Wireless Powered Cooperation-Assisted Mobile Edge Computing

Xiaoyan Hu 🆔, *Student Member, IEEE*, Kai-Kit Wong, *Fellow, IEEE*, and Kun Yang, *Senior Member, IEEE*

Abstract— This paper studies a mobile edge computing (MEC) system in which two mobile devices are energized by the wireless power transfer (WPT) from an access point (AP) and they can offload part or all of their computation-intensive latency-critical tasks to the AP connected with an MEC server or an edge cloud. This harvest-then-offload protocol operates in an optimized time-division manner. To overcome the doubly near-far effect for the farther mobile device, cooperative communications in the form of relaying via the nearer mobile device is considered for offloading. Our aim is to minimize the AP's total transmit energy subject to the constraints of the computational tasks. We illustrate that the optimization is equivalent to a min–max problem, which can be optimally solved by a two-phase method. The first phase obtains the optimal offloading decisions by solving a sum-energy-saving maximization problem for given an energy transmit power. In the second phase, the optimal minimum energy transmit power is obtained by a bisection search method. Numerical results demonstrate that the optimized MEC system utilizing cooperation has significant performance improvement over systems without cooperation.

Index Terms— Mobile edge computing, user cooperation, wireless power transfer, energy consumption, resource management.

I. Introduction

THE rapid developments of Internet-of-things (IoT) and 5G communication technologies, as well as a wide range of emerging mobile applications, from highly-interactive online gaming to facial recognition, and virtual reality, etc., have driven the increasing computing demands for mobile devices. Such computationally intensive applications easily exceed the ability of resource-limited mobile devices, not to mention the fact that they will drain the power quickly. To tackle this issue, in recent years, mobile edge computing (MEC) has emerged as a promising concept, which promotes to use cloud-computing facilities at the edge of mobile networks by integrating MEC servers at the wireless access points (APs). This paradigm of computation offloading is motivated by ultralow latency, high bandwidth, and real-time access to radio network information, which is widely considered as an effective means to liberate the mobile devices from heavy computation workloads, e.g., [1]–[3].

A. Prior Works

MEC, with proximate access, is a promising complementary counterpart of centralized mobile cloud computing. The cross-disciplinary nature of MEC lays the important role of resource management in achieving energy-efficient or delay-optimal MEC. Recent years have witnessed encouraging progress on this topic for both single-user [4]–[8] as well as multi-user [9]–[14] MEC systems. For single-user MEC systems, the energy-optimal mobile cloud computing under stochastic wireless channel was considered in [4]. Later in [5], a dynamic offloading scheme with link selection was proposed to improve the energy efficiency. Another dynamic offloading scheme with energy harvesting was addressed in [7] to reduce the delay cost. In [8], a Markov decision process approach was adopted to handle a delay minimization problem. As for the multi-user MEC systems, joint radio-and-computational resource management becomes more complicated. A multi-cell MEC offloading system was considered in [9] in order to minimize the overall energy consumption of users. In [10], the distributed offloading decision making problem was formulated as a multiuser computation offloading game. Optimal energy-efficient resource allocation for multiple users was addressed in [11] based on time-division multiple-access (TDMA) and orthogonal frequency-division multiple-access (OFDMA) systems. The cooperation among clouds was investigated in [12] to maximize the revenues of clouds. In [14], a stochastic resource management resorting Lyapunov optimization was considered to minimize the power consumption.

Taking the full benefits of powerful computational resources at the edges, nonetheless, faces several challenges. Insufficient power supply is one major limitation for battery-based devices and mobile applications will be terminated if the battery is running out. It therefore makes sense to leverage the technology of wireless power transfer (WPT) so that mobile devices are not power-limited by their batteries but can be energized remotely, e.g., [15]–[19]. WPT particularly in the form of wireless powered communication networks (WPCNs) [17]–[19]

have recently been considered as an important paradigm to provide genuine sustainability for mobile communications. In addition, many works have seen the possible synergy integrating MEC with WPT [6], [13]. An interesting work in [6] considered a wireless powered single-user MEC system, in which binary offloading was investigated, i.e., either local computing only or fully offloading, so as to maximize the computing probability. More recently in [13], an energy-efficient wireless powered multiuser MEC system combining with a multi-antenna AP was considered. The optimal transmit energy beamforming of the AP, offloading decision and resource allocation for minimizing the energy consumption at the AP were obtained.

However, WPCNs are susceptible to suffer from the so-called "double-near-far" effect, which occurs because a farther user from an AP harvests less energy and is also required to communicate in longer distances [17], [18]. It is known that user cooperation has been extensively investigated in wireless communications for its ability to enhance data rate under unfavourable channel conditions, e.g., [18]–[22]. Particularly the efforts in [18]–[20] focused on the effect of cooperation between near-far users, trying to improve the performance of WPCNs. Most recently, user cooperation was also considered in MEC [23], where a three-node MEC system was considered to exploit joint computation and communication cooperation for reducing the energy consumption.

B. Our Contributions

In this paper, we study a wireless powered MEC system to complete the computation-intensive latency-critical tasks of two near-far users[1] exploiting cooperative communication, where the entire process is solely powered by the AP. Note that as far as green computing is concerned, minimizing the carbon footprint of the AP has appeared to be the priority in WPT-MEC systems. Hence, our objective is to minimize the total transmit energy of the AP with jointly optimal power and time allocation to fully explore the benefits of user cooperation in enhancing the performance of the WPT-MEC system. Our contributions are summarized as follows:

- A harvest-then-offload protocol with a block-based time division mechanism that leverages cooperative communications to overcome the doubly near-far effect in WPT-MEC systems, is proposed.
- We first formulate the AP's transmit energy minimization (APTEM) problem, and then transform it into an equivalent min-max optimization problem (that also turned out to be equivalent to the AP's transmit power minimization (APTPM) problem). The problem is optimally tackled by a two-phase approach. In the first phase, the inner sum-energy-saving maximization (SESM) problem based on a given energy transmit power is solved by the Lagrangian method, where the optimal offloading decisions with joint power and time allocation are found in closed or semi-closed form. Then in the second phase, a simple bisection search is adopted to obtain the minimum-energy transmit power based on the solution of the SESM problem, resulting the joint-optimal solution.

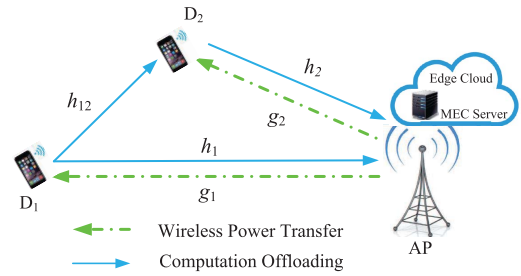[1]In this paper, we use "device" and "user" interchangeably.

Fig. 1.   The model for the two-user WPT-MEC system.

- Further, we prove that the optimal offloaded data sizes of the two users have threshold-based structures in relation to some offloading priority indicators, and the thresholds are determined by the users' energy harvesting potentials, reflecting the effect of user cooperation. It is also verified that the optimal WPT time-duration is a monotonic non-decreasing function of the AP's transmit power, which further shows the equivalence between the APTEM and APTPM problems. Moreover, we prove that at least one user makes no energy saving when the minimum-energy transmit power is employed at the AP.
- A low-complexity algorithm is proposed to solve the APTEM problem, and we show that the complexity is at most with the order of $\mathcal{O}(1)\ln(1/\sigma)\ln(1/\delta)$, where $\sigma, \delta > 0$ respectively denote the computational accuracies of two tiers of one-dimensional search in the algorithm.
- Simulation results verify the theoretical analysis of the proposed cooperative computation offloading scheme by comparing with two baselines. It is shown that the proposed scheme not only achieves significant performance improvement, but also demonstrates the effectiveness of handling computation-intensive latency-critical tasks and resisting the double-near-far effect in WPCNs.

The rest of this paper is organized as follows. In Section II, we introduce the system model and the problem formulation. The proposed two-phase method for energy-efficient resource allocation with user cooperation is presented in Section III. Section IV provides the simulation results. Some possible extensions will be discussed in Section V and we conclude our paper in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a wireless powered MEC system shown in Fig. 1 that consists of a single-antenna AP (with an integrated MEC server), and two single-antenna mobile devices, denoted by D_1 and D_2, both operating in the same frequency band and each having a computation-intensive latency-critical task to be completed. A block-based TDMA structure is adopted where each block has a duration of T seconds. During each block, AP energizes the mobile devices in the downlink via WPT. Using the harvested energy, the two devices accomplish their computation tasks in a partial offloading fashion [2], where the task-input bits are bit-wise independent and can be arbitrarily divided to facilitate parallel trade-offs between local computing at the mobile devices and computation offloading
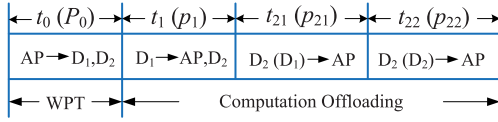
Fig. 2. The time division structure for the harvest-then-offload protocol.

to the MEC server. After the AP computes the offloaded data, it sends the results back to the devices. Note that local computing and downlink WPT can be performed simultaneously while wireless communications (for offloading) and WPT are non-overlapping in time considering half-duplex transmission for both two users. As a result, the harvest-then-transmit protocol proposed in [17] is employed in our model but for wireless powered computation offloading, which we refer to it as the harvest-then-offload protocol.

Assuming that the AP has the perfect knowledge of all the channels and task-related parameters which can be obtained by feedback, the AP is designed to make offloading decisions and allocate both radio and computational resources optimally. Our aim is to minimize the total transmit energy of the AP for completing the computation tasks of the two users.

A. Computation Task Model

Each user D_i ($i \in \{1, 2\}$) has a computation-intensive and latency-critical task in each block, fully characterized by a positive parameter tuple $\langle I_i, C_i, O_i, T_i \rangle$, where I_i denotes the size (in bits) of the computation input data (e.g., the program codes and input parameters), C_i is the amount of required computational resource for computing 1-bit of input data (i.e., the number of CPU cycles required), O_i is the output data size which is proportional to but much less than I_i, and T_i is the maximum tolerable latency. A mobile user can apply the methods (e.g., call graph analysis) in [24] and [25] to obtain the information of I_i and C_i. Note that this model allows rich task modelling flexibility in practice and can be easily extended to consider other kinds of resources by introducing more parameters in the tuple. In this paper, we assume that the maximum tolerable latency for two users is one block length, i.e., $T_1 = T_2 = T$.

B. User Cooperation Model for Computation Offloading

For computation-intensive tasks with large input data size I_i, it would be difficult to rely upon local computing to satisfy the latency constraint, and thus computation offloading may be necessary. Considering the double-near-far effect in our considered WPCN, cooperation amongst near-far users during offloading will help to improve the computation performance. Without loss of generality, it is assumed that D_2 is nearer to the AP than D_1, and we denote the distances between AP and D_1, AP and D_2, D_1 and D_2 as d_1, d_2, and d_{12}, respectively, with $d_2 \leq d_1$. We also assume that $d_{12} \leq d_1$, and therefore it will be easier for D_2 to decode the information sent by D_1 than the AP, which makes such cooperative communications useful. For an arbitrary single block, the time division structure is shown in Fig. 2. During the first period t_0, AP broadcasts wireless power to both D_1 and D_2 in the downlink with transmit power P_0. Assume that the two devices have enough

battery storages, and thus the energy harvested by each device during the WPT period is given by

$$E_i = v_i g_i P_0 t_0, \ i \in \{1, 2\}, \tag{1}$$

where g_i is the downlink channel power gain from the AP to D_i and $0 < v_i \leq 1$ is the energy conversion efficiency for D_i. Note that no other sources of energy are available to carry out the computation tasks except from WPT of the AP.

After the WPT period, D_1 transmits its input-data-bearing information with average power p_1 from its harvested energy during the subsequent period t_1, and both the AP and D_2 decode their respective received signals from D_1. To overcome the doubly near-far effect, during the remaining time of the block, the nearer user D_2 will first relay the farther user D_1's information with average power p_{21} over t_{21} amount of time and then transmits its own input-data-bearing information to the AP with average power p_{22} over period t_{22}, all using its harvested energy. We denote the time allocation and power allocation vectors as $\mathbf{t} = [t_0, t_1, t_{21}, t_{22}]$ and $\mathbf{p} = [p_1, p_{21}, p_{22}]$, respectively. According to the results (Theorems 1–5) in [22], with a given pair of $\mathbf{t}$ and $\mathbf{p}$, the offloaded data size of D_1 for remote computation at the AP should be the smaller value between the decoded data sizes at the AP and D_2, i.e.,

$$L_1(\mathbf{t}, \mathbf{p}) = \min \{L_{1,1}(\mathbf{t}, \mathbf{p}) + L_{1,2}(\mathbf{t}, \mathbf{p}), L_{1,12}(\mathbf{t}, \mathbf{p})\}, \tag{2}$$

where $L_{1,1}(\mathbf{t}, \mathbf{p})$, $L_{1,2}(\mathbf{t}, \mathbf{p})$ and $L_{1,12}(\mathbf{t}, \mathbf{p})$ denote D_1's offloaded data size from D_1 to the AP, from D_2 to the AP, and from D_1 to D_2, respectively, which are given by

$$L_{1,1}(\mathbf{t}, \mathbf{p}) = t_1 r_{1,1}(\mathbf{p}) = t_1 B \log_2 \left(1 + \frac{p_1 h_1}{N_0}\right), \tag{3}$$

$$L_{1,2}(\mathbf{t}, \mathbf{p}) = t_{21} r_{1,2}(\mathbf{p}) = t_{21} B \log_2 \left(1 + \frac{p_{21} h_2}{N_0}\right), \tag{4}$$

$$L_{1,12}(\mathbf{t}, \mathbf{p}) = t_1 r_{1,12}(\mathbf{p}) = t_1 B \log_2 \left(1 + \frac{p_1 h_{12}}{N_2}\right), \tag{5}$$

where $r_{1,1}(\mathbf{p})$, $r_{1,2}(\mathbf{p})$, and $r_{1,12}(\mathbf{p})$ are the transmission rates according to the channel achievable rates for offloading D_1's input data. In the above expressions, h_1, h_2 are the uplink channel power gains from D_1 and D_2 to the AP, respectively, and h_{12} is the device-to-device channel power gain from D_1 to D_2.[2] Also, B is the channel bandwidth. N_0 and N_2 are respectively the receiver noise power at the AP and D_2, and we further assume that $N_2 = N_0$ without loss of generality. Similarly, the offloaded data size of D_2 for computing at the AP is described as

$$L_2(\mathbf{t}, \mathbf{p}) = t_{22} r_2(\mathbf{p}) = t_{22} B \log_2 \left(1 + \frac{p_{22} h_2}{N_0}\right), \tag{6}$$

where $r_2(\mathbf{p})$ denotes the transmission rate for offloading D_2's input data. According to the task model, the offloaded data size of each user should not be greater than its corresponding input data size, i.e., $L_i(\mathbf{t}, \mathbf{p}) \leq I_i$, for $i \in \{1, 2\}$.

---

[2]All the channels mentioned in this paper are quasi-static block fading channels. In order to investigate the effect of user cooperation in resisting the double-near-far problem caused by path loss, we mainly consider the case of $h_1 < h_{12}$, and thus $L_{1,1}(\mathbf{t}, \mathbf{p}) < L_{1,12}(\mathbf{t}, \mathbf{p})$ always holds.

In practice, the MEC-integrated AP will provide sufficient CPU capability and transmit power, while the computed results are usually of much smaller sizes especially compared with the input data size, i.e., $O_i \ll I_i$. Hence, the decoding and computation time spent at the AP as well as the time consumed for delivering the computed results are negligible. For the nearer user D_2, the decoding time for D_1's information is also negligible compared with the uplink offloading time for both D_1 and D_2's information. For these reasons, we only consider the WPT time and the uplink offloading time as the total latency of the WPT-MEC system, and thus we obtain a latency constraint given by

$$t_0 + t_1 + t_{21} + t_{22} \leq T. \tag{7}$$

For each user, the energy required to receive its computed results from the AP is also considered negligible. Therefore, the energy consumption of D_1 and D_2 for computation offloading equals to the energy consumed for wireless transmissions, given by

$$\begin{cases} E_{\text{off},1}(\mathbf{t}, \mathbf{p}) &= p_1 t_1, \\ E_{\text{off},2}(\mathbf{t}, \mathbf{p}) &= p_{21} t_{21} + p_{22} t_{22}. \end{cases} \tag{8}$$

C. Local Computing Model

Given a pair of time and power allocation vectors $(\mathbf{t}, \mathbf{p})$, the offloaded data sizes $\{L_i(\mathbf{t}, \mathbf{p})\}$ will be known, and hence the remaining input data of the corresponding computation tasks, i.e., $I_i - L_i(\mathbf{t}, \mathbf{p})$, should be computed locally at D_i, $i \in \{1, 2\}$. For local computing, we assume that the CPU frequency is fixed as f_i for D_i, which means that the two mobile devices are of limited computing resources. In order to satisfy the latency constraint, i.e., $(I_i - L_i(\mathbf{t}, \mathbf{p})) C_i / f_i \leq T$, the offloaded data for D_i should have a minimum size of $L_i(\mathbf{t}, \mathbf{p}) \geq M_i^+$ with $M_i = I_i - f_i T / C_i$ where $(x)^+ = \max\{x, 0\}$. Under the assumption of a low CPU voltage that normally holds for low-power devices, the energy consumption per CPU cycle for local computing at D_i can be denoted as $Q_i = \kappa_i f_i^2$, where κ_i is the effective capacitance coefficient that depends on the chip architecture. Hence, the energy consumption of D_i for local computing can be expressed as

$$E_{\text{loc},i}(\mathbf{t}, \mathbf{p}) = (I_i - L_i(\mathbf{t}, \mathbf{p})) C_i Q_i, \quad i \in \{1, 2\}. \tag{9}$$

D. Problem Formulation

Based on the model, the energy saving for $D_i, i \in \{1, 2\}$ is

$$E_{\text{s},i}(P_0, \mathbf{t}, \mathbf{p}) = v_i g_i P_0 t_0 - E_{\text{off},i}(\mathbf{t}, \mathbf{p}) - E_{\text{loc},i}(\mathbf{t}, \mathbf{p}). \tag{10}$$

Furthermore, the APTEM problem for minimizing AP's transmit energy can be formulated as problem (P1) below

$$(\text{P1}): \quad \min_{P_0 > 0, \mathbf{t}, \mathbf{p}} \quad P_0 t_0 \tag{11a}$$

$$\text{s.t.} \quad T - (t_0 + t_1 + t_{21} + t_{22}) \geq 0, \tag{11b}$$

$$E_{\text{s},1}(P_0, \mathbf{t}, \mathbf{p}) \geq 0, \tag{11c}$$

$$E_{\text{s},2}(P_0, \mathbf{t}, \mathbf{p}) \geq 0, \tag{11d}$$

$$M_1^+ \leq L_1(\mathbf{t}, \mathbf{p}) \leq I_1, \tag{11e}$$

$$M_2^+ \leq L_2(\mathbf{t}, \mathbf{p}) \leq I_2, \tag{11f}$$

$$t_0 \geq 0, \quad t_1 \geq 0, \quad t_{21} \geq 0, \quad t_{22} \geq 0, \tag{11g}$$

$$p_1 \geq 0, \quad p_{21} \geq 0, \quad p_{22} \geq 0, \tag{11h}$$

where (11c) and (11d) represent the energy harvesting constraints for D_1 and D_2, respectively. Note that problem (P1) is a nonconvex optimization problem in the above form because of the expressions of $L_1(\mathbf{t}, \mathbf{p})$ and $L_2(\mathbf{t}, \mathbf{p})$, and the product of $P_0 t_0$. Actually, problem (P1) can be equivalently transformed into the following min-max problem (P2)[3]

$$(\text{P2}): \quad \min_{P_0 > 0} \max_{\mathbf{t}, \mathbf{p}} \quad E_{\text{s},1}(\mathbf{t}, \mathbf{p}) + E_{\text{s},2}(\mathbf{t}, \mathbf{p})$$
$$\text{s.t.} \quad (11\text{b})\text{–}(11\text{h}). \tag{12}$$

However, problem (P2) is still nonconvex in this form. In order to make this problem solvable and facilitate further analysis, we propose a two-phase method. In the first phase, we solve the inner sub-problem with a given P_0 where the sum-energy-saving (SES), i.e., $E_{\text{s},1}(\mathbf{t}, \mathbf{p}) + E_{\text{s},2}(\mathbf{t}, \mathbf{p})$ is maximized under the constraints in (P1), referred to as the SESM problem (P3):

$$(\text{P3}): \quad \max_{\mathbf{t}, \mathbf{p}} \quad E_{\text{s},1}(\mathbf{t}, \mathbf{p}) + E_{\text{s},2}(\mathbf{t}, \mathbf{p})$$
$$\text{s.t.} \quad (11\text{b})\text{–}(11\text{h}), \tag{13}$$

through which the optimal time and power allocation corresponding to the given P_0 can be obtained. In the second phase, we will find the minimum P_0 by a bi-section search method. In the following section, we will demonstrate the details of the problem-solving process with the two-phase method.

III. THE TWO-PHASE METHOD FOR COOPERATIVE MEC

Here, we focus on the two-phase method for joint power and time allocation for cooperative MEC. The process of operating the first phase with a given P_0 is presented in Sections III-A to III-D, where the optimal offloaded data size, the power allocation of D_1 (in semi-closed from) and D_2 (in closed from) as well as the optimal time allocation are obtained for each sub-problem. Besides, the equivalence between problem (P1) and (P2) is given in Section III-E. Finally, the second phase is described in Section III-F, where the minimum $P_0^\star$ is achieved.

A. Transforming the SESM Problem (P3) Into Convex

To make the nonconvex SESM problem (P3) solvable with a given P_0, we first introduce the variables $q_1 = \frac{p_1 t_1}{v_1 g_1 P_0}$ and $q_{21} = \frac{p_{21} t_{21}}{v_2 g_2 P_0}$. By denoting $\mathbf{q} = [q_1, q_{21}]$, $L_{1,1}(\mathbf{t}, \mathbf{p})$, $L_{1,2}(\mathbf{t}, \mathbf{p})$ and $L_{1,12}(\mathbf{t}, \mathbf{p})$ described in (3)–(5) can then be re-expressed as functions of $\mathbf{t}$ and $\mathbf{q}$ as

$$L_{1,1}(\mathbf{t}, \mathbf{q}) = t_1 B \log_2 \left(1 + \beta_1 P_0 \frac{q_1}{t_1} \right), \tag{14}$$

$$L_{1,2}(\mathbf{t}, \mathbf{q}) = t_{21} B \log_2 \left(1 + \beta_2 P_0 \frac{q_{21}}{t_{21}} \right), \tag{15}$$

$$L_{1,12}(\mathbf{t}, \mathbf{q}) = t_1 B \log_2 \left(1 + \beta_{12} P_0 \frac{q_1}{t_1} \right), \tag{16}$$

---

[3]The proof of verifying the equivalence between problems (P1) and (P2) will be given in Section III-E after solving the inner sub-problem (P3) since the proof needs some results obtained through solving problem (P3).

where $\beta_1 = \frac{v_1 g_1 h_1}{N_0}$, $\beta_2 = \frac{v_2 g_2 h_2}{N_0}$, and $\beta_{12} = \frac{v_1 g_1 h_{12}}{N_2}$. Note that the above three functions equal to 0 when $t_1 = 0$, $t_{21} = 0$ and $t_1 = 0$, respectively. Using the property of perspective function [26], it is easily verified that $L_{1,1}(\mathbf{t}, \mathbf{q})$, $L_{1,2}(\mathbf{t}, \mathbf{q})$ and $L_{1,12}(\mathbf{t}, \mathbf{q})$ are all joint concave functions of $\mathbf{t}$ and $\mathbf{q}$. Besides, they are all monotonically increasing functions over each element of (t_1, q_1), (t_{21}, q_{21}) and (t_1, q_1), respectively. Next, we introduce a new variable

$$\overline{L}_1 = \min\left\{L_{1,1}(\mathbf{t}, \mathbf{q}) + L_{1,2}(\mathbf{t}, \mathbf{q}), L_{1,12}(\mathbf{t}, \mathbf{q})\right\} \quad (17)$$

to replace $L_1(\mathbf{t}, \mathbf{p})$ in problem (P3) with two additional convex constraints, $L_{1,1}(\mathbf{t}, \mathbf{q}) + L_{1,2}(\mathbf{t}, \mathbf{q}) \geq \overline{L}_1$ and $L_{1,12}(\mathbf{t}, \mathbf{q}) \geq \overline{L}_1$. Thus, the expression of energy saving for D_1 in the objective function of (P3) (and its corresponding constraints) has been turned into concave (convex). However, even though we can use a similar variable-changing method to convert $L_2(\mathbf{t}, \mathbf{p})$ into a concave function $L_2(\mathbf{t}, \mathbf{q})$, the corresponding constraint $L_2(\mathbf{t}, \mathbf{q}) \leq I_2$ in (11f) is still nonconvex. To tackle this, we redefine the offloaded data size of D_2 as an independent variable L_2, and then by defining a function $g(x) = N_0(2^{\frac{x}{B}} - 1)$, $x \geq 0$, the offloading power p_{22} can be described as a function of L_2 and t_{22} according to (6), given by

$$p_{22} = \frac{1}{h_2} g\left(\frac{L_2}{t_{22}}\right). \quad (18)$$

Hence, the energy savings for D_1 and D_2 with a given P_0 can be rewritten as

$$E_{s,1}\left(\mathbf{t}, \mathbf{q}, \overline{L}_1\right) = v_1 g_1 P_0 (t_0 - q_1) - (I_1 - \overline{L}_1) C_1 Q_1, \quad (19)$$

$$E_{s,2}\left(\mathbf{t}, \mathbf{q}, L_2\right) = v_2 g_2 P_0 (t_0 - q_{21})$$
$$- \frac{t_{22}}{h_2} g\left(\frac{L_2}{t_{22}}\right) - (I_2 - L_2) C_2 Q_2. \quad (20)$$

Therefore, the SESM problem (P3) can be equivalently reformulated as another SESM problem (P4)

$$\text{(P4)}: \quad \max_{\mathbf{t}, \mathbf{q}, \overline{L}_1, L_2} \; E_{s,1}\left(\mathbf{t}, \mathbf{q}, \overline{L}_1\right) + E_{s,2}\left(\mathbf{t}, \mathbf{q}, L_2\right) \quad (21a)$$

$$\text{s.t. } T - (t_0 + t_1 + t_{21} + t_{22}) \geq 0, \quad (21b)$$
$$E_{s,1}\left(\mathbf{t}, \mathbf{q}, \overline{L}_1\right) \geq 0, \quad (21c)$$
$$E_{s,2}\left(\mathbf{t}, \mathbf{q}, L_2\right) \geq 0, \quad (21d)$$
$$L_{1,1}(\mathbf{t}, \mathbf{q}) + L_{1,2}(\mathbf{t}, \mathbf{q}) \geq \overline{L}_1, \quad (21e)$$
$$L_{1,12}(\mathbf{t}, \mathbf{q}) \geq \overline{L}_1, \quad (21f)$$
$$M_1^+ \leq \overline{L}_1 \leq I_1, \quad (21g)$$
$$M_2^+ \leq L_2 \leq I_2, \quad (21h)$$
$$t_0 \geq 0, \; t_1 \geq 0, \; t_{21} \geq 0, \; t_{22} \geq 0, \quad (21i)$$
$$q_1 \geq 0, \; q_{21} \geq 0. \quad (21j)$$

As $g(x)$ is a convex function, its perspective function $t_{22} g\left(\frac{L_2}{t_{22}}\right)$ is a joint convex function of t_{22} and L_2 considering both the cases of $t_{22} > 0$ and $t_{22} = 0$ [26]. Therefore, the objective function is concave and all the constraints are convex, constituting a convex optimization problem (P4).

B. Problem-Solving With Lagrange Method

To gain more insights of the solution, we next solve problem (P4) optimally by leveraging the Lagrange method [26]. The partial Lagrange function of (P4) is defined as

$$\mathcal{L}(\mathbf{t}, \mathbf{q}, \overline{L}_1, L_2, \eta, \boldsymbol{\lambda})$$
$$\triangleq (1 + \lambda_1) E_{s,1}\left(\mathbf{t}, \mathbf{q}, \overline{L}_1\right) + (1 + \lambda_2) E_{s,2}\left(\mathbf{t}, \mathbf{q}, L_2\right)$$
$$+ \eta\left(T - (t_0 + t_1 + t_{21} + t_{22})\right)$$
$$+ \lambda_3\left(L_{1,1}(\mathbf{t}, \mathbf{q}) + L_{1,2}(\mathbf{t}, \mathbf{q}) - \overline{L}_1\right)$$
$$+ \lambda_4\left(L_{1,12}(\mathbf{t}, \mathbf{q}) - \overline{L}_1\right), \quad (22)$$

where $\eta \geq 0$ and $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_4] \succeq \mathbf{0}$ ($\succeq$ denotes the componentwise inequality) consist of the Lagrange multipliers associated with the constraints (21b) and (21c)-(21f) in problem (P4), respectively. In order to facilitate the analysis in the sequel, we define another two functions

$$f(x) = \ln(1 + x) + \frac{1}{1 + x} - 1, \quad x \geq 0, \quad (23)$$
$$h(x) = g(x) - x g'(x), \quad x \geq 0, \quad (24)$$

where $g'(x)$ denotes the first-order derivative of $g(x)$, and thus the following two lemmas are established.

Lemma 1: $f(x)$ is a monotonic increasing function of $x \geq 0$ with $f(0) = 0$. Given $C > 0$, there exists a unique positive solution for the equation $f(x) = C$, given by

$$x^* = -\left(1 + \frac{1}{W_0(-e^{(-(C+1))})}\right), \quad (25)$$

where $W_0(z)$ is the principal branch of the lambert W function defined as the solution for $W_0(z) e^{W_0(z)} = z$ [27], and e is the base of the natural logarithm.

Proof: See Appendix A. $\square$

Lemma 2: $h(x)$ is a monotonic decreasing function of $x \geq 0$ with $h(0) = 0$. Given $G < 0$, there exists a unique positive solution for the equation $h(x) = G$, given by

$$x^* = \frac{B}{\ln 2}\left[W_0\left(\frac{G/N_0 + 1}{-e}\right) + 1\right]. \quad (26)$$

Proof: See Appendix B. $\square$

We first assume that problem (P4) is feasible with the given P_0 and let $(\mathbf{t}^*, \mathbf{q}^*, \overline{L}_1^*, L_2^*)$ denote the optimal solution of (P4) and η^*, $\boldsymbol{\lambda}^*$ denote the optimal Lagrange multipliers. Then applying the Karush-Kuhn-Tucker (KKT) conditions [26] leads to the following necessary and sufficient conditions:

$$\frac{\partial \mathcal{L}}{\partial t_0^*} = (1 + \lambda_1^*) v_1 g_1 P_0 + (1 + \lambda_2^*) v_2 g_2 P_0 - \eta^* = 0, \quad (27)$$

$$\frac{\partial \mathcal{L}}{\partial t_1^*} = \frac{B \lambda_3^*}{\ln 2} f\left(\beta_1 P_0 \frac{q_1^*}{t_1^*}\right)$$
$$+ \frac{B \lambda_4^*}{\ln 2} f\left(\beta_{12} P_0 \frac{q_1^*}{t_1^*}\right) - \eta^* \begin{cases} < 0, & t_1^* = 0, \\ = 0, & t_1^* > 0, \end{cases} \quad (28)$$

$$\frac{\partial \mathcal{L}}{\partial t_{21}^*} = \frac{B \lambda_3^*}{\ln 2} f\left(\beta_2 P_0 \frac{q_{21}^*}{t_{21}^*}\right) - \eta^* \begin{cases} < 0, & t_{21}^* = 0, \\ = 0, & t_{21}^* > 0, \end{cases} \quad (29)$$

$$\frac{\partial \mathcal{L}}{\partial t_{22}^*} = -(1 + \lambda_2^*) \frac{1}{h_2} h\left(\frac{L_2^*}{t_{22}^*}\right) - \eta^* \begin{cases} < 0, & t_{22}^* = 0, \\ = 0, & t_{22}^* > 0, \end{cases} \quad (30)$$

$$\frac{\partial \mathcal{L}}{\partial q_1^*} = -(1 + \lambda_1^*)v_1 g_1 P_0 + \frac{B}{\ln 2}$$

$$\times \left(\frac{\lambda_3^* \beta_1 P_0}{1 + \beta_1 P_0 \frac{q_1^*}{t_1^*}} + \frac{\lambda_4^* \beta_{12} P_0}{1 + \beta_{12} P_0 \frac{q_1^*}{t_1^*}} \right) \begin{cases} < 0, & q_1^* = 0, \\ = 0, & q_1^* > 0, \end{cases} \tag{31}$$

$$\frac{\partial \mathcal{L}}{\partial q_{21}^*} = -(1 + \lambda_2^*)v_2 g_2 P_0$$

$$+ \frac{B}{\ln 2} \left(\frac{\lambda_3^* \beta_2 P_0}{1 + \beta_2 P_0 \frac{q_{21}^*}{t_{21}^*}} \right) \begin{cases} < 0, & q_{21}^* = 0, \\ = 0, & q_{21}^* > 0, \end{cases} \tag{32}$$

$$\frac{\partial \mathcal{L}}{\partial \overline{L}_1^*} = (1 + \lambda_1^*)C_1 Q_1 - \lambda_3^* - \lambda_4^* \begin{cases} < 0, & \overline{L}_1^* = M_1^+, \\ = 0, & \overline{L}_1^* \in (M_1^+, I_1), \\ > 0, & \overline{L}_1^* = I_1, \end{cases} \tag{33}$$

$$\frac{\partial \mathcal{L}}{\partial L_2^*} = (1 + \lambda_2^*)$$

$$\times \left[C_2 Q_2 - \frac{1}{h_2} g' \left(\frac{L_2^*}{t_{22}^*} \right) \right] \begin{cases} < 0, & L_2^* = M_2^+, \\ = 0, & L_2^* \in (M_2^+, I_2), \\ > 0, & L_2^* = I_2, \end{cases} \tag{34}$$

$$\eta^* \left(T - (t_0^* + t_1^* + t_{21}^* + t_{22}^*) \right) = 0, \tag{35}$$

$$\lambda_1^* E_{s,1} \left(\mathbf{t}^*, \mathbf{q}^*, \overline{L}_1^* \right) = 0, \tag{36}$$

$$\lambda_2^* E_{s,2} \left(\mathbf{t}^*, \mathbf{q}^*, L_2^* \right) = 0, \tag{37}$$

$$\lambda_3^* \left(L_{1,1}(\mathbf{t}^*, \mathbf{q}^*) + L_{1,2}(\mathbf{t}^*, \mathbf{q}^*) - \overline{L}_1^* \right) = 0, \tag{38}$$

$$\lambda_4^* \left(L_{1,12}(\mathbf{t}^*, \mathbf{q}^*) - \overline{L}_1^* \right) = 0. \tag{39}$$

Note that $t_0^* + t_1^* + t_{21}^* + t_{22}^* = T$ must hold; otherwise, we can always allocate the remaining time to t_0^* to further increase the energy saving of the two users, and thus $\eta^* > 0$ holds for sure. Furthermore, the following lemma describes an important result concerning $\mathbf{t}^*$, $\mathbf{q}^*$ and $\overline{L}_1^*$:

Lemma 3: The optimal time and power allocation $(\mathbf{t}^*, \mathbf{q}^*)$ ensures the following property of D_1's offloaded data size, $\overline{L}_1^*$.

$$\overline{L}_1^* = L_{1,1}(\mathbf{t}^*, \mathbf{q}^*) + L_{1,2}(\mathbf{t}^*, \mathbf{q}^*) \le L_{1,12}(\mathbf{t}^*, \mathbf{q}^*). \tag{40}$$

Proof: See Appendix C. $\square$

Remark 1 (Intuitive Explanation): Lemma 3 sheds light on the fact that the optimal offloaded data size of D_1, i.e., $\overline{L}_1^*$ should be the sum of the decoded data sizes at the AP, i.e., $(L_{1,1}(\mathbf{t}^*, \mathbf{q}^*) + L_{1,2}(\mathbf{t}^*, \mathbf{q}^*))$ rather than $L_{1,12}(\mathbf{t}^*, \mathbf{q}^*)$, which simplifies the expression of $\overline{L}_1$ compared with that in (2).

Based on the result of Lemma 3, we can derive that $\lambda_3^* > 0$ and $\lambda_4^* = 0$. Furthermore, for $\mathbf{t}^* \succ 0$ and $\mathbf{q}^* \succ 0$, it can be derived from the KKT conditions (28), (29) and the result of Lemma 1 that

$$\beta_1 \frac{q_1^*}{t_1^*} = \beta_2 \frac{q_{21}^*}{t_{21}^*} = -\frac{1}{P_0} \left(1 + \left(W_0 \left(-e^{-\left(\frac{\eta^* \ln 2}{\lambda_3^* B} + 1 \right)} \right) \right)^{-1} \right). \tag{41}$$

Moreover, through the KKT conditions (31) and (32), we can respectively derive that

$$\beta_1 \frac{q_1^*}{t_1^*} = \frac{\lambda_3^* B \beta_1}{(1 + \lambda_1^*)v_1 g_1 P_0 \ln 2} - \frac{1}{P_0}, \tag{42}$$

$$\beta_2 \frac{q_{21}^*}{t_{21}^*} = \frac{\lambda_3^* B \beta_2}{(1 + \lambda_2^*)v_2 g_2 P_0 \ln 2} - \frac{1}{P_0}. \tag{43}$$

Based on (41)-(43), we obtain that $(1 + \lambda_1^*)v_1 g_1 P_0 = \frac{\beta_1}{\beta_2}(1 + \lambda_2^*)v_2 g_2 P_0$. Combining the condition (27), the optimal Lagrange multipliers have the following property:

$$(1 + \lambda_i^*)v_i g_i P_0 = \frac{\beta_i \eta^*}{\beta_1 + \beta_2}, \quad i \in \{1, 2\}. \tag{44}$$

Hence, by substituting (44) into (42) and (43), we obtain

$$\beta_1 \frac{q_1^*}{t_1^*} = \beta_2 \frac{q_{21}^*}{t_{21}^*} = \frac{B \lambda_3^* (\beta_1 + \beta_2)}{\eta^* \ln 2} - \frac{1}{P_0}. \tag{45}$$

Based on these results, the optimal resource allocation of problem (P4) for a given feasible P_0 is characterized in the following subsections.

C. Optimal Offloading Decisions With Power Allocation

First, we define an offloading priority indicator for D_i as

$$\mu_i \triangleq \frac{B h_i C_i Q_i}{N_0 \ln 2}, \quad i \in \{1, 2\}. \tag{46}$$

Note that μ_i depends on the corresponding variables quantifying uplink channel (h_i), local computing overhead ($C_i Q_i$), and it is an monotonically increasing function of h_i, C_i and Q_i. The relationship between the optimal offloaded data size and power allocation for each user with the corresponding offloading priority indicator is shown in the following Theorem.

Theorem 1 (Optimal Cooperative Computation Offloading Decisions With Power Allocation):

1) If $M_1^+ > 0$ or $\mu_1 \ge (\beta_1 + \beta_2)P_0/z^*$, the optimal $\overline{L}_1^*$, p_1^* and p_{21}^* (all in semi-closed from) can be expressed as

$$\overline{L}_1^* \begin{cases} = M_1^+, & \mu_1 < \frac{(\beta_1 + \beta_2)P_0}{z^*}, \\ \in (M_1^+, I_1), & \mu_1 = \frac{(\beta_1 + \beta_2)P_0}{z^*}, \\ = I_1, & \mu_1 > \frac{(\beta_1 + \beta_2)P_0}{z^*}, \end{cases} \tag{47}$$

$$p_1^* = \frac{N_0}{h_1} \left(\frac{(\beta_1 + \beta_2)P_0}{z^*} - 1 \right) > 0, \tag{48}$$

$$p_{21}^* = \frac{N_0}{h_2} \left(\frac{(\beta_1 + \beta_2)P_0}{z^*} - 1 \right) > 0, \tag{49}$$

in which z^* is the unique solution of the equation given by $e^{\left(\frac{1}{(\beta_1 + \beta_2)P_0} - 1 \right)z} - \frac{e}{(\beta_1 + \beta_2)P_0} z = 0$ on the specific range of $z \in (0, (\beta_1 + \beta_2)P_0)$. If $M_1^+ = 0$ and $\mu_1 < (\beta_1 + \beta_2)P_0/z^*$, it is optimal to set $\overline{L}_1^* = 0$, $p_1^* = 0$, and $p_{21}^* = 0$.

2) If $M_2^+ > 0$ or $\rho(\mu_2) \ge (\beta_1 + \beta_2)P_0$, the optimal L_2^* and p_{22}^* (all in closed form) are given by

$$L_2^* \begin{cases} = M_2^+, & \rho(\mu_2) < (\beta_1 + \beta_2)P_0, \\ \in (M_2^+, I_2), & \rho(\mu_2) = (\beta_1 + \beta_2)P_0, \\ = I_2, & \rho(\mu_2) > (\beta_1 + \beta_2)P_0, \end{cases} \tag{50}$$

$$p_{22}^* = \frac{1}{h_2} g\left(\frac{B}{\ln 2}\left[W_0\left(\frac{(\beta_1 + \beta_2)P_0 - 1}{e}\right) + 1\right]\right), \quad (51)$$

where $\rho(\mu_2) \triangleq \mu_2 \ln \mu_2 - \mu_2 + 1$. If $M_2^+ = 0$ and $\rho(\mu_2) < (\beta_1 + \beta_2)P_0$, it is optimal to set $L_2^* = 0$ and $p_{22}^* = 0$.

Proof: See Appendix D. □

Lemma 4 (Quick Offloading Decisions for the Minimum Offloaded Data Size of D_1 and D_2): When $\mu_1 \leq 1$ (or $\mu_2 \leq 1$), the optimal offloaded data size for D_1 (or D_2) is the minimum, i.e., $\overline{L}_1^* = M_1^+$ (or $L_2^* = M_2^+$). In these two cases, we can get the optimal $\overline{L}_1^*$ (or L_2^*) just according to the value of μ_1 (or μ_2) without making comparisons as in (47) (or (50)).

Proof: See Appendix E. □

Remark 2 (Whether Computation Offloading is Necessary?): According to Theorem 1, it is easy to note that the offloading decision and power allocation of each user depend on their corresponding offloading priority indicator μ_i as well as the minimum required offloaded data size M_i^+. If $M_1^+ = 0$ and $\mu_1 < (\beta_1 + \beta_2)P_0/z^*$, then operating the whole computation task locally is optimal for D_1; otherwise computation offloading is required. Similarly, if $M_2^+ = 0$ and $\rho(\mu_2) < (\beta_1 + \beta_2)P_0$, then fulfilling the whole computation task locally is optimal for D_2; otherwise computation offloading is necessary.

Remark 3 (Effects of Parameters on the Offloading Priority): It is easy to note that $\rho(\mu_2)$ is a monotonic increasing function of μ_2 for $\mu_2 > 1$ (as for $\mu_2 \leq 1$, $L_2^* = M_2^+$), and thus it also monotonically increases with parameters C_2, Q_2 and h_2 in this case, according to the monotonicity rule of compound function. The results in Theorem 1 show that the optimal offloaded data sizes for the two cooperative users D_i, $i \in \{1, 2\}$ grow with increasing μ_i, which is consistent with the intuition that more resources should be scheduled to computation offloading when users have good channels (i.e., large h_i) or endure high local computing energy consumption (i.e., large C_i and Q_i), so as to save energy.

Remark 4 (Binary Structure of the Offloading Decisions for Two Cooperative Users): Theorem 1 reveals that the optimal offloading decisions for both D_1 and D_2 have a similar threshold-based structure when computation offloading saves energy. Moreover, since the exact cases of $\mu_1 = (\beta_1 + \beta_2)P_0/z^*$ in (47) and $\rho(\mu_2) = (\beta_1 + \beta_2)P_0$ in (50) rarely occur in practice, the optimal offloading decisions have a binary structure for both cooperative users.

Remark 5 (Effects of Parameters on the Thresholds of the Offloading Decisions): The same item in the thresholds of the offloading decisions for the two users in Theorem 1, i.e., $(\beta_1 + \beta_2) = (\nu_1 g_1 h_1 + \nu_2 g_2 h_2)/N_0$ reflects the energy harvesting potentials of the two users (i.e., $\nu_1 g_1$ and $\nu_2 g_2$) and the quality of uplink offloading channels for the users (i.e., h_1 and h_2), which demonstrates the effect of user cooperation that either user's offloading decision is affected by the other user's energy-harvesting ability and offloading-channel quality.

Lemma 5: For the case of $\overline{L}_1^* > 0$, the optimal transmit rates of D_1 and D_2 for offloading D_1's input data are same, which is expressed as

$$r_{1,1}(\mathbf{p}^*) = r_{1,2}(\mathbf{p}^*) = B \log_2 \frac{(\beta_1 + \beta_2)P_0}{z^*}. \quad (52)$$

Proof: It is easy to verify the result in Lemma 5 by substituting the optimal transmit power in (48) and (49) into the expressions of $r_{1,1}(\mathbf{p})$ and $r_{1,2}(\mathbf{p})$ in (3) and (4), respectively. □

D. Optimal Energy-Efficient Time Allocation

Using Theorem 1, we have obtained the optimal offloaded data size, i.e., $(\overline{L}_1^*, L_2^*)$ and the optimal power allocation, i.e., $\mathbf{p}^* = (p_1^*, p_{21}^*, p_{22}^*)$ for the SESM problem (P3) under a given feasible P_0. In this subsection, we focus on obtaining the corresponding optimal time allocation, i.e., $\mathbf{t}^* = (t_0^*, t_1^*, t_{21}^*, t_{22}^*)$, which is summarised in Theorem 2.

Theorem 2 (Optimal Time Allocation for WPT and Cooperative Computation Offloading):

1) The optimal time allocation for offloading D_2's input data is given by

$$t_{22}^* = \frac{\ln 2 \times L_2^*}{B\left[W_0\left(\frac{(\beta_1 + \beta_2)P_0 - 1}{e}\right) + 1\right]}. \quad (53)$$

2) The optimal WPT duration time can be derived as

$$t_0^* = \begin{cases} T - t_{22}^* - \overline{L}_1^*/r_{1,1}(\mathbf{p}^*), & \overline{L}_1^* > 0, \\ T - t_{22}^*, & \overline{L}_1^* = 0. \end{cases} \quad (54)$$

3) The optimal time allocation for offloading D_1's input data, i.e., (t_1^*, t_{21}^*) can be expressed as[4]

$$\begin{cases} t_1^* = \frac{\overline{L}_1^*}{r_{1,12}(\mathbf{p}^*)}, \\ t_{21}^* = \frac{\overline{L}_1^*}{r_{1,1}(\mathbf{p}^*)} - t_1^*, \end{cases} \quad (55)$$

where $(t_1^*, t_{21}^*) = (0, 0)$ when $\overline{L}_1^* = 0$.

Proof: See Appendix F. □

E. The Equivalence Between Problem (P1) and (P2)

In this part, we proceed to show the equivalence between the original APTEM problem (P1) and the min-max problem (P2). First, an important property of the optimal WPT duration time t_0^* is given in the following Lemma 6.

Lemma 6: The optimal WPT duration time t_0^* is a monotonic non-decreasing function of P_0.

Proof: See Appendix G. □

Remark 6 (The Effect of P_0 and t_0 on Maximizing SES): The result of Lemma 6 shows that t_0^* is small when P_0 is relatively small, since in this case the extra energy harvested by increasing t_0 cannot compensate the extra energy consumed by reducing the time for computation offloading (i.e., $T - t_0$), leading to a smaller SES. On the contrary, when P_0 becomes large, t_0^* increases accordingly to obtain more SES.

Theorem 3: The APTEM problem (P1) is equivalent to the min-max problem (P2).

Proof: We first introduce a transitional problem (P5), denoted as the AP's transmit power minimization (APTPM) problem

$$(P5): \min_{P_0 > 0, \mathbf{t}, \mathbf{p}} P_0, \quad \text{s.t. (11b)–(11h).} \quad (56)$$

---

[4]In this paper, we mainly consider the case of $h_1 < h_2$, which is most likely to happen based on our assumption. Actually, if the rare case of $h_1 > h_2$ does happen, we can simply exchange the roles of the two devices to apply the proposed scheme, which will achieve similar performance.

In the sequel, we first try to prove the equivalence between problem (P2) and (P5), and then show the equivalence of problem (P5) and (P1) to finally verify the theorem.

Problem (P5) is a general problem for minimizing the WPT transmit power P_0 by jointly optimizing P_0, $\mathbf{t}$ and $\mathbf{p}$, while (P2) gives a specific method for obtaining the minimum P_0. (P2) is solved by a two-phase method where the minimum $P_0^\star$ can be obtained through a one-dimensional (bisection) search by solving problem (P3) (or P4) with each given P_0. It is easy to understand that if we assume the given P_0 is the minimum $P_0^\star$, then the optimal $\mathbf{t}^\star$ and $\mathbf{p}^\star$ of (P5) can be obtained by solving the SESM problem (P3) with the given $P_0^\star$. If we find the minimum given $P_0^\star$ that maximizes the sum-energy-saving with all the constraints being satisfied through a bisection search, then the obtained $(P_0^\star, \mathbf{t}^\star, \mathbf{p}^\star)$, i.e., the optimal solution of (P2), is actually the joint-optimal solution of (P5). Hence, we can say that problem (P2) and (P5) are equivalent for obtaining the joint-optimal $(P_0^\star, \mathbf{t}^\star, \mathbf{p}^\star)$.

According to the result of Lemma 6, the optimal WPT duration time of the SESM problem (P3), i.e., t_0^*, is a monotonic non-decreasing function of P_0, which indicates that $P_0 t_0^*(P_0)$ is a monotonic increasing function of P_0. Hence, we can conclude that the minimum P_0 of the APTPM problem (P5) is same as the optimal P_0 for minimizing $P_0 t_0$ in the original APTEM problem (P1), which means that (P1) and (P5) are equivalent, finally proving the equivalence between problem (P1) and (P2). This indicates that when the minimum feasible $P_0^\star$ is used in (P3) (or P4), the obtained maximum SES reaches its minimum with respect to P_0. □

F. Optimal Resource Allocation for Obtaining $P_0^\star$

In this section, we will discuss the second phase of solving problem (P2). It is easy to note that with a larger feasible P_0, as extra $\triangle P_0 > 0$ is available, the feasible region of problem (P3) (or P4) will be larger as well, and thus more extra energy, at least $v_1 g_1 \triangle P_0 t_0 + v_2 g_2 \triangle P_0 t_0$ will be saved, which means that the maximum SES obtained by (P3) (or P4) is a monotonic increasing function of P_0 as long as (P3) (or P4) is feasible. Hence, the minimum $P_0^\star$ of the original APTEM problem (P1) can be obtained through a bisection search of P_0.

As a matter of fact, the optimal time allocation parameters should satisfy the latency constraint (7). Note that t_{22}^* monotonically decreases with P_0, and thus a lower bound of P_0, denoted as P_0^L, can be obtained by solving the equation $t_{22}^*(P_0) = T$ with $L_2^* = I_2$. Based on this P_0^L, we can further obtain a proper upper bound of P_0, denoted as P_0^U, which should make problem (P4) feasible and lead to positive energy savings for both of the users. The optimal $P_0^\star$ must be in the range of (P_0^L, P_0^U), and the following lemma shows a property of $P_0^\star$ which gives a stopping criterion of the bisection search.

Lemma 7: When the minimum feasible $P_0^\star$ is used in problem (P3) (or P4), at least one of the two users should use up all its harvested energy, i.e., $E_{s,1}^*(P_0^\star) = 0$ or $E_{s,2}^*(P_0^\star) = 0$.

Proof: The above lemma can be proved by the method of contradiction. If both $E_{s,1}^*(P_0^\star) > 0$ and $E_{s,2}^*(P_0^\star) > 0$

hold, then at least $\triangle P_0 = \min \left\{ \frac{E_{s,1}^*(P_0^\star)}{v_1 g_1 t_0^*}, \frac{E_{s,2}^*(P_0^\star)}{v_2 g_2 t_0^*} \right\} > 0$ can be reduced to minimize P_0, which will make $E_{s,1}^*(P_0^\star - \triangle P_0) = 0$ or $E_{s,2}^*(P_0^\star - \triangle P_0) = 0$. □

The whole process of solving the original APTEM problem (P1) is summarized in Algorithm 1, where the final optimal $P_0^\star$ and the corresponding offloaded data size $(\overline{L}_1^\star, L_2^\star)$, and power-time allocation $(\mathbf{p}^\star, \mathbf{t}^\star)$ can all be obtained.

Algorithm 1 Joint Power and Time Allocation Algorithm for Solving the APTEM Problem (P1)

1: **Input**: $v_i, g_i, h_i, f_i, (I_i, C_i, T), Q_i = \kappa_i f_i^2, M_i = I_i - f_i T / C_i,$ $i \in \{1, 2\}$, and $\omega > 1, \delta > 0, B, T, N_0, h_{12}$.
2: Initialize $P_0^U = P_0^L, \theta = 0$, where $\theta \in \{0, 1\}$ is an indicator for the feasibility of problem (P4);
3: **while** $\theta = 0$ **do**
4:     Set $P_0^U = \omega P_0^U$, and then obtain the corresponding $(\overline{L}_1^*, L_2^*, \mathbf{p}^*)$ and $\mathbf{t}^*$ according to Theorem 1 and Theorem 2;
5:     Calculating $E_{s,1}^*(P_0^U)$ and $E_{s,2}^*(P_0^U)$ according to (10);
6:     **if** $t_0^*(P_0^U) > 0, E_{s,1}^*(P_0^U) > 0$, and $E_{s,2}^*(P_0^U) > 0,$ **then** $\theta = 1$;
7:     **else** $\theta = 0$;
8:     **end if**
9: **end while**
10: **while** $P_0^U - P_0^L > \delta$ **do**
11:     Set $P_0 = (P_0^L + P_0^U)/2$, and then obtain the corresponding $(\overline{L}_1^*, L_2^*, \mathbf{p}^*)$ and $\mathbf{t}^*$ according to Theorem 1 and Theorem 2;
12:     Calculating $E_{s,1}^*(P_0)$ and $E_{s,2}^*(P_0)$ according to (10);
13:     **if** $t_0^*(P_0) > 0, E_{s,1}^*(P_0) > 0$, and $E_{s,2}^*(P_0) > 0,$ **then** $P_0^U = P_0$;
14:     **elseif** $t_0^*(P_0) \leq 0$ or $E_{s,1}^*(P_0) < 0$ or $E_{s,2}^*(P_0) < 0,$ **then** $P_0^L = P_0$;
15:     **else** break;
16:     **end if**
17: **end while**
18: **Output**: $P_0^\star = P_0$, and the corresponding $\overline{L}_1^\star, L_2^\star, \mathbf{t}^\star = (t_0^\star, t_1^\star, t_{21}^\star, t_{22}^\star), \mathbf{p}^\star = (p_1^\star, p_{21}^\star, p_{22}^\star)$.

Remark 7 (Low-Complexity Algorithm): Through implementing Algorithm 1, the optimal solutions of the original APTEM problem (P1) can be obtained with closed or semi-closed form by substituting the optimal $P_0^\star$ into Theorem 1 and Theorem 2. At most two tiers of one-dimensional (bisection) search are needed to execute Algorithm 1. The inner tier one is for obtaining z^* in Theorem 1-1) and the outer tier one is for acquiring the optimal $P_0^\star$ following the step 2-step 15. Therefore, the complexity of Algorithm 1 is at most with the order of $\mathcal{O}(1) \ln(1/\sigma) \ln(1/\delta)$, where $\sigma, \delta > 0$ denote the computational accuracies of the two tiers of one-dimensional search. Compared with the traditional block-coordinate descending algorithm where iterative optimization should be operated, the proposed Algorithm 1 is of much lower complexity.

IV. SIMULATION RESULTS

In this section, the performance of the proposed wireless powered computation offloading scheme with user cooperation by jointly optimizing power and time allocation is investigated by computer simulations. We will refer to our scheme as

"UC-JOPT" in the figures for comparison. Also, we include the results of the following two baselines:

1) A simplified wireless powered computation offloading scheme with user cooperation where D_1 and D_2 use same transmit time to offload D_1's input data ("UC-ET"). In this scheme, $\mathbf{p}$, $\overline{L}_1$, L_2, t_{22} and t_1 are assigned as the optimal solutions obtained from Theorem 1 and Theorem 2. As for t_{21}, D_2 chooses to use the same time duration as t_1 to relay D_1's input-data information, and thus $t_0 = T - t_1 - t_{21} - t_{22}$, which is suboptimal when compared with the optimal resource allocation in the proposed UC-JOPT scheme.

2) Wireless powered computation offloading scheme with inactive user cooperation by letting $t_{21} = 0$ and $t_1 = \overline{L}_1^*/r_{1,1}(\mathbf{p}^*)$ ("IUC" or cooperation is disabled).

The simulation settings are set as follows unless specified otherwise. The bandwidth and the time block length are set as $B = 10\text{MHz}$ and $T = 0.2\text{s}$, respectively. It is assumed that the channel reciprocity holds for the downlink and uplink, and thus $g_1 = h_1$, $g_2 = h_2$. The channel power gain is modeled as $h_j = 10^{-3}d_j^{-\alpha}\phi_j$, $j \in \{1, 2, 12\}$, where ϕ_j represents the short-term fading which is assumed to be an exponentially distributed random variable with unit mean (Rayleigh fading). For distance d_j in meters with the same path-loss exponent α, a 30dB average signal power attenuation is assumed for all the channels at reference of 1m. We assume that $d_1 = 10\text{m}$, $d_2 = 6\text{m}$, $d_{12} = 6\text{m}$ and $\alpha = 2$. The noise at the AP and D_2 is assumed to have a white power of $N_0 = 10^{-9}\text{W}$. For each user D_i, $i \in \{1, 2\}$, the CPU frequency f_i is uniformly selected from the set of $\{0.1, 0.2, \ldots, 1.0\}\text{GHz}$. We set $\nu_i = 0.8$ and $\kappa_i = 10^{-28}$, respectively. As for the computation tasks, the input data size and the required number of CPU cycles per bit follow the uniform distribution with $I_i \in [100, 500]$ KB and $C_i \in [1000, 2000]$ cycles/bit, respectively. The figures by simulations in the following subsections are based on 1000 independent realizations, in which h_j, f_i, I_i and C_i are randomly selected according to the above assumptions in each realization, modeling the real heterogeneous computing scenarios.

A. The Equivalence of Problem (P1) and (P2)

In this subsection, we will verify the equivalence of problem (P1) and (P2) by simulations. The average minimum transmit energy (AMTE) combining with the corresponding average minimum transmit power (AMTP) results at the AP are shown in Fig. 3 and Fig. 4, versus the block length T and the same input data size $I = I_1 = I_2$, respectively.

From Fig. 3, we can observe that the corresponding curves of AMTE and AMTP illustrate the same trend and property, verifying the equivalence of these two criteria in problem (P1) and (P2). It is shown that the proposed UC-JOPT scheme obviously outperforms the baselines. Specifically, the curves of UC-JOPT are much lower than those of UC-ET, indicating the effectiveness of the optimization for time allocation. Besides, the AMTE and AMTP of UC-JOPT are even less than half of those for IUC, which further displays the significance of user cooperation in handling the doubly near-far effect. It is valuable to note that the gaps of AMTE (AMTP) between
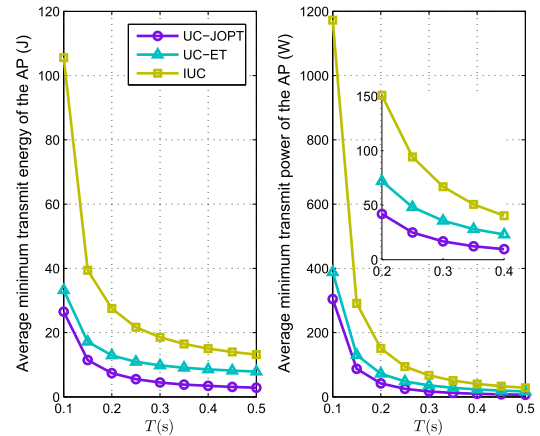

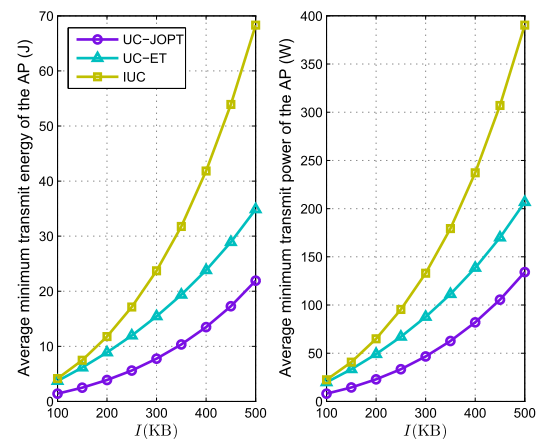Fig. 3. Average minimum transmit energy and power of the AP versus T.

Fig. 4. Average minimum transmit energy and power of the AP versus I.

different schemes become more significant for a shorter block length, demonstrating the superiority of the proposed UC-JOPT scheme in handling the latency-critical tasks.

Fig. 4 also shows the equivalence between problem (P1) and (P2) by depicting both AMTE and AMTP versus the same input data size I. The AMTE and AMTP of all the schemes increase gradually with I, as expected. Besides, the performance improvement of the proposed UC-JOPT scheme is clearly displayed, and we can obtain similar results as those reported in Fig. 3. Also, it is noted that the reduction of AMTE (AMTP) between different schemes become more obvious as I increases, which further indicates the advantage of the proposed UC-JOPT scheme in completing computation-intensive tasks.

The above results verify that the proposed UC-JOPT scheme is highly capable of dealing with computation-intensive latency-critical tasks and resisting the double-near-far effect in WPCNs by fully taking the benefits of joint-optimal resource allocation and user cooperation.

B. The Effect of Path Loss

From the expression of the channel power gain described above, it is understood that the path-loss exponent α and the distances d_1, d_2 and d_{12} have great influence on the value of h_1, h_2 and h_{12}, and thus further affect the AMTE (AMTP) of
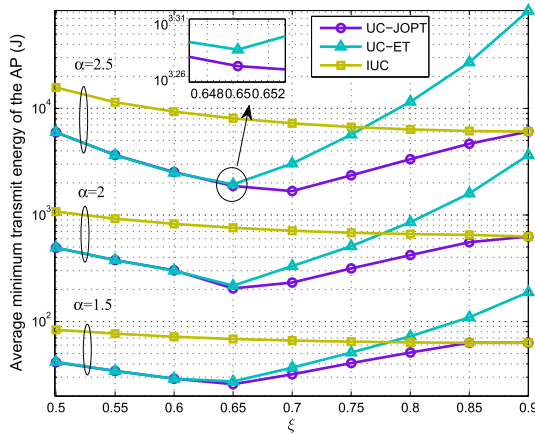
Fig. 5.   Average minimum transmit energy of the AP versus ξ.

each scheme. In this part of simulations, we set same short-term fading parameters for D_1 and D_2, i.e., $\phi_1 = \phi_2$, and focus on the effect of α and distances on the AMTE. Setting $d_1 = 10$m, $d_2 = \xi d_1$, and $d_{12} = (1 - \xi)d_1$, Fig. 5 depicts the AMTE with respect to ξ for $\alpha = 1.5, 2, 2.5$.

From the results in Fig. 5, we have the observation that the performance of the proposed UC-JOPT scheme is superior to the benchmarks, and the corresponding improvements are even more pronounced with a larger α, indicating that the UC-JOPT scheme is highly effective in resisting the attenuation caused by path loss. It is also noticed that the AMTE curves of the two cooperative schemes, i.e., UC-JOPT and UC-ET, first decrease then increase with ξ, and there is a saddle point of ξ in each curve achieving the minimum AMTE. This is due to the fact that for the cooperative computation offloading schemes, the performance depends not only on h_2 but also h_{12}, and there exists a tradeoff between these two values. When ξ is small, the performance is limited by the value of h_{12}, and the AMTE curves decrease with ξ since h_{12} increases accordingly. Around the saddle point, the performance of both two cooperative schemes degrades with ξ as the decreasing h_2 plays a dominant role in this situation. This figure also shows that when ξ is less than the saddle point, the gaps between the two cooperative schemes are not that obvious, while the gaps widen obviously as ξ goes beyond the saddle point. It is interesting to note that the performance of the proposed UC-JOPT scheme converges to that of IUC as ξ gradually tends to 1 since both D_1 and D_2 suffer from severe signal attenuation, and $t_{21}^{\star}$ gradually approaches to 0. However, the performance of the UC-ET scheme is even worse than that of the IUC scheme when ξ becomes larger approaching to 1, which shows the importance and effect of optimizing the offloading time fraction.

V. EXTENSIONS

This work focuses on the wireless powered cooperation-assisted MEC model for only a three-node scenario, with an AP, and two near-far mobile devices D_1, D_2, all with single antenna. However, extensions to other more complex

scenarios are possible. This section discusses some straight-forward approaches to extend the proposed system to more general settings.

- *Multi-antenna AP*—In this case, the design of the transmit energy beamforming and the receive signal combining at the AP will be considered to improve the network performance giving the multiple antenna capability of the AP. Such design can be easily achieved by using maximum ratio transmission for wireless power transfer and maximum ratio combining for data reception at the AP. The formulation and approach will be more or less the same except that the resulting channel coefficients after the antenna processing is considered.
- *More Mobile Devices*—Our proposed method in its current form addresses the near-far problem by pairing two mobile devices (one "near" user and another "far" user) for cooperation. Therefore, a natural approach would be to list, then rank and pair users according to their distances from the AP. Communications among different pairs can be dealt with over orthogonal channels within the same cell covered by the AP. By doing so, our proposed solution could be adopted directly. Not allowing different pairs to occupy the same radio channels makes sense because the intra-cell interference would be too much to bear unless advanced interference mitigation techniques are in place. In that case, user pairing has to be done with consideration of the interference levels, as this will affect the energy consumption at the mobile devices and the AP. Same goes to extend the proposed work to a multi-cell scenario where inter-cell interference is an important factor. After a proper user pairing with consideration of interference control and balancing, our proposed method can be directly applied, although the pairing will be more challenging.
- *Computing Resource Sharing*—Another possible extension is to allow users to share not only the radio resources (i.e., power and relaying cooperation as in our current work) but also the computing resources, where the users with stronger computation capacity can help weaker users complete their computational tasks. In this scenario, the required optimization will be much more complex because the energy consumption for carrying out tasks for others and sending back the results to others will need to be evaluated and compared with that for simply relaying the decoded data to the AP. The overall optimization problem can be formulated in a similar manner with the emphasis on minimizing the transmit energy of the AP but the required optimization is not believed to be convex. The exact way to tackle this will require further analysis and will be considered in our future work.

VI. CONCLUSION

In this paper, we investigated the use of cooperative communications in computation offloading for a WPT-MEC system, in which an AP acts as an energy source via WPT and serves as an MEC server to assist two near-far mobile devices to complete their computation-intensive

latency-critical tasks. Joint power and time allocation for cooperative computation offloading has been considered based on a block-based harvest-then-offload protocol, with the aim to minimize the transmit energy of the AP for completing the computation tasks of the two users. A two-phase method was proposed to find the optimal solution. Simulation results revealed that the proposed scheme greatly outperforms the baselines.

APPENDIX A
PROOF OF LEMMA 1

It is easy to verify that $f(x)$ is a monotonic increasing function for $x \geq 0$ with $f(0) = 0$ by simply deriving its first-order derivative. Hence, the equation $f(x) = C$ with $C > 0$ has a unique solution. Through derivation, $f(x) = C$ can be equivalently expressed as

$$-\frac{1}{1+x}e^{-\frac{1}{1+x}} = -e^{-(C+1)} \in (-e^{-1}, 0). \tag{57}$$

By using the definition and property of Lambert function [27], we obtain the solution $x^* = -\left(1 + \frac{1}{W_0(-e^{(-(C+1))})}\right) > 0$, where $W_0(-e^{(-(C+1))}) \in (-1, 0)$.

APPENDIX B
PROOF OF LEMMA 2

Similar to Lemma 1, by deriving the first-order derivative of $h(x)$, we can verify that $h(x)$ is a monotonic decreasing function of $x \geq 0$ with $h(0) = 0$. Hence, the equation $h(x) = G$ with $G < 0$ has a unique solution. Through derivation, $h(x) = G$ can be equivalently expressed as

$$\left(\frac{\ln 2}{B}x - 1\right)e^{(\frac{\ln 2}{B}x - 1)} = \frac{G/N_0 + 1}{-e}. \tag{58}$$

Therefore, we obtain $x^* = \frac{B}{\ln 2}\left[W_0\left(\frac{G/N_0+1}{-e}\right) + 1\right] > 0$ by using the definition and property of Lambert function [27], where $W_0\left(\frac{G/N_0+1}{-e}\right) > W_0(-e^{-1}) = -1$.

APPENDIX C
PROOF OF LEMMA 3

According to the constraint (21d) and condition (30), we know that $\frac{\partial(t_{22}g(\frac{L_2}{t_{22}}))}{\partial t_{22}} = h\left(\frac{L_2}{t_{22}}\right) < 0$ for $t_{22} > 0$, which indicates that $t_{22}g(\frac{L_2}{t_{22}})$ is a monotonically decreasing function of t_{22}. It is easy to prove that the inequality $L_{1,1}(\mathbf{t}^*, \mathbf{q}^*) < L_{1,12}(\mathbf{t}^*, \mathbf{q}^*)$ always holds for the considered case of $h_1 < h_{12}$. If $L_{1,1}(\mathbf{t}^*, \mathbf{q}^*) + L_{1,2}(\mathbf{t}^*, \mathbf{q}^*) > L_{1,12}(\mathbf{t}^*, \mathbf{q}^*)$ holds, we can always allocate part of t_{21}^* to t_{22}^* while maintaining the same $\overline{L}_1^*, L_2^*, \mathbf{q}^*, t_0^*, t_1^*$ and the sum of t_{21}^*, t_{22}^*, which will decrease $L_{1,2}(\mathbf{t}^*, \mathbf{q}^*)$ until the equality holds. This operation will result in an increased $E_{s,2}\left(\mathbf{t}^*, \mathbf{q}^*, L_2^*\right)$ by decreasing $t_{22}^*g(\frac{L_2^*}{t_{22}^*})$ without reducing $E_{s,1}(\mathbf{t}^*, \mathbf{q}^*, \overline{L}_1^*)$, and thus will increase the objective function of problem (P4). Hence, expression (40) always holds with the optimal solution of problem (P4).

APPENDIX D
PROOF OF THEOREM 1

1) In order to prove the first result of Theorem 1, we need the following lemma.

Lemma 8: For function $q(z) = e^{(m-1)z} - emz = 0$, there exists a unique root on $z \in (0, \frac{1}{m})$, where $m > 0$ is a constant.

Proof: Note that $q(0) = 1 > 0$ and $q(\frac{1}{m}) = e(e^{-1/m} - 1) < 0$, indicating that there exists at least one root for $q(z) = 0$ on $z \in (0, 1/m)$. Besides, the second-order derivative of $q(z)$ is non-negative, which means that $q(z)$ is a convex function of z. Hence, we can conclude that there exists one and only one root on $(0, \frac{1}{m})$ for $q(z) = 0$, and it can be easily obtained by a bisection search on $z \in (0, \frac{1}{m})$. □

We will next show that for the cases of $M_1^+ > 0$ or $\mu_1 \geq (\beta_1 + \beta_2)P_0/z^*$, computation offloading for D_1 is necessary, and thus $\overline{L}_1^* > 0$, $t_1^* > 0$, $q_1^* > 0$. From the two expressions of $\beta_{1,1}\frac{q_1^*}{t_1^*}$ in (41) and (45), we can get the equation given below

$$W_0\left(-e^{-\left(\frac{\eta^* \ln 2}{\lambda_3^* B} + 1\right)}\right) = \frac{-\eta^* \ln 2}{\lambda_3^* B(\beta_1 + \beta_2)P_0}. \tag{59}$$

Denoting $z^* = \frac{\eta^* \ln 2}{\lambda_3^* B} > 0$ and using the definition of the Lambert function, the above equation can be rewritten as

$$e^{\left(\frac{1}{(\beta_1+\beta_2)P_0} - 1\right)z^*} - \frac{e}{(\beta_1+\beta_2)P_0}z^* = 0. \tag{60}$$

Note that $\beta_1\frac{q_1^*}{t_1^*} = \frac{\lambda_3^* B(\beta_1+\beta_2)}{\eta^* \ln 2} - \frac{1}{P_0} = \frac{(\beta_1+\beta_2)}{z^*} - \frac{1}{P_0} > 0$, which means that the above equation should have a unique root z^* on $(0, (\beta_1 + \beta_2)P_0)$ because the optimal Lagrange multipliers λ_3^* and η^* are uniquely determined in the convex optimization problem (P4). According to Lemma 8, solving (60) is equivalent to finding the unique root of $q(z) = 0$ on $z \in (0, (\beta_1 + \beta_2)P_0)$ with $m = 1/(\beta_1 + \beta_2)P_0$, and this unique root always exists which can be obtained through a bisection search on $z^* \in (0, (\beta_1 + \beta_2)P_0)$. Therefore, λ_3^* can be expressed by η^* as $\lambda_3^* = \frac{\eta^* \ln 2}{Bz^*}$.

Substituting the expressions of λ_3^*, $(1+\lambda_1^*)$ (in (44)) related to η^* and β_1 into the condition (33) leads to

$$\frac{\partial \mathcal{L}}{\partial \overline{L}_1^*} = \frac{\ln 2}{B}\left(\frac{\mu_1}{(\beta_1 + \beta_2)P_0} - \frac{1}{z^*}\right)\eta^*. \tag{61}$$

Comparison between $\frac{\mu_1}{(\beta_1+\beta_2)P_0}$ and $\frac{1}{z^*}$ according to the result in (33) establishes the result of $\overline{L}_1^*$ in (47).

Similarly, substituting $\lambda_3^* = \frac{\eta^* \ln 2}{Bz^*}$ into (45), the expressions of $\frac{q_1^*}{t_1^*}$ and $\frac{q_{21}^*}{t_{21}^*}$ can be obtained as

$$\frac{q_1^*}{t_1^*} = \frac{1}{\beta_1}\left(\frac{\beta_1 + \beta_2}{z^*} - \frac{1}{P_0}\right) > 0, \tag{62}$$

$$\frac{q_{21}^*}{t_{21}^*} = \frac{1}{\beta_2}\left(\frac{\beta_1 + \beta_2}{z^*} - \frac{1}{P_0}\right) > 0. \tag{63}$$

Based on these, we can further obtain p_1^* and p_{21}^* through the variable revivification, i.e., $p_1^* = \nu_1 g_1 P_0 \frac{q_1^*}{t_1^*}$ and $p_{21}^* = \nu_2 g_2 P_0 \frac{q_{21}^*}{t_{21}^*}$, which leads to the results in (48) and (49).

For the case of $M_1^+ = 0$, $\mu_1 < (\beta_1 + \beta_2)P_0/z^*$, it can be derived that $\overline{L}_1^* = 0$ according to condition (33), which means

that fulfilling D_1's computation task locally saves more energy, and thus we have $p_1^* = 0$, $p_{21}^* = 0$.

2) Next, we will prove the second result of Theorem 1. Similarly, we also first show that for the cases of $M_2^+ > 0$ or $\rho(\mu_2) \geq (\beta_1 + \beta_2)P_0$, computation offloading for D_2 is necessary, and thus $L_2^* > 0$, $t_{22}^* > 0$, $q_{22}^* > 0$. According to Lemma 2, the optimal transmission rate for offloading D_2's input data, i.e., $\frac{L_2^*}{t_{22}^*}$ can be obtained through (30) as

$$r_2^* = \frac{L_2^*}{t_{22}^*} = \frac{B}{\ln 2}\left[W_0\left(\frac{\frac{-h_2\eta^*}{(1+\lambda_2^*)N_0}+1}{-e}\right)+1\right]$$
$$\overset{(a)}{=} \frac{B}{\ln 2}\left[W_0\left(\frac{(\beta_1+\beta_2)P_0-1}{e}\right)+1\right] > 0, \quad (64)$$

where (a) is obtained through the property of λ_2^* in (44) and the definition of β_2. Based on the expression of $g(x)$, its first-order derivative can be expressed as $g'(x) = \frac{N_0 \ln 2}{B}2^{\frac{x}{B}}$, which is a monotonically increasing function of x. Through the KKT condition (34), we can derive that the cases $\frac{\partial\mathcal{L}}{\partial L_2^*}(<,=,>)0$ hold if and only if $\frac{L_2^*}{t_{22}^*}(>,=,<)\frac{B}{\ln 2}\ln\mu_2$, respectively. Hence, the result of L_2^* in (50) can be obtained by comparing the expression of $\frac{L_2^*}{t_{22}^*}$ in (64) and $\frac{B}{\ln 2}\ln\mu_2$, where the definition and property of the Lambert function W_0 [27] should be used. According to (18), the optimal transmit power for offloading D_2's data is $p_{22}^* = \frac{1}{h_2}g\left(\frac{L_2^*}{t_{22}^*}\right)$, giving the result in (51).

For the case of $M_2^+ = 0$, $\rho(\mu_2) < (\beta_1 + \beta_2)P_0$, it can be derived that $L_2^* = 0$ according to (34), which means that fulfilling D_2's task locally saves more energy, thus $p_{22}^* = 0$.

APPENDIX E
PROOF OF LEMMA 4

Based on the expression of $\frac{\partial\mathcal{L}}{\partial L_1}$ in (61) and the range of $z^* \in (0, (\beta_1 + \beta_2)P_0)$, we can verify that $\frac{\partial\mathcal{L}}{\partial L_1^*} < 0$ when $\mu_1 \leq 1$, and thus $\overline{L}_1^* = M_1^+$. As for D_2, $\frac{L_2^*}{t_{22}^*} > \frac{B}{\ln 2}\ln\mu_2$ always holds when $\mu_2 \leq 1$ and $L_2^* > 0$, which is equivalent to $\frac{\partial\mathcal{L}}{\partial L_2^*} < 0$ according to the proof of Theorem 1, and thus $L_2^* = M_2^+$, which completes the proof.

APPENDIX F
PROOF OF THEOREM 2

Based on the results of Theorem 1, we can easily derive the expression of t_{22}^* by leveraging the fact of $t_{22}^* = \frac{L_2^*}{r_2^*}$ with the expression of r_2^* in (64). With the result of t_{22}^*, we can further derive the optimal WPT duration time t_0^* as follows.

For the case of $\overline{L}_1^* = 0$, we understand that $t_1^* = 0$ and $t_{21}^* = 0$, and thus $t_0^* = T - t_{22}^*$. For the case of $\overline{L}_1^* > 0$, combining the results of Lemma 3, Lemma 5, and the active time-sharing constraint in (21b), establishes the following equation

$$t_1^* + t_{21}^* = \frac{\overline{L}_1^*}{r_{1,1}(\mathbf{p}^*)} = T - t_{22}^* - t_0^*, \quad (65)$$

which leads to the results in (54).

As for the derivation of (t_1^*, t_{21}^*) when $\overline{L}_1^* > 0$, we resort to the results of Lemma 3 and Theorem 1, and further derive the following Lemma.

Lemma 9: The optimal time allocation (t_1^*, t_{21}^*) for cooperatively offloading D_1's input data satisfies

$$\overline{L}_1^* = L_{1,1}(t_1^*) + L_{1,2}(t_{21}^*) = L_{1,12}(t_1^*). \quad (66)$$

Proof: According to Lemma 3 and Lemma 5, we know that

$$\overline{L}_1^* = (t_1^* + t_{21}^*)r_{1,1}(p_1^*) \leq t_1^* r_{1,12}(p_1^*), \quad (67)$$

where $\overline{L}_1^*$ and p_1^* have been obtained in Theorem 1. Since we assume that $h_1 < h_{12}$, then $r_{1,1}(p_1^*) < r_{1,12}(p_1^*)$ holds for sure. With a given feasible P_0 and the corresponding optimal t_0^*, t_{22}^*, and p_1^*, p_{21}^*, p_{22}^*, $\overline{L}_1^*$, L_2^* obtained in Theorem 1, maximizing the SES is equivalent to minimizing the following energy consumption for offloading D_1's input data, i.e.,

$$\min_{t_1, t_{21}} \ p_1^* t_1 + p_{21}^* t_{21}$$
$$\text{s.t. (67),} \quad t_1 \geq 0, \ t_{21} \geq 0. \quad (68)$$

In order to make the cooperative computation offloading strategy effective, we mainly consider the case of $h_1 < h_2$, and thus the offloading power satisfies $p_1^* > p_{21}^*$ according to the result of Theorem 1. If $\overline{L}_1^* = (t_1^* + t_{21}^*)r_{1,1}(p_1^*) < t_1^* r_{1,12}(p_1^*)$ holds, we can always increase t_{21} meanwhile decreasing t_1 with the fixed $t_1 + t_{21} = \overline{L}_1^*/r_{1,1}(p_1^*)$ until $\overline{L}_1^* = (t_1^* + t_{21}^*)r_{1,1}(p_1^*) = t_1^* r_{1,12}(p_1^*)$ holds, which will lead to a smaller objective value of problem (68). Hence, expression (66) always holds with the optimal time allocation (t_1^*, t_{21}^*). □

From the result of the above lemma, we can deduce the optimal time division parameters (t_1^*, t_{21}^*) as in (55).

APPENDIX G
PROOF OF LEMMA 6

According to the expression of t_0^* in (54), its monotonicity with respect to P_0 is determined by the monotonicity of $\overline{L}_1^*/r_{1,1}(\mathbf{p}^*)$ and $t_{22}^* = L_2^*/r_2^*$ when $\overline{L}_1^* > 0$ or $L_2^* > 0$. From the expression of r_2^* in (64), it is clear that r_2^* is a monotonic increasing function of P_0 due to the fact that the first-branch of Lambert function $W_0(\cdot)$ is a monotonic increasing function. Next, we will prove that P_0/z^* is also a monotonic increasing function of P_0 to further proceed this proof.

From the equation used to obtain z^* in (60), it is easy to note that z^* is an implicit function of P_0. Besides, equation (60) can be transformed into another form given by

$$\ln\left(\frac{z^*}{(\beta_1+\beta_2)P_0}\right) = \frac{z^*}{(\beta_1+\beta_2)P_0} - z^* - 1. \quad (69)$$

As such, the first-order derivative of z^* on P_0 can be found as

$$\frac{dz^*}{dP_0} = \frac{z^*\left[(\beta_1+\beta_2)P_0 - z^*\right]}{P_0\left[(\beta_1+\beta_2)P_0 - z^* + (\beta_1+\beta_2)P_0 z^*\right]} \quad (70)$$

through applying the differentiation rule of the implicit function on the equation (69). Note that $\frac{dz^*}{dP_0} > 0$ always holds

since z^* is in the range of $(0, (\beta_1 + \beta_2)P_0)$. Thus, the first-order derivative of P_0/z^* can then be expressed as

$$\frac{\mathrm{d}\,(P_0/z^*)}{\mathrm{d}\,P_0} = \frac{(\beta_1 + \beta_2)P_0}{(\beta_1 + \beta_2)P_0 - z^* + (\beta_1 + \beta_2)P_0 z^*}, \quad (71)$$

which is always positive for $z^* \in (0, (\beta_1 + \beta_2)P_0)$. Hence, we can conclude that P_0/z^* monotonically increases with P_0. Then we further prove that $r_{1,1}(\mathbf{p}^*)$ in (52) is also a monotonic increasing function of P_0 according to the monotonicity rule of compound function. Note that the thresholds of the offloading decisions for two users in Theorem 1, i.e., $(\beta_1 + \beta_2)P_0/z^*$ and $(\beta_1 + \beta_2)P_0$, monotonically increase with P_0, which means that $\overline{L}_1^*$ and L_2^* are two non-increasing piecewise functions of P_0 each with two constant values. Therefore, it is natural that $t_{22}^* = L_2^*/r_2^*$ and $\overline{L}_1^*/r_{1,1}(\mathbf{p}^*)$ are two monotonic decreasing functions of P_0. Therefore, we can conclude that the optimal WPT duration t_0^* in (54) is a monotonic increasing function of P_0 for the cases of $\overline{L}_1^* > 0$ or $L_2^* > 0$. When $\overline{L}_1^* = 0$ and $L_2^* = 0$ hold simultaneously, we have $t_1^* = t_{21}^* = t_{22}^* = 0$, and thus t_0^* is fixed as $t_0^* = T$. In conclusion, the WPT duration t_0^* is a monotonic non-decreasing function of P_0.

REFERENCES

[1] "Mobile-edge computing-introductory technical white paper," ETSI, Sophia Antipolis, France, White Paper 1 (Issue 1), Sep. 2014. [Online]. Available: https://portal.etsi.org/portals/0/tbpages/mec/docs/mobile-edge_computing_-_introductory_technical_white_paper_v1%2018-09-14.pdf

[2] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.

[3] S. Barbarossa, S. Sardellitti, and P. D. Lorenzo, "Communicating while computing: Distributed mobile cloud computing over 5G heterogeneous networks," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 45–55, Nov. 2014.

[4] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4569–4581, Sep. 2013.

[5] X. Xiang, C. Lin, and X. Chen, "Energy-efficient link selection and transmission scheduling in mobile cloud computing," *IEEE Wireless Commun. Lett.*, vol. 3, no. 2, pp. 153–156, Apr. 2014.

[6] 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. 1757–1771, May 2016.

[7] Y. Mao, J. Zhang, and K. B. Letaief, "Dynamic computation offloading for mobile-edge computing with energy harvesting devices," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3590–3605, Dec. 2016.

[8] J. Liu, Y. Mao, J. Zhang, and K. B. Letaief, "Delay-optimal computation task scheduling for mobile-edge computing systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 1451–1455.

[9] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 1, no. 2, pp. 89–103, Jun. 2015.

[10] 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.

[11] C. You, K. Huang, H. Chae, and B.-H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1397–1411, Mar. 2017.

[12] R. Kaewpuang, D. Niyato, P. Wang, and E. Hossain, "A framework for cooperative resource management in mobile cloud computing," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 2685–2700, Dec. 2013.

[13] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," *IEEE Trans. Wireless Commun.*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/8234686/

[14] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Stochastic joint radio and computational resource management for multi-user mobile-edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5994–6009, Sep. 2017.

[15] K. Huang and V. K. N. Lau, "Enabling wireless power transfer in cellular networks: Architecture, modeling and deployment," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 902–912, Feb. 2014.

[16] S. Bi, C. K. Ho, and R. Zhang, "Wireless powered communication: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 117–125, Apr. 2015.

[17] 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.

[18] H. Ju and R. Zhang, "User cooperation in wireless powered communication networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Austin, TX, USA, Dec. 2014, pp. 1430–1435.

[19] H. Chen, Y. Li, J. L. Rebelatto, B. F. Uchôa-Filho, and B. Vucetic, "Harvest-then-cooperate: Wireless-powered cooperative communications," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1700–1711, Apr. 2015.

[20] H. Liang, C. Zhong, H. A. Suraweera, G. Zheng, and Z. Zhang, "Optimization and analysis of wireless powered multi-antenna cooperative systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3267–3281, May 2017.

[21] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[22] Y. Liang and V. V. Veeravalli, "Gaussian orthogonal relay channels: Optimal resource allocation and capacity," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3284–3289, Sep. 2005.

[23] X. Cao, F. Wang, J. Xu, R. Zhang, and S. Cui. (Apr. 22, 2017). "Joint computation and communication cooperation for mobile edge computing." [Online]. Available: https://arxiv.org/abs/1704.06777

[24] E. Cuervo *et al.*, "MAUI: Making smartphones last longer with code offload," in *Proc. 8th Int. Mobile Syst., Appl., Services*, San Francisco, CA, USA, 2010, pp. 49–62.

[25] L. Yang, J. Cao, Y. Yuan, T. Li, A. Han, and A. Chan, "A framework for partitioning and execution of data stream applications in mobile cloud computing," *Perform. Eval. Rev.*, vol. 40, no. 4, pp. 23–32, 2013.

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

[27] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Adv. Comput. Math.*, vol. 5, no. 1, pp. 329–359, 1996.

Xiaoyan Hu (S'16) received the M.S. degree in information and communication engineering from Xi'an Jiaotong University, Xi'an, China, in 2016. She is currently pursuing the Ph.D. degree in wireless communications with the Department of Electronic and Electrical Engineering, University College London, London, U.K. Her research interests are in the areas of mobile edge computing, wireless energy harvesting, cooperative communications, and physical-layer security.

Kai-Kit Wong (M'01–SM'08–F'16) received the B.Eng., M.Phil., and Ph.D. degrees in electrical and electronic engineering from The Hong Kong University of Science and Technology, Hong Kong, in 1996, 1998, and 2001, respectively. After graduation, he took up academic and research positions at The University of Hong Kong, Lucent Technologies, Bell-Labs, Holmdel, NJ, USA, the Smart Antennas Research Group of Stanford University, and the University of Hull, U.K. He is currently the Chair of wireless communications, Department of Electronic and Electrical Engineering, University College London, U.K.

His current research centers around 5G and beyond mobile communications, including topics such as massive MIMO, full-duplex communications, millimetre-wave communications, edge caching and fog networking, physical-layer security, wireless power transfer and mobile computing, V2X communications, cognitive radios, fluid antenna communications systems, and remote ECG detection. He was a co-recipient of the 2013 IEEE Signal Processing Letters Best Paper Award and the 2000 IEEE VTS Japan Chapter Award at the IEEE Vehicular Technology Conference in Japan in 2000, and a few other international best paper awards.

He is a fellow of IET and is also on the editorial board of several international journals. He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 2009 to 2012 and an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2005 to 2011. He was also a Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Special Issue on Virtual MIMO in 2013, and he is currently a Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Special Issue on Physical Layer Security for 5G. He has been serving as a Senior Editor for the IEEE COMMUNICATIONS LETTERS since 2012 and also for the IEEE WIRELESS COMMUNICATIONS LETTERS since 2016.
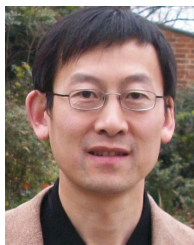
Kun Yang (SM'08) received the M.Sc. and B.Sc. degree from the Computer Science Department, Jilin University, China, and the Ph.D. degree from the Department of Electronic and Electrical Engineering, University College London (UCL), U.K. He was with UCL on several European Union (EU) research projects for several years. He is currently a Chair Professor with the School of Computer Science and Electronic Engineering, University of Essex, U.K., where he is also leading the Network Convergence Laboratory. He is also an Affiliated Professor at UESTC, China. He manages research projects funded by various sources, such as UK EPSRC, EU FP7/H2020, and industries. He has published 100 journal papers. His main research interests include wireless networks and communications, data and energy integrated networks, and computation-communication cooperation. He has been a fellow of IET since 2009. He serves on the editorial boards of both IEEE and non-IEEE journals.