Energy-Efficient Beamforming and Resource Allocation for Multi-Antenna MEC Systems

HUNWOO LIM AND TAEWON HWANG
School of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, South Korea
Corresponding author: Taewon Hwang (twhwang@yonsei.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant by the Korean Government under Grant 2019R1A2C1009826 and Grant 2021R1F1A1063934.

ABSTRACT This paper studies an energy-efficient beamforming and resource allocation for multi-access edge computing (MEC) systems consisting of multi-antenna access points (APs) and single-antenna users. We consider maximizing energy efficiency (EE) of a MEC system, defined as the total computed bits per total energy consumption of the MEC system. To enhance the EE performance, we employ multiple antennas at APs to exploit multiplexing- or (receive) beamforming-gain in the uplink and perform download beamforming for transmitting computation results in the downlink. We consider both spatial-division multiple access (SDMA) based MEC system and time-division multiple access (TDMA) based MEC system and compare their EE performance. We formulate EE maximization problems for the SDMA-based and TDMA-based MEC systems, which are nonconvex and thus cannot be solved by standard convex optimization techniques. We first transform the problems by applying semidefinite relaxation (SDR). Then, we solve the relaxed problems for the SDMA-based MEC system and TDMA-based MEC system by using Dinkelbach method and difference-of-concave programming algorithm. We observe that the SDMA-based MEC system outperforms the TDMA-based MEC system in EE performance most of the cases. However, for the special case when i) only users’ energy consumption is counted (i.e., energy consumption of AP and MEC server is not considered) for EE, and ii) there is no minimum throughput requirement, we show that the TDMA-based MEC system outperforms the SDMA-based MEC system. Simulation results demonstrate that the proposed schemes significantly enhance the EE of MEC systems.

INDEX TERMS Internet of Things (IoT), multi-access edge computing (MEC), energy efficiency (EE), spatial-division multiple access (SDMA), time-division multiple access (TDMA).

I. INTRODUCTION
With recent advancement of Internet of Things (IoT) and increasing popularity of mobile devices, including smart phones, wearable devices, and smart sensors, users expect to be able to run a wide-range of new applications such as face recognition, virtual reality, and unmanned driving on their devices. However, since mobile devices are equipped with relatively low-performance processor and capacity-limited battery due to their size constraint, it is challenging for them to execute these emerging applications, which usually require highly intensive computation and stringent latency. To tackle these problems, multi-access edge computing (MEC) has appeared as a promising approach [1]–[3]. MEC can significantly enhance the computing capability of mobile devices by allowing the devices to offload their computation-intensive tasks to nearby MEC servers located at the network edge, e.g., base stations (BSs) and Wi-Fi access points (APs). Then, the MEC servers compute the offloaded computation tasks remotely and return the result of computation to the devices. In the MEC literature, joint allocation of the communication and computation resources has been intensively studied [4]–[14]. In [4], the computation rate was maximized in a wireless powered MEC system under the binary computation offloading mode. The authors in [5] proposed a distributed computation offloading scheme by using game theory to minimize latency and energy consumption of users. The authors in [6] jointly optimized offloading ratio, transmission power, and CPU frequency for minimizing energy consumption or latency in a single-user MEC system.
Most of the existing works on MEC [4]− [14] focused on minimizing energy consumption, minimizing computation latency, or maximizing computation rate without considering energy efficiency (EE) [15]− [18]. EE is increasingly important for wireless networks due to economic reason, environmental concern, and limited battery capacity. Originally, EE has been defined for communication systems as the transmission bits per energy consumption. Extended from EE for communication systems, EE for MEC systems is defined as the achievable computation bits per energy consumption, where the computation bits can be achieved by both local computing and offloading. The work in [19] maximized the sum of the individual users’ EEs for a time-division multiple access (TDMA) based MEC system, where multiple users offload their computation tasks to an MEC server in a time-division manner. The authors in [20] investigated a system-wise EE maximization problem of a wireless powered MEC network, where non-orthogonal multiple access (NOMA) is employed for uplink task offloading. However, these works on enhancing the EE of MEC systems are limited to the scenarios with a single-antenna AP, which cannot exploit the advantages of multiple antennas in offloading and downloading efficiency.

In this paper, we propose EE maximization schemes for multi-antenna MEC systems. By employing multiple antennas at APs, we can exploit multiplexing or beamforming gain to increase the uplink throughput for offloading and the downlink throughput for computation-result transmission. In our work, we first consider spatial-division multiple access (SDMA) for a multiple accessing scheme. It is well-known that SDMA can improve spectral efficiency by using multiple antennas, but it is not clear if SDMA can still improve EE compared with TDMA for multi-antenna MEC systems. Hence, we also consider TDMA-based MEC system and compare its EE performance with SDMA-based MEC system.

The main contributions of this paper are summarized as follows:

- We consider multi-antenna MEC systems and formulate EE maximization problems to find the jointly optimal beamforming and communication/computation resource allocation schemes. The EE maximization problems are non-convex and thus challenging to obtain the optimal solutions. In particular, compared to the existing works on MEC, which ignored the process of computed results downloading [4], [19], [20] or considered only the transmit power of the AP for the downloading process [23], the SDMA-based MEC system in our work explicitly optimizes downlink beamforming vector as well as transmit power of the AP for transmitting computed results and thus the optimization problems are more challenging. To address the challenge, we first resolve the non-convexity issue due to the beamforming vectors by applying semidefinite relaxation (SDR), while guaranteeing the tightness of the relaxed problems. Then, we divide the problems into their inner and outer problems. We first obtain the optimal solutions of the inner problems in closed-forms. Then, we solve the outer problems by using the Dinkelbach method and the difference-of-concave (DC) programming algorithm. We show that employing multiple antennas at MEC systems can significantly improve the EE performance of MEC systems.

- We consider SDMA as well as TDMA for offloading in multi-antenna MEC systems and compare their EE performance. This is different from the existing works on MEC systems that maximize EE which considered either TDMA [19] or NOMA [20] for offloading in a single-antenna MEC system. We observe that the SDMA-based MEC system in general outperforms the TDMA-based MEC system in EE performance. However, when i) only users’ energy consumption is counted (i.e., energy consumption of AP and MEC server is not considered) for EE, and ii) there is no minimum throughput requirement, we analytically show that the TDMA-based MEC system performs better than the SDMA-based MEC system.

- We include the practical parameters in our system model which are commonly neglected in the existing works on MEC. First, we include the computation time of the MEC server and the computation-result downloading time into our time model. Since EE is adversely affected by the energy consumption of the MEC server and AP during MEC computation and downloading, respectively, it may not be always optimal to reduce the computing and downloading time by increasing the power for MEC computation and downloading from AP. Under this non-negligible time model, we designed the timeline for the TDMA-based MEC system such that the MEC server can compute users’ tasks while it is receiving tasks offloaded by other users or transmitting computed results to them, which can improve the efficiency of the MEC system. Second, we include non-negligible energy consumption by leakage currents at the local processors of the users. Under this energy model, we show that using all the available time for local computing is not necessarily optimal to maximize EE, which is not true when there is no leakage-current energy consumption. Accordingly, we optimize local computing times of the users, which makes the problems more difficult.

The rest of this paper is organized as follows. Section II introduces the system models of our MEC systems. Section III and Section IV investigate the EE maximization problems for the SDMA-based MEC system and the TDMA-based MEC system, respectively. Section V provides simulation results to show the performances of the proposed EE maximization schemes. Finally, Section VI concludes the paper.

**Notation:** Scalars are denoted by lower-case letters, vectors by bold-face lower-case letters, and matrices by bold-face upper-case letters. \( \textbf{I} \) and \( \textbf{0} \) denote an identity matrix and
all-zero matrix, respectively, with appropriate dimensions. For a vector \( \mathbf{x} \), \( \mathbf{x} \succeq \mathbf{0} \) means that \( \mathbf{x} \) is element-wise non-negative. For a square matrix \( \mathbf{S} \), \( \text{tr}(\mathbf{S}) \) denotes the trace of \( \mathbf{S} \), while \( \mathbf{S} \succeq \mathbf{0} \) and \( \mathbf{S} \preceq \mathbf{0} \) mean that \( \mathbf{S} \) is positive semidefinite and negative semidefinite, respectively. For a matrix \( \mathbf{M} \) of arbitrary size, \( \text{rank}(\mathbf{M}) \), \( \mathbf{M}^T \), and \( \mathbf{M}^H \) denote the rank, transpose, and conjugate transpose of \( \mathbf{M} \), respectively. \( \mathbb{C}^{x \times y} \) denotes the space of \( x \times y \) complex matrices. \( \| \mathbf{x} \| \) denotes the Euclidean norm of a complex vector \( \mathbf{x} \), and \( |z| \) denotes the magnitude of a complex number \( z \).

II. SYSTEM MODEL

We consider a multi-antenna MEC system, which consists of an \( M \)-antenna AP and \( K \) single-antenna users, denoted by \( U_i \) for \( i \in \mathcal{K} \triangleq \{1, \ldots, K\} \), as shown in Fig. 1. The AP is integrated with an MEC server which can execute the computation tasks offloaded from the \( K \) users. The system operates using a slotted communication protocol where each block has time duration \( T \). The uplink/downlink channels between the AP and the users can be assumed to be with block fading, where they remain unchanged during a block duration \( T \). The uplink channel from \( U_i \) to the AP and the downlink channel from the AP to \( U_i \) are denoted by \( \mathbf{g}_i \in \mathbb{C}^{M \times 1} \) and \( \mathbf{h}_i \in \mathbb{C}^{M \times 1} \), respectively. We assume that the AP perfectly knows the channels \( \mathbf{g}_i \)'s and \( \mathbf{h}_i \)'s.\(^1\) Also, we assume that computation offloading in the uplink, and computed results downloading in the downlink are operated over the same frequency band.

In each block, each \( U_i \) executes its computation task via both local computing and offloading. Here, we assume that the computation bits of each user’s task can be partitioned into: 1) local computing bits to be computed locally and 2) offloading bits to be offloaded to the MEC server. Each user processes the local computing bits by operating its own processor and transmits the offloading bits to the AP. Then, the MEC server (collocated with the AP) processes the offloading bits. Therefore, the number of bits offloaded from the users to the MEC server is equal to the number of computation bits processed by the MEC server. Finally, the AP transmits the computed results of the MEC server to the users. In the following subsections, we derive the EEs for the TDMA-based MEC system and the SDMA-based MEC system.

A. LOCAL COMPUTING

The computing unit of each user is assumed to be separate from the communication circuit. Therefore, each user can perform local computing and task offloading at the same time. Let \( \tau_i^l \) denote the local computing time of \( U_i \), then it follows that \( 0 \leq \tau_i^l \leq T \). Let \( C_i \) and \( f_i \) denote the number of CPU cycles for computing one input bit of \( U_i \)'s task and the CPU frequency of \( U_i \), respectively. Then, the executed computation bits and consumed energy by local computing at \( U_i \) during \( \tau_i^l \) can be written as

\[
B_{loc,i} = \frac{f_i \tau_i^l}{C_i}
\]

and

\[
E_{loc,i} = (k_i f_i^3 + p_i^{l,c}) \tau_i^l,
\]

respectively [26], [27], where \( k_i \) is the effective capacitance coefficient of the processor’s chip at \( U_i \) and \( p_i^{l,c} \) is the power consumed by leakage currents.

B. TASK OFFLOADING

In addition to local computing, the users can offload their tasks to the MEC server by transmitting the input bits of their tasks to the AP. For the offloading and downloading of the users’ tasks, we consider two multiple access schemes: SDMA and TDMA.

1) SDMA-BASED OFFLOADING AND DOWNLOADING

With SDMA, users can perform multiple accessing in a simultaneous way, and times of \( \tau_i^u \in (0, T) \) and \( \tau_i^d \in (0, T) \)

\(^1\) In practice, the AP needs to transmit pilot signals and receive feedback signals from the users to obtain channel state information. The detailed procedure for estimating channels and optimizing pilot transmission is provided in [25].
are allocated for offloading and downloading the users’ data, respectively, as shown in Fig. 2(a). Let \(p_i^c \in [0, P_i]\) denote the transmit power of \(U_i\) for offloading, where \(P_i\) is the maximum transmit power of \(U_i\). Then, the energy consumed by \(U_i\) for task offloading is given by

\[
E_{\text{off},i}^{\text{sdma}} = (p_i^u + p_i^c) \tau^u,
\]

where \(p_i^u, p_i^c\) is the circuit power consumed by \(U_i\) in its RF chains, baseband signal processing, etc., which is independent of the transmit power \(p_i^u\). Here, we assume that the AP employs the zero-forcing receiver to decode the users’ signals, where \(G = [g_1, \ldots, g_K]\). Then, the number of bits that \(U_i\) can transmit by computation offloading is given by

\[
B_{\text{off},i}^{\text{sdma}} = W \tau^u \log_2 \left(1 + \gamma_i^{\text{sdma}} p_i^u \right),
\]

where \(W\) is the system bandwidth and \(\gamma_i^{\text{sdma}} = \frac{1}{\sigma^2 \| \mathbf{z}_i^H \|^2}\) is the effective channel gain of \(U_i\)’s signal with \(\sigma^2\) and \(\mathbf{z}_i^H\) denoting the noise power at the receiver of the AP and the \(i\)th row vector of \(\mathbf{Z}\), respectively. During offloading time \(\tau^u\), the AP consumes energy \(M p_A^c \tau^u\) for receiving the offloaded bits, where \(p_A^c\) is the circuit power consumption per AP antenna to receive signal [17]. Then, the MEC server computes \(U_i\)’s task and generates the output result of \(O_i B_{\text{off},i}^{\text{sdma}}\) bits, where \(O_i\) is the number of output bits per input bit. The energy and time consumed by the MEC server for executing \(U_i\)’s task can be written as

\[
E_i^c = k_M C_i B_{\text{off},i}^{\text{sdma}} r^2 = \xi_i \tau^u \log_2 \left(1 + \gamma_i^{\text{sdma}} p_i^u \right)
\]

and

\[
\tau_i^c = \frac{C_i B_{\text{off},i}^{\text{sdma}}}{f_M} = \frac{C_i W \tau^u \log_2 \left(1 + \gamma_i^{\text{sdma}} p_i^u \right)}{f_M},
\]

respectively, where \(\xi_i = k_M C_i W f_M^2\) with \(k_M\) depending on the chip architecture of the MEC server.

The downlink signal from the AP to the users for transmitting the computation results can be written as \(x_0 = \sum_{j=1}^{K} v_j s_j\), where \(v_j \in \mathbb{C}^{M \times 1}\) is the beamforming vector for \(U_i\) and \(s_j\) is its corresponding data signal. Then, the energy consumed by the AP for transmitting the results can be written as \(\tau^d (\sum_{j=1}^{K} \| v_j \|^2 + M p_A^c)\), where \(p_A^c\) is the circuit power consumption per AP antenna to transmit signal. Also, the number of bits that the AP can transmit to \(U_i\) during downloading time \(\tau^d\) is given by

\[
B_{\text{dn},i}^{\text{sdma}} = W \tau^d \log_2 \left(1 + \frac{\| v_i^H \|_2^2}{\sum_{k \neq i} \| v_k^H \|_2^2 + \sigma^2} \right).
\]

The amount of energy consumed by \(U_i\) for receiving the computed results can be written as \(p_i^c \tau^d\), where \(p_i^c\) is the circuit power consumed by \(U_i\) for receive signal processing.

From above, the EE of the SDMA-based MEC system, which is defined as the ratio of the total achievable computation bits to the total energy consumption of the MEC system, can be expressed as \(\beta\), shown at the bottom of the next page. In (8), \(\beta \geq 0\) is a weight that controls the contribution of the energy consumption of the AP and MEC server to that of the whole MEC system in computing EE. For example, \(\beta = 0\) indicates that the energy consumption of the AP and MEC server is not considered.

2) TDMA-BASED OFFLOADING AND DOWNLOADING

For TDMA, each offloading time \(\tau_i^d \in (0, T)\) is allocated to each user \(U_i\). Different from the existing works on MEC, we consider non-negligible computation time of the MEC server in our system model. Under this assumption, the MEC server for the TDMA-based system can compute users’ tasks while it is receiving tasks from other users as shown in Fig. 2(b). However, the MEC server can begin computing \(U_i\)’s task only after it i) receives all the input bits of \(U_i\)’s task and ii) completes computing \(U_{i-1}\)’s task, i.e.,

\[
\begin{align*}
\tau_i^u &\leq \tau_i^c, \\
\max \left(\tau_i^c + \tau_{i-1}^c, \sum_{j=1}^{K} \tau_j^d\right) &\leq \tau_i^d, \\
\tau_i^d &\leq T,
\end{align*}
\]

where \(\tau_i^u\) and \(\tau_i^c\) denote the time instance at which the MEC server can begin computing \(U_i\)’s task and time allocated to \(U_i\)’s task, respectively. Similarly, the AP can begin transmitting \(U_i\)’s computed results after it i) completes computing \(U_i\)’s task and ii) completes transmitting \(U_{i-1}\)’s computed results (or receives all users’ tasks for \(i = 1\), i.e.,

\[
\begin{align*}
\max \left(\tau_i^c + \tau_j^c, \sum_{j=1}^{K} \tau_j^d\right) &\leq \tau_i^d, \\
\max (\tau_i^d + \tau_i^c, \tau_{i-1}^d + \tau_{i-1}^c) &\leq \tau_i^d, \\
\tau_i^d &\leq T,
\end{align*}
\]

where \(\tau_i^d\) and \(\tau_i^d\) denote the time instance at which \(U_i\) can begin downloading its computed results and time allocated to \(U_i\) for downloading its computed results, respectively. Since the last user \(U_K\) should complete downloading its computed results within \(T\), we have

\[
\tau_K^d + \tau_K^c \leq T.
\]

The energy consumed by \(U_i\) for task offloading is given by

\[
E_{\text{off},i}^{\text{sdma}} = (p_i^u + p_i^c) \tau_i^u,
\]

Also, the number of bits that \(U_i\) can transmit by task offloading is given by

\[
B_{\text{off},i}^{\text{sdma}} = W \tau_i^d \log_2 \left(1 + \gamma_i^{\text{sdma}} p_i^u \right),
\]

where \(\gamma_i^{\text{sdma}} \triangleq \frac{\| v_i^H \|_2^2}{\sigma^2}\) is \(U_i\)’s uplink channel signal-to-noise ratio (SNR). During offloading time \(\sum_{j=1}^{K} \tau_j^d\), the AP consumes energy \(\sum_{j=1}^{K} M p_A^c \tau_j^d\) for receiving the offloaded bits. Once the AP receives input bits of \(U_i\), the MEC server computes the task and generates the output result of \(O_i B_{\text{off},i}^{\text{sdma}}\) bits. Assuming a constant CPU frequency \(f_M\) of the MEC server,\(^2\) the energy and time consumed by the MEC server for executing \(U_i\)’s task can be written as \(f_M\).

\(^2\)Note that the optimal \(f_M\) can be found by a one-dimensional search after solving the EE maximization problems for a given PD.
\[ E_i^c = kM C_i^\text{ldma} \tau_i^2 = \xi_i \tau_i^u \log_2 \left( 1 + \gamma_i^\text{ldma} p_i^u \right) \] (14)

and

\[ r_i^c = \frac{C_i P_i^\text{ldma}}{f_M} = \frac{C_i W \tau_i^u \log_2 \left( 1 + \gamma_i^\text{ldma} p_i^u \right)}{f_M}, \] (15)

respectively. The number of bits that the AP can transmit to \( U_i \) during downloading time \( r_i^d \) is given by

\[ P_{\text{ldma},i} = W \tau_i^d \log_2 \left( 1 + \gamma_{\text{ldma},i} p_i^d \right). \] (16)

with consumed energy \( \sum_{i=1}^{K} r_i^d p_i^d \), where \( \gamma_{\text{ldma},i} \triangleq \frac{[h_i]^2}{\sigma^2} \) is \( U_i \)'s downlink channel SNR. During downloading time \( r_i^d \), \( U_i \) consumes energy for receiving the results of the TDMA, which can be written as \( p_i^f \), \( \tau_i^d \). From above, the EE of the TDMA-based MEC system can be written as (17), shown at the bottom of the page.

Now, we formulate and investigate EE maximization problems for both the SDMA-based MEC system and the TDMA-based MEC system in the following sections.

### III. EE Maximization for the SDMA-Based MEC System

In this section, we study a EE maximization problem for the SDMA-based MEC system and find its optimal solution. Specifically, we jointly optimize downloading time \( r_i^d \), offloading time \( r_i^u \), offloading transmit power \( p_i^d \) of the users, local CPU frequency \( f_i \), local computing time \( \tau_i^d \), and downlink beamforming vector \( \mathbf{v}_i \) for \( i \in \mathcal{K} \) to maximize EE. Let \( x = \{\tau_i^d, \tau_i^u, p_i^d, f_i, \tau_i^d, \mathbf{v}_i \mid i \in \mathcal{K}\} \), then the problem is

\[
\begin{align*}
\max_{x} & \quad \eta_i^\text{sdma}(x) \\
\text{s.t.} & \quad (k f_i^3 + p_i^f + p_i^f c) \tau_i^d + (p_i^d + p_i^f c) \tau_i^u + p_i^f c \tau_i^d \leq E_i, \quad i \in \mathcal{K}, \\
& \quad \frac{f_i \tau_i^d}{C_i} + W \tau_i^u \log_2 \left( 1 + \gamma_i^\text{sdma} p_i^u \right) \geq R_{\text{min},i}, \quad i \in \mathcal{K}, \\
& \quad \tau_i^d \log_2 \left( 1 + \frac{[h_i]^2}{\sum_{k \neq i} [h_k]^2 + \sigma^2} \right) \geq O_i \tau_i^d \log_2 \left( 1 + \gamma_i^\text{sdma} p_i^u \right), \quad i \in \mathcal{K},
\end{align*}
\] (18a)

\[
\begin{align*}
\eta_i^\text{sdma} &= \sum_{i=1}^{K} \left[ \frac{f_i \tau_i^d}{C_i} + W \tau_i^u \log_2 \left( 1 + \gamma_i^\text{sdma} p_i^u \right) \right] \\
&\leq \sum_{i=1}^{K} \left[ \left( k f_i^3 + p_i^f + p_i^f c \right) \tau_i^d + \left( p_i^d + p_i^f c \right) \tau_i^u + p_i^f c \tau_i^d \right] + \beta \sum_{i=1}^{K} \left[ \xi_i \tau_i^u \log_2 \left( 1 + \gamma_i^\text{sdma} p_i^u \right) + \tau_i^d \left( \|v_i\|^2 + \frac{M p_i^f c}{K} + \frac{M p_i^f c}{K} \right) \right],
\end{align*}
\] (8)

where \( \xi_i \) is the maximum available energy of \( U_i \), \( R_{\text{min},i} \) is the minimum computation bit requirement for \( U_i \), and \( P_A \) is the maximum transmit power of the AP. Besides, (18b), (18c), (18d), (18e), and (18g) are the energy constraints of the users, the minimum computation bit constraints of the users, the download bits causality constraint, the transmit power constraint of the AP, and the total time constraint, respectively.

Note that problem (18) is non-convex due to the fractional-form objective function and the non-convex terms including the product terms of control variables in (18a)–(18h).

To resolve this non-convex issue, we introduce the auxiliary variables i) \( t_i^d \), \( t_i^u \) with constraints in (19a) and (19b), ii) \( e_i^u = \tau_i^u p_i^u \), iii) \( b_i \) to replace \( \tau_i^d \log_2 \left( 1 + \gamma_i^\text{sdma} p_i^u \right) \) with additional constraints in (19c), iv) \( \Gamma_i = \text{tr}(H_i S_i) \) with constraints in (19a), and v) \( \Gamma_{ij} = \sum_{i \neq j} \text{tr}(H_i S_j) \) with constraints in (19f) for \( i \in \mathcal{K} \), where \( H_i = h_i h_i^H \). Then, problem (18) can be rewritten as

\[
\begin{align*}
\max_{\mathbf{y}} & \quad \eta_i^\text{sdma}(\mathbf{y}) \\
\text{s.t.} & \quad (k f_i^3 + p_i^f + p_i^f c) t_i^d + e_i^u + p_i^f c t_i^u + p_i^f c t_i^d \leq E_i, \quad i \in \mathcal{K}, \\
& \quad \frac{f_i t_i^d}{C_i} + W t_i^u \log_2 \left( 1 + \gamma_i^\text{sdma} p_i^u \right) \geq R_{\text{min},i}, \quad i \in \mathcal{K}, \\
& \quad t_i^d \log_2 \left( 1 + \frac{[h_i]^2}{\sum_{k \neq i} [h_k]^2 + \sigma^2} \right) \geq O_i t_i^d \log_2 \left( 1 + \gamma_i^\text{sdma} p_i^u \right), \quad i \in \mathcal{K}, \\
& \quad t_i^u \log_2 \left( 1 + \frac{\gamma_i^\text{sdma} e_i^u}{t_i^u} \right) \geq b_i, \quad i \in \mathcal{K}
\end{align*}
\] (19a)

\[
\eta_i^\text{sdma} = \sum_{i=1}^{K} \left[ \frac{f_i t_i^d}{C_i} + W t_i^u \log_2 \left( 1 + \gamma_i^\text{sdma} p_i^u \right) \right] \\
= \sum_{i=1}^{K} \left[ \left( k f_i^3 + p_i^f + p_i^f c \right) t_i^d + \left( p_i^d + p_i^f c \right) t_i^u + p_i^f c t_i^d \right] + \beta \sum_{i=1}^{K} \left[ \xi_i t_i^u \log_2 \left( 1 + \gamma_i^\text{sdma} p_i^u \right) + t_i^d \left( \|v_i\|^2 + \frac{M p_i^f c}{K} + \frac{M p_i^f c}{K} \right) \right].
\] (17)
The inner problem in (20) is non-convex. However, we always obtain the optimal solution of problem (19), we can always obtain the optimal solution of problem (18) using the relation 

\[ S_i = \tau_i^d v_i^H h_i, \quad p_i^u = \frac{e_i}{\tau_i^d}, \quad \tau_i = \text{tr}(H_i S_i) = |h_i| h_i^H v_i|^2, \quad \text{and} \quad \Gamma_i = \text{tr}(H_i S_i) = |h_i|^2, \]

achieves the same optimal value while satisfying all the constraints in (18b) − (18h).

Now, it is clear that the constraint functions in (19c)−(19k) are convex. However, due to the product terms of \( f_i \) and \( \tau_i^d \) in (19a)−(19c) and the rank constraints in (19l), problem (19) is still non-convex. To tackle this issue, we apply SDR technique [29], which ignores the rank constraints in (19l), thereby causing a feasible set expansion. Therefore, the optimal value of problem (19) without (19l) serves as an upper bound on that of problem (19) in general. However, after optimizing \( S_i \)s of the relaxed problem, we will verify that \( S_i \)'s still satisfy the rank constraints as will be shown in Theorem 2. Besides, by introducing local computing energy \( e_i^L \), we divide problem (19) into the inner problem in (20) and the outer problem in (21). The inner problem optimizes local computing frequency \( f_i \) and local computing time \( \tau_i^d \) for a given local computing energy \( e_i^L \) to maximize local computing bits \( b_i^L \), while the outer problem optimally controls \( e_i^L \) to maximize the EE. Specifically, the inner problem can be written as

\[
\max_{f_i, \tau_i^d} \frac{f_i \tau_i^d}{C_i} \quad \text{s.t.} \quad (k d_i^j + p_i^L c_i) \tau_i^d \leq e_i^L, \quad f_i \geq 0, \quad 0 \leq \tau_i^d \leq T. \tag{20a}
\]

while the outer problem can be written as

\[
\max \sum_{i=1}^{K} A_i(z) \quad \text{s.t.} \quad \sum_{i=1}^{K} p_i^L e_i^L \leq E_i, \quad i \in K, \tag{21a}
\]

\[
b_i^L (e_i^L) + W b_i \geq R_{\min, i}, \quad i \in K, \quad \tau_i^d \geq 0, \quad \tau_i^d \leq T, \quad e_i^L \geq 0, \quad i \in K \tag{21b}
\]

\[ (19d) - (19k) \]

where \( z = (\tau_i^d, \tau_i^u, e_i^L, e_i^L, b_i, \Gamma_i, \Gamma_i, S_i) | i \in K \) and

\[ A_i(z) = b_i^L (e_i^L) + W b_i, \quad b_i^L (e_i^L) = e_i^L + p_i^L c_i \tau_i^d + p_i^L c_i \tau_i^d, \quad \beta_0 \left( c_i b_i + \text{tr}(S_i) + \frac{M P_i^L c_i \tau_i^d + M P_i^L c_i \tau_i^d}{K} \right). \tag{22}\]

Here, we briefly show the equivalence between problem (19) and its inner/outdoor problem in (20) and (21). Once we obtain the optimal solution \( z^* = \{ (\tau_i^d)^*, (\tau_i^u)^*, (e_i^L)^*, (e_i^L)^*, b_i^*, \Gamma_i^*, S_i^* \} | i \in K \) of the outer problem in (21), we can obtain a solution of problem (19), denoted by \( z^* = (\tau_i^d)^*, (\tau_i^u)^*, (e_i^L)^*, (e_i^L)^*, b_i^*, \Gamma_i^*, S_i^* \) where \( f_i^* \) and \( (\tau_i^d)^* \) is the optimal solution of the inner problem in (20) with \( (e_i^L)^* \). Then, the optimal value of problem (19) achieved by \( y^* \) is same as that of problem (21) achieved by \( z^* \).

The inner problem in (20) is non-convex. However, we derive its optimal solution in a closed-form as stated in the following theorem.

**Theorem 1:** The optimal value of problem (20) is given by

\[
b_i^L (e_i^L) = \left\{ \begin{array}{ll}
\frac{1}{C_i} \left( \frac{p_i^L c_i}{2k_i} \right)^{1/3} \frac{e_i^L}{2 p_i^L c_i} & \tau_i^d < 3 \frac{p_i^L c_i}{2} T \\
\frac{e_i^L}{T} \left( \frac{3 p_i^L c_i}{k_i T} - \frac{p_i^L c_i}{k_i} \right)^{1/3} & \tau_i^d \geq 3 \frac{p_i^L c_i}{2} T
\end{array} \right. \tag{22}
\]

Moreover, \( b_i^L (e_i^L) \) is a concave function.

**Proof:** Please refer to Appendix A.

The outer problem in (21) is a fractional programming. Since \( b_i^L (e_i^L) \) is a concave function and the denominator of the objective function in (21a) is affine, the objective function in (21a) is quasi-concave. Now, we find the optimal solution of the outer problem in (21) by using the Dinkelbach method [30]. The procedure to solve problem (21) is provided in the following. According to [30], the fractional programming problem in (21) is equivalent to the following substitute-form problem with \( q = q^* \):

\[
\max \sum_{i=1}^{K} A_i(z) - q \sum_{i=1}^{K} B_i(z) \tag{23a}
\]

\[
\text{s.t.} \quad \tau_i^d \log_2 \left( \frac{\Gamma_i}{\tau_i^d} + \frac{\Gamma_{i,i}^*}{\tau_i^d} + \sigma^2 \right) - Q_i(d_i) \geq O_{i} b_i, \quad i \in K \tag{23b}
\]

\[
(21b) - (21d), \quad (19e) - (19k) \tag{23c}
\]

where \( Q_i(d_i) = \tau_i^d \log_2 \left( \frac{\Gamma_i^*}{\tau_i^d} + \sigma^2 \right) \) for \( i \in K \) with \( d_i = \Gamma_{i,i}^* \). Here, \( q^* \) denotes the optimal value of
problem (21) and makes the optimal value of problem (23) to be zero. To obtain \( q^* \) and the corresponding optimal solution \( z^* \), we repeat the following steps until the optimal value of problem (23) converges to zero [30].

**Step 1** For a given \( q \), solve problem (23) and obtain its solution \( z \).

**Step 2** Update

\[
q = \sum_{i=1}^{K} A_i(z) - \sum_{i=1}^{K} B_i(z) \tag{24}
\]

by using \( z \) obtained from **Step 1**.

In the following, we discuss the details of **Step 1**, i.e., the approach to solve problem (23) for a given \( q \). Now, the objective function of problem (23) is concave, but problem (23) is still non-convex due to the non-concave functions in (23b). However, since the non-concave functions have DC structures, problem (23) is a DC programming [35]. Therefore, we use the DC algorithm [35] to solve problem (23). The DC algorithm is guaranteed to converge to the stationary points that satisfy the Karush-Kuhn-Tucker (KKT) conditions of the original problem [36], [37]. Although the DC algorithm does not guarantee a global optimality of solution theoretically, it often provides a global optimal solution in practice [38].

The basic idea of the DC algorithm is to iteratively solve approximate convex problems, which can be obtained by approximating the second concave functions in the DC structured functions with affine functions. Specifically, to solve problem (23), we iteratively solve the following problem:

\[
\begin{align*}
\max_{z} & \quad \sum_{i=1}^{K} A_i(z) - q \sum_{i=1}^{K} B_i(z) \\
\text{s.t.} & \quad \tau^d \log_2 \left( \frac{\Gamma_i}{\tau^d} + \frac{\Gamma_{i,l}}{\tau^d} + \sigma^2 \right) \\
& \quad - \hat{Q}_i(d_i; d_i^{(l)}) \geq Q_i(b_i), \quad i \in \mathcal{K} \\
& \quad (21b) - (21d), \quad (19c) - (19k) \tag{25a}
\end{align*}
\]

where \( l \) is an iteration index, and

\[
\hat{Q}_i(d_i; d_i^{(l)}) = Q_i(d_i^{(l)}) + \left[ \nabla Q_i(d_i^{(l)}) \right]^T (d_i - d_i^{(l)}), \quad i \in \mathcal{K} \tag{25b}
\]

are affine functions with \( d_i^{(l)} = \left[ \Gamma_{i,j}^l (\tau^d)^l \right]^T \). Here, \( \nabla Q_i(d_i^{(l)}) = [\nabla Q_{i,1}^{(l)}, \nabla Q_{i,2}^{(l)}]^T \) is the gradient of \( Q_i \) with respect to \( d_i \) at \( d_i^{(l)} \). We iteratively solve problem (25) until the solution converges, i.e., \( |z_i^{(l+1)} - z_i^{(l)}| < \epsilon \), where \( z_i^{(l)} \) is the optimal solution of problem (25) at the \( l \)th iteration and \( \epsilon \) is the stopping criterion. Since problem (25) is convex, we solve it by using the Lagrange dual method. The partial Lagrangian [31] of problem (25) can be written as (27), shown at the bottom of the page, where \( \psi = \{ \lambda_i, v_i, \theta_i, \delta_i, \alpha, \mu, i, \phi_i | i \in \mathcal{K} \} \) is a set of the nonnegative Lagrange multipliers corresponding to the constraints in (21b), (21c), (23b), (19e), (19f), (19h), (19i), and (19j), respectively. Then, the dual problem of problem (25) is

\[
\min_{\psi \geq 0} g(\psi), \tag{28}
\]

where \( g(\psi) \) is the dual function and can be obtained by solving the following problem:

\[
\begin{align*}
g(\psi) &= \max_z L(z, \psi) \tag{29a} \\
& \quad \text{s.t.} (19g), \quad (19k), \quad (21d) \tag{29b}
\end{align*}
\]

Due to the convexity of problem (29), we can find the optimal \( z \) by using the Karush-Kuhn-Tucker (KKT) conditions. The optimal \( (e_i^*) \) can be obtained by solving \( \frac{\partial L}{\partial e_i^*} = 0 \), i.e.,

\[
\tau^d \int_e^{e_i^*} \frac{\partial L}{\partial e_i^*} \Big|_{e_i^*} = \frac{\psi_i + v_i}{1 + v_i}. \quad \text{Since} \quad \frac{\partial L}{\partial e_i^*} \quad \text{is a monotonically decreasing function of} \quad e_i^* \quad \text{as stated in Appendix A, we can find (e_i^*) by using the bisection method over e_i^*}. \quad \text{Also, we can obtain the optimal} \quad S_l \quad \text{as shown in the following theorem.}
\]

**Theorem 2:** The optimal \( S_l \) that maximizes \( L(z, \psi) \) is given by

\[
S_l = \sqrt{a_i u_i u_i^H}, \quad \text{(30)}
\]

where \( a_i \geq 0 \) and \( u_i \) is the unit-norm eigenvector of \( B_i \triangleq i H_i - \sum_{k \neq i}^{K} \phi_k H_k \) corresponding to its largest eigenvalue \( \lambda_1(B_i) \).

**Proof:** Please refer to Appendix B. □
From Theorem 2, we can see that the optimal \( S_i = \sqrt{\alpha_i} u_i u_i^H \) is rank-one for \( a_i > 0 \) or rank-zero for \( a_i = 0 \), which satisfies the rank constraint in (19).

Next, we find the optimal \( \Gamma_i \) and \( \Gamma_i^* \) that maximizes the Lagrangian \( L \). The optimization problem for finding the optimal \( \Gamma_i \) and \( \Gamma_i^* \) can be written as

\[
\max_{\Gamma_i, \Gamma_i^*} \sum_{i=1}^K \theta_i \left[ e^d \log_2 \left( \frac{\Gamma_i}{\tau^d} + \frac{\Gamma_i^*}{\tau^d} + \sigma^2 \right) - \nabla Q_{i,i}^{(l)} \right] + \sum_{i=1}^K \left( \phi_i \Gamma_i - t_i \Gamma_i^* \right) \tag{31}
\]

where \( \nabla Q_{i,i}^{(l)} \) is the first element of \( \nabla Q_i \left( d_i^{(l)} \right) \). Define \( \tilde{\Gamma}_i = \frac{\Gamma_i}{\tau} \) and \( \tilde{\Gamma}_i^* = \frac{\Gamma_i^*}{\tau} \) for \( i \in \mathcal{K} \). Then, problem (31) can be reformulated as

\[
\max_{\tilde{\Gamma}_i, \tilde{\Gamma}_i^*} \sum_{i=1}^K \theta_i \left[ \log_2 \left( \tilde{\Gamma}_i + \tilde{\Gamma}_i^* + \sigma^2 \right) - \nabla Q_{i,i}^{(l)} \right] + \sum_{i=1}^K \left( \phi_i \tilde{\Gamma}_i - t_i \tilde{\Gamma}_i^* \right) \tag{32}
\]

Since problem (32) is convex, we can find its optimal solution by using the gradient ascent method [32], [33].

Besides, from \( \frac{\partial L}{\partial e_i^*} = 0 \) with considering constraints in (19g), we have

\[
(p_i^*)^* = \left( e_i^* \right)^* = \left[ \frac{b_{i}}{\ln(1 \div 2 + x_i \tilde{\tau})} - \frac{1}{\gamma_i^* \tilde{\tau}_i^*} \right]^{p_i}, \tag{33}
\]

where \([x]_{0}^{\mathcal{K}} = \max(0, \min(x, s_U)) \). Using the results above, \( \frac{\partial L}{\partial \tau^d} \) and \( \frac{\partial L}{\partial \tau^u} \) can be written as

\[
\frac{\partial L}{\partial \tau^d} = -q \left( e_i^* \right)^* \tau^d + \frac{\nabla Q_{i,i}^{(l)}}{\ln 2(\tilde{\Gamma}_i + \tilde{\Gamma}_i^* + \sigma^2)} + a \mu \xi_{\mathcal{P}A}^c - \beta M_{\mathcal{P}A}^c \tilde{\tau}_i^d \tag{34}
\]

\[
\frac{\partial L}{\partial \tau^u} = -q \left( e_i^* \right)^* \tau^u + \frac{\nabla Q_{i,i}^{(l)}}{\ln 2(\tilde{\Gamma}_i + \tilde{\Gamma}_i^* + \sigma^2)} + a \mu \xi_{\mathcal{P}A}^c - \beta M_{\mathcal{P}A}^c \tilde{\tau}_i^d \tag{35}
\]

Then, from (34) and (35), we can see that the Lagrangian \( L \) is linear in \( \tau^d \) and \( \tau^u \). Also, from (27), we can easily see that the Lagrangian \( L \) is linear in \( [b_i]_{K_i}^{b_i} \). Therefore, problem (25) is a linear programming (LP) with respect to \( w = \{ \tau^d, \tau^u, b_i \} \). Substituting \( e_i^* = (e_i^*)^* \), \( S_i = \sqrt{\alpha_i} u_i u_i^H \), \( \Gamma_i = \tau^d \Gamma_i^* \),

\[
\Gamma_{i,J} = \tau^d \Gamma_{i,J}^* \text{, and } e_i^* = (p_i^*)^* \tau^u \text{ into problem (25), we have the following LP problem:}
\]

\[
\max_{w, \{u_i\}} \sum_{i=1}^K \left[ b_i^* \left( (e_i^*)^* \right) + Wb_i \right] - q \sum_{i=1}^K \left[ (e_i^*)^* \tau^d + p_i^* \tau^u + p_i^c \tau^u + p_i^c \tau^d \right] + \beta \left( \xi_{b_i} + a_i + \frac{M_{\mathcal{P}A}^c \tau^d + M_{\mathcal{P}A}^c \tau^u}{K} \right) \tag{36a}
\]

s.t. \( (e_i^*)^* \tau^d + p_i^* \tau^u + p_i^c \tau^u + p_i^c \tau^d \leq E_i \), \( i \in \mathcal{K} \) \tag{36b}

\[
b_i^* \left( (e_i^*)^* \right) + Wb_i \geq R_{\text{min}}, \quad i \in \mathcal{K} \tag{36c}
\]

\[
\tau^d \log_2 \left( \tilde{\Gamma}_i + \tilde{\Gamma}_i^* + \sigma^2 \right) - \tilde{Q}_i(\tau^d, d_i^*) \geq \alpha b_i, \quad i \in \mathcal{K} \tag{36d}
\]

\[
\tau^u \log_2 \left( 1 + \gamma_i^s d_{\text{dma}}(p_i^*)^* \right) \geq b_i, \quad i \in \mathcal{K} \tag{36e}
\]

\[
0 \leq a_i \leq \tau^d P_A, \quad i \in \mathcal{K} \tag{36f}
\]

\[
\tau^u + \sum_{i=1}^K \frac{C_i Wb_i}{f_{MT}} + \tau^d \leq T, \tag{36g}
\]

\[
a_i |u_i^H h_i|^2 \geq \tau^d \tilde{\Gamma}_i^*, \quad i \in \mathcal{K} \tag{36h}
\]

\[
\sum_{k \neq i} d_k |u_k^H h_i|^2 \geq \tau^d \tilde{\Gamma}_i^*, \quad i \in \mathcal{K} \tag{36i}
\]

\[
\tau^u \geq 0, \quad \tau^d \geq 0, \tag{36j}
\]

To solve the LP problem in (36), standard linear optimization techniques such as the simplex method [32] can be employed.

Based on the above optimal solution, we can obtain the dual function \( g(\psi) \) for given dual variables \( \psi \). Then, the dual problem in (28) can be solved by using the subgradient method. e.g., the ellipsoid method [34]. With alternate optimization of primal variables \( z \) and dual variables \( \psi \) above, we can obtain the optimal solution of problem (25). The proposed algorithm for solving problem (21) is summarized in Algorithm 1.

The computational complexity for solving problem (21) can be evaluated as follows. Since there are 8\( K \) primal variables, the complexity for obtaining primal variables for a given dual vector \( \psi \) linearly increases with the number of users \( K \), and thus can be written as \( O(K) \). Next, the complexity of the subgradient method is \( O(K^2) \) because the complexity of the subgradient method quadratically increases with the number of control variables [33] and there are 10\( K \) control variables in problem (28). Using the above results and denoting \( N_{\text{dc}} \) and \( N_{\text{df}} \) to be number of the DC algorithm iterations and the Dinkelbach iterations, respectively, the overall complexity for solving problem (21) is given by \( O(N_{\text{dc}} N_{\text{df}} K^3) \).
Algorithm 1 Algorithm for Solving a EE Maximization Problem for the SDMA-Based MEC System

1: Initialize $q = 0$
2: Repeat
3: Initialize $l = 0$
4: Repeat
5: Initialize $\psi$
6: Repeat
7: Obtain $e_l^d$ by bisection method.
8: Obtain $S_l$ from Theorem 2.
9: Obtain $\Gamma_l$ and $\Gamma_{l,i}$ by solving problem (32).
10: Obtain $p_l^d$ from (33).
11: Obtain $w = \{\tau^d, \tau^n, b_i | i \in K\}$ by solving problem (36).
12: Update $\psi$ by using the subgradient method.
13: Until $\psi$ converges
14: Update $z(l)$
15: $l = l + 1$
16: Until $z(l)$ converges
17: Update $q$ from (24).
18: Until $q$ converges

IV. EE MAXIMIZATION FOR THE TDMA-BASED MEC SYSTEM

In this section, we study a EE maximization problem for the TDMA-based MEC system to compare its EE performance with the SDMA-based MEC system. Note that we can solve the problem by using the similar approach used in Section III, but we describe its procedure for the completeness of the paper. We jointly optimize downloading time $\tau_i^d$, offloading time $\tau_i^n$, MEC server’s computing time $\tau_i^c$, time instances $\tau_i^j$ at which users can initiate downloading or computing results, time instances $\tau_i^k$ at which MEC server can initiate users’ task execution, downloading transmit power $p_i^d$ of the AP, offloading transmit power $p_i^n$ of the local CPU, local computing power $p_i^c$, and local computing time $\tau_i^c$ for $i \in K$ to maximize EE. Let $x = \{\tau_i^d, \tau_i^n, \tau_i^c, \tau_i^j, \tau_i^k, p_i^d, p_i^n, p_i^c, \psi, \tau_i^c\}_{i=1}^{K}$, then the problem can be formulated as

$$\max_x \eta_{\text{tdma}}(x)$$ (37a)

s.t. $$(kf_i^3 + p_i^c + p_i^n + p_i^d)\tau_i^d + (p_i^c + p_i^d)\tau_i^n + p_i^d \tau_i^c \leq E_i, \quad i \in K$$ (37b)

$$\frac{f_i \tau_i^d}{C_i} + W \tau_i^n \log \left(1 + \frac{\tau_i^d \psi}{p_i^d} \right) \geq R_{\text{min}}, \quad i \in K$$ (37c)

$$\tau_i^c \log \left(1 + \gamma_{\text{tdma},i}^d \frac{p_i^d}{n_i^d} \right) \geq 0, \quad \tau_i^n \log \left(1 + \gamma_{\text{tdma},i}^n \frac{p_i^n}{n_i^n} \right) \geq 0, \quad \tau_i^d \leq T, \quad i \in K$$ (37d)

$$\tau_i^d = \frac{C_iW \tau_i^n \log \left(1 + \gamma_{\text{tdma},i}^d \frac{p_i^d}{n_i^d} \right)}{f_i}, \quad i \in K$$ (37e)

$$0 \leq p_i^d \leq P_A, \quad 0 \leq p_i^n \leq P_i, \quad i \in K$$ (37f)

$$\tau_i^d \geq 0, \quad \tau_i^c \geq 0, \quad i \in K$$ (37g)

$$f_i \geq 0, \quad 0 \leq \psi \leq T, \quad i \in K$$ (37h)

where (37b), (37c), (37d), (37e), and (37f) are the energy constraints of the users, the minimum throughput constraints of the users, the download bits causality constraints, the MEC server time constraints, and the maximum transmit power constraints, respectively.

Note that problem (37) is non-convex due to the fractional-form objective function and the non-convex terms including the product terms of the optimization variables in (37a)–(37d). To resolve this non-convexity issue, we introduce auxiliary variables, i) $e_i^d = \tau_i^d \psi$ meaning the energy consumed by $U_i$ for uplink data transmission, ii) $e_i^c = \tau_i^c \psi$ meaning the energy consumed by the AP for transmitting the computed result of $U_i$ for $i \in K$, and iii) $b_i$ to replace $\tau_i^c \log \left(1 + \gamma_{\text{tdma},i}^d \frac{p_i^d}{n_i^d} \right)$ with constraints in (38e). Further, by introducing local computing energy $e_i^l$, problem (37) can be divided into its inner problem in (20) and its outer problem in (38) as follows:

$$\max_z \sum_{i=1}^{K} A_i(z)$$ (38a)

s.t. $e_i^d + e_i^c + p_i^d \tau_i^d + p_i^c \tau_i^c \leq E_i, \quad i \in K$ (38b)

$$b_i^d(e_i^d) + W_i^d \geq R_{\text{min}}, \quad i \in K$$ (38c)

$$\tau_i^c \log \left(1 + \gamma_{\text{tdma},i}^d \frac{p_i^d}{n_i^d} \right) \geq 0, \quad i \in K$$ (38d)

$$\tau_i^n \log \left(1 + \gamma_{\text{tdma},i}^n \frac{p_i^n}{n_i^n} \right) \geq b_i, \quad i \in K$$ (38e)

$$t_i^d - 1 + C_i W_i b_{i-1} \leq t_i^d, \quad i = 2, \ldots, K$$ (38f)

$$\sum_{j=1}^{i} t_i^j \leq t_i^d, \quad i \in K$$ (38g)

$$\sum_{i=1}^{K} t_i^d \leq t_i^d$$ (38h)

$$t_i^d - C_i W_i b_i \leq t_i^d, \quad i \in K$$ (38i)

$$t_i^d - 1 + t_{i-1}^d \leq t_i^d, \quad i = 2, \ldots, K$$ (38j)

$$\tau_i^d \leq T, \quad i \in K$$ (38k)

Note that constraints (38f)–(38j) can be obtained from (9) and (10) by removing max operation and using $\tau_i^d = \frac{C_i W_i b_i}{f_i}$ from (37e). To solve the fractional programming in problem (38), we apply the Dinkelbach method as shown in the following procedure:
Step 1): For a given $q$, solve the following substractive-form problem
\[
\begin{align*}
\max_z & \quad \sum_{i=1}^{K} A_i(z) - q \sum_{i=1}^{K} B_i(z) \\
\text{s.t.} & \quad (38b) - (38m)
\end{align*}
\]

(39a) - (39b)

Step 2): Update
\[
q = \frac{\sum_{i=1}^{K} A_i(z)}{\sum_{i=1}^{K} B_i(z)}
\]

(40)

by using $z$ obtained from Step 1).

By repeating the above two steps, we can obtain $q^*$, which corresponds to the optimal value of problem (38), and the corresponding optimal solution $z^*$ [30]. Since problem (39) is convex, we can optimally solve it by using the similar approach used for solving problem (23). Here, we omit the detailed procedure for solving problem (39).

The computational complexity for solving problem (38) is similar to that of problem (21): The complexity for obtaining primal variables and dual variables are given by $O(K)$ and $O(K^2)$, respectively. Denoting $N_d$ to be number of the Dinkelbach iterations for obtaining $q^*$, the overall complexity for solving problem (38) is given by $O(N_d K^3)$.

In the above, we have solved EE maximization problems for both the TDMA-based MEC system and the SDMA-based MEC system. Now, we compare the EE performance of the TDMA-based MEC system and the SDMA-based MEC system assuming i) only users’ energy consumption is counted, i.e., energy consumption of AP and MEC server is not considered for EE and ii) there is no minimum throughput constraint, i.e., $\beta = 0$ and $R_{\min,i} = 0$ for $i \in K$.

Theorem 3: When $\beta = 0$ and $R_{\min,i} = 0$ for $i \in K$, the optimal EE of the TDMA-based MEC system is always greater than that of the SDMA-based MEC system.

Proof: Please refer to Appendix C. \hfill \Box

The intuition behind Theorem 3 is as follows. Compared to the TDMA-based MEC system, each user of the SDMA-based MEC system has longer offloading and downloading time, which can increase offloading throughput at the cost of more transmit energy consumption and circuit energy consumption. However, when $R_{\min,i} = 0$ for $i \in K$, allocating longer offloading time and downloading time to each user by employing SDMA is not necessarily beneficial to user EE. Moreover, the SDMA-based MEC system suffers a loss in SNR compared to the TDMA-based MEC system due to the loss in receive beamforming gain in the uplink and inter-user interference in the downlink. From above, the TDMA-based MEC system performs better than the SDMA-based MEC system in EE when $\beta = 0$ and $R_{\min,i} = 0$ for $i \in K$.

V. SIMULATION RESULTS

In this section, we present the simulation results of the proposed EE maximization scheme for the SDMA-based MEC system (EEM-S) and that for the TDMA-based MEC system (EEM-T). We consider a multi-antenna MEC system with $M = 4$ antennas at the AP, and a single antenna at each user. The users are located at a distance of $d_i = 50$ m from the AP for $i \in K$ unless specified otherwise. The channels $h_i$’s are assumed to be with Rayleigh fading and the pathloss is given by $10^{-3} \cdot d_i^{PL}$ for $i \in K$, where $d_i$ is the distance from the AP to $U_i$ and $PL = 3.6$ is the pathloss exponent. Channel reciprocity is assumed for the uplink and downlink channels, i.e., $h_i = g_i$ for $i \in K$. The system bandwidth is set to $W = 400$ kHz, time block length is set to $T = 1$ second, and $\beta$ is set to 0.1, unless specified otherwise. The following parameters of the users are set identically as follows: $C_i = 10^3$ cycles/bit, $k_i = 10^{-28}$, $O_i = 0.1$, $p_{l,i}^c = 1$ mW, $p_{l,i}^t = 2$ mW, and $p_{f,i}^c = 1$ mW for $i \in K$. The circuit power consumption per AP antenna is set to $p_{AP}^c = 10$ mW, $p_{AP}^t = 100$ mW. The minimum throughput requirements are set to $R_{\min} = 100$ kbits for all the users, unless specified otherwise. For comparison, we also provide the EE performances of the following benchmark schemes:

1) EE maximization with local computing only (EEM-X-LO): this scheme maximizes EE under the constraint that all the users $U_i$’s compute their task by only local computing;

2) EE maximization with offloading only (EEM-X-OO): this scheme maximizes EE under the constraint that all the users $U_i$’s compute their task by only offloading;

3) Computation bit maximization (CBM-X) [4]: this scheme maximizes the total achievable computation bits;

4) Energy minimization (EM-X) [8]: this scheme minimizes the total energy consumption of the users;

5) Communication EE maximization (CEEM-X): this scheme maximizes the communication EE, which is defined as the transmission bits per energy consumed for transmission.

6) EE maximization with single-antenna AP (EEM-T-SA) [19]: this scheme maximizes EE of the TDMA-based MEC system with the single-antenna AP.

In the above acronyms, $X$ is used to denote $S$ or $T$: $X = S$ for the SDMA-based MEC system and $X = T$ for the TDMA-based MEC system.

Fig. 3 depicts the EE performance of the proposed schemes versus system bandwidth $W$. Fig. 3(a) compares the proposed EEM-S with EEM-S-LO and EEM-S-OO for the SDMA-based MEC system, while Fig. 3(b) compares the proposed EEM-T with EEM-T-LO and EEM-T-OO for the TDMA-based MEC system. It is observed that EEM-S outperforms EEM-S-LO and EEM-S-OO. Also, the EEs of EEM-S and EEM-S-OO increase as $W$ increases, while that of EEM-S-LO remains unchanged. This is because EEM-S and EEM-S-OO employ offloading whose throughput bits $W r_i^u \log_2 \left(1 + \gamma_{off,i}^{ud} \right)$ linearly increases as $W$ increases, while EEM-S-LO employs only local computing whose throughput bits is independent of $W$. From above, the efficiency of offloading depends on $W$; when $W$ is small, offloading is less efficient than local computing, while when $W$ is large, it is more efficient than local computing.
H. Lim, T. Hwang: Energy-Efficient Beamforming and Resource Allocation for Multi-Antenna MEC Systems

Therefore, when $W$ is small, EEM-S-LO outperforms EEM-S-OO, and EEM-S computes most of its task by local computing. In contrast, when $W$ is large, EEM-S-OO outperforms EEM-S-LO, and EEM-S computes most of its task by offloading. For the TDMA-based MEC system, similar trends can be observed as shown in Fig. 3(b).

Fig. 4 illustrates the EE performance of various schemes versus the time block length $T$. Fig. 4(a) compares EEM-S with CBM-S, EM-S, and CEEM-S for the SDMA-based MEC system, while Fig. 4(b) compares EEM-T with CBM-T, EM-T, and CEEM-T for the TDMA-based MEC system. We can observe that EEM-S and EEM-T respectively outperform their corresponding benchmark schemes. Also, we can observe that at small values of $T$, the EE performance of EEM-S is close to that of EM-S. It can be explained as follows. When $T$ is small, the users of EEM-S may not have sufficient time for completing their tasks, and therefore, they need to use high transmit power (for offloading) and high local CPU frequency (for local computing) to meet the minimum throughput requirement $R_{\text{min}}$. Since the users have to consume much energy to meet $R_{\text{min}}$, consuming additional energy to increase the throughput beyond $R_{\text{min}}$ is not beneficial to EE. As a result, EE maximization behaves like energy minimization for small $T$. Similar trend can be observed for the TDMA-based MEC system as shown in Fig. 4(b).

Fig. 5 shows the EEs of the proposed EEM-S and EEM-T versus $R_{\text{min}}$. We can observe that the EEs of EEM-S and EEM-T decrease as $R_{\text{min}}$ increases. This is because as $R_{\text{min}}$ increases, the feasible sets of the EE maximization problems for the SDMA-based and TDMA-based MEC system reduce, which leads to a decrease in their optimal values. Also, as shown in the figure, EEM-T outperforms EEM-S when $R_{\text{min}} = 0$, which agrees with our analysis in Theorem 3. However, as $R_{\text{min}}$ increases above $R_{\text{min}} = 600$ kbits, the EE of EEM-T becomes less than that of EEM-S. This can be explained as follows. As $R_{\text{min}}$ increases, the EE maximization problems behave like computation bit maximization problems. It is well-known that SDMA outperforms TDMA in spectral efficiency in general, which implies that EEM-S outperforms EEM-T in computation bit performance. As a result, as $R_{\text{min}}$ increases, EEM-S can
consume less energy compared to EEM-T to meet higher minimum computation bit constraint, and thus shows a slower decrease in EE. Meanwhile, for the case of $\beta = 0.2$, where energy consumption of the AP is counted for EE, it can be observed that EEM-S outperforms EEM-T for all values of $R_{\min}$. This is because the AP of EEM-S can reduce its offloading and downloading time by serving multiple users at the same time, thereby saving circuit energy consumption of the AP, while the AP of EEM-T has to serve users one at a time, which leads to an inevitable increase in circuit energy consumption of the AP.

Fig. 6 depicts the EE performance of EEM-S, EEM-T, and EEM-T-SA versus the numbers of the AP antennas $M$. Since EEM-T-SA employs a single given antenna, the channels between the AP and users do not change as $M$ increases and thus the EE of EEM-T-SA remains unchanged. It can be seen that both EEM-S and EEM-T outperform EEM-T-SA, which directly show that employing multiple antennas at MEC systems can significantly improve the EE performance of MEC systems. Besides, as shown in the figure, the EEs of the proposed EEM-S and EEM-T first increase and then decrease with the number of antennas at the AP. This is because when the number of antennas becomes too large, the increased circuit energy consumption due to the increased number of antennas has more impact on EE than the improvement of beamforming gain does, which leads to decrease in EE.

VI. CONCLUSION

In this paper, we have studied EE maximization problems for the multi-antenna MEC systems based on SDMA and TDMA, which are both non-convex. To address this challenge, we have proposed an efficient two-stage approach that divides the problems into their inner and outer problems. We have derived the optimal solution of the inner problems in closed-forms and used the Dinkelbach method and DC programming algorithm to solve the outer problems. We observe that the SDMA-based MEC system in general outperforms the TDMA-based MEC system in EE performance. However, in the special case when i) only the users’ energy consumption is considered for EE ($\beta = 0$) and ii) there is no minimum throughput requirement ($R_{\min,i} = 0$ for $i \in K$), we analytically show that the TDMA-based MEC system outperforms the SDMA-based MEC system. Simulation results have shown that employing multiple antennas at MEC systems can significantly enhance the EE performance of MEC systems.

APPENDIX A

PROOF OF THEOREM 1

From (20b), we have $\tau_i^l \leq \frac{e_i^l}{k_d^l + p_i^r}$, where equality holds when (20b) satisfies with equality. Since $\frac{\partial}{\partial \tau_i^l} \left( \frac{e_i^l f_i}{k_d^l + p_i^r} \right) = \frac{-2k e_i^l (e_i^l + p_i^r e_i^l)}{(k_d^l + p_i^r e_i^l)^2}$, from first-order necessary condition, the maximum value of $\frac{e_i^l f_i}{k_d^l + p_i^r}$ can be achieved when $f_i = \left( \frac{p_i^r}{2k^l} \right)^{1/3}$. Thus, ignoring the constraint $0 \leq \tau_i^l \leq T$ in (20c), the optimal solution of problem (20) is given by $f_i^* = \left( \frac{p_i^r}{2k^l} \right)^{1/3}$ and $(\tau_i^l)^* = \frac{e_i^l}{k_d^l (e_i^l)^3 + p_i^r} = \frac{2e_i^l}{3p_i^r}$.  

Now, considering the constraint $0 \leq \tau_i^l \leq T$ in (20c), we consider the following two cases.

i) $\frac{2e_i^l}{3p_i^r} < T \left( e_i^l < \frac{3p_i^r}{2T} \right)$

In this case, we have $f_i^* = \left( \frac{p_i^r}{2k^l} \right)^{1/3}$ and $(\tau_i^l)^* = \frac{2e_i^l}{3p_i^r}$. Then, the optimal value of problem (20) can be obtained as

$$b_i^T(e_i^l) = \frac{f_i^* (\tau_i^l)^*}{C_i} = \frac{1}{C_i} \left( \frac{p_i^r}{2k^l} \right)^{1/3} \frac{2e_i^l}{3p_i^r}$$

ii) $\frac{2e_i^l}{3p_i^r} \geq T \left( e_i^l \geq \frac{3p_i^r}{2T} \right)$

In this case, considering the constraints $0 \leq \tau_i^l \leq T$ in (20c), we have $(\tau_i^l)^* = T$. Then, from (20b),
Next, we show the concavity of $f_i = \left(\frac{e_i}{k_i} - \frac{p_i^c}{k_i^c}\right)^{1/3}$. Since the objective function $f_i^* = \frac{f_i^*}{C_i}$ increases with $f_i$, the optimal $f_i$ is given by $f_i^* = \left(\frac{e_i}{k_i} - \frac{p_i^c}{k_i^c}\right)^{1/3}$. Then, the optimal value of problem (20) can be obtained as

$$
b^*_i(e^*_i) = \frac{\tau_i^*}{C_i} = \frac{e_i^*}{k_i T} - \frac{p_i^c}{k_i}.
$$

From above, $b^*_i(e^*_i)$ can be expressed as shown in (22). Next, we show the concavity of $b^*_i(e^*_i)$. The derivative of $b^*_i(e^*_i)$ exists and it can be written by

$$D_i(e^*_i) = \frac{db^*_i}{de^*_i} = \begin{cases}
\frac{1}{C_i} \left(\frac{p_i^c}{2k_i}\right)^{1/3} & e^*_i < \frac{3p_i^c}{2} T
\end{cases}
$$

To prove the concavity of $b_i(e_i^*)$, it is sufficient to show that [32]

$$b_i^*(b) \leq b_i^*(a) + D_i(a)(b - a), \quad \forall a, b \in [0, \infty). \quad (41)
$$

We first consider the case of $a \leq b$. Note that $D_i(e_i^*)$ first remains constant as $\frac{1}{C_i} \left(\frac{p_i^c}{2k_i}\right)^{1/3} \frac{2}{3p_i^c} \frac{e_i^*}{k_i} T$ and then decreases in $e_i^* \geq \frac{3p_i^c}{2} T$. Since $D_i(e_i^*)$ is a non-increasing function in $e_i^*$, we have $D_i(x) \leq D_i(a)$ for $a \leq x \leq b$. Then, it follows that

$$\int_a^b D_i(x) \, dx \leq \int_a^b D_i(a) \, dx,
$$

or equivalently, $b_i^*(b) - b_i^*(a) \leq D_i(a)(b - a)$. Then, we have $b_i^*(b) \leq b_i^*(a) + D_i(a)(b - a)$. For the case of $b \leq a$, we can prove (41) in a similar way.

**APPENDIX B**

**PROOF OF THEOREM 2**

The Lagrangian in (27) can be rewritten as $L(z, \psi) = \sum_{i=1}^K \text{tr}(A_iS_i) + L_K(z - \{S_i\}_{i=1}^K, \psi)$, where

$$A_i^* \triangleq v_i H_i - \sum_{k \neq i} \phi_k H_k - (\alpha_i + q\beta)I = B_i - (\alpha_i + q\beta)I$$

$$B_i^* \triangleq v_i H_i - \sum_{k \neq i} \phi_k H_k.$$

To ensure that the dual function in (29) is bounded from above, we only consider the set of dual variables $\psi$ such that $A_i \preceq 0$. From $A_i \preceq 0$ and $S_i \succeq 0$, we have $\text{tr}(A_iS_i) \leq 0$. Then, the optimal $S_i^*$ that maximizes $L$ should satisfy

$$\text{tr}(A_iS_i^*) = 0 \iff A_iS_i^* = 0, \quad (42)$$

where $(a)$ is due to $A_i \preceq 0$ and $S_i^* \succeq 0$.

Now, we consider the eigenvalues of $A_i$ and $B_i$, denoted by $\{\lambda_1(A_i) \geq \cdots \geq \lambda_M(A_i)\}$ and $\{\lambda_1(B_i) \geq \cdots \geq \lambda_M(B_i)\}$, respectively. Note that $B_i$ can be rewritten as $B_i = \mu_iH_i - \sum_{k \neq i} \phi_k H_k = F_iD_iF_i^H$, where $F_i = \left[\sqrt{\beta_i}, \sqrt{\beta_i} H_i, \cdots, \sqrt{\beta_i} H_K\right]$ is an $M \times K$ matrix and $D_i = \text{diag}(1, -1, \cdots, -1)$ is an $K \times K$ matrix with $K \leq M$. Then, by adding $M - K$ vectors $f_i, \cdots, f_{M-K}$ to $F_i$ such that $F_i = [F_i, f_1, \cdots, f_{M-K}]$ is an invertible $M \times M$ matrix, we can rewrite $B_i = B_i \triangleq \bar{F}_iD_i\bar{F}_i^H$, where $D_i = \text{diag}(1, -1, \cdots, -1, 0, \cdots, 0)$. From Sylvester’s law of inertia [39], we can see that $B_i$ has one positive eigenvalue, $K - 1$ negative eigenvalues, and $M - K$ zero eigenvalues, i.e., $\{\lambda_1(B_i) > 0 = \cdots = 0 > \lambda_{M-K+2}(B_i) \geq \cdots \geq \lambda_{M}(B_i)\}$. Then, the eigenvalues of $A_i = B_i - (\alpha_i + q\beta)I$ can be expressed as $\{\lambda_1(A_i) - (\alpha_i + q\beta) > - (\alpha_i + q\beta) = \cdots = - (\alpha_i + q\beta) > \lambda_{M-K+2}(B_i) - (\alpha_i + q\beta) \geq \lambda_{M-K}(A_i) - (\alpha_i + q\beta)\}$, with the eigenvectors same as those of $B_i$.

From $A_i \preceq 0$, it follows that $\lambda_1(B_i) - (\alpha_i + q\beta) \leq 0$. Here, we consider the following two cases.

Case i) $\lambda_1(B_i) - (\alpha_i + q\beta) < 0$: In this case, $A_i$ has full rank. Then, from (42), we have $S_i^* = 0$, which corresponds to $S_i^* = \sqrt{\mu_i}H_i$ with $a_i = 0$.

Case ii) $\lambda_1(B_i) - (\alpha_i + q\beta) = 0$: In this case, rank($A_i$) = $M - 1$. Denote $u_i$ to be the unit-norm eigenvector of $B_i$ corresponding to its largest eigenvalue $\lambda_1(B_i)$. Then, $u_i$ is also the eigenvector of $A_i$ corresponding to its maximum eigenvalue $\lambda_1(A_i) - (\alpha_i + q\beta) = 0$, which implies that $u_i$ spans the null space of $A_i$, i.e., $A_iu_i = 0$. Then, $S_i^* \succeq 0$ and (42) can be satisfied with $S_i^* = \sqrt{\mu_i}H_i$ for any $a_i \geq 0$.

**APPENDIX C**

**PROOF OF THEOREM 3**

Denote $(\tau^{\text{dma}})_i = \{(	au^d)^*_i, (\tau^u)^*_i, (\tau^s)^*_i, (\tau^c)^*_i, a_i^*\}$, $\tau^{\text{dma}}_i, \tau^{\text{ldma}}_i, S_i^* \in K$ to be the optimal solution of problem (21) that achieves $(\eta^{\text{dma}})_i$. Then, we consider a solution of $(\tau^{\text{ldma}})_i = \{\tilde{\tau}^{\text{ldma}}_i, \tilde{\tau}^{\text{ldma}}_i, \tilde{\tau}^{\text{ldma}}_i, \tilde{\tau}^{\text{ldma}}_i, \tilde{\tau}^{\text{ldma}}_i\}$ for problem (38), with $\tilde{\tau}^{\text{ldma}}_i = \frac{\tau^{\text{ldma}}_i}{K}, \tilde{\tau}^{\text{ldma}}_i = \frac{\tau^{\text{ldma}}_i}{K}, \tilde{\tau}^{\text{ldma}}_i = \frac{\tau^{\text{ldma}}_i}{K}, \tilde{\tau}^{\text{ldma}}_i = \frac{\tau^{\text{ldma}}_i}{K}$, and $\tilde{\tau}^{\text{ldma}}_i = \frac{\tau^{\text{ldma}}_i}{K}$ for $i \in K$. Besides, we set $\tilde{\tau}^{\text{ldma}}_i = \tilde{\tau}^{\text{ldma}}_i, \tilde{\tau}^{\text{ldma}}_i = \max(\tilde{\tau}^{\text{ldma}}_i + \frac{C_iW_i}{f_M}, \sum_{j=1}^K \tilde{\tau}^{\text{ldma}}_j)$, and $\tilde{\tau}^{\text{ldma}}_i = \max(\tilde{\tau}^{\text{ldma}}_i + \frac{C_iW_i}{f_M}, \sum_{j=1}^K \tilde{\tau}^{\text{ldma}}_j)$ for $i = 2, \cdots, K$, which satisfies (38)–(38). Preliminarily, we calculate the upper bounds of $\tilde{\tau}^{\text{ldma}}_i$ and $\tilde{\tau}^{\text{ldma}}_i$. It follows that

$$\tilde{\tau}^{\text{ldma}}_2 = \max\left(\tilde{\tau}^{\text{ldma}}_2 + \frac{C_iW_i}{f_M}, \tilde{\tau}^{\text{ldma}}_2 + \frac{\tilde{\tau}^{\text{ldma}}_2}{(a)}\right)
$$

$$= \max\left(\tilde{\tau}^{\text{ldma}}_2 + \frac{C_iW_i}{f_M}, \tilde{\tau}^{\text{ldma}}_2 + \frac{\tilde{\tau}^{\text{ldma}}_2}{(a)}\right)
$$

$$= \tilde{\tau}^{\text{ldma}}_2 + \frac{\tilde{\tau}^{\text{ldma}}_2}{(a)} + \frac{C_iW_i}{f_M} \geq \tilde{\tau}^{\text{ldma}}_2 + \frac{\tilde{\tau}^{\text{ldma}}_2}{(a)} \quad (43)$$
where \((a)\) holds from \(\max(A, B) \leq A + B\) for \(A \geq 0\) and \(B \geq 0\). Repeating the above manipulations for \(\tilde{t}_i^d\), \(i = 3, \ldots, K\), we have

\[
\tilde{t}_K^d \leq \sum_{i=1}^{K} \tilde{t}_i^u + \sum_{i=1}^{K-1} C_i W \tilde{b}_i \tag{44}
\]

Also, it follows that

\[
\tilde{t}_i^d = \max \left( \tilde{t}_i^u + \frac{C_i W \tilde{b}_i}{f_m}, \sum_{i=1}^{K} \tilde{t}_i^u + \frac{C_i W \tilde{b}_i}{f_m} \right) \leq \max \left( \tilde{t}_i^u + \frac{C_i W \tilde{b}_i}{f_m}, \sum_{i=1}^{K} \tilde{t}_i^u + \frac{C_i W \tilde{b}_i}{f_m} \right)
\]

\[
\leq \sum_{i=1}^{K} \tilde{t}_i^u + \sum_{i=1}^{K} C_i W \tilde{b}_i \tag{45}
\]

where \((b)\) holds from (43) and (45). Repeating the above manipulations for \(\tilde{t}_i^d\), \(i = 3, \ldots, K\), we have

\[
\tilde{t}_K^d \leq \sum_{i=1}^{K} \tilde{t}_i^u + \sum_{i=1}^{K} C_i W \tilde{b}_i \tag{47}
\]

Now, we show that \(\tilde{z}\) is feasible for problem (38). The solution \(\tilde{z}\) satisfies (38b) because

\[
\tilde{\gamma}_i^d(e_i^d)^* + (e_i^d)^* + p_{i,c} \tilde{z}_i^d \leq (e_i^d)^* + (e_i^d)^* + p_{i,c} \tilde{z}_i^d \leq E_i \tag{c) \leq E_i}
\]

where \((c)\) holds from (21b). Also, \(\tilde{z}\) satisfies (38k) because

\[
\tilde{t}_K^d \leq \sum_{i=1}^{K} \tilde{t}_i^u + \sum_{i=1}^{K} C_i W \tilde{b}_i \leq \sum_{i=1}^{K} \tilde{t}_i^u + \sum_{i=1}^{K} C_i W \tilde{b}_i \leq \sum_{i=1}^{K} \tilde{t}_i^u + \sum_{i=1}^{K} C_i W \tilde{b}_i \leq (e_i^d)^* + (e_i^d)^* + p_{i,c} \tilde{z}_i^d \leq T, \tag{e)
\]

where \((d)\) and \((e)\) holds from (47) and (19b), respectively. The solution \(\tilde{z}\) also satisfies (38d) and (38e) because

\[
\tilde{t}_i^d \log_2 \left( 1 + \frac{\gamma_{d_i}^{\text{dma}}}{\tilde{t}_i^d} \right) = \frac{(e_i^d)^*}{K} \log_2 \left( 1 + \frac{\|h_i^d\|^2}{\tau^d} \cdot \text{tr}(S_i^d) \right)
\]

Here, \((l)\) holds due to the concavity of \(b_i^d(e_i^d)^*\), which is stated in Theorem 1.
Y. Kwon, T. Hwang, and X. Wang, “Energy-efficient transmit power control for multi-tier MIMO HetNets,” IEEE J. Sel. Areas Commun., vol. 33, no. 10, pp. 2070–2086, Oct. 2015.

H. Sun, F. Zhou, and R. Q. Hu, “Joint offloading and computation energy efficiency maximization in a mobile edge computing system,” IEEE Trans. Veh. Technol., vol. 68, no. 3, pp. 3052–3056, Mar. 2019.

L. Shi, Y. Ye, X. Chu, and G. Lu, “Computation energy efficiency maximization for a NOMA-based WPT-MEC network,” IEEE Internet Things J., vol. 8, no. 13, pp. 10731–10744, Jul. 2021.

S. Mao, N. Zhang, L. Liu, J. Wu, M. Dong, K. Ota, T. Liu, and D. Wu, “Computation rate maximization for intelligent reflecting surface enhanced wireless powered mobile edge computing networks,” IEEE Trans. Veh. Technol., vol. 70, no. 10, pp. 10820–10831, Oct. 2021.

J. Wen, C. Ren, and A. K. Sangaiha, “Energy-efficient device-to-device edge computing network: An approach offloading both traffic and computation,” IEEE Commun. Mag., vol. 56, no. 9, pp. 96–102, Sep. 2018.

Y. Pan, M. Chen, Z. Yang, N. Huang, and M. Shikh-Bahaei, “Energy-efficient NOMA-based mobile edge computing offloading,” IEEE Commun. Lett., vol. 23, no. 2, pp. 310–313, Feb. 2019.

M. Zeng, W. Hao, O. A. Dobre, and H. V. Poor, “Delay minimization for massive MIMO assisted mobile edge computing,” IEEE Trans. Veh. Technol., vol. 69, no. 6, pp. 6788–6792, Jun. 2020.

T. H. Nguyen, T. V. Chien, H. Q. Ngo, X. N. Tran, and E. Bjornson, “Pilot assignment for joint uplink-downlink spectral efficiency enhancement in massive MIMO systems with spatial correlation,” IEEE Trans. Veh. Technol., vol. 70, no. 8, pp. 8292–8297, Aug. 2021.

M. Andrews, A. Fernandez Anta, L. Zhang, and W. Zhao, “Routing for energy minimization in the speed scaling model,” in Proc. IEEE INFOCOM, San Diego, CA, USA, Mar. 2010, pp. 1–9.

M. Andrews, A. F. Anta, L. Zhang, and W. Zhao, “Routing for power minimization in the speed scaling model,” IEEE/ACM Trans. Netw., vol. 20, no. 1, pp. 285–294, Feb. 2012.

F. R. Yu and V. Leung, Advances in Mobile Cloud Computing Systems. Boca Raton, FL, USA: CRC Press, 2015.

Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, “Semidefinite relaxation of quadratic optimization problems,” IEEE Signal Process. Mag., vol. 27, no. 3, pp. 20–34, May 2010.

D. Winkelbach, “On nonlinear fractional programming,” Manage. Sci., vol. 13, no. 7, pp. 492–498, Mar. 1967.

D. G. Luenberger and Y. Ye, Linear and Nonlinear Programming, 3rd ed. New York, NY, USA: Springer, 2009.

E. K. P. Chong and S. H. Zak, An Introduction to Optimization, 3rd ed. New York, NY, USA: Wiley, 2008.

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

S. Boyd, “Ellipsoid method,” Stanford Univ., Stanford, CA, USA, Lect. Notes EE364b, 2018. [Online]. Available: http://web.stanford.edu/class/ee364b/lectures/ellipsoid_method_slides.pdf

S. Boyd, L. Xiao, A. Mutapic, and J. Mattingley, “Sequential convex programming,” Stanford Univ., Stanford, CA, USA, Tech. Rep. EE364b, 2007.

G. R. Lanckriet and B. K. Sriperumbudur, “On the convergence of the concave-convex procedure,” in Proc. Adv. Neural Inf. Process. Syst., 2009, pp. 1759–1767.

N. Vucic, S. Shi, and M. Schubert, “DC programming approach for resource allocation in wireless networks,” in Proc. Int. Symp. Modeling Optim. Mobile Ad Hoc Wireless Netw., May 2010, pp. 380–386.

L. T. H. An and P. D. Tao, “Network utility maximisation: A DC programming approach for sigmoidal utility function,” in Proc. Int. Conf. Adv. Technol. Commun. (ATC), Ho Chi Minh City, Vietnam, Oct. 2013, pp. 50–54.

J. Carrell, Groups, Matrices, and Vector Spaces: A Group Theoretic Approach to Linear Algebra. New York, NY, USA: Springer, 2017.

HUNWOO LIM received the B.S. and M.S. degrees in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2015 and 2017, respectively, where he is currently pursuing the Ph.D. degree. His research interests include energy-efficient communications, wireless energy transfer, multi-access edge computing, and vehicular communications.

TAEWON HWANG received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 1993, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1995 and 2005, respectively. From 1995 to 2000, he was a Member of Technical Staff with the Electronics and Telecommunications Research Institute, Daejeon, South Korea. Since 2006, he has been with the School of Electrical and Electronic Engineering, Yonsei University, where he is currently a Professor. His research interests include energy-efficient communications, heterogeneous networks, cognitive radio, and MIMO communications. He has been a technical program committee member of many international conferences. He has served as an Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Cognitive Radio Series.

IEEE Access