Connection Density Maximization of Narrowband IoT Systems with NOMA

Ahmed Elhamy Mostafa, *Student Member, IEEE*, Yong Zhou, *Member, IEEE*, and Vincent W.S. Wong, *Fellow, IEEE*

***Abstract*—Narrowband Internet of Things (NB-IoT) provides energy-efficient communications with extended coverage for low data rate IoT devices. In this paper, we propose a power-domain non-orthogonal multiple access (NOMA) scheme for NB-IoT systems to enhance the connection density by allowing multiple IoT devices to simultaneously access one subcarrier. We consider both single-tone and multi-tone transmission modes of NB-IoT systems, where each device can access a single subcarrier or a bond of contiguous subcarriers, respectively. We formulate joint subcarrier and power allocation problems for both transmission modes to maximize the connection density while taking the quality of service requirements and the transmit power constraints of IoT devices into account. We solve the single-tone nonconvex mixed integer programming problem by transforming it into a mixed integer linear programming problem to obtain the optimal solution. The multi-tone problem is solved by using the difference of convex programming approach to obtain a close-to-optimal solution. We also propose low-complexity heuristic algorithms to solve both problems in a suboptimal manner. Simulations results show that our proposed scheme increases the connection density of NB-IoT systems by 87% in the single-tone mode and by 24% in the multi-tone mode compared to orthogonal multiple access.**

***Index Terms*—Connection density maximization, massive Internet of Things (mIoT), narrowband IoT (NB-IoT), non-orthogonal multiple access (NOMA), power and subcarrier allocation.**

I. Introduction

Machine-type communications (MTC) [1] enable devices having sensing, processing and communication capabilities to provide a wide variety of services with minimal human intervention. The Long Term Evolution-Advanced (LTE-A) Pro cellular technology is a candidate solution to connect the MTC devices to the Internet, pushing forward the emerging paradigm of the Internet of Things (IoT). The International Telecommunications Union (ITU) introduces three use cases for the fifth generation (5G) networks, which are enhanced mobile broadband (eMBB), ultra-reliable and low latency communications (URLLC) and massive IoT (mIoT) [2]. URLLC and mIoT are the use cases associated with enabling the IoT. Providing cellular coverage for IoT devices in URLLC and mIoT is different from that for conventional user equipment (UE) devices in eMBB. In particular, the density of IoT devices is expected to reach 10^6 devices per km^2 [3], which is much higher than that of UEs, as the number of IoT devices worldwide may reach 3.9 billion by 2022 [4]. For IoT devices, uplink data transmission represents a considerable share of data traffic while UEs mostly receive data in the downlink [5].

Although URLLC and mIoT have common features, such as uplink data transmission and operation with minimal human intervention, they have different characteristics. URLLC devices are subject to high reliability requirements (i.e., probability of successful packet delivery $\geq$ 0.99999) and strict latency constraints in the order of several milliseconds [5]. These devices are usually associated with mission-critical applications, such as e-health and autonomous driving. On the other hand, mIoT devices (e.g., environmental sensors and wearables) usually perform delay-tolerant tasks with relaxed latency requirements in the order of seconds or hours [5]. High energy-efficiency and high connection density are the main requirements of mIoT devices [5]. Narrowband IoT (NB-IoT) is standardized by the Third Generation Partnership Project (3GPP) to support low-rate energy-efficient communication for mIoT devices [6, Sec. 16, pp. 506−533].

In LTE-A Pro networks, NB-IoT can support an uplink data rate up to 250 kbps [7]. NB-IoT allocates subcarriers of a narrow bandwidth (e.g., 3.75 kHz or 15 kHz) to IoT devices [8, p. 201]. By communicating over a narrow bandwidth, the noise power is reduced and the network coverage can be enhanced by 20 dB [7]. The coverage enhancement improves the reliability of NB-IoT signals for cell-edge and indoor IoT devices. Hence, NB-IoT is a candidate technology for many IoT applications, such as smart metering and smart hospitals [7]. Enhancing NB-IoT is one of the suggested work items in Release 15 and Release 16 of the 3GPP New Radio [9], [10]. These work items include enhancing NB-IoT by supporting small cell operation, improving the uplink and downlink transmission efficiency and reducing the latency and the energy consumption. In addition, 3GPP New Radio will support non-stand-alone operation to coexist on top of the LTE network and it can benefit from the enhancement made to LTE supporting technologies, such as NB-IoT [10]. Potential approaches of enhancing NB-IoT to achieve the goals of the work items include improving the random access procedure and the medium access control for data transmission.

In NB-IoT, random access is performed in a way similar to

Manuscript received on Aug. 16, 2018; revised on Nov. 20, 2018 and May 3, 2019; accepted on Jun. 26, 2019. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. This paper has been published in part in the *Proceedings of IEEE International Conference on Communications (ICC)*, Paris, France, May 2017. The editor coordinating the review of this paper and approving it for publication was Jinhong Yuan.

A. E. Mostafa and V. W.S. Wong are with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, V6T 1Z4, Canada (e-mail: {ahmedem, vincentw}@ece.ubc.ca).

Y. Zhou is with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China (e-mail: zhouyong@shanghaitech.edu.cn).

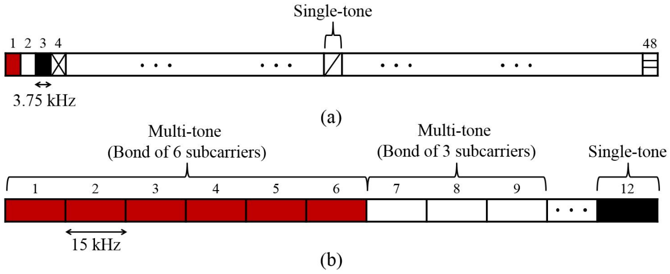


Fig. 1. NB-IoT transmission modes: (a) The PRB is divided into 48 subcarriers. The subcarrier bandwidth is 3.75 kHz. Only single-tone mode is supported. (b) The PRB is divided into 12 subcarriers. The subcarrier bandwidth is 15 kHz. Both single-tone and multi-tone modes are supported.

conventional LTE-A. An IoT device sends a single preamble consisting of four symbol groups multiple times over a set of subcarriers of the narrowband physical random access channel (NPRACH) with a certain frequency hopping sequence [8, pp. 208−211]. The random access procedure in NB-IoT systems may result in preamble collisions due to the use of a single preamble sequence for all NB-IoT devices and a limited number of subcarriers for NPRACH. When a preamble collision occurs, the device performs backoff and retransmission. The settings of NB-IoT coverage enhancement levels and preamble retransmission parameters are studied in [11] to increase the random access success probability. The random access success probability can also be improved by configuring the number of preamble transmission repetitions [12].

For data transmission in NB-IoT, the system bandwidth is equivalent to the bandwidth of one physical resource block (PRB), i.e., 180 kHz. The PRB can be divided into either 48 or 12 subcarriers [8, p. 201]. For the case with 48 subcarriers, each IoT device can be allocated a single subcarrier, referred to as the *single-tone* mode, as shown in Fig. 1(a). For the case with 12 subcarriers, each IoT device can be allocated either a single subcarrier (single-tone mode), or a bond of 3, 6, or 12 contiguous subcarriers, referred to as the *multi-tone* mode, as shown in Fig. 1(b). These modes enable medium access with single-carrier frequency division multiple access (SC-FDMA), i.e., an orthogonal multiple access (OMA) scheme, which may not be sufficient to support the use case of mIoT in the near future since each subcarrier can only be accessed by a single IoT device at a given time. To overcome this problem, an algorithm is proposed in [13] to offload NB-IoT traffic to small cells operating in the unlicensed bands in order to improve the aggregate data rate and secrecy for some IoT applications (e.g., surveillance cameras). In [14], it is proposed to adopt a compressed signaling technique termed fast orthogonal frequency division multiplexing to use spectrum more efficiently and provide connection for more IoT devices.

To further support mIoT in NB-IoT systems, it is required to allow a large number of devices to transmit data simultaneously over a limited number of subcarriers. Non-orthogonal multiple access (NOMA) [15]–[19] is a promising technology that can be used to enable multiple devices to access a single subcarrier concurrently for data transmission. In uplink power-domain NOMA, multiple devices, each with different transmit power, transmit their messages using the same subcarrier at the same time. After receiving these messages, the base station exploits the received power level difference to decode them sequentially using successive interference cancellation (SIC). The message with the highest received power level is decoded first. Subsequently, the decoded message is subtracted from the combined message to facilitate the decoding of the messages with lower received power levels.

Uplink NOMA transmission has been investigated in many works [20]–[27]. Power allocation strategies are designed for uplink transmission of multiple-input multiple-output (MIMO) NOMA systems based on the cognitive radio concept in [20]. In [21], power allocation is designed to guarantee the quality of service (QoS) requirements in MIMO-NOMA systems. In [22], a joint power and resource block allocation problem is formulated for data rate maximization in uplink NOMA systems and a heuristic solution is proposed. The power allocation and the resource block allocation problems are formulated as two separate problems in [23]. First, many-to-many matching is used to pair users and allocate channels. Then, for a given resource allocation, the power allocation problem is solved using either an iterative water-filling algorithm or geometric programming. Optimal joint power and subcarrier allocation to maximize the aggregate data rate of full-duplex NOMA systems is obtained using monotonic optimization in [24]. In [25], a user pairing scheme is proposed for uplink single-antenna NOMA systems. This scheme is based on obtaining tight upper and lower bounds of the aggregate rate per subcarrier in low signal-to-noise ratio (SNR) regimes and it is proven to result in a near-optimal sum rate performance.

The aforementioned studies focus on the data rate maximization of human type communications. However, in mIoT use case, maximizing the connection density is the main requirement. Random NOMA is proposed in [26], where each IoT device randomly selects a subcarrier for data transmission, and the maximum arrival rate to maintain the number of devices transmitting data below a certain limit is derived. In our previous work [27], uplink power-domain NOMA is used to improve the connection density of NB-IoT systems. Heuristic algorithms are proposed to allocate the transmit power and subcarriers with a higher priority given to those IoT devices with higher data rate requirement.

Another approach for supporting uplink NOMA is to use code-domain NOMA, e.g., sparse code multiple access (SCMA) [28]. SCMA enables multiple devices to use complex sparse spreading codes for message bits. Message passing algorithm is used to decode messages with near-optimal performance. Power-domain NOMA is used in our work, rather than code-domain NOMA, due to the simplicity of receiver implementation. To reduce the complexity of the SCMA receiver, the overloading factor should be decreased, i.e., the number of devices per communication channel should be decreased (e.g., six devices accessing four communication channels with an overloading factor of 1.5). With power-domain NOMA, we can enable two devices to access a single communication channel so that we may have $2N$ devices accessing N channels even if this would reduce the achievable data rate of each device compared to the case when N devices access N channels with either OMA or NOMA. However, high data rates are not required for many mIoT applications as long as a minimum data rate threshold is achieved.

The existing works on joint subcarrier and power allocation for NOMA mainly focused on maximizing the aggregate data rate of human-type communications. However, achieving a high aggregate data rate for all the devices is not a major requirement of many mIoT applications, where the mIoT devices are required to transmit data with a minimum data rate threshold. Transmitting data with a data rate that is higher than the minimum threshold may not be necessary since it increases the energy consumption. Furthermore, achieving a lower transmission delay by transmitting data at a higher data rate is not beneficial for mIoT devices that perform delay-tolerant tasks. On the other hand, enabling the maximum number of devices to transmit data with a minimum data rate threshold is more compliant with the connection density requirements of the mIoT use case.

In this paper, we consider an NB-IoT system that provides cellular connections to IoT devices and develop algorithms to maximize the connection density during data transmission using uplink power-domain NOMA. We take into account the minimum data rate and the maximum transmit power constraints of IoT devices. Each subcarrier is accessed by two IoT devices. The data of both devices are decoded sequentially at the base station using SIC. Our objective is to serve the maximum number of IoT devices in a single time slot such that they meet their minimum data rate requirements, i.e., maximizing the connection density. NOMA can enable two devices to access a single subcarrier, and hence the number of simultaneously served devices can be greater than the number of available subcarriers. The main contributions are summarized as follows:

- We formulate joint subcarrier and transmit power allocation problems for single-tone and multi-tone NB-IoT systems to maximize the connection density.
- We obtain the optimal solution of the single-tone problem by transforming it into an equivalent mixed integer linear programming problem. We also obtain a close-to-optimal solution of the multi-tone problem by using difference of convex programming.
- We propose heuristic algorithms that provide suboptimal solutions of the single-tone and multi-tone problems with lower computational complexity by formulating simpler binary integer programming problems.

Simulations show that using NOMA in NB-IoT systems can increase the connection density by up to 87% compared to OMA in single-tone mode and by up to 24% in multi-tone mode. In addition, the proposed heuristic algorithms achieve similar performance to the optimal and close-to-optimal solutions of the single-tone and multi-tone problems, respectively. The proposed algorithms are also shown to outperform NOMA with heuristic pairing of near devices to the base station with far devices from the base station.

The rest of this paper is organized as follows. In Section II, we present the system model. In Section III, we formulate the connection density maximization problem for the single-tone mode in NB-IoT systems and propose an algorithm to solve it. We present and solve the connection density maximization problem for the multi-tone mode in NB-IoT systems in Section IV. Performance evaluation is conducted in Section V. Finally, Section VI concludes this paper.

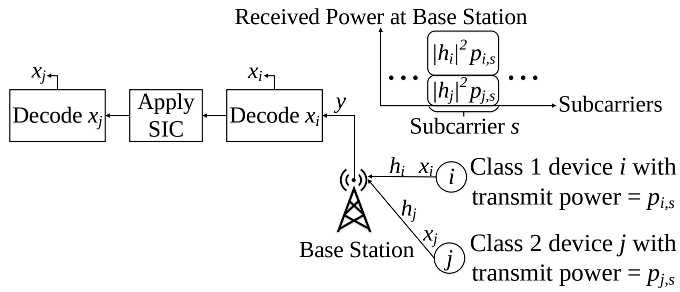


Fig. 2. System model. Each subcarrier can be accessed by one class 1 device and one class 2 device. Symbols from class 1 device are decoded first. Symbols from class 2 device are then decoded subsequently after applying SIC.

II. System Model

Consider a cellular coverage area with a single base station and multiple IoT devices. One PRB is used to provide cellular coverage for IoT devices based on the NB-IoT standard [8, pp. 200−201]. We focus on the mIoT use case in this work. Hence, all the IoT devices have no strict latency requirements. Each device only has a minimum data rate requirement and a maximum transmit power constraint that are known by both the IoT device and the base station. Each IoT device is assigned to a particular class from the set of classes $\mathcal{C} = \{1, 2\}$ based on its data rate requirements. The set of class 1 devices ($c = 1$) is denoted as $\mathcal{D}_1$, and the set of class 2 devices ($c = 2$) is denoted as $\mathcal{D}_2$. The sets of class 1 and class 2 devices are disjoint, i.e., $\mathcal{D}_1 \cap \mathcal{D}_2 = \emptyset$. Without loss of generality, class 1 devices have higher data rate requirements than class 2 devices. To support more IoT devices, we use power-domain NOMA to allow two IoT devices to simultaneously access the same subcarrier. The classes determine the decoding order at the base station such that the data symbols of the higher data rate devices of class 1 are decoded first. Note that devices from both classes have the same priority in our scheme.

In an NB-IoT system, the total bandwidth is divided into S subcarriers. Each subcarrier has a bandwidth of B Hz. We have a set of subcarriers $\mathcal{S} = \{1, \ldots, S\}$. Each subcarrier can be allocated to one class 1 device $i \in \mathcal{D}_1$ and one class 2 device $j \in \mathcal{D}_2$, as shown in Fig. 2. At a given subcarrier $s \in \mathcal{S}$, device $i \in \mathcal{D}_1$ transmits symbol x_i with transmit power $p_{i,s}$ and device $j \in \mathcal{D}_2$ transmits symbol x_j with transmit power $p_{j,s}$ to the base station. We assume $\mathbb{E}[|x_i|^2] = 1$, for all $i \in \mathcal{D}_c, c \in \mathcal{C}$. The base station receives a combined symbol y with the additive white Gaussian noise σ given by

$$y = \sqrt{p_{i,s}} h_i x_i + \sqrt{p_{j,s}} h_j x_j + \sigma, \tag{1}$$

where h_i represents the channel gain between an IoT device $i \in \mathcal{D}_c$, $c \in \mathcal{C}$ and the base station. Flat fading is assumed over the whole narrow system bandwidth of 180 kHz. The channel gain incorporates the effects of path loss and fading. Without loss of generality, we assume $|h_i|^2 p_{i,s} > |h_j|^2 p_{j,s}$, so that the base station can decode symbol x_i sent by class 1 device i first. Subsequently, the base station decodes symbol

x_j sent by class 2 device j after applying SIC. Throughout this work, class 1 set, $\mathcal{D}_1$, includes the IoT devices that have their data decoded first at the base station. Whereas, class 2 set, $\mathcal{D}_2$, includes the IoT devices that have their data decoded after SIC at the base station.

Consider a class 1 device $i \in \mathcal{D}_1$ and a class 2 device $j \in \mathcal{D}_2$ access a given subcarrier $s \in \mathcal{S}$. The signal-to-interference-plus-noise ratio (SINR) of device i is given by

$$\gamma_{i,s} = \frac{|h_i|^2 p_{i,s}}{|h_j|^2 p_{j,s} + N_o B}, \quad i \in \mathcal{D}_1,\ s \in \mathcal{S}, \tag{2}$$

where N_o is the noise power spectral density. On the other hand, the SNR of device j can be expressed as

$$\gamma_{j,s} = \frac{|h_j|^2 p_{j,s}}{N_o B}, \quad j \in \mathcal{D}_2,\ s \in \mathcal{S}. \tag{3}$$

The achievable data rate of class 2 device j, denoted as r_j, can be expressed as the sum of the achievable data rate over all subcarriers

$$r_j = \sum_{s \in \mathcal{S}} B \log_2 \left(1 + \frac{|h_j|^2 p_{j,s}}{N_o B}\right), \quad j \in \mathcal{D}_2. \tag{4}$$

The achievable data rate of class 1 device i, denoted as r_i, is given by

$$r_i = \sum_{s \in \mathcal{S}} B \log_2 \left(1 + \frac{|h_i|^2 p_{i,s}}{\sum_{j \in \mathcal{D}_2} |h_j|^2 p_{j,s} + N_o B}\right), \quad i \in \mathcal{D}_1, \tag{5}$$

where $\sum_{j \in \mathcal{D}_2} |h_j|^2 p_{j,s}$ represents the interference caused by class 2 devices at subcarrier s.

The base station uses its knowledge of the minimum data rate requirements and maximum transmit power constraints of IoT devices to control their transmit power and allocate the subcarriers according to their channel conditions. We assume perfect channel state information (CSI) is available at the base station such that the channel gain $h_i, i \in \mathcal{D}_c, c \in \mathcal{C}$, can be obtained by using the demodulation reference signals [8, pp. 201–203] sent within the narrowband physical uplink shared channel (NPUSCH). A given subcarrier can be allocated to either (a) a pair consisting of one class 1 device and one class 2 device using NOMA or (b) only one device from class 1 using OMA or (c) only one device from class 2 using OMA.

We introduce a binary subcarrier allocation matrix K, where element $k_{i,s}^c$ is equal to one if device $i \in \mathcal{D}_c,\ c \in \mathcal{C}$, is allocated subcarrier $s \in \mathcal{S}$, and $k_{i,s}^c$ is equal to zero otherwise. Each subcarrier can be accessed by at most one class 1 IoT device and one class 2 IoT device, i.e.,

$$\sum_{i \in \mathcal{D}_c} k_{i,s}^c \leq 1, \quad s \in \mathcal{S}, c \in \mathcal{C}. \tag{6}$$

An IoT device can transmit data if it is allocated at least one subcarrier in the current time slot (e.g., transmission time interval (TTI) [7]).

The total transmit power of device i should not exceed the maximum transmit power, P_i. Hence, we have

$$0 \leq \sum_{s \in \mathcal{S}} p_{i,s} \leq P_i, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C}. \tag{7}$$

An IoT device allocates its transmit power only to the allocated subcarriers to it. That is

$$p_{i,s} \leq P_i k_{i,s}^c, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C},\ s \in \mathcal{S}. \tag{8}$$

Constraint (8) indicates that if device i is not allocated subcarrier s (i.e., $k_{i,s}^c = 0$), then $p_{i,s}$ is forced to be zero. If $k_{i,s}^c$ is equal to one, then $p_{i,s}$ can take any value in the range $(0, P_i]$.

Device i is required to satisfy a minimum data rate requirement R_i if it is allocated sufficient subcarriers

$$r_i \geq \min\left(\sum_{s \in \mathcal{S}} k_{i,s}^c, 1\right) R_i \quad i \in \mathcal{D}_c,\ c \in \mathcal{C}. \tag{9}$$

This condition implies that an IoT device is allocated subcarriers if and only if its QoS requirement can be met. If device i can satisfy its data rate requirement, it can be allocated at least one subcarrier, i.e., $\sum_{s \in \mathcal{S}} k_{i,s}^c \geq 1$. Constraint (9) becomes $r_i \geq R_i$ which enforces that the transmit power over allocated subcarriers takes a value in the range $(0, P_i]$ according to constraints (4), (5) and (8). On the other hand, if device i cannot satisfy its data rate requirement, it cannot be allocated any subcarriers, i.e., $\sum_{s \in \mathcal{S}} k_{i,s}^c = 0$. Constraint (9) becomes $r_i \geq 0$ but the transmit power over all subcarriers takes a value of zero, i.e., $p_{i,s} = 0,\ s \in \mathcal{S}$, according to constraint (8). This results in forcing $r_i = 0$ according to constraints (4) and (5).

III. Single-tone Connection Density Maximization

In this section, we focus on the connection density maximization problem when subcarrier bandwidth B is equal to 3.75 kHz, i.e., $S = 48$ subcarriers, as shown in Fig. 1(a). The single-tone mode is the only available transmission mode in this case.

A. Problem Formulation

For the single-tone mode, each IoT device is allocated at most a single subcarrier, i.e.,

$$\sum_{s \in \mathcal{S}} k_{i,s}^c \leq 1, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C}. \tag{10}$$

Constraint (10) implies that $\sum_{s \in \mathcal{S}} k_{i,s}^c \in \{0, 1\}$. Hence, constraint (9) can be written without the $\min$ operator as follows:

$$r_i \geq \sum_{s \in \mathcal{S}} k_{i,s}^c R_i, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C}. \tag{11}$$

The connection density maximization problem in single-tone mode aims to maximize the number of IoT devices that are allocated one subcarrier and meet their QoS requirements while satisfying the transmit power constraints. To achieve this goal, it is required to control the transmit power such that class 1 and class 2 devices are paired using NOMA (i.e., for a given subcarrier s, the best case is to have $\sum_{i \in \mathcal{D}_1} k_{i,s}^1 = \sum_{j \in \mathcal{D}_2} k_{j,s}^2 = 1$ according to (6)). In this case, each subcarrier can be accessed by two devices which can increase the connection density. The problem is formulated as follows:

$$\underset{P_{\mathcal{D}_1},\, P_{\mathcal{D}_2},\, K}{\text{maximize}} \quad \sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{D}_c} \sum_{s \in \mathcal{S}} k_{i,s}^c \tag{12a}$$

$$\text{subject to} \quad k^c_{i,s} \in \{0,1\}, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C},\ s \in \mathcal{S} \quad \text{(12b)}$$
$$\text{constraints (4)–(8), (10)–(11),}$$

where $P_{\mathcal{D}_1} = [p_{i,s}] \in \mathbf{R}^{|\mathcal{D}_1| \times S}$ and $P_{\mathcal{D}_2} = [p_{j,s}] \in \mathbf{R}^{|\mathcal{D}_2| \times S}$ are matrices of the allocated transmit power per subcarrier of class 1 and class 2 IoT devices, respectively. Note that $|\mathcal{D}_c|$ is the cardinality of the set $\mathcal{D}_c,\ c \in \mathcal{C}$. Problem (12) is a mixed integer nonconvex problem due to nonconvex constraints (5) and (11) for all IoT devices $i \in \mathcal{D}_1$.

B. Problem Transformation

In order to obtain an optimal solution for problem (12), we formulate an equivalent mixed integer linear programming problem by making use of the fact that each IoT device is allocated a single subcarrier. Based on (5), constraint (11) for device $i \in \mathcal{D}_1$ can be expressed as (Please refer to Appendix for detailed steps):

$$|h_i|^2 \sum_{s \in \mathcal{S}} p_{i,s} \geq \left(2^{\frac{R_i}{B}} - 1\right) \left(N_o B + \sum_{s \in \mathcal{S}} k^1_{i,s} \sum_{j \in \mathcal{D}_2} |h_j|^2 p_{j,s}\right) - \left(1 - \sum_{s \in \mathcal{S}} k^1_{i,s}\right)\left(2^{\frac{R_i}{B}} - 1\right) N_o B,\ i \in \mathcal{D}_1. \quad (13)$$

We introduce the variable $I^1_{i,s}$ to represent the interference suffered by class 1 device i at subcarrier s such that

$$I^1_{i,s} = k^1_{i,s} \sum_{j \in \mathcal{D}_2} |h_j|^2 p_{j,s}, \quad i \in \mathcal{D}_1,\ s \in \mathcal{S}. \quad (14)$$

Constraint (14) can be expressed as a group of inequalities using the big-M formulation [24], [29, Ch. 2]

$$I^1_{i,s} \leq \sum_{j \in \mathcal{D}_2} |h_j|^2 p_{j,s}, \quad i \in \mathcal{D}_1, s \in \mathcal{S} \quad \text{(15a)}$$
$$I^1_{i,s} \leq k^1_{i,s} \sum_{j \in \mathcal{D}_2} |h_j|^2 P_j, \quad i \in \mathcal{D}_1, s \in \mathcal{S} \quad \text{(15b)}$$
$$I^1_{i,s} \geq 0, \quad i \in \mathcal{D}_1, s \in \mathcal{S} \quad \text{(15c)}$$
$$I^1_{i,s} \geq \sum_{j \in \mathcal{D}_2} |h_j|^2 p_{j,s} - (1 - k^1_{i,s}) \sum_{j \in \mathcal{D}_2} |h_j|^2 P_j, \quad i \in \mathcal{D}_1, s \in \mathcal{S}. \quad \text{(15d)}$$

If $k^1_{i,s}$ is equal to zero, then $I^1_{i,s}$ is also equal to zero. On the other hand, if $k^1_{i,s}$ is equal to one, then $I^1_{i,s}$ is equal to $\sum_{j \in \mathcal{D}_2} |h_j|^2 p_{j,s}$. Hence, we can write constraint (13) as a function of $I^1_{i,s}$ as

$$|h_i|^2 \sum_{s \in \mathcal{S}} p_{i,s} \geq \left(2^{\frac{R_i}{B}} - 1\right) \left(N_o B + I^1_{i,s}\right) - \left(1 - \sum_{s \in \mathcal{S}} k^1_{i,s}\right)\left(2^{\frac{R_i}{B}} - 1\right) N_o B,\ i \in \mathcal{D}_1. \quad (16)$$

Furthermore, the QoS constraint of class 2 devices, which is a convex constraint, can be written in an affine form using constraints (4), (8) and (10) as

$$p_{j,s} |h_j|^2 \geq k^2_{j,s} N_o B \left(2^{\frac{R_j}{B}} - 1\right), \quad j \in \mathcal{D}_2,\ s \in \mathcal{S}. \quad (17)$$

The equivalent problem of problem (12) can be formulated as follows:

$$\underset{P_{\mathcal{D}_1},\, P_{\mathcal{D}_2},\, I_{\mathcal{D}_1},\, K}{\text{maximize}} \quad \sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{D}_c} \sum_{s \in \mathcal{S}} k^c_{i,s} \quad (18)$$
$$\text{subject to} \quad \text{constraints (6)–(8), (10), (12b), (15a)–(15d), (16), (17),}$$

where $I_{\mathcal{D}_1} = [I^1_{i,s}] \in \mathbf{R}^{|\mathcal{D}_1| \times S}$ is the matrix of the interference suffered by class 1 devices per subcarrier. Problem (18) is a mixed integer linear programming problem, which can be solved by using CVX with Gurobi (or Mosek) solver to obtain the optimal solution with the branch and bound method. This method requires solving a linear programming problem for each combination of the binary variables. In this problem, we have a_{bin} binary variables, a_{con} continuous variables and b constraints. Then, we need to solve $2^{a_{\text{bin}}}$ linear programming problems where the complexity order of each is $\mathcal{O}(a^2_{\text{con}} b)$ [30, p. 6]. Hence, the overall algorithm complexity is $\mathcal{O}(2^{a_{\text{bin}}} a^2_{\text{con}} b)$.

C. Heuristic Solution with Binary Integer Programming

In this subsection, we present a low-complexity heuristic algorithm to obtain a feasible solution of problem (12) by formulating a binary integer programming problem. Our heuristic algorithm is based on determining the transmit power of IoT devices according to their QoS requirements and decoding order. Then, these values are used to perform subcarrier allocation such that the maximum number of devices are paired.

From (4) and (11), we can obtain the minimum transmit power, denoted as $\hat{p}_j$, of class 2 device $j \in \mathcal{D}_2$ to meet its data rate requirement

$$\hat{p}_j = \frac{(2^{R_j/B} - 1) N_o B}{|h_j|^2}, \quad j \in \mathcal{D}_2. \quad (19)$$

If $\hat{p}_j > P_j$ for a given class 2 device j, then this device is unable to meet its QoS requirement given the maximum transmit power budget. Device j should not be considered for subcarrier allocation (i.e., $k^2_{j,s} = 0$ for all $s \in \mathcal{S}$). Note that this transmit power is allocated to a single subcarrier since we consider the single-tone mode. By using exactly the minimum transmit power to satisfy the QoS constraint, we reduce the interference caused by class 2 devices to class 1 devices. This enables us to have more class 2 devices that can be paired with class 1 devices and use NOMA. Achieving a higher data rate is not an objective for most mIoT applications. Consequently, there is no advantage of using a transmit power higher than $\hat{p}_j$ for class 2 devices.

On the other hand, we set the initial transmit power of class 1 devices, denoted as $\hat{p}_i$, to the maximum value according to (7), i.e.,

$$\hat{p}_i = P_i, \quad i \in \mathcal{D}_1. \quad (20)$$

Using the maximum transmit power, class 1 devices have a higher chance to meet their QoS requirements and can tolerate more interference from class 2 devices. We estimate the maximum tolerable interference for each device $i \in \mathcal{D}_1$, denoted as $\hat{I}_i$, when using the maximum transmit power to achieve the minimum data rate threshold R_i from (5) and (11). Hence, we have

$$\hat{I}_i = \frac{P_i|h_i|^2}{2^{R_i/B}-1} - N_o B, \quad i \in \mathcal{D}_1. \tag{21}$$

Note that a negative value of $\hat{I}_i$ for a device $i \in \mathcal{D}_1$ indicates that this device has poor channel conditions and cannot achieve its required data rate even if there is no interference from class 2 devices. Hence, it should not be considered for subcarrier allocation (i.e., $k^1_{i,s} = 0$ for all $s \in \mathcal{S}$).

Finally, we formulate a binary integer programming problem to match a class 2 device $j \in \mathcal{D}_2$ with initial transmit power $\hat{p}_j$ to a class 1 device $i \in \mathcal{D}_1$ that can tolerate an interference power of $\hat{p}_j|h_j|^2$ from this device (i.e., $\hat{p}_j|h_j|^2 \le \hat{I}_i$ when $\hat{I}_i > 0$. Note that always $\hat{p}_j > 0$ since $R_j > 0$). Hence, for a given subcarrier $s \in \mathcal{S}$,

$$\sum_{j\in\mathcal{D}_2} \hat{p}_j|h_j|^2 k^2_{j,s} \le \sum_{i\in\mathcal{D}_1} \hat{I}_i k^1_{i,s} + \omega\left(1 - \sum_{i\in\mathcal{D}_1} k^1_{i,s}\right),\ s \in \mathcal{S}, \tag{22}$$

where ω is a scaling factor of a very large value. The term $w(1 - \sum_{i\in\mathcal{D}_1} k^1_{i,s})$ has a value greater than zero only if subcarrier s is not allocated to any of the class 1 devices. When a given subcarrier is not allocated to any of the class 1 devices, there will be no upper limit on the interference that can be caused by class 2 devices (i.e., $\hat{p}_j|h_j|^2 < \infty$). ω can be set to take any value greater than the maximum interference that can be caused by class 2 devices at a given subcarrier s, i.e., ω should be greater than or equal to $\max_{j\in\mathcal{D}_2} P_j|h_j|^2$. The single-tone connection density maximization problem can be formulated as follows:

$$\begin{aligned}
&\underset{K}{\text{maximize}} && \sum_{c\in\mathcal{C}}\sum_{i\in\mathcal{D}_c}\sum_{s\in\mathcal{S}} k^c_{i,s} && \text{(23a)}\\
&\text{subject to} && k^2_{j',s} = 0,\ j' \in \{j \in \mathcal{D}_2 \mid \hat{p}_j > P_j\}, s \in \mathcal{S} && \text{(23b)}\\
& && k^1_{i',s} = 0,\ i' \in \{i \in \mathcal{D}_1 \mid \hat{I}_i < 0\}, s \in \mathcal{S} && \text{(23c)}\\
& && \text{constraints (6), (10), (12b), (22).}
\end{aligned}$$

Problem (23) is a binary integer programming problem, which can be solved by using CVX with Gurobi solver. Note that we restrict the domain of $p_{i,s}$ to be $\{0, \hat{p}_i\}$ for all IoT devices $i \in \mathcal{D}_c,\ c \in \mathcal{C}$. Algorithm 1 shows the detailed steps to solve this problem. In Steps 2 – 4, the initial values of the transmit power of class 1 and class 2 devices are calculated. Problem (23) is solved in Step 5. Step 6 is used to assign the transmit power to the allocated subcarrier for each device according to the solution of problem (23).

The complexity of evaluating a feasible solution of the binary integer programming problem with a binary variables and b constraints is $\mathcal{O}(ab)$. Hence, the computational complexity of checking all the possible binary combinations is $\mathcal{O}(2^a ab)$. Although the algorithm has an exponential complexity, in our simulations, it has a small average running time of 0.49 seconds per simulation using MATLAB/CVX with Gurobi solver in a PC with processor Intel(R) Core(TM) i7-2600K CPU@3.4 GHz.

Algorithm 1: Heuristic algorithm for joint subcarrier and power allocation in single-tone mode

1. **Input**: $\omega,\ h_i, R_i, P_i,\ i \in \mathcal{D}_c,\ c \in \mathcal{C}$
2. Determine $\hat{p}_j,\ j \in \mathcal{D}_2$ using (19)
3. Determine $\hat{p}_i,\ i \in \mathcal{D}_1$ using (20)
4. Determine $\hat{I}_i,\ i \in \mathcal{D}_1$ using (21)
5. Solve problem (23) to obtain $k^c_{i,s},\ i \in \mathcal{D}_c,\ c \in \mathcal{C}, s \in \mathcal{S}$
6. $p_{i,s} := \hat{p}_i k^c_{i,s},\ i \in \mathcal{D}_c,\ c \in \mathcal{C}, s \in \mathcal{S}$
7. **Output**: $P_{\mathcal{D}_1}, P_{\mathcal{D}_2}$, and K

IV. Multi-tone Connection Density Maximization

In this section, we formulate the connection density maximization problem when the subcarrier bandwidth B is equal to 15 kHz, i.e., $S = 12$ subcarriers, as shown in Fig. 1(b). Both single-tone and multi-tone modes can be used in this case.

A. Problem Formulation

The NB-IoT standard [6, p. 524] enforces two constraints on subcarrier allocation for the multi-tone mode. The first constraint is that each IoT device can be allocated either 0, 1, 3, 6, or 12 subcarriers. That is,

$$\sum_{s\in\mathcal{S}} k^c_{i,s} \in \{0, 1, 3, 6, 12\}, \quad \forall\, i \in \mathcal{D}_c,\ c \in \mathcal{C}. \tag{24}$$

The second constraint is that the subcarriers allocated to each IoT device should be contiguous and follow a specific combination m out of the set of all possible combinations $\mathcal{M} = \{1, 2, \ldots, M\}$, which has $M = 19$ combinations as follows:

- One subcarrier $\left(\sum_{s\in\mathcal{S}} k^c_{i,s} = 1,\ i \in \mathcal{D}_c,\ c \in \mathcal{C}\right)$. There are twelve combinations.
- Three subcarriers $\left(\sum_{s\in\mathcal{S}} k^c_{i,s} = 3,\ i \in \mathcal{D}_c,\ c \in \mathcal{C}\right)$ starting from subcarrier s to subcarrier $s + 2$ and $s \in \{1, 4, 7, 10\}$. There are four combinations, e.g., the bond of subcarriers $\{7, 8, 9\}$.
- Six subcarriers $\left(\sum_{s\in\mathcal{S}} k^c_{i,s} = 6,\ i \in \mathcal{D}_c,\ c \in \mathcal{C}\right)$ starting from subcarrier s to subcarrier $s + 5$ and $s \in \{1, 7\}$. There are two combinations, e.g., the bond of subcarriers $\{1, \ldots, 6\}$.
- The whole 12 subcarriers $(\sum_{s\in\mathcal{S}} k^c_{i,s} = 12,\ i \in \mathcal{D}_c,\ c \in \mathcal{C})$. There is one combination.

If the summation $\sum_{s\in\mathcal{S}} k^c_{i,s}$ is equal to zero, then no subcarriers are allocated to device $i \in \mathcal{D}_c,\ c \in \mathcal{C}$. Let $v^c_{i,m} = 1$ indicate if device $i \in \mathcal{D}_c,\ c \in \mathcal{C}$ is allocated subcarriers according to combination m, and $v^c_{i,m} = 0$ otherwise. Each device can be allocated subcarriers in a way that suits only one of the aforementioned combinations. Hence, we have

$$\sum_{m\in\mathcal{M}} v^c_{i,m} \le 1, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C}. \tag{25}$$

In addition, each combination can be assigned to at most one class 1 and one class 2 device

$$\sum_{i\in\mathcal{D}_c} v^c_{i,m} \le 1, \quad c \in \mathcal{C},\ m \in \mathcal{M}. \tag{26}$$

Furthermore, $v^c_{i,m}$ can be written in terms of the subcarrier allocation matrix elements $\{k^c_{i,s}\}$ to enforce the subcarrier allocation combinations of the NB-IoT standard,

$$v_{i,m}^c = k_{i,m}^c - v_{i,13}^c - v_{i,17}^c - v_{i,19}^c, \quad 1 \le m \le 3, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27a}$$

$$v_{i,m}^c = k_{i,m}^c - v_{i,14}^c - v_{i,17}^c - v_{i,19}^c, \quad 4 \le m \le 6, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27b}$$

$$v_{i,m}^c = k_{i,m}^c - v_{i,15}^c - v_{i,18}^c - v_{i,19}^c, \quad 7 \le m \le 9, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27c}$$

$$v_{i,m}^c = k_{i,m}^c - v_{i,16}^c - v_{i,18}^c - v_{i,19}^c, \quad 10 \le m \le 12, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27d}$$

$$v_{i,13}^c \le k_{i,s}^c - v_{i,17}^c - v_{i,19}^c, \quad 1 \le s \le 3, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27e}$$

$$v_{i,13}^c \ge \sum_{s=1}^{3} k_{i,s}^c - v_{i,17}^c - v_{i,19}^c - 2, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27f}$$

$$v_{i,14}^c \le k_{i,s}^c - v_{i,17}^c - v_{i,19}^c, \quad 4 \le s \le 6, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27g}$$

$$v_{i,14}^c \ge \sum_{s=4}^{6} k_{i,s}^c - v_{i,17}^c - v_{i,19}^c - 2, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27h}$$

$$v_{i,15}^c \le k_{i,s}^c - v_{i,18}^c - v_{i,19}^c, \quad 7 \le s \le 9, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27i}$$

$$v_{i,15}^c \ge \sum_{s=7}^{9} k_{i,s}^c - v_{i,18}^c - v_{i,19}^c - 2, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27j}$$

$$v_{i,16}^c \le k_{i,s}^c - v_{i,18}^c - v_{i,19}^c, \quad 10 \le s \le 12, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27k}$$

$$v_{i,16}^c \ge \sum_{s=10}^{12} k_{i,s}^c - v_{i,18}^c - v_{i,19}^c - 2, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27l}$$

$$v_{i,17}^c \le k_{i,s}^c - v_{i,19}^c, \quad 1 \le s \le 6, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27m}$$

$$v_{i,17}^c \ge \sum_{s=1}^{6} k_{i,s}^c - v_{i,19}^c - 5, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27n}$$

$$v_{i,18}^c \le k_{i,s}^c - v_{i,19}^c, \quad 7 \le s \le 12, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27o}$$

$$v_{i,18}^c \ge \sum_{s=7}^{12} k_{i,s}^c - v_{i,19}^c - 5, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27p}$$

$$v_{i,19}^c \le k_{i,s}^c, \quad 1 \le s \le 12, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C} \tag{27q}$$

$$v_{i,19}^c \ge \sum_{s=1}^{12} k_{i,s}^c - 11, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C}. \tag{27r}$$

Constraints (27a)–(27d) correspond to combinations $m = 1, \ldots, 12$, in which device i is allocated one subcarrier. For example, when device i is allocated subcarrier 2, then $v_{i,2}^c = k_{i,2}^c = 1$, and $v_{i,1}^c = v_{i,3}^c = \cdots = v_{i,19}^c = 0$. Constraints (27e)–(27l) correspond to combinations $m = 13, \ldots, 16$, in which device i is allocated three subcarriers. There are two constraints per combination to enforce contiguous subcarrier allocation according to the combination. For example, when device i is allocated subcarriers 1, 2 and 3 according to combination $m = 13$, then $k_{i,1}^c = k_{i,2}^c = k_{i,3}^c = 1$ and $v_{i,13}^c = 1$ while $v_{i,1}^c = v_{i,2}^c = v_{i,3}^c = \cdots = v_{i,12}^c = v_{i,14}^c = \cdots = v_{i,19}^c = 0$. Similarly constraints (27m)–(27r) correspond to combinations $m = 17, \ldots, 19$. Note that when device i is allocated subcarriers $1, \ldots, 6$ according to combination $m = 17$, constraints (27a)–(27b), (27e)–(27h) enforce that $v_{i,1}^c = \cdots = v_{i,6}^c = v_{i,13}^c = v_{i,14}^c = 0$. This is due to the fact that the combinations $m = 1, \ldots, 6, 13, 14$ are associated with a subset of the allocated subcarriers to device i and they cannot be assigned to other devices from the same class.

When a device is allocated a bond of subcarriers, the transmit power is divided equally over all those allocated subcarriers due to the usage of SC-FDMA. Hence, we have

$$p_{i,s} = \frac{\sum_{s' \in \mathcal{S}} p_{i,s'}}{\sum_{s' \in \mathcal{S}} k_{i,s'}^c} k_{i,s}^c, \quad i \in \mathcal{D}_c, c \in \mathcal{C}, s \in \mathcal{S}. \tag{28}$$

In multi-tone case, a device is served if it is allocated subcarriers according to one of the multi-tone NB-IoT combinations in (27). Hence, the QoS constraints in (11) can be written as:

$$r_i \ge \sum_{m \in \mathcal{M}} v_{i,m}^c R_i, \quad i \in \mathcal{D}_c, c \in \mathcal{C}. \tag{29}$$

The connection density maximization problem of the multi-tone mode can be formulated as

$$\underset{P_{\mathcal{D}_1},\, P_{\mathcal{D}_2},\, K,\, V}{\text{maximize}} \quad \sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{D}_c} \sum_{m \in \mathcal{M}} v_{i,m}^c \tag{30a}$$

$$\text{subject to} \quad k_{i,s}^c \in \{0,1\},\ \ i \in \mathcal{D}_c,\ c \in \mathcal{C},\ s \in \mathcal{S} \tag{30b}$$

$$v_{i,m}^c \in \{0,1\},\ i \in \mathcal{D}_c,\ c \in \mathcal{C},\ m \in \mathcal{M} \tag{30c}$$

$$\text{constraints (4)–(8), (25)–(29),}$$

where $V = (v_{i,m}^1, v_{j,m}^2)$, $[v_{i,m}^1] \in \mathbf{R}^{|\mathcal{D}_1| \times M}$, $[v_{j,m}^2] \in \mathbf{R}^{|\mathcal{D}_2| \times M}$, is a matrix of the binary variables $v_{i,m}^c$. The formulated problem of the multi-tone mode is also a mixed integer nonconvex problem due to the nonconvex constraints (5), (28) and (29). In order to solve this problem, we use difference of convex programming to obtain a close-to-optimal solution. In addition, we propose a low-complexity heuristic algorithm to solve problem (30) in a way similar to the heuristic algorithm of the single-tone problem.

B. Difference of Convex Programming Method

First, we handle the equal power division nonconvex constraint (28) by introducing a new variable $\tilde{p}_{i,m}$, which is the transmit power used by device $i \in \mathcal{D}_c$ when allocated the subcarriers according to combination m. Hence, we can represent the transmit power per subcarrier $p_{i,s}$ as a function of $\tilde{p}_{i,m}$ as follows:

$$p_{i,s} = \begin{cases} \tilde{p}_{i,s} + \frac{1}{3}\tilde{p}_{i,13} + \frac{1}{6}\tilde{p}_{i,17} + \frac{1}{12}\tilde{p}_{i,19}, & 1 \le s \le 3, \\ \tilde{p}_{i,s} + \frac{1}{3}\tilde{p}_{i,14} + \frac{1}{6}\tilde{p}_{i,17} + \frac{1}{12}\tilde{p}_{i,19}, & 4 \le s \le 6, \\ \tilde{p}_{i,s} + \frac{1}{3}\tilde{p}_{i,15} + \frac{1}{6}\tilde{p}_{i,18} + \frac{1}{12}\tilde{p}_{i,19}, & 7 \le s \le 9, \\ \tilde{p}_{i,s} + \frac{1}{3}\tilde{p}_{i,16} + \frac{1}{6}\tilde{p}_{i,18} + \frac{1}{12}\tilde{p}_{i,19}, & 10 \le s \le 12, \end{cases} \quad i \in \mathcal{D}_c,\ c \in \mathcal{C}. \tag{31}$$

This constraint enforces the overall transmit power used by device i, i.e., $\tilde{p}_{i,m}$, to be divided equally over the allocated subcarriers. For example, if device i is allocated three subcarriers according to combination $m = 13$, then the total transmit power is $\tilde{p}_{i,13}$ and the transmit powers of the first three subcarriers $p_{i,1}, p_{i,2}$, and $p_{i,3}$ are equal to $\frac{1}{3}\tilde{p}_{i,13}$. Note that $\tilde{p}_{i,1} = \cdots = \tilde{p}_{i,12} = \tilde{p}_{i,14} = \cdots = \tilde{p}_{i,19} = 0$ since only one combination is allowed per device. Similarly, if device i is allocated one subcarrier according to combination $m = 3$, then the total transmit power is allocated to the third subcarrier, i.e., $\tilde{p}_{i,3} = p_{i,3}$. In addition, $\tilde{p}_{i,m}$ has a nonzero value only if device i is allocated a set of subcarriers according to combination m, i.e.,

$$\tilde{p}_{i,m} \leq P_i v^c_{i,m}, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C},\ m \in \mathcal{M}. \tag{32}$$

Second, the nonconvex QoS constraint of class 1 devices (i.e., constraints (5) and (29)) can be written as the difference between two negative log expressions as follows:

$$\begin{aligned} &\sum_{s\in\mathcal{S}} -\log\left(1 + p_{i,s}g_i + \sum_{j\in\mathcal{D}_2} p_{j,s}g_j\right) \\ &-\sum_{s\in\mathcal{S}} -\log\left(1 + \sum_{j\in\mathcal{D}_2} p_{j,s}g_j\right) \leq -\log(2)\frac{R_i}{B}\sum_{m\in\mathcal{M}} v^1_{i,m}, \\ &\hspace{20em} i \in \mathcal{D}_1, \end{aligned} \tag{33}$$

where $g_i = \frac{|h_i|^2}{N_o B},\ i \in \mathcal{D}_c,\ c \in \mathcal{C}$. The difference between two convex functions of an arbitrary variable $\mathbf{z}$, $f_1(\mathbf{z}) - f_2(\mathbf{z})$, can be convexified by approximating it using the first order Taylor expansion as $f_1(\mathbf{z}) - f_2(\mathbf{z}^{(t)}) - \nabla_{\mathbf{z}} f_2(\mathbf{z}^{(t)})^T(\mathbf{z} - \mathbf{z}^{(t)})$, where $\mathbf{z}^{(t)}$ is a feasible value of the arbitrary variable $\mathbf{z}$ at iteration t. Hence, the QoS constraint in (33) can be written as follows:

$$\begin{aligned} &\sum_{s\in\mathcal{S}} -\log\left(1 + p_{i,s}g_i + \sum_{j\in\mathcal{D}_2} p_{j,s}g_j\right) \\ &-\sum_{s\in\mathcal{S}} -\log\left(1 + \sum_{j\in\mathcal{D}_2} p^{(t)}_{j,s}g_j\right) \\ &-\sum_{j\in\mathcal{D}_2}\sum_{s\in\mathcal{S}} -\frac{g_j(p_{j,s} - p^{(t)}_{j,s})}{1 + \sum_{j\in\mathcal{D}_2} p^{(t)}_{j,s}g_j} \leq -\log(2)\frac{R_i}{B}\sum_{m\in\mathcal{M}} v^1_{i,m}, \\ &\hspace{20em} i \in \mathcal{D}_1. \end{aligned} \tag{34}$$

This requires initializing $p^{(t)}_{j,s}$ with an arbitrary value for all $j \in \mathcal{D}_2,\ s \in \mathcal{S}$ and solving the problem multiple times till the value of this variable converges, i.e., $p^{(t+1)}_{j,s} - p^{(t)}_{j,s} < \epsilon$, where ϵ is the convergence tolerance threshold.

Third, we use the difference of convex programming to handle the binary variables as well. We relax all the binary variables into continuous variables such that $k^c_{i,s} \in \{0, 1\}$ satisfies the following constraints

$$0 \leq k^c_{i,s} \leq 1, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C},\ s \in \mathcal{S} \tag{35a}$$

$$k^c_{i,s} - (k^c_{i,s})^2 = 0, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C},\ s \in \mathcal{S}, \tag{35b}$$

and $v^c_{i,m} \in \{0, 1\}$ satisfies the following constraints

$$0 \leq v^c_{i,m} \leq 1, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C},\ m \in \mathcal{M} \tag{36a}$$

$$v^c_{i,m} - (v^c_{i,m})^2 = 0, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C},\ m \in \mathcal{M}. \tag{36b}$$

We include constraints (35b) and (36b) as penalty terms in the objective function as they are nonconvex. In addition, constraints (35b) and (36b) are expressed as difference of two convex functions and can be approximated using first order Taylor series approximation in order to make the objective function convex even after adding these constraints as penalty terms. Hence, we have

$$k^c_{i,s} - {k^c_{i,s}}^{(t)}{k^c_{i,s}}^{(t)} - 2{k^c_{i,s}}^{(t)}(k^c_{i,s} - {k^c_{i,s}}^{(t)}) = 0, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C},\ s \in \mathcal{S} \tag{37a}$$

$$v^c_{i,m} - {v^c_{i,m}}^{(t)}{v^c_{i,m}}^{(t)} - 2{v^c_{i,m}}^{(t)}(v^c_{i,m} - {v^c_{i,m}}^{(t)}) = 0, \quad i \in \mathcal{D}_c,\ c \in \mathcal{C},\ m \in \mathcal{M}. \tag{37b}$$

This ensures variables $k^c_{i,s}$ and $v^c_{i,m}$ to take values of 0 or 1 after a sufficient number of iterations using difference of convex programming. By applying all these changes, the multi-tone power and subcarrier allocation problem can be written as a convex problem for a certain iteration $t + 1$ as given in (38), shown at the top of the next page, and a suboptimal solution can be obtained by using successive convex programming [24], [31], where η_k and η_v are the weights of the penalty terms (37a) and (37b), respectively. $\tilde{P}_{\mathcal{D}_1} = [\tilde{p}_{i,m}] \in \mathbf{R}^{|\mathcal{D}_1| \times M}$ and $\tilde{P}_{\mathcal{D}_2} = [\tilde{p}_{j,m}] \in \mathbf{R}^{|\mathcal{D}_2| \times M}$ are matrices of the allocated transmit power per combination of class 1 and class 2 IoT devices, respectively. Problem (38) is a convex optimization problem. The values of variables ${p_{j,s}}^{(t)}$, ${k^c_{i,s}}^{(t)}$ and ${v^c_{i,m}}^{(t)}$ are initialized randomly at $t = 0$. Then, the problem is solved for multiple iterations till the convergence of the variables $p_{j,s},\ j \in \mathcal{D}_2,\ s \in \mathcal{S},\ k^c_{i,s},\ i \in \mathcal{D}_c,\ c \in \mathcal{C},\ s \in \mathcal{S}$, and $v^c_{i,m},\ i \in \mathcal{D}_c,\ c \in \mathcal{C},\ m \in \mathcal{M}$ for a given tolerance threshold ϵ. The solution of this problem may be a local optimal solution and it may be sensitive to the initialization of the aforementioned variables. Hence, the problem may need to be solved multiple times with different initializations. The complexity of solving a convex problem with a variables and b constraints using T_{int} iterations of the interior-point method is $\mathcal{O}(T_{\text{int}} \max\{a^3, a^2 b, F\})$ [30, p. 8], where F is the cost of evaluating the first and second derivatives of the objective and constraint functions. We solve the convex problem for T iterations with R random initial solutions within each iteration. Hence, the overall complexity is $\mathcal{O}(TRT_{\text{int}} \max\{a^3, a^2 b, F\})$.

C. Heuristic Solution with Binary Integer Programming

Similar to the single-tone mode, a heuristic algorithm that is based on allocating initial transmit power is proposed. Then, subcarrier allocation is obtained by solving a binary integer programming problem. The minimum transmit power of the class 2 device to meet its minimum data rate requirement depends on the number of allocated subcarriers. Hence, we express the minimum transmit power for a given combination m as

$$\hat{p}_{j,m} = \begin{cases} \frac{(2^{R_j/B} - 1)N_o B}{|h_j|^2}, & 1 \leq m \leq 12, \\ 3\frac{(2^{R_j/3B} - 1)N_o B}{|h_j|^2}, & 13 \leq m \leq 16, \\ 6\frac{(2^{R_j/6B} - 1)N_o B}{|h_j|^2}, & 17 \leq m \leq 18, \\ 12\frac{(2^{R_j/12B} - 1)N_o B}{|h_j|^2}, & m = 19, \end{cases} \quad j \in \mathcal{D}_2. \tag{39}$$

If $\hat{p}_{j,m} > P_j$ for a given class 2 device j, then this device is unable to meet its QoS requirement using this subcarrier

$$
\begin{aligned}
\underset{P_{\mathcal{D}_1},\, P_{\mathcal{D}_2},\, \hat{P}_{\mathcal{D}_1},\, \hat{P}_{\mathcal{D}_2},\, K,\, V}{\text{maximize}} \quad & \sum_{i\in\mathcal{D}_c}\sum_{c\in\mathcal{C}}\sum_{m\in\mathcal{M}} v_{i,m}^c \\
& - \eta_k \sum_{i\in\mathcal{D}_c}\sum_{c\in\mathcal{C}}\sum_{s\in\mathcal{S}} \Big(k_{i,s}^c - {k_{i,s}^c}^{(t)} {k_{i,s}^c}^{(t)} - 2{k_{i,s}^c}^{(t)} (k_{i,s}^c - {k_{i,s}^c}^{(t)}) \Big) \\
& - \eta_v \sum_{i\in\mathcal{D}_c}\sum_{c\in\mathcal{C}}\sum_{m\in\mathcal{M}} \Big(v_{i,m}^c - {v_{i,m}^c}^{(t)} {v_{i,m}^c}^{(t)} - 2{v_{i,m}^c}^{(t)} (v_{i,m}^c - {v_{i,m}^c}^{(t)}) \Big) \qquad (38) \\
\text{subject to} \quad & \text{constraints (4), (6)–(8), (25)–(27), (29), (31)–(32), (34), (35a), (36a),}
\end{aligned}
$$

allocation combination even if it uses the maximum transmit power. Hence, such subcarrier allocation combinations can be excluded for this device. Note that this transmit power is calculated based on equal power division over allocated subcarriers. On the other hand, we assume that the initial transmit power of the class 1 devices at a given combination $\hat{p}_{i,m}$ takes the maximum value according to (7), i.e.,

$$\hat{p}_{i,m} = P_i, \quad i \in \mathcal{D}_1,\ m \in \mathcal{M}. \qquad (40)$$

We estimate the maximum tolerable interference for each device $i \in \mathcal{D}_1$ at a certain subcarrier allocation combination m, denoted as $\hat{I}_{i,m}$. It can be expressed as

$$\hat{I}_{i,m} = \begin{cases} \frac{\hat{p}_{i,m}|h_i|^2}{2^{R_i/B}-1} - N_oB, & 1 \le m \le 12, \\ \frac{\hat{p}_{i,m}|h_i|^2}{2^{R_i/3B}-1} - N_oB, & 13 \le m \le 16, \\ \frac{\hat{p}_{i,m}|h_i|^2}{2^{R_i/6B}-1} - N_oB, & 17 \le m \le 18, \\ \frac{\hat{p}_{i,m}|h_i|^2}{2^{R_i/12B}-1} - N_oB, & m = 19, \end{cases} \quad i \in \mathcal{D}_1. \qquad (41)$$

Note that a negative value of $\hat{I}_{i,m}$ for a given class 1 device $i \in \mathcal{D}_1$ indicates that this device cannot achieve its required data rate even if class 2 devices do not cause any interference to it when it is allocated subcarriers according to combination m.

Similar to the single-tone case, we formulate a binary integer programming problem to match one or a group of class 2 devices $j \in \mathcal{D}_2$ to one or more class 1 devices $i \in \mathcal{D}_1$ such that the maximum tolerable interference power at a given subcarrier s is not exceeded by the received power from class 2 devices at the same subcarrier. That is, for a given subcarrier $s \in \mathcal{S}$,

$$\sum_{m\in\mathcal{M}_s}\sum_{j\in\mathcal{D}_2} \frac{1}{n_m}\hat{p}_{j,m}|h_j|^2 v_{j,m}^2 \le \sum_{m\in\mathcal{M}_s}\sum_{i\in\mathcal{D}_1} \frac{1}{n_m}\hat{I}_{i,m} v_{i,m}^1 + \omega\Bigg(1 - \sum_{m\in\mathcal{M}_s}\sum_{i\in\mathcal{D}_1} v_{i,m}^1\Bigg), \quad s \in \mathcal{S}, \qquad (42)$$

where $\mathcal{M}_s$ is the set of combinations that include subcarrier s. For example, when $s = 1$, then $\mathcal{M}_s = \{1, 13, 17, 19\}$. n_m is the number of subcarriers allocated by following combination m, e.g., when $m = 13$, then $n_m = 3$. ω is a scaling factor of a very large value. The term $w(1 - \sum_{m\in\mathcal{M}_s}\sum_{i\in\mathcal{D}_1} v_{i,m}^1)$ has a value greater than zero only if subcarrier s is not allocated to any of the class 1 devices. Hence, there will be no upper limit on the interference that can be caused by class 2 devices (i.e., $\sum_{m\in\mathcal{M}_s}\sum_{j\in\mathcal{D}_2} \frac{1}{n_m}\hat{p}_{j,m}|h_j|^2 v_{j,m}^2 < \infty$). ω can be set to take any value greater than the maximum interference that can be caused by class 2 devices at a given subcarrier s, i.e., ω should be greater than or equal to $\max_{j\in\mathcal{D}_2} P_j|h_j|^2$.

Some combinations cannot be assigned to different devices from a given class simultaneously. For example, if a class 1 (class 2) device is allocated subcarriers according to combination $m = 13$, then the other class 1 (class 2) devices cannot be allocated subcarriers according to combinations $m = 1, 2, 3, 17,$ or 19. Hence, we have

$$\sum_{m\in\mathcal{M}_s}\sum_{i\in\mathcal{D}_c} v_{i,m}^c \le 1, \quad s \in \mathcal{S},\ c \in \mathcal{C}. \qquad (43)$$

The multi-tone connection density maximization problem can be formulated as follows:

$$
\begin{aligned}
\underset{V}{\text{maximize}} \quad & \sum_{c\in\mathcal{C}}\sum_{i\in\mathcal{D}_c}\sum_{m\in\mathcal{M}} v_{i,m}^c && (44a) \\
\text{subject to} \quad & v_{j',m'}^2 = 0, \\
& (j', m') \in \{(j,m) \in \mathcal{D}_2 \times \mathcal{M} \mid \hat{p}_{j,m} > P_j\} && (44b) \\
& v_{i',m'}^1 = 0, \\
& (i', m') \in \{(i,m) \in \mathcal{D}_1 \times \mathcal{M} \mid \hat{I}_{i,m} < 0\} && (44c) \\
& \text{constraints (25)–(26), (30b)–(30c), (42)–(43).}
\end{aligned}
$$

Problem (44) is a binary integer programming problem, which can be solved using CVX with Gurobi solver. Algorithm 2 shows the detailed steps to solve the problem. In Steps 2 – 4, we calculate the initial values of power allocation for the class 1 and class 2 devices. The problem is solved in Step 5. Steps 6 – 8 are used to assign the transmit power to the allocated subcarriers for each device according to the solution of problem (44). Similar to the single-tone mode, the computational complexity of Algorithm 2 is $\mathcal{O}(2^a ab)$. Despite the exponential complexity, it has a small average running time of 0.82 seconds per simulation using MATLAB/CVX with Gurobi solver in a PC with processor Intel(R) Core(TM) i7-2600K CPU@3.4 GHz.

V. Performance Evaluation

In this section, we present the simulation results of the proposed algorithms. We consider a single cell where class 1 and class 2 devices are uniformly distributed within a 1 km^2 square region. We consider flat Rayleigh fading channels since the total system bandwidth is as narrow as 180 kHz. The distance-dependent path loss $PL(D)$ at 900 MHz carrier frequency is calculated by $PL(D) = 120.9 + 37.6\log(D/1000) + LI + AG$

Algorithm 2: Heuristic algorithm for joint subcarrier and power allocation in multi-tone mode

Input: $\omega,\ h_i,\ R_i, P_i,\ i \in \mathcal{D}_c,\ c \in \mathcal{C},\ n_m, \mathcal{S}_m,\ m \in \mathcal{M}, \mathcal{M}_s,\ s \in \mathcal{S}$
Determine $\hat{p}_{j,m}, j \in \mathcal{D}_2, m \in \mathcal{M}$ using (39)
Determine $\hat{p}_{i,m}, i \in \mathcal{D}_1, m \in \mathcal{M}$ using (40)
Determine $\hat{I}_{i,m}, i \in \mathcal{D}_1, m \in \mathcal{M}$ using (41)
Solve problem (44) to obtain $v^c_{i,m}$
$\tilde{p}_{i,m} := \hat{p}_{i,m} v^1_{i,m},\ i \in \mathcal{D}_1,\ m \in \mathcal{M}$
$\tilde{p}_{j,m} := \hat{p}_{j,m} v^2_{j,m},\ j \in \mathcal{D}_2,\ m \in \mathcal{M}$
$p_{i,s} := \sum_{m \in \mathcal{M}_s} (\tilde{p}_{i,m} v^c_{i,m})/n_m,\ i \in \mathcal{D}_c,\ c \in \mathcal{C},\ s \in \mathcal{S}$
$k^c_{i,s} := \sum_{m \in \mathcal{M}_s} v^c_{i,m},\ i \in \mathcal{D}_c,\ c \in \mathcal{C},\ s \in \mathcal{S}$
Output: $P_{\mathcal{D}_1}, P_{\mathcal{D}_2}, \tilde{P}_{\mathcal{D}_1}, \tilde{P}_{\mathcal{D}_2}, K,$ and V

[32, p. 481]. D is the distance between an IoT device and the base station in meters, AG is the transmit antenna gain of -4 dB, and LI is the indoor penetration loss that is assumed to be 20 dB for 80% of IoT devices (indoor devices) and 0 dB for the remaining 20% IoT devices (outdoor devices). We consider additive white Gaussian noise with power spectral density -174 dBm/Hz and noise figure of 5 dB. The maximum transmit power, P_i for all $i \in \mathcal{D}_c,\ \ c \in \mathcal{C}$ is 23 dBm [32, p. 481]. Note that class 1 devices set $\mathcal{D}_1$ represents the set of devices with high data rate requirements and class 2 devices set $\mathcal{D}_2$ represents the set of devices with low data rate requirements. The minimum data rate requirement of class 1 devices is denoted as R_h and the minimum data rate requirement of class 2 devices is denoted as R_l, where $R_h > R_l$.

To evaluate the performance, we consider the connection density metric, which is defined as the number of IoT devices satisfying their QoS requirements. The connection density is calculated as $\sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{D}_c} \sum_{s \in \mathcal{S}} k^c_{i,s}$ in the single-tone mode and $\sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{D}_c} \sum_{m \in \mathcal{M}} v^c_{i,m}$ in the multi-tone mode. We consider the supported connection density in a single time slot. With the expected large number of devices in cellular networks, supporting a higher connection density in a single time slot results in supporting simultaneous data transmission of a larger number of devices over a longer period of time. In addition, if we consider serving a fixed number of devices, then simultaneous data transmission with NOMA can enable a group of devices to send their payloads in a shorter period of time and utilize the radio resources more efficiently. Furthermore, with a massive number of devices per km^2, the number of devices sending data in a single time slot should be sufficiently greater than the number of available subcarriers in NB-IoT to exploit NOMA.

We compare the performance of our proposed NOMA scheme with that of OMA in both single-tone and multi-tone modes. In OMA, each subcarrier can only be allocated to a single device $d \in \mathcal{D}_1 \cup \mathcal{D}_2$ as long as this device can use the allocated subcarrier(s) to meet its minimum data rate requirement. In addition, we compare the performance of our proposed NOMA scheme with other NOMA schemes from the literature.

A. Single-tone Mode Performance Evaluation

In the single-tone mode, we consider a network where $S = 48$ subcarriers are available to serve 48 high data rate devices and 48 low data rate devices. Single-tone mode enables devices to achieve an uplink data rate up to 20 kbps [33]. Fig. 3 shows the achievable connection density in the single-tone mode for a single time slot with varying QoS requirements of class 1 and class 2 devices. Results show that NOMA can support much more IoT devices than OMA in NB-IoT systems. NOMA can help satisfy the QoS requirements of up to 90 IoT devices in a single time slot. This results in a connection density gain that ranges between 73% and 87% compared to OMA, which can only support a maximum of 48 IoT devices (one device per subcarrier). To reflect the aforementioned connection density gains on a scenario with a larger number of devices, if NB-IoT can support 52,500 devices sending small payload in an area of 0.86 km^2 with OMA [32, p. 483], then NB-IoT can generally support a number of devices that ranges between 90,825 and 98,175 with NOMA.

The supported connection density decreases as the data rate requirements increase. For class 1 devices, more devices require a higher received SINR for successful decoding at the base station and may not be able to tolerate interference from class 2 devices. Hence, those devices require accessing their subcarriers using OMA to meet their minimum data rate requirements and cannot be paired with any of the class 2 devices using NOMA. This reduces the supported connection density. We also compare the proposed optimal and heuristic pairing schemes with a conventional NOMA scheme that pairs the closest (second closest, . . . , furthest) class 1 device to the base station with the furthest (second furthest, . . . , closest) class 2 device from the base station, which we denote as near-far pairing (NFP). Our proposed optimal and heuristic pairing schemes outperform the NFP scheme.

In addition, our proposed optimal and heuristic schemes achieve almost the same connection density. The heuristic scheme is based on reducing the number of feasible solutions of the connection density maximization problem. For example, the set of the feasible values of the transmit powers of IoT devices is reduced from $[0, P_i]$ to $\{0, P_i\}$ for class 1 devices and $\{0, \tilde{p}_j\}$ (obtained using (19)) for class 2 devices. However, this reduction does not affect the value of the objective function. Class 1 device needs to maximize its SINR to meet its minimum data rate requirement. This goal is achieved by increasing its transmit power (e.g., using the maximum transmit power) and suffering from minimal interference from a class 2 device according to (5). On the other hand, class 2 device causes minimum interference when it uses the minimum transmit power that enables it to meet its minimum data rate requirement according to (4) and (19). With the aforementioned conditions, the chance that these two devices can access a single subcarrier with NOMA is maximized which maximizes the overall connection density as a result.

Fig. 4 shows the impact of the transmit power budget $P_i,\ i \in \mathcal{D}_c,\ c \in \mathcal{C}$ on the supported connection density in a single time slot. By reducing the transmit power of the devices, the energy consumption of IoT devices is reduced. However, a lower transmit power budget makes it less flexible to control the transmit power of class 1 and class 2 devices. Hence, a sufficient difference of received power at the base station to employ NOMA cannot be easily guaranteed. Then,

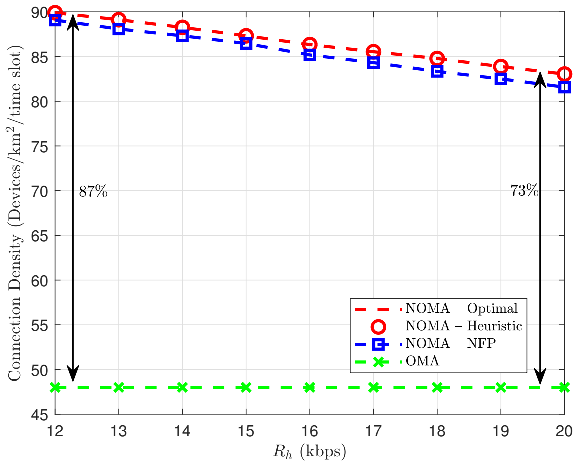


Fig. 3. Connection density in the single-tone mode versus the data rate requirement of class 1 devices R_h. R_l is set to 6 kbps.

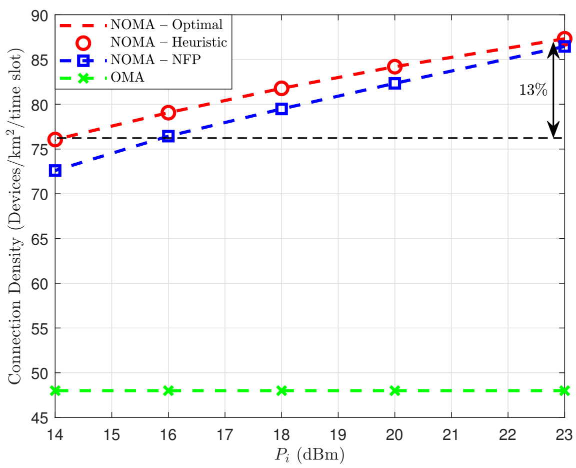


Fig. 4. Connection density in the single-tone mode versus the transmit power budget of all devices P_i. R_h and R_l are set to 15 kbps and 6 kbps, respectively.

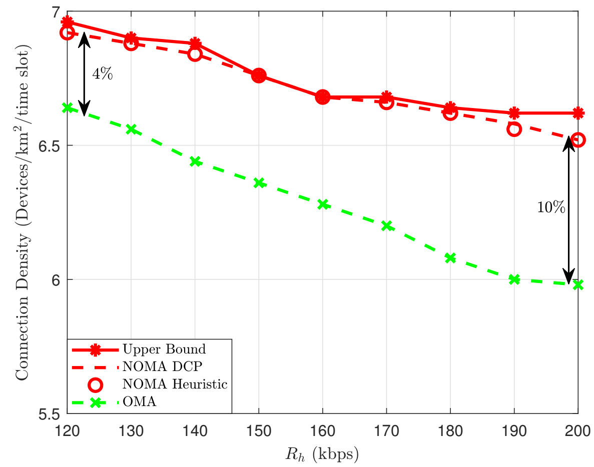


Fig. 5. Connection density in the multi-tone mode versus the data rate requirement of class 1 devices R_h. R_l is set to 50 kbps. $|\mathcal{D}_1| = |\mathcal{D}_2| = 4$ devices.

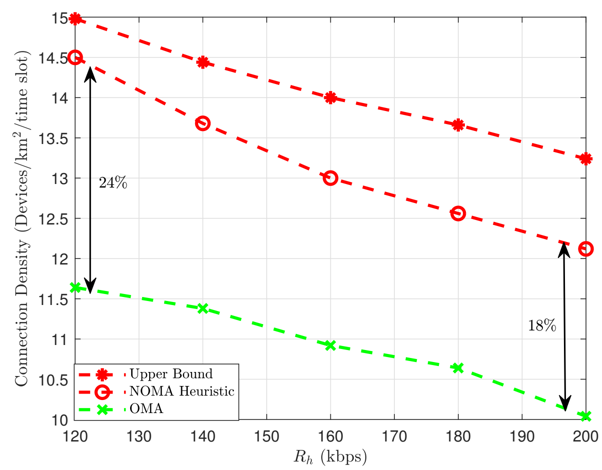


Fig. 6. Connection density in the multi-tone mode versus the data rate requirement of class 1 devices R_h. R_l is set to 50 kbps. $|\mathcal{D}_1| = |\mathcal{D}_2| = 10$ devices.

it will be less probable to pair class 1 and class 2 devices and the connection density decreases. For example, the decrease of the transmit power from 23 dBm to 14 dBm results in reducing the connection density by 13% as shown in Fig. 4.

NOMA requires class 1 devices to use a higher transmit power than OMA in order to mitigate the interference from class 2 devices. However, NOMA reduces the average queueing delay and transmission delay of IoT data packets. Hence, the IoT devices can switch to the sleep mode earlier and prolong their average battery lifetime. This results in reducing the long-term power consumption of the network.

B. Multi-tone Mode Performance Evaluation

In multi-tone mode, we consider the allocation of $S = 12$ subcarriers to high data rate and low data rate IoT devices. Multi-tone mode enables NB-IoT to support relatively higher data rates (up to 250 kbps [33]) for IoT devices compared to the single-tone mode by allocating a bond of subcarriers to these devices but with much less connection density. We consider high data rate requirements that may require the IoT devices to access multiple subcarriers based on the channel conditions. We evaluate the connection density enhancement due to the use of NOMA in the existence of few devices with high data rate requirements. For the difference of convex programming (DCP) method, we solve problem (38) with 200 random initializations for up to 30 iterations each. The tolerance threshold ϵ is set to be 0.001. The penalty terms η_k and η_v are set to $|\mathcal{D}_1| + |\mathcal{D}_2| + 1$.

Figs. 5 and 6 show the connection density of multi-tone NB-IoT with different QoS requirements of a varying number of class 1 and class 2 devices in a single time slot. DCP and the heuristic algorithm almost result in the same connection density in Fig. 5, which is very close to the upper bound. The upper bound is an infeasible solution obtained by assuming that class 1 and class 2 devices have separate sets of subcarriers and do not interfere with each other. With NOMA, NB-IoT can support a higher connection density than OMA by up to 24% when $R_l = 50$ kbps.

C. Distance-based Class Assignment

In this subsection, we assign devices to different classes based on their distances from the base station rather than their QoS requirements. The closest 50% of the IoT devices (i.e., near devices) are assigned to class 1 and the furthest 50% (i.e.,

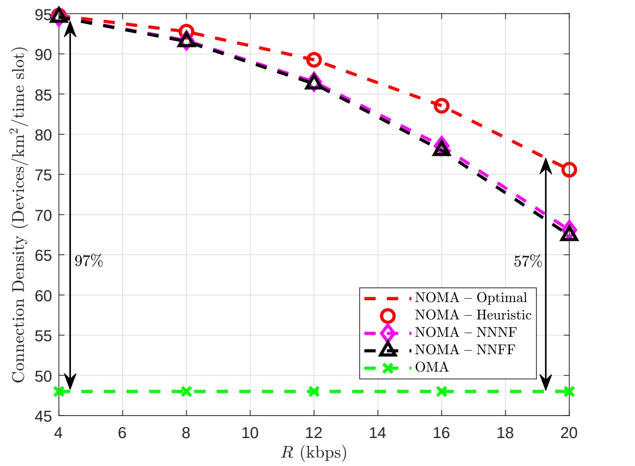


Fig. 7. Connection density in the single-tone mode versus the data rate requirement of all devices R. P_i is set to 23 dBm, $i \in \mathcal{D}_c,\ c \in \mathcal{C}$.

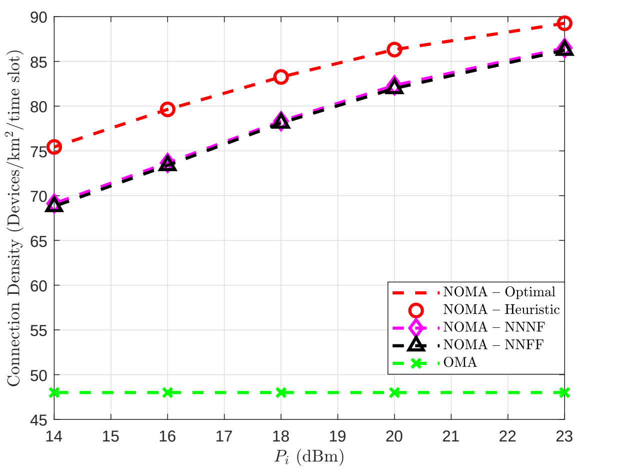


Fig. 8. Connection density in the single-tone mode versus the transmit power budget of all devices P_i. R_i is set to 12 kbps, $i \in \mathcal{D}_c,\ c \in \mathcal{C}$.

far devices) are assigned to class 2. The data from the near devices are decoded first. We solve problems (18) and (23) in the single-tone mode to allocate subcarriers and transmit power. We compare our obtained solutions with the solutions provided by nearest-near-nearest-far (NNNF) and nearest-near-furthest-far (NNFF) algorithms [34]. In NNNF, the closest (second closest, . . . , furthest) near device is paired with the closest (second closest, . . . , furthest) far device. In NNFF, the closest (second closest, . . . , furthest) near device is paired with the furthest (second furthest, . . . , closest) far device.

In the simulations, we consider 96 IoT devices with the same minimum data rate requirement R in the single-tone case. Figs. 7 and 8 show the supported connection density with different QoS requirements and maximum transmit power constraints, respectively. Distance-based device pairing achieves a connection density gain that ranges between 57% and 97% compared to OMA with different data rate requirements. The proposed device pairing scheme achieves a higher connection density than both NNNF and NNFF. When the data rate requirements increase or the transmit power budgets decrease, the connection density gap between the optimal pairing and both the NNNF and NNFF schemes becomes larger (i.e., a higher connection density gain compared to NNNF and

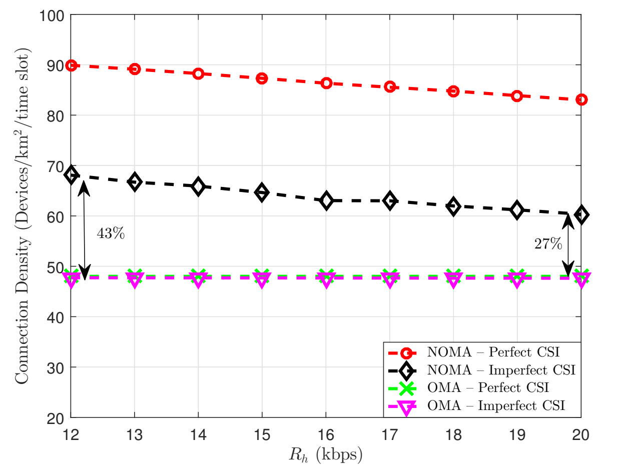


Fig. 9. Connection density in the single-tone mode versus the data rate requirement of class 1 devices R_h while considering both perfect and imperfect CSI. R_l is set to 6 kbps and σ_e^2 is set to 0.01. $|\mathcal{D}_1| = |\mathcal{D}_2| = 48$ devices.

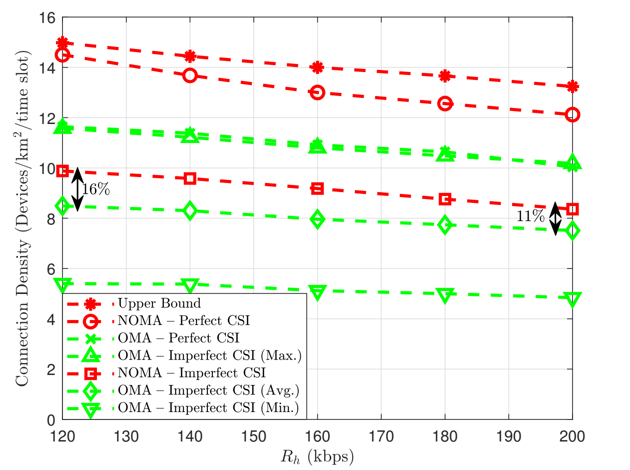


Fig. 10. Connection density in the multi-tone mode versus the data rate requirement of class 1 devices R_h while considering both perfect and imperfect CSI. R_l is set to 50 kbps and σ_e^2 is set to 0.01. $|\mathcal{D}_1| = |\mathcal{D}_2| = 10$ devices.

NNFF is achieved). This indicates that optimal device pairing becomes more essential to serve more devices in case of decreasing transmit power or increasing data rate demand, i.e., when power and subcarrier resources are limited with respect to the QoS demands.

D. Impact of Imperfect CSI

In this subsection, we investigate the impact of imperfect CSI on the connection density supported by the proposed NOMA scheme in both single-tone and multi-tone modes. We use the imperfect CSI model presented in [35], where $h_i = \hat{h}_i + e_i$. $\hat{h}_i$ denotes the imperfect channel gain estimate of h_i for a given IoT device $i \in \mathcal{D}_c, c \in \mathcal{C}$, and e_i is the complex Gaussian channel estimation error of device i with zero mean and variance σ_e^2.

In both single-tone and multi-tone modes, we consider the same network setup in Subsections V.A and V.B, respectively. We set σ_e^2 to 0.01 throughout this subsection. Simulation results in Fig. 9 show the connection density in the single-tone mode for both perfect and imperfect CSI cases. Imperfect

CSI causes the base station to make inaccurate subcarrier and power allocation decisions. Hence, some IoT devices may not be able to meet their minimum data rate requirements in spite of the allocated subcarrier and transmit power resources, which reduces the number of IoT devices supported by the proposed NOMA scheme. In case of imperfect CSI, a lower connection density gain is obtained, and it ranges between 27% and 43%. The connection density of the multi-tone mode is shown in Fig. 10. To compare with OMA in case of imperfect CSI, we consider three OMA schemes, where IoT devices meet their minimum data rate requirements by using different values of transmit power of $\frac{\left(2^{R_j/B}-1\right)N_oB}{|\hat{h}_j|^2}$ (denoted as OMA – Imperfect CSI (Min.)), P_i (denoted as OMA – Imperfect CSI (Max.)) and $0.5\left(\frac{\left(2^{R_j/B}-1\right)N_oB}{|\hat{h}_j|^2}+P_i\right)$ (denoted as OMA – Imperfect CSI (Avg.)). NOMA can still outperform OMA when the IoT devices transmit data with the minimum or the average transmit power. In the latter case, a connection density gain up to 16% can be achieved. However, OMA can be more robust against imperfect CSI by allowing devices to transmit data with maximum transmit power. Note that a channel estimation error in OMA may cause one device not be able to meet its data rate requirement. On the other hand, a channel estimation error in NOMA may cause a pair of devices fail to achieve the required data rate.

VI. Conclusion

In this paper, we employed the concept of NOMA to enhance the connection density in NB-IoT systems to support the mIoT use case. Each subcarrier is accessed by two devices with diverse QoS requirements. We formulated joint subcarrier and transmit power allocation problems for both the single-tone and multi-tone modes, and both problems were found to be mixed integer nonconvex problems. The single-tone problem was transformed into an equivalent mixed integer linear programming problem and the optimal solution was obtained. The multi-tone problem was solved using difference of convex programming to obtain a suboptimal solution by relaxing the binary variables and expressing the nonconvex constraint as a difference of two convex functions. In addition, we proposed heuristic algorithms with low complexity for both the single-tone and multi-tone modes and they resulted in close performance to the optimal and suboptimal solutions in both cases, respectively. Simulation results showed that using NOMA increased the connection density by up to 87% compared to OMA in the single-tone mode. In multi-tone mode, connection density was also increased by up to 24%.

For future work, we will use uplink NOMA in grant-free multiple access systems to support the IoT. Grant-free multiple access enables data transmission without granting uplink radio resources by the base station. Hence, the IoT devices should decide their transmit power levels and the subcarriers to target in order to maximize the probability of successful data transmission.

Appendix

Based on (5), the minimum data rate requirement constraint (11) for device $i \in \mathcal{D}_1$ in the single-tone mode can be written as

$$\log_2\left(1+\frac{|h_i|^2 p_{i,1}}{\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,1}+N_oB}\right)+\cdots + \log_2\left(1+\frac{|h_i|^2 p_{i,S}}{\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,S}+N_oB}\right) \geq \frac{R_i}{B}\sum_{s\in\mathcal{S}} k^1_{i,s}, \quad i \in \mathcal{D}_1. \tag{45}$$

Given constraints (8) and (10), if $p_{i,s}>0$, then $p_{i,s'}=0$ for all $s' \neq s$. This implies that if $\log_2\left(1+\frac{|h_i|^2 p_{i,s}}{\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,s}+N_oB}\right)>0$, then $\log_2\left(1+\frac{|h_i|^2 p_{i,s'}}{\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,s'}+N_oB}\right)=0$, i.e., there can be only one log expression in (45) that has a nonzero value. Let $a=\frac{|h_i|^2 p_{i,s}}{\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,s}+N_oB}$ and $b=\frac{|h_i|^2 p_{i,s'}}{\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,s'}+N_oB}$. If $b=0$, then $\log(1+a)+\log(1+b)$ can be equivalently written as $\log(1+a+b)$. Thus, based on (8) and (10), constraint (45) can be expressed as

$$\log_2\left(1+\frac{|h_i|^2 p_{i,1}}{\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,1}+N_oB}+\cdots + \frac{|h_i|^2 p_{i,S}}{\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,S}+N_oB}\right) \geq \frac{R_i}{B}\sum_{s\in\mathcal{S}} k^1_{i,s}, \ i \in \mathcal{D}_1, \tag{46}$$

which can be simplified as

$$\frac{|h_i|^2 p_{i,1}}{\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,1}+N_oB}+\cdots + \frac{|h_i|^2 p_{i,S}}{\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,S}+N_oB} \geq 2^{\frac{R_i}{B}\sum_{s\in\mathcal{S}} k^1_{i,s}}-1, \ i \in \mathcal{D}_1. \tag{47}$$

Constraint (47) can be written equivalently as

$$\frac{|h_i|^2 p_{i,1}}{k^1_{i,1}\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,1}+N_oB}+\cdots + \frac{|h_i|^2 p_{i,S}}{k^1_{i,S}\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,S}+N_oB} \geq 2^{\frac{R_i}{B}\sum_{s\in\mathcal{S}} k^1_{i,s}}-1, \ i \in \mathcal{D}_1. \tag{48}$$

The equivalence between (47) and (48) is due to the fact that given inequality (8), if a device is not allocated subcarrier s, it will not assign any power to it and will not suffer interference at this subcarrier. Thus, there can be only one nonzero fraction expression. Let $a=\frac{|h_i|^2 p_{i,s}}{k^1_{i,s}\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,s}+N_oB}$ and $b=\frac{|h_i|^2 p_{i,s'}}{k^1_{i,s'}\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,s'}+N_oB}$, where $s \neq s'$. When $p_{i,s}>0$ and $k^1_{i,s}>0$ (i.e., $a>0$), then both $p_{i,s'}$ and $k^c_{i,s'}$ are equal to zero (i.e., $b=0$). Then $a+b$ can be expressed as $\frac{|h_i|^2 p_{i,s}+|h_i|^2 p_{i,s'}}{k^1_{i,s}\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,s}+k^1_{i,s'}\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,s'}+N_oB}$. Hence, constraint (48) becomes

$$\frac{|h_i|^2 p_{i,1}+\cdots+|h_i|^2 p_{i,S}}{N_oB+k^1_{i,1}\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,1}+\cdots+k^1_{i,S}\sum_{j\in\mathcal{D}_2}|h_j|^2 p_{j,S}} \geq 2^{\frac{R_i}{B}\sum_{s\in\mathcal{S}} k^1_{i,s}}-1, \quad i \in \mathcal{D}_1, \tag{49}$$

which can further be expressed as given by (50) which is shown at the top of the next page.

Note that (50) is equivalent to (48) in single-tone mode.

$$|h_i|^2 \sum_{s\in\mathcal{S}} p_{i,s} \geq \left(2^{\frac{R_i}{B}\sum_{s\in\mathcal{S}} k_{i,s}^1} - 1\right)\left(N_o B + \sum_{s\in\mathcal{S}}\left(k_{i,s}^1 \sum_{j\in\mathcal{D}_2} |h_j|^2 p_{j,s}\right)\right), \quad i \in \mathcal{D}_1. \tag{50}$$

$$|h_i|^2 \sum_{s\in\mathcal{S}} p_{i,s} \geq \left(2^{\frac{R_i}{B}} - 1\right)\left(N_o B + \sum_{s\in\mathcal{S}} k_{i,s}^1 \sum_{j\in\mathcal{D}_2} |h_j|^2 p_{j,s}\right) - \left(1 - \sum_{s\in\mathcal{S}} k_{i,s}^1\right)\left(2^{\frac{R_i}{B}} - 1\right) N_o B, \quad i \in \mathcal{D}_1. \tag{51}$$

If device $i \in \mathcal{D}_1$ is allocated a given subcarrier s, then $p_{i,s} > 0$, and $p_{i,s'} = 0$ for all $s' \neq s$. In addition, $k_{i,s}^1 \sum_{j\in\mathcal{D}_2} |h_j|^2 p_{j,s} \geq 0$ and $k_{i,s'}^1 \sum_{j\in\mathcal{D}_2} |h_j|^2 p_{j,s'} = 0$ for all subcarriers $s' \neq s$. An equivalent affine constraint of (50) is given by (51). Note that the summation $\sum_{s\in\mathcal{S}} k_{i,s}^1$ is equal to either zero or one. If $\sum_{s\in\mathcal{S}} k_{i,s}^1$ is equal to 1 (i.e., device i is allocated one subcarrier), then $(1 - \sum_{s\in\mathcal{S}} k_{i,s}^1) = 0$ and the second term of the right hand side of (51) is equal to zero. In this case, device i is required to meet its minimum data rate requirement. On the other hand, if $\sum_{s\in\mathcal{S}} k_{i,s}^1$ is equal to 0 (i.e., device i is not allocated any subcarrier), then $(1 - \sum_{s\in\mathcal{S}} k_{i,s}^1) = 1$, $\sum_{s\in\mathcal{S}} p_{i,s} = 0$ and $\sum_{s\in\mathcal{S}} k_{i,s}^1 \sum_{j\in\mathcal{D}_2} |h_j|^2 p_{j,s} = 0$. Hence, the two sides of (51) are equal to zero which satisfies the constraint. This is equivalent to the case in (50) when $\sum_{s\in\mathcal{S}} k_{i,s}^1$ is assumed to take a value of zero or one. This makes the constraints (50) and (51) to be equivalent.

References

[1] F. Ghavimi and H. H. Chen, "M2M communications in 3GPP LTE/LTE-A networks: Architectures, service requirements, challenges, and applications," *IEEE Commun. Surveys & Tuts.*, vol. 17, no. 2, pp. 525–549, Oct. 2014.

[2] ITU-R M.2083, "IMT vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," Sep. 2015.

[3] ITU, "Minimum requirements related to technical performance for IMT-2020 radio interface(s)," Feb. 2017.

[4] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2017 – 2022," Feb. 2019.

[5] Next Generation Mobile Networks (NGMN) Alliance, "NGMN 5G White Paper," Feb. 2015.

[6] 3GPP TS 36.213 V15.5.0, "Evolved universal terrestrial radio access (E-UTRA): Physical layer procedures (Release 15)," Mar. 2019.

[7] Y. P. E. Wang, X. Lin, A. Adhikary, A. Grovlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi, "A primer on 3GPP narrowband Internet of Things," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 117–123, Mar. 2017.

[8] 3GPP TS 36.211 V15.5.0, "Evolved universal terrestrial radio access (E-UTRA): Physical channels and modulation (Release 15)," Mar. 2019.

[9] 3GPP RP-171428, "Further NB-IoT enhancements," Jun. 2017.

[10] 3GPP RP-181451, "Rel-16 enhancements for NB-IoT," Jun. 2018.

[11] R. Harwahyu, R. G. Cheng, C. H. Wei, and R. F. Sari, "Optimization of random access channel in NB-IoT," *IEEE Internet of Things J.*, vol. 5, no. 1, pp. 391–402, Feb. 2018.

[12] N. Jiang, Y. Deng, M. Condoluci, W. Guo, A. Nallanathan, and M. Dohler, "RACH preamble repetition in NB-IoT network," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1244–1247, Jan. 2018.

[13] X. Yang, X. Wang, Y. Wu, L. Qian, W. Lu, and H. Zhou, "Small-cell assisted secure traffic offloading for narrow-band Internet of Things (NB-IoT) systems," *IEEE Internet of Things J.*, vol. 5, no. 3, pp. 1516–1526, Jun. 2018.

[14] T. Xu and I. Darwazeh, "Non-orthogonal narrowband Internet of Things: A design for saving bandwidth and doubling the number of connected devices," *IEEE Internet of Things J.*, vol. 5, no. 3, pp. 2120–2129, Jun. 2018.

[15] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[16] Y. Zhou, V. W. S. Wong, and R. Schober, "Stable throughput regions of opportunistic NOMA and cooperative NOMA with full-duplex relaying," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5059–5075, May 2018.

[17] ——, "Dynamic decode-and-forward based cooperative NOMA with spatially random users," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3340–3356, Aug. 2018.

[18] B. Chen, Y. Chen, Y. Chen, Y. Cao, N. Zhao, and Z. Ding, "A novel spectrum sharing scheme assisted by secondary NOMA relay," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 732–735, Oct. 2018.

[19] H. Lin, F. Gao, S. Jin, and G. Y. Li, "A new view of multi-user hybrid massive MIMO: Non-orthogonal angle division multiple access," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2268–2280, Oct. 2017.

[20] 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.

[21] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, Nov. 2016.

[22] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, "Uplink non-orthogonal multiple access for 5G wireless networks," in *Proc. of IEEE Int. Symp. Wireless Communications Systems (ISWCS)*, Barcelona, Spain, Aug. 2014.

[23] R. Ruby, S. Zhong, H. Yang, and K. Wu, "Enhanced uplink resource allocation in non-orthogonal multiple access systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1432–1444, Mar. 2018.

[24] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[25] M. A. Sedaghat and R. R. Muller, "On user pairing in uplink NOMA," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3474–3486, May 2018.

[26] M. Shirvanimoghaddam, M. Condoluci, M. Dohler, and S. J. Johnson, "On the fundamental limits of random non-orthogonal multiple access in cellular massive IoT," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2238–2252, Oct. 2017.

[27] A. E. Mostafa, Y. Zhou, and V. W. S. Wong, "Connectivity maximization for narrowband IoT systems with NOMA," in *Proc. of IEEE Int. Conf. Communications (ICC)*, Paris, France, May 2017.

[28] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. of IEEE Int. Symp. Personal, Indoor, and Mobile Radio Communications (PIMRC)*, London, UK, Sep. 2013.

[29] J. Lee and S. Leyffer, *Mixed Integer Nonlinear Programming*. Springer, 2012.

[30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[31] Q. T. Dinh and M. Diehl, "Local convergence of sequential convex programming for nonconvex optimization," *Recent Advances in Optimization and Its Applications in Engineering*, pp. 93–102, 2010.

[32] 3GPP TR 45.820 V13.1.0, "Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)," Nov. 2015.

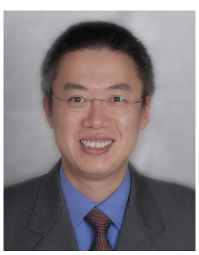
[33] Keysight Technologies, "NB-IoT and LTE Cat-M1 field measurements and SLA verification," Jan. 2018.

[34] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[35] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the performance of non-orthogonal multiple access systems with partial channel information," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 654–667, Feb. 2016.

Ahmed Elhamy Mostafa (S'15) received the B.Sc. degree (with highest honors) in electronics engineering and the M.Sc. degree (with highest honors) in electronics and communications engineering from the American University in Cairo (AUC), Cairo, Egypt, in 2012 and 2015, respectively. He was a wireless application engineer with the Wireless Communications Lab, Intel Corporation, from 2012 to 2014. He is currently pursuing the Ph.D. degree with the University of British Columbia (UBC), Vancouver, BC, Canada. He has been a recipient of the Four Year Doctoral Fellowship (4YF) at UBC since 2015. He was also awarded the Graduate Support Initiative (GSI) Award at UBC in 2016 and 2017. His research interests include the Internet of Things (IoT), machine learning for wireless communications, 5G wireless networks, and radio resource allocation. Mr. Mostafa served as a TPC Member of the *IEEE VTC-Spring'*17. He also served as a Reviewer of *IEEE Transactions on Vehicular Technology*, *IEEE Access*, *IEEE Communications Letters*, *IEEE Wireless Communications Letters*, and several conferences.

Yong Zhou (S'13, M'16) received the B.Sc. and M.Eng. degrees from Shandong University, Jinan, China, in 2008 and 2011, respectively, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2015. From 2015 to 2017, he worked as a postdoctoral research fellow in the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada. He is currently an Assistant Professor in the School of Information Science and Technology, ShanghaiTech University, China. His research interests include performance analysis and resource allocation of 5G and Internet of Things (IoT) networks. He has served as a technical program committee (TPC) member for several conferences.



Vincent W.S. Wong (S'94, M'00, SM'07, F'16) received the B.Sc. degree from the University of Manitoba, Winnipeg, MB, Canada, in 1994, the M.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, in 1996, and the Ph.D. degree from the University of British Columbia (UBC), Vancouver, BC, Canada, in 2000. From 2000 to 2001, he worked as a systems engineer at PMC-Sierra Inc. (now Microchip Technology Inc.). He joined the Department of Electrical and Computer Engineering at UBC in 2002 and is currently a Professor. His research areas include protocol design, optimization, and resource management of communication networks, with applications to wireless networks, smart grid, mobile edge computing, and Internet of Things. Currently, Dr. Wong is an Executive Editorial Committee Member of *IEEE Transactions on Wireless Communications*, an Area Editor of *IEEE Transactions on Communications*, and an Associate Editor of *IEEE Transactions on Mobile Computing*. He has served as a Guest Editor of *IEEE Journal on Selected Areas in Communications* and *IEEE Wireless Communications*. He has also served on the editorial boards of *IEEE Transactions on Vehicular Technology* and *Journal of Communications and Networks*. He was a Tutorial Co-Chair of *IEEE Globecom*'18, a Technical Program Co-chair of *IEEE SmartGridComm*'14, as well as a Symposium Co-chair of *IEEE ICC*'18, *IEEE SmartGridComm* ('13, '17) and *IEEE Globecom*'13. He is the Chair of the IEEE Vancouver Joint Communications Chapter and has served as the Chair of the IEEE Communications Society Emerging Technical Sub-Committee on Smart Grid Communications. He received the 2014 UBC Killam Faculty Research Fellowship. He is an IEEE Communications Society Distinguished Lecturer (2019 - 2020).