Connectivity Maximization in Non-orthogonal Network Slicing Enabled Industrial Internet-of-Things with Multiple Services

Bo Yin, Jianhua Tang, *Member, IEEE,* and Miaowen Wen, *Senior Member, IEEE*

***Abstract*—Industrial Internet of Things (IIoT) is a technological revolution that is profoundly reshaping the visage of industry. Facing the explosively increasing number of multi-service devices, traditional industrial network technology is no longer applicable. The advent of the fifth-generation (5G) wireless networks brings unprecedented possibilities for deploying the anticipated IIoT. To address the two main issues, i.e., connection density and multi-service requirements, in 5G empowered IIoT, we consider the non-orthogonal network slicing in this work. In particular, we jointly utilize network slicing to incorporate different types of services and exploit non-orthogonal multiple access (NOMA) to enhance the connection density. We formulate the connectivity maximization problem with joint sub-carrier association and power allocation as a mixed-integer nonlinear programming (MINLP). To tackle the intractable MINLP, we first transform it into a mixed-integer linear programming (MILP) and then simplify the MILP into an integer linear programming (ILP) by developing a simple yet effective pairing guideline. In order to further reduce the computational complexity, we then propose the alternating selection best-effort pairing (AS-BEP) algorithm with low complexity to solve the ILP effectively. Our analyses are supplemented by comprehensive simulation results that illustrate the performance superiority of the proposed algorithms to the benchmark schemes.**

***Index Terms*—Connectivity maximization, network slicing, Industrial Internet of Things (IIoT), non-orthogonal multiple access (NOMA), resource allocation.**

I. Introduction

RECENTLY, the growing popularity of the Internet of Things (IoT) is providing an up-and-coming scheme not only for the development of various home automation systems but also for different industrial applications. By grasping the advantages, Industrial Internet of Things (IIoT) is becoming a buzz word. IIoT is a group of interconnected static/mobile objects, such as devices equipped with communication, networked sensors, and control modules, connected through the Internet. The concept of IIoT paves the way for a better understanding of the manufacturing process, allowing for more efficient and sustainable production [2]. To accelerate the growth of Industry 4.0, which primarily focuses on the manufacturing industry, IIoT urgently requires in possession of strong communication characteristics such as high density with multiple services, real-time performance, high reliability, and energy efficiency. Nevertheless, indeed because of these characteristics, it is challenging to deploy an IIoT.

An increasing number of communication technologies, such as bluetooth in low energy personal area networks, Zigbee in home automation systems, WiFi, low power wide area networks (LPWAN) and widely-deployed cellular technologies (such as Long-term Evolution (LTE) and LTE-advanced (LTE-A) - 4G machine-to-machine (M2M)) were developed to support the fast growth diversity applications and communication requirements of IoT/IIoT. However, these technologies have limits in terms of connection density and coverage when it comes to covering very broad regions in the IIoT [3]–[6].

The newly emerged fifth generation (5G) wireless communication is expected to support a wide variety of usage scenarios, including enhanced mobile broadband (eMBB), ultra-reliable and low latency communications (URLLC) and massive machine type communications (mMTC) [7]. Meanwhile, the industry urgently needs to build a wireless IIoT to connect the machines, consumers, and data, which is consistent with two of the 5G usage scenarios: URLLC and mMTC [8]. From this perspective, the advent of 5G cellular networks, with the availability of a genuinely ubiquitous, heterogeneous, reliable, scalable, and cost-effective wireless communication technology, is viewed as a potentially critical driver for establishing IIoT.

There are many works on 5G empowered IIoT [9]–[12]. [9] evaluates present research efforts in IIoT from the perspective of three critical system aspects: control, networking, and computation, and concludes that 5G has the potential to aid IIoT applications by providing the needed network capacity and throughput, as well as low latency. A 5G-based communication architecture is described to facilitate the deployment of IoT system with a central controller in which multiple sensors and actuators may establish full-duplex communication connections with the central controller [10]. The authors in [11] propose an edge intelligence and blockchain-enabled IIoT architecture to accomplish flexible and highly secure edge service management. The authors of [12] offer a co-design of state estimation and wireless transmission in 5G mMTC enabled IIoT systems to improve the accuracy of estimating state parameters over resource-constrained and complicated

B. Yin and Jianhua Tang are with the Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, China. M. Wen is with the School of Electronic and Information Engineering, South China University of Technology, China. E-mails: bborg.yin@gmail.com, jtang4@e.ntu.edu.sg, eemwwen@scut.edu.cn. The corresponding author is Jianhua Tang.

The work of Jianhua Tang was supported in part by the National Nature Science Foundation of China under Grant 62001168 and in part by the Foundation and Application Research Grant of Guangzhou under Grant 202102020515. The work of Miaowen Wen was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021B1515120067. Part of this paper [1] has been presented at the 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, December 2021.

wireless networks.

In general, the majority of IIoT applications have a low throughput need, thus data rate is not a major constraint. Connecting a large number of devices to the IIoT network at a cheap cost, with limited hardware capabilities and energy consumption is the primary concern. In addition, in a real factory, IIoT can provide multiple data services, such as collecting massive process data, communicating with industrial robots, and tracking machines/parts/products on the factory [13]. Therefore, in this work, we consider a scenario with a large number of machine type communications devices (MTCDs) that simultaneously provide multiple types of services. To accomplish which, we jointly consider and resort to the following two emerging technologies:

- Non-orthogonal multiple access (NOMA). Traditional orthogonal multiple access (OMA) technologies, such as time division multiple access (TDMA) and orthogonal frequency-division multiple access (OFDMA) are unable to meet the high density requirements of IIoT due to their reliance on orthogonal resources (time, frequency) to differentiate different users [14]. In recent years, NOMA has been regarded as an attractive solution. By allocating one time-frequency channel to multiple users (utilizing superposition coding (SC) at the transmitter and successive interference cancellation (SIC) at the receiver) at the same time within the same cell, NOMA brings a number of advantages, including higher spectrum efficiency, improved network capacity, and higher cell-edge throughput, which is a promising technology in IIoT [15]–[18]. Furthermore, due to the distinctive merits, the integration of NOMA with other technologies, such as unmanned aerial vehicle (UAV), millimeter-wave (mmWave), and reconfigurable intelligent surface (RIS), is able to significantly improve wireless system performance [19]–[21].
- Network slicing. By facilitating multiple logical self-contained networks on top of a common physical network, it enables various service types for IIoT [22]. In particular, network slicing can improve the efficiency of wireless resource utilization by dynamically adjusting physical time-frequency resource between the slices. It is possible to implement autonomous systems for such resource allocations and dynamic management of resources. This technique can be used to enhance the scalability of IIoT services [23], [24].

A. Related Works

In reality, there is a scarcity of relevant research on applying NOMA and network slicing to address the issue of connection maximization for IoT/IIoT networks. However, there has already been research on using the power-domain NOMA scheme into the 5G enabled IoT/IIoT to achieve high density and massive connectivity[1]. In [25]–[28], the maximum access problem (MAP) for uplink networks with NOMA is explored. More specifically, the authors of [25] offer a method for solving MAP that incorporates admission control, user clustering, channel assignment, and power management. In graph theory, this issue is referred to as the maximal independent set problem. In [26], a scheme of uplink power-domain NOMA is proposed to allow multiple ultra-reliable MTC (uMTC) devices and mMTC devices to share the same sub-carrier. The optimization problem is decomposed into two subproblems by the order of MTC devices selection. In [27], [28], different modes (single-tone and multi-tone) and more general sub-carrier allocation schemes for maximizing the connection density in Narrowband-Internet of Things (NB-IoT) with uplink power-domain NOMA have been discussed, respectively.

With the advancement of 5G, one of the key technologies, network slicing, is expected to create different service capabilities for IIoT. [29] provides an information-theoretic framework for the non-orthogonal sharing of radio access network (RAN) resources in uplink communications from a set of eMBB, mMTC, and URLLC devices and proposes the concepts of orthogonal and non-orthogonal network slicing on the physical layer. The investigations in [30] and [31] incorporate two different types of services, i.e., URLLC and eMBB, into a single physical network by wireless network slicing. With effective admission control, the network operator obtains the maximum revenue. However, the studies in [30] and [31] are still in the category of orthogonal slicing, i.e., OMA is employed to isolate different slices.

It can be foreseen that, high connection density and heterogeneous networks with multiple services (i.e., scalability and flexibility [5]) are two critical requirements for deploying massive IIoT networks. Unfortunately, the works presented in [25]–[31] focus on only one of the two aspects, which is insufficient to tackle the issue in this context. In this paper, we combine NOMA and network slicing to construct non-orthogonal network slicing that enables the coexistence of multiple services while maximizing the connectivity, which significantly improves the scalability of the network and the spectrum efficiency of the system. As an extension of our previous work [1], which only considers uplink non-orthogonal slicing, in this paper, we take into account both uplink and downlink non-orthogonal slicing, to further improving the utilization of time-frequency resources.

B. Our Contributions

In work, we consider the non-orthogonal slicing in IIoT to support three distinct types of services. Our aim is to maximize the number of accessed devices, i.e., maximizing the connectivity. To satisfy the Quality of Service (QoS) requirements for each service/slice with high connectivity density, we utilize power-domain NOMA scheme for both the uplink and downlink. The main contributions of this paper are summarized as follows:

- We build up the system with the coexistence of NOMA and OMA to provide three distinct services. In particular, we consider the non-orthogonal network slicing to

[1]In our paper, connectivity is defined as the ability of the network to connect/interact with devices simultaneously, i.e., the number of devices that can be successfully connected to the network during a transmission time interval. Compared to other metrics, such as the sum-rate, transmit power, and spectral efficiency, etc., maximizing the connectivity, as our objective, is to satisfy the requirements of future 5G/B5G enabled IIoT for massive connection density.

provide different services for MTCDs. For this purpose, all MTCDs are classified into different sets and the same time-frequency resources are shared among the different services/slices devices.
- We formulate a connectivity maximization problem by jointly considering device-sub-carrier association and transmit power allocation of MTCDs and base station (BS). The problem is formulated as a mixed-integer nonlinear programming (MINLP).
- We transform the MINLP into a mixed-integer linear programming (MILP) and then further reduce the MILP to integer linear programming (ILP) by devising a simple but effective transmit power allocation scheme. Moreover, to further reduce the computational complexity of the ILP, we propose the *alternating selection best-effort pairing (AS-BEP)* algorithm to solve the ILP. The comprehensive simulations show that our proposed low-complexity BEP/AS-BEP algorithms outperform other existing methods and have close performance to the global optimal result (which is computationally prohibited).

C. Organization and Notations

The remainder of this article is organized as follows. Section II elaborate the general system model, including basic uplink and downlink NOMA framework and the main constraints of system. In Section III, we formulate the connectivity maximization optimization problem and transform it into an MILP. In Section IV, we propose the low-complexity algorithms to solve the simplified MILP by given power scheme. The performance of the proposed algorithms is evaluated in Section V. Finally, Section VI concludes this paper.

Notations: We use calligraphy letters to represent the sets. Vectors and matrices are in the form of lowercase and uppercase bold letters, respectively. $\mathbf{I}_N$ means N-dimensional identity matrix. $\mathbb{E}(\cdot)$ and $\|\cdot\|_2$ stand for the expectation and the Euclidean norm, respectively. $(\cdot)^T$ and $(\cdot)^H$ denote transpose and complex conjugate transpose, respectively. The distribution of a circularly symmetric complex Gaussian random variable with zero mean and variance σ^2 is denoted by $\mathcal{CN}\left(0,\sigma^2\right)$. The N-dimensional complex, real, and natural spaces are denoted by $\mathbb{C}^N$, $\mathbb{R}^N$, and $\mathbb{N}^N$, respectively. $|\mathcal{A}|$ represents the cardinality of set $\mathcal{A}$. The $\log(\cdot)$ function is the logarithm function with base 2.

II. System Model And Main Constraints

We explore a NOMA-assisted multiuser uplink/downlink Industrial Internet-of-Things scenario, in which a BS is communicating with multiple MTCDs. The BS is equipped M-antennas and each MTCD is equipped with single antenna. The MTCDs are assumed to be randomly distributed in a cell, and we denote the collection of all MTCDs as $\mathcal{C}$. In which, there are three types of services that these MTCDs can provide:

- Type A service: uploading collected data to the BS and also acquiring data from the BS. For example, the automated guided vehicles (AGVs).
- Type B service: uploading collected data only. For example, the temperature sensors.

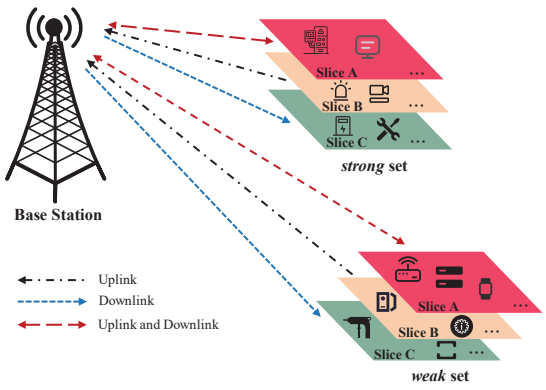


Fig. 1. System framework.

- Type C service: downloading collected data only. For example, the actuators.

We assume that each MTCD can only provide one type of service and hence we can classify MTCDs into three disjoint sets, i.e., type A set, $\mathcal{C}^A$, type B set, $\mathcal{C}^B$, and type C set, $\mathcal{C}^C$. To enable these three distinct types of services in a single IIoT, we employ three disparate network slices: slice A, slice B and slice C, respectively.

Within one transmission time interval (TTI), all MTCDs share the physical resource block (PRB), which is evenly divided into S_{tol} sub-carriers with each sub-carrier of W Hz bandwidth. We denote the set of all sub-carriers as $\mathcal{S} \triangleq \{1,\ldots,S_{tol}\}$. To fully utilize the physical resources, power-domain NOMA for both uplink and downlink is involved in this work:

- NOMA for uplink, to connect different uplink MTCDs (i.e., type A and type B devices) as many as possible.
- NOMA for downlink, to connect different downlink MTCDs (i.e., type A and type C devices) as many as possible.

Given the decoding complexity of SIC and error propagation during SIC process in practical systems, we assume that each sub-carrier can accommodate at most two devices [25]–[28], [32], i.e., one device from each of the two decoding groups. Based on this, we assume that the devices are divided into two decoding groups. A device belongs to the *strong* or *weak* set if it has a large or small channel gain respectively [2]. Thus, we denote $\mathcal{C}^A_{\mathscr{S}}$, $\mathcal{C}^B_{\mathscr{S}}$ and $\mathcal{C}^C_{\mathscr{S}}$ as type A, type B and type C devices in the *strong* set respectively, and then define $\mathcal{C}^{A,B}_{\mathscr{S}} = \mathcal{C}^A_{\mathscr{S}} \cup \mathcal{C}^B_{\mathscr{S}}$ and $\mathcal{C}^{A,C}_{\mathscr{S}} = \mathcal{C}^A_{\mathscr{S}} \cup \mathcal{C}^C_{\mathscr{S}}$. Similarly, we denote $\mathcal{C}^A_{\mathscr{W}}$, $\mathcal{C}^B_{\mathscr{W}}$, $\mathcal{C}^C_{\mathscr{W}}$, $\mathcal{C}^{A,B}_{\mathscr{W}}$ and $\mathcal{C}^{A,C}_{\mathscr{W}}$ as the corresponding devices in the *weak* set respectively. The system framework is shown in Fig. 1.

A. NOMA Uplink

We consider an uplink single-input multiple-output (SIMO) scenario with power-domain NOMA to access type A and B devices from different sets, and denote $\mathcal{C}^{A,B} = \mathcal{C}^A \cup \mathcal{C}^B$.

[2]The concept of dividing devices into *strong* and *weak* sets by means of the corresponding channel gain is motivated by the perspective that a better performance gain can be achieved if two users in a cluster have a significant disparity in channel gain [33].

The devices share the same uplink channel and can transmit their own information on the same frequency block within the same time slot. We define a binary variable $\alpha_{k,s}$ to identify the association between device and sub-carrier, i.e., $\alpha_{k,s} = 1$ if the s-th sub-carrier is allocated to the k-th device, and $\alpha_{k,s} = 0$ otherwise. The uplink channel between the k-th device and the BS is denoted as $\boldsymbol{h}_k^u \in \mathbb{C}^{M\times 1}$, whose small-scale fading is assumed to be frequency-flat[3]. Assume that x_k is the transmitted symbols of the k-th device satisfying $\mathbb{E}\left(|x_k|^2\right) = 1$ and $\mathbb{E}\left(x_i, x_j\right) = 0\ (\forall i \neq j)$. The M-dimensional uplink received signal vector at BS is represented as

$$\boldsymbol{y}=\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{C}_{\mathscr{S}}^{A,B}}\alpha_{i,s}\sqrt{p_{i,s}^u}\boldsymbol{h}_i^u x_i+\sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{C}_{\mathscr{W}}^{A,B}}\alpha_{j,s}\sqrt{p_{j,s}^u}\boldsymbol{h}_j^u x_j+\boldsymbol{n},$$

where $\boldsymbol{y} \in \mathbb{C}^{M\times 1}$; $p_{k,s}^u$ is the uplink transmit power for the k-th device over the s-th sub-carrier; $\boldsymbol{n} \sim \mathcal{CN}\left(0, N_0\mathbf{I}_M\right)$ represents an additive white Gaussian noise (AWGN) vector at the BS and N_0 is power spectral density of the AWGN. By invoking the optimal minimum mean square error (MMSE)-SIC receiver structure at the BS [15], the signals in the *strong* set are decoded first with interference, and then the signals in the *weak* set are decoded without interference (since the interferences are subtracted). As a result, for any given $s \in \mathcal{S}$, the received instantaneous signal-to-interference-plus-noise ratios (SINR) at the BS corresponding to the device in the *strong* and *weak* set are given by

$$\gamma_{i,s}^u = \frac{\alpha_{i,s}p_{i,s}^u\|\boldsymbol{h}_i^u\|_2^2}{\sum_{j\in\mathcal{C}_{\mathscr{W}}^{A,B}}\alpha_{j,s}p_{j,s}^u\|\boldsymbol{h}_j^u\|_2^2+N_0W}, \forall i \in \mathcal{C}_{\mathscr{S}}^{A,B}, \quad (1)$$

and

$$\gamma_{j,s}^u = \frac{\alpha_{j,s}p_{j,s}^u\|\boldsymbol{h}_j^u\|_2^2}{N_0W}, \quad \forall j \in \mathcal{C}_{\mathscr{W}}^{A,B}, \quad (2)$$

respectively. Then, the achievable transmission rate for the k-th device can be written as

$$r_k^u = \sum_{s\in\mathcal{S}} W\log\left(1+\gamma_{k,s}^u\right), \quad \forall k \in \mathcal{C}^{A,B}. \quad (3)$$

B. NOMA Downlink

We consider a downlink multiple-input single-output (MISO) system where an M-antenna BS transmits information to multiple single-antenna type A and C devices via NOMA, and denote $\mathcal{C}^{A,C} = \mathcal{C}^A \cup \mathcal{C}^C$. Similarly, we define a binary variable $\beta_{k,s}$ to indicate this association, i.e., $\beta_{k,s} = 1$ if the s-th sub-carrier is allocated to the k-th device, and $\beta_{k,s} = 0$ otherwise. Assume that z_k is the message to be received by k-th device with $\mathbb{E}\left(|z_k|^2\right) = 1$ and $\boldsymbol{w}_k \in \mathbb{C}^{M\times 1}$ represents the beamforming vector to the k-th device. In the downlink NOMA systems, the BS transmits the superposition coding and hence the received signal y_k at the k-th device can be expressed as

$$y_k = \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{C}^{A,C}}\beta_{k,s}\sqrt{p_{k,s}^d}\boldsymbol{h}_k^{dH}\boldsymbol{w}_i z_i + n_k, \quad \forall k \in \mathcal{C}^{A,C},$$

where $\boldsymbol{h}_k^d (\in \mathbb{C}^{M\times 1}) \sim \mathcal{CN}\left(0, \mathbf{I}_M\right)$ represents the channel coefficients from the BS to k-th device, and $p_{k,s}^d$ represents the downlink transmit power allocated by BS to the k-th device over the s-th sub-carrier; $n_k \sim \mathcal{CN}\left(0, \sigma^2\right)$ is the AWGN noise. And, we assume the decoding order as the ascending order of the channel gains [16], [32]. For simplicity, the maximum ratio transmission (MRT) linear processing is adopted at BS, and thus the beamforming vectors are designed as [34]

$$\boldsymbol{w}_k = \frac{\boldsymbol{h}_k^d}{\|\boldsymbol{h}_k^d\|_2}, \quad \forall k \in \mathcal{C}^{A,C}.$$

By employing SIC at the receiver, the signals in the *weak* set are decoded first with interference, and then the signals in the *strong* set are decoded without interference (since the interferences are subtracted). Then, for any given $s \in \mathcal{S}$, the received instantaneous SINR corresponding to the device in the *strong* and *weak* set are calculated by

$$\gamma_{i,s}^d = \frac{\beta_{i,s}p_{i,s}^d\|\boldsymbol{h}_i^d\|_2^2}{N_0W}, \quad \forall i \in \mathcal{C}_{\mathscr{S}}^{A,C}, \quad (4)$$

and

$$\gamma_{j,s}^d = \frac{\beta_{j,s}p_{j,s}^d\|\boldsymbol{h}_j^d\|_2^2}{\sum_{i\in\mathcal{C}_{\mathscr{S}}^{A,C}}\beta_{i,s}p_{i,s}^d\|\boldsymbol{h}_j^d\|_2^2+N_0W}, \forall j \in \mathcal{C}_{\mathscr{W}}^{A,C}, \quad (5)$$

respectively. Thus, the achievable transmission rate for the k-th device can be written as

$$r_k^d = \sum_{s\in\mathcal{S}} W\log\left(1+\gamma_{k,s}^d\right), \quad \forall k \in \mathcal{C}^{A,C}. \quad (6)$$

Remark 1: In the uplink, the signals in the *strong* set are decoded first with interference. Such a decoding order maximizes the channel gain advantage of the *strong* set since the signals in the *strong* set have better channel conditions, and it can accommodate more interference under the condition of the same transmit power. Whereas, to improve the fairness of devices in NOMA, the downlink decoding order is assumed as the ascending order of the channel gains, which is the completely opposite of the uplink decoding order. More details about the decoding order can be found in [16], [32]. Note that finding the optimal ordering for SIMO/MISO-NOMA is a complex optimization problem [35]. For brevity, we take the above efficient decoding order, which is in line with many literature studies [15]–[18], [25]–[28], [32].

C. Main Constraints

There are some primary constraints to be considered while using NOMA for the uplink and downlink systems.

1) Constraints for each sub-carrier: Each sub-carrier can be accessed by at most one *strong* set device and one *weak*

[3]With the notation $\boldsymbol{h}_k^u$, we can define the term *strong* or *weak* set mathematically. That is, $\mathcal{C}_{\mathscr{S}}^A \triangleq \{k|\left\|\boldsymbol{h}_k^u\right\|_2 \geq \psi^A,\ k \in \mathcal{C}^A\}$ and $\mathcal{C}_{\mathscr{W}}^A \triangleq \{k|\left\|\boldsymbol{h}_k^u\right\|_2 < \psi^A,\ k \in \mathcal{C}^A\}$, where ψ^A is a threshold. And $\mathcal{C}_{\mathscr{S}}^B$, $\mathcal{C}_{\mathscr{S}}^C$, $\mathcal{C}_{\mathscr{W}}^B$, and $\mathcal{C}_{\mathscr{W}}^C$ can be defined similarly.

set device, i.e., we have

$$0 \leq \sum_{k \in \mathcal{C}_{\mathscr{S}}^{A,B}} \alpha_{k,s} \leq 1, 0 \leq \sum_{k \in \mathcal{C}_{\mathscr{W}}^{A,B}} \alpha_{k,s} \leq 1, \forall s \in \mathcal{S}, \quad (7)$$

$$0 \leq \sum_{k \in \mathcal{C}_{\mathscr{S}}^{A,C}} \beta_{k,s} \leq 1, 0 \leq \sum_{k \in \mathcal{C}_{\mathscr{W}}^{A,C}} \beta_{k,s} \leq 1, \forall s \in \mathcal{S}, \quad (8)$$

in the uplink and downlink, respectively. More importantly, to ensure that each sub-carrier can be only assigned to either downlink or uplink, for a given $s \in \mathcal{S}$, the relationship between $\alpha_{k,s}$ and $\beta_{k,s}$ can be expressed as

$$\left(\sum_{k \in \mathcal{C}^{A,B}} \alpha_{k,s}\right)\left(\sum_{k \in \mathcal{C}^{A,C}} \beta_{k,s}\right) = 0. \quad (9)$$

Remark 2: Since each sub-carrier can be used by at most two devices, there are many possible device combinations on each sub-carrier. Using A, B and C to represent a type A, B and C device respectively. For an uplink sub-carrier, the combinations can be $\{A, B\}$, $\{A, A\}$, $\{B, B\}$, $\{A\}$, $\{B\}$ or none; for a downlink sub-carrier, the combinations can be $\{A, C\}$, $\{A, A\}$, $\{C, C\}$, $\{A\}$, $\{C\}$ or none.

2) Constraints for each device: For either uplink or downlink, each device can only occupy at most one sub-carrier for transmission[4], i.e.,

$$\begin{aligned} 0 \leq \sum_{s \in \mathcal{S}} \alpha_{k,s} \leq 1, \quad \forall k \in \mathcal{C}^{A,B}, \\ 0 \leq \sum_{s \in \mathcal{S}} \beta_{k,s} \leq 1, \quad \forall k \in \mathcal{C}^{A,C}. \end{aligned} \quad (10)$$

3) Constraints among devices/sub-carriers: Type A slice provides simultaneous uploading and downloading service for type A devices. Thus, when a type A device is accessed, it takes up both uplink and downlink sub-carriers, i.e.,

$$\sum_{s \in \mathcal{S}} \alpha_{k,s} = \sum_{s \in \mathcal{S}} \beta_{k,s}, \quad \forall k \in \mathcal{C}^{A}. \quad (11)$$

In addition, since the total number of sub-carriers is limited, not every device can be accessed, i.e.,

$$\frac{1}{2} \sum_{k \in \mathcal{C}^{A,B}} \sum_{s \in \mathcal{S}} \alpha_{k,s} + \frac{1}{2} \sum_{k \in \mathcal{C}^{A,C}} \sum_{s \in \mathcal{S}} \beta_{k,s} \leq \mathcal{S}_{tol}. \quad (12)$$

4) Constraints for power: The uplink transmit power of the k-th device over the s-th sub-carrier should not exceed the maximum transmit power of the device, P_D, i.e.,

$$0 \leq p_{k,s}^{u} \leq \alpha_{k,s} P_D, \quad \forall k \in \mathcal{C}^{A,B}, s \in \mathcal{S}, \quad (13)$$

and the transmit power allocated by BS to the k-th device over the s-th sub-carrier should not exceed the maximum power of the BS, P_B, i.e.,

$$0 \leq p_{k,s}^{d} \leq \beta_{k,s} P_B, \quad \forall k \in \mathcal{C}^{A,C}, s \in \mathcal{S}. \quad (14)$$

[4]In NB-IoT systems, there are single-tone and multi-tone transmission modes, where each device can access a single sub-carrier or a bond of contiguous sub-carriers, respectively [28]. Given the fact that multiple services exist in our system, we explore just single-mode here, which also reduces the complexity of system.

Furthermore, the total downlink transmit power should satisfy

$$0 \leq \sum_{k \in \mathcal{C}^{A,C}} \sum_{s \in \mathcal{S}} p_{k,s}^{d} \leq P_B. \quad (15)$$

III. Problem Formulation and Transformation

A. Problem Formulation

Our goal is to maximize the connectivity while satisfying the QoS of different type devices. Hence, the optimization problem is formulated as follows[5]:

$$\begin{aligned} \mathscr{P}1: \quad \underset{\mathcal{P},\mathcal{A},\mathcal{B}}{\text{maximize}} \quad & \xi \textstyle\sum_{k \in \mathcal{C}^A} \sum_{s \in \mathcal{S}} \alpha_{k,s} \\ & + \omega \textstyle\sum_{k \in \mathcal{C}^B} \sum_{s \in \mathcal{S}} \alpha_{k,s} \\ & + (1 - \xi - \omega) \textstyle\sum_{k \in \mathcal{C}^C} \sum_{s \in \mathcal{S}} \beta_{k,s} \\ \text{s.t.} \quad & C1: r_k^u \geq \left(\textstyle\sum_{s \in \mathcal{S}} \alpha_{k,s}\right) \bar{\lambda}_u, \quad \forall k \in \mathcal{C}^{A,B}, \\ & C2: r_k^d \geq \left(\textstyle\sum_{s \in \mathcal{S}} \beta_{k,s}\right) \bar{\lambda}_d, \quad \forall k \in \mathcal{C}^{A,C}, \\ & C3: \alpha_{k,s} \in \{0,1\}, \quad \forall k \in \mathcal{C}^{A,B}, s \in \mathcal{S}, \\ & C4: \beta_{k,s} \in \{0,1\}, \quad \forall k \in \mathcal{C}^{A,C}, s \in \mathcal{S}, \\ & (7)\text{–}(15), \end{aligned}$$

where $\mathcal{P}$, $\mathcal{A}$, $\mathcal{B}$ are the set of variables $\{p_{k,s}^u, p_{k,s}^d\}$, $\{\alpha_{k,s}\}$, $\{\beta_{k,s}\}$, respectively, and $\xi, \omega \in [0,1]$ with $\xi + \omega \leq 1$, are weights to indicate the preference of different slices; $\bar{\lambda}_u$ and $\bar{\lambda}_d$ are the minimum required data rate of devices in the uplink and downlink, respectively.

Note that problem $\mathscr{P}1$ is an MINLP and NP-hard [36]. Therefore, it is necessary to transform problem $\mathscr{P}1$ into a more tractable optimization problem.

B. Transformation

In this subsection, to address the difficulties of problem $\mathscr{P}1$, we first transform the nonlinear constraints to linear constraints. For simplicity, if not specified, all the subscript s in the remaining context are arbitrary, i.e., for any $s \in \mathcal{S}$, due to the frequency-flat channel.

1) Linearize constraint (9)*:* Based on (7) and (8), $\sum_{k \in \mathcal{C}^{A,B}} \alpha_{k,s}$ and $\sum_{k \in \mathcal{C}^{A,C}} \beta_{k,s}$ can only take values in $\{0, 1, 2\}$. Thus, (9) is equivalent to the following constraints, i.e.,

$$\sum_{k \in \mathcal{C}^{A,B}} \alpha_{k,s} + \sum_{k \in \mathcal{C}^{A,C}} \beta_{k,s} \leq 2, \quad (16)$$

and,

$$\begin{cases} \text{Either} & 1 \leq \left|\sum_{k \in \mathcal{C}^{A,B}} \alpha_{k,s} - \sum_{k \in \mathcal{C}^{A,C}} \beta_{k,s}\right|, \\ \text{or} & \sum_{k \in \mathcal{C}^{A,B}} \alpha_{k,s} + \sum_{k \in \mathcal{C}^{A,C}} \beta_{k,s} \leq 0. \end{cases} \quad (17)$$

Removing the absolute value sign, (17) can be further rewritten as

$$\begin{cases} \text{Either} & \sum_{k \in \mathcal{C}^{A,B}} \alpha_{k,s} - \sum_{k \in \mathcal{C}^{A,C}} \beta_{k,s} \leq -1, \\ \text{or} & \sum_{k \in \mathcal{C}^{A,C}} \beta_{k,s} - \sum_{k \in \mathcal{C}^{A,B}} \alpha_{k,s} \leq -1, \\ \text{or} & \sum_{k \in \mathcal{C}^{A,B}} \alpha_{k,s} + \sum_{k \in \mathcal{C}^{A,C}} \beta_{k,s} \leq 0. \end{cases} \quad (18)$$

[5]Since our objective is to obtain a theoretical upper bound on the connectivity and our proposed algorithms (for deterministic problem) can also be extended to the case of imperfect channel state information (CSI) by invoking the Markov inequality, the perfectly estimated CSI obtained at devices as well as BS is considered in this work.

By introducing auxiliary binary variables, $y_{i,s}$ ($i = \{1,2,3\}$, $s \in \mathcal{S}$), Either-Or constraints (18) can be replaced by [37]

$$\begin{cases} \sum_{k\in\mathcal{C}^{A,B}} \alpha_{k,s} - \sum_{k\in\mathcal{C}^{A,C}} \beta_{k,s} \le -1 + M_1(1-y_{1,s}), \\ \sum_{k\in\mathcal{C}^{A,C}} \beta_{k,s} - \sum_{k\in\mathcal{C}^{A,B}} \alpha_{k,s} \le -1 + M_1(1-y_{2,s}), \\ \sum_{k\in\mathcal{C}^{A,B}} \alpha_{k,s} + \sum_{k\in\mathcal{C}^{A,C}} \beta_{k,s} \le M_1(1-y_{3,s}), \\ 1 \le \sum_i^3 y_{i,s}, \text{ and } y_{i,s} \in \{0,1\}, \text{ for } i = 1,2,3, \end{cases} \tag{19}$$

where M_1 is an extremely large positive number. Finally, the nonlinear constraints (9) can be equivalently replaced by the linear constraints (16) and (19).

2) Linearize constraint $C1$: Due to (1)–(6), $C1$ and $C2$ in $\mathscr{P}1$ are both non-convex. For type A and type B devices in the uplink, i.e., $k \in \mathcal{C}^{A,B}$, $C1$ can be replaced by (see Appendix A for the derivation)

$$\alpha_{k,s}\left(I_{k,s} + N_0 W\right)\left(2^{\frac{\bar{\lambda}_u}{W}} - 1\right) \le p^u_{k,s}\|\boldsymbol{h}^u_k\|_2^2, \forall s \in \mathcal{S}, \tag{20}$$

where $I_{k,s}$ is,

$$I_{k,s} = \begin{cases} \sum_{i\in\mathcal{C}^{A,B}_{\mathscr{W}}} \alpha_{i,s} p^u_{i,s}\|\boldsymbol{h}^u_i\|_2^2, & \text{for } k \in \mathcal{C}^{A,B}_{\mathscr{S}}, \\ 0, & \text{for } k \in \mathcal{C}^{A,B}_{\mathscr{W}}. \end{cases}$$

According to (10) and (13), we can remove the $\alpha_{i,s}$ variable from $I_{k,s}$, i.e.,

$$I_{k,s} = \begin{cases} \sum_{i\in\mathcal{C}^{A,B}_{\mathscr{W}}} p^u_{i,s}\|\boldsymbol{h}^u_i\|_2^2, & \text{for } k \in \mathcal{C}^{A,B}_{\mathscr{S}}, \\ 0, & \text{for } k \in \mathcal{C}^{A,B}_{\mathscr{W}}. \end{cases} \tag{21}$$

In order to linearize (20), we introduce a new variable, $\tilde{\alpha}_{k,s} = \alpha_{k,s} I_{k,s}$ ($k \in \mathcal{C}^{A,B}_{\mathscr{S}}, s \in \mathcal{S}$). When $\alpha_{k,s} = 1$, $\tilde{\alpha}_{k,s} = I_{k,s}$, otherwise $\alpha_{k,s} = \tilde{\alpha}_{k,s} = 0$. Then, firstly, for $k \in \mathcal{C}^{A,B}_{\mathscr{S}}$, (20) can be rewritten as

$$\left(\tilde{\alpha}_{k,s} + \alpha_{k,s} N_0 W\right)\left(2^{\frac{\bar{\lambda}_u}{W}} - 1\right) \le p^u_{k,s}\|\boldsymbol{h}^u_k\|_2^2, \forall s \in \mathcal{S}, \tag{22}$$

Using the Big-M method [28], together with (21), $\tilde{\alpha}_{k,s}$ can be further expressed by the follows (for $k \in \mathcal{C}^{A,B}_{\mathscr{S}}$):

$$\begin{aligned} &0 \le \tilde{\alpha}_{k,s} \le \sum_{i\in\mathcal{C}^{A,B}_{\mathscr{W}}} p^u_{i,s}\|\boldsymbol{h}^u_i\|_2^2, \\ &\tilde{\alpha}_{k,s} \le \alpha_{k,s} \sum_{i\in\mathcal{C}^{A,B}_{\mathscr{W}}} P_D\|\boldsymbol{h}^u_i\|_2^2, \text{ and,} \\ &(\alpha_{k,s} - 1)\sum_{i\in\mathcal{C}^{A,B}_{\mathscr{W}}} P_D\|\boldsymbol{h}^u_i\|_2^2 + \sum_{i\in\mathcal{C}^{A,B}_{\mathscr{W}}} p^u_{i,s}\|\boldsymbol{h}^u_i\|_2^2 \le \tilde{\alpha}_{k,s}. \end{aligned} \tag{23}$$

Secondly, for $k \in \mathcal{C}^{A,B}_{\mathscr{W}}$, $I_{k,s} = 0$ (due to the fact that the devices in the *weak* set do not experience the interference in the uplink), (20) can be rewritten as

$$\alpha_{k,s} N_0 W\left(2^{\frac{\bar{\lambda}_u}{W}} - 1\right) \le p^u_{k,s}\|\boldsymbol{h}^u_k\|_2^2, \quad \forall s \in \mathcal{S}. \tag{24}$$

Hence, the non-convex constraint $C1$ can ultimately be equivalently represented by (22)–(24).

3) Linearize constraint $C2$: Similarly, for type A and type C devices in the downlink, i.e., $k \in \mathcal{C}^{A,C}$, $C2$ can be replaced by

$$\beta_{k,s}\left(\bar{I}_{k,s} + N_0 W\right)\left(2^{\frac{\bar{\lambda}_d}{W}} - 1\right) \le p^d_{k,s}\|\boldsymbol{h}^d_k\|_2^2, \forall s \in \mathcal{S}, \tag{25}$$

where $\bar{I}_{k,s}$ is,

$$\bar{I}_{k,s} = \begin{cases} 0, & \text{for } k \in \mathcal{C}^{A,C}_{\mathscr{S}}, \\ \sum_{i\in\mathcal{C}^{A,C}_{\mathscr{S}}} \beta_{i,s} p^d_{i,s}\|\boldsymbol{h}^d_k\|_2^2, & \text{for } k \in \mathcal{C}^{A,C}_{\mathscr{W}}. \end{cases}$$

Based on (10) and (14), $\bar{I}_{k,s}$, can be simplified as

$$\bar{I}_{k,s} = \begin{cases} 0, & \text{for } k \in \mathcal{C}^{A,C}_{\mathscr{S}}, \\ \sum_{i\in\mathcal{C}^{A,C}_{\mathscr{S}}} p^d_{i,s}\|\boldsymbol{h}^d_k\|_2^2, & \text{for } k \in \mathcal{C}^{A,C}_{\mathscr{W}}. \end{cases} \tag{26}$$

To linearize (25), we also introduce a new variable $\tilde{\beta}_{k,s} = \beta_{k,s}\bar{I}_{k,s}$ ($k \in \mathcal{C}^{A,C}_{\mathscr{W}}, s \in \mathcal{S}$). Then, firstly, for $k \in \mathcal{C}^{A,C}_{\mathscr{W}}$, we can readily get (25) transformed as

$$\left(\tilde{\beta}_{k,s} + \beta_{k,s} N_0 W\right)\left(2^{\frac{\bar{\lambda}_d}{W}} - 1\right) \le p^d_{k,s}\|\boldsymbol{h}^d_k\|_2^2, \quad \forall s \in \mathcal{S}. \tag{27}$$

Invoking the Big-M method and (26), $\tilde{\beta}_{k,s}$ is reformulated as (for $k \in \mathcal{C}^{A,C}_{\mathscr{W}}$)

$$\begin{aligned} &0 \le \tilde{\beta}_{k,s} \le \sum_{i\in\mathcal{C}^{A,C}_{\mathscr{S}}} p^d_{i,s}\|\boldsymbol{h}^d_k\|_2^2, \\ &\tilde{\beta}_{k,s} \le \beta_{k,s} \sum_{i\in\mathcal{C}^{A,C}_{\mathscr{S}}} P_B\|\boldsymbol{h}^d_k\|_2^2, \text{ and,} \\ &(\beta_{k,s} - 1)\sum_{i\in\mathcal{C}^{A,C}_{\mathscr{S}}} P_B\|\boldsymbol{h}^d_k\|_2^2 + \sum_{i\in\mathcal{C}^{A,C}_{\mathscr{S}}} p^d_{i,s}\|\boldsymbol{h}^d_k\|_2^2 \le \tilde{\beta}_{k,s}. \end{aligned} \tag{28}$$

Secondly, for $k \in \mathcal{C}^{A,C}_{\mathscr{S}}$, $\bar{I}_{k,s} = 0$ (due to the fact that the devices in the *strong* set do not experience the interference in the downlink), we have,

$$\beta_{k,s} N_0 W\left(2^{\frac{\bar{\lambda}_d}{W}} - 1\right) \le p^d_{k,s}\|\boldsymbol{h}^d_k\|_2^2, \quad \forall s \in \mathcal{S}. \tag{29}$$

Therefore, the non-convex constraint $C2$ can be equivalently represented by (27)–(29).

Since all constraints are characterized into their equivalent linear form, problem $\mathscr{P}1$ is now equivalently transformed as

$$\begin{aligned} \mathscr{P}2: \quad \underset{\mathcal{P},\tilde{\mathcal{A}},\tilde{\mathcal{B}},\mathcal{Y}}{\text{maximize}} \quad & \xi \sum_{k\in\mathcal{C}^A}\sum_{s\in\mathcal{S}} \alpha_{k,s} \\ & + \omega \sum_{k\in\mathcal{C}^B}\sum_{s\in\mathcal{S}} \alpha_{k,s} \\ & + (1-\xi-\omega)\sum_{k\in\mathcal{C}^C}\sum_{s\in\mathcal{S}} \beta_{k,s} \\ \text{s.t.} \quad & (7), (8), (10)\text{–}(16), (19), (22)\text{–}(24), \\ & (27)\text{–}(29), \text{C3 and C4}, \end{aligned}$$

where $\tilde{\mathcal{A}}$, $\tilde{\mathcal{B}}$, $\mathcal{Y}$ are the set of variables $\{\alpha_{k,s}, \tilde{\alpha}_{k,s}\}$, $\{\beta_{k,s}, \tilde{\beta}_{k,s}\}$, and $\{y_{i,s}\}$, respectively. Problem $\mathscr{P}2$ is an MILP.

Although problem $\mathscr{P}2$ is an MILP which can be solved by Generalised Benders Decomposition and Branch-and-Bound [38]. However, these algorithms incur huge computational complexity, especially when the number of variables is large. The acquired results can be only used as the benchmark. In problem $\mathscr{P}2$, we have $|\mathcal{C}|\, S_{tol}$ binary variables, $|\mathcal{C}|\, S_{tol}$ continuous variables and $\max\{\left|\mathcal{C}^{A,B}\right|, \left|\mathcal{C}^{A,C}\right|\} S_{tol}$ constraints. The

overall computational complexity for problem $\mathscr{P}2$ is [39]

$$\mathcal{O}\left(2^{(|\mathcal{C}|S_{tol})}\cdot\left(|\mathcal{C}|\,S_{tol}\right)^2\cdot\max\left\{\left|\mathcal{C}^{A,B}\right|,\left|\mathcal{C}^{A,C}\right|\right\}S_{tol}\right).$$

Therefore, we propose efficient algorithms to tackle problem $\mathscr{P}2$ in the following section.

IV. Proposed Algorithms

In this section, we first simplify MILP $\mathscr{P}2$ by pre-allocating transmit power in the uplink and downlink, respectively, and then present low-complexity algorithms to solve the simplified $\mathscr{P}2$.

A. Given Power Scheme

Since our objective is to maximize the connectivity, and the transmit power from both BS and MTCDs is not our main concern, we can derive a straightforward approach to find out the optimal $p^u_{k,s}$ and sub-optimal $p^d_{k,s}$.

From the perspective of QoS, when a device is accessed, the minimum transmit power $p^u_{k,s}$ and $p^d_{k,s}$ are defined as

$$\begin{cases} p^{u(min)}_{k,s} \triangleq \frac{N_0W(2^{\frac{\bar{\lambda}_u}{W}}-1)}{\|\boldsymbol{h}^u_k\|_2^2}, & \text{for } k\in\mathcal{C}^{A,B},\\ p^{d(min)}_{k,s} \triangleq \frac{N_0W(2^{\frac{\bar{\lambda}_d}{W}}-1)}{\|\boldsymbol{h}^d_k\|_2^2}, & \text{for } k\in\mathcal{C}^{A,C}, \end{cases}$$

respectively. Apparently, when a device is accessed, we should have,

$$\begin{aligned} &p^{u(min)}_{k,s}\le P_D, \quad \forall k\in\mathcal{C}^{A,B}, \text{ and,}\\ &p^{d(min)}_{k,s}\le P_B, \quad \forall k\in\mathcal{C}^{A,C}, \end{aligned} \tag{30}$$

Lemma 1: *The optimal transmit power, $p^{u*}_{k,s}$, for type A and type B devices in the uplink is*

$$\begin{cases} p^{u*}_{k,s}=P_D, & \text{for } k\in\mathcal{C}^{A,B}_{\mathscr{S}},\\ p^{u*}_{k,s}=p^{u(min)}_{k,s}, & \text{for } k\in\mathcal{C}^{A,B}_{\mathscr{W}}. \end{cases} \tag{31}$$

Proof: Based on (20), when a device in the *strong* set is accessed, the interference, $I_{k,s}$, that it can tolerate is

$$I_{k,s}=\frac{p^u_{k,s}\|\boldsymbol{h}^u_k\|_2^2}{2^{\frac{\bar{\lambda}_u}{W}}-1}-N_0W, \quad \forall k\in\mathcal{C}^{A,B}_{\mathscr{S}}, \tag{32}$$

and, with (21), when a device in the *weak* set is accessed, the interference that it introduces to a device in the *strong* set is

$$p^u_{k,s}\|\boldsymbol{h}^u_k\|_2^2, \quad \forall k\in\mathcal{C}^{A,B}_{\mathscr{W}}. \tag{33}$$

For any given sub-carrier $s\in\mathcal{S}$, if two devices are accessed over the same sub-carrier at the same time, then we have

$$\frac{p^u_{i,s}\|\boldsymbol{h}^u_i\|_2^2}{2^{\frac{\bar{\lambda}_u}{W}}-1}-N_0W\ge p^u_{j,s}\|\boldsymbol{h}^u_j\|_2^2 \quad \forall i\in\mathcal{C}^{A,B}_{\mathscr{S}}, j\in\mathcal{C}^{A,B}_{\mathscr{W}}. \tag{34}$$

To ensure that inequality (34) holds, a device in the *strong* set should allocate as much power as possible to accommodate interference, while the device in the *weak* set should allocate as less power as possible to reduce the interference it causes. Adopting the power allocation scheme in (31), the left hand side of (34) can obtain its maximum value and the right hand side of (34) achieves the minimum value. The Lemma is proven. □

Lemma 2: *The optimal transmit power, $p^{d*}_{k,s}$, for devices from* strong *set in the downlink is*

$$p^{d*}_{k,s}=p^{d(min)}_{k,s}, \quad \text{for } k\in\mathcal{C}^{A,C}_{\mathscr{S}}. \tag{35}$$

Proof: Based on (26), when a device in the *strong* set is accessed, the interference that it can cause to a device in the *weak* set is

$$p^d_{k,s}\|\boldsymbol{h}^d_k\|_2^2, \quad \forall k\in\mathcal{C}^{A,B}_{\mathscr{S}}. \tag{36}$$

Thus, the device from the *strong* set can cause the minimum interference by using $p^d_{k,s}=p^{d(min)}_{k,s}$. □

However, there is no optimal power pre-allocation strategy for devices in the *weak* set since the total downlink transmit power is limited at the BS. In reality, the power allocation in downlink NOMA is a challenging issue, which is determined by users' channel conditions, availability of CSI, QoS requirements, total power constraint, and system objective [40]. To make problem $\mathscr{P}2$ tractable, we adopt the fractional transmit power control (FTPC) [40], according to its channel gain powered with a decaying factor, as the sub-optimal power allocation (denoted as $p^{d+}_{k,s}$) for devices in the *weak* set,

$$p^{d+}_{k,s}=\frac{\rho P_B}{\sum_{i\in\mathcal{C}^{A,C}_{\mathscr{W}}}\left(\left\|\boldsymbol{h}^d_k\right\|_2^2\right)^{-\tau}}\left(\left\|\boldsymbol{h}^d_k\right\|_2^2\right)^{-\tau}, \quad k\in\mathcal{C}^{A,C}_{\mathscr{W}}, \tag{37}$$

where τ ($-1\le\tau\le1$) and ρ ($1\le\rho$) are the decay factor and gain factor, respectively. The case of $\tau=0$ corresponds to equal transmit power allocation among the devices. The more τ is increased, the more power is allocated to the device with lower channel gain.

Obviously, when a device is accessed, we should have[6],

$$p^{d(min)}_{k,s}\le p^{d+}_{k,s}, \quad \forall k\in\mathcal{C}^{A,C}_{\mathscr{W}}. \tag{38}$$

B. Pairing Guideline

Exploiting (31), (35) and (37) power allocation schemes, we are ready to define $\Lambda_{k,s}$, $\Omega_{k,s}$ as the maximum (or fixed) interference that a device can tolerate and the minimum interference that a device can cause, respectively. Based on (20), (21), (25), and (26), when a device is accessed, we have

$$\Lambda_{k,s}=\begin{cases}\frac{p^{u*}_{k,s}\|\boldsymbol{h}^u_k\|_2^2}{2^{\frac{\bar{\lambda}_u}{W}}-1}-N_0W, & \forall k\in\mathcal{C}^{A,B}_{\mathscr{S}},\\ \frac{p^{d+}_{k,s}\|\boldsymbol{h}^d_k\|_2^2}{2^{\frac{\bar{\lambda}_d}{W}}-1}-N_0W, & \forall k\in\mathcal{C}^{A,C}_{\mathscr{W}},\end{cases}$$

and,

$$\Omega_{k,s}=\begin{cases}p^{u*}_{k,s}\|\boldsymbol{h}^u_k\|_2^2, & \forall k\in\mathcal{C}^{A,B}_{\mathscr{W}},\\ p^{d*}_{k,s}\sum_{i\in\mathcal{C}^{A,C}_{\mathscr{W}}}\beta_{i,s}\|\boldsymbol{h}^d_i\|_2^2, & \forall k\in\mathcal{C}^{A,C}_{\mathscr{S}}.\end{cases}$$

Lemma 3: *For any given sub-carrier $s\in\mathcal{S}$, if two devices (from strong and weak set respectively) are accessed over the*

[6]Since $p^{d(min)}_{k,s}>p^{d+}_{k,s}$ can also happen, we should introduce a new constraint to guarantee that all accessed devices satisfy the inequality (38). Specifically, the new constraint can be $\beta_{k,s}p^{d(min)}_{k,s}\le p^{d+}_{k,s}$, where $\beta_{k,s}=1$ if the inequality (38) is satisfied, and $\beta_{k,s}=0$ otherwise.

same sub-carrier at the same time, then $\Lambda_{i,s} \geq \Omega_{j,s}$ must be satisfied, for $i \in \mathcal{C}_{\mathscr{S}}^{A,B}$ and $j \in \mathcal{C}_{\mathscr{W}}^{A,B}$ in the uplink or $i \in \mathcal{C}_{\mathscr{W}}^{A,C}$ and $j \in \mathcal{C}_{\mathscr{S}}^{A,C}$ in the downlink.

Proof: In the uplink, based on (20) and (21), accessing two devices i and j over the s-th sub-carrier yields ($i \in \mathcal{C}_{\mathscr{S}}^{A,B}$ and $j \in \mathcal{C}_{\mathscr{W}}^{A,B}$)

$$\left(p_{j,s}^{u}\|\boldsymbol{h}_{j}^{u}\|_2^2 + N_0 W\right)\left(2^{\frac{\bar{\lambda}_u}{W}} - 1\right) \leq p_{i,s}^{u}\|\boldsymbol{h}_{i}^{u}\|_2^2. \tag{39}$$

According to **Lemma 1**, (39) produces

$$p_{j,s}^{u*}\|\boldsymbol{h}_{j}^{u}\|_2^2 \leq \frac{p_{i,s}^{u*}\|\boldsymbol{h}_{i}^{u}\|_2^2}{2^{\frac{\bar{\lambda}_u}{W}} - 1} - N_0 W \Rightarrow \Omega_{j,s} \leq \Lambda_{i,s}.$$

In the downlink, with (25) and (26), assume that the i- and j-th, $i \in \mathcal{C}_{\mathscr{W}}^{A,C}$, $j \in \mathcal{C}_{\mathscr{S}}^{A,C}$, devices are accessed, we have,

$$\left(p_{j,s}^{d}\|\boldsymbol{h}_{i}^{d}\|_2^2 + N_0 W\right)\left(2^{\frac{\bar{\lambda}_d}{W}} - 1\right) \leq p_{i,s}^{d}\|\boldsymbol{h}_{i}^{d}\|_2^2. \tag{40}$$

According to **Lemma 2**, (40) can obtain

$$p_{j,s}^{d*}\|\boldsymbol{h}_{i}^{d}\|_2^2 \leq \frac{p_{i,s}^{d+}\|\boldsymbol{h}_{i}^{d}\|_2^2}{2^{\frac{\bar{\lambda}_d}{W}} - 1} - N_0 W \Rightarrow \Omega_{j,s} \leq \Lambda_{i,s},$$

where $\sum_{i \in \mathcal{C}_{\mathscr{W}}^{A,C}} \beta_{i,s}\|\boldsymbol{h}_{i}^{d}\|_2^2 = \|\boldsymbol{h}_{i}^{d}\|_2^2$, if the i-th device get accessed. □

On the basis of **Lemma 3**, we can get the basic pairing guideline for NOMA. For any given sub-carrier $s \in \mathcal{S}$, there are three different access situations: (a) Two devices from *strong* and *weak* set respectively are accessed at the same time; (b) Only one device from *strong* or *weak* set is accessed; (c) No device get accessed.

In the uplink, all the three access situations can be concluded by

$$\sum_{k \in \mathcal{C}_{\mathscr{W}}^{A,B}} \alpha_{k,s}\Omega_{k,s} \leq \sum_{k \in \mathcal{C}_{\mathscr{S}}^{A,B}} \alpha_{k,s}\Lambda_{k,s} + M_2\left(1 - \sum_{k \in \mathcal{C}_{\mathscr{S}}^{A,B}} \alpha_{k,s}\right), \tag{41}$$

where $M_2 \geq \max_{k \in \mathcal{C}_{\mathscr{W}}^{A,B}}\{\Omega_{k,s}\}$ is a large positive number. In detail, when sub-carrier s is not allocated to any of the *strong* set devices, i.e., $\sum_{k \in \mathcal{C}_{\mathscr{S}}^{A,B}} \alpha_{k,s} = 0$, there will be no upper limit on the interference that can be caused by *weak* set devices, which implies $M_2 \geq \max_{k \in \mathcal{C}_{\mathscr{W}}^{A,B}}\{\Omega_{k,s}\}$.

In the downlink, using **Lemma 3**, we can readily get

$$\sum_{k \in \mathcal{C}_{\mathscr{S}}^{A,C}} \beta_{k,s}\Omega_{k,s} \leq \sum_{k \in \mathcal{C}_{\mathscr{W}}^{A,C}} \beta_{k,s}\Lambda_{k,s} + M_3\left(1 - \sum_{k \in \mathcal{C}_{\mathscr{W}}^{A,C}} \beta_{k,s}\right),$$

in which $\Omega_{k,s}$ can be written as

$$\Omega_{k,s} = p_{k,s}^{d*}\sum_{i \in \mathcal{C}_{\mathscr{W}}^{A,C}} \beta_{i,s}\|\boldsymbol{h}_{i}^{d}\|_2^2 \leq p_{k,s}^{d*}\max\left\{\|\boldsymbol{h}_{i}^{d}\|_2^2\right\}_{i \in \mathcal{C}_{\mathscr{W}}^{A,C}}.$$

Therefore, all the three access situations in the downlink can be given by

$$\sum_{k \in \mathcal{C}_{\mathscr{S}}^{A,C}} \beta_{k,s}p_{k,s}^{d*}\Delta \leq \sum_{k \in \mathcal{C}_{\mathscr{W}}^{A,C}} \beta_{k,s}\Lambda_{k,s} + M_3\left(1 - \sum_{k \in \mathcal{C}_{\mathscr{W}}^{A,C}} \beta_{k,s}\right), \tag{42}$$

where $\Delta = \max\left\{\|\boldsymbol{h}_{i}^{d}\|_2^2\right\}_{i \in \mathcal{C}_{\mathscr{W}}^{A,C}}$; M_3 is a large positive number satisfying $M_3 \geq \max_{k \in \mathcal{C}_{\mathscr{S}}^{A,C}}\{p_{k,s}^{d*}\Delta\}$.

With (31)–(42), the variable $\mathcal{P}$ in problem $\mathscr{P}2$ is removed, and problem $\mathscr{P}2$ is now simplified as

$$\begin{aligned}
\mathscr{P}3: \quad & \underset{\mathcal{A},\mathcal{B},\mathcal{Y}}{\text{maximize}} \quad \xi\sum_{k \in \mathcal{C}^A}\sum_{s \in \mathcal{S}} \alpha_{k,s} \\
& \qquad + \omega\sum_{k \in \mathcal{C}^B}\sum_{s \in \mathcal{S}} \alpha_{k,s} \\
& \qquad + (1 - \xi - \omega)\sum_{k \in \mathcal{C}^C}\sum_{s \in \mathcal{S}} \beta_{k,s} \\
\text{s.t.} \quad & C5: \alpha_{k,s}p_{k,s}^{u(min)} \leq P_D, \quad \forall k \in \mathcal{C}^{A,B}, s \in \mathcal{S}, \\
& C6: \beta_{k,s}p_{k,s}^{d(min)} \leq P_B, \quad \forall k \in \mathcal{C}^{A,C}, s \in \mathcal{S}, \\
& C7: \beta_{k,s}p_{k,s}^{d(min)} \leq p_{k,s}^{d+}, \quad \forall k \in \mathcal{C}_{\mathscr{W}}^{A,C}, s \in \mathcal{S}, \\
& C8: \sum_{s \in \mathcal{S}}\left(\sum_{k \in \mathcal{C}_{\mathscr{S}}^{A,C}} \beta_{k,s}p_{k,s}^{d*} + \right. \\
& \qquad \left. \sum_{k \in \mathcal{C}_{\mathscr{W}}^{A,C}} \beta_{k,s}p_{k,s}^{d+}\right) \leq P_B, \\
& (7), (8), (10)\text{–}(12), (16), (19), (41), (42), \\
& \text{C3 and C4},
\end{aligned}$$

where $C5$, $C6$ and $C7$ force constraints (30) and (38) to be satisfied when a device is accessed; $C8$ indicates that the total power allocated in the downlink using FTPC method cannot exceed the total transmit power of the BS. Problem $\mathscr{P}3$ is an integer linear programming (ILP). Note that problem $\mathscr{P}3$ produces the sub-optimal solution for problem $\mathscr{P}2$.

From problem $\mathscr{P}2$ to problem $\mathscr{P}3$, we reduce the MILP to ILP, whose computational complexity becomes $\mathcal{O}\left(2^a ab\right)$, where a and b are the number of binary variables and constraints, respectively [39]. In problem $\mathscr{P}3$, $a = |\mathcal{C}|\, S_{tol}$ and $b = \max\{\left|\mathcal{C}^{A,B}\right|, \left|\mathcal{C}^{A,C}\right|\}S_{tol}$. However, the computational complexity is still exponential, we next propose efficient algorithms to solve problem $\mathscr{P}3$.

C. Alternating Selection Best-effort Pairing

In this subsection, we present the alternating selection best-effort pairing algorithm (AS-BEP) to solve problem $\mathscr{P}3$. It is a greedy heuristic scheme with low-complexity. Prior to introducing AS-BEP, we first present the best-effort pairing (BEP) algorithm.

To maximize the connectivity, in some ways, is to maximize the number of NOMA pairs. That is, if we want to access as many devices as possible, the more pairs, the better. Based on (41) and (42), if we can find two devices satisfying $\Lambda_{k,s} \geq \Omega_{k,s}$, it realizes a pair. To be more specific, our best-effort pairing algorithm has two principles:

- A device having large $\Lambda_{k,s}$ should also accommodate a device having large $\Omega_{k,s}$.
- Devices with smaller downlink transmit power should be paired and accessed with higher priority.

1) Best-effort pairing for uplink: Define ordered sets $\mathcal{U}_u$ and $\mathcal{V}_u$ as (we drop the subscript s here since we consider the flat fading channel)

$$\begin{aligned}
\mathcal{U}_u &= \left\{k \mid \frac{\Lambda_k}{p_k} \geq \frac{\Lambda_{k+1}}{p_{k+1}}\right\}, k \in \mathcal{C}_{\mathscr{S}}^{A,B}, \text{ and,} \\
\mathcal{V}_u &= \left\{k \mid \frac{p_k}{\Omega_k} \leq \frac{p_{k+1}}{\Omega_{k+1}}\right\}, k \in \mathcal{C}_{\mathscr{W}}^{A,B},
\end{aligned} \tag{43}$$

where the devices in $\mathcal{U}_u$ are sorted in descending order in terms of $\frac{\Lambda_k}{p_k}$ and in $\mathcal{V}_u$ are sorted in ascending order in terms

of $\frac{p_k}{\Omega_k}$; $p_k = p_k^{d*}$ for $k \in \mathcal{C}_{\mathscr{S}}^A$ and $p_k = p_k^{d+}$ for $k \in \mathcal{C}_{\mathscr{W}}^A$. Since type B devices do not require downlink transmission, to make (43) defined, we set a virtual p_k for $k \in \mathcal{C}^B$. To reasonably set p_k, there are two facts to be considered:

- When ω changes from 0 to 1, the preference of type B devices is increasing. From aforementioned second principle, we should set the virtual p_k as an decreasing function of ω.
- Since type A and type B devices share the power at the BS side, when designing the virtual p_k for $k \in \mathcal{C}^B$, p_k for $k \in \mathcal{C}^A$ should be used as the references.

On these bases, we design a virtual value $p_k = \Gamma$, for $k \in \mathcal{C}^B$, as follows:

$$\Gamma = \begin{cases} \min\{p_i^{d*}\}_{i\in\mathcal{C}_{\mathscr{S}}^A} \left(\frac{\max\{p_i^{d*}\}_{i\in\mathcal{C}_{\mathscr{S}}^A}}{\min\{p_i^{d*}\}_{i\in\mathcal{C}_{\mathscr{S}}^A}}\right)^{1-\omega}, & \text{for } k \text{ in } \mathcal{C}_{\mathscr{S}}^B, \\ \min\{p_j^{d+}\}_{j\in\mathcal{C}_{\mathscr{W}}^A} \left(\frac{\max\{p_j^{d+}\}_{j\in\mathcal{C}_{\mathscr{W}}^A}}{\min\{p_j^{d+}\}_{j\in\mathcal{C}_{\mathscr{W}}^A}}\right)^{1-\omega}, & \text{for } k \text{ in } \mathcal{C}_{\mathscr{W}}^B. \end{cases}$$

2) Best-effort pairing for downlink: We similarly define ordered sets, $\mathcal{U}_d$ and $\mathcal{V}_d$ as

$$\begin{aligned} \mathcal{U}_d &= \left\{k \mid \frac{\Lambda_k}{p_k} \geq \frac{\Lambda_{k+1}}{p_{k+1}}\right\}, \quad k \in \mathcal{C}_{\mathscr{W}}^{A,C}, \text{ and,} \\ \mathcal{V}_d &= \left\{k \mid \frac{p_k}{\Omega_k} \leq \frac{p_{k+1}}{\Omega_{k+1}}\right\}, \quad k \in \mathcal{C}_{\mathscr{S}}^{A,C}, \end{aligned} \tag{44}$$

where $p_k = p_k^{d*}$ for $k \in \mathcal{C}_{\mathscr{S}}^A$ and $p_k = p_k^{d+}$ for $k \in \mathcal{C}_{\mathscr{W}}^A$. Due to the existence of the preference factor ξ and ω, we need to define a virtual p_k for $k \in \mathcal{C}^C$. Besides the above facts, we also take the original downlink transmit power of type C devices as the references and normalize it to get virtual p_k as follows:

$$p_k = \begin{cases} \frac{\Theta p_k^{d*}}{\max\{p_i^{d*}\}_{i\in\mathcal{C}_{\mathscr{S}}^C}}, & k \in \mathcal{C}_{\mathscr{S}}^C, \\ \frac{\Theta p_k^{d+}}{\max\{p_i^{d+}\}_{i\in\mathcal{C}_{\mathscr{W}}^C}}, & k \in \mathcal{C}_{\mathscr{W}}^C, \end{cases}$$

where

$$\Theta = \begin{cases} \min\{p_i^{d*}\}_{i\in\mathcal{C}_{\mathscr{S}}^A} \left(\frac{\max\{p_i^{d*}\}_{i\in\mathcal{C}_{\mathscr{S}}^A}}{\min\{p_i^{d*}\}_{i\in\mathcal{C}_{\mathscr{S}}^A}}\right)^{\xi+\omega}, & \text{for } k \text{ in } \mathcal{C}_{\mathscr{S}}^C, \\ \min\{p_j^{d+}\}_{j\in\mathcal{C}_{\mathscr{W}}^A} \left(\frac{\max\{p_j^{d+}\}_{j\in\mathcal{C}_{\mathscr{W}}^A}}{\min\{p_j^{d+}\}_{j\in\mathcal{C}_{\mathscr{W}}^A}}\right)^{\xi+\omega}, & \text{for } k \text{ in } \mathcal{C}_{\mathscr{W}}^C. \end{cases}$$

Now, the best-effort pairing algorithm for uplink/downlink is ready to be devised. For example, a device with the largest Λ_k and smallest p_k (hence largest Λ_k/p_k) and a device with the smallest p_k and largest Ω_k (hence smallest p_k/Ω_k) can be paired first. We define paired matrices, $\mathbf{Q}_u \in \mathbb{N}^{2\times l_1}$ and $\mathbf{Q}_d \in \mathbb{N}^{2\times l_2}$ as the output of **Algorithm 1**, where the entries in the first and second rows of $\mathbf{Q}_u$ are constructed by the device numbers from $\mathcal{C}_{\mathscr{S}}^{A,B}$ and $\mathcal{C}_{\mathscr{W}}^{A,B}$, respectively. Similarly, the entries in the first and second rows of $\mathbf{Q}_d$ are constructed by the device numbers from $\mathcal{C}_{\mathscr{S}}^{A,C}$ and $\mathcal{C}_{\mathscr{W}}^{A,C}$, respectively; l_1 and l_2 are dynamic variables which represent the number of occupied sub-carriers for uplink and downlink, respectively. We use $col(\mathbf{Q}_u)$ and $col(\mathbf{Q}_d)$ to indicate the number of columns of the matrix $\mathbf{Q}_u$ and $\mathbf{Q}_d$, respectively. We present the best-effort pairing algorithm for uplink/downlink in **Algorithm 1**, which has time complexity

$$\mathcal{O}\left(\max\left\{\left|\mathcal{C}_{\mathscr{S}}^{A,B}\right|\left|\mathcal{C}_{\mathscr{W}}^{A,B}\right|, \left|\mathcal{C}_{\mathscr{S}}^{A,C}\right|\left|\mathcal{C}_{\mathscr{W}}^{A,C}\right|\right\}\right).$$

Algorithm 1 BEP Algorithm for Uplink/Downlink

Input: P_B, P_D, $\mathcal{U}_u$ ($\mathcal{U}_d$), $\mathcal{V}_u$ ($\mathcal{V}_d$)
Output: Paired matrix, $\mathbf{Q}_u$, $\mathbf{Q}_d$

Initialization: $p \leftarrow 0$, $\mathcal{Q}_u = \mathcal{Q}_d \leftarrow \emptyset$
Eliminate devices that does not meet (30) and (38)
for $i \in \mathcal{U}_u$ ($\mathcal{U}_d$) **do**
  for $j \in \mathcal{V}_u$ ($\mathcal{V}_d$) **do**
    if $\Lambda_i \geq \Omega_j$ **then**
      $p \leftarrow p + p_i^{d*} + p_j^{d*}$ (p_k^{d+} for $k \in \mathcal{C}_{\mathscr{W}}^{A,C}$)
      if $p \leq P_B$ **then**
        $\mathcal{V}_u \leftarrow \mathcal{V}_u \setminus \{j\}$ ($\mathcal{V}_d \leftarrow \mathcal{V}_d \setminus \{j\}$)
        $\mathbf{Q}_u \leftarrow \begin{bmatrix} \mathbf{Q}_u & \begin{bmatrix} i \\ j \end{bmatrix} \end{bmatrix}$ $\mathbf{Q}_d \leftarrow \begin{bmatrix} \mathbf{Q}_d & \begin{bmatrix} i \\ j \end{bmatrix} \end{bmatrix}$
        if $col(\mathbf{Q}_u) \geq S_{tol}$ ($col(\mathbf{Q}_d) \geq S_{tol}$) **then**
          termination
        end if
        break
      else
        $p \leftarrow p - p_i^{d*} - p_j^{d*}$ (p_k^{d+} for $k \in \mathcal{C}_{\mathscr{W}}^{A,C}$)
      end if
    end if
  end for
end for

Remark 3: It is worth mentioning that, **Algorithm 1** results in a mixed NOMA and OMA system. For example, for the uplink BEP case, if we have a type A device accessed in the uplink with NOMA, then the downlink of this device should also be accessed, while with OMA[7].

As we can learn from **Algorithm 1** and Remark 3, BEP algorithm can be used for non-orthogonally accessing devices over either one of uplink or downlink, i.e., the other link should be orthogonally accessed. However, in fact, both uplink and downlink transmission can exist simultaneously in our system, and both of them require NOMA to maximize the connectivity. In order to make the BEP algorithm still effective in this scenario, we next propose the AS-BEP algorithm. We highlight the main ideas of the AS-BEP algorithm as follows:

- Firstly, regarding uplink and downlink as two separate processes, i.e., construct their own paired matrices, $\mathbf{Q}_u$ and $\mathbf{Q}_d$ by **Algorithm 1**, independently.
- Between the obtained $\mathbf{Q}_u$ and $\mathbf{Q}_d$, alternatively selecting (which is regarded as alternating selection) their columns that satisfy the criterion[8] to guarantee the consistent service for different types of devices.

Denote the final paired matrix as $\mathbf{Q}_f \in \mathbb{N}^{2\times l}$ as the output of AS-BEP, where the entries of the first and second rows of

[7]In **Algorithm 1**, we omitted the description of OMA downlink/uplink to keep this algorithm more consistent with **Algorithm 2**. For the BEP algorithm that considers OMA downlink/uplink, it can be found in our conference version [1].

[8]Since an accessed type A device should occupy both uplink and downlink sub-carriers, if we find a type A device only appears in one column of $\mathbf{Q}_u$ or $\mathbf{Q}_d$, then this type A device should not be accessed.

Algorithm 2 AS-BEP Algorithm for $\mathscr{P}3$

Input: P_D, P_B, W, S_{tol}, $\bar{\lambda}_u$, $\bar{\lambda}_d$, $\boldsymbol{h}_k^u$, $\boldsymbol{h}_k^d$, $\forall k \in \mathcal{C}$
Output: Final Paired matrix, $\mathbf{Q}_f$

Initialization: $\mathcal{Q}_u = \mathcal{Q}_d = \mathcal{Q}_f \leftarrow \emptyset$
Calculate $\mathcal{U}_u$ ($\mathcal{U}_d$) and $\mathcal{V}_u$ ($\mathcal{V}_d$) by (43) and (44)
Calculate paired matrices $\mathbf{Q}_u$, $\mathbf{Q}_d$ by using **Algorithm** 1
while $col(\mathbf{Q}_f) < S_{tol}$ **and** $pow(\mathbf{Q}_f) \leq P_B$ **do**
  if $\mathcal{Q}_u(1) \cap \mathcal{C}^A \neq \emptyset$ **then**
    if $\mathcal{Q}_u(1) \cap \mathcal{Q}_d(m) \neq \emptyset, m = 1, \cdots, col(\mathbf{Q}_d)$ **then**
      $\mathcal{Q}_u \leftarrow \mathcal{Q}_u \backslash \mathcal{Q}_u(1)$, $\mathcal{Q}_d \leftarrow \mathcal{Q}_d \backslash \mathcal{Q}_d(m)$
      $\mathbf{Q}_f \leftarrow [\mathbf{Q}_f \ \mathbf{Q}_u(1) \ \mathbf{Q}_d(m)]$
    else
      $\mathcal{Q}_u \leftarrow \mathcal{Q}_u \backslash \mathcal{Q}_u(1)$
    end if
  else
    $\mathcal{Q}_u \leftarrow \mathcal{Q}_u \backslash \mathcal{Q}_u(1)$, $\mathbf{Q}_f \leftarrow [\mathbf{Q}_f \ \mathbf{Q}_u(1)]$
  end if
  if $col(\mathbf{Q}_f) \geq S_{tol}$ **or** $pow(\mathbf{Q}_f) > P_B$ **then**
    break
  end if
  if $\mathcal{Q}_d(1) \cap \mathcal{C}^A \neq \emptyset$ **then**
    if $\mathcal{Q}_d(1) \cap \mathcal{Q}_u(n) \neq \emptyset, n = 1, \cdots, col(\mathbf{Q}_u)$ **then**
      $\mathcal{Q}_d \leftarrow \mathcal{Q}_d \backslash \mathcal{Q}_d(1)$, $\mathcal{Q}_u \leftarrow \mathcal{Q}_u \backslash \mathcal{Q}_u(n)$
      $\mathbf{Q}_f \leftarrow [\mathbf{Q}_f \ \mathbf{Q}_d(1) \ \mathbf{Q}_u(n)]$
    else
      $\mathcal{Q}_d \leftarrow \mathcal{Q}_d \backslash \mathcal{Q}_d(1)$
    end if
  else
    $\mathcal{Q}_d \leftarrow \mathcal{Q}_d \backslash \mathcal{Q}_d(1)$, $\mathbf{Q}_f \leftarrow [\mathbf{Q}_f \ \mathbf{Q}_d(1)]$
  end if
end while

$\mathbf{Q}_f$ are constructed by the device numbers from *strong* and *weak* set respectively; l is a dynamic variable and represents total the number of occupied sub-carriers. We use $col(\mathbf{Q}_f)$ to indicate the number of columns of the matrix $\mathbf{Q}_f$ and denote $\mathcal{Q}_f$ as the set of all entries in matrix $\mathbf{Q}_f$. Also, we denote $pow(\mathbf{Q}_f)$ as the total downlink transmit power consumed by the accessed devices in the matrix $\mathbf{Q}_f$, i.e., $pow(\mathbf{Q}_f) = \sum_{k \in \mathcal{K}_1} p_k^{d*} + \sum_{k \in \mathcal{K}_2} p_k^{d+}$, where $\mathcal{K}_1 = \mathcal{Q}_f \cap \mathcal{C}_{\mathscr{S}}^{A,C}$ and $\mathcal{K}_2 = \mathcal{Q}_f \cap \mathcal{C}_{\mathscr{W}}^{A,C}$. $\mathbf{Q}(n)$ represents the n-th column of the matrix $\mathbf{Q}$ and $\mathcal{Q}(n)$ denotes the set constructed by the entries of $\mathbf{Q}(n)$.

Now the AS-BEP algorithm is ready to be devised. For example, we first select $\mathbf{Q}_u(1)$, and judge whether it has a type A device. If no, we just put into $\mathbf{Q}_u(1)$ into $\mathbf{Q}_f$. If yes, we need check more: If $\mathcal{Q}_u(1) \cap \mathcal{Q}_d(m) \neq \emptyset$, for $m = 1, \cdots, col(\mathbf{Q}_d)$, then select both $\mathbf{Q}_u(1)$ and $\mathbf{Q}_d(m)$ and put the two columns into $\mathbf{Q}_f$, otherwise $\mathbf{Q}_u(1)$ is not selected. The similar procedure is then implemented for $\mathbf{Q}_d$ at the next turn. By alternately selecting devices between $\mathbf{Q}_u$ and $\mathbf{Q}_d$, we obtain the final paired matrix $\mathbf{Q}_f$. The proposed AS-BEP algorithm is presented in **Algorithm 2**, which has time complexity $\mathcal{O}\left((S_{tol}+1) \cdot \max\left\{\left|\mathcal{C}_{\mathscr{S}}^{A,B}\right|\left|\mathcal{C}_{\mathscr{W}}^{A,B}\right|, \left|\mathcal{C}_{\mathscr{S}}^{A,C}\right|\left|\mathcal{C}_{\mathscr{W}}^{A,C}\right|\right\}\right)$.

In summary, the key idea of the AS-BEP algorithm is, instead of seeking the maximum number of accessed devices directly, we choose to achieve as many accessed NOMA pairs as possible. In addition, for a type A device, its uplink or downlink should be either both accessed or both not accessed. The proposed AS-BEP algorithm for problem $\mathscr{P}3$ obtains the number of accessed NOMA pairs, which can eventually be translated into the number of accessed devices. The logic flow to solve problem $\mathscr{P}1$ is shown in Fig. 2.

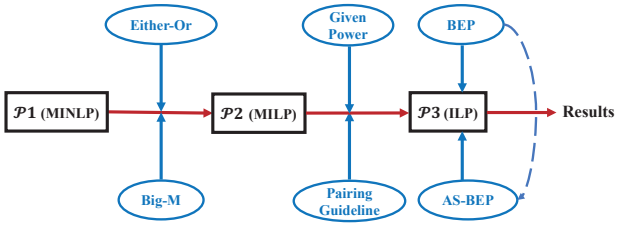


Fig. 2. The logic flow to solve problem $\mathscr{P}1$.

TABLE I
SIMULATION PARAMETERS

Parameters	Values
D	[0, 500] m
M	4
N_0	-174 dBm/Hz
W	6 kHz
S_{tol}	30
$PL(D)$	120.9 + 37.6log10 (D [km]) + L + G
L & G	20 dB & -4 dB
P_D	23 dBm
P_B	30 dBm

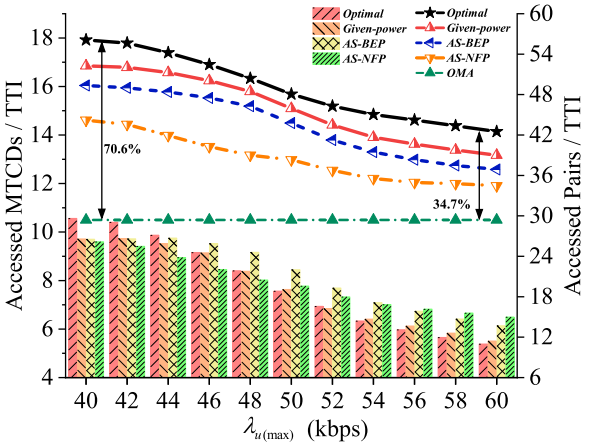
V. SIMULATION RESULTS

In this section, we evaluate the system performance of the proposed algorithms. We consider a single cell where all devices are randomly distributed within the radius of 500 m. The frequency of operation is taken to be 900 MHz. Flat Rayleigh fading channels are considered since the total system bandwidth is as narrow as 180 kHz. The distance-dependent path loss, $PL(D)$ is modeled as [41]

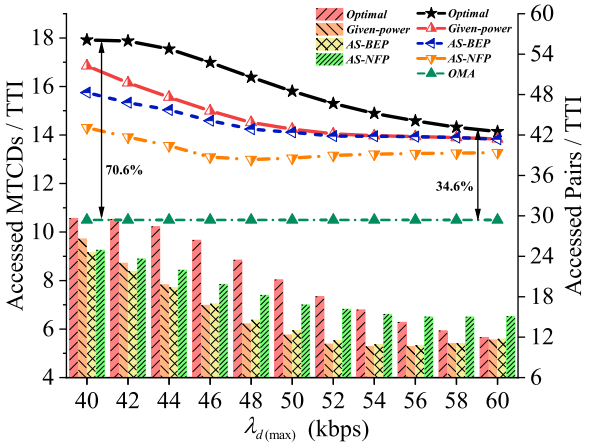
$$PL(D) = 120.9 + 37.6 \log\left(\frac{D}{1000}\right) + L + G,$$

where D is the distance between MTCDs and the BS in meter, and L is the indoor penetration loss that is assumed to be 20 dB. G is the antenna gain of -4 dB. The power spectral density of the noise is -174 dBm/Hz and noise figure is 5 dB. The maximum transmit power for each device, P_D, is set to 23 dBm and the maximum transmit power of BS, P_B, is set to 30 dBm. All results are simulated with 100 independent channel realizations. The main simulation parameters used in this paper are summarized in Table I.

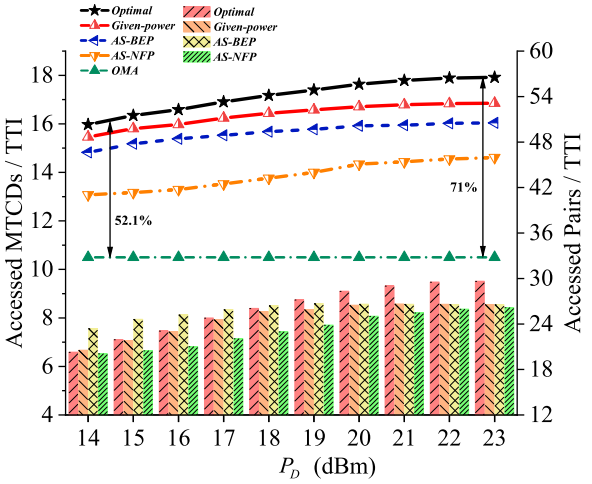
To estimate the system performance, we consider two metrics: 1) the connectivity density metric, which is defined as the number of accessed MTCDs; 2) the pairing density metric, which is defined as the number of accessed pairings/occupied sub-carriers. In addition, to verify the efficiency of our scheme



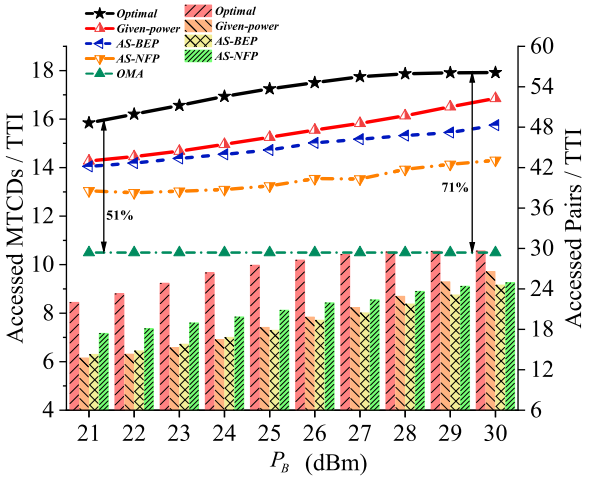
(a) $\bar{\lambda}_d$ is set to 40 kbps. P_D and P_B are 23 dBm and 30 dBm, respectively.



(b) $\bar{\lambda}_u$ is set to 40 kbps. P_D and P_B are 23 dBm and 30 dBm, respectively.



(c) $\bar{\lambda}_u$ and $\bar{\lambda}_d$ are set to 40 kbps.



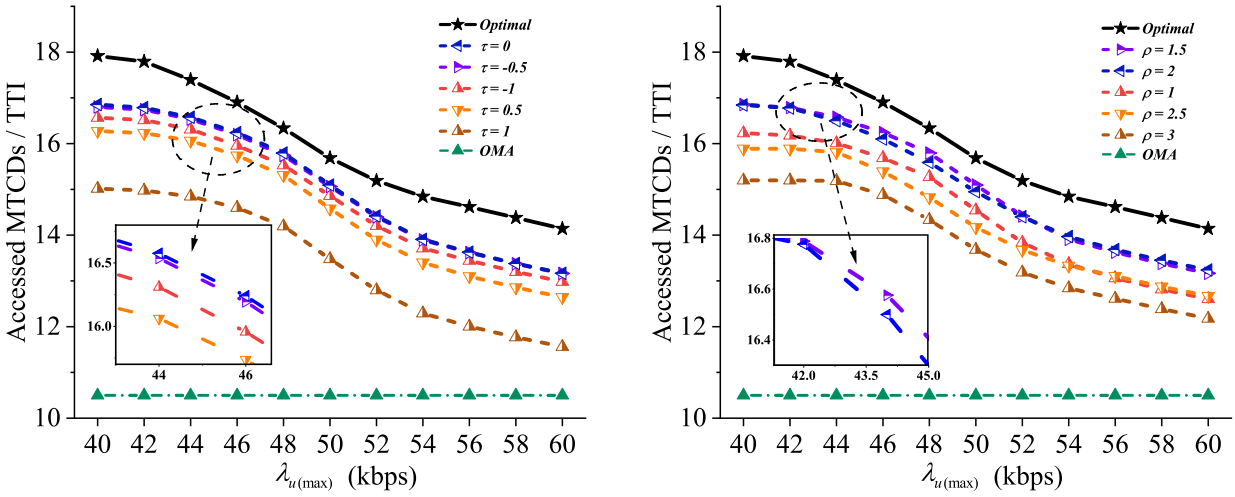
(d) $\bar{\lambda}_u$ and $\bar{\lambda}_d$ are set to 40 kbps.

Fig. 3. Connection density (in curve with left-axis) and pairing density (in histogram with right-axis) versus $\lambda_{u(max)}$, $\lambda_{d(max)}$, P_D, and P_B, respectively. ξ and ω are set to 0.7 and 0.2, respectively. τ and ρ are set to 0 and 1.5, respectively.

for radio resource management, the number of MTCDs sending data simultaneously is assumed to be sufficiently greater than the number of available sub-carriers. Thus, we assume there are 30 sub-carriers, each with bandwidth 6 kHz for either uplink or downlink, to serve 30 *strong* set devices and 30 *weak* set devices. To show the effectiveness of our proposed algorithm, we compare our simulation results with the following benchmark algorithms:

- Optimal: The result from solving MILP $\mathscr{P}2$ by toolbox.
- Given-power: The result from solving ILP $\mathscr{P}3$ by toolbox.
- Near-far Pairing (NFP): It is a conventional NOMA scheme that pairs the closest (second closest, ...) device in the *strong* set with the furthest (second furthest, ...) device in the *weak* set, respectively. This scheme can be used for either uplink or downlink, and cannot apply to the case that the devices have both uplink and downlink transmission simultaneously.
- Alternating Selection Near-far Pairing (AS-NFP): It is a scheme that extend the NFP to the system where both uplink and downlink exist simultaneously.
- Refined Near-far Pairing (R-NFP): It is a scheme that refines the pairs from NFP with respect to the downlink transmit power. Since the pairs in NFP may require more power than the maximum power of BS, in R-NFP, we only choose the devices that do not violate the maximum power constraint when pairing.
- Orthogonal Multiple Access (OMA): In OMA, each sub-carrier can only be allocated to a single MTCD as long as this device can meet its minimum data rate requirement.

Fig. 3 shows the performance of AS-BEP algorithm with variations in transmit power and QoS requirements in the coexistence of three types of services scenarios, i.e., type A, B and C. In Fig. 3(a) and 3(b), $\bar{\lambda}_u$ and $\bar{\lambda}_d$ are randomly distributed in the range $\big(38, \lambda_{u(max)}\big)$ and $\big(38, \lambda_{d(max)}\big)$ respectively. We can conclude that, as $\lambda_{u(max)}$ and $\lambda_{d(max)}$ increase, the number of accessed MTCDs and pairs decreases gradually. With NOMA, there is an approximately 34.6% to



(a) $\rho = 1.5$ (b) $\tau = 0$

Fig. 4. Connection density with varying τ and ρ versus $\lambda_{u(max)}$. ξ and ω are set to 0.7 and 0.2, respectively. $\bar{\lambda}_d$ is set to 40 kbps.

70.6% increase w.r.t connectivity density compared to OMA. We can also learn from Fig. 3(c) and 3(d), the connectivity density and pairing density increase with the increase of transmit power, P_D and P_B. In comparison to OMA, there is also an improvement from 51% to 71% in connectivity density. It is evident that the proposed AS-BEP algorithm can perform much better than the conventional OMA scheme and AS-NFP scheme. In addition, from Fig. 3, we can learn that AS-BEP has close performance to Given-power and outperforms AS-NFP w.r.t. the connectivity density. And in some cases, AS-BEP has better performance than Optimal and Given-power w.r.t. the pairing density in Fig. 3(a) and 3(c), which verifies the original intention of our algorithm.

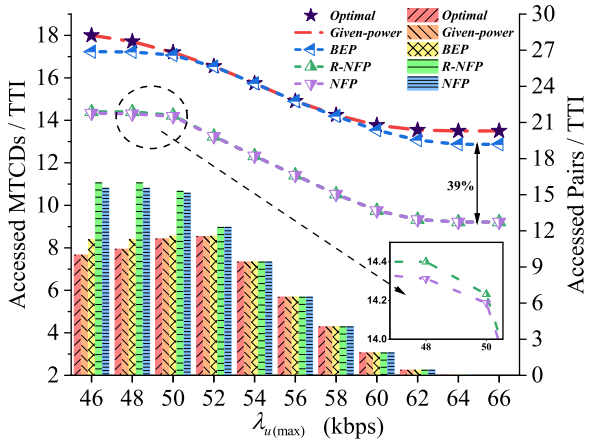
It is worth nothing that the performance of Given-power in Fig. 3(a) and 3(c) is closer to Optimal compared than that of Fig. 3(b) and 3(d). This is because, with Given-power scheme, we can find out the optimal transmit power for MTCDs in the uplink associated with $\bar{\lambda}_u$ and P_D. In contrast, there is no optimal power allocation for MTCDs in the downlink associated with $\bar{\lambda}_d$ and P_B, and it depends on parameters τ and ρ in FTPC. As $\bar{\lambda}_d$ and P_B change, τ and ρ, in theory, should be dynamically set to achieve better performance.

Fig. 4 describes the impact of the parameters τ and ρ in FTPC on the Given-power scheme. We can learn from Fig. 4(a) that when ρ is fixed, Given-power has the best performance with $\tau = 0$ which means that the principle of average power allocation is adopted in FTPC. As τ increases, allocating more power to the device with lower channel gain reduces the overall ability of the *weak* set devices to accommodate interference. As a result, the performance of Given-power deteriorates. From Fig. 4(b), when τ is fixed, it is observed that Given-power performs best with ρ taking a value around 1.5. This is because an appropriate increase in ρ can make the inequality (38) hold for more devices. Meanwhile, due to the fact that the power limitation at BS, the value of ρ cannot be infinitely large or it will result in a decrease in performance.

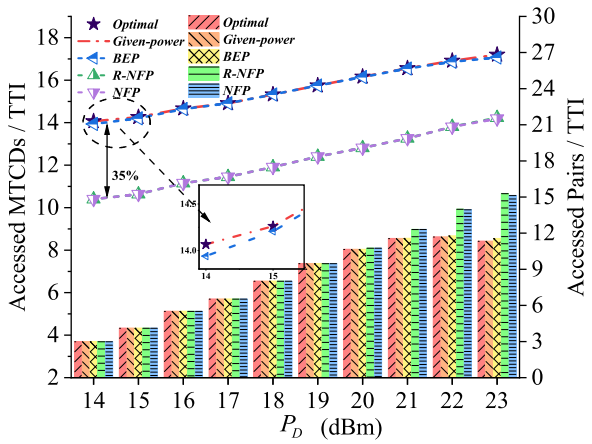
Fig. 5 illustrates the performance of Given-power scheme and BEP algorithm in the situation where only type A and type B services exist in the system simultaneously. In such a case, AS-BEP and AS-NFP algorithms degenerate into BEP and NFP algorithms, respectively. In Fig. 5, we examine the impact of uplink transmission rate and power of MTCDs on the connectivity, respectively. We can learn from Fig. 5 that the connectivity performance of Given-power is identical with the Optimal, which validates our approach to optimize the uplink transmit power, i.e., Lemma 1 is optimal for uplink NOMA. In addition, BEP has close performance to Optimal and Given-power w.r.t. the accessed MTCDs, especially in Fig. 5(b), and BEP also has better performance than Optimal and Given-power w.r.t. the accessed pairs, which shows the efficiency of our proposed algorithm. Besides, the performance of BEP surpasses both the celebrated NFP and R-NFP. For example, at $\xi = 0.9$, BEP accesses up to 39% and 35% more devices in Fig. 5(a) and 5(b), respectively, compared to NFP.

VI. Conclusions and Future Works

In this paper, we employed NOMA and network slicing to enhance the connection density with multiple services in IIoT. In particular, we leveraged the non-orthogonal network slicing to support three distinct types of services. To maximize the number of accessed MTCDs, we formulated joint sub-carrier association and transmit power allocation optimization problems as an MINLP and transformed it into a MILP. To further reduce the computational complexity, we then reduced the MILP into an ILP by designing an effective transmit power allocation scheme, and proposed the low-complexity AS-BEP algorithm to solve the ILP. Finally, we evaluated the performance of the proposed schemes by comparing with a wide range of benchmark algorithms and concluded that our proposed algorithms achieved the close-to-optimal performance and enjoyed great improvements against the existing methods. In the future work, we aim to extend our algorithms to more general cases, and address the complicated



(a) $\bar{\lambda}_d$ is set to 50 kbps. P_D and P_B are 23 dBm and 30 dBm, respectively.



(b) $\bar{\lambda}_u$ and $\bar{\lambda}_d$ are set to 50 kbps and 60 kbps, respectively.

Fig. 5. Connection density (in curve with left-axis) and pairing density (in histogram with right-axis) versus $\lambda_{u(max)}$ and P_D. ξ and ω are set to 0.9 and 0.1, respectively.

optimization problem by using machine learning approaches. Moreover, from a more practical perspective, imperfect CSI will also be the next direction of our research.

Appendix A
Derivation for constraint $C1$ in (20)

Based on (3) and (21), for $k \in \mathcal{C}^{A,B}$, $C1$ can be rewritten as

$$\sum_{s\in\mathcal{S}} W\log_2\left(1+\frac{\alpha_{k,s}p^u_{k,s}\|\boldsymbol{h}^u_k\|_2^2}{I_{k,s}+N_0W}\right) \geq \left(\sum_{s\in\mathcal{S}}\alpha_{k,s}\right)\bar{\lambda}_u. \quad (45)$$

According to (10), there is at most one item being zero on the left side of formula (45). Based on $\log(1+a)+\log(1+b)=\log(1+a+b)$ if one of a and b is 0, (45) can be rewritten as

$$\sum_{s\in\mathcal{S}} W\log_2\left(1+\frac{\alpha_{k,s}p^u_{k,s}\|\boldsymbol{h}^u_k\|_2^2}{I_{k,s}+N_0W}\right) \geq \left(\sum_{s\in\mathcal{S}}\alpha_{k,s}\right)\bar{\lambda}_u \Rightarrow$$
$$W\log_2\left(1+\frac{\alpha_{k,1}p^u_{k,1}\|\boldsymbol{h}^u_k\|_2^2}{I_{k,1}+N_0W}+\cdots+\frac{\alpha_{k,S_{tol}}p^u_{k,S_{tol}}\|\boldsymbol{h}^u_k\|_2^2}{I_{k,S_{tol}}+N_0W}\right)$$
$$\geq \left(\sum_{s\in\mathcal{S}}\alpha_{k,s}\right)\bar{\lambda}_u. \quad (46)$$

Note that

$$1+\frac{\alpha_{k,1}p^u_{k,1}\|\boldsymbol{h}^u_k\|_2^2}{I_{k,1}+N_0W}+\cdots+\frac{\alpha_{k,S_{tol}}p^u_{k,S_{tol}}\|\boldsymbol{h}^u_k\|_2^2}{I_{k,S_{tol}}+N_0W} = 1+ \frac{\alpha_{k,1}p^u_{k,1}\|\boldsymbol{h}^u_k\|_2^2+\cdots+\alpha_{k,S_{tol}}p^u_{k,S_{tol}}\|\boldsymbol{h}^u_k\|_2^2}{\alpha_{k,1}I_{k,1}+\cdots+\alpha_{k,S_{tol}}I_{k,S_{tol}}+N_0W}. \quad (47)$$

From (46) and (47), we can transform $C1$ into the equivalent constraints, that is,

$$\left(2^{\frac{\left(\sum_{s\in\mathcal{S}}\alpha_{k,s}\right)\bar{\lambda}_u}{W}}-1\right) \leq \frac{\alpha_{k,1}p^u_{k,1}\|\boldsymbol{h}^u_k\|_2^2+\cdots+\alpha_{k,S_{tol}}p^u_{k,S_{tol}}\|\boldsymbol{h}^u_k\|_2^2}{\alpha_{k,1}I_{k,1}+\cdots+\alpha_{k,S_{tol}}I_{k,S_{tol}}+N_0W} \Rightarrow$$
$$\left(2^{\frac{\left(\sum_{s\in\mathcal{S}}\alpha_{k,s}\right)\bar{\lambda}_u}{W}}-1\right)\left(N_0W+\sum_{s\in\mathcal{S}}\alpha_{k,s}I_{k,s}\right) \leq \|\boldsymbol{h}^u_k\|_2^2\sum_{s\in\mathcal{S}}\alpha_{k,s}p^u_{k,s}. \quad (48)$$

Based on (48), if the k-th ($k \in \mathcal{C}^{A,B}$) device is to be accessed over the s-th sub-carrier, we can finally obtain

$$\alpha_{k,s}\left(I_{k,s}+N_0W\right)\left(2^{\frac{\bar{\lambda}_u}{W}}-1\right) \leq p^u_{k,s}\|\boldsymbol{h}^u_k\|_2^2. \quad (49)$$

References

[1] B. Yin, J. Tang, and M. Wen, "Maximizing the connectivity of wireless network slicing enabled industrial Internet-of-Things," in *Proc. IEEE GLOBECOM*, Madrid, Spain, Dec. 2021, pp. 1–6.

[2] H. Boyes, B. Hallaq, C. J *et al.*, "The industrial Internet of Things (IIoT): An analysis framework," *Computers in Industry*, vol. 101, pp. 1–12, Oct. 2018.

[3] M. R. Palattella, M. Dohler, A. Grieco *et al.*, "Internet of Things in the 5G era: Enablers, architecture, and business models," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 510–527, Mar. 2016.

[4] M. Aazam, S. Zeadally, and K. A. Harras, "Deploying fog computing in industrial Internet of Things and industry 4.0," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4674–4682, Oct. 2018.

[5] E. Sisinni, A. Saifullah, S. Han *et al.*, "Industrial Internet of Things: Challenges, opportunities, and directions," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 4724–4734, Nov. 2018.

[6] J. Du, F. R. Yu, G. Lu *et al.*, "MEC-assisted immersive VR video streaming over terahertz wireless networks: A deep reinforcement learning approach," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9517–9529, Oct. 2020.

[7] M. Shafi, A. F. Molisch, P. J. Smith *et al.*, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[8] T. Jiang, J. Zhang, P. Tang *et al.*, "3GPP standardized 5G channel model for IIoT scenarios: A survey," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 8799–8815, Jun. 2021.

[9] H. Xu, W. Yu, D. Griffith *et al.*, "A survey on industrial Internet of Things: A cyber-physical systems perspective," *IEEE Access*, vol. 6, pp. 78 238–78 259, Dec. 2018.

[10] S. Li, Q. Ni, Y. Sun *et al.*, "Energy-efficient resource allocation for industrial cyber-physical IoT systems in 5G era," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2618–2628, Jun. 2018.

[11] K. Zhang, Y. Zhu, S. Maharjan *et al.*, "Edge intelligence and blockchain empowered 5G beyond for the industrial Internet of Things," *IEEE Netw.*, vol. 33, no. 5, pp. 12–19, Sept. 2019.

[12] L. Lyu, C. Chen, S. Zhu *et al.*, "5G enabled codesign of energy-efficient transmission and estimation for industrial IoT systems," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2690–2704, Jun. 2018.

[13] Y. Liu, M. Kashef, K. B. Lee *et al.*, "Wireless network design for emerging IIoT applications: Reference framework and use cases," *Proc. IEEE*, vol. 107, no. 6, pp. 1166–1192, Jun. 2019.

[14] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple access techniques for 5G wireless networks and beyond*. Springer, 2019, vol. 159.

[15] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[16] Y. Saito, Y. Kishiyama, A. Benjebbour *et al.*, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE VTC*, Dresden, Germany, Jun. 2013, pp. 1–5.

[17] S. M. R. Islam, N. Avazov, O. A. Dobre *et al.*, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quarter. 2017.

[18] L. Dai, B. Wang, Z. Ding *et al.*, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quarter. 2018.

[19] T. Hou, Y. Liu, Z. Song *et al.*, "Reconfigurable intelligent surface aided NOMA networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2575–2588, Nov. 2020.

[20] X. Pang, J. Tang, N. Zhao *et al.*, "Energy-efficient design for mmwave-enabled NOMA-UAV networks," *Sci. China Inf. Sci.*, vol. 64, no. 4, pp. 1–14, Jul. 2021.

[21] N. Zhao, Y. Li, S. Zhang *et al.*, "Security enhancement for NOMA-UAV networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 3994–4005, Feb. 2020.

[22] I. Afolabi, T. Taleb, K. Samdanis *et al.*, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2429–2453, 3rd Quarter. 2018.

[23] P. V. Kafle, Y. Fukushima, P. Martinez-Julia *et al.*, "Adaptive virtual network slices for diverse IoT services," *IEEE Commun. Stand. Mag.*, vol. 2, no. 4, pp. 33–41, Dec. 2018.

[24] S. Wijethilaka and M. Liyanage, "Survey on network slicing for Internet of Things realization in 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 957–994, 2nd Quarter. 2021.

[25] D. Zhai and R. Zhang, "Joint admission control and resource allocation for multi-carrier uplink NOMA networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 922–925, Dec. 2018.

[26] Z. Zhang, Y. Hou, Q. Wang *et al.*, "Joint sub-carrier and transmission power allocation for MTC under power-domain NOMA," in *Proc. IEEE ICC*, Kansas City, MO, USA, May 2018, pp. 1–6.

[27] S. Mishra, L. Salaün, and C. S. Chen, "Maximizing connection density in NB-IoT networks with NOMA," in *Proc. IEEE VTC*, Antwerp, Belgium, May 2020, pp. 1–6.

[28] A. E. Mostafa, Y. Zhou, and V. W. S. Wong, "Connection density maximization of narrowband IoT systems with NOMA," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4708–4722, Oct. 2019.

[29] P. Popovski, K. F. Trillingsgaard, O. Simeone *et al.*, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55 765–55 779, Sept. 2018.

[30] J. Tang, B. Shim, T. H. Chang *et al.*, "Incorporating URLLC and multicast eMBB in sliced cloud radio access network," in *Proc. IEEE ICC*, Shanghai, China, May 2019, pp. 1–7.

[31] J. Tang, B. Shim, and T. Q. S. Quek, "Service multiplexing and revenue maximization in sliced C-RAN incorporated with URLLC and multicast eMBB," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 881–895, Apr. 2019.

[32] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4438–4454, Jun. 2016.

[33] Z. Zhang, H. Sun, and R. Q. Hu, "Downlink and uplink non-orthogonal multiple access in a dense wireless network," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2771–2784, Dec. 2017.

[34] G. Zhu, C. Zhong, H. A. Suraweera *et al.*, "Outage probability of dual-hop multiple antenna AF systems with linear processing in the presence of co-channel interference," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2308–2321, Apr. 2014.

[35] Y. Liu, H. Xing, C. Pan *et al.*, "Multiple-antenna-assisted non-orthogonal multiple access," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 17–23, Apr. 2018.

[36] M. Tawarmalani and N. V. Sahinidis, *Convexification and global optimization in continuous and mixed-integer nonlinear programming: theory, algorithms, software, and applications*. Springer Science & Business Media, 2013, vol. 65.

[37] F. S. Hillier and G. J. Lieberman, *Introduction to operations research*. Tata McGraw-Hill Education, 2012.

[38] S. Burer and A. N. Letchford, "Non-convex mixed-integer nonlinear programming: A survey," *Surveys in Operations Research and Management Science*, vol. 17, no. 2, pp. 97–106, 2012.

[39] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[40] S. M. R. Islam, M. Zeng, O. A. Dobre *et al.*, "Resource allocation for downlink NOMA systems: Key techniques and open issues," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 40–47, Apr. 2018.

[41] 3GPP, "Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)," 3GPP TR 45.820 V13.1.0, Nov. 2015.

Bo Yin received the B.Eng. degree in information engineering and M.Eng. degree in information and communication engineering from South China University of Technology, Guangzhou, China, in 2019 and 2022, respectively. He is currently pursuing the Ph.D. degree in electronics and ICT engineering with the Department of Information Technology at Ghent University, Belgium. His research interests include network planning, machine learning, and wireless communications.

Jianhua Tang (S'11-M'15) received the B.E. degree in communications engineering from Northeastern University, China, in 2010, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2015. He was a Post-Doctoral Research Fellow with the Singapore University of Technology and Design from 2015 to 2016, and a Research Assistant Professor with the Department of Electrical and Computer Engineering, Seoul National University, South Korea, from 2016 to 2018. He is currently an Associate Professor with the Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, China. His research interests include edge computing, network slicing and industrial Internet of Things.

He was honored with the 2020 IEEE Communications Society Stephen O. Rice Prize. He is currently serving as an Editor for the IEEE WIRELESS COMMUNICATIONS LETTERS.

Miaowen Wen (SM'18) received the Ph.D. degree from Peking University, Beijing, China, in 2014. From 2019 to 2021, he was with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, as a Post-Doctoral Research Fellow. He is currently a Professor with the South China University of Technology, Guangzhou, China. He has published two books and more than 170 journal articles. His research interests include a variety of topics in the areas of wireless and molecular communications.

He was a recipient of the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2020, and five Best Paper Awards from the IEEE ITST'12, the IEEE ITSC'14, the IEEE ICNC'16, the IEEE ICCT'19, and the EAI QSHINE'22. He was the Winner in data bakeoff competition (Molecular MIMO) from the 2019 IEEE Communication Theory Workshop (CTW), Selfoss, Iceland. He served as a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He is currently serving as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS, and the IEEE COMMUNICATIONS LETTERS.