Minimizing Energy and Latency in FD MEC Through Multi-objective Optimization

Mahmoud T. Kabir[1], Muhammad R. A. Khandaker[2], and Christos Masouros[1]

[1]Dept. of Electronic and Electrical Engineering, University College London, United Kingdom

[2]School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh

Email: kabir.tukur.15@ucl.ac.uk, m.khandaker@hw.ac.uk, c.masouros@ucl.ac.uk

Abstract—In this paper, we consider a multi-user full duplex (FD) mobile edge computing (MEC) system, where a FD base station (BS) integrated with a MEC server, simultaneously transmits information signals to downlink users through the downlink and receive computation tasks for execution by the MEC server from mobile devices in the uplink. We study the trade-off between the offloading energy and latency by minimize the total offloading energy and latency via a weighted multi-objective optimization problem. First, we propose a design strategy based on the traditional interference suppression in the downlink, and then, we proposed a second design based on downlink interference exploitation. We employ Lagrangian methods to solve the two non-convex designs via an iterative algorithm. Simulation results not only show the trade-off between the offloading energy and latency, but also show the substantial gains achieved by the proposed FD schemes compared with the baseline half duplex schemes.

I. INTRODUCTION

The next generation 5G network aims at providing higher data rate and low latency. Full duplex (FD) communications is widely recognised as one of the key technologies for providing higher data rate. Following the considerable breakthroughs in terms of self interference (SI) cancellation [1], various practical implementation issues such as protocols and resource allocation algorithms have been investigated. [2] proposed a power efficient resource allocation design to study the trade-off between the UL and DL power consumption for a multiuser FD communication system with a FD BS serving uplink (UL) and downlink (DL) users, simultaneously. In [3], the authors studied the resource allocation for distributed antenna systems in a similar system setup. In [**?**], [4], [5], the authors used the knowledge of the DL signals at the FD BS to exploit multi-user interference (MUI) instead treating the MUI as unwanted as in traditional interference suppression techniques. In addition, it was shown that by exploiting the downlink MUI, the downlink transmit power is reduced significantly, although the downlink MUI is exploited, the power gains extend to the uplink through the self-interference serving as the link [4].

On the other hand, the unprecedented growth of mobile data traffic calls for significant increase in the wireless network capacity. Furthermore, the data traffic growth is not matched by improvement on mobile devices in terms of energy consumption and data storage capabilities. A promising solution is to enable mobile devices (MDs) offload their intensive and latency-critical computation tasks to nearby servers for execution through mobile edge computing (MEC). In this way,

the battery life, data storage capabilities and reliability of the MDs are enhanced [6]. In quest to reap the benefits of the MEC, several resource allocation designs have been proposed. In [7], an offline heuristic algorithm was proposed to minimize the average completion time of multiple users for partitioning and scheduling the offloading of their computations. [8] investigated resource allocation design for MEC systems based on time- and orthogonal frequency-division multiple access offloading by considering the local computation capabilities of the users. In [9], a wireless powered multiuser MEC system was proposed where the devices depend on their harvested energy to compute locally or offload tasks to the MEC server while the energy consumption of the MEC server is minimized. However, in all the above works, the authors focused on single-objectives i.e., either energy consumption or latency objectives. In [10], the authors studied the effects of using multiple access points (APs) with computation capabilities for offloading tasks in order to minimize the energy consumption and latency for fixed and elastic central processing unit (CPU) frequency. However, the authors assumed fixed transmitting and receiving power in their analysis, and in addition, the authors like the authors in [7]–[9] focused on half duplex transmission.

Accordingly, in this paper, we study a multiuser FD MEC-supported system. The main contributions of this paper are: 1) the system model: which comprises a FD BS equipped with a MEC server. The FD BS sends data signals in the DL and receive computation tasks to be executed by the MEC server through the UL, 2) the optimisation problem formulation: which involves minimising two desirable but conflicting system objectives, namely the total offloading energy and latency, that are handled through a weighted multi-objective formulation in order to study their trade-off which is highly dependent on the optimisation variables. To the best of our knowledge these have not been studied before in the literature.

II. SYSTEM MODEL

The system consists of a FD BS, integrated with a MEC server, with N transmit and N receive antennas simultaneously serving K single-antenna downlink users and J single-antenna mobile devices. In this system, the downlink users receive information signals from the FD BS, while, the mobile devices leverage the MEC server at the FD BS to offload its latency-sensitive computation tasks, which can not be locally executed,

to be executed by the MEC server. In this paper, we assume that the FD BS knows all the channel state information (CSI) from and to all the users in the system. We focus on slow fading channel scenario, where the channels change at the beginning of each frame. Thus, to facilitate the channel realization in practice, handshaking is performed between the FD BS and all users. As the channel changes slowly, pilot signals are usually embedded in the data packets, which allows the FD BS to constantly update the CSI estimation of the transmission links of the users and devices.

A. Downlink Transmission

For the transmission of information signal from the FD BS to the i-th downlink user, let $\mathbf{h}_i \in \mathbb{C}^{N \times 1}$ and $\mathbf{w}_i \in \mathbb{C}^{N \times 1}$ be the channel vector and the beamforming vector between the FD BS and the i-th downlink user, respectively. Hence, the received signals at the i-th downlink user is given by

$$y_i = \mathbf{h}_i^H \sum_{k=1}^{K} \mathbf{w}_k + \sum_{j=1}^{J} \sqrt{p_j} \ell_{j,i} + n_i, \tag{1}$$

where, p_j and $n_i \sim \mathcal{CN}\left(0, \sigma_i^2\right)$ denote the transmit power for the j-th mobile device and the additive white Gaussian noise at the i-th user, respectively. $\ell_{j,i}$ is the channel between the j-th mobile device and the i-th downlink user.

B. Computation Offloading

We denote the computation task to be offloaded to the MEC server for execution in bits at the j-th mobile device as q_j, which are classified as either energy consuming or time consuming tasks for the battery-constrained and time-constrained mobile device [6]. Computation offloading to the MEC server involves the transmission of the computation tasks to the FD BS by each mobile device and downloading[1] of the results by each user. Hence, we define the transmission rate of the j-th mobile device with bandwidth B as

$$r_j = B \log_2 \left(1 + \gamma_j\right), \tag{2}$$

where

$$\gamma_j = \frac{p_j \left| \mathbf{g}_j^H \mathbf{u}_j \right|^2}{\sum_{n \neq j}^{J} p_n \left| \mathbf{g}_n^H \mathbf{u}_j \right|^2 + s_j + \sigma_j^2 \left\| \mathbf{u}_j \right\|^2}. \tag{3}$$

In addition, $\mathbf{g}_j \in \mathbb{C}^{N \times 1}$ denotes the channel between the FD BS and the j-th mobile device and $\mathbf{n}_j \sim \mathcal{CN}(0, \sigma_j^2)$ is the additive white Gaussian noise at the FD BS. We denote $\mathbf{u}_j \in^{N \times 1}$ as the receive beamforming vector for the j-th mobile device. In this paper, to reduce complexity, we adopt zero-forcing (ZF) beamforming at the FD BS for the detection of the offloaded tasks. Furthermore, due to the simultaneous transmission and reception at the FD BS, there is a strong interference called self-interference (SI) that degrades the reception of the of-floaded computation tasks at the FD BS. In the literature,

there are different SI mitigation techniques which could be employed to reduce the effects of SI. In order to isolate our proposed scheme from the specific implementation of any passive or active SI mitigation techniques, we model the SI after cancellation as $s_j = \sum_{i=1}^{K} \text{Tr} \left\{ \mathbf{w}_i \mathbf{w}_i^H \mathbf{H}_{\text{SI}}^H \mathbf{u}_j \mathbf{u}_j^H \mathbf{H}_{\text{SI}} \right\}$ [2], where the matrix $\mathbf{H}_{\text{SI}} \in \mathbb{C}^{N \times N}$ denotes the SI channel at the FD BS. Accordingly, given the computation task q_j to be offloaded by the j-th mobile device, the total offloading latency is defined as the time taken to offload the task q_j to the FD BS plus the time taken for the FD BS to compute the corresponding result. This can be expressed as

$$T_j^{\text{total}} = \underbrace{\frac{q_j}{r_j}}_{t_{off,j}} + \underbrace{\frac{q_j L_{\text{BS},j}}{f_{\text{BS}}}}_{t_{\text{BS},j}}. \tag{4}$$

We denote $t_{off,j}$ as the time is takes to offload task q_j to the FD BS and $t_{\text{BS},j}$ as the computation time at the FD BS for task q_j, where $L_{\text{BS},j}$ (cycles/bit) is the number of CPU cycles required to compute 1 input bit of q_j at the FD BS and f_{BS} (cycles/s) is the CPU frequency of the FD BS. Thus, the corresponding total computation energy consumed in the offloading process by all the mobile devices is

$$E_{off} = \sum_{j=1}^{J} p_j t_{off,j}. \tag{5}$$

We note here the dependency of the transmit power of the mobile devices and the downlink beamforming vectors, in that, p_j through the SI term (s_j) is a function of $\mathbf{w}_i$, which in turn is a function of p_j through (1).

III. PROBLEM FORMULATION

Our main objective in this paper is to study the trade-off between two important and desirable system objectives, namely, the total offloading energy and the total offloading latency. In practice, there always exist a trade-off between these two objectives, in that, on one hand, an increase in the offloading energy implies increase in transmit power of the mobile devices and in essence, leads to a decrease in the offloading latency and vice versa. In order to efficiently analyse and address this trade-off between these objectives, we adopt a weighted multi-objective optimization (MOO) that aims at minimizing the two objectives by jointly optimizing the downlink beamforming vectors and the transmit power for each mobile device, while satisfying the total offload-ing latency requirement constraint and downlink users QoS constraints as well as the power constraints. In the following subsections, we present two strategies for the trade-off design, one based on classical interference suppression and one based on interference exploitation.

A. Trade-off Optimization based on Interference Suppression

First, in this section, we define the signal-to-interference plus noise ratio (SINR) at the i-th downlink user that promotes interference suppression (IS) transmission based on (1) as

$$\Gamma_i^{\text{DL}} = \frac{|\mathbf{h}_i^H \mathbf{w}_i|^2}{\sum_{k \neq i}^{K} |\mathbf{h}_i^H \mathbf{w}_k|^2 + \sum_{j=1}^{J} p_j |\ell_{j,i}|^2 + \sigma_i^2}. \tag{6}$$

---

[1]As the computational results are usually small, in our analysis we ignore the downloading time and the power consumed during transmitting and receiving the results [6], [9]. Thus, in this paper, our focus is particularly on the offloading of the tasks. We set aside the consideration of the results downloading from the FD BS to the mobile devices for our future work.

Thus, based on the DL SINR expression in (6) the MOOP based on IS can be mathematically formulated as

$$P(1): \min_{\{\mathbf{w}_i\},\{p_j\}} c_1 \cdot E_{off} + c_2 \cdot \sum_{j=1}^{J} t_{off,j}$$

$$\text{s.t.} \quad A1: \frac{q_j}{r_j} + \frac{q_j L_{\text{BS},j}}{f_{\text{BS}}} \leq T, \forall j, \quad A2: \Gamma_i^{\text{DL}} \geq \Gamma_i, \forall i, \quad (7)$$

$$A3: 0 \leq p_j \leq P_{\max}^{\text{MD}}, \forall j, \quad A4: \sum_{i=1}^{K} \|\mathbf{w}_i\|^2 \leq P_{\max}^{\text{DL}},$$

where c_1 and c_2 are the weights given to the two objectives, respectively, with $c_1 + c_2 = 1$. Constraints A1 ensures the total offloading latency of each mobile device does not exceed the required threshold T. Constraint A2 ensures a certain QoS for the downlink user and constraints A3 and A4 are the maximum power constraints for each mobile device and for downlink transmission, respectively.

At this point, we emphasize the flexibility provided by the MOOP (7) with respect to optimization variables. There is a strong interdependency between the optimization variables, in that, increasing the transmit power of the mobile devices in order to satisfy the latency constraints increases the co-channel interference (CCI) to the downlink users. At the same time, increasing the downlink transmit power to satisfy the downlink SINR constraints due to the increase in CCI, increases the SI power, which hinders the reception of the offloaded computation tasks.

The optimization problem (7) is non-convex and in general difficult to solve partly due to the fractional objective functions. In order to solve (7), in the following we develop a tractable approach to obtain the optimal resource allocation in an iterative manner.

First, given a fixed power p_j for each mobile device, the problem reduces to obtaining the beamforming vectors for the downlink users. Thus, it can be seen that obtaining the beamforming vectors $\mathbf{w}_i$, for $i = 1,\ldots,K$ in (7) aims at minimizing the downlink transmit power in order minimize the SI power to satisfy the constraints. Thus, this can be obtained by solving the following subproblem

$$P(1.1): \min_{\{\mathbf{w}_i\}} \sum_{i=1}^{K} \|\mathbf{w}_i\|^2 \quad (8)$$

$$\text{s.t.} \quad A1, A2, A4.$$

The optimization problem (8) is non-convex but can be easily solved through semidefinite relaxation (SDR). By defining $\mathbf{W}_i = \mathbf{w}_i \mathbf{w}_i^H, \mathbf{H}_i = \mathbf{h}_i \mathbf{h}_i^H$, the SDR formulation of (8) is given by

$$\widetilde{P}(1.1): \min_{\{\mathbf{W}_i \geq 0\}} \sum_{i=1}^{K} \text{Tr}\{\mathbf{W}_i\}$$

$$\text{s.t.} \widetilde{A1}: \tau_j - \gamma_j \leq 0, \forall j, \quad \widetilde{A4}: \sum_{i=1}^{K} \text{Tr}(\mathbf{W}_i) \leq P_{\max}^{\text{DL}},$$

$$\widetilde{A2}: \frac{\text{Tr}(\mathbf{H}_i \mathbf{W}_i)}{\Gamma_i} \geq \sum_{k \neq i}^{K} \text{Tr}(\mathbf{H}_i \mathbf{W}_k) + \sum_{j=1}^{J} p_j |\ell_{j,i}|^2 + \sigma_i^2, \forall i, \quad (9)$$

where $\tau_j = 2^{\frac{q_j}{B\left(T - \frac{q_j L_{\text{BS},j}}{f_{\text{BS}}}\right)}} - 1$.

The SDR formulation (9) is convex and can be solved by standard convex solvers. Please note that, the formulation in (9) is a relaxed form of (8) where the rank one constraint on $\mathbf{W}_i$ has been dropped. If the resulting solution $\mathbf{W}_i$ after solving (9) is rank one, the optimal $\mathbf{w}_i$ can be obtained by applying eigenvalue-decomposition (EVD), otherwise, ramdomization technique [11] can be used to retrieve $\mathbf{w}_i$.

Accordingly, for fixed downlink beamforming vectors $\mathbf{w}_i$, for $i = 1,\ldots,K$, the transmit power for the mobile devices can be obtained by solving the following subproblem

$$P(1.2): \min_{\substack{\{p_j\},\{a_j\}, \\ \{b_j\}}} c_1 \cdot \left(\sum_{j=1}^{J} q_j a_j\right) + c_2 \cdot \left(\sum_{j=1}^{J} q_j b_j\right)$$

$$\text{s.t.} \quad A5: \frac{p_j}{r_j} \leq a_j, \quad A6: \frac{1}{r_j} \leq b_j, \quad (10)$$

$$\widetilde{A1}: \tau_j - \gamma_j \leq 0, \forall j, \quad A3: 0 \leq p_j \leq P_{\max}^{\text{MD}}, \forall j.$$

Here, we introduce auxiliary variables a_j and b_j for $j = 1,\ldots,J$. In order to solve (10) we analyse the problem using Lagrangian method. Thus, the Lagrange function of problem (10) is

$$\mathcal{L}\left(p_j, a_j, b_j, \lambda_j, \mu_j, \beta_j, \nu_j\right) = c_1 \sum_{j=1}^{J} q_j a_j + c_2 \sum_{j=1}^{J} q_j b_j$$

$$+ \sum_{j=1}^{J} \lambda_j \left(p_j - a_j r_j \left(\mathbf{w}_i, p_j\right)\right) + \sum_{j=1}^{J} \mu_j \left(1 - b_j r_j \left(\mathbf{w}_i, p_j\right)\right)$$

$$+ \sum_{j=1}^{J} \beta_j \left(\tau_j - \gamma_j \left(\mathbf{w}_i, p_j\right)\right) + \sum_{j=1}^{J} \nu_j \left(p_j - P_{\max}^{\text{MD}}\right), \quad (11)$$

where $\lambda_j, \mu_j, \beta_j, \nu_j$ are the Lagrange multipliers for constraints A5, A6, $\widetilde{A1}$ and A3, respectively. Based on the definition of Karush-Kuhn-Tucker (KKT) conditions, we have

$$\frac{\partial \mathcal{L}}{\partial p_j} = \lambda_j - \lambda_j a_j \frac{\partial r_j}{\partial p_j} - \mu_j b_j \frac{\partial r_j}{\partial p_j} - \beta_j \frac{\partial \gamma_j}{\partial p_j} + \nu_j = 0, \quad (12)$$

$$\frac{\partial \mathcal{L}}{\partial a_j} = c_1 q_j - \lambda_j r_j = 0, \quad \frac{\partial \mathcal{L}}{\partial b_j} = c_2 q_j - \mu_j r_j = 0, \quad (13)$$

$$\lambda_j \left(p_j - a_j r_j\right) = 0, \quad \mu_j \left(1 - b_j r_j\right) = 0, \quad (14)$$

$$\beta_j \left(\tau_j - \gamma_j\right) = 0, \quad \nu_j \left(p_j - P_{\max}^{\text{MD}}\right) = 0. \quad (15)$$

From (13) and (14) we have $\lambda_j = \frac{c_1 q_j}{r_j}$, $\mu_j = \frac{c_2 q_j}{r_j}$, $a_j = \frac{p_j}{r_j}$ and $b_j = \frac{1}{r_j}$, respectively. Furthermore, notice that the optimal solution (p_j^*, a_j^*, b_j^*) of problem (10) satisfies the KKT conditions of the following J subproblems

$$\min_{p_j} \quad \lambda_j p_j - \lambda_j a_j r_j \left(\mathbf{w}_i, p_j\right) - \mu_j b_j r_j \left(\mathbf{w}_i, p_j\right)$$

$$\text{s.t.} \quad \widetilde{A1}: \tau_j - \gamma_j \leq 0, \quad (16)$$

$$A3: 0 \leq p_j \leq P_{\max}^{\text{MD}}.$$

Algorithm 1 Iterative algorithm for solving problem (7)

1: **Initialization:**
   Set $p_j = P_{\max}^{\text{MD}}$, for $j = 1, \ldots, J$,
   Obtain $\mathbf{w}_i$, for $i = 1, \ldots, K$, by solving subproblem (9)
   Repeat
   Loop
2: Compute λ_j, μ_j, a_j and b_j, for $j = 1, \ldots, J$,
3: Update p_j, β_j and ν_j, for $j = 1, \ldots, J$,
   until convergence. End Loop
4: Update $\mathbf{w}_i$, for $i = 1, \ldots, K$ through solving (9)
   Until stopping criterion is satisfied.
5: **Output**: $\{\mathbf{w}_i^*\}$ and $\{p_j^*\}$.

It is easy to see that the KKT conditions for the subproblem (16) are the same as that of problem (10) and are given by

$$\lambda_j - \lambda_j a_j \frac{B}{\ln 2} \frac{\Xi_j}{(1+\gamma_j)} - \mu_j b_j \frac{B}{\ln 2} \frac{\Xi_j}{(1+\gamma_j)} - \beta_j \Xi_j + \nu_j = 0,$$
(17)

$$\beta_j (\tau_j - \gamma_j) = 0,$$
(18)

$$\nu_j \left(p_j - P_{\max}^{\text{MD}}\right) = 0.$$
(19)

where $\Xi_j = \frac{\left|\mathbf{g}_j^H \mathbf{u}_j\right|^2}{s_j + \sigma_j^2 \|\mathbf{u}_j\|^2}$. From (17), we see that the optimal p_j^* is

$$p_j^* = \frac{B}{\ln 2} \frac{\lambda_j a_j + \mu_j b_j}{\lambda_j - \beta_j^* \Xi_j + \nu_j^*} - \frac{1}{\Xi_j},$$
(20)

where β_j^* and ν_j^* satisfy the KKT conditions (18) and (19), respectively. In the following, we examine 3 cases in order to obtain $\{p_j^*, \beta_j^*, \nu_j^*\}$:

1) From (18) and (19) we have $p_j^* \in \left(\frac{\tau_j}{\Xi_j}, P_{\max}^{\text{MD}}\right)$ for $\beta_j^* = \nu_j^* = 0$. In this case, $p_j^* = M_j$ where $M_j = \frac{B}{\ln 2} \frac{\lambda_j a_j + \mu_j b_j}{\lambda_j} - \frac{1}{\Xi_j}$ according to (20). Thus, we have $p_j^* = M_j$ and $\beta_j^* = \nu_j^* = 0$ if $M_j \in \left[\frac{\tau_j}{\Xi_j}, P_{\max}^{\text{MD}}\right]$.

2) If $M_j < \frac{\tau_j}{\Xi_j}$ implies that $\beta_j^* > 0$ from the constraints (20). Therefore, $p_j^* = \frac{\tau_j}{\Xi_j}$ and $\nu_j^* = 0$ according to (18) and (19). By substituting these in (20) gives $\beta_j^* = \frac{\lambda_j}{\Xi_j} - \frac{B}{\ln 2} \frac{\lambda_j a_j + \mu_j b_j}{\tau_j + 1}$.

3) Similarly, if $M_j > P_{\max}^{\text{MD}}$ implies that $\nu_j^* > 0$. In this case, $p_j^* = P_{\max}^{\text{MD}}$ and $\beta_j^* = 0$ according to (19) and (18) and putting these into (20) gives $\nu_j^* = \frac{B}{\ln 2} \frac{\Xi_j(\lambda_j a_j + \mu_j b_j)}{P_{\max}^{\text{MD}} \Xi_j + 1} - \lambda_j$.

Accordingly, from these cases the solution to the subproblem shown in (10) are given by

$$p_j^* = \begin{cases} \frac{\tau_j}{\Xi_j}, & \text{for } M_j < \frac{\tau_j}{\Xi_j}, \\ M_j, & \text{for } \frac{\tau_j}{\Xi_j} \leq M_j \leq P_{\max}^{\text{MD}}, \\ P_{\max}^{\text{MD}}, & \text{for } M_j > P_{\max}^{\text{MD}}, \end{cases}$$
(21)

$$\beta_j^* = \begin{cases} \frac{\lambda_j}{\Xi_j} - \frac{B}{\ln 2} \frac{\lambda_j a_j + \mu_j b_j}{\tau_j + 1}, & \text{for } M_j < \frac{\tau_j}{\Xi_j}, \\ 0, & \text{elsewhere}, \end{cases}$$
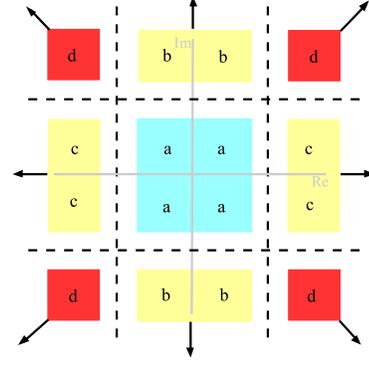(22)

Fig. 1: 16QAM constellation points

$$\nu_j^* = \begin{cases} 0, & \text{for } M_j \leq P_{\max}^{\text{MD}}, \\ \frac{B}{\ln 2} \frac{\Xi_j(\lambda_j a_j + \mu_j b_j)}{P_{\max}^{\text{MD}} \Xi_j + 1} - \lambda_j, & \text{elsewhere}. \end{cases}$$
(23)

Algorithm 1 summarizes the step by step procedure for solving the optimization (7) based on IS.

B. Trade-off Optimization based on Constructive Interference

In this section, we formulate the MOOP based on constructive interference (CI). The basic idea of CI is that, the knowledge of the downlink data signals at the FD BS can be used to exploit the multiuser interference rather than suppress it as in Section III-A. The concept of CI has been thoroughly studied in the literature for both PSK and QAM modulation in [12]–[14] and references therein, where analytical criteria are also derived. For notational convenience, we focus on QAM here. To formulate the MOOP based on CI, we first rewrite the received signal at the i-th downlink user as

$$\widetilde{y}_i = \mathbf{h}_i^H \left(\sum_{k=1}^K \mathbf{w}_k d_k\right) = \mathbf{h}_i^H \mathbf{x},$$
(24)

where we have omitted the noise term, $\mathbf{x} = \sum_{k=1}^K \mathbf{w}_k d_k$ and the unit-energy QAM symbol for the i-th downlink user is represented as d_k.

To illustrate the concept of CI for QAM modulation we provide a schematic representation for 16QAM constellation points in Fig. 1. Based on [13], to guarantee CI for the constellation points, we write the SINR constraints for the downlink users to exploit the specific detection regions for each group of constellation points separately as detailed below. From Fig. 1, to ensure CI is achieved and the constellation points are received in the correct detection region for the downlink users, the following constraints are adopted. Note that the dotted lines in Fig. 1 represent the decision boundaries.

- For the group of constellation points in the box labelled "a" in Fig. 1, since they are all surrounded by the decision boundaries, the constraints should guarantee that the received signals achieve the exact constellation point so as not to exceed the decision boundaries. The constraints are

$$\text{B2a}: \Re(\widetilde{y}_i) = \sqrt{\Gamma_i \sum_{j=1}^J p_j^{\text{CI}} |\ell_{j,i}|^2 + \Gamma_i \sigma_i^2} \Re(d_i),$$

$$\text{B2b} : \mathfrak{I}(\widetilde{y}_i) = \sqrt{\Gamma_i \sum_{j=1}^{J} p_j^{\text{CI}} |\ell_{j,i}|^2 + \Gamma_i \sigma_i^2} \, \mathfrak{I}(d_i),$$

where $\mathfrak{R}$ and $\mathfrak{I}$ are the real and imaginary parts, respectively.

- For the group of constellation points labelled "b" in Fig. 1, the constraints should guarantee that the received signals fall in the detection region away from the decision boundaries, which is the real axis. The constraints are

$$\text{B2a} : \mathfrak{R}(\widetilde{y}_i) = \sqrt{\Gamma_i \sum_{j=1}^{J} p_j^{\text{CI}} |\ell_{j,i}|^2 + \Gamma_i \sigma_i^2} \, \mathfrak{R}(d_i),$$

$$\text{B2b} : \mathfrak{I}(\widetilde{y}_i) \gtrless \sqrt{\Gamma_i \sum_{j=1}^{J} p_j^{\text{CI}} |\ell_{j,i}|^2 + \Gamma_i \sigma_i^2} \, \mathfrak{I}(d_i).$$

- For the group of constellation points labelled "c" in Fig. 1, the constraints should guarantee that the received signals fall in the detection region away from the decision boundaries, which is the imaginary axis. The constraints are

$$\text{B2a} : \mathfrak{R}(\widetilde{y}_i) \gtrless \sqrt{\Gamma_i \sum_{j=1}^{J} p_j^{\text{CI}} |\ell_{j,i}|^2 + \Gamma_i \sigma_i^2} \, \mathfrak{R}(d_i),$$

$$\text{B2b} : \mathfrak{I}(\widetilde{y}_i) = \sqrt{\Gamma_i \sum_{j=1}^{J} p_j^{\text{CI}} |\ell_{j,i}|^2 + \Gamma_i \sigma_i^2} \, \mathfrak{I}(d_i).$$

- For the group of constellation points labelled "d" in Fig. 1, the constraints should guarantee that the received signals fall in the detection region away from the decision boundaries. Here, the constellation points are not surrounded by the decision boundaries and therefore have a larger detection region that extend infinitely. The constraints are

$$\text{B2a} : \mathfrak{R}(\widetilde{y}_i) \gtrless \sqrt{\Gamma_i \sum_{j=1}^{J} p_j^{\text{CI}} |\ell_{j,i}|^2 + \Gamma_i \sigma_i^2} \, \mathfrak{R}(d_i),$$

$$\text{B2b} : \mathfrak{I}(\widetilde{y}_i) \gtrless \sqrt{\Gamma_i \sum_{j=1}^{J} p_j^{\text{CI}} |\ell_{j,i}|^2 + \Gamma_i \sigma_i^2} \, \mathfrak{I}(d_i).$$

Therefore, by adopting the required DL SINR constraints B2a and B2b for the corresponding group constellation points, the MOOP based on CI can be mathematically formulated as

$$\text{P(2)} : \min_{\mathbf{x}, \{p_j^{\text{CI}}\}} \quad c_1 \cdot E_{off}^{\text{CI}} + c_2 \cdot \sum_{j=1}^{J} t_{off,j}^{\text{CI}}$$

$$\text{s.t.} \quad \text{B1} : \frac{q_j}{r_j^{\text{CI}}} + \frac{q_j L_{\text{BS},j}}{f_{\text{BS}}} \leq T, \forall j, \quad \text{B2a, B2b,} \forall i, \tag{25}$$

$$\text{B3} : p_j^{\text{CI}} \leq P_{\max}^{\text{MD}}, \forall j, \quad \text{B4} : \|\mathbf{x}\|^2 \leq P_{\max}^{\text{DL}}.$$

Here, $t_{off,j}^{\text{CI}} = \frac{q_j}{r_j^{\text{CI}}}$ and $E_{off}^{\text{CI}} = \sum_{j=1}^{J} p_j^{\text{CI}} t_{off,j}^{\text{CI}}$, where $r_j^{\text{CI}} = B \log_2 \left(1 + \gamma_j^{\text{CI}} \right), \gamma_j^{\text{CI}} = \frac{p_j |\mathbf{g}_j^H \mathbf{u}_j|^2}{s_j^{\text{CI}} + \sigma_j^2 \|\mathbf{u}_j\|^2}$ and $s_j^{\text{CI}} = \left| \mathbf{u}_j^H \mathbf{H}_{\text{SI}} \mathbf{x} \right|^2$. The

MOOP (25) is non-convex. We solve (25) in a similar fashion to Section III-A.

For fixed power p_j^{CI}, the variable $\{\mathbf{x}\}$ can be obtained by solving the following subproblem

$$\text{P(2.1)} : \min_{\mathbf{x}} \quad \|\mathbf{x}\|^2$$
$$\text{s.t.} \quad \text{B1, B2a, B2b, B4.} \tag{26}$$

Unlike the conventional scheme, the subproblem (26) is convex and can be solved using standard convex solvers. Accordingly, given the variable $\{\mathbf{x}\}$, the transmit power for the mobile devices can be obtained by solving the following subproblem

$$\text{P(2.2)} : \min_{\substack{\{p_j^{\text{CI}}\}, \{a_j^{\text{CI}}\}, \\ \{b_j^{\text{CI}}\}}} c_1 \cdot \left(\sum_{j=1}^{J} q_j a_j^{\text{CI}} \right) + c_2 \cdot \left(\sum_{j=1}^{J} q_j b_j^{\text{CI}} \right)$$

$$\tag{27}$$

$$\text{s.t.} \quad \text{B5} : \frac{p_j^{\text{CI}}}{r_j^{\text{CI}}} \leq a_j^{\text{CI}}, \quad \text{B6} : \frac{1}{r_j^{\text{CI}}} \leq b_j^{\text{CI}},$$

$$\widetilde{\text{B1}} : \tau_j - \gamma_j^{\text{CI}} \leq 0, \forall j, \quad \text{B3} : p_j^{\text{CI}} \leq P_{\max}^{\text{MD}}, \forall j,$$

To solve (27), we analyse the problem using Lagrangian method in a similar fashion to Section III-A. Accordingly, we obtain the following as the corresponding solutions to the problem (27)

$$\lambda_j^{\text{CI}} = \frac{c_1 q_j}{r_j^{\text{CI}}}, \; \mu_j^{\text{CI}} = \frac{c_2 q_j}{r_j^{\text{CI}}}, \; a_j^{\text{CI}} = \frac{p_j^{\text{CI}}}{r_j^{\text{CI}}}, \; b_j^{\text{CI}} = \frac{1}{r_j^{\text{CI}}},$$

$$p_j^{\text{CI}^*} = \begin{cases} \frac{\tau_j}{\Xi_j^{\text{CI}}}, & \text{for } M_j^{\text{CI}} < \frac{\tau_j}{\Xi_j^{\text{CI}}}, \\ M_j^{\text{CI}}, & \text{for } \frac{\tau_j}{\Xi_j^{\text{CI}}} \leq M_j^{\text{CI}} \leq P_{\max}^{\text{MD}}, \\ P_{\max}^{\text{MD}}, & \text{for } M_j^{\text{CI}} > P_{\max}^{\text{MD}}, \end{cases}$$

$$\beta_j^{\text{CI}^*} = \begin{cases} \frac{\lambda_j^{\text{CI}}}{\Xi_j^{\text{CI}}} - \frac{B}{\ln 2} \frac{\lambda_j^{\text{CI}} a_j^{\text{CI}} + \mu_j^{\text{CI}} b_j^{\text{CI}}}{\tau_j + 1}, & \text{for } M_j^{\text{CI}} < \frac{\tau_j}{\Xi_j^{\text{CI}}}, \\ 0, & \text{elsewhere,} \end{cases}$$

$$\nu_j^{\text{CI}^*} = \begin{cases} 0, & \text{for } M_j^{\text{CI}} \leq P_{\max}^{\text{MD}}, \\ \frac{B}{\ln 2} \frac{\Xi_j^{\text{CI}}(\lambda_j^{\text{CI}} a_j^{\text{CI}} + \mu_j^{\text{CI}} b_j^{\text{CI}})}{P_{\max}^{\text{MD}} \Xi_j^{\text{CI}} + 1} - \lambda_j^{\text{CI}}, & \text{elsewhere,} \end{cases}$$

where $\Xi_j^{\text{CI}} = \frac{|\mathbf{g}_j^H \mathbf{u}_j|^2}{s_j^{\text{CI}} + \sigma_j^2 \|\mathbf{u}_j\|^2}$ and $M_j^{\text{CI}} = \frac{B}{\ln 2} \frac{\lambda_j^{\text{CI}} a_j^{\text{CI}} + \mu_j^{\text{CI}} b_j^{\text{CI}}}{\lambda_j^{\text{CI}}} - \frac{1}{\Xi_j^{\text{CI}}}$. Please note that, a summary of the Algorithm to solve (25) based on CI is omitted due to space constraints, however, (25) can be solved by following the same steps as shown in Algorithm 1 with the corresponding CI based solutions shown in Section III-B.

IV. SIMULATION RESULTS

We consider the system with the FD BS at the centre of a cell with $N = 6$ antennas, each for transmitting and receiving. We assume $K = 4$ DL users and $J = 2$ MDs, are randomly and uniformly distributed between the distance of 2m and 20m. We model the channels to the MDs and DL users as Rayleigh fading. The SI channel is modelled as Rician fading channel
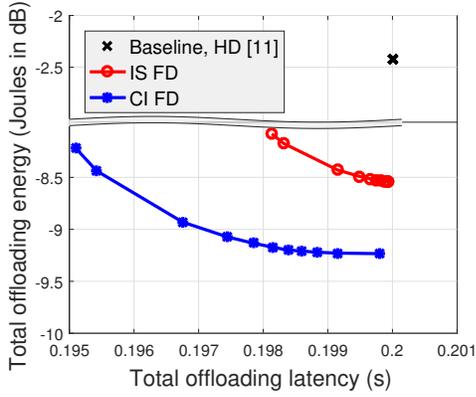
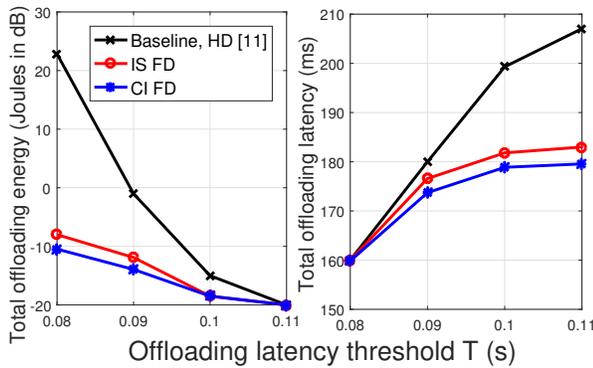Fig. 2: Trade-off plot energy vs latency

Fig. 3: Energy and latency vs latency threshold

with Rician factor 6dB. Furthermore, we consider $\sigma_i = \sigma_j = -60$dBm, $P_{max}^{MD} = 32$dBm, $P_{max}^{DL} = 40$dBm, $T = 100$ms, $\Gamma_i = 4$dB, $B = 1$MHz, $q_j = 10^5$, $L_{BS,j} = 10^3$ and $f_{BS} = 10^{10}$ [6]. Our baseline is the HD scheme in [10]. For fair comparison, here, the data rate of HD is set equal to the one for FD which requires that the individual data rate of the mobile devices and downlink users are double the ones for the FD case, due to the slotted HD transmission.

Fig. 2 shows the trade-off between the total offloading energy and latency by varying the weights c_1 and c_2 between 0 to 1, respectively. First, it can be seen that an increase in the energy leads to the decrease in the latency and vice versa. This is as a result of the dependency of the optimization variables. On one hand, increasing the transmit power of the mobile devices in order to satisfy the latency constraints and minimize the offloading latency, increases the CCI to the downlink users. Hence, the downlink transmit power is increased to accommodate for the increase in CCI, which in turn increases the SI. In essence, this leads a continuous increase in the uplink and downlink transmit power, thus, the offloading energy increases. On the contrary, reducing the transmit powers in order to reduce the CCI and SI, and minimize the offloading energy, gives rise to an increase in the offloading latency. In addition, the proposed CI scheme consumes less energy and time as compared to the IS scheme.

This is because less DL transmit power is required to satisfy the DL SINR constraint, hence, reduced SI, as compared to the IS scheme where interference is rather suppressed. More importantly, it can be seen that the FD schemes out perform their HD counterpart. Furthermore, Fig. 3 shows the total offloading energy and latency for different latency thresholds where we set $c_1 = c_2 = 0.5$. In terms of offloading energy, for all schemes, it can be seen that increasing T reduces the energy consumption, while in terms of offloading latency, all schemes are proportional to increase in latency. Besides, the FD schemes outperforms the HD scheme in both cases. This further highlights the effectiveness of the proposed FD schemes.

V. CONCLUSION

In this paper, we studied the offloading energy and latency trade-off in a multiuser FD system that performs both data transmission and MEC. The proposed FD schemes show a promising performance improvement over the HD scheme.

REFERENCES

[1] M. Duarte, A. Sabharwal, V. Aggarwal, R. Jana, K. Ramakrishnan, C. W. Rice, and N. Shankaranarayanan, "Design and characterization of a full-duplex multiantenna system for WiFi networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1160–1177, 2014.

[2] Y. Sun, D. W. K. Ng, J. Zhu, and R. Schober, "Multi-objective optimization for robust power efficient and secure full-duplex wireless communication systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 8, pp. 5511–5526, 2016.

[3] D. W. K. Ng, Y. Wu, and R. Schober, "Power efficient resource allocation for full-duplex radio distributed antenna networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2896–2911, 2016.

[4] M. T. Kabir, M. R. A. Khandaker, and C. Masouros, "Reducing self-interference in full duplex transmission by interference exploitation," *IEEE Global Communications Conference*, pp. 1–6, Dec 2017.

[5] M. T. Kabir, M. R. Khandaker, and C. Masouros, "Robust energy harvesting FD transmission: Interference suppression vs exploitation," *IEEE Communications Letters*, vol. 22, no. 9, pp. 1866–1869, 2018.

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

[7] L. Yang, J. Cao, H. Cheng, and Y. Ji, "Multi-user computation partitioning for latency sensitive mobile cloud applications," *IEEE Transactions on Computers*, vol. 64, no. 8, pp. 2253–2266, 2015.

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

[9] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1784–1797, 2018.

[10] T. Q. Dinh, J. Tang, Q. D. La, and T. Q. Quek, "Offloading in mobile edge computing: Task allocation and computational frequency scaling," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3571–3584, 2017.

[11] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[12] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Transactions on Signal processing*, vol. 63, no. 14, pp. 3628–3640, 2015.

[13] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Symbol-level multiuser miso precoding for multi-level adaptive modulation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5511–5524, 2017.

[14] M. R. Khandaker, C. Masouros, and K.-K. Wong, "Constructive interference based secure precoding: A new dimension in physical layer security," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2256–2268, 2018.