Feng Wang

School of Information Engineering, Guangdong University of Technology, Guangzhou, China Email: fengwang13@gdut.edu.cn

Abstract—Integrating mobile edge computing (MEC) and wireless power transfer (WPT) has been regarded as a promising technique to improve computation capabilities for self-sustainable Internet of Things (IoT) devices. This paper investigates a wireless powered multiuser MEC system, where a multi-antenna access point (AP) (integrated with an MEC server) broadcasts wireless power to charge multiple users for mobile computing. We consider a time-division multiple access (TDMA) protocol for multiuser computation offloading. Under this setup, we aim to maximize the weighted sum of the computation rates (in terms of the number of computation bits) across all the users, by jointly optimizing the energy transmit beamformer at the AP, the task partition for the users (for local computing and offloading, respectively), and the time allocation among the users. We derive the optimal solution in a semi-closed form via convex optimization techniques. Numerical results show the merit of the proposed design over alternative benchmark schemes.

Index Terms—Mobile edge computing (MEC), wireless power transfer, computation offloading, energy beamforming.

I. INTRODUCTION

As an enabling technique to provide cloud-like computing for various low-latency and computation-extensive Internet of Things (IoT) applications such as augmented reality and autonomous driving, mobile edge computing (MEC) has received growing attentions from both industry and academia [1]–[12]. At the edge of radio access networks such as access points (APs), MEC servers are deployed therein. The IoT devices are generally of small size and low power. Depending on the computation task is partitionable or not, the resource-limited IoT devices can offload part or all of their computation tasks to the APs, respectively; then the installed MEC servers can execute the offloaded tasks on behalf of these devices.

On the other hand, radio-frequency (RF) signal based wireless power transfer (WPT) provides a viable solution for powering self-sustainable IoT devices by deploying dedicated energy transmitters to broadcast energy wirelessly [13]–[15]. Wireless powered communication networks (WPCNs) and simultaneous wireless information and power transfer (SWIPT) paradigms can achieve ubiquitous wireless communications in a self-sustainable way. It is also expected that WPT can facilitate self-sustainable ubiquitous computing for IoT devices.

Note that the prior works in [9], [10] investigated the wireless powered MEC systems with one or more wireless IoT devices in a self-sustainable way. Specifically, in [9] the single

user maximizes the probability of successful computation with binary offloading, where each task is not partitionable but offloaded as a whole or locally computed by user itself. Assuming that each task is partitionable and a time-division multiple access (TDMA) protocol is adopted for the partial offloading, [10] minimized the total energy consumption for the AP by jointly optimizing the energy transmit beamforming, the task partition and the central processing unit (CPU) frequency per user, as well as the TDMA based time allocations for computation offloading across the users.

Different from the above works, this paper considers a wireless powered MEC system with *limited* resources. The AP employs energy transmit beamforming to simultaneously charge multiple users for mobile computing. The downlink WPT and the computation offloading are operated over orthogonal frequency bands. Suppose that the partial offloading is allowed for each user. As in [10], a TDMA protocol is employed to coordinate multiuser computation offloading. Under this setup, we aim to maximize the weighted sum of computation rates (in terms of the number of computation bits over a particular time block) across all the users subject to the limited MEC computation capacity and the AP transmit power budget constraints. We jointly optimize the energy transmit beamformer at the AP, the task partition for local computing and offloading for each user, as well as the time allocation among the users. Using the Lagrange dual method, we derive the optimal solution in a semi-closed form. Numerical results show the merit of the proposed joint design over alternative benchmark schemes.

The remainder of the paper is organized as follows. Section II presents the system model and formulates the weighted sum of computation rates maximization problem of our interest. Section III develops an efficient algorithm to obtain the optimal solution in a semi-closed form. Section IV provides numerical results, followed by the conclusion in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a wireless powered multiuser MEC system, where an N-antenna AP (with an integrated MEC server) employs RF signal based transmit energy beamforming to charge a set $\mathcal{K} \triangleq \{1, \ldots, K\}$ of single-antenna users. The downlink WPT and the wireless communication (for offloading) are assumed to operate over orthogonal frequency bands simultaneously. We consider a time block of finite duration T for both the WPT and the MEC. Assume that the task is partitionable

^{*}The author would like to thank his colleague Prof. Jie Xu for the careful discussions and comments in this work.

for each of the K users. Relying on the harvested energy in one block, each user can then arbitrarily partition its task into two parts for local computing and for offloading to the MEC server, respectively. Note that the computation task at each user must be accomplished before the end of this block; hence, the number of computation bits over the block can measure the computation rate. In addition, we assume that the AP perfectly knows the computation information of all the K users, as well as the channel state information (CSI) from/to the K users.

A. Energy Transmit Beamforming from AP to Users

Let $s \in \mathbb{C}^{N \times 1}$ and $Q \triangleq \mathbb{E}[ss^H]$ denote the energy-bearing transmit signal by the AP and the energy transmit covariance matrix, respectively, where $\mathbb{E}[\cdot]$ and the superscript H denote the stochastic expectation and the Hermitian transpose, respectively. Let P_{\max} denote the maximum transmit power at the AP. We then have the following energy transmit beamforming constraint at the AP:

$$\operatorname{tr}(\boldsymbol{Q}) \triangleq \mathbb{E}[\|\boldsymbol{s}\|^2] \le TP_{\max},\tag{1}$$

where $\|\cdot\|$ and $\operatorname{tr}(\cdot)$ denote the Euclidean norm and the trace operation, respectively. In general, the AP can use multiple energy beams to deliver the wireless energy, i.e., Q can be of any rank [13]. Let $h_i \in \mathbb{C}^{N \times 1}$ denote the channel vector from the AP to user $i \in \mathcal{K}$, and define $H_i \triangleq h_i h_i^H$, $\forall i \in \mathcal{K}$. The harvested energy amount by user i over this block is then

$$E_{i} = T\eta \mathbb{E}\left[\left|\boldsymbol{h}_{i}^{H}\boldsymbol{s}\right|^{2}\right] = T\eta \mathrm{tr}(\boldsymbol{Q}\boldsymbol{H}_{i}), \qquad (2)$$

where $|\cdot|$ denotes the absolute value of a scalar and $0 < \eta \leq 1$ is the energy conversion efficiency per user. Both the local computing and offloading for user $i \in \mathcal{K}$ in the block are powered the harvested energy E_i .

B. Computation Task Execution for Users

Over the duration-T block, the computation task for user $i \in \mathcal{K}$ is partitioned into two parts with $\ell_i \ge 0$ and $q_i \ge 0$ bits for offloading and local computing, respectively, in parallel.

1) Computation Offloading from Users to the AP: Consider a TDMA protocol for the K users' offloading, where the block is divided into 2K time slots each with duration t_i , $\forall i \in \{1, \ldots, 2K\}$. These users offload their computation bits to the AP one by one in the first K time slots. After the offloaded tasks having been executed at the MEC server, the AP sends the computation results back to the K users in the next K time slots sequentially. As in [10], the computation time consumed at the MEC server is negligible and the user can download the computation results immediately after the first K time slots, i.e., $t_i \approx 0$, $\forall i \in \{K+1, \ldots, 2K\}$. In addition, we ignore the energy consumption for transmitting/receiving the computation results in this paper. As a result, the TDMA based offloading time allocation across the K users is

$$\sum_{i=1}^{K} t_i \le T. \tag{3}$$

Let $g_i \in \mathbb{C}^{N \times 1}$ denote the channel vector from user *i* to the AP and $p_i \ge 0$ the transmit power for user *i*'s offloading in

time slot t_i . The maximum ratio combining (MRC) receiver is further assumed for the AP to decode the information. The achievable offloading rate (in bits/sec) for user *i* is then

$$r_i = B \log_2 \left(1 + \frac{p_i \tilde{g}_i}{\Gamma \sigma^2} \right), \quad \forall i \in \mathcal{K},$$
 (4)

where *B* denotes the bandwidth, $\tilde{g}_i \triangleq ||\mathbf{g}_i||^2$ denotes the effective channel power gain from user *i* to the AP, σ^2 is the noise power at the receiver of the AP, and $\Gamma \ge 1$ is a constant accounting for the gap from the channel capacity due to a practical coding and modulation scheme. For simplicity, $\Gamma = 1$ is assumed throughout this paper. As a result, the number of offloaded bits ℓ_i by user *i* to the AP can be expressed as

$$\ell_i = r_i t_i, \quad \forall i \in \mathcal{K}. \tag{5}$$

Consider an MEC server with limit computation capacity. Let L_{max} be the maximum number of computation bits that can be executed at the MEC server over the block. We then have the following computation capacity constraint:

$$\sum_{i=1}^{K} \ell_i \le L_{\max}.$$
(6)

It is worth noting that computation offloading incurs energy consumption at both the K users and the AP. Per user $i \in \mathcal{K}$, in addition to the transmit power p_i , a constant circuit power $p_{c,i} > 0$ (by the digital-to-analog converter (DAC), filter, etc.) is consumed. The offloading energy consumption at user i is

$$E_{\text{offl},i} = \frac{t_i}{\tilde{g}_i} \beta\left(\frac{\ell_i}{t_i}\right) + p_{c,i} t_i, \tag{7}$$

where $\beta(x) \triangleq \sigma^2(2^{\frac{x}{B}} - 1)$ is a monotonically increasing and convex function with respect to x. Note that to avoid the issue of dividing by zero, we define $\beta\left(\frac{\ell_i}{t_i}\right) = 0$ when either $\ell_i = 0$ or $t_i = 0$ holds.

2) Local Computing at Users: We next consider the local computing for executing q_i computation bits at each user $i \in \mathcal{K}$. Let C_i be the number of CPU cycles required for computing one computation bit at user i. Then the total number of CPU cycles for the q_i bits is C_iq_i . By applying the dynamic voltage and frequency scaling (DVFS) technique [2], user i can adjust the CPU frequency $f_{i,n}$ for each cycle $n \in \{1, \ldots, C_iq_i\}$, where $0 < f_{i,n} \leq f_i^{\max}$ and f_i^{\max} denotes user i's maximum CPU frequency. As the local computing should be accomplished before the end of the block, we have the following computation latency requirements:

$$\sum_{n=1}^{C_{i}q_{i}} \frac{1}{f_{i,n}} \le T, \quad \forall i \in \mathcal{K}.$$
(8)

Accordingly, the consumed energy for local computing at user $i \in \mathcal{K}$ is given by

$$E_{\text{loc},i} = \sum_{n=1}^{C_i q_i} \zeta_i f_{i,n}^2, \tag{9}$$

where $\zeta_i > 0$ is the effective capacitance coefficient that depends on the chip architecture at user *i*. In order to minimize the energy consumption while satisfying the latency constraint,

it is optimal for each user to set the CPU frequencies to be identical for different CPU cycles (see [10, Lemma 3.1]). By using this fact and letting the constraints in (8) be met with strict equality, we have

$$f_{i,1} = \dots = f_{i,C_iq_i} = C_i q_i / T, \quad \forall i \in \mathcal{K}.$$
 (10)

As the maximum CPU frequency f_i^{\max} is specified for each user $i \in \mathcal{K}$, we have the following constraints for the numbers of computation bits by local computing:

$$0 \le q_i \le \frac{T f_i^{\max}}{C_i}, \quad \forall i \in \mathcal{K}.$$
(11)

By substituting (10) in (9), the energy consumption $E_{\text{loc},i}$ is re-expressed as

$$E_{\text{loc},i} = \frac{\zeta_i C_i^3 q_i^3}{T^2}.$$
 (12)

C. Energy Harvesting Constraints at Users

To achieve self-sustainable operation, the energy harvesting constraint needs to be imposed such that the totally consumed energy at the user cannot exceed the harvested energy E_i in (2) per block [14]. By combining the computation offloading energy in (7) and the local computation energy in (12), the total energy consumed by user *i* within the block is $E_{\text{offl},i} + E_{\text{loc},i}$. Therefore, we have

$$\frac{\zeta_i C_i^3 q_i^3}{T^2} + \frac{t_i}{\tilde{g}_i} \beta\left(\frac{\ell_i}{t_i}\right) + p_{c,i} t_i \le T \eta \operatorname{tr}\left(\boldsymbol{QH}_i\right), \forall i \in \mathcal{K}.$$
(13)

D. Problem Formulation

We are interested in utilizing the constrained communication/computation resources at the AP to maximize the total computation rates across all the users. Different users usually have different priorities. Accordingly, we maximize the weighted sum-number of users' computation bits subject to the MEC computation capacity and the transmit power constraints at the AP. Let $\omega_i > 0$ denote the positive weight for user *i* that characterizes the priority of its computation task. Mathematically, the weighted sum of computation rates maximization problem is formulated as

$$(\mathcal{P}1): \max_{\boldsymbol{Q} \succeq \boldsymbol{0}, \boldsymbol{t}, \boldsymbol{\ell}, \boldsymbol{q}} \sum_{i=1}^{K} \omega_i (q_i + \ell_i)$$
(14a)
s.t. $0 < t_i < T, \quad 0 < \ell_i < L_{\max}, \quad \forall i \in \mathcal{K}$ (14b)

where $\mathbf{t} \triangleq [t_1, \ldots, t_K]^{\dagger}$, $\boldsymbol{\ell} \triangleq [\ell_1, \ldots, \ell_K]^{\dagger}$, and $\mathbf{q} \triangleq [q_1, \ldots, q_K]^{\dagger}$ with the superscript \dagger being the transpose operation; $\mathbf{Q} \succeq \mathbf{0}$ guarantees \mathbf{Q} to be positive semidefinite. Note that problem ($\mathcal{P}1$) is convex and can thus be efficiently solved by the interior-point method [16]. Nevertheless, to gain more engineering insights, we next leverage the Lagrange dual method to obtain the optimal solution in a semi-closed form.

III. Optimal Solution to Problem $(\mathcal{P}1)$

In this section, we obtain the optimal solution to $(\mathcal{P}1)$ in a semi-closed form and develop an efficient algorithm.

Let $\rho \ge 0$, $\mu \ge 0$, $\theta \ge 0$, and $\lambda_i \ge 0$ be the Lagrange multipliers for (1), (3), (6), and the *i*-th constraint in (13), $\forall i \in \mathcal{K}$, respectively. The partial Lagrangian of problem ($\mathcal{P}1$) is given by

$$\mathcal{L}(\boldsymbol{Q}, \boldsymbol{t}, \boldsymbol{\ell}, \boldsymbol{q}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\rho}, \boldsymbol{\theta}) = \operatorname{tr}\left(\left(\sum_{i=1}^{K} T\eta\lambda_{i}\boldsymbol{H}_{i} - \boldsymbol{\rho}\boldsymbol{I}\right)\boldsymbol{Q}\right) + \boldsymbol{\mu}T + \boldsymbol{\rho}TP_{\max} + \boldsymbol{\theta}L_{\max} + \sum_{i=1}^{K} \left(\omega_{i}q_{i} - \frac{\lambda_{i}\zeta_{i}C_{i}^{3}q_{i}^{3}}{T^{2}}\right) + \sum_{i=1}^{K} \left((\omega_{i} - \boldsymbol{\theta})\boldsymbol{\ell}_{i} - \frac{\lambda_{i}t_{i}}{\tilde{g}_{i}}\boldsymbol{\beta}\left(\frac{\boldsymbol{\ell}_{i}}{t_{i}}\right) - \boldsymbol{\mu}t_{i} - \lambda_{i}p_{\mathrm{c},i}t_{i}\right).$$
 (15)

The dual function of problem $(\mathcal{P}1)$ is then

$$\Phi(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\rho}, \boldsymbol{\theta}) = \max_{\boldsymbol{Q} \succeq \boldsymbol{0}, \boldsymbol{t}, \boldsymbol{\ell}, \boldsymbol{q}} \mathcal{L}_2(\boldsymbol{Q}, \boldsymbol{t}, \boldsymbol{\ell}, \boldsymbol{q}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\rho}, \boldsymbol{\theta}) \quad (16)$$

s.t. (11) and (14b).

Consequently, the dual problem of $(\mathcal{P}1)$ is expressed as

$$(\mathcal{D}1): \min_{\boldsymbol{\lambda}, \mu, \rho, \theta} \Phi(\boldsymbol{\lambda}, \mu, \rho, \theta)$$
(17a)

s.t.
$$\boldsymbol{G}(\boldsymbol{\lambda}, \rho) \triangleq \sum_{i=1}^{N} \eta \lambda_i \boldsymbol{H}_i - \rho \boldsymbol{I} \preceq \boldsymbol{0}$$
 (17b)

$$\mu \ge 0, \ \rho \ge 0, \ \theta \ge 0, \ \lambda_i \ge 0, \ \forall i \in \mathcal{K}.$$
 (17c)

Note that the constraint (17b) is to ensure that the dual function $\Phi(\lambda, \mu, \rho, \theta)$ is bounded from above (see Appendix A). Denote by S the set of $(\lambda, \mu, \rho, \theta)$ characterized by (17b) and (17c).

Since problem ($\mathcal{P}1$) is convex and satisfies the Slater's condition, strong duality holds between ($\mathcal{P}1$) and ($\mathcal{D}1$) [16]. As a result, we can solve ($\mathcal{P}1$) by equivalently solving the dual problem ($\mathcal{D}1$). For convenience of presentation, we denote (Q^*, t^*, ℓ^*, q^*) as the optimal solution to problem (16) under given ($\lambda, \mu, \rho, \theta$) $\in S$, ($Q^{\text{opt}}, t^{\text{opt}}, \ell^{\text{opt}}, q^{\text{opt}}$) as the optimal solution to ($\mathcal{P}1$), ($\lambda^{\text{opt}}, \mu^{\text{opt}}, \rho^{\text{opt}}, \theta^{\text{opt}}$) as the optimal solution to ($\mathcal{D}1$).

A. Evaluating the Dual Function $\Phi(\lambda, \mu, \rho, \theta)$

First, we obtain the dual function $\Phi(\lambda, \mu, \rho, \theta)$ under any given $(\lambda, \mu, \rho, \theta) \in S$. Note that problem (16) can be decomposed into (2K + 1) subproblems as follows, one for optimizing Q, K for optimizing q_i 's, and another K for jointly optimizing t_i 's and ℓ_i 's, respectively.

$$\max_{\boldsymbol{Q}} \operatorname{tr} (\boldsymbol{G}(\boldsymbol{\lambda}, \rho)\boldsymbol{Q}) \quad \text{s.t.} \quad \boldsymbol{Q} \succeq \boldsymbol{0}$$
(18)

$$\max_{q_i} \omega_i q_i - \frac{\lambda_i \zeta_i C_i^3 q_i^3}{T^2} \quad \text{s.t.} \quad 0 \le q_i \le \frac{T f_i^{\max}}{C_i} \tag{19}$$

$$\max_{\substack{t_i, \ell_i}} (\omega_i - \theta) \ell_i - \frac{\lambda_i t_i}{\tilde{g}_i} \beta\left(\frac{\ell_i}{t_i}\right) - \mu t_i - \lambda_i p_{c,i} t_i$$

s.t. $0 \le t_i \le T, \quad 0 \le \ell_i \le L_{\max},$ (20)

where $i \in \mathcal{K}$.

For problem (18), under the condition of $G(\lambda, \rho) \leq 0$, the optimal value of (18) is zero and the optimal solution Q^* to (18) can be *any* positive semidefinite matrix in the null space of $G(\lambda, \rho)$. Note that the optimal solution $Q^* = 0$ of (18) is used only for evaluating the dual function $\Phi(\lambda, \mu, \rho, \theta)$.

Next, consider the *i*-th subproblems in (19) and (20). As both problems are convex and satisfy the Slater's condition, one can obtain their solutions in semi-closed forms based on the Karush-Kuhn-Tucker (KKT) conditions [16], as stated in the following two lemmas, respectively.

<u>Lemma</u> 3.1: Under given $(\lambda, \mu, \rho, \theta) \in S$, the optimal solution of the number of local computing bits q_i^* to problem (19) can be obtained as follows.

- If $\lambda_i = 0$, we have $q_i^* = T f_i^{\max} / C_i$;
- If $\lambda_i > 0$, we have

$$q_i^* = \min\left(\sqrt{\frac{\omega_i T^2}{3\lambda_i \zeta_i C_i^3}}, \frac{T f_i^{\max}}{C_i}\right).$$
(21)

Proof: See Appendix B.

<u>Lemma</u> 3.2: The optimal solution of the offloading time t_i^* and the number of offloading bits ℓ_i^* to problem (20) are given below.

- If λ_i = 0 and ω_i−θ > 0, we have t^{*}_i = 0 and ℓ^{*}_i = L_{max};
- If λ_i = 0 and ω_i − θ ≤ 0, we have t^{*}_i = 0 and ℓ^{*}_i = 0;
 If λ_i > 0 and ω_i − θ ≤ λ_iσ² ln 2/(Bğ̃_i), we have t^{*}_i = 0
- If $\lambda_i > 0$ and $\omega_i \theta \le \lambda_i \sigma^2 \ln 2/(Bg_i)$, we have $t_i^* = 0$ and $\ell_i^* = 0$;
- If $\lambda_i > 0$ and $\omega_i \theta > \lambda_i \sigma^2 \ln 2/(B\tilde{g}_i)$, the optimal offloading rate is given by

$$r_i^* = B \log_2 \left(\frac{(\omega_i - \theta) B \tilde{g}_i}{\lambda_i \sigma^2 \ln 2} \right), \tag{22}$$

and $\ell_i^* = r_i^* t_i^*$, where

$$t_{i}^{*} \begin{cases} =0, & \text{if } \frac{\lambda_{i}}{\tilde{g}_{i}} \left(\beta(r_{i}^{*}) - r_{i}^{*}\beta'(r_{i}^{*})\right) + \mu + \lambda_{i}p_{c,i} > 0\\ \in [0,T], & \text{if } \frac{\lambda_{i}}{\tilde{g}_{i}} \left(\beta(r_{i}^{*}) - r_{i}^{*}\beta'(r_{i}^{*})\right) + \mu + \lambda_{i}p_{c,i} = 0\\ =T, & \text{otherwise}, \end{cases}$$

$$(23)$$

and $\beta'(x)$ denotes the first-order derivative with respect to x.

Proof: See Appendix C.

As stated in Lemma 3.2, if $\frac{\lambda_i}{\hat{g}_i} (\beta(r_i) - r_i \beta'(r_i)) + \mu + \lambda_i p_{c,i} = 0$, $t_i^* \in [0,T]$ is generally *not* a unique solution; in this case, we set $t_i^* = 0$ to facilitate the dual function evaluation. An additional procedure will be used in Section III-C to retrieve the optimal primary t^{opt} , together with Q^{opt} .

By combining $Q^* = 0$ and Lemmas 3.1–3.2, the dual function $\Phi(\lambda, \mu, \rho, \theta)$ in (D1) can be evaluated.

B. Obtaining $(\boldsymbol{\lambda}^{\text{opt}}, \mu^{\text{opt}}, \rho^{\text{opt}}, \theta^{\text{opt}})$ to Minimize $\Phi(\boldsymbol{\lambda}, \mu, \rho, \theta)$

Note that the dual function $\Phi(\lambda, \mu, \rho, \theta)$ is convex but non-differentiable in general. As a result, problem $(\mathcal{D}1)$ can be solved by subgradient based methods such as the ellipsoid method [17]. For the objective function in (17a), the subgradient with respect to $(\lambda, \mu, \rho, \theta)$ is given by $\left[-\frac{\zeta_1 C_1^3 q_1^{*3}}{T^2} - \frac{t_1^*}{\tilde{g}_1} \beta\left(\frac{\ell_1^*}{t_1^*}\right) - p_{c,1}t_1^*, \dots, -\frac{\zeta_K C_K^3 q_K^{*3}}{T^2} - \frac{t_K^*}{\tilde{g}_K} \beta\left(\frac{\ell_K^*}{t_K^*}\right) -$

 $p_{c,K}t_K^*, T - \sum_{i=1}^{K} t_i^*, TP_{\max}, L_{\max} - \sum_{i=1}^{K} \ell_i^* \Big]^{\dagger}$. As in [10], the subgradient for the constraint in (17b) is given by $[\eta v^H H_1 v, \ldots, \eta v^H H_K v, 0, -1, 0]^{\dagger}$, where $v \in \mathbb{C}^{N \times 1}$ is an eigenvector corresponding to the largest eigenvalue of $G(\lambda, \rho)$ and is given by $v = \arg \max_{\|\boldsymbol{\xi}\|=1} \boldsymbol{\xi}^H G(\lambda, \rho) \boldsymbol{\xi}$. Furthermore, the subgradients for the constraints in (17c) are given by $e_{K+1}, e_{K+2}, e_{K+3}$, and $e_i, \forall i \in \mathcal{K}$, respectively, where e_i is the standard unit vector with one in the *i*-th entry and zeros elsewhere in $\mathbb{R}^{(K+3) \times 1}$.

C. Finding $(Q^{\text{opt}}, t^{\text{opt}}, \ell^{\text{opt}}, q^{\text{opt}})$ to Problem $(\mathcal{P}1)$

With the optimal dual $(\lambda^{\text{opt}}, \mu^{\text{opt}}, \rho^{\text{opt}}, \theta^{\text{opt}})$, it remains to determine the optimal solution for $(\mathcal{P}1)$. Specifically, by replacing λ with λ^{opt} in Lemma 3.1, one can obtain the optimal q^{opt} to $(\mathcal{P}1)$, and accordingly find the optimal CPU frequencies as $f_{i,1}^{\text{opt}} = \cdots = f_{i,C_i q_i}^{\text{opt}} = C_i q_i^{\text{opt}}/T$, $\forall i \in \mathcal{K}$. By replacing (λ, μ, θ) with $(\lambda^{\text{opt}}, \mu^{\text{opt}}, \theta^{\text{opt}})$ in Lemma 3.2, we obtain the offloading rate r_i^{opt} and accordingly $\ell_i^{\text{opt}} = r_i^{\text{opt}} t_i^{\text{opt}}$, $\forall i \in \mathcal{K}$. Nevertheless, as t_i^* is generally non-unique in (23), one cannot obtain t_i^{opt} (and ℓ_i^{opt}) directly here but requires an additional procedure. By substituting q^{opt} and $\ell_i = r_i^{\text{opt}} t_i$, $\forall i \in \mathcal{K}$, in problem $(\mathcal{P}1)$, we have the following semidefinite program (SDP) to obtain Q^{opt} and t^{opt} :

$$(\boldsymbol{Q}^{\text{opt}}, \boldsymbol{t}^{\text{opt}}) = \underset{\boldsymbol{Q} \succeq \boldsymbol{0}, \ \boldsymbol{t}}{\operatorname{arg\,max}} \sum_{i=1}^{K} \omega_i (q_i^{\text{opt}} + t_i r_i^{\text{opt}})$$
(24)

s.t.
$$\frac{\zeta_i C_i^3 (q_i^{\text{opt}})^3}{T^2} + \frac{t_i}{\tilde{g}_i} \beta \left(r_i^{\text{opt}} \right) + p_{\text{c},i} t_i \leq T \eta \text{tr} \left(\boldsymbol{Q} \boldsymbol{H}_i \right), \forall i \in \mathcal{K}$$

(1), (3),
$$\sum_{i=1}^K t_i r_i^{\text{opt}} \leq L_{\text{max}}, \text{ and } 0 \leq t_i \leq T, \forall i \in \mathcal{K}.$$

Note that the SDP in (24) can be efficiently solved via CVX [18]. Then the optimal ℓ_i^{opt} 's are obtained as $\ell_i^{\text{opt}} = r_i^{\text{opt}} t_i^{\text{opt}}$, $\forall i \in \mathcal{K}$. By combining Q^{opt} , t^{opt} and ℓ^{opt} here, together with q^{opt} , the optimal solution to problem ($\mathcal{P}1$) is found.

<u>*Remark*</u> 3.1: It can be readily shown that the optimal solution to $(\mathcal{P}1)$ has the following properties:

- 1) For all users, it is optimal to leave a strictly positive number of bits for local computing, i.e., $q_i^{\text{opt}} > 0, \forall i$.
- 2) At the optimality, each user fully exploits its harvested energy, i.e., the energy harvesting constraints in (13) are active for all users.

Note that the first property is due to the fact that the marginal energy consumption of local computing is nearly zero when $q_i^{\text{opt}} \rightarrow 0$; hence, it is always beneficial to leave some bits for local computing. Intuitively, the second property indicates that all users' harvested energy must be used up for ($\mathcal{P}1$) to maximize the weighted sum-number of computed bits subject to the limited resource and computation latency constraints.

IV. NUMERICAL RESULTS

In this section, numerical results are provided to validate the performance of the proposed design with joint WPT and computation offloading optimization, as compared to the following three benchmark schemes.



Fig. 1. The average number of computation bits per user versus the maximum transmit power P_{\max} at the AP.

1) Local computing only: Each user $i \in \mathcal{K}$ accomplishes its computation task by only local computing. This scheme corresponds to solving ($\mathcal{P}1$) by setting $\ell_i = 0$, $\forall i \in \mathcal{K}$.

2) Computation offloading only: Each user $i \in \mathcal{K}$ accomplishes its computation task by fully offloading the computation bits to the AP. This scheme corresponds to solving $(\mathcal{P}1)$ by setting $q_i = 0, \forall i \in \mathcal{K}$.

3) Joint design with isotropic WPT: The AP radiates the RF energy isotropically over all directions by setting Q = pI, where p denotes the transmit power at each antenna and I is an $N \times N$ identity matrix. This scheme corresponds to solving ($\mathcal{P}1$) by replacing Q as pI with p being an optimized variable.

In the simulations, the AP is equipped with N = 4 antennas. The energy conversion efficiency is $\eta = 0.8$. All channels are modeled as independent Rayleigh fading with an average power loss of 5×10^{-6} (i.e., -53 dB) which corresponds to a distance of about 5 meters from users to the AP in the urban environment. We set $C_i = 10^3$ cycles/bit, $\zeta_i = 10^{-28}$, and the circuit power as $p_{c,i} = 10^{-4}$ Watt (W) for $i \in \mathcal{K}$. The receiver noise power at the AP is set as $\sigma^2 = 10^{-9}$ W and the bandwidth for offloading as B = 2 MHz. We set the maximum number of computation bits supported by the MEC and the maximum CPU frequency for each user $i \in \mathcal{K}$ as $L_{\rm max} = 2 \times 10^5$ bits and $f_i^{\rm max} = 0.1$ GHz, respectively. In addition, the weights are set to be identical for different users, i.e., $\omega_i = 1/K$, $i \in \mathcal{K}$. Accordingly, the objective is to maximize the average number of computation bits per user. The results are obtained by averaging over 500 randomized channel realizations.

Fig. 1 shows the average number of computation bits versus the AP transmit power P_{max} , where K = 10 and T = 0.1 sec. It is observed that the proposed joint design achieves significant performance gains over the three benchmark schemes, and the isotropic WPT design is clearly suboptimal. The performances achieved by all the schemes improve significantly as P_{max} increases. When $P_{\text{max}} \ge 50$ dBm, the average numbers of computation bits for all the schemes are bounded from above. This is expected since in this case, the number of computable bits is fundamentally limited by both the



Fig. 2. The average number of computation bits per user versus the user number K.

computation capability and users' local CPU frequencies. It is also observed that the local-computing-only scheme achieves a near optimal performance close to that with the proposed joint design at small $P_{\rm max}$ values. This suggests that most users prefer computing locally in this case.

Fig. 2 shows the average number of computation bits versus the user number K, where $P_{\text{max}} = 40$ dBm and T = 0.1 sec. In general, we have similar observations as in Fig. 1. Particularly, as K increases, the average number of computation bits per user by all the schemes decreases and the decreasing with the full-offloading-only scheme is more significantly than the other three schemes. This is due to the fact that all users share the finite time block in the full-offloading-only scheme, thereby leading to drastically decreasing of the number of offloaded bits when K becomes large.

V. CONCLUSION

In this paper, we investigated a unified WPT-MEC design for a wireless powered multiuser MEC system. Specifically, we developed an efficient design framework to maximize the weighted sum of the computation rates across all the users subject to the limited computation/communication resource constraints. Using the Lagrange dual method, we obtained the optimal solution in a semi-closed form. Numerical results demonstrate the merit of the proposed joint design over alternative benchmark schemes.

APPENDICES

A. Proof of $G(\lambda, \rho) \preceq 0$

The condition $G(\lambda, \rho) \leq 0$ can be verified by contradiction. Assume that $G(\lambda, \rho)$ is not negative semidefinite. Denote by $\vartheta \in \mathbb{C}^{N \times 1}$ an eigenvector corresponding to one positive eigenvalue of $G(\lambda, \rho)$. By setting $Q = \tau \vartheta \vartheta^H \succeq 0$ with τ going to positive infinity, it follows that

$$\lim_{\tau \to +\infty} \operatorname{tr}(\boldsymbol{G}(\boldsymbol{\lambda}, \rho)\boldsymbol{Q}) = \lim_{\tau \to +\infty} \tau \boldsymbol{\vartheta}^{H} \boldsymbol{G}(\boldsymbol{\lambda}, \rho) \boldsymbol{\vartheta} = +\infty, \quad (25)$$

which in turn implies that the value $\Phi(\lambda, \mu, \rho, \theta)$ in (16) is unbounded from above over $Q \succeq 0$. Hence, to ensure that $\Phi(\lambda, \mu, \rho, \theta)$ is bounded from above, it requires $G(\lambda, \rho) \preceq 0$.

B. Proof of Lemma 3.1

Given λ , we solve problem (19) for each user $i \in \mathcal{K}$. When $\lambda_i = 0$, the objective function in (19) becomes $\omega_i q_i$. It is evident that $q_i^* = T f_i^{\max} / C_i$ is optimal for problem (19).

For $\lambda_i > 0$, the Lagrangian of (19), denoted by \mathcal{L}_i , is then given as

$$\bar{\mathcal{L}}_i = \omega_i q_i - \frac{\lambda_i \zeta_i C_i^3 q_i^3}{T^2} + \underline{\eta}_i q_i + \bar{\eta}_i \left(\frac{T f_i^{\max}}{C_i} - q_i\right), \quad (26)$$

where $\underline{\eta}_i \geq 0$ and $\overline{\eta}_i \geq 0$ are the Lagrange multipliers associated with $q_i \geq 0$ and $q_i \leq T f_i^{\max}/C_i$, respectively. It can be verified that (19) satisfies the Slater's condition. Based on the KKT conditions [16], it follows that

$$\underline{\eta}_{i}^{*}q_{i}^{*} = 0, \quad \bar{\eta}_{i}^{*}\left(\frac{Tf_{i}^{\max}}{C_{i}} - q_{i}^{*}\right) = 0$$
(27a)

$$\omega_i - \frac{3\lambda_i \zeta_i C_i^3 (q_i^*)^2}{T^2} + \underline{\eta}_i^* - \bar{\eta}_i^* = 0, \qquad (27b)$$

where $\underline{\eta}_i^*$ and $\bar{\eta}_i^*$ are the optimal dual variables, (27a) collects the complementary slackness conditions, and (27b) is the firstorder derivative condition for $\overline{\mathcal{L}}_i$ with respect to q_i . From (27a) and (27b), it thus follows that

$$q_i^* = \min\left(\sqrt{\frac{\omega_i T^2}{3\lambda_i \zeta_i C_i^3}}, \frac{T f_i^{\max}}{C_i}\right).$$
(28)

C. Proof of Lemma 3.2

When $\lambda_i = 0$, the objective function in (20) becomes $(\omega_i - \theta)\ell_i - \mu t_i$. Evidently, if $\omega_i - \theta > 0$, it follows that $t_i^* = 0$ and $\ell_i^* = L_{\text{max}}$ are optimal for (20); otherwise, $t_i^* = \ell_i^* = 0$. For $\lambda_i > 0$, the Lagrangian of (20) is given by

$$\underline{\mathcal{L}}_{i} = (\omega_{i} - \theta)\ell_{i} - \frac{\lambda_{i}t_{i}}{\tilde{g}_{i}}\beta\left(\frac{\ell_{i}}{t_{i}}\right) - \mu t_{i} - \lambda_{i}p_{c,i}t_{i} + \bar{a}_{i}t_{i} \\
+ \bar{b}_{i}\ell_{i} + \bar{c}_{i}(T - t_{i})$$
(29)

where $\bar{a}_i \geq 0$, $\bar{b}_i \geq 0$, and $\bar{c}_i \geq 0$ are the Lagrangian multipliers associated with $t_i \geq 0$, $\ell_i \geq 0$, and $t_i \leq T$, respectively.

Based on the KKT conditions [16], it follows that

$$\bar{a}_{i}^{*}t_{i}^{*} = 0, \quad \bar{b}_{i}^{*}\ell_{i}^{*} = 0, \quad \bar{c}_{i}^{*}(T - t_{i}^{*}) = 0$$

$$\lambda : \left(\begin{array}{c} \ell^{*} \\ \ell^{*} \end{array} \right) \quad \ell^{*} \quad \ell^{*} \\ \lambda : \left(\begin{array}{c} \ell^{*} \\ \ell^{*} \end{array} \right) \quad \lambda^{*} \quad \ell^{*} \\ \lambda^{*} \quad \ell^{*} \\ \lambda^{*} \\ \lambda^{*}$$

$$-\frac{\lambda_i}{\tilde{g}_i} \left(\beta \left(\frac{\ell_i}{t_i^*} \right) - \frac{\ell_i}{t_i^*} \beta' \left(\frac{\ell_i}{t_i^*} \right) \right) - \mu - \lambda_i p_{\mathrm{c},i} + \bar{a}_i^* - \bar{c}_i^* = 0$$
(30b)

$$(\omega_i - \theta) - \frac{\lambda_i}{\tilde{g}_i} \beta' \left(\frac{\ell_i^*}{t_i^*}\right) + \bar{b}_i^* = 0, \qquad (30c)$$

where $(\bar{a}_i^*, \bar{b}_i^*, \bar{c}_i^*)$ denotes the optimal dual solution, (30a) collects the complementary slackness conditions, and the left-hand-side (LHS) terms of (30b) and (30c) are the first-order derivatives of $\underline{\mathcal{L}}_i$ with respect to t_i^* and ℓ_i^* , respectively. Let $r_i^* \triangleq \ell_i^*/t_i^*$, and define $r_i^* = 0$ if either $\ell_i^* = 0$ or $t_i^* = 0$. From (30a) and (30c), we have

$$r_i^* = \begin{cases} 0, & \text{if } (\omega_i - \theta) B \tilde{g}_i \le \lambda_i \sigma^2 \ln 2\\ B \log_2 \left(\frac{(\omega_i - \theta) B \tilde{g}_i}{\lambda_i \sigma^2 \ln 2} \right), & \text{otherwise.} \end{cases}$$
(31)

Furthermore, substituting (31) into (30b), we obtain that

$$\bar{a}_i^* - \bar{c}_i^* = \frac{\lambda_i}{\tilde{g}_i} \left(\beta \left(r_i^* \right) - r_i \beta' \left(r_i^* \right) \right) + \mu + \lambda_i p_{\mathrm{c},i}.$$
(32)

Clearly, $\bar{a}_i^* - \bar{c}_i^* > 0$ and $\bar{a}_i^* - \bar{c}_i^* < 0$ imply that $\bar{a}_i^* > 0$ and $\bar{c}_i^* > 0$, respectively. In addition, when $\bar{a}_i^* - \bar{c}_i^* = 0$, we have $\bar{a}_i^* = \bar{c}_i^* = 0$. Based on the complementary slackness conditions in (30a), it follows that

$$t_{i}^{*} \begin{cases} = 0, & \text{if } \frac{\lambda_{i}}{\tilde{g}_{i}} \left(\beta\left(r_{i}^{*}\right) - r_{i}\beta'\left(r_{i}^{*}\right)\right) + \mu + \lambda_{i}p_{\text{c},i} > 0\\ \in [0,T], & \text{if } \frac{\lambda_{i}}{\tilde{g}_{i}} \left(\beta\left(r_{i}^{*}\right) - r_{i}\beta'\left(r_{i}^{*}\right)\right) + \mu + \lambda_{i}p_{\text{c},i} = 0\\ = T, & \text{otherwise}, \end{cases}$$

and $\ell_i^* = r_i^* t_i^*$. Now it completes the proof of Lemma 3.2.

REFERENCES

- M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," *IEEE Internet Thing J.*, vol. 3, no. 6, pp. 854–864, Jun. 2016.
- [2] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," 2017. [Online]. Available: https: //arxiv.org/abs/1702.05309
- [3] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "Mobile edge computing: The communication perpective," 2017. [Online]. Available: https://arxiv.org/abs/1701.01090
- [4] S. Barbarossa, S. Sardellitti, and P. D. Lorenzo, "Communicating while computing: Distributed mobile cloud computing over 5G heterogenous networks," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 45–55, Nov. 2014.
- [5] ETSI white paper (2015). Mobile edge computing: A key technology towards 5G. [Online]. Available: http://www.etsi.org/images/files/ ETSIWhitePapers/etsi_wp11_mec_a_key_technology_towards_5g.pdf
- [6] J. Liu, Y. Mao, J. Zhang, and K. B. Letaief, "Delay-optimal computation task scheduling for mobile-edge computing systems," in *Proc. IEEE ISIT*, Barcelona, Spain, Jun. 2016, pp. 1451–1455.
- [7] M.-H. Chen, B. Liang, and M. Dong, "Joint offloading decision and resource allocation for mobile cloud with computing access point," in *Proc. IEEE ICASSP*, Shanghai, China, Mar. 2016, pp. 3516–3520.
- [8] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2795–2808, Oct. 2016.
- [9] C. You, K. Huang, and H. Chae, "Energy efficient mobile cloud computing powered by wireless energy transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1757–1770, May 2016.
- [10] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wirless powered mobile-edge computing system," in *Proc. IEEE ICC*, Paris, France, May 2017, pp. 1–6.
- [11] X. Cao, F. Wang, J. Xu, R. Zhang, and S. Cui, "Joint computation and communication cooperation for mobile edge computing," 2017. [Online]. Available: https://arxiv.org.org/1704.06777
- [12] F. Wang, J. Xu, and Z. Ding, "Optimized multiuser computation offloading with multi-antenna NOMA," 2017. [Online]. Available: https: //arxiv.org/abs/1707.02486
- [13] 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.
- [14] H. Li, J. Xu, R. Zhang, and S. Cui, "A general utility optimization framework for energy harvesting based wireless communications," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 79–85, Apr. 2015.
- [15] F. Wang, T. Peng, Y. Huang, and X. Wang, "Robust transceiver optimization for power-splitting based downlink MISO SWIPT systems," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1492–1496, Sep. 2015.
- [16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [17] S. Boyd, "Ellipsoid method," Stanford University, California, USA. [Online]. Available: http://stanford.edu/class/ee364b/lectures/ellipsoid_ method_notes.pdf
- [18] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," 2009. [Online]. Available: http://cvxr.com/cvx/