Multicast Wirelessly Powered Network With Large Number of Antennas via First-Order Method

Shuai Wang, *Student Member, IEEE*, Minghua Xia, *Member, IEEE*, and Yik-Chung Wu, *Senior Member, IEEE*

***Abstract*— To prolong the lifetime of energy constrained devices in Internet of Things, devices can harvest wireless energy from the control signal multicast from the access point. Unfortunately, hampered by the path-loss, the efficiency of such multicast wirelessly powered network is low. While large-scale antennas at access point can be used to improve the efficiency, the beamforming design problem in multicast wirelessly powered network is known to be NP-hard, and the traditional difference of convex programming becomes prohibitively time consuming in large-scale settings. On the other extreme, by using the assumption of infinite number of antennas and applying the law of large numbers, simple beamforming solution is possible. However, when applied to scenarios with finite number of antennas, the performance of such asymptotic solution is far from that of difference of convex programming. To resolve this apparent complexity-performance dilemma, this paper develops an algorithm which reduces the computation time by orders of magnitude, while still guaranteeing the same performance compared with the difference of convex programming. In particular, the proposed algorithm consists of two fast-convergent iterative procedures and is guaranteed to obtain a Karush–Kuhn–Tucker solution. Furthermore, in each iteration, the algorithm only requires the computation of inner products between channel vectors and can be run in parallel for all the users. Thus, the complexity scales linearly with the number of antennas at access point. Finally, numerical results validate the performance and the speed of the proposed scheme.**

***Index Terms*— Wirelessly powered communication network (WPCN), large-scale, first-order method, homogeneous quadratically constrained quadratic programming (QCQP), nonlinear energy harvesting model.**

I. Introduction

By 2020, the number of inter-connected Internet of Things (IoT) devices is expected to exceed 20 billion [1], and collecting information from these massive IoT devices becomes a fundamental task. Unfortunately, with limited sizes, IoT devices such as sensors and tags are usually energy constrained [2]. As the access point needs to coordinate the massive IoT devices by multicasting control signals [3]–[6], these signals provide opportunities for IoT devices to harvest energy [7]–[11], and the harvested energy can prolong the lifetimes of IoT devices and also benefit the information collection procedure [12]–[21]. As a consequence, multicast wirelessly powered communication network (WPCN) is very promising for IoT applications.

Currently, the biggest challenge in multicast WPCNs is the high propagation path-loss during energy transmission [12]. To mitigate the path-loss, a viable solution is to employ beamforming with large-scale antennas at the access point [22]–[25]. But unfortunately, the beamforming design problem in multicast WPCNs is NP-hard (i.e., the optimal solution cannot be found in polynomial time), and a common approach is to apply difference of convex (DC) programming [27]–[32]. However, since DC programming requires the computation of the inverse of Hessian matrices, DC programming becomes extremely time consuming and cannot be used in practice if the number of antennas is in the range of hundreds or more. On the other hand, existing methods assuming infinite number of antennas and exploiting the law of large numbers could lead to simple beamforming designs [23]–[25]. Nonetheless, such beamforming designs [23]–[25] suffer from significant performance loss compared to DC programming when applied in scenarios with large but finite number of antennas.

This paper proposes a method that reduces the computation time by orders of magnitude compared to DC programming while still guaranteeing the same performance. In particular, the beamforming design problem in multicast WPCNs is first transformed into a two-stage optimization problem, representing the uplink and downlink problems, respectively. While the uplink problem is a challenging multi-objective problem that usually has multiple Pareto solutions, it is shown in this paper that there exists a unique Parato solution and this solution can be obtained by simply computing a series of conjugate gradients (CG), with the iteration number of the proposed CG method no larger than the number of users. Based on the obtained solution of the uplink problem, the downlink problem can be transformed into a large-scale homogeneous quadratically constrained quadratic programming (QCQP) problem. By applying an accelerated first-order method to this problem, an iterative algorithm is proposed and it is proved to converge to a Karush-Kuhn-Tucker (KKT) solution, which is the best solution one can obtain in polynomial time for homogeneous

Manuscript received October 9, 2017; revised January 4, 2018 and February 28, 2018; accepted March 5, 2018. Date of publication March 22, 2018; date of current version June 8, 2018. This work was supported by the National Natural Science Foundation of China under Grant 61671488. The associate editor coordinating the review of this paper and approving it for publication was X. Zhou.

S. Wang and Y.-C. Wu are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: swang@eee.hku.hk; ycwu@eee.hku.hk).

M. Xia is with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China (e-mail: xiamingh@mail.sysu.edu.cn).

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

Digital Object Identifier 10.1109/TWC.2018.2816062

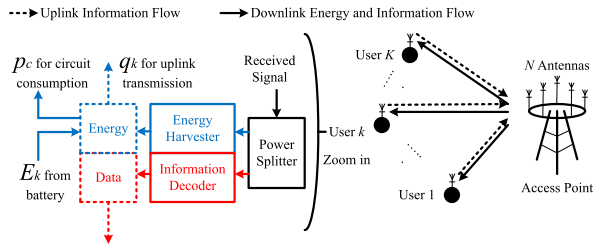


Fig. 1. System model of multicast wirelessly powered network.

QCQP problems. Furthermore, the iteration complexity of the proposed accelerated first-order method is proved to touch the lower bound derived for any first-order method. Therefore, the proposed method is among *the fastest algorithms for solving large-scale homogeneous QCQP problems.* Since both procedures for solving the uplink and downlink problems only require the computation of inner products between channel vectors, the total computational complexity for solving the two-stage problem is linear in terms of the number of antennas at access point. With the proposed algorithms for uplink and downlink problems capable of running in parallel for all the users, its computation time can be further reduced in practice. Finally, numerical results validate the low-complexity nature of the proposed method, and show that the optimization performance of the proposed scheme is equivalent to that of DC programming.

The rest of the paper is organized as follows. In Section II, system model is described and the problem is formulated. In Section III, the problem is solved via a first-order method, and the proposed method is proved to converge to a KKT solution with very low complexity. Connection of the proposed method to the design of hybrid beamformer and time allocation is given in Section IV. Simulation results are presented in Section V, and finally conclusions are drawn in Section VI.

Notation: Italic letters, small bold letters, and capital bold letters represent scalars, vectors, and matrices, respectively. The operators $\mathrm{Tr}(\cdot), (\cdot)^T, (\cdot)^H, \mathrm{Rank}(\cdot), (\cdot)^{-1}$ take the trace, transpose, Hermitian, rank, and inverse of a matrix, respectively. The operator $[x]^+ = \max(x, 0)$. Symbol $\mathbf{I}_N$ represents the $N \times N$ identity matrix. Finally, $\mathbb{E}(\cdot)$ represents the expectation of a random variable, and $\exp(\cdot)$ represents the exponential function of a scalar.

II. System Model and Problem Formulation

A. System Model

In this paper, we consider a multi-user network consisting of an access point with N antennas, and K single-antenna users. As shown in Fig. 1, the access point intends to multicast common information to K users, and the K users intend to send individual information to the access point. The transmission therefore involves two phases, i.e., downlink multicasting phase and uplink multiple access phase. Below, we give the details of each transmission phase.

In the first downlink phase, the access point multicasts a signal s with $\mathbb{E}[|s|^2] = 1$ to all the users through the transmit beamforming vector $\mathbf{v} \in \mathbb{C}^{N\times 1}$ with power $||\mathbf{v}||^2$. Accordingly, the received signal $r_k \in \mathbb{C}$ at the user k is $r_k = \mathbf{g}_k^H \mathbf{v} s + n_k$, where $\mathbf{g}_k^H \in \mathbb{C}^{1\times N}$ is the downlink channel vector from the access point to the k^{th} user, and $n_k \in \mathbb{C}$ is the Gaussian noise at the k^{th} user with power σ_u^2. The received signal r_k is further split into two branches, one for the information decoder and the other for the energy harvester.

At the information decoder side, the signal is given by $\tilde{r}_k = \sqrt{\beta_k}\mathbf{g}_k^H \mathbf{v} s + \sqrt{\beta_k} n_k + z_k$, where $\beta_k \in [0, 1]$ is the splitting factor, and $z_k \in \mathbb{C}$ is Gaussian noise introduced by the power splitter, with $\mathbb{E}[|z_k|^2] = \sigma_z^2$. Based on the expression of $\tilde{r}_k$, the downlink signal to interference plus noise ratio (SINR) of the k^{th} user is

$$\Gamma_k^{\text{DL}} = \frac{\beta_k |\mathbf{g}_k^H \mathbf{v}|^2}{\beta_k \sigma_u^2 + \sigma_z^2}. \tag{1}$$

On the other hand, at the energy harvester of user k, the input power can be expressed as $(1 - \beta_k)\mathbb{E}[|r_k|^2]$. Based on the expression of r_k, the input power can be further expressed as $(1 - \beta_k)|\mathbf{g}_k^H \mathbf{v}|^2$, provided that the noise power σ_u^2 is negligible compared to the signal power of $\mathbf{g}_k^H \mathbf{v} s$. Accordingly, the harvested power is denoted by $\Upsilon\left((1-\beta_k)|\mathbf{g}_k^H \mathbf{v}|^2\right)$, where Υ is the function representing the energy conversion process [33]–[38]. While there are many possible ways to model Υ, we adopt a recent non-linear model [39], [40]:

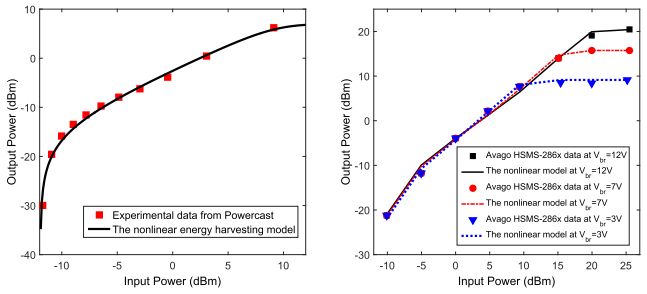
$$\Upsilon(P_{\text{in}}) = \left[\frac{P_{\max}}{\exp(-\tau P_0 + \nu)} \left(\frac{1 + \exp(-\tau P_0 + \nu)}{1 + \exp(-\tau P_{\text{in}} + \nu)} - 1 \right) \right]^+, \tag{2}$$

where the parameter P_0 denotes the harvester's sensitivity threshold and $P_{\max}$ refers to the maximum harvested power when the energy harvesting circuit is saturated [39]. The parameters τ and ν are used to capture the nonlinear dynamics of energy harvesting circuits.

In the second uplink phase, all the users transmit data symbols to the access point simultaneously via spatial division multiple access (SDMA), with the k^{th} user symbol being $x_k \in \mathbb{C}$ with $\mathbb{E}[|x_k|^2] = 1$ and the user transmit power being q_k. The received signal $\mathbf{y} \in \mathbb{C}^{N\times 1}$ at the access point is given by $\mathbf{y} = \sum_{k=1}^K \mathbf{h}_k \sqrt{q_k} x_k + \mathbf{m}$, where $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ is the uplink channel vector, and $\mathbf{m} \in \mathbb{C}^{N\times 1}$ is the Gaussian noise at access point with $\mathbb{E}[\mathbf{m}\mathbf{m}^H] = \sigma_a^2 \mathbf{I}_N$. In order to detect the signal of user k while limiting the array processing complexity, a linear receive beamforming vector $\mathbf{w}_k^H \in \mathbb{C}^{1\times N}$ is applied to $\mathbf{y}$, and we can express the uplink SINR of the k^{th} user as

$$\Gamma_k^{\text{UL}} = \frac{q_k |\mathbf{w}_k^H \mathbf{h}_k|^2}{\sum_{l\neq k} q_l |\mathbf{w}_k^H \mathbf{h}_l|^2 + \sigma_a^2 ||\mathbf{w}_k||^2}. \tag{3}$$

Remark 1: Since the antenna number is large, the downlink channels can be estimated via channel reciprocity. Notice that the wirelessly powered system considered in this paper can be employed in scenarios beyond traditional cellular networks. For example, the wirelessly powered system could be employed in IoT and millimeter wave networks, where the pilot contamination effect can be safely ignored. In particular, in IoT networks, since the coverage of the access point is small [41, Table III], the access points are sparsely distributed and



(a) (b)

Fig. 2. (a) Comparison between the Powercast data and the nonlinear energy harvesting model. The parameters in the model are given by $\tau = 274$, $\nu = 0.29$, $P_{\max} = 0.004927$ W and $P_0 = 0.000064$ W. (b) Comparison between the Avago HSMS-286x data and the nonlinear energy harvesting model. The parameters in the model are given as follows. (i) $V_{\rm br} = 3$V: $\tau = 411, \nu = 2.2, P_{\max} = 0.0082$ W and $P_0 = 0.00008$ W; (ii) $V_{\rm br} = 7$V: $\tau = 116, \nu = 2.3, P_{\max} = 0.0375$ W and $P_0 = 0.00008$ W; (iii) $V_{\rm br} = 12$V: $\tau = 47, \nu = 2.4, P_{\max} = 0.11$ W and $P_0 = 0.00008$ W.

the inter-cell interference is weak. On the other hand, in millimeter wave networks, since broadband frequency spectrum is used, such systems would be operated in noise limited regime according to the measurements in [42, Fig. 1]. In both cases, the pilot contamination effect is not the utmost issue.

Remark 2: To ensure that the model in (2) can adequately represent the energy conversion process, based on least squares criterion, we fit the model to the experimental data from the Powercast energy harvester P2110 [36] at 915 MHz with transmitter-receiver distance ranging from 1 m to 15 m. The output power versus input power is shown in Fig. 2a. It can be observed that with the choice of $\tau = 274$, $\nu = 0.29$, $P_{\max} = 0.004927$ W and $P_0 = 0.000064$ W, the non-linear energy harvesting model in (2) matches the experimental data very well. To further demonstrate the versatility of the model, we also fit the model to the experimental data of Avago HSMS-286x diode, which is obtained from [37, Fig. 6b]. It can be observed from Fig. 2b that the non-linear energy harvesting model in (2) matches the data of this circuit very well for all values of breakdown voltage $V_{\rm br}$.

Remark 3: While time division multiple access (TDMA) may outperform SDMA if there are no local powers at the users [18], the result cannot be directly applied to the case when users have local powers. Moreover, SDMA is an appealing technology for delay sensitive applications and for user fairness provisioning, since SDMA is a simultaneous transmission scheme [18]. As a result, it is important to study SDMA for wirelessly powered networks. Notice that even under the TDMA scheme, the proposed Algorithm 2 in Section III-B can still be applied to design the downlink energy beamformer.

B. Problem Formulation

In the considered network, the design variables that can be controlled are the transmit-receive beamformers $\{\mathbf{v}, \mathbf{w}_k\}$, users' transmit powers $\{q_k\}$, and users' power splitting ratios $\{\beta_k\}$. Since a fundamental quality of service (QoS) requirement in a communication system is the guaranteed SINR [43], [44], our aim is to provide reliable communication for all the users at their requested SINR targets, with the uplink and downlink SINR requirements for the $k^{\rm th}$ user denoted by α_k and θ_k, respectively. On the other hand, since the uplink SINR $\Gamma_k^{\rm UL}$ in (3) depends on the user transmit power q_k, which in part was harvested from the downlink wireless signal, we must have $\frac{1}{2}\Upsilon\Big((1-\beta_k)|\mathbf{g}_k^H\mathbf{v}|^2\Big) + E_k \geq \frac{1}{2}q_k + p_c$, where E_k is the local power per symbol-time available at the $k^{\rm th}$ user in duration T, and p_c is the circuit power consumption per symbol-time at user terminals (the coefficient $1/2$ is due to the two phases of transmission with equal duration).

Having the QoS and energy harvesting requirements satisfied, it is crucial to minimize the total transmit power at access point and users[1] because saving power translates to cost reduction and environmental benefits. As a result, by accounting for all the factors mentioned above, an optimization problem can be formulated as:

$$\begin{aligned}
\mathrm{P1}: \min_{\{\mathbf{v},\mathbf{w}_k,q_k,\beta_k\}} \ & ||\mathbf{v}||^2 + \sum_{k=1}^{K} q_k \\
\text{s.t. } & q_k|\mathbf{w}_k^H\mathbf{h}_k|^2 \geq \alpha_k\Big(\sum_{l\neq k} q_l|\mathbf{w}_k^H\mathbf{h}_l|^2 + \sigma_a^2||\mathbf{w}_k||^2\Big), \\
& \qquad\qquad \forall k = 1,\ldots,K \\
& \beta_k|\mathbf{g}_k^H\mathbf{v}|^2 \geq \theta_k\Big(\beta_k\sigma_u^2 + \sigma_z^2\Big), \quad \forall k = 1,\ldots,K \\
& \frac{1}{2}\Upsilon\Big((1-\beta_k)|\mathbf{g}_k^H\mathbf{v}|^2\Big) + E_k \geq \frac{1}{2}q_k + p_c, \\
& \qquad\qquad \forall k = 1,\ldots,K \\
& q_k \geq 0, \ \beta_k \in [0,1], \quad \forall k = 1,\ldots,K,
\end{aligned}$$

where the first and second constraints are the uplink and downlink SINR QoS requirements, respectively. Unfortunately, problem P1 is non-convex due the coupled terms of $\{q_k, \mathbf{w}_k\}$ and $\{\beta_k, \mathbf{v}\}$, as well as the nonlinear function Υ. In fact, P1 is NP-hard in general since its special case of $\{\alpha_k = 0\}$ is NP-hard according to [3, Claim 1].

Remark 4: Since a large number of antennas at the access point may incur significant antenna circuit power consumption [26], it is also interesting to consider the energy efficiency maximization problem at access point. In such a case, the objective of P1 becomes

$$\max \ \frac{\sum_{k=1}^{K}\log_2\left(1 + \frac{q_k|\mathbf{w}_k^H\mathbf{h}_k|^2}{\sum_{l\neq k} q_l|\mathbf{w}_k^H\mathbf{h}_l|^2 + \sigma_a^2||\mathbf{w}_k||^2}\right)}{||\mathbf{v}||^2 + P_0},$$

where P_0 is the antenna circuit power consumption at access point. By using the Dinkelbach method [26], the above fractional objective function can be transformed into a subtractive form, and the resultant problem can be solved using a similar approach to that of P1.

III. Proposed First-Order Method for P1

To resolve the non-convexity, a traditional way is to reformulate P1 into a DC programming problem (by introducing substitution variables $\chi_k = \frac{1}{q_k}$ to replace q_k) [27]–[31]. However, DC programming needs to solve a convex quadratic problem in each iteration, and each problem has at least

[1]The user transmit powers are minimized to save energy for the battery.

$KN + N + 2K$ variables and K^2 second-order cone (SOC) constraints of dimension 2. Therefore, DC programming requires a complexity of $O(K^4N^3)$ in each iteration [45], and such a method would be extremely slow when N is large. To overcome the drawback of DC programming, a fast first-order method is presented below.

To provide a fast algorithm for solving P1, the first step is to decompose the problem P1 into smaller pieces. To achieve this goal, it is first observed from the third constraint that smaller $(q_1, \ldots, q_K)$ loosens the constraints on $\mathbf{v}$ since Υ is a monotonically increasing function. As smaller $(q_1, \ldots, q_K)$ also helps in reducing the objective value, the optimal $(q_1^*, \ldots, q_K^*)$ must therefore be within the set $\mathcal{Q}$, which is the Pareto optimal set for the following multi-objective uplink problem

$$\mathrm{U}: \min_{\{b_k \geq 0, \mathbf{w}_k\}} \left\{ (b_1, \ldots, b_K) : b_k |\mathbf{w}_k^H \mathbf{h}_k|^2 \geq \alpha_k \left(\sum_{l \neq k} b_l |\mathbf{w}_k^H \mathbf{h}_l|^2 + \sigma_a^2 ||\mathbf{w}_k||^2 \right), \quad \forall k = 1, \ldots, K \right\}.$$

By defining the feasible set of problem U as $\mathcal{Y}$, $\mathcal{Q}$ can be expressed as

$$\mathcal{Q} = \left\{ (b_1, \ldots, b_K) \in \mathcal{Y} : \nexists (x_1, \ldots, x_K) \in \mathcal{Y} \text{ with } (x_1, \ldots, x_K) \prec (b_1, \ldots, b_K) \right\},$$

where $\prec$ represents "Pareto dominate" (for minimization problems) [46]. Based on the above observation, the feasible set of $(q_1, \ldots, q_K)$ can be restricted into the Pareto set $\mathcal{Q}$ and P1 can be equivalently transformed into the following problem:

$$\begin{aligned} \mathrm{P2}: \min_{\{\mathbf{v}, \beta_k \in [0,1], q_k\}} \; & ||\mathbf{v}||^2 + \sum_{k=1}^{K} q_k \\ \text{s.t. } & \beta_k |\mathbf{g}_k^H \mathbf{v}|^2 \geq \theta_k \left(\beta_k \sigma_u^2 + \sigma_z^2 \right), \quad \forall k \\ & \frac{1}{2} \Upsilon \left((1 - \beta_k) |\mathbf{g}_k^H \mathbf{v}|^2 \right) + E_k \geq \frac{1}{2} q_k + p_c, \; \forall k \\ & \left(q_1, \ldots, q_K \right) \in \mathcal{Q}. \end{aligned}$$

By inspection, the problems U and P2 appears to be inseperable. This is because the uplink problem U is a multi-objective problem, which may have many Pareto solutions [46]. Since different Pareto solutions (of the uplink problem) would impose different constraints on downlink beamformer $\mathbf{v}$ (observing from the second constraint of P2) and lead to different objective values of the entire problem P2, we have to choose a solution among all the Pareto solutions of the uplink problem such that the total power of P2 is minimized. In fact, such uplink-downlink coupling is an inevitable issue in wirelessly powered system [17]. But fortunately, we have the following property to decouple U and P2.

Property 1: If $\mathrm{Rank}([\mathbf{h}_1, \ldots, \mathbf{h}_K]) + \sum_{k=1}^{K} \frac{1}{1+\alpha_k} > K$, *then* $|\mathcal{Q}| = 1$.

Proof: If $\mathrm{Rank}([\mathbf{h}_1, \ldots, \mathbf{h}_K]) + \sum_{k=1}^{K} \frac{1}{1+\alpha_k} > K$, we have $\mathcal{Q} \neq \emptyset$ based on [43, Th. III.1] and there must exist some points $\left(q_1, \ldots, q_K \right)$ in $\mathcal{Q}$. Furthermore, according to the fixed point equation in [43, Appendix A], the points $\left(q_1, \ldots, q_K \right)$ in $\mathcal{Q}$ must satisfy the following:

$$\frac{\alpha_k}{\mathbf{h}_k^H \left(\sum_{l \neq k} q_l \mathbf{h}_l \mathbf{h}_l^H + \sigma_a^2 \mathbf{I}_N \right)^{-1} \mathbf{h}_k} - q_k = 0. \tag{4}$$

Since the left hand side of (4) is strictly radially quasi-concave for all k, there is at most one solution to (4) according to [47, Corollary 1]. Therefore, $|\mathcal{Q}| = 1$. ■

Property 1 indicates that there exists a unique Pareto solution for the uplink problem under certain condition. Notice that if the access point is equipped with large number of antennas, we must have $N \gg K$ and $\mathrm{Rank}([\mathbf{h}_1, \ldots, \mathbf{h}_K]) = K$. Together with the fact that α_k is the QoS requirement of the k^{th} user and must be positive, $\mathrm{Rank}([\mathbf{h}_1, \ldots, \mathbf{h}_K]) + \sum_{k=1}^{K} \frac{1}{1+\alpha_k} > K$ is always satisfied, implying that the unique solution is guaranteed. Therefore, we only need to find the unique point in $\mathcal{Q}$ and put it back into P2 for subsequent derivation.

A. Solving the Uplink Problem

To find the unique Pareto point in $\mathcal{Q}$, an outer approximation algorithm can be adopted, which generates a sequence of lower bounds $\{(\gamma_1^{[0]}, \ldots, \gamma_K^{[0]}), (\gamma_1^{[1]}, \ldots, \gamma_K^{[1]}), \ldots\}$ to approach the solution. As a valid lower bound, we can set $(\gamma_1^{[0]}, \ldots, \gamma_K^{[0]}) = \mathbf{0}$. Furthermore, consider the following update of $\{\mathbf{z}_k, \gamma_k\}$ at the n^{th} iteration:

$$\mathbf{z}_k^{[n+1]} = \left(\sum_{l \neq k} \gamma_l^{[n]} \mathbf{h}_l \mathbf{h}_l^H + \sigma_a^2 \mathbf{I}_N \right)^{-1} \mathbf{h}_k, \tag{5}$$

$$\gamma_k^{[n+1]} = \frac{\alpha_k \left(\sum_{l \neq k} \gamma_l^{[n]} |(\mathbf{z}_k^{[n+1]})^H \mathbf{h}_l|^2 + \sigma_a^2 ||\mathbf{z}_k^{[n+1]}||^2 \right)}{|(\mathbf{z}_k^{[n+1]})^H \mathbf{h}_k|^2}, \tag{6}$$

and the following property, which is based on the fixed-point iteration [43], can be established.

Property 2: If $\mathrm{Rank}([\mathbf{h}_1, \ldots, \mathbf{h}_K]) + \sum_{k=1}^{K} \frac{1}{1+\alpha_k} > K$, *then with* $\gamma_k^{[0]} = 0$ *for all* k, *the sequence* $\gamma_k^{[n]}$ *converges to the unique limit point* $(\gamma_1^\diamond, \ldots, \gamma_K^\diamond) \in \mathcal{Q}$.

For the above alternating minimization (AM) procedure (5)-(6), while the first iteration is straightforward (due to $\gamma_k^{[0]} = 0$), the computational complexity for $n \geq 1$ is dominated by the inverse operation of the $N \times N$ matrix $\sum_{l \neq k} \gamma_l^{[n]} \mathbf{h}_l \mathbf{h}_l^H + \sigma_a^2 \mathbf{I}_N$. This leads to a cubic complexity $O(KN^3)$ in terms of the antenna numbers [17], [44], and thus direct implementation of (5) is not suitable for systems with large number of antennas.

To get around the matrix inversion, we will apply the first-order method (i.e., gradient-based method) for the implementation of (5), which has linear complexity in terms of antenna numbers [48]. In particular, we first reformulate (5) as an unconstrained least squares problem

$$\mathbf{z}_k^{[n+1]} = \underset{\mathbf{x}}{\operatorname{argmin}} \underbrace{\mathbf{x}^H \left(\sum_{l \neq k} \gamma_l^{[n]} \mathbf{h}_l \mathbf{h}_l^H + \sigma_a^2 \mathbf{I}_N \right) \mathbf{x} - 2\mathrm{Re}\left(\mathbf{h}_k^H \mathbf{x} \right)}_{:= \Phi^{[n]}(\mathbf{x})}. \tag{7}$$

Then it can be observed from (7) that the Hessian matrix $\Big(\sum_{l\neq k}\gamma_l^{[n]}\mathbf{h}_l\mathbf{h}_l^H+\sigma_a^2\mathbf{I}_N\Big)$ of $\Phi^{[n]}(\mathbf{x})$ is a sum of $(K-1)$ rank-one Hermitian matrices and an identity matrix. Thus the matrix $\Big(\sum_{l\neq k}\gamma_l^{[n]}\mathbf{h}_l\mathbf{h}_l^H+\sigma_a^2\mathbf{I}_N\Big)$ has at most K distinct eigenvalues, and problem (7) can be solved by the conjugate gradient (CG) method within K iterations [48, Th. 5.4]. More specifically, CG method is an iterative algorithm which generates a set of searching directions that are orthogonal to each other in the Euclidean space weighted by $\Big(\sum_{l\neq k}\gamma_l^{[n]}\mathbf{h}_l\mathbf{h}_l^H+\sigma_a^2\mathbf{I}_N\Big)$. By doing so, CG method is guaranteed to converge much faster than the gradient descent method while still maintaining a low per-iteration complexity.

In order to apply CG method to problem (7), the warm-start initialization $\mathbf{d}_0=\mathbf{z}_k^{[n]}$ is used. Then the initial searching direction can be set to $\mathbf{c}_0=\nabla\Phi^{[n]}(\mathbf{d}_0)$ ($\nabla:=\frac{\partial}{\partial\ \mathrm{conj}(\mathbf{x})}$), where $\nabla\Phi^{[n]}(\mathbf{x})=\big(\sum_{l\neq k}\gamma_l^{[n]}\mathbf{h}_l\mathbf{h}_l^H+\sigma_a^2\mathbf{I}_N\big)\mathbf{x}-\mathbf{h}_k$. At the m^{th} iteration, $\mathbf{d}_{m+1}$ and $\mathbf{c}_{m+1}$ are updated as follows [50]:

$$\mathbf{d}_{m+1}=\mathbf{d}_m-\delta_m\mathbf{c}_m, \tag{8}$$

$$\mathbf{c}_{m+1}=\nabla\Phi^{[n]}(\mathbf{d}_{m+1})-\eta_m\mathbf{c}_m. \tag{9}$$

In equation (8), the term δ_m is the line search step-size, and the optimal step-size can be found by exact minimization, which is given by

$$\delta_m=\frac{\Big|\Big|\sum_{l\neq k}\gamma_l^{[n]}\mathbf{h}_l\mathbf{h}_l^H\mathbf{d}_m+\sigma_a^2\mathbf{d}_m-\mathbf{h}_k\Big|\Big|^2}{\sum_{l\neq k}\gamma_l^{[n]}\mathbf{c}_m^H\mathbf{h}_l\mathbf{h}_l^H\mathbf{c}_m+\sigma_a^2\mathbf{c}_m^H\mathbf{c}_m}. \tag{10}$$

On the other hand, in equation (9), the factor η_m is a carefully chosen coefficient such that $\mathbf{c}_i^H\Big(\sum_{l\neq k}\gamma_l^{[n]}\mathbf{h}_l\mathbf{h}_l^H+\sigma_a^2\mathbf{I}_N\Big)\mathbf{c}_j=0$ holds for all $i\neq j$. In order to satisfy such orthogonal property, one possibility is to use Fletcher-Rieves factor [50]:

$$\begin{aligned}\eta_m&=-\frac{||\nabla\Phi^{[n]}(\mathbf{d}_{m+1})||^2}{||\nabla\Phi^{[n]}(\mathbf{d}_m)||^2}\\&=-\frac{\Big|\Big|\sum_{l\neq k}\gamma_l^{[n]}\mathbf{h}_l\mathbf{h}_l^H\mathbf{d}_{m+1}+\sigma_a^2\mathbf{d}_{m+1}-\mathbf{h}_k\Big|\Big|^2}{\Big|\Big|\sum_{l\neq k}\gamma_l^{[n]}\mathbf{h}_l\mathbf{h}_l^H\mathbf{d}_m+\sigma_a^2\mathbf{d}_m-\mathbf{h}_k\Big|\Big|^2}.\end{aligned} \tag{11}$$

Following a similar proof to that of [48, Th. 5.3], we have $\mathbf{c}_i^H\Big(\sum_{l\neq k}\gamma_l^{[n]}\mathbf{h}_l\mathbf{h}_l^H+\sigma_a^2\mathbf{I}_N\Big)\mathbf{c}_j=0$ holds for all $i\neq j$. Using the above orthogonal property and according to [48, Th. 5.1], the sequence $\{\mathbf{d}_0,\mathbf{d}_1,\ldots\}$ is guaranteed to converge and the converged point $\mathbf{d}^\diamond$ is the optimal solution of (7). More importantly, according to [48, Th. 5.4], the number of iterations for CG method to converge must be less than or equal to K. Based on the CG method, the entire procedure to compute the optimal solution of the uplink problem U is summarized in Algorithm 1.

Algorithm 1 Computing the Optimal Solution for the Uplink Problem U

Input $\{\mathbf{h}_k,\alpha_k\},\sigma_a^2$.
Initialize $\{\gamma_k^{[0]}=0,\ \forall k\}$ and set $n=0$.
Update $\mathbf{z}_k^{[1]}=\mathbf{h}_k/\sigma_a^2$ and $\gamma_k^{[1]}=\alpha_k\sigma_a^2/||\mathbf{h}_k||^2$ for all k. Update $n=1$.
Repeat
  For $k=1:K$
    Initialize $\mathbf{d}_0=\mathbf{z}_k^{[n]}$.
    Initialize $\mathbf{c}_0=\big(\sum_{l\neq k}\gamma_l^{[n]}\mathbf{h}_l\mathbf{h}_l^H+\sigma_a^2\mathbf{I}_N\big)\mathbf{z}_0-\mathbf{h}_k$.
    Set $m=0$.
    Repeat
      Update $\mathbf{d}_{m+1}$ using (8).
      Update $\mathbf{c}_{m+1}$ using (9).
      Set $m:=m+1$.
    Until $m=K$ and the converged point is $\mathbf{d}^\diamond=\mathbf{d}_K$. Set $\mathbf{z}_k^{[n+1]}=\mathbf{d}^\diamond$.
  End
  Update $\gamma_k^{[n+1]}$ using (6) for all k.
  Set $n:=n+1$.
Until $\sqrt{\sum_k(\gamma_k^{[n]}-\gamma_k^{[n-1]})^2}<10^{-6}$. The converged point is $\{\gamma_k^\diamond,\mathbf{z}_k^\diamond\}$.
Output $\{q_k^*=\gamma_k^\diamond,\mathbf{w}_k^*=\mathbf{z}_k^\diamond\}$.

B. Solving the Downlink Problem

By substituting the optimal solution of the uplink problem $\{q_k=\gamma_k^\diamond\}$ into P2, the problem P2 is equivalently transformed into

$$\begin{aligned}\mathrm{D}:\ \min_{\{\mathbf{v},\beta_k\in[0,1]\}}\ &||\mathbf{v}||^2\\ \text{s.t. }&|\mathbf{g}_k^H\mathbf{v}|^2\geq\theta_k(\sigma_u^2+\frac{\sigma_z^2}{\beta_k}),\quad\forall k=1,\ldots,K\\&|\mathbf{g}_k^H\mathbf{v}|^2\geq\frac{\mu_k}{1-\beta_k},\quad\forall k=1,\ldots,K,\end{aligned}$$

where $\mu_k=\Upsilon^\dagger(\gamma_k^\diamond+2p_c-2E_k)$ and

$$\Upsilon^\dagger(x)=\begin{cases}+\infty, & \text{if } x\geq P_{\max}\\ \dfrac{\nu}{\tau}-\dfrac{1}{\tau}\ln\Big(\dfrac{1+\exp(-\tau P_0+\nu)}{1+P_{\max}^{-1}\exp(-\tau P_0+\nu)x}-1\Big), & \text{if } 0<x<P_{\max}\\ 0, & \text{if } x\leq 0\end{cases}$$

can be considered as pesudo-inverse of $\Upsilon(x)$.

To solve the downlink problem D, we will first determine the optimal solution of β_k. In particular, if $\theta_k=0$, then the first constraint of D always holds. From the second constraint of D, the feasibility set of $\mathbf{v}$ is maximized by using the minimum value of β_k, which leads to $\beta_k^*=0$. On the other hand, if $\theta_k>0$, we need to divide the discussion into three cases, since $\Upsilon^\dagger(x)$ is a piecewise function consisting of three cases.

(i) If $\gamma_k^\diamond+2p_c-2E_k\geq P_{\max}$, then $\mu_k\to+\infty$ and the problem D is infeasible.

(ii) If $\gamma_k^\diamond+2p_c-2E_k\leq 0$, then $\mu_k=0$ and the second constraint of D always holds. On the other hand, from the first constraint of D, the feasibility set of $\mathbf{v}$ is maximized by using the maximum value of β_k, which leads to $\beta_k^*=1$.

(iii) If $0<\gamma_k^\diamond+2p_c-2E_k<P_{\max}$, then β_k and $\mathbf{v}$ would be involved in both constraints in D. Taking the intersection

of the two inequalities, they can be combined as

$$|\mathbf{g}_k^H\mathbf{v}|^2 \geq \max\Big[\theta_k(\sigma_u^2+\frac{\sigma_z^2}{\beta_k}), \frac{\mu_k}{1-\beta_k}\Big]. \tag{12}$$

Inside the max function of (12), the first term is a decreasing function of β_k while the second term is an increasing function of β_k. Therefore, the minimum of p_k is obtained when

$$\theta_k(\sigma_u^2+\frac{\sigma_z^2}{\beta_k}) = \frac{\mu_k}{1-\beta_k}. \tag{13}$$

Solving (13) for β_k leads to

$$\beta_k^* = \frac{2\sigma_z^2}{\sigma_z^2-\sigma_u^2+\frac{\mu_k}{\theta_k}+\sqrt{\Big(\sigma_z^2-\sigma_u^2+\frac{\mu_k}{\theta_k}\Big)^2+4\sigma_u^2\sigma_z^2}}. \tag{14}$$

Since putting $\mu_k = 0$ into (14) also leads to $\beta_k^* = 1$, we can combine the cases of (ii)-(iii) into one, and the solution of β_k^* for the case of $\theta_k > 0$ is given by (14). Furthermore, by putting β_k^* for the cases of $\theta_k = 0$ and $\theta_k > 0$ into the constraints of D and defining

$$\xi_k = \begin{cases} \mu_k, & \text{if } \theta_k = 0 \\ \frac{\theta_k}{2}\Big(\sigma_z^2+\sigma_u^2+\frac{\mu_k}{\theta_k} \\ \quad +\sqrt{(\sigma_z^2-\sigma_u^2+\frac{\mu_k}{\theta_k})^2+4\sigma_u^2\sigma_z^2}\Big), & \text{if } \theta_k > 0 \end{cases} \tag{15}$$

problem D becomes

$$\begin{aligned} \text{D1}: \ &\min_{\mathbf{v}} \ ||\mathbf{v}||^2 \\ &\text{s.t. } |\mathbf{g}_k^H\mathbf{v}|^2 \geq \xi_k, \quad k=1,\ldots,K. \end{aligned}$$

Now, due to the nonconvexity of the term $|\mathbf{g}_k^H\mathbf{v}|^2$, problem D1 is generally NP-hard [3, Claim 1]. To resolve the nonconvexity, one traditional way is to apply semidefinite relaxation (SDR) [3], [10], which results in a semidefinite programming (SDP) problem with K variables and one semidefinite constraint of dimension $N\times N$. However, the SDR method requires a complexity $O\left(\sqrt{N}(K^3+K^2N^2+KN^3)\right)$ [48], and could only provide a lower bound to the objective function of D1. A slightly better way to resolve the nonconvexity is applying the successive linear approximation (SLA) method [4] to the constraints of D1. This would lead to an iterative algorithm with the number of iterations being M_{SLA}. In each iteration, we need to solve a second order cone programming problem with N variables and K linear constraints. Therefore, the SLA method requires a complexity of $O\left(M_{SLA}\sqrt{K}(N^3+2NK)\right)$, which is still too large for massive MIMO applications.

To overcome the drawbacks of SDR and SLA methods, we propose an accelerated primal-dual gradient (APDG) algorithm for D1, which is a parallel first-order algorithm for solving large-scale homogeneous QCQP problems. In particular, starting from a feasible point $\mathbf{v}^{[0]}$ (one possibility is to choose $\mathbf{v}^{[0]} = \sqrt{p_0}\left(\sum_{l=1}^K\sqrt{\xi_l}\frac{\mathbf{g}_l}{||\mathbf{g}_l||^2}\right)$, where $p_0 = \max_k \frac{\xi_k}{\left|\sum_{l=1}^K\sqrt{\xi_l}\frac{\mathbf{g}_k^H\mathbf{g}_l}{||\mathbf{g}_l||^2}\right|^2}$), consider the following iteration of $\{\varphi_k\}$ and $\mathbf{v}$:

$$\begin{aligned} &\left(\varphi_1^{[n+1]},\ldots,\varphi_K^{[n+1]}\right) \\ &= \underset{\{\lambda_k\geq 0\}}{\text{argmax}} \ \underbrace{\sum_{k=1}^K\lambda_k(\xi_k+|\mathbf{g}_k^H\mathbf{v}^{[n]}|^2) - \Big|\Big|\sum_{k=1}^K\lambda_k\mathbf{g}_k\mathbf{g}_k^H\mathbf{v}^{[n]}\Big|\Big|^2}_{:=\Theta^{[n]}(\{\lambda_1,\ldots,\lambda_K\})}, \end{aligned} \tag{16}$$

$$\mathbf{v}^{[n+1]} = \sum_{k=1}^K\varphi_k^{[n+1]}\mathbf{g}_k\mathbf{g}_k^H\mathbf{v}^{[n]}, \tag{17}$$

and the following theorem (proved in Appendix) can be established.

Theorem 1: With a feasible $\mathbf{v}^{[0]}$*, the sequence* $\{\mathbf{v}^{[0]}, \mathbf{v}^{[1]},\ldots\}$ *is convergent, and the converged point* $\mathbf{v}^\diamond$ *is a Karush-Kuhn-Tucker solution of* D1.

For the above iterative procedure, equations (16) and (17) are the dual and primal updates, respectively. Interestingly, the iteration (16)-(17) can also be interpreted as alternating optimization, where (16) optimizes the combining coefficients and (17) optimizes the beamforming vector.

The goal of applying Theorem 1 is to transform the non-smooth optimization problem D1 (as the constraints of D1 can be viewed as indicator functions) into successive smooth optimization problems (16)-(17). Such a smoothing procedure is very important because the first-order methods converge very slow for non-smooth problems [49]. In contrast, since the gradients of smooth problems are Lipschitz continuous, it is possible to use the Lipschitz condition for acceleration [50].

More specifically, observing that the objective function in (16) is a smooth concave function, we will propose an accelerated projected gradient method to solve problem (16) based on Nesterov's acceleration [50]. In particular, starting from $\{\lambda_k^{[0]}=0\}_{k=1}^K$ (for $n\geq 1$, a warm start of $\{\lambda_k^{[0]}=\varphi_k^{[n]}\}_{k=1}^K$ is applied), the following accelerated projected gradient method can be applied to update $\lambda_k^{[m+1]}$ at the m^{th} iteration [51], [52]:

$$\begin{aligned} &\lambda_k^{[m+1]} \\ &= \Pi_{\mathbb{R}_+}\Big(\rho_k^{[m]}+\frac{1}{L}\nabla_{\lambda_k}\Theta^{[n]}(\{\lambda_1,\ldots,\lambda_K\})\Big|_{\{\lambda_k=\rho_k^{[m]}\}_{k=1}^K}\Big) \\ &= \Big\{\rho_k^{[m]}+\frac{1}{L}\Big(\xi_k+|\mathbf{g}_k^H\mathbf{v}^{[n]}|^2 \\ &\quad -2\sum_{l=1}^K\text{Re}\left[\rho_l^{[m]}(\mathbf{v}^{[n]})^H\mathbf{g}_k\mathbf{g}_k^H\mathbf{g}_l\mathbf{g}_l^H\mathbf{v}^{[n]}\right]\Big)\Big\}^+, \end{aligned} \tag{18}$$

where $\Pi_{\mathbb{R}_+}$ is the projection onto the set $\mathbb{R}_+$ and L is the Lipschitz constant of the gradient $\nabla\Big(||\sum_{k=1}^K\lambda_k\mathbf{g}_k\mathbf{g}_k^H\mathbf{v}^{[n]}||^2\Big)$ [51]. Moreover, $\rho_k^{[m]}$ is a linear combination of $\lambda_k^{[m]}$ and $\lambda_k^{[m-1]}$ as

$$\rho_k^{[m]} = \lambda_k^{[m]}+\frac{c^{[m-1]}-1}{c^{[m]}}(\lambda_k^{[m]}-\lambda_k^{[m-1]}), \tag{19}$$

and $c^{[m]}$ is a particularly tuned parameter satisfying

$$c^{[0]}=1, \quad c^{[m]} = \frac{1+\sqrt{1+4(c^{[m-1]})^2}}{2}.$$

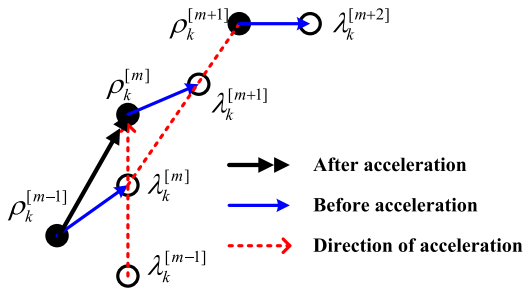


Fig. 3. Illustration of acceleration.

The key insight of the accelerated gradient method is that the traditional gradient method is too conservative, and we need to add some overshoots (the quantity $\lambda_k^{[m]} - \lambda_k^{[m-1]}$ in equation (19)) as shown in Fig. 3 (if we drop the quantity $\lambda_k^{[m]} - \lambda_k^{[m-1]}$, then the accelerated gradient method reduces to the gradient method). However, we also need to guarantee that the overshoots would not be too "confident" such that we miss the optimal point. Therefore, we need to design a monotonically increasing sequence $\{c^{[m]}\}$, which can be viewed as a damping system [53], to represent how much we trust in the overshoots. In particular, at the beginning, $c^{[m]}$ is small and over-damping is used to push the solution points forward. As $c^{[m]}$ becomes larger, under-damping is used to pull the solution points back to the optimal point.

Notice that the step size $1/L$ in (18) is the inverse of L, which can be computed as follows. In particular, by computing the Hessian matrix $\mathbf{G}^{[n]} = \nabla^2\left(||\sum_{k=1}^K \lambda_k \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}||^2\right)$, it can be shown that the $(k,l)^{\text{th}}$ element of $\mathbf{G}^{[n]}$ is equal to $2\text{Re}\left[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{g}_l \mathbf{g}_l^H \mathbf{v}^{[n]}\right]$. Then, due to $\mathbf{G}^{[n]} \preceq \lambda_{\max}(\mathbf{G}^{[n]})\mathbf{I}_K$, where $\lambda_{\max}(\mathbf{G}^{[n]})$ takes the largest eigenvalue of $\mathbf{G}^{[n]}$, the Lipschitz constant of the gradient is $L = \lambda_{\max}(\mathbf{G}^{[n]})$.

With the obtained L, $\lambda_k^{[m]}$ computed using (18) would converge to the optimal solution of (16) with an iteration complexity of [50]:

$$O\left(\sqrt{L\sum_{k=1}^K \left(\lambda_k^\star - \lambda_k^{[0]}\right)^2} \cdot \frac{1}{\sqrt{\epsilon}}\right),$$

where $\sqrt{\sum_k (\lambda_k^\star - \lambda_k^{[0]})^2}$ is the Euclidean distance between the converged point $\{\lambda_k^\star\}$ and the initial point $\{\lambda_k^{[0]}\}$ of (18), $L = \lambda_{\max}(\mathbf{G}^{[n]})$ is the Lipchitz constant of gradients, and ϵ is the target solution accuracy. This iteration complexity is significantly smaller than that of the sub-gradient method and the alternating direction method of multipliers (ADMM) method.[2] In fact, the iteration complexity touches the lower bound derived in [50, Th. 2.1.6] for any first-order method, meaning that the proposed algorithm is among the lowest complexity in the class of first-order methods.

Since solving the problem D is equivalent to solving for $\{\beta_k\}$ using (14) and solving for $\mathbf{v}$ using D1, the entire procedure for computing the solution of problem D using the APDG algorithm is summarized in Algorithm 2.

[2]The iteration complexity of the sub-gradient method is $O(\frac{1}{\epsilon^2})$ [54] and that of a standard ADMM method is $O(\frac{1}{\epsilon})$ [55].

Algorithm 2 Computing a KKT Solution for the Problem D

Input $\{\mathbf{g}_k, \theta_k, \}, \sigma_u, \sigma_z$, and $\{q_k^* = \gamma_k^\diamond\}$.
Compute $\{\beta_k^*\}$ using (14) and $\{\xi_k\}$ using (15).
Initialize $\mathbf{v}^{[0]} = \sqrt{p_0}\left(\sum_{l=1}^K \sqrt{\xi_l}\frac{\mathbf{g}_l}{||\mathbf{g}_l||^2}\right)$ with

$$p_0 = \max_{k=1,\ldots,K} \frac{\xi_k}{\left|\sum_{l=1}^K \sqrt{\xi_l}\frac{\mathbf{g}_k^H \mathbf{g}_l}{||\mathbf{g}_l||^2}\right|^2}.$$

Set $\{\varphi_k^{[0]} = 0\}$ and $n = 0$.
Repeat
  Initialize $\lambda_k^{[1]} = \lambda_k^{[0]} = \varphi_k^{[n]}$ for all k. Compute $|\mathbf{g}_k^H \mathbf{v}^{[n]}|^2$ and $\text{Re}[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{g}_l \mathbf{g}_l^H \mathbf{v}^{[n]}]$.
  Set $c^{[0]} = 1$ and $m = 1$.
  Repeat
    For $k = 1 : K$
      Update $c^{[m]} = \frac{1+\sqrt{1+4(c^{[m-1]})^2}}{2}$.
      Update $\rho_k^{[m]}$ using (19).
      Update $\lambda_k^{[m+1]}$ using (18).
      Set $m := m + 1$.
    End
  Until $|\Theta^{[n]}(\{\lambda_k^{[m]}\}_{k=1}^K) - \Theta^{[n]}(\{\lambda_k^{[m-1]}\}_{k=1}^K)| < 10^{-10}$ and set $\{\varphi_k^{[n+1]} = \lambda_k^{[m]}\}$.
  Set $\mathbf{v}^{[n+1]} = \sum_{k=1}^K \varphi_k^{[n+1]} \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}$.
  Set $n := n + 1$.
Until $\left|||\mathbf{v}^{[n]}||^2 - ||\mathbf{v}^{[n-1]}||^2\right| < 10^{-6}$ and the converged point is $\mathbf{v}^\diamond$.
Output $\mathbf{v} = \mathbf{v}^\diamond, \{\beta_k = \beta_k^*\}$.

C. Overall Algorithm and Complexity Analysis

Since the problem P1 can be decoupled into two sub-problems according to **Property 1**, and each sub-problem can be further solved by Algorithms 1 and 2, respectively, the overall algorithm is to execute Algorithm 1 and Algorithm 2 sequentially. As Algorithm 1 is based on conjugate gradient (CG), and Algorithm 2 is based on accelerated primal-dual gradient (APDG), we called the overall algorithm CG-ADPG. Furthermore, according to **Property 2** and **Theorem 1**, the CG-APDG method is guaranteed to converge to a KKT solution of problem P2. Since we have shown that P1 can be equivalently transformed into P2, the obtained solution is also a KKT solution for P1.

For complexity analysis, it can be seen from Algorithm 1 that the computational complexity is dominated by (8) and (9), which require $O(KN)$ floating-point operations. Therefore, with K iterations for the CG procedure and K users for computation, the total complexity is $O(M_1 K^3 N)$, where M_1 is the number of outer iterations for Algorithm 1 to converge. On the other hand, in each iteration of Algorithm 2, we need to first compute $|\mathbf{g}_k^H \mathbf{v}^{[n]}|^2$ and $\text{Re}[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{g}_l \mathbf{g}_l^H \mathbf{v}^{[n]}]$, which requires $O(N)$ operations for each user. After that we need to perform $O\left(\frac{1}{\sqrt{\epsilon}}\right)$ scalar

TABLE I
SUMMARY OF COMPLEXITY

Scheme	Complexity
CG-APDG	$O\left(M_1K^3N + M_2(KN + K\frac{1}{\sqrt{\epsilon}})\right)$
DC programming	$O\left(M_2K^4N^3\right)$
AM-SDR	$O\left(M_1KN^3 + \sqrt{N}(K^3 + K^2N^2 + KN^3)\right)$
AM-SLA	$O\left(M_1KN^3 + M_2\sqrt{K}(N^3 + 2NK)\right)$

operations[3] to update the beamforming coefficients for each user, where ϵ is the target solution accuracy. Therefore, the complexity of Algorithm 2 is $O\left(M_2(KN + K\frac{1}{\sqrt{\epsilon}})\right)$, where M_2 is the number of outer iterations for Algorithm 2 to converge. Based on the complexities of Algorithms 1 and 2, the proposed CG-APDG method requires a complexity of $O\left(M_1K^3N + M_2(KN + K\frac{1}{\sqrt{\epsilon}})\right)$. Notice that with $\epsilon = 10^{-4}$, the term $\frac{1}{\sqrt{\epsilon}} = 100$ would be in the same order as N.

Based on the above complexity analysis and the complexity results of DC-programming, AM, SDR, and SLA, the complexities of the proposed methods and alternative solutions are summarized in Table I.[4] It can be seen from Table I that if $N = 100$ and $K = 20$, the proposed CG-APDG method requires $O(10^6)$ operations (the complexity is dominated by $O(M_1K^3N)$). On the other hand, the AM-SDR requires $O(10^8)$ operations, and the AM-SLA requires $O(10^7)$ operations. This indicates that the proposed method reduces the computational complexity by at least 90% compared to alternative methods. Moreover, it can be seen from (5)-(6) and (18) that the proposed algorithms for uplink and downlink problems are capable of running in parallel for all the users. Therefore, its computation time can be further reduced in practice.

IV. Practical Considerations

A. Hybrid Beamforming

In massive MIMO systems, a common assumption is that the RF chains are limited compared to the large number of antennas. In such a case, a hybrid beamforming design, which consists of an analog beamformer and a digital beamformer, is required. There are two ways to handle the design of hybrid beamformer (the hybrid receiver can be designed similarly). The first one is to design a traditional beamformer $\mathbf{v}^\diamond$ and then factorize it into the multiplication of $\mathbf{V}^{RF}$ and $\mathbf{v}^{BB}$, where $\mathbf{V}^{RF} \in \mathbb{C}^{N\times N_{RF}}$ is the analog beamformer controlling the phase shifts, N_{RF} is the number of RF chains, and $\mathbf{v}^{BB} \in \mathbb{C}^{N_{RF}\times 1}$ is the digital beamfomer. Mathematically, the above factorization can be written as [57], [58]:

$$\min_{\{\mathbf{V}^{RF},\mathbf{v}^{BB}\}} \ ||\mathbf{v}^\diamond - \mathbf{V}^{RF}\mathbf{v}^{BB}||^2$$
$$\text{s.t. } |\mathbf{V}^{RF}_{i,j}| = 1, \quad \forall i,j, \ ||\mathbf{V}^{RF}\mathbf{v}^{BB}|| = ||\mathbf{v}^\diamond||, \quad (20)$$

where $\mathbf{v}^\diamond$ is the traditional beamformer. In this context, the proposed algorithm represents an efficient design for $\mathbf{v}^\diamond$. Once $\mathbf{v}^\diamond$ is obtained, problem (20) can be solved efficiently by alternating minimization [57], [58].

The second way to design a hybrid beamformer is to apply a two-stage approach [59]. In particular, the first stage designs the analog beamformer $\mathbf{V}^{RF}$ based on beamsteering codebooks [59] or Fourier matrix (which is asymptotically optimal if $N \to \infty$ [59]). With the result of $\mathbf{V}^{RF}$ in the first stage and by replacing $\mathbf{v}$ with $\mathbf{V}^{RF}\mathbf{v}^{BB}$ in problem P1, the second stage problem is given by

$$\min_{\{\mathbf{v}^{BB},\mathbf{w}_k,q_k,\beta_k\}} \ ||\mathbf{V}^{RF}\mathbf{v}^{BB}||^2 + \sum_{k=1}^{K} q_k$$
$$\text{s.t. } \ q_k|\mathbf{w}_k^H\mathbf{h}_k|^2 \geq \alpha_k\Big(\sum_{l\neq k} q_l|\mathbf{w}_k^H\mathbf{h}_l|^2 + \sigma_a^2||\mathbf{w}_k||^2\Big), \ \forall k$$
$$\beta_k|\mathbf{g}_k^H\mathbf{V}^{RF}\mathbf{v}^{BB}|^2 \geq \theta_k\Big(\beta_k\sigma_u^2 + \sigma_z^2\Big), \qquad \forall k$$
$$\frac{1}{2}\Upsilon\Big((1-\beta_k)|\mathbf{g}_k^H\mathbf{V}^{RF}\mathbf{v}^{BB}|^2\Big) + E_k \geq \frac{1}{2}q_k + p_c, \ \forall k$$
$$q_k \geq 0, \quad \beta_k \in [0,1], \quad \forall k.$$

It can be seen that the above problem has exactly the same structure as P1, and we can execute the proposed algorithm in Section III to solve this problem.

In conclusion, no matter which method to employ for hybrid beamforming design, the proposed algorithm in Section III constitutes an indispensable part.

B. Time Allocation

In practice, the users may not need to receive information in the downlink phase. In such a case, users could operate in an idle state and it is beneficial to reduce the duration of uplink transmission. To address this problem, we can introduce a scalar variable $0 \leq z \leq 1$ to represent the proportion of the downlink phase. Moreover, since the downlink and uplink phases are unequal, we need to minimize the weighted average transmit power and the problem P1 is modified as

$$\text{Q}: \min_{\{\mathbf{v},\mathbf{w}_k,q_k,\beta_k\},z} \ z||\mathbf{v}||^2 + (1-z)\sum_{k=1}^{K} q_k$$
$$\text{s.t. } (1-z)\log_2\left(1 + \frac{q_k|\mathbf{w}_k^H\mathbf{h}_k|^2}{\sum_{l\neq k} q_l|\mathbf{w}_k^H\mathbf{h}_l|^2 + \sigma_a^2||\mathbf{w}_k||^2}\right)$$
$$\geq \widetilde{\alpha}_k, \quad \forall k = 1,\ldots,K$$
$$z\log_2\left(1 + \frac{\beta_k|\mathbf{g}_k^H\mathbf{v}|^2}{\beta_k\sigma_u^2 + \sigma_z^2}\right) \geq \widetilde{\theta}_k, \quad \forall k = 1,\ldots,K$$
$$z\Upsilon\Big((1-\beta_k)|\mathbf{g}_k^H\mathbf{v}|^2\Big) + E_k \geq (1-z)q_k$$
$$+ \Big[z\mathbb{1}_{\widetilde{\theta}_k} + (1-z)\Big]p_c, \quad \forall k = 1,\ldots,K$$
$$q_k \geq 0, \quad \beta_k \in [0,1], \ \forall k = 1,\ldots,K, \ 0 \leq z \leq 1.$$

[3]Before $\mathbf{v}^{[n]}$ falls into the convergence region, it is not necessary to accurately solve (16), and the inner loop can be terminated after a fixed number of iterations [56, Sec. II-D]. As a result, the complexity of the APDG method is dominated by the part after $\mathbf{v}^{[n]}$ falls into the convergence region. In such a case, $\sqrt{L\sum_k(\lambda_k^* - \lambda_k^{[0]})^2}$ is a small constant, and the complexity becomes $O\left(\frac{1}{\sqrt{\epsilon}}\right)$.

[4]The number of outer iterations for Algorithm 1 to converge is equal to that for AM [43], i.e., $M_1 = M_{AM}$. On the other hand, the number of outer iterations for Algorithm 2 to converge is equal to that for DC programming and SLA [32], i.e., $M_2 = M_{DC} = M_{SLA}$. As shown by simulations, M_1 and M_2 are much smaller than N.

where $\widetilde{\alpha}_k$ and $\widetilde{\theta}_k$ represent the uplink and downlink data-rate targets, respectively. The constant $\mathbb{1}_{\widetilde{\theta}_k} = 1$ if $\widetilde{\theta}_k > 0$ and $\mathbb{1}_{\widetilde{\theta}_k} = 0$ if $\widetilde{\theta}_k = 0$. This constant $\mathbb{1}_{\widetilde{\theta}_k}$ is used to model the idle state during the downlink phase if users do not need to receive signals.

It can be seen that problem Q has the same structure with problem P1 when the variable z is fixed. Furthermore, since z is a scalar and is bounded by $0 \leq z \leq 1$, a one-dimensional search can be applied to find the optimal z^* [17], with each iteration running Algorithm 1 and Algorithm 2 in Section III.

V. Simulation Results and Discussions

This section presents simulation results to verify the performance of the proposed scheme. The distance-dependent pathloss model of the k^{th} user $\varrho_k = \varrho_0 \cdot (\frac{d_k}{d_0})^{-2.7}$ is adopted [16], where $\varrho_0 = 10^{-3}$, d_k is the distance from the k^{th} user to the access point, and $d_0 = 1\text{m}$ is the reference distance [16]. In the simulations, $d_k \sim \mathcal{U}(1, 10)$ in meter [23], [27], where $\mathcal{U}$ represents the uniform distribution, and $\mathbf{h}_k$ is generated according to $\mathcal{CN}(\mathbf{0}, \varrho_k \mathbf{I}_N)$. Moreover, due to channel reciprocity in time division duplex (TDD) systems,[5] we have $\mathbf{g}_k = \mathbf{h}_k$. The parameters in the energy harvesting model are obtained by fitting the experimental data from the Powercast energy harvester P2110 to the model (2), and they are given by $\tau = 274$, $\nu = 0.29$, $P_{\max} = 0.004927$ W and $P_0 = 0.000064$ W [39]. It is assumed that noise power $\sigma_a^2 = \sigma_u^2 = \sigma_z^2 = -40\text{dBm}$ (corresponding to power spectral density -110 dBm/Hz with 10 MHz bandwidth), which includes thermal noise, intermodulation noise, crosstalk and impulse noise. The circuit power consumption at users is set to $p_c = 5\text{dBm}$; the available energy at users is $E_k \sim \mathcal{U}(2, 8)$ in dBm; and the users' SINR requirements are $\theta_k = \alpha_k = 3\text{dB}$. Each point in the figures is obtained by averaging over 100 simulation runs, with independent channels between consecutive runs. All problem instances are solved by Matlab R2015b on a desktop with Intel Core i5-4570 CPU at 3.2 GHz and 8GB RAM.

A. Convergence Behavior

To verify the convergence of Algorithm 1 in Section III-A, Fig. 4 shows the transmit powers of different users versus number of outer iterations (the number of inner iterations is smaller than or equal to K) when $K = 20$ and $N = 256$. To avoid too much clutter in the figure, we only plot the results for four of the users, as the results for the other users are similar. It can be seen that all the transmit powers converge fast and stabilize after 3 iterations. This result verifies the convergence property in **Property 2** and also indicates that the number of outer iterations M_1 for Algorithm 1 to converge is small.

Next, in order to verify the convergence of the inner loop (i.e., line 8 to line 15 in Algorithm 2) of the proposed APDG method in Section III-B, Fig. 5 shows the function value of

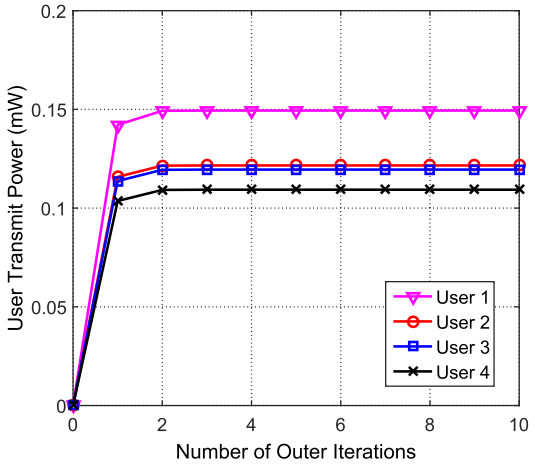


Fig. 4. User transmit powers versus the number of outer iterations in Algorithm 1 for the case of $K = 20$ and $N = 256$.

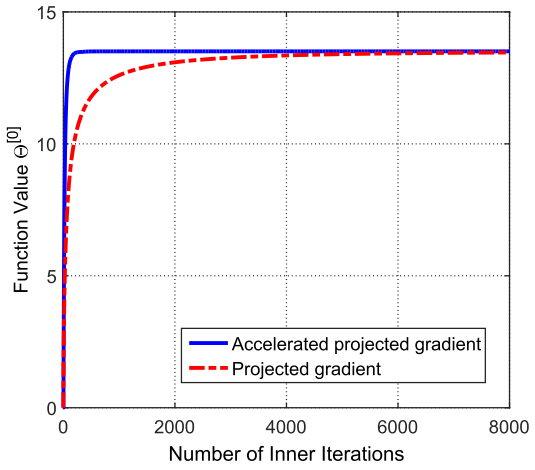


Fig. 5. Function value $\Theta^{[0]}$ of (16) versus number of inner loop iterations in Algorithm 2 when $K = 20$ and $N = 256$.

$\Theta^{[0]}(\{\lambda_k^{[m]}\}_{k=1}^K)$ in (16) versus number of inner iterations $(m = 0, \ldots, 8000)$ when $K = 20$ and $N = 256$. It can be seen that the accelerated projected gradient method and the projected gradient method converge to the same value. However, the accelerated projected gradient method converges significantly faster than the projected gradient method, which reveals the improved convergence rate brought by the acceleration. Notice that since the convergence behavior for $\{\Theta^{[1]}(\{\lambda_k^{[m]}\}_{k=1}^K), \Theta^{[2]}(\{\lambda_k^{[m]}\}_{k=1}^K), \ldots\}$ would be similar, they are not repeated here. On the other hand, to verify the convergence of the outer loop (i.e., line 5 to line 18 in Algorithm 2) of the APDG method in Section III-B, Fig. 6 shows the transmit power at access point versus number of outer iterations $(n = 0, \ldots, 50)$ when $K = 20$ and $N = 256$. As observed from the figure, the proposed algorithm converges after 30 iterations, which corroborates with the convergence result of **Theorem 1** and indicates that the number of outer iterations M_2 for Algorithm 2 to converge is small.

[5]The results for frequency division duplex (FDD) systems, where $\mathbf{g}_k$ and $\mathbf{h}_k$ are independently generated, are similar to that of TDD systems. Therefore, the results for FDD systems are not reported here.

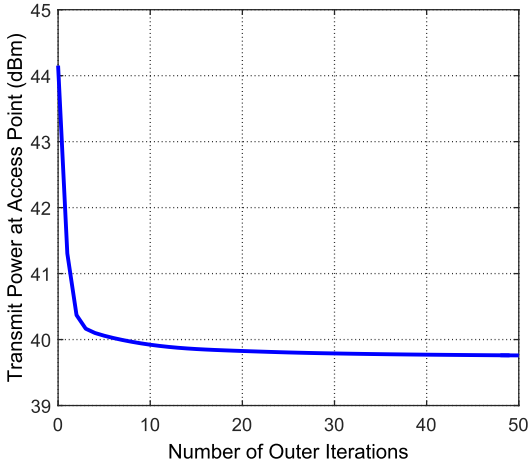


Fig. 6. Transmit power at access point versus number of outer loop iterations in Algorithm 2 when $K = 20$ and $N = 256$.

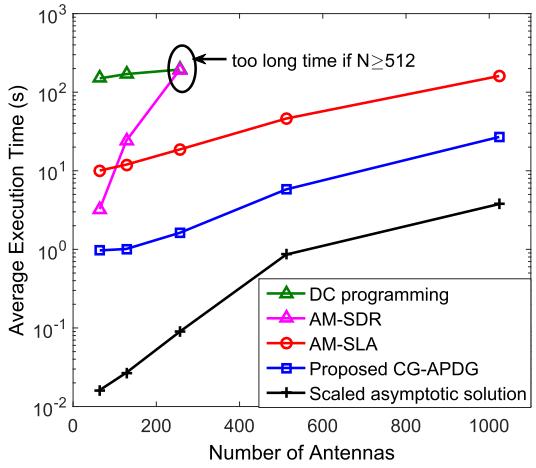


Fig. 8. Average execution time versus number of antennas for the case of $K = 20$.

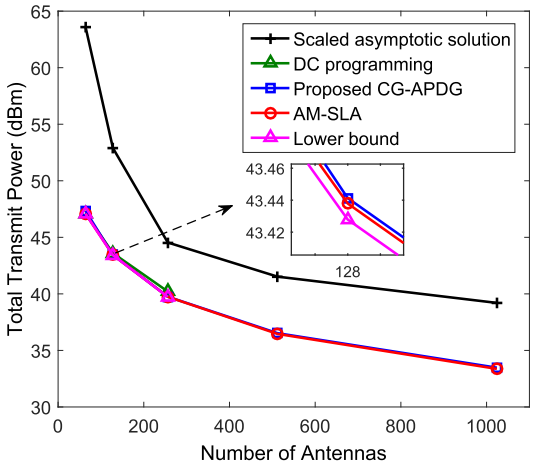


Fig. 7. Total transmit power versus number of antennas for the case of $K = 20$.

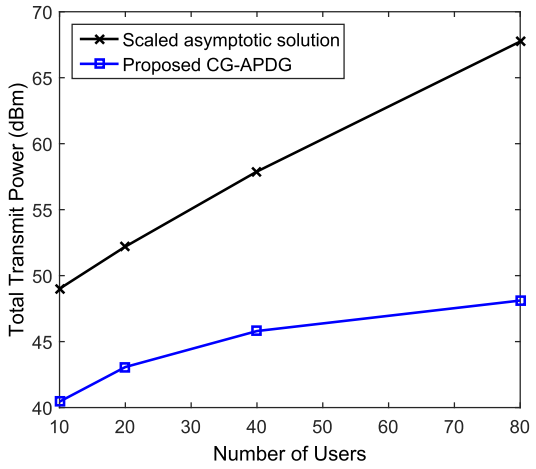


Fig. 9. Total transmit power versus number of users for the case of $N = 256$.

B. Performance and Running Time

Now, in order to verify the performance and complexity of the overall algorithm CG-APDG, we focus on the case of $K = 20$ with $N \in \{64, 128, 256, 512, 1024\}$. Besides the methods in Table I, we also simulate the scaled asymptotic solution.[6] It can be observed from Fig. 7 that the proposed CG-APDG achieves the same transmit power as the AM-SLA and the DC programming method, and they are very close to the lower bound (see the description below problem D1). Moreover, all these algorithms significantly outperform the scaled asymptotic solution. On the other hand, Fig. 8 shows the average execution time versus the number of antennas at access point. Compared to the DC programming method, the proposed CG-APDG algorithm reduces the computation time by orders of magnitude, which demonstrates the effectiveness of the proposed two-stage optimization. Furthermore, compared to AM-SDR and AM-SLA, the CG-APDG algorithm saves at least 90% of their computation times at $N = 128$, which corroborates with the complexity analysis in Section III-C. Under all the simulated values of N, the CG-APDG algorithm saves at least 80% of the computation times of AM-SLA, revealing the low-complexity nature of the CG-APDG algorithm. Notice that for DC-programming and AM-SDR, their simulation times are so large that the results cannot be obtained in a reasonable time when $N \geq 512$.

To study the relationship between the total transmit power and the number of users, we consider the case of $N = 256$ with $K \in \{10, 20, 40, 80\}$. Since CG-APDG, AM-SDR, AM-SLA, and DC programming have the same transmit power, we only simulate the proposed CG-APDG algorithm, and compare it to the scaled asymptotic solution. It can be observed from Fig. 9 that CG-APDG outperforms the scaled asymptotic solution under all the settings of K. Furthermore, the performance gap between CG-APDG and the scaled asymptotic solution grows with the number of users K. This is because larger K results in stronger interference, which departs more from the zero interference assumption in the scaled asymptotic solution.

[6]For scaled asymptotic solution, it is assumed that $N \to \infty$. In such a case, based on the law of large numbers, all the user channels would be asymptotically orthogonal (i.e., $\mathbf{h}_i^H \mathbf{h}_j = 0$ and $\mathbf{g}_i^H \mathbf{g}_j = 0$ for any $i \neq j$). However, since the number of antennas is finite in practice, the obtained solution is not feasible for P1, and the transmit power needs to be scaled up until all the constraints are satisfied. If scaling up the power still cannot make the constraints satisfied, this setting is removed from the simulation.

VI. Conclusions

This paper studied the large-scale beamforming design problem in multicast WPCNs. By reformulating the original problem into a two-stage optimization problem, a fast parallel iterative algorithm was proposed and it was proved to converge to a KKT solution. As each iteration of the proposed algorithm only involves computation of inner products between channel vectors, its complexity scales linearly in terms of the number of antennas at the access point. Simulation results showed that the proposed algorithm reduces the execution time by orders of magnitude compared to the DC programming method while guaranteeing the same performance.

Appendix

Defining the set

$$\mathcal{V}_n = \Big\{\mathbf{x} : \xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2 - 2\mathrm{Re}[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{x}] \le 0, \forall k\Big\},$$

we have the following two lemmas.

Lemma 1: $\mathbf{v}^{[n+1]} = \mathrm{argmin}_{\mathbf{x}\in\mathcal{V}_n} \ ||\mathbf{x}||^2$ *for all* $n \ge 0$.

Proof: Since

$$\Big|\Big|\sum_{k=1}^{K} \lambda_k \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}\Big|\Big|^2 = \sum_{k=1}^{K} \lambda_k \mathrm{Re}[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H (\sum_{l=1}^{K} \lambda_l \mathbf{g}_l \mathbf{g}_l^H \mathbf{v}^{[n]})], \quad (21)$$

the function inside argmax of (16) is

$$\begin{aligned}&\sum_{k=1}^{K} \lambda_k (\xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2) - \Big|\Big|\sum_{k=1}^{K} \lambda_k \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}\Big|\Big|^2 \\ &= \Big|\Big|\sum_{k=1}^{K} \lambda_k \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}\Big|\Big|^2 + \sum_{k=1}^{K} \lambda_k \Big(\xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2 \\ &\quad - 2\mathrm{Re}[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H (\sum_{l=1}^{K} \lambda_l \mathbf{g}_l \mathbf{g}_l^H \mathbf{v}^{[n]})]\Big). \end{aligned} \quad (22)$$

Observing that $\mathbf{x} = \sum_{k=1}^{K} \lambda_k \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}$ is the optimal solution to the problem

$$\min_{\mathbf{x}} \ ||\mathbf{x}||^2 + \sum_{k=1}^{K} \lambda_k \Big(\xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2 - 2\mathrm{Re}[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{x}]\Big),$$

which can be proved by setting the derivative of the above objective function to zero, we have

$$\begin{aligned}&\max_{\{\lambda_k \ge 0\}} \ \sum_{k=1}^{K} \lambda_k (\xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2) - \Big|\Big|\sum_{k=1}^{K} \lambda_k \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}\Big|\Big|^2 \\ &= \max_{\{\lambda_k \ge 0\}} \ \min_{\mathbf{x}} \ ||\mathbf{x}||^2 \\ &\quad + \sum_{k=1}^{K} \lambda_k \Big(\xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2 - 2\mathrm{Re}[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{x}]\Big). \end{aligned} \quad (23)$$

Since the function in (23) is convex in both λ_k and $\mathbf{x}$, strong duality holds, and we can exchange the operators max and min [60]. Further noticing that

$$\begin{aligned}&\max_{\{\lambda_k \ge 0\}} \ ||\mathbf{x}||^2 + \sum_{k=1}^{K} \lambda_k \Big(\xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2 - 2\mathrm{Re}[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{x}]\Big) \\ &= \begin{cases} ||\mathbf{x}||^2, & \text{if } \xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2 - 2\mathrm{Re}[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{x}] \le 0 \\ \infty, & \text{otherwise,} \end{cases} \end{aligned} \quad (24)$$

taking the $\min_{\mathbf{x}}$ on both sides of (24), it is obvious that $\mathbf{x}$ cannot be chosen so that the second case in (24) holds. Therefore, the right hand side of (23) is equal to $\min_{\mathbf{x}\in\mathcal{V}_n} ||\mathbf{x}||^2$, and we have the following holds:

$$\max_{\{\lambda_k \ge 0\}} \ \sum_{k=1}^{K} \lambda_k (\xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2) - \Big|\Big|\sum_{k=1}^{K} \lambda_k \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}\Big|\Big|^2 = \min_{\mathbf{x}\in\mathcal{V}_n} \ ||\mathbf{x}||^2. \quad (25)$$

Now, it can be observed from (25) that $\mathbf{x}$ and $\{\lambda_k\}$ are a pair of primal and dual variables, with the Lagrangian given by

$$\mathcal{L} = ||\mathbf{x}||^2 + \sum_{k=1}^{K} \lambda_k (\xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2 - 2\mathrm{Re}[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{x}]).$$

Therefore, the primal optimal solution $\mathbf{x}^*$ and the dual optimal solution $\{\lambda_k^*\}$ should together satisfy the KKT condition $\partial\mathcal{L}/\partial \ \mathrm{conj(x)} = 0$, which leads to $\mathbf{x}^* = \sum_{k=1}^{K} \lambda_k^* \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}$. Finally, as $\lambda_k^* = \varphi_k^{[n+1]}$ from (16) and $\mathbf{v}^{[n+1]} = \sum_{k=1}^{K} \varphi_k^{[n+1]} \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}$ from (17), we immediately have $\mathbf{v}^{[n+1]} = \mathbf{x}^*$. ■

Lemma 2: $\mathbf{v}^{[n]} \in \mathcal{V}_n$ *for all* $n \ge 0$.

Proof: We prove the lemma by considering two cases.

(i) $\underline{n = 0}$. Since $\mathbf{v}^{[0]}$ is feasible for D1, we have $\xi_k - |\mathbf{g}_k^H \mathbf{v}^{[0]}|^2 \le 0$, which is equivalent to $\xi_k + |\mathbf{g}_k^H \mathbf{v}^{[0]}|^2 - 2\mathrm{Re}[(\mathbf{v}^{[0]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[0]}] \le 0$. This indicates $\mathbf{v}^{[0]} \in \mathcal{V}_0$.

(ii) $\underline{n \ge 1}$. Based on **Lemma 1**, $\mathbf{v}^{[n]} = \mathrm{argmin}_{\mathbf{x}\in\mathcal{V}_{n-1}} \ ||\mathbf{x}||^2$, and from the definition of the set $\mathcal{V}_{n-1}$, we have

$$\xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n-1]}|^2 - 2\mathrm{Re}[(\mathbf{v}^{[n-1]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}] \le 0. \quad (26)$$

Moreover, due to $|\mathbf{g}_k^H (\mathbf{v}^{[n-1]} - \mathbf{v}^{[n]})|^2 \ge 0$, expanding the square gives

$$|\mathbf{g}_k^H \mathbf{v}^{[n-1]}|^2 - 2\mathrm{Re}[(\mathbf{v}^{[n-1]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}] \ge -|\mathbf{g}_k^H \mathbf{v}^{[n]}|^2. \quad (27)$$

Subtracting (27) from (26), we obtain $\xi_k - |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2 \le 0$, which can be rewritten as $\xi_k + |\mathbf{g}_k^H \mathbf{v}^{[n]}|^2 - 2\mathrm{Re}[(\mathbf{v}^{[n]})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^{[n]}] \le 0$. Therefore, $\mathbf{v}^{[n]} \in \mathcal{V}_n$.

Combining (i)-(ii), the proof is completed. ■

Using $\mathbf{v}^{[n+1]} = \mathrm{argmin}_{\mathbf{x}\in\mathcal{V}_n} \ ||\mathbf{x}||^2$ according to **Lemma 1** and $\mathbf{v}^{[n]} \in \mathcal{V}_n$ according to **Lemma 2**, we immediately have $||\mathbf{v}^{[n+1]}|| \le ||\mathbf{v}^{[n]}||$. Therefore, the sequence $\{||\mathbf{v}^{[0]}||, ||\mathbf{v}^{[1]}||, \ldots\}$ is monotonically decreasing. As the norm is lower bounded by zero, the sequence $\{||\mathbf{v}^{[0]}||, ||\mathbf{v}^{[1]}||, \ldots\}$ must converge.

Finally, we will show that the converged point $\mathbf{v}^\diamond$ satisfies the KKT condition of D1. According to **Lemma 1**, $\mathbf{v}^\diamond$ is the optimal solution to

$$\min_{\mathbf{x}} \ \Big\{||\mathbf{x}||^2 : \xi_k + |\mathbf{g}_k^H \mathbf{v}^\diamond|^2 - 2\mathrm{Re}[(\mathbf{v}^\diamond)^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{x}] \le 0, \ \forall k\Big\}. \tag{28}$$

Based on (28), we can obtain two results. Firstly, since $\mathbf{v}^\diamond$ must satisfy the constraint in (28), putting it into the constraint leads to $\xi_k - |\mathbf{g}_k^H \mathbf{v}^\diamond|^2 \le 0$, indicating that $\mathbf{v}^\diamond$ is feasible for D1. Secondly, due to $\mathbf{v}^\diamond$ being the optimal solution of (28), there always exists Lagrange multipliers $\{\zeta_k \ge 0\}$ such that the following two equations hold:

$$\begin{cases} \mathrm{conj}(\mathbf{v}^\diamond) + \sum_{k=1}^{K} \zeta_k \dfrac{\partial}{\partial \mathbf{x}} \Big(\xi_k \\ \quad + |\mathbf{g}_k^H \mathbf{v}^\diamond|^2 - 2\mathrm{Re}[(\mathbf{v}^\diamond)^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{x}]\Big)\Big|_{\mathbf{x}=\mathbf{v}^\diamond} = 0 \\ \zeta_k \Big(\xi_k + |\mathbf{g}_k^H \mathbf{v}^\diamond|^2 - 2\mathrm{Re}[(\mathbf{v}^\diamond)^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^\diamond]\Big) = 0, \quad \forall k. \end{cases} \tag{29}$$

By simple calculations, it can be proved that

$$\frac{\partial}{\partial \mathbf{x}} \Big(\xi_k + |\mathbf{g}_k^H \mathbf{v}^\diamond|^2 - 2\mathrm{Re}[(\mathbf{v}^\diamond)^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{x}]\Big)\Big|_{\mathbf{x}=\mathbf{v}^\diamond} = \frac{\partial}{\partial \mathbf{x}} \Big(\xi_k - |\mathbf{g}_k^H \mathbf{x}|^2\Big)\Big|_{\mathbf{x}=\mathbf{v}^\diamond}$$

and

$$\zeta_k \Big(\xi_k + |\mathbf{g}_k^H \mathbf{v}^\diamond|^2 - 2\mathrm{Re}[(\mathbf{v}^\diamond)^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}^\diamond]\Big) = \zeta_k (\xi_k - |\mathbf{g}_k^H \mathbf{v}^\diamond|^2).$$

Therefore, (29) becomes

$$\begin{cases} \mathrm{conj}(\mathbf{v}^\diamond) + \sum_{k=1}^{K} \zeta_k \dfrac{\partial}{\partial \mathbf{x}} \Big(\xi_k - |\mathbf{g}_k^H \mathbf{x}|^2\Big)\Big|_{\mathbf{x}=\mathbf{v}^\diamond} = 0 \\ \zeta_k \Big(\xi_k - |\mathbf{g}_k^H \mathbf{v}^\diamond|^2\Big) = 0, \quad \forall k. \end{cases} \tag{30}$$

Based on (30) and since we have shown that $\mathbf{v}^\diamond$ is feasible for D1, $\mathbf{v}^\diamond$ is a KKT solution of D1.

REFERENCES

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generat. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, 2013.

[2] S. Timotheou, I. Krikidis, G. Zheng, and B. Ottersten, "Beamforming for MISO interference channels with QoS and RF energy transfer," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2646–2658, May 2014.

[3] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.

[4] L.-N. Tran, M. F. Hanif, and M. Juntti, "A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 114–117, Jan. 2014.

[5] M. R. A. Khandaker and K.-K. Wong, "SWIPT in MISO multicasting systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 3, pp. 277–280, Jun. 2014.

[6] D. W. K. Ng, R. Schober, and H. Alnuweiri, "Secure layered transmission in multicast systems with wireless information and power transfer," in *Proc. IEEE ICC*, Jun. 2014, pp. 5389–5395.

[7] M. Xia and S. Aissa, "On the efficiency of far-field wireless power transfer," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2835–2847, Jun. 2015.

[8] K. Huang and X. Zhou, "Cutting the last wires for mobile communications by microwave power transfer," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 86–93, Jun. 2015.

[9] J. Xu, L. Liu, and R. Zhang, "Multiuser MISO beamforming for simultaneous wireless information and power transfer," *IEEE Trans. Signal Process.*, vol. 62, no. 18, pp. 4798–4810, Sep. 2014.

[10] Q. Shi, L. Liu, W. Xu, and R. Zhang, "Joint transmit beamforming and receive power splitting for MISO SWIPT systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3269–3280, Jun. 2014.

[11] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Relaying Protocols for wireless energy harvesting and information processing," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3622–3636, Jul. 2013.

[12] H. Ju and R. Zhang, "Throughput maximization in wireless powered communication networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 418–428, Jan. 2014.

[13] Q. Wu, W. Chen, D. W. K. Ng, J. Li, and R. Schober, "User-centric energy efficiency maximization for wireless powered communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6898–6912, Oct. 2016.

[14] H. Xing, K.-K. Wong, Z. Chu, and A. Nallanathan, "To harvest and jam: A paradigm of self-sustaining friendly jammers for secure AF relaying," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6616–6631, Dec. 2015.

[15] N. Zlatanov, D. W. K. Ng, and R. Schober, "Capacity of the two-hop relay channel with wireless energy transfer from relay to source and energy transmission cost," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 647–662, Jan. 2017.

[16] S. Wang, M. Xia, and Y.-C. Wu, "Quality of service constrained wirelessly powered communication with multiple antennas," in *Proc. IEEE GLOBECOM Workshop WEHCN*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[17] L. Liu, R. Zhang, and K.-C. Chua, "Multi-antenna wireless powered communication with energy beamforming," *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4349–4361, Dec. 2014.

[18] Q. Wu, W. Chen, D. W. K. Ng, and R. Schober, "Spectral and energy efficient wireless powered IoT networks: NOMA or TDMA?" *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2018.2799947.

[19] Z. Ding *et al.*, "Application of smart antenna technologies in simultaneous wireless information and power transfer," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 86–93, Apr. 2015.

[20] S. Bi, Y. Zeng, and R. Zhang, "Wireless powered communication networks: An overview," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 10–18, Apr. 2016.

[21] Q. Wu, G. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable green 5G networks," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 72–80, Aug. 2017.

[22] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[23] G. Yang, C. K. Ho, R. Zhang, and Y. L. Guan, "Throughput optimization for massive MIMO systems powered by wireless energy transfer," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 8, pp. 1640–1650, Aug. 2015.

[24] Y. Zhu, L. Wang, K.-K. Wong, S. Jin, and Z. Zheng, "Wireless power transfer in massive MIMO-aided HetNets with user association," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4181–4195, Oct. 2016.

[25] X. Chen, J. Chen, and T. Liu, "Secure transmission in wireless powered massive MIMO relaying systems: Performance analysis and optimization," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8025–8035, Dec. 2016.

[26] Q. Wu, M. Tao, D. W. K. Ng, W. Chen, and R. Schober, "Energy-efficient resource allocation for wireless powered communication networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2312–2327, Mar. 2016.

[27] S. Wang, M. Xia, and Y.-C. Wu, "Multipair two-way relay network with harvest-then-transmit users: Resolving pairwise uplink-downlink coupling," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 8, pp. 1506–1521, Dec. 2016.

[28] D. W. K. Ng and R. Schober, "Secure and green SWIPT in distributed antenna networks with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5082–5097, Sep. 2015.

[29] G. Zheng, "Joint beamforming optimization and power control for full-duplex MIMO two-way relay channel," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 555–566, Feb. 2015.

[30] A. A. Nasir, D. T. Ngo, X. Zhou, R. A. Kennedy, and S. Durrani, "Joint resource optimization for multicell networks with wireless energy harvesting relays," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6168–6183, Aug. 2016.

[31] A. A. Nasir, H. D. Tuan, D. T. Ngo, T. Q. Duong, and H. V. Poor, "Beamforming design for wireless information and power transfer systems: Receive power-splitting versus transmit time-switching," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 876–889, Feb. 2017.

[32] L. T. H. An and P. D. Tao, "The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems," *Ann. Oper. Res.*, vol. 133, nos. 1–4, pp. 23–46, 2005.

[33] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, "Practical non-linear energy harvesting model and resource allocation for SWIPT systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2082–2085, Dec. 2015.

[34] B. Clerckx and E. Bayguzina, "Waveform design for wireless power transfer," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6313–6328, Dec. 2016.

[35] Y. Huang and B. Clerckx, "Large-scale multiantenna multisine wireless power transfer," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5812–5827, Nov. 2017.

[36] (2016). *Powercast Wireless Power Calculator (Version 1.5)*. [Online]. Available: http://www.powercastco.com/ power-calculator/

[37] C. R. Valenta and G. D. Durgin, "Harvesting wireless power: Survey of energy-harvester conversion efficiency in far-field, wireless power transfer systems," *IEEE Microw. Mag.*, vol. 15, no. 4, pp. 108–120, Jun. 2014.

[38] S. D. Assimonis, S.-N. Daskalakis, and A. Bletsas, "Sensitive and efficient RF harvesting supply for batteryless backscatter sensor networks," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 4, pp. 1327–1338, Apr. 2016.

[39] S. Wang, M. Xia, K. Huang, and Y.-C. Wu, "Wirelessly powered two-way communication with nonlinear energy harvesting model: Rate regions under fixed and mobile relay," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8190–8204, Dec. 2017.

[40] S. Wang, M. Xia, and Y.-C. Wu, "Space-time signal optimization for SWIPT: Linear versus nonlinear energy harvesting model," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 408–411, Feb. 2018.

[41] S. Kashyap, E. Björnson, and E. G. Larsson, "On the feasibility of wireless energy transfer using massive antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3466–3480, May 2016.

[42] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[43] R. Hunger and M. Joham, "A complete description of the QoS feasibility region in the vector broadcast channel," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3870–3878, Jul. 2010.

[44] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[45] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization* (MPS/SIAM Series on Optimizations). Philadelphia, PA, USA: SIAM, 2013.

[46] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidisciplinary Optim.*, vol. 26, no. 6, pp. 369–395, Apr. 2004.

[47] J. Kennan, "Uniqueness of positive fixed points for increasing concave functions on R^n: An elementary result," *Rev. Econ. Dyn*, vol. 4, no. 4, pp. 893–899, Oct. 2001.

[48] J. Nocedal and S. J. Wright, *Numerical Optimal*, 2nd ed. New York, NY, USA: Springer-Verlag, 2006.

[49] S. Bubeck, "Convex optimization: Algorithms and complexity," *Found. Trends Mach. Learn.*, vol. 8, nos. 3–4, pp. 231–357, 2015.

[50] Y. Nesterov, *Introductory Lectures Convex Optimization: A Basic Course* (Applied Optimization). Springer, 2004.

[51] Y. Nesterov, "A method for solving the convex programming problem with convergence rate $O(1/k^2)$" *Dokl. Akad. Nauk SSSR*, vol. 269, pp. 543–547, 1983.

[52] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[53] W. Su, S. Boyd, and E. Candes, "A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 2510–2518.

[54] A. Konar and N. D. Sidiropoulos, "Fast approximation algorithms for a class of non-convex QCQP problems using first-order methods," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3494–3509, Jul. 2017.

[55] E. Chen and M. Tao, "ADMM-based fast algorithm for multi-group multicast beamforming in large-scale wireless systems," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2685–2698, Jun. 2017.

[56] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[57] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[58] R. Rajashekar and L. Hanzo, "Iterative matrix decomposition aided block diagonalization for mm-wave multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1372–1384, Mar. 2017.

[59] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[60] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

Shuai Wang (S'16) received the B.S. and M.S. degrees in electronic engineering from the Beijing University of Posts and Telecommunications in 2011 and 2014, respectively. He is currently pursuing the Ph.D. degree in electrical and electronic engineering with The University of Hong Kong. His research interests include signal processing and wireless communications.

Minghua Xia (M'12) received the Ph.D. degree in telecommunications and information systems from Sun Yat-sen University, Guangzhou, China, in 2007. He has been a Professor with Sun Yat-sen University since 2015.

From 2007 to 2009, he was with the Electronics and Telecommunications Research Institute (ETRI) of South Korea, Beijing R&D Center, Beijing, China, where he was a Member and then a Senior Member of Engineering Staff and participated in the projects on the physical layer design of 3GPP LTE mobile communications. From 2010 to 2014, he was in sequence with The University of Hong Kong, Hong Kong, the King Abdullah University of Science and Technology, Jeddah, Saudi Arabia, and the Institut National de la Recherche Scientifique, University of Quebec, Montreal, Canada, as a Postdoctoral Fellow. He holds two patents granted in China. His research interests are in the general area of 5G wireless communications, and in particular the design and performance analysis of multi-antenna systems, cooperative relaying systems, and cognitive relaying networks, and recently focus on the design and analysis of wireless power transfer and/or energy harvesting systems, and massive MIMO and small cells.

Dr. Xia received the Professional Award at the IEEE TENCON, Macau, in 2015. He also received the Exemplary Reviewer Award from the IEEE TRANSACTIONS ON COMMUNICATIONS in 2014, the IEEE COMMUNICATIONS LETTERS in 2014, and the IEEE WIRELESS COMMUNICATIONS LETTERS in 2014 and 2015, respectively.

Yik-Chung Wu (S'99–M'05–SM'14) received the B.Eng. (EEE) degree and the M.Phil. degree from The University of Hong Kong (HKU), in 1998 and 2001, respectively, and the Ph.D. degree from Texas A&M University, College Station, TX, USA, in 2005. From 2005 to 2006, he was with the Thomson Corporate Research, Princeton, NJ, USA, as a Member of Technical Staff. Since 2006, he has been with HKU, where he is currently an Associate Professor. He was a Visiting Scholar at Princeton University, in 2011 and 2015. His research interests are in general areas of signal processing, machine learning, and communication systems, and in particular distributed signal processing and robust optimization theories with applications to communication systems and smart grid. He served as an Editor for the IEEE COMMUNICATIONS LETTERS and the IEEE TRANSACTIONS ON COMMUNICATIONS. He is currently an Editor of the *Journal of Communications and Networks*.