# Energy-Efficient Mobile-Edge Computation Offloading over Multiple Fading Blocks

Rongfei Fan, Fudong Li, Song Jin, Gongpu Wang, Hai Jiang, and Shaohua Wu

Abstract—By allowing a mobile device to offload computation-intensive tasks to a base station, mobile edge computing (MEC) is a promising solution for saving the mobile device's energy. In real applications, the offloading may span multiple fading blocks. In this paper, we investigate energy-efficient offloading over multiple fading blocks with random channel gains. An optimization problem is formulated, which optimizes the amount of data for offloading to minimize the total expected energy consumption of the mobile device. Although the formulated optimization problem is non-convex, we prove that the objective function of the problem is piecewise convex, and accordingly develop an optimal solution for the problem. Numerical results verify the correctness of our findings and the effectiveness of our proposed method.

*Index Terms*—Mobile edge computing, fading, mobile computation offloading.

## I. Introduction

Due to the limited computation resources at a mobile device, it is difficult for the mobile device to run computation-intensive applications that require low latency, such as natural language processing, virtual reality (VR), and augmented reality (AR). Mobile edge computing (MEC) provides a solution, in which a mobile device is allowed to offload part of or all its data to a base station for computing, as the base station often has high computation capability. After the base station completes the computing tasks, it feeds back the computed results to the mobile device [1].

Since the offloading is via wireless links, the wireless transmission for offloading involves power consumption and leads to latency. How to minimize power consumption of the mobile device while satisfying latency requirements of computing has been investigated in the literature [2]–[10]. For a single mobile device with a single antenna, the work in [2] considers that the computation task can be partitioned to two parts, which are computed locally at the mobile device and are offloaded to the base station, respectively. The optimal offloading ratio (i.e., the percentage of the computation task that is offloaded) and the mobile device's transmit power level are derived. Also for a single mobile device, the work in [3]

R. Fan and S. Jin are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, P. R. China (email: {fanrongfei, 15120075709}@bit.edu.cn). F. Li and H. Jiang are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada (email: {fudong1,hai1}@ualberta.ca). G. Wang is with the School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, P. R. China (email: gpwang@bjtu.edu.cn). S. Wu is with the Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, 518055, P. R. China (email: hitwush@hit.edu.cn).

investigates a similar research problem when both the mobile device and the base station have multiple antennas. The work in [4] supposes that the base station's CPU alternates between busy state and idle state and assumes that the base station's idle intervals are known in advance. The offloaded data in each epoch (during which the base station's CPU keeps idle or keeps busy) is optimized so as to minimize the mobile device's total energy consumption. The work in [5] considers multiple single-antenna mobile devices working in time division multiple access (TDMA) mode. Each mobile device's offloading ratio, transmit power allocation, and time ratio for channel access are optimized. The work in [6] considers a multiple-input and multiple-output (MIMO) MEC system with one and multiple mobile devices. All the data for computing are offloaded to the base station. Pre-coding matrix is designed for the single-device case, while precoding matrix design and the base station's computation resource allocation are investigated for the multiple-device case. In [7], a mobile device can be served by one of multiple base stations. Base station selection is investigated so as to minimize the mobile device's energy consumption. The work in [8] considers a relay between a mobile device and its base station, and optimizes the offloading ratio and the transmit power levels of the mobile device and the relay.

Note that in all these works [2]–[8], the wireless channel gain is assumed to keep unchanged in the whole process of data offloading. However, the maximal allowable delay of some applications can be longer than one fading block. For example, the application of VR requires the processing delay to be within 20-60 ms [11], while the length of one fading block, within which wireless channel gain can be considered unchanged, can be at the order of 1 ms [12]. Thus, the data offloading process may span multiple fading blocks, which means that the wireless channel gain varies during the data offloading process, and therefore, the methods in [2]–[8] cannot be used.

To the best of our knowledge, only two works in the literature consider data offloading over multiple fading blocks, and both works suppose there is a single mobile device in the system. The work in [9] investigates the problem of optimal offloading over multiple fading blocks for one mobile device that is wireless-powered by the base station. It is assumed that the channel state information (CSI) of multiple future fading blocks is non-causal, i.e., the CSI of multiple future fading blocks is known in advance, which may not be practical.

The work in [10] also deals with offloading over multiple fading blocks. Data for computation is not partitioned. Thus, the whole computation task is offloaded to the base station or computed locally, depending on which side leads to less energy consumption. The wireless channel gain is assumed to take a "good" state or a "bad" state. The computation time at the base station is not taken into account (i.e., the computation time is assumed to be zero). By approximating the energy consumption for offloading one bit of data as a monomial function, a dynamic optimization problem is formulated to minimize the total energy consumption for offloading, if the mobile device decides to offload rather than locally computing its data.

In this paper, we also investigate the MEC offloading over multiple fading blocks. Different from [9], we assume that the CSI of multiple future fading blocks is causal, i.e., the CSI of multiple future fading blocks is not known in advance. In addition, the research problem in [9] is a convex problem, which is solved by using the KKT condition. But our considered research problem is a non-convex problem. Differences of our work from [10] are as follows. 1) We consider that the data for computation can be partitioned and executed in parallel at the mobile device and a base station. This model follows most of existing works [2], [5] and is a more general model compared to the non-partitioning model in [10]. 2) We take into account computation time at the base station.

The contributions of our work are summarized as follows. Considering computation offloading of a mobile device over multiple fading blocks, we formulate an optimization problem to minimize the total expected energy consumption of the mobile device by deciding the amount of data to be offloaded. The formulated problem is a non-convex optimization problem, and thus, is hard to solve. To solve the non-convex problem, we discover and theoretically prove that the objective function of the problem is piecewise convex, and thus, the formulated problem can be optimally solved by first finding solutions for individual intervals within which convexity is kept and then picking up the best solution.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider one mobile device with a computation-intensive computation task. The mobile device is connected to a base station, and can offload part of its data to the base station for computing. After receiving all the offloaded data, the base station starts computation and then feeds back the computed results to the mobile device.

The wireless channel between the mobile device and the base station is assumed to be block-faded. Specifically, the channel gain remains constant in a fading block, but varies randomly and independently from fading block to fading block. Denote the *normalized channel gain* as h, which is the ratio of the wireless channel gain to the background noise

variance. h is independently and identically distributed over fading blocks with probability density function (PDF) f(h). Since the maximal allowable latency of the computation task can be much longer than the duration of one fading block, the computation offloading process may span multiple fading blocks.

At the beginning of a fading block, the mobile device needs to estimate the normalized channel gain in the fading block (for example, the mobile device sends pilot signals to the base station; then the base station estimates the normalized channel gain and feeds back to the mobile device). The time overhead to estimate the normalized channel gain is normally very short, and thus, is not considered in our work. Nevertheless, it is straightforward to take this time overhead into account in our system model.

Detailed computation and communication models are given as follows.

## A. Computation Model

The computation task, denoted as  $\mathcal{T}$ , can be described by a profile of three parameters  $(T,D,c_0)$ , where T is the maximal allowable latency, D is the amount of computation input data (in unit of nat for presentation simplicity), and  $c_0$  is the number of CPU cycles required for computing one input nat. With such a description, the total number of CPU cycles required for completing task  $\mathcal{T}$  is  $c_0D$ . The computation task  $\mathcal{T}$  is separable. Denote the amount of local task (i.e., the data to be computed at the mobile device) and external task (i.e. the data to be computed at the base station) as  $D_l$  and  $D_e$ , respectively. Then we have

$$D = D_l + D_e. (1)$$

For the mobile device, let  $f_l$  denote its computation capability, which is in unit of CPU cycles per second and is also called *CPU frequency*. Similar to [2],  $f_l$  is adjustable and is not larger than an upper limit denoted as  $f_l^U$ . To complete the local task within duration T, the mobile device's computation capability should be set as  $f_l = \frac{c_0 D_l}{T}$  (recalling that the number of CPU cycles for computing the local task is  $c_0 D_l$ ). Due to the constraint  $f_l \leq f_l^U$ , we have

$$D_l \le \frac{f_l^U T}{c_0}. (2)$$

 $^1$  As shown in [7], [9], [10], for a given data size D, the number of CPU cycles required for computing the data can be written as  $C_0D$ , where  $C_0$  is a random variable whose distribution, denoted as  $g(C_0)$ , depends on the nature of the application. To facilitate the analysis, a fixed value  $c_0$  is selected in [7], [9], [10] to guarantee that  $\Pr\left(C_0D \leq c_0D\right) \geq (1-\varepsilon)$  where  $\Pr(\cdot)$  means probability and  $\varepsilon$  is a predefined small value. Thus,  $c_0$  can be calculated by using knowledge of  $g(C_0)$ . Accordingly, the number of CPU cycles required for computing the data with size D can be expressed as  $c_0D$ .

According to [2], the consumed energy for the mobile device to finish the local task with CPU frequency  $f_l$  within duration T can be written as

$$E_l(D_l) = kf_l^3 T = \frac{kc_0^3 D_l^3}{T^2}$$
 (3)

where k is a fixed coefficient depending on the architecture of the mobile device's CPU.

At the base station, denote the computation capability for serving the mobile device as  $f_e$ . Thus, to complete the external task with  $c_0D_e$  CPU cycles, the associated computational time can be expressed as

$$T_e(D_e) = \frac{c_0 D_e}{f_e}. (4)$$

Thus, the offloading should be finished within duration  $T - T_e(D_e)$ .

## B. Communication Model

Data offloading is performed over multiple fading blocks. Denote the time duration of one fading block as  $T_f$ . Similar to [5], the time for feeding back the computed results from the base station to the mobile device is negligible and thus, is not considered. This is reasonable since the computed results are usually of small data size and the base station can transmit with high power. The total number of fading blocks for offloading  $D_e$  data nats is a function of  $D_e$  and can be written as

$$N(D_e) = \left\lceil \frac{T - T_e(D_e)}{T_f} \right\rceil = \left\lceil \frac{Tf_e - c_0 D_e}{T_f f_e} \right\rceil$$
 (5)

where  $\lceil \cdot \rceil$  is the ceiling function. For the ease of discussion, these  $N(D_e)$  fading blocks are indexed inversely, i.e., the mobile device starts offloading data in fading block  $N(D_e)$  and ends offloading data in fading block 1. Denote the set of the  $N(D_e)$  fading blocks as  $\mathcal{N}(D_e) \triangleq \{1, 2, ..., N(D_e)\}$ .

Note that the mobile device transmits in partial duration in fading block 1. Denote transmission time duration in fading block n as  $t_n(D_e)$ . Then we have  $t_n(D_e) = T_f$ , for  $n \in \{2, 3, ..., N(D_e)\}$ , and

$$t_{1}(D_{e}) = T - T_{e}(D_{e}) - (N(D_{e}) - 1)T_{f}$$

$$= \left(T - \frac{c_{0}D_{e}}{f_{e}}\right) - \left[\left(T - \frac{c_{0}D_{e}}{f_{e}}\right)/T_{f}\right]T_{f} + T_{f}.$$
(6)

In one fading block, recall that h denotes the normalized channel gain. Consider transmission of d data nats with transmission duration t over the spectrum with bandwidth W. By using the Shannon channel capacity formula [12], the transmit power at the mobile device should be

$$p(d, h, t, W) = \frac{\left(e^{\frac{d}{tW}} - 1\right)}{h} \tag{7}$$

and the consumed energy can be written as

$$e(d, h, t, W) = p(d, h, t, W)t = \frac{\left(e^{\frac{d}{tW}} - 1\right)t}{h}.$$
 (8)

The mobile device should offload  $D_e$  data nats to the base station within  $N(D_e)$  fading blocks. We should minimize the energy consumption in the offloading. We have two definitions:

- $J_n(d)$   $(n \in \mathcal{N}(D_e))$  is defined as the minimum expected energy consumption for transmitting totally d data nats from fading block n to fading block 1;
- $J_n(d,h_n)$  is defined as the minimum expected energy consumption for transmitting totally d data nats from fading block n to fading block 1 given that the normalized channel gain in fading block n is a specific value  $h_n$ .

Thus,  $J_n(d)$  and  $J_n(d,h_n)$  satisfy the following equation  $J_n(d)=\int_0^\infty J_n(d,h_n)f(h_n)dh_n.$ 

According to Bellman's equation in dynamic programming [13], we have the following iterative calculation formulas:

$$J_n(d) = \int_0^\infty J_n(d, h_n) f(h_n) dh_n, \quad n \in \mathcal{N}(D_e), \quad (9)$$

in which

$$J_{n}(d, h_{n}) = \min_{0 \leq d_{n} \leq d} \left( e(d_{n}, h_{n}, T_{f}, W) + \int_{0}^{\infty} J_{n-1} (d - d_{n}, h_{n-1}) f(h_{n-1}) dh_{n-1} \right)$$

$$= \min_{0 \leq d_{n} \leq d} \left( e(d_{n}, h_{n}, t_{n}, W) + J_{n-1, m} (d - d_{n}) \right),$$

$$\forall n > 1.$$
(10)

and 
$$J_1(d, h_1) = e(d, h_1, t_1, W)$$
.

In (10),  $d_n$  means the offloaded data nats within fading block n. Then by following the iterative calculation formulas in (9) and (10),  $J_{N(D_e)}(D_e)$  can be calculated, which is the minimum expected energy consumption for transmitting totally  $D_e$  data nats from fading block  $N(D_e)$  to fading block 1. To achieve  $J_{N(D_e)}(D_e)$ , in fading block n ( $n \in \mathcal{N}(D_e)$ ), with  $h_n$  measured and d remaining data nats to be offloaded, the mobile device only needs to determine  $d_n$  by solving the optimization problem defined in (10).

## C. Problem Formulation

Given the system model, the total energy consumption of the mobile device to complete the task  $\mathcal{T}$  is  $\left(J_{N(D_e)}(D_e)+E_l(D_l)\right)$ . Our target is to minimize the total energy consumption by setting  $D_l$  and  $D_e$  optimally before the mobile device starts the data offloading, i.e., to solve the following optimization problem

Problem 1:

$$\min_{D_{e},D_{l}} \quad J_{N(D_{e})}(D_{e}) + E_{l}(D_{l}) 
\text{s.t.} \quad D_{e} \ge 0, \qquad (11) 
D_{l} \ge 0, \qquad (12) 
\text{Constraints } (1), (2).$$

## III. OPTIMAL SOLUTION

Problem 1 is not a convex optimization problem since  $N(D_e)$  in  $J_{N(D_e)}(D_e)$  involves the ceiling function of  $D_e$ . In this section, we will show how to solve Problem 1 optimally. The following lemmas can be expected.

Lemma 1: With  $d \in [0, D_e]$ ,  $J_n(d)$  and  $J_n(d, h_n)$  are convex with d for  $n \in \{1, 2, ..., N(D_e) - 1\}$ .

*Proof:* We use the induction method. For n=1, it can be easily checked that

$$J_1(d, h_1) = e(d, h_1, t_1, W)$$

$$= \frac{t_1(D_e)}{h_1} \left( e^{\frac{d}{t_1(D_e)W}} - 1 \right)$$
(13)

and

$$= \int_{0}^{J_{1}(d)} J_{1}(d, h_{1}) f(h_{1}) dh_{1}$$

$$= t_{1}(D_{e}) \sigma^{2} W\left(e^{\frac{d}{t_{1}(D_{e})W}} - 1\right) \int_{0}^{\infty} \frac{1}{h_{1}} f(h_{1}) dh_{1},$$
(14)

both of which are convex functions with d.

Now, suppose both  $J_n(d)$  and  $J_n(d,h_n)$  are convex with d for  $n \in \{1,2,...,N(D_e)-2\}$ . Then it can be derived that  $J_{n+1}(d,h_n)$  is still convex with d since the infimal convolution of convex functions is still convex [14, Theorem 5.4].

Given that  $J_n(d,h_n)$  is convex with d, it is straightforward to see the convexity of  $J_n(d)$  with d for  $n \in \{1,2,...,N(D_e)-1\}$  since  $J_n(d)$  can be interpreted as nonnegative weighted sum of  $J_n(d,h_n)$  with weighting coefficient  $f(h_n)dh_n$ , which is still a convex function [15].

This completes the proof.

Lemma 2:  $J_{N(D_e)}(D_e)$  is piecewise convex with  $D_e$ .

*Proof:* From (6), we notice that  $t_1(D_e)$  is piecewise linear decreasing function with  $D_e$ . Suppose  $\mathcal{I} \triangleq \{0,1,...,I^*\}$  where  $I^*$  is the maximal integer i such that  $(T-iT_f)$  is larger than 0. Define

$$\mathcal{D}_{i} \triangleq \left\{ \begin{array}{l} \left( (T - (i+1)T_{f}) \frac{f_{e}}{c_{0}}, (T - iT_{f}) \frac{f_{e}}{c_{0}} \right], \\ \forall i \in \mathcal{I} \setminus \{I^{*}\}, \\ \left( 0, (T - iT_{f}) \frac{f_{e}}{c_{0}} \right], i = I^{*}. \end{array} \right.$$

Then when  $D_e$  falls within  $\mathcal{D}_i$ , we have

$$t_1(D_e) = T - iT_f - \frac{c_0 D_e}{f_e}.$$
 (15)

Without loss of generality, we assume  $D_e$  falls into  $\mathcal{D}_i$ , where  $i \in \mathcal{I}$ . Thus  $N(D_e) = i+1$ . In this case, denote the optimal solution of d achieving  $J_{i+1}(D_e,h_{i+1})$  as  $d_{i+1}^*(D_e,h_{i+1})$ . Suppose  $D_e^\dagger \in \mathcal{D}_i$  and  $D_e^\dagger \in \mathcal{D}_i$ , thus  $i+1=N(D_e^\dagger)=N(D_e^\dagger)$ . For any  $\alpha \in [0,1]$ , the convexity of  $J_{i+1}(D_e,h_{i+1})$  can be proved in (16) on the top of the next page, where (a) is due to the convexity of  $J_i(d,h_i)$  with d according to Lemma 1 and the convexity of e(d,h,t,W) with d, and (b) holds since

$$\begin{array}{l} \left(\alpha d_{i+1}^*(D_e^\dagger,h_{i+1}) + (1-\alpha)\,d_{i+1}^*(D_e^\dagger,h_{i+1})\right) \\ \in \left[0,\alpha D_e^\dagger + (1-\alpha)D_e^\dagger\right] \end{array}$$

and  $d_{i+1}^*(\alpha D_e^\dagger + (1-\alpha)D_e^\dagger, h_{i+1})$  is the optimal solution of  $d_{i+1}$  for achieving  $J_{i+1}(\alpha D_e^\dagger + (1-\alpha)D_e^\dagger, h_{i+1})$ .

This proves the convexity of  $J_{i+1}(D_e, h_{i+1})$  with  $D_e$  for  $D_e \in \mathcal{D}_i$ . Thus  $J_{N(D_e)}(D_e, h_{N(D_e)})$  is a piecewise convex function with  $D_e$ .

Then by following the similar discussion in the proof of Lemma 1, the piecewise convexity of  $J_{N(D_e)}(D_e)$  with  $D_e$  can be also proved.

This completes the proof.

According to Lemma 2 and by checking (3), the objective function of Problem 1 is also a piecewise convex function with  $D_e$ . Thus to solve Problem 1, we only need to solve one convex optimization problem (whose global optimal solution is achievable) in each interval of  $D_e$ , i.e., for  $D_e \in \mathcal{D}_i, \, \forall i \in \mathcal{I}$ , and select the minimum total energy consumption among  $|\mathcal{I}|$  intervals. Specifically, define the following problem:

Problem 2:

$$E^*(i) \triangleq \min_{D_e} \quad (J_{i+1}(D_e) + E_l(D - D_e))$$
s.t. 
$$\left(D - \frac{f_l^U T}{c_0}\right) \le D_e \le D, \qquad (17)$$

$$D_e \in \mathcal{D}_i. \qquad (18)$$

To find  $E^*(i)$ , bisection search method can be used since both  $J_{i+1}(D_e)$  and  $E_l(D-D_e)$  are convex with  $D_e$  and Problem 2 is a convex optimization problem.

Then we only need to select the minimum  $E^*(i)$  among  $i \in \mathcal{I}$ , which is the minimal expected energy consumption of the mobile device. In other words, Problem 1 is solved optimally.

## IV. NUMERICAL RESULTS

In this section, numerical results are presented. Similar to [2], [5], [6], system parameters are as follows:  $c_0=40$ , T=20ms,  $T_f=2$ ms,  $D=4\times 10^4$  nats,  $f_l^U=0.5$ GHz,  $f_e=1$ GHz,  $k=10^{-23}$ , and W=1MHz. The normalized channel gain h is exponentially distributed (i.e., Rayleigh fading is assumed) with mean value 100.

In Fig. 1, the functions  $J_{N(D_e)}(D_e)$ ,  $J_{N(D_e)}(D_e,h)$ ,  $J_n(D_e)$ , and  $J_n(D_e,h)$  are plotted for  $n\in\{4,5\}$  and h=100. It can be seen that the functions  $J_n(D_e)$  and  $J_n(D_e,h)$  are convex with  $D_e$ , the functions  $J_{N(D_e)}(D_e)$  and  $J_{N(D_e)}(D_e,h)$  are piecewise convex with  $D_e$ , which support the claims in Lemma 1 and Lemma 2, respectively.

In Fig. 2, the total energy consumption achieved by our proposed method is plotted for  $D \in [5 \times 10^3, 4 \times 10^4]$ . Fig. 2 also shows the total energy consumption of the method in [6], which offloads all the data to the base station, and the method in [10], which computes all data locally or offloads all data to the base station, whichever has less energy consumption. It can be found that as D (the amount of data to be processed) goes higher, the total energy consumption grows, which is intuitive.

$$\alpha J_{i+1} \left( D_{e}^{\dagger}, h_{i+1} \right) + (1 - \alpha) J_{i+1} \left( D_{e}^{\dagger}, h_{i+1} \right) \\
= \alpha \cdot e \left( d_{i+1}^{*} (D_{e}^{\dagger}, h_{i+1}), h_{i+1}, T_{f}, W \right) + (1 - \alpha) \cdot e \left( d_{i+1}^{*} (D_{e}^{\dagger}, h_{i+1}), h_{i+1}, T_{f}, W \right) \\
+ \alpha J_{i} \left( D_{e}^{\dagger} - d_{i+1}^{*} (D_{e}^{\dagger}, h_{i+1}) \right) + (1 - \alpha) J_{i} \left( D_{e}^{\dagger} - d_{i+1}^{*} (D_{e}^{\dagger}, h_{i+1}) \right) \\
\geq e \left( \alpha d_{i+1}^{*} \left( D_{e}^{\dagger}, h_{i+1} \right) + (1 - \alpha) d_{i+1}^{*} \left( D_{e}^{\dagger}, h_{i+1} \right), h_{i+1}, T_{f}, W \right) \\
+ J_{i} \left( \alpha \left( D_{e}^{\dagger} - d_{i+1}^{*} (D_{e}^{\dagger}, h_{i+1}) \right) + (1 - \alpha) \left( D_{e}^{\dagger} - d_{i+1}^{*} (D_{e}^{\dagger}, h_{i+1}) \right) \right) \\
\geq e \left( d_{i+1}^{*} \left( \alpha D_{e}^{\dagger} + (1 - \alpha) D_{e}^{\dagger}, h_{i+1} \right), h_{i+1}, T_{f}, W \right) + J_{i} \left( \alpha D_{e}^{\dagger} + (1 - \alpha) D_{e}^{\dagger} - d_{i+1}^{*} (\alpha D_{e}^{\dagger} + (1 - \alpha) D_{e}^{\dagger}, h_{i+1}) \right) \\
= J_{i+1} \left( \alpha D_{e}^{\dagger} + (1 - \alpha) D_{e}^{\dagger}, h_{i+1} \right).$$
(16)

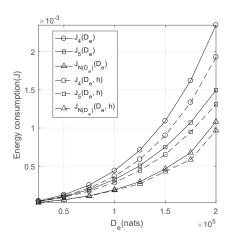


Fig. 1: Verification of Lemma 1 and Lemma 2.

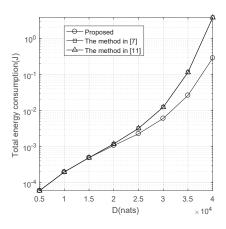


Fig. 2: Total energy consumption versus D.

In Fig. 3, the total energy consumption achieved by our proposed method and the methods in [6], [10] are plotted versus the deadline T. It can be seen that as T increases, the total energy consumption tends to decrease. This is due to the fact that as the deadline T for processing data becomes less urgent, the mobile device can have more time to complete its local computing and its data offloading, both of which can

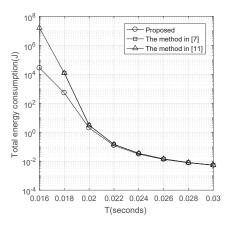


Fig. 3: Total energy consumption versus T.

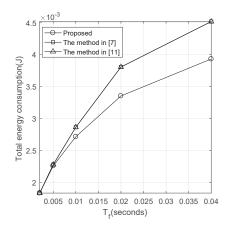


Fig. 4: Total energy consumption versus  $T_f$ .

help to save energy consumption.

In Fig. 4, T is set as 40ms. The total energy consumption achieved by our proposed method and the methods in [6], [10] are plotted versus  $T_f$ . It shows that as the duration of one fading block  $T_f$  decreases, the total energy consumption decreases. This can be explained as follows. When there are more fading blocks within duration T, the mobile user has more flexibility in adjusting the transmit power over

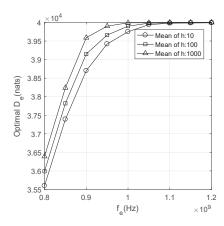


Fig. 5: Optimal offloading data  $D_e$  versus  $f_e$ .

time, i.e., transmit with high power in case of good channel condition and transmit with low power in case of bad channel condition. Thus, more efficient utilization of the transmit power is achieved, which leads to energy saving.

From Fig. 2, Fig. 3, and Fig. 4, it can be seen that our proposed method always has less energy consumption than the methods in [6], [10]. The reason is as follows. The methods in [6], [10] set  $D_e = D$  or  $D_l = D$ , which means that they only provide a feasible solution to Problem 1. Our method provides the optimal solution to Problem 1, and thus, has superior performance than the methods in [6], [10].

Fig. 5 plots the optimal offloading data  $D_e$  under various computation capability  $f_e$  and various mean values of normalized channel gain h. From Fig. 5, it can be seen that as the computation capability  $f_e$  goes up, the mobile device tends to offload more data to the base station until all the data is offloaded to the base station. It can be also seen that as the mean of h grows, the mobile device would like to offload more data to the base station. Indeed, with better wireless channel, the mobile device can send more data to the base station without increasing its power consumption.

## V. CONCLUSION

In this paper, we study the energy-efficient offloading in MEC over multiple fading blocks. An offloading strategy is proposed, which targets minimizing the total expected energy consumption of the mobile device, by selecting the amount of data nats for offloading. An optimization problem is formulated, which is non-convex. We show that the objective function of the optimization problem is piecewise convex and develop an optimal solution for the problem. Since privacy is an important issue for the application of MEC [16], [17], in the future, by combining some interesting discussions on security and privacy over data offloading as in [18]–[20], we wish to extend our work in MEC.

#### REFERENCES

- X. Ma, Y. Zhao, L. Zhang, H. Wang, and L. Peng, "When mobile terminals meets the clouds: Computation offloading as the bridge," *IEEE Netw.*, vol. 27, no.5, pp. 28-33, Sep./Oct. 2013.
- [2] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-edge computing: Partial computation offloading using dynamic voltage scaling," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4268-4282, Oct. 2016.
- [3] O. Muñoz, A. Pascual-Iserte, and J. Vidal, "Optimization of radio and computational resources for energy efficiency in latency-constrained application offloading," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp.4738-4755, Oct. 2015.
- [4] C. You and K. Huang, "Exploiting non-causal CPU-state information for energy-efficient mobile cooperative computing," *IEEE Trans. Wireless Commn.*, vol. 17, no. 6, pp. 4104-4117, Jun. 2018.
- [5] C. You, K. Huang, H. Chae, and B. H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans. Wireless Commn.*, vol. 16, no. 3, pp. 1397-1411, Mar. 2017.
- [6] 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, June 2015.
- [7] Z. Sheng, C. Mahapatra, V. C. M. Leung, M. Chen, and P. K. Sahu, "Energy efficient cooperative computing in mobile wireless sensor networks," *IEEE Trans. Cloud Comput.*, vol. 6, no. 1, pp. 114-126, Jan.-Mar., 2018.
- [8] X. Cao, F. Wang, J. Xu, R. Zhang, and S. Cui, "Joint computation and communication cooperation for energy-efficient mobile edge computing," *IEEE Internet of Things Journal*, accepted.
- [9] C. You, K. Huang, and H. Chae, "Energy efficient mobile cloud computing powered by wireless energy transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1757-1771, May 2016.
- [10] 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 Commn.*, vol. 12, no. 9, pp. 4569-4581, Sept. 2013.
- [11] J. Zhao, R. S. Allison, M. Vinnikov, and S. Jennings, "Estimating the motion-to-photon latency in head mounted displays," in *Proc. 5th IEEE* Virtual Reality, Los Angeles, CA, U.S., Mar. 2017.
- [12] D. Tse, and P. Viswanath, Fundamentals of Wireless Communication. New York: Cambridge University Press, 2005.
- [13] D. P. Bertsekas, Dynamic Programming and Optimal Control. Nashua: Athena Scientific, 2017.
- [14] R. T. Rockafellar, Convex Analysis. Princeton: Princeton Univ. Press, 1970.
- [15] S. P. Boyd and L. Vandenberghe, Convex Optimization. Cambridge: Cambridge University Press, 2004.
- [16] W. Tu and L. Lai, "Keyless authentication and authenticated capacity, IEEE Trans. Inf. Theory, vol. 64, no. 5, pp. 3696-3714, May 2016.
- [17] W. Tu, M. Goldenbaum, L. Lai, and H. V. Poor, "On simultaneously generating multiple keys in a joint source-channel model, *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 2, pp. 298-308, Feb. 2017.
- [18] W. Tu and L. Lai, "On the simulatability condition in key generation over a non-authenticated public channel, in *Proc. IEEE International Symposium on Information Theory (ISIT)*, pp. 720-724, Hongkong, China, 2015.
- [19] W. Tu and L. Lai, "Function computation with privacy constraints, in Proc. 51st Asilomar Conference on Signals, Systems, and Computers, pp. 1672-1676, Pacific Grove, CA, 2017.
- [20] W. Tu and L. Lai, "Keyless authentication over noisy channel, in *Proc. 50st Asilomar Conference on Signals, Systems, and Computers*, pp. 1665-1669, Pacific Grove, CA, 2016.