



HAL
open science

Maximum Completion Time Optimization in Uplink Multi-subcarrier NOMA Systems

Yongna Guo, Chi Wan Sung, Ye Liu, Chung Shue Chen

► **To cite this version:**

Yongna Guo, Chi Wan Sung, Ye Liu, Chung Shue Chen. Maximum Completion Time Optimization in Uplink Multi-subcarrier NOMA Systems. IEEE Wireless Communications Letters, 2023. hal-04071599

HAL Id: hal-04071599

<https://hal.science/hal-04071599>

Submitted on 17 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Maximum Completion Time Optimization in Uplink Multi-subcarrier NOMA Systems

Yongna Guo, *Student Member, IEEE*, Chi Wan Sung, *Senior Member, IEEE*, Ye Liu, *Member, IEEE*,
and Chung Shue Chen, *Senior Member, IEEE*

Abstract—Maximum completion time is a key performance metric for future wireless applications that have stringent latency requirements. This letter studies its optimization in uplink multi-subcarrier non-orthogonal multiple access (NOMA) networks. The joint optimization problem of power allocation, user pairing and scheduling is analyzed and proved to be NP-hard. To tackle the problem, the power allocation subproblem for a single NOMA pair is first derived. Next, for the joint user pairing and scheduling subproblem, an optimal algorithm for a special case is constructed and a suboptimal algorithm for the general case is designed. Our proposed scheme is applicable to data aggregation in wireless federated learning networks, and simulation results validate its superior performance.

Index Terms—NOMA, scheduling, maximum completion time, power allocation, user pairing

I. INTRODUCTION

Non-orthogonal multiple access (NOMA) has been considered in many practical applications, such as mobile edge computing (MEC) [1], cache-enabled networks [2], and wireless federated learning [3]. Since many applications have stringent delay requirements, the delay minimization problem is investigated in NOMA-enabled Internet of Things (IoT), MEC, and mmWave-MEC networks in [4], [5], [6], and [7], respectively. Those works, however, consider only the single-subcarrier case. For multi-subcarrier systems, the problem is more complicated and the performance metric can be defined in different ways. The work [8] formulates the problem and adopts the maximum completion time as the optimization criterion. It assumes for the superposition of two packets, when a user finishes its transmission, the other user can transmit without interference and increase the code rate for the remaining part of his packet. Such an assumption, however, increases overhead and complexity in practical systems.

In this letter, we re-formulate the completion-time maximization problem with the more practical assumption of constant transmission rate of one packet. In addition, we consider the power allocation problem, which is ignored in [8]. Power allocation and user pairing are to be jointly optimized with the scheduling problem, and the scheduling algorithm is redesigned. We derive an optimal power allocation scheme for one NOMA pair. We prove that the joint optimization problem

is NP-hard. Especially, we propose an optimal scheduling algorithm for the special case when the number of users is twice of the number of subcarriers. A suboptimal algorithm is also given for the general case. In simulation, we further validate the good performance of our proposed scheme compared to orthogonal multiple access (OMA) and the algorithm proposed in [8]. Moreover, we apply our proposed NOMA scheme for data aggregation in practical wireless federated learning scenario and show its good performance in terms of transmission time.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider the uplink of a wireless network with one base station (BS) and K IoT devices indexed by $\mathcal{K} \triangleq \{1, 2, \dots, K\}$. The total bandwidth of the system is denoted by B , which consists of M subcarriers, each of bandwidth $W = B/M$. We assume that these subcarriers are narrowband and undergo flat block fading [9]. We denote the channel gain between each IoT device k and the BS by h_k , which may include both large scale fading and small scale fading, and is assumed to be constant during *one transmission round*, as in [10]. The transmission round is defined as the time period until all K devices finish transmitting one packet. The noise power of each subcarrier at the BS is denoted by $n = N_0W$, where $N_0/2$ is the noise spectral density. Besides, each IoT device k has one packet of L_k bits, to be uploaded to the BS. We assume that each IoT device can only occupy one subcarrier to upload the data and all L_k bits are to be transmitted to the BS via one packet transmission without splitting. For example, in practical synchronous distributed learning or computing systems, the devices will have some computed data to transmit to the server for data aggregation. The transmit data of each device will be sent in one packet without splitting to reduce overheads. The power constraint for each IoT device is denoted by P . To shorten the transmission delay, we assume NOMA is applied and there are at most two packets superposed for one transmission considering the decoding complexity and the error propagation problem of the successive interference cancellation (SIC) receiver. Without loss of generality, K is assumed an even number, for otherwise we may add one dummy device with 0 bit to transmit.

Suppose that devices $i, j \in \mathcal{K}$ are superposed to transmit in one subcarrier with power p_i and p_j , respectively, and the SIC receiver is applied at the BS. Assume that the receiver decodes the packet of device i first, subtracts it from the received signal, and then decodes the packet of device j . The transmission rates of devices i and j are given by:

Yongna Guo and Chi Wan Sung are with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong. (e-mail:yongnaguo2-c@my.cityu.edu.hk; albert.sung@cityu.edu.hk). This work was supported in part by City University of Hong Kong (Project 7005619). (*Corresponding author: Chi Wan Sung.*)

Ye Liu is with the Department of Communications and Networking, Xi'an Jiaotong-Liverpool University, China. (e-mail:ye.liu02@xjtlu.edu.cn)

Chung Shue Chen is with the Department of Machine Learning and Systems, Nokia Bell Labs, Paris-Saclay, Nozay 91620, France. (e-mail: chung_shue.chen@nokia-bell-labs.com)

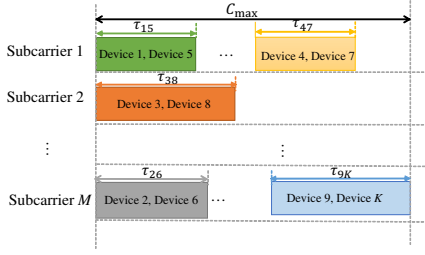


Fig. 1. Illustration of maximum completion time.

$$R_i = W \log_2 \left(1 + \frac{h_i p_i}{h_j p_j + n} \right), \quad (1)$$

$$R_j = W \log_2 \left(1 + \frac{h_j p_j}{n} \right) \quad (2)$$

The transmission time for device k and the completion time for device pair (i, j) are $T_k = \frac{L_k}{R_k}$ and $T_{ij} = \max(T_i, T_j)$, respectively. Note that T_{ji} represents the completion time for decoding the packet of device j first and is not equal to T_{ij} in general. Let $\tau_{ij} \triangleq \min(T_{ij}, T_{ji})$ be the completion time of the NOMA pair (i, j) and thus $\tau_{ij} = \tau_{ji}$. Besides, the transmission rate of each device remains unchanged for any single packet because we assume the channel coding and the modulation scheme cannot be adaptively changed within a packet.

Let the binary indicator $I_{i,j}^m = I_{j,i}^m = 1$ when devices $i, j \in \mathcal{K}$, $i \neq j$, are superposed to transmit on subcarrier $m \in \mathcal{M}$, otherwise $I_{i,j}^m = I_{j,i}^m = 0$. The total completion time of subcarrier m , denoted by C_m , is given by:

$$C_m = \sum_{i=1}^K \sum_{j=i+1}^K I_{i,j}^m \tau_{ij}. \quad (3)$$

In this paper, we consider how to pair up these devices and schedule these pairs on the subcarriers to minimize the maximum completion time, i.e., $C_{\max} \triangleq \max_{m \in \mathcal{M}} C_m$, see Fig. 1. The problem can be formulated mathematically as:

$$\min_{\mathbf{I}, \mathbf{p}} C_{\max}, \quad (4)$$

subject to

$$\begin{aligned} C1: & 0 \leq p_k \leq P, \forall k \in \mathcal{K}, \\ C2: & I_{i,j}^m \in \{0, 1\}, \forall i, j \in \mathcal{K}, \forall m \in \mathcal{M}, \\ C3: & \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{K}} I_{i,j}^m = 1, \forall i \in \mathcal{K}, \end{aligned} \quad (5)$$

where $\mathbf{I} = (\mathbf{I}^1, \mathbf{I}^2, \dots, \mathbf{I}^M)$, $\mathbf{I}^m = [I_{i,j}^m]$ is a symmetric matrix and $\mathbf{p} = (p_1, p_2, \dots, p_K)$. $C1$ represents the power constraint, $C2$ represents the binary constraint on $I_{i,j}^m$, and $C3$ ensures that every device is paired up with exactly one device and assigned to exactly one channel, and thus the packets of all devices will be uploaded to the BS. In this letter, we will investigate the joint optimization problem of power allocation, user pairing, and the scheduling of the NOMA pairs to minimize the maximum completion time in uplink multicarrier NOMA NB-IoT systems. We call this problem MIN-NOMA-DELAY.

III. POWER ALLOCATION SCHEME OF A DEVICE PAIR

To investigate the optimization problem, we first consider the power allocation subproblem to minimize the transmission time of a pair of devices sharing the same channel using NOMA, which is defined as the time that both devices to complete their transmissions.

Consider devices i and j , which form such a pair. The completion time for device pair (i, j) , τ_{ij} , depends on p_i and p_j , and the decoding order of the pair. To minimize τ_{ij} , we need to solve the following problem:

$$\min \tau_{ij} \triangleq \min(T_{ij}, T_{ji}), \quad (6)$$

subject to

$$0 \leq p_i \leq P, 0 \leq p_j \leq P. \quad (7)$$

First, consider the optimal power allocation for minimizing T_{ij} , where the BS decodes the packet of device i before that of device j .

Proposition 1. To minimize T_{ij} , an optimal power allocation scheme for user i and j is $(p_i^*, p_j^*) = (P, \min(\tilde{p}_j, P))$, where \tilde{p}_j is the root of the following equation:

$$\left(1 + \frac{h_j p_j}{n} \right)^{L_i} = \left(1 + \frac{h_i P}{h_j p_j + n} \right)^{L_j}. \quad (8)$$

Proof. Since the signal of device i is decoded first and will not affect the data rate of device j as can be seen from (2), it is clear from (1) that we can maximize R_i by setting $p_i^* = P$, which is optimal.

Note that (8) is obtained from (1) and (2) by letting $T_i = T_j$, i.e.,

$$L_i / R_i = L_j / R_j. \quad (9)$$

When p_j increases from 0 to ∞ , the left-hand expression increases monotonically from 1 to ∞ while the right-hand expression decreases monotonically from a certain value to 1. Therefore, there is a unique positive real number \tilde{p}_j such that (8) and (9) hold when $p_j = \tilde{p}_j$. Note that the value of \tilde{p}_j can be numerically obtained by one-dimensional search methods such as the bisection search.

Now we prove by contradiction that $p_j^* = \min(\tilde{p}_j, P)$ is optimal. First, consider the case where $p_j^* = \tilde{p}_j \leq P$. Suppose the optimal solution for p_j is $\hat{p}_j \neq p_j^*$. If $\hat{p}_j < \tilde{p}_j$, from (2), we have $\hat{R}_j < \tilde{R}_j$ and thus $\hat{T}_j > \tilde{T}_j$. If $\hat{p}_j > \tilde{p}_j$, from (1), we have $\hat{R}_i < \tilde{R}_i$ and thus $\hat{T}_i > \tilde{T}_i$. Therefore, for both conditions, $\hat{T}_{ij} > T_{ij}^*$, which leads to a contradiction. The second case where $p_j^* = P < \tilde{p}_j$ can be proved similarly. \square

With p_i^* and p_j^* , the minimum transmission time T_{ij}^* of the (i, j) pair can be determined for decoding i first. Following the same procedure, we can calculate the minimum transmission time T_{ji}^* , for which the packet of j is decoded first. Thus, we obtain $\tau_{ij}^* = \min(T_{ij}^*, T_{ji}^*)$.

For some special cases, we can decide the optimal decoding order without determining explicitly both T_{ij}^* and T_{ji}^* .

Proposition 2. If a NOMA pair $(i, j) \in \mathcal{K}^2$ satisfies the following condition:

$$\left(1 + \frac{h_j P}{n} \right)^{L_i} \leq \left(1 + \frac{h_i P}{h_j P + n} \right)^{L_j}, \quad (10)$$

decoding the packet of device i first is optimal.

Proof. From (10), we can obtain that:

$$\frac{L_i}{W \log_2(1 + \frac{h_i P}{h_j P + n})} \leq \frac{L_j}{W \log_2(1 + \frac{h_j P}{n})} = T_{ij}^*. \quad (11)$$

Besides, we have:

$$\frac{L_j}{W \log_2(1 + \frac{h_j P}{n})} \leq \frac{L_j}{W \log_2(1 + \frac{h_j P}{h_i p_i^* + n})} \leq T_{ji}^*, \quad (12)$$

where $\frac{L_j}{W \log_2(1 + \frac{h_j P}{h_i p_i^* + n})}$ represents the transmission time for device j when decoding the packet of device j first and it is no more than T_{ji}^* . Hence, it is obvious that $T_{ij}^* \leq T_{ji}^*$ and thus we should always decode the packet of device i first if device pair (i, j) satisfies (10). \square

IV. NP-HARDNESS OF MIN-NOMA-DELAY

In this section, we will show the NP-hardness of MIN-NOMA-DELAY by proving that its decision version, NOMA-DELAY, is NP-complete. The *question* of the decision version is this: Can the devices, which are paired up via NOMA and assigned to transmit on each subcarrier, finish all the transmissions within time T ?

We will show the NP-completeness of NOMA-DELAY by a reduction from the OMA-DELAY problem, which is equivalent to the multiprocessor scheduling problem, an NP-complete problem [8]. The instance of OMA-DELAY is the same as that of NOMA-DELAY, while the *question* of OMA-DELAY is: Can the packets of the devices, assigned to transmit on each subcarrier, finish all the transmissions within time T ? To prove the NP-completeness of NOMA-DELAY, for clarity, we first calculate the minimum transmission time for each device $k \in \mathcal{K}$ in OMA-DELAY, denoted by T_{OMA}^k , as $T_{\text{OMA}}^k = \frac{L_k}{W \log_2(1 + \frac{h_k P}{n})}$. In wireless networks, the channel gains are random, and thus T_{OMA}^k 's are all distinct with probability 1. Assume the devices are indexed in such a way that $T_{\text{OMA}}^1 < T_{\text{OMA}}^2 < \dots < T_{\text{OMA}}^K$.

Proposition 3. NOMA-DELAY is NP-complete.

Proof. Given any instance of OMA-DELAY, we reduce it to an instance of NOMA-DELAY by adding one dummy device $k' \in \mathcal{K}'$ for each device $k \in \mathcal{K}$, where \mathcal{K}' denotes the set of the dummy devices. For each device $k \in \mathcal{K}$, the corresponding dummy device $k' \in \mathcal{K}'$ is with channel gain $h_{k'} = \delta$ and $L_{k'} = \mu T_{\text{OMA}}^k$, where δ is an arbitrarily small positive number, and $\mu = W \log_2(1 + \frac{\delta P}{n - \delta P})$. Besides, let the noise power at the BS be $n - \delta P$. The reduction can be done in polynomial time. We claim that after some optimal NOMA pairing algorithm, the transmission time for all NOMA pairs of the instance from NOMA-DELAY are the same as the transmission time of all OMA devices of the instance from OMA-DELAY.

When devices $k \in \mathcal{K}$ and $k' \in \mathcal{K}'$ transmit the packets to the BS on one subcarrier using OMA, respectively, the minimum transmission time for k and k' , denoted by T_k^* and $T_{k'}^*$, can be derived as follows:

$$T_k^* = \frac{L_k}{W \log_2(1 + \frac{h_k P}{n - \delta P})}, T_{k'}^* = \frac{L_{k'}}{W \log_2(1 + \frac{\delta P}{n - \delta P})}. \quad (13)$$

It is obvious that $T_k^* < T_{k'}^* = T_{\text{OMA}}^k$. Therefore, we have $T_1^* < T_{1'}^* < T_2^* < T_{2'}^* < \dots < T_K^* < T_{K'}^*$, since we can let δ be arbitrarily small to make sure that $T_{k'}^* < T_{k+1}^*$. Suppose k and k' are paired up and the packet of device k is decoded first. We have:

$$\frac{L_k}{W \log_2(1 + \frac{h_k P}{\delta P + n - \delta P})} = \frac{L_{k'}}{W \log_2(1 + \frac{\delta P}{n - \delta P})}, \quad (14)$$

where the left term and the right term represent the transmission time for device k and k' when decoding the packet of device k first, respectively. Besides, (14) satisfies (9) and thus we have $T_{kk'}^* = T_{k'}^* = T_{\text{OMA}}^k$. Since the transmission time of a NOMA pair cannot be smaller than its constituents, $T_{kk'}^* = T_{\text{OMA}}^k$ is the minimum transmission time for pair (i, j) .

For the instance of NOMA-DELAY, the devices are paired up into NOMA pairs and scheduled for transmission. After some NOMA pairing algorithm, there are K NOMA pairs. Without loss of generality, we assume the NOMA pairs are ordered in the ascending order of the transmission time, i.e., $\tau_1 \leq \tau_2 \leq \dots \leq \tau_K$, where τ_k represents the transmission time for the k -th NOMA pair. Suppose devices (i_k, j_k) form the k -th NOMA pair. Since the transmission time of each NOMA pair cannot be smaller than the transmission time of its constituents, i.e., $\tau_{ij}^* \geq \max(T_i^*, T_j^*)$, we have $\tau_1 \geq \max(T_{i_1}^*, T_{j_1}^*) \geq \max(T_{1'}^*, T_1^*) = T_{1'}^*$. Then we have $\tau_2 \geq \max(T_{i_2}^*, T_{j_2}^*)$ and $\tau_2 \geq \tau_1 \geq \max(T_{i_1}^*, T_{j_1}^*)$, and thus $\tau_2 \geq \max(T_{i_2}^*, T_{j_2}^*, T_{i_1}^*, T_{j_1}^*) \geq \max(T_{2'}^*, T_2^*, T_{1'}^*, T_1^*) = T_{2'}^*$. By induction, we have that $\tau_k \geq T_{k'}^*$, $k = 1, 2, \dots, K$, no matter how the devices are paired up. An optimal NOMA pairing algorithm is to pair up each device $k \in \mathcal{K}$ with the corresponding dummy device $k' \in \mathcal{K}'$ since we can obtain $\tau_k = T_{k'}^* = T_{\text{OMA}}^k$ by (14), which also proves our claim. Therefore, it is obvious that all transmissions of the derived NOMA-DELAY instance can finish within time T if and only if all transmissions of OMA-DELAY can finish within time T .

The problem NOMA-DELAY is clearly in NP, since it is efficiently verifiable. Therefore, it is NP-complete. \square

V. USER PAIRING AND SCHEDULING ALGORITHM

In this section, we propose user pairing and scheduling algorithms for MIN-NOMA-DELAY. Though MIN-NOMA-DELAY is NP-hard, we can solve it optimally with an efficient algorithm when $K = 2M$. Then we propose a suboptimal two-step algorithm for the general optimization problem.

A. An Optimal Scheduling Algorithm for $K = 2M$

When $K = 2M$, after some NOMA pairing algorithm, there are M NOMA pairs. It is optimal to schedule exactly one NOMA pair to one distinct subcarrier. The optimization problem degenerates to the NOMA pairing problem, which is to pair the devices up to minimize the maximum transmission time for the NOMA pairs. The problem can be represented by a graph. Let $G = (V, E)$ be a complete weighted graph, where V represents the set of the K devices. Define $w_{ij} = \tau_{ij}^*$ as the weight of each edge $e_{ij} \in E$. To minimize the maximum transmission time, we can simply remove the edges in the graph G one by one in descending order of w_{ij} until we find the edge without which the remaining edges do not contain a perfect matching. Hence, the weight of this edge,

Algorithm 1 Binary Search for NOMA Pairing

Input: $G = (V, E)$;
Output: C_{\max} ;
1: Initialize $l := |V|/2, r := |E|$;
2: Sort all edges e_{ij} 's in descending order of w_{ij} 's;
3: **while** TRUE **do**
4: let $m := \lfloor \frac{r+l}{2} \rfloor$;
5: **if** $I_{\text{pm}}(V, E(m)) = 0$ **then**
6: let $l := m + 1$;
7: **else**
8: **if** $I_{\text{pm}}(V, E(m-1)) = 0$ **then**
9: **return** $C_{\max} = \max_{e_{ij} \in E(m)} w_{ij}$;
10: **else**
11: let $r := m - 1$;
12: **end if**
13: **end if**
14: **end while**

ω_{ij}^* , is exactly the optimal solution to MIN-NOMA-DELAY when $K = 2M$, i.e., $C_{\max}^* = \omega_{ij}^*$. However, this sequential searching method is slow, and thus we further apply binary search to find ω_{ij}^* . Let $E(n)$ denote the set of n edges with the n smallest weights. Let $I_{\text{pm}}(V, E)$ be the indicator function defined by $I_{\text{pm}}(V, E) = 1$ if there exists a perfect matching in the graph $G = (V, E)$, and $I_{\text{pm}}(V, E) = 0$ otherwise. Denote the left and right boundaries of the binary search range by l and r , respectively. We initialize $l := |V|/2$ since a perfect matching exists only if the graph contains at least $|V|/2$ edges and let $r := |E|$. The edges are sorted in descending order of weights, w_{ij} 's. The search procedure is as follows: Let the middle element in the search range be $m := \lfloor \frac{r+l}{2} \rfloor$. If $(V, E(m))$ does not include a perfect matching, we let $l := m + 1$ and repeat the search procedure. Otherwise, we consider two sub-cases: (a) if $(V, E(m-1))$ does not include a perfect matching, the search is done, and returns the maximum completion time $C_{\max} = \max_{e_{ij} \in E(m)} w_{ij}$; (b) if $(V, E(m-1))$ includes a perfect matching, we let $r := m - 1$ and repeat the procedure. The pseudo-code for the binary search NOMA pairing is shown as Algorithm 1. The complexity of finding the perfect matching is the same as finding a maximum cardinality matching, which can be solved in $O(\sqrt{VE})$ time [11]. The complexity of the binary search algorithm is $O(\log E)$. Therefore, the complexity of Algorithm 1 is $O(\sqrt{VE} \log E) = O(K^{2.5} \log K)$ since $V = K$ and $E = \binom{K}{2}$.

B. A Suboptimal Two-step Scheduling Algorithm for $K > 2M$

For the general case, we propose a two-step scheduling algorithm, which first matches the users into NOMA pairs and then schedules these pairs to the subcarriers. In the first step, as a heuristic rule, we find the NOMA pairs that minimize the total completion time. This is done by finding the maximum weight matching in $G' = (V', E')$, where G' represents the complete weighted graph and V' represents the set of K devices. Define $w_{ij} = \Delta - \tau_{ij}^*$ as the weight of each edge $e_{ij}' \in E'$, where Δ is a number larger than all τ_{ij}^* 's. After the first step, the users have been paired up and the remaining subproblem degenerates to the identical-machines scheduling

problem. The longest-processing-time-first (LPT) algorithm is a classical scheduling algorithm to minimize the maximum completion time in identical-machines scheduling problems and the worst-case bound of LPT is $4/3 - 1/(3M)$ [12]. Therefore, in the second step, we apply the LPT algorithm to schedule the NOMA pairs as follows: The NOMA pairs are sorted in *descending* order of τ_{ij}^* and placed in a queue. The first pair in the queue is then assigned to the subcarrier that has the earliest idle time. The idle time of that subcarrier is then updated accordingly. The procedure is repeated until the queue becomes empty. The complexity of maximum weight matching is $O(K^3)$ [11] and the complexity of LPT is $O(K)$. The overall complexity of the heuristic scheduling algorithm is $O(K^3)$.

VI. NUMERICAL RESULTS

In this section, we first compare the performance of our proposed scheduling algorithm with OMA scheduling algorithm. Besides, we conduct more experiments considering the practical transmission time minimization problem in wireless federated learning scenario.

A. Scheduling Performance

We consider an NB-IoT system in a single cell with radius equal to 500 meters, where the BS is located at the center of the cell and the K IoT devices are randomly distributed in the cell. Each device wants to deliver a packet of size uniformly distributed between 8 bits and 1024 bits. Following the NB-IoT standard, the system bandwidth is assumed to be 180 kHz, so the bandwidth W of each channel is $180/M$ kHz. The number of subcarriers, M , varies from 5 to 25. The maximum transmit power P of each user is assumed to be 20 dBm. The channel between the IoT devices and the BS is modeled as the combination of the large-scale fading and the small-scale Rayleigh fading with variance equal to 1. The large-scale path loss effect is $128.1 + 37.6 \log_{10} d$, where d represents the distance from the IoT devices to the BS in kilometers. The noise spectral density is -174 dBm/Hz.

First, we examine the performance of our proposed scheme under $K = 2M$. The OMA scheme is used as a benchmark, which applies the LPT scheduling scheme. Besides, the UPaS NOMA scheduling algorithm proposed in [8] is also used as a benchmark, which is based on the shortest-processing-time-first (SPT) scheme. In contrast to LPT, SPT schedules the users in *ascending* order of their transmission times. The scheme OMA-SPT used in [8] is adopted as another benchmark. Fig. 2 shows the transmission time versus the number of subcarriers. The number of users is set to be $K = 2M$. From Fig. 2, we can see that our proposed optimal NOMA-opt scheduling scheme performs better than other schemes. When the number of subcarriers is $K = 15$, NOMA reduces the completion times of OMA-LPT, UPaS, and OMA-SPT by 36.8%, 43.1%, and 53.3%, respectively. Besides, OMA-LPT performs better than UPaS and OMA-SPT. The reason is that LPT-based scheduling outperforms SPT-based scheduling in terms of maximum completion time.

Next, we examine the performance of our proposed heuristic algorithm for the general $K > 2M$ case. The OMA-LPT, OMA-SPT, and UPaS are used as benchmarks. We let $M = 5$

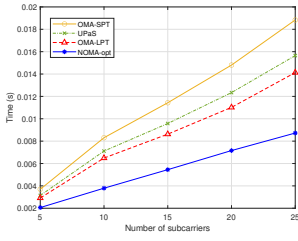


Fig. 2. The maximum completion time vs. the number of subcarriers M ($K = 2M$).

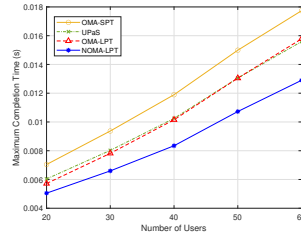


Fig. 3. The maximum completion time vs. the number of users K ($K > 2M$, $M = 5$).

and K varies from 20 to 60. Fig. 3 shows the maximum completion time versus the number of users for the general case. We can see that our proposed NOMA-LPT algorithm has shorter completion time than the other schemes. When the number of users are $K = 20$ and $K = 60$, NOMA-LPT reduces the completion time of OMA-LPT by 11.7% and 18.3%, respectively. Besides, it reduces the completion time of UPaS by 17.9% when $K = 50$.

B. Experimental Study on Federated Learning

To demonstrate the performance of our proposed algorithm in practical systems, we implement our proposed scheme on a practical wireless federated learning (FL) system for data aggregation to mitigate the transmission delay. We train a convolutional neural network (CNN) on the MNIST dataset for handwritten digit recognition. The CNN network consists of two 5×5 convolution layers followed by the ReLU activations and the maxpooling layers with 32 and 64 neurons, respectively, and final 10 fully connected layers with softmax activation. There are 259,466 learnable parameters for transmission. Following [13], we quantize each parameter into 12 bits. We apply FedAVG algorithm. The total number of users or workers, is set to be 100. The large-scale channel gains are generated in the same way as in Section VI-A while small-scale channel gains experience independent Rayleigh fading for each communication round of the learning process. The total bandwidth is set to be 5 MHz and the number of subcarriers is 5. The learning parameters are set according to [14]. The total number of the training data samples is 60,000 and the whole data set is randomly partitioned into 100 equal-sized subsets. Each user stores one subset, i.e., 600 images. The number of the communication rounds and local epochs are set to be 300 and 5, respectively. The learning rate is 0.1. The minibatch size is 10. The client fraction is 0.1, i.e., in each round, 10 users are allowed to transmit the local updated learnable variables to the BS for data aggregation. Besides, the users are selected using random selection [9]. The selected users transmit the learnable parameters to the server by either OMA or NOMA. Fig. 4 shows the accuracy and maximum completion time versus the number of communication rounds. It can be seen that the performance of NOMA is much better than OMA in terms of the transmission time. When the number of communication rounds are 200 and 300, NOMA outperforms OMA by 44.6% and 42.7%, respectively. Fig. 5 shows the accuracy versus the transmission time. It can be seen that NOMA uses shorter transmission time to achieve the same accuracy as OMA.

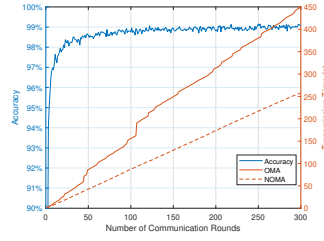


Fig. 4. Model accuracy and the transmission times of OMA and NOMA.

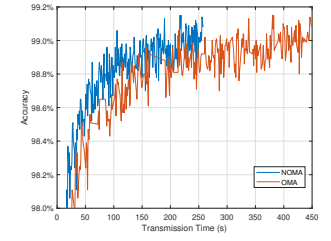


Fig. 5. Model accuracy vs. the transmission time.

VII. CONCLUSION

This work proposes a joint power allocation, user pairing and scheduling algorithm to minimize the maximum completion time in uplink multi-subcarrier NOMA networks. An optimal scheduling algorithm and an efficient heuristic are proposed for a special case and the general case, respectively. The simulation results show its good delay performance of the data aggregation of wireless federated learning. Hence, it can be considered for many practical applications, such as distributed learning and computing.

REFERENCES

- [1] Z. Ding, D. W. K. Ng, R. Schober, and H. V. Poor, "Delay minimization for NOMA-MEC offloading," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1875–1879, 2018.
- [2] Z. Ding, P. Fan, G. K. Karagiannidis, R. Schober, and H. V. Poor, "NOMA assisted wireless caching: Strategies and performance analysis," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4854–4876, 2018.
- [3] H. Sun, X. Ma, and R. Q. Hu, "Adaptive federated learning with gradient compression in uplink NOMA," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16325–16329, 2020.
- [4] D. Zhai, R. Zhang, L. Cai, and F. R. Yu, "Delay minimization for massive internet of things with non-orthogonal multiple access," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 553–566, 2019.
- [5] Y. Wu, L. P. Qian, K. Ni, C. Zhang, and X. Shen, "Delay-minimization nonorthogonal multiple access enabled multi-user mobile edge computation offloading," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 392–407, 2019.
- [6] F. Fang, Y. Xu, Z. Ding, C. Shen, M. Peng, and G. K. Karagiannidis, "Optimal resource allocation for delay minimization in NOMA-MEC networks," *IEEE Trans. Wireless Commun.*, vol. 68, no. 12, pp. 7867–7881, 2020.
- [7] J. Shi, Y. Zhou, Z. Li, Z. Zhao, Z. Chu, and P. Xiao, "Delay minimization for NOMA-mmW scheme based MEC offloading," *IEEE Internet Things J.*, 2022.
- [8] M. Mohsenivatani, Y. Liu, M. Derakhshani, S. Parsaeefard, and S. Lambotharan, "Completion-time-driven scheduling for uplink NOMA-enabled wireless networks," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1775–1779, 2020.
- [9] H. H. Yang, Z. Liu, T. Q. Quek, and H. V. Poor, "Scheduling policies for federated learning in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 68, no. 1, pp. 317–333, 2019.
- [10] W. Shi, S. Zhou, Z. Niu, M. Jiang, and L. Geng, "Joint device scheduling and resource allocation for latency constrained wireless federated learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 453–467, 2020.
- [11] C. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: algorithms and complexity*. Courier Corporation, 1998.
- [12] J. Y. Leung, *Handbook of scheduling: algorithms, models, and performance analysis*. CRC press, 2004.
- [13] X. Mo and J. Xu, "Energy-efficient federated edge learning with joint communication and computation design," *Journal of Communications and Information Networks*, vol. 6, no. 2, pp. 110–124, 2021.
- [14] Y. Li, Y. Zhou, A. Jolfaei, D. Yu, G. Xu, and X. Zheng, "Privacy-preserving federated learning framework based on chained secure multi-party computing," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6178–6186, 2020.