled with power-demanding analog-to-digital converters (ADCs) are employed. To account for practical hardware complexity, we consider that the access point (AP) is equipped with resolution-adaptive analog-to-digital converters (RADCs). We maximize the achievable rate under hardware constraints by jointly optimizing the ADC quantization bits, the RIS phase shifts, and the beam selection matrix. The formulated problem is non-convex. To efficiently tackle this problem, a block coordinated descent (BCD)-based algorithm is proposed. Simulations demonstrate that an RIS can mitigate the hardware loss due to the use of RADCs, and that the proposed BCD-based algorithm outperforms state-of-the-art algorithms.

Index Terms—Reconfigurable intelligent surface, millimeter-wave communication, resolution-adaptive analog-to-digital converter, block coordinated descent algorithm.

I. INTRODUCTION

MILLIMETER-WAVE (mmWave) communication systems play an important role in fifth generation (5G) wireless networks. MMW communication systems can offer a higher transmission capacity compared with their microwave counterpart. However, they are impaired by blockages, which affect their reliability, especially in urban environments characterized by the presence of large and densely deployed buildings [1], [2].

To enhance the reliability of mmWave communication systems, several solutions can be employed [2]. Recently, the emerging technology of reconfigurable intelligent surfaces (RISs) has been proposed for enhancing the system performance, especially at high frequency bands [1]. In particular, by appropriately co-phasing the incident signals, RISs provide high beamforming gains that increase the coverage of mmWave communication systems, without the need of using power amplifiers and multiple radio frequency (RF) chains that are, on the other hand, needed if relays are employed [3]–[5], [18], [19]. Considering these advantages, RIS-aided communication systems have been investigated in several works. In [6], the authors proposed a projected gradient method (PGM) for maximizing the achievable rate of RIS-aided multiple-input multiple-output (MIMO) systems. In [7], the authors analyzed the outage probability of RIS-aided non-orthogonal multiple access (NOMA) systems. In [8], the authors studied the physical layer security of RIS-aided systems, and a low-complexity iterative algorithm was proposed. A beamforming design was proposed for RIS-aided simultaneous wireless information and power transfer (SWIPT) systems in [9].

In mmWave communication systems, the prohibitive cost and power consumption of the hardware components at the access points (APs) make the realization of fully-digital solutions very difficult. For these reasons, hybrid analog-digital processing schemes are usually employed to reduce the number of RF chains [8]. In addition, low-resolution analog-to-digital converters (ADCs) are often employed in order to further reduce the hardware cost and power consumption. In particular, a viable solution to find a good trade-off between hardware complexity, power consumption, cost, and performance is to use resolution adaptive ADCs (RADCs) [10]. However, no research work has yet investigated the design and optimization of RIS-aided mmWave communication systems with RADCs. Therefore, there is scope for further research on sum-rate maximization for RIS-aided mmWave systems with RADCs. This letter studies the sum rate optimization problem of the uplink of RIS-aided mmWave systems with RADCs.

To the best of our knowledge, no research work has yet investigated the design and optimization of RIS-aided mmWave communication systems with RADCs. Therefore, it is meaningful to study the impact of using RADCs for application to RIS-aided uplink mmWave communication systems. Specifically, we consider the problem of maximizing the achievable rate by jointly optimizing the beam selection matrix at the AP, the RIS phase shifts, and the ADC quantization bits at the AP. The resulting optimization problem is non-convex, thus, it is difficult to solve. To circumvent this issue, we propose a block coordinated descent (BCD)-based algorithm.
where $\mathcal{F}(\cdot)$ denotes the quantization operator and $F_\alpha = \alpha I_M \in \mathbb{C}^{M \times M}$, where $I_M$ is an $M \times M$ identity matrix. $\alpha = \frac{2\sqrt{3}}{4} b$ denotes the normalized quantization error for $b$ quantization bits when $b > 5$ [14]. When $b \leq 5$, $\alpha$ can be set as shown in [15, Table I]. Also, $n_q$ denotes the quantized noise, whose mean and covariance matrix are 0 and $A_u = F_\alpha F_k \text{diag}(F^H \mathbf{G} \mathbf{H}^H F^H + \sigma^2 F^H F)$, $F_b = (1 - \alpha) I_M \in \mathbb{C}^{M \times M}$. According to this model, the signal detected by the $k$th user is given by

$$\hat{s}_k = u_k^H F_\alpha F^H \mathbf{G} \mathbf{H} x + u_k^H F_\alpha F^H n + u_k^H n_q, \quad (4)$$

where $u_k$ denotes the decoding vector of the $k$th user, and $x$ is the data stream. For convenience, we define $w = \text{vec}(W) \in \mathbb{C}^{N_{RF} \times 1}$ and $u = [u_1^T, \ldots, u_K^T]^T \in \mathbb{C}^{MK \times 1}$. Then, the achievable rate of the $k$th user is given in (5), shown at the bottom of the page. To maximize the uplink achievable rate, the problem can be formulated as

$$\max_{\Theta, b, u, w} \sum_{k=1}^{K} R_k \tag{6a}$$

subject to

$$\sum_{s=1}^{S} w_{s,m} = 1, \quad \sum_{m=1}^{M} w_{s,m} \leq 1, \quad (6b)$$

$$|\theta_i| = 1, \quad (6c)$$

$$b_{\min} \leq b \leq b_{\max}, \quad b \text{ is an integer.} \quad (6d)$$

The problem in (5) is non-convex due to the non-convexity of (6b) and (6d). To handle the non-convex discrete constraint (6d), we relax it into a continuous constraint, i.e.,

$$b_{\min} \leq b \leq b_{\max}. \quad (8)$$

The problem in (7) is non-convex due to the non-convexity of (6b) and (6d). To handle the non-convex discrete constraint (6d), we relax it into a continuous constraint, i.e.,

$$b_{\min} \leq b \leq b_{\max}. \quad (8)$$

According to [10], $b$ is rounded as

$$b(\delta) = \begin{cases} \lfloor b^* \rfloor, & \text{if } b^* - \lfloor b^* \rfloor \leq \delta \\ \lceil b^* \rceil, & \text{otherwise} \end{cases} \quad (9)$$

where $0 \leq \delta \leq 1$ is chosen based on [10]. In addition, in order to address the difficulties caused by constraint (6b), an appropriate transformation is required. In particular, (6b) can
be transformed into the following equivalent form.

\[ \mathbf{w}^T e_m = \hat{w}_{s,m}, \sum_{s=1}^{S} \mathbf{w}^T e_m = 1, \quad (10) \]

\[ \mathbf{w}^T e_m (1 - \hat{w}_{s,m}) = 0, \quad (11) \]

\[ \mathbf{w} \succeq 0, \quad \mathbf{w}^T \mathbf{1}_M \leq 1, \quad 0 \leq \hat{w}_{s,m} \leq 1, \quad (12) \]

where \( e_m = I(\cdot, m) \) and \( \mathbf{1}_M = [1, \ldots, 1]^T \in \mathbb{R}^{M \times 1} \). To deal with the bilinear variables, \( \mathbf{w}^T e_m (1 - \hat{w}_{s,m}) = 0 \) can be transformed into the following constraints by the Schur complement [13].

\[ \left[ \begin{array}{c} \mathbf{w}^T e_m \\ \mathbf{r}_{s,m} - \mathbf{r}_{s,m} \\ 1 - \hat{w}_{s,m} \end{array} \right] \succeq 0, \quad (13) \]

and

\[ \mathbf{w}^T e_m (1 - \hat{w}_{s,m}) \leq \mathbf{r}_{s,m}^2, \quad (14) \]

where \( \mathbf{r}_{s,m} \) is an auxiliary variable. To deal with the bilinear function on the left-hand side of (14), the successive convex approximation (SCA) method based on the arithmetic-geometric mean (AGM) inequality is adopted. Accordingly, (14) can be rewritten as

\[ \mathbf{w}^T e_m (1 - \hat{w}_{s,m}) \leq \frac{1}{2} \left( (\mathbf{w}^T e_m \eta_{s,m})^2 + (1 - \hat{w}_{s,m})^2 \right) \leq \mathbf{r}_{s,m}^2, \quad (15) \]

where \( \eta_{s,m} \) is a feasible point. To tighten the upper bound, \( \eta_{s,m} \) is iteratively updated. In particular, at the \( n \)-th iteration, \( \eta_{s,m} \) is expressed as

\[ \eta_{s,m} = \sqrt{1 - \hat{w}_{s,m}^{(n-1)}} / ((\mathbf{w}^T e_m^{(n-1)}) \mathbf{e}_m). \quad (16) \]

However, the constraint in (15) is still non-convex. Then, we use the SCA method to transform (15) into the following convex constraint.

\[ \frac{1}{2} \left( (\mathbf{w}^T e_m \eta_{s,m})^2 + (1 - \hat{w}_{s,m})^2 \right) - \mathbf{r}_{s,m} (\mathbf{r}_{s,m} - \mathbf{r}_{s,m}) \leq 0. \quad (17) \]

Next, by introducing the auxiliary variables \( \omega_k \) and \( \zeta \) to deal with the non-convex objective function in (7a), \( R_k \) is rewritten as

\[ R_k = \log \left( 1 + \frac{\omega_k \mathbf{A}_k}{\omega_k} \right), \quad (18) \]

where

\[ \omega_k = \sum_{i \neq k} |u_k^H \mathbf{F}_i \mathbf{G} \mathbf{h}_i |^2 + \sigma^2 \| u_k^H \mathbf{F}_a \mathbf{F}_a^H \| ^2 + u_k^H \mathbf{A}_k u_k, \quad (19) \]

\[ \mathbf{A}_k = (D^H \mathbf{G} \mathbf{h}_k \mathbf{h}_k^T \otimes (u_k^H \mathbf{F}_a) \mathbf{H}), \quad (20) \]

\[ \log_2((\pi \sqrt{3})/(2\zeta)) - b = 0. \quad (21) \]

To tackle the non-convex constraint in (21), we use the SCA method to transform (21) in the following convex constraints

\[ b + \frac{2\zeta}{\pi \sqrt{3}} \ln \frac{\pi \sqrt{3}}{2\sqrt{2} (\zeta - \zeta)} \leq 0, \quad \log_2 \left( \frac{\pi \sqrt{3}}{2\zeta} \right) - b \geq 0. \quad (22) \]

Due to the coupling between \( \zeta \) and \( \omega \), the objective function in (18) is still intractable. Then, we transform (18) to

\[ R_k = \log(1 + \rho_k), \quad (23) \]

where

\[ \rho_k \leq \zeta |\omega_k \mathbf{A}_k|^2 / \omega_k. \quad (24) \]

**Algorithm 1: SCA-Based Algorithm for Problem (7)**

1. **Initialization**: \( \theta_k^{(0)}, \omega_k^{(0)}, \gamma_k^{(0)} \), \( k \).
2. **Repeat**
3. Update \( \{b(n), \mathbf{w}(n), \mathbf{u}(n), \mathbf{w}_m(n), \mathbf{u}_m(n), \rho_k \}, \gamma_k^{(n)}, \beta_k^{(n)} \) with fixed \( \theta_k^{(n)}, \omega_k^{(n-1)}, \gamma_k^{(n-1)}, \beta_k^{(n-1)} \) by solving (30).
4. Update \( \theta_k(n), \omega_k(n), \gamma_k^{(n)} \) based on (16) and (29).
5. Update \( n = n + 1 \).
6. **Until** Convergence.
7. **Output**: \( \mathbf{w}^*, \beta^* \).

By using a similar line of thought, we have

\[ \begin{bmatrix} \zeta \\ \omega_k \end{bmatrix} \succeq 0, \quad q = |\omega \mathbf{A}_k|^2, \quad (25) \]

\[ \frac{t_k^2}{\omega_k} \geq \rho_k. \quad (26) \]

Similarly, \( q = |\omega \mathbf{A}_k|^2 \) is transformed as

\[ q + |\bar{w} \mathbf{A}_k|^2 - 2\Re(\bar{w} \mathbf{A}_k^H w^H) \leq 0, \quad q - |\omega \mathbf{A}_k|^2 \geq 0. \quad (27) \]

Then, we use the SCA method based on the first-order Taylor expansion to tackle (26). Specifically, the left-hand side of (26) is non-convex with respect to \( t_k \) and \( \omega_k \), and thus it can be tightly bounded from below with its first-order Taylor approximation. In particular, for any fixed points \( (t_k, \omega_k) \), we have

\[ \frac{t_k^2}{\omega_k} \geq \tilde{t}_k + \frac{2\tilde{t}_k}{\omega_k} t_k - \frac{\tilde{t}_k^2}{\omega_k^2} \omega_k \geq \rho_k. \quad (28) \]

By applying the SCA method in [5], we iteratively update \( \tilde{t}_k \) and \( \bar{w}_k \) at the \( n \)-th iteration as

\[ \tilde{t}_k^{(n)} = \tilde{t}_k^{(n-1)}, \quad \bar{w}_k^{(n)} = \bar{w}_k^{(n-1)}. \quad (29) \]

Therefore, the problem in (7) is transformed into the following convex problem

\[ \max_{b, w, r, w_m, \rho_k, q_k, \zeta, t_k} \sum_{k=1}^{K} \log_2(1 + \rho_k) \quad (30a) \]

s.t. \( (10), (11), (12), (13), (14), (17), (22), (25), (27), (28) \). \( (30b) \)

The SCA-based algorithm is given in Algorithm 1.

**B. Optimization of \( u \) and \( \Theta \)**

For given \( b, \mathbf{w}, \Theta \), the problem in (6) is rewritten as

\[ \max_{u_k} \sum_{k=1}^{K} R_k. \quad (31) \]

Since the achievable rate of the \( k \)-th user is only related to the decoding vector of the \( k \)-th user, maximizing \( \sum_{k=1}^{K} R_k \) is equivalent to maximizing \( R_k \) of each user by optimizing the decoding vector \( u_k \) of each user. Therefore, the problem in (31) can be recast as

\[ \max_{u_k} \sum_{k=1}^{K} \frac{u_k^H \mathbf{B}_k u_k}{u_k^H \mathbf{D}_k u_k}, \quad \forall k. \quad (32) \]

where \( \mathbf{B}_k \) and \( \mathbf{D}_k \) are given in (32) and (33) at the bottom of the next page. We use the MM algorithm to solve (32) [11].

Specifically, let \( y = u_k^H \mathbf{D}_k u_k, \) \( h(u_k) = \frac{u_k^H \mathbf{B}_k u_k}{y} \) is jointly
Algorithm 2: MM-Based Algorithm for (31)

1 Initialization: \( u_k^{(0)}, \ t = 0 \).
2 Repeat:
3 Update \( u_k \) (1) based on (43).
4 Update \( t = t + 1 \)
5 Until: Convergence.
6 Output: the solution \( u_k \).

convex in \( u_k \) and \( y \) because \( B_k \) is positive definite. Thanks to the convexity, we have

\[
h(u_k) \geq 2 \text{Re}(u_k^H B_k u_k) - \frac{u_k^H B_k u_k}{(u_k^H D_k u_k)^2} u_k^H D_k u_k + c. \tag{35}
\]

where \( u_k \) is the initial variable value. Based on [11, Lemma 1], a lower bound of \( u_k^H D_k u_k \) is

\[
u_k^H D_k u_k \leq u_k^H \lambda_{\text{max}}(D_k) u_k + 2 \text{Re}(u_k^H (D_k - \lambda_{\text{max}}(D_k) I) u_k), \tag{36}
\]

where \( \lambda_{\text{max}}(D_k) \) is the maximum eigenvalue of matrix \( D_k \). Substituting (36) into (35), we have

\[
\frac{u_k^H B_k u_k}{u_k^H D_k u_k} \geq g(u_k | u_k) + [h(u_k) - g(u_k | u_k)], \tag{37}
\]

where

\[
g(u_k | u_k) = 2 \text{Re}(u_k^H B_k u_k) - \frac{u_k^H B_k u_k}{(u_k^H D_k u_k)^2} u_k^H \lambda_{\text{max}}(D_k) u_k + 2 \text{Re}(u_k^H (D_k - \lambda_{\text{max}}(D_k) I) u_k).
\]

Since \( h(u_k) - g(u_k | u_k) \) is a constant, the decoding vector optimization problem in each iteration of the majorize-minimize (MM) algorithm is equivalent to solving the following problem

\[
\max_{u_k} g(u_k | u_k) = \text{Re}[(v(t)+H u_k - \beta(t) u_k^H u_k], \tag{40}
\]

where

\[
v(t) = \frac{B_k u_k(t)}{(u_k(t))^H D_k u_k} - \frac{(u_k(t))^H B_k u_k(t) [D_k - \lambda_{\text{max}}(D_k) I]}{[(u_k(t))^H D_k u_k(t)]^2} \times \frac{u_k}{(u_k(t))^H D_k u_k(t)}, \tag{41}
\]

\[
\beta(t) = \frac{\lambda_{\text{max}}(D_k) u_k^H B_k u_k}{(u_k^H D_k u_k)^2}. \tag{42}
\]

By checking the first-order optimality condition of (41), we have

\[
v_k^{t+1} = v(t) / \beta(t). \tag{43}
\]

The MM-based algorithm is given in Algorithm 2.

We are then left with the optimization of \( \Theta \). For given \( b \), \( u_k \) and \( W \), (6) is rewritten as

\[
\max_{\Theta} \sum_{k=1}^{K} R_k \tag{44a}
\]

\[s.t. \ (6c). \tag{44b}
\]

The problem in (44) can be solved by using the manifold optimization (MO) algorithm [12]. Finally, the BCD-based algorithm for solving problem (6) is summarized in Algorithm 3.

C. Complexity Analysis

In this section, we compare the computational complexity of the proposed BCD-based algorithm with the following state-of-the-art algorithms:

- Hybrid combining-based scheme with MO (SHC-MO): This is a codebook-based hybrid combination scheme, and the RIS phase shifts are obtained by using the MO algorithm.
- Minimum mean-square error quantization bit allocations with MO (MMSQE-BA-MO): It is a variant of the algorithm in [10]. The MO algorithm is used to obtain the RIS phase shifts.

The total computational complexity of each iteration of the BCD-based algorithm is \( O(SM^{1.5} + K^2.5 + KN_1 + N_1) \). Algorithm 3 has lower computational complexity than the SHC-MO algorithm, which amounts to \( O(M^0 + KN_1 + N_1) \), and the MMSQE-BA-MO algorithm, which amounts to \( O(M^0 + M^2 N + KN_1^2) \). Therefore, our proposed BCD-based algorithm provides a better compromise between computational complexity and performance as confirmed by the numerical results illustrated in the next section.

IV. Numerical Results

As shown in Fig. 1, we consider a single-cell system where \( K = 10 \) users are uniformly distributed in the area (2, 30, 0) m and (2, 90, 0) m. The simulation setup is \( N_l = 64 \), \( N_r = 16 \), \( S = 12 \), \( N_{RF} = 8 \), \( b_{\min} = 1 \), \( b_{\max} = 5 \), \( \sigma^2 = -110 \) dBm.

The location of the AP is (0, 0, 0) m and the location of the RIS is (0, 40, 20) m. The mmWave channels from the AP to the RIS, and from the RIS to the \( k \)th user are expressed as

\[
G = \sqrt{1/\beta_l I_1} \sum_{l=0}^{L_1-1} \alpha_a a_T(N_r, \theta_l) a^T_R(N_r, \varphi_l, \phi_l), \tag{45}
\]

\[
h_k = \sqrt{1/\beta_k L_k} \sum_{l=0}^{L_k-1} \hat{a}_k a_R(N_r, \varphi_k), \tag{46}
\]

where \( \beta \) and \( \hat{\beta}_k \) denote the large-scale fading coefficients, \( \beta \) and \( \hat{\beta}_k \) are generated according to a complex Gaussian

\[
B_k = \sum_{l=1}^{K} F_{\alpha} F^H G \Theta h_l h^H l I \Theta H F F^H + \sigma^2 F_{\alpha} F^H F F^H + A_a \tag{33}
\]

\[
D_k = \sum_{l \neq k} F_{\alpha} F^H G \Theta h_l h^H l I \Theta H F F^H + \sigma^2 F_{\alpha} F^H F F^H + A_a \tag{34}
\]
distribution \([10, 20]\)

\[
\beta = C N(0, 10^{-0.1\kappa}) \\
\hat{\beta}_k = C N(0, 10^{-0.1\kappa_k})
\]  

(47)  

(48)

where \(\kappa = 72 + 29.2 \log_{10} d + \zeta\) and \(\kappa_k = 72 + 29.2 \log_{10} d + \zeta_k\). \(d\) denotes the propagation distance, \(\zeta \sim C N(0, 1)\) and \(\zeta_k \sim C N(0, 1)\) account for the log-normal shadowing \([1]\), \(\alpha\) and \(\alpha_k\) denote the small-scale fading coefficients whose distribution is \(C N(0, 1)\) \([8]\). In addition to the SCH-MO and MMSQE-BA-MO schemes, we compare the following two schemes with Algorithm 3:

- **Power allocation and hybrid combining with MO (PHC-MO):** It is a variant of \([21]\).
- **NO-RIS:** In this scheme, the RIS is not used. However, the quantization bits and the beam selection are optimized by using Algorithm 1 and Algorithm 2.

We investigate the convergence behavior of Algorithm 3 in Fig. 3. We observe that the algorithm converges quickly, and, in general, only a few iterations are needed for ensuring the convergence. This shows that the proposed algorithm has low complexity. Fig. 3(b) exhibits the achievable sum rate (ASR) as a function of the number of reflecting elements of the RIS. We observe that the proposed algorithm achieves the best ASR. When \(N_r = 12\), we observe that the proposed RIS-aided schemes yield the same ASR as the NO-RIS scheme with the high resolution of the ADC. This finding validates the feasibility of using RISs to mitigate the impact of low-resolution ADCs. Moreover, we consider the RIS-aided system without considering the quantization noise of the LDACs. It is not difficult to find that the quantization noise degrades the sum rate of the system, which demonstrates that it is advantageous to consider the impact of the quantization noise for system design.

Fig. 4(a) shows the ASR as a function of the number of RF chains. Compared with all the other baseline schemes, the proposed BCD-based algorithm yields the best ASR. From Fig. 4(a), we observe that the ASR increases with \(N_{RF}\). This is because a large number of RF chains can yield a high signal gain and can mitigate the interference. In addition, when \(N_{RF} = 4\) and \(N_{RF} = 6\), we observe that the proposed scheme and the conventional SCH-MO and MMSQE-BA-MO schemes aided by the RIS outperform the NO-RIS scheme with resolution of the ADCs.

V. CONCLUSION AND FUTURE WORKS

The uplink achievable rate optimization problem for RIS-aided mmWave communications with hardware limitations at the AP was investigated. Specifically, the RIS phase shifts, the beam selection matrix, the decoding vector, and the quantization bits are jointly optimized to maximize the sum rate. To deal with the formulated non-convex problem, a BCD-based algorithm is proposed. Simulation results showed that the proposed algorithm outperforms conventional algorithms in terms of ASR.

REFERENCES


