



# Downlink Connection Density Maximization for NB-IoT Networks using NOMA with Perfect and Partial CSI

Shashwat Mishra, Lou Salaun, Chi Wan Sung, Chung Shue Chen

## ► To cite this version:

Shashwat Mishra, Lou Salaun, Chi Wan Sung, Chung Shue Chen. Downlink Connection Density Maximization for NB-IoT Networks using NOMA with Perfect and Partial CSI. IEEE Internet of Things Journal, 2021. hal-03119407v2

HAL Id: hal-03119407

<https://inria.hal.science/hal-03119407v2>

Submitted on 25 Mar 2021

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

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

# Downlink Connection Density Maximization for NB-IoT Networks using NOMA with Perfect and Partial CSI

Shashwat Mishra, Lou Salaün, *Member, IEEE*, Chi Wan Sung, *Senior Member, IEEE*, and Chung Shue Chen, *Senior Member, IEEE*

**Abstract**—We address the issue of maximizing the number of connected devices in a Narrowband Internet of Things (NB-IoT) network using non-orthogonal multiple access (NOMA) in the downlink. We first propose an optimal joint sub-carrier and power allocation strategy assuming perfect channel state information (CSI) called Stratified Device Allocation (SDA), that maximizes the connectivity under data rate, power and bandwidth constraints. Then, we generalize the connectivity maximization problem to the case of partial CSI, where only the distance-dependent path-loss component of the channel gain is available at the base station (BS). We introduce a novel framework called the Stochastic Connectivity Optimization (SCO) framework. In this framework, we propose a heuristic improvement to SDA namely SDA with Excess Power (SDA-EP) algorithm for operation under partial CSI. Furthermore, we derive a concave approximation (SCO-CA) algorithm of near-optimal performance to SCO given the same amount of CSI. Through computer simulations, we show that SDA-EP and SCO-CA outperform conventional NOMA and OMA schemes in the presence of partial CSI over a wide range of service scenarios.

**Index Terms**—NB-IoT, NOMA, Connectivity Maximization, Stochastic Optimization, mMTC, MTCD, 5G, Concave Approximation, Partial CSI.

## I. INTRODUCTION

A N important goal of 5G is to provide massive connectivity. Under this paradigm, we wish to connect a huge number of devices with limited available resources. This issue has gained huge interest in recent times due to the emergence of technologies like NB-IoT and massive machine type communication (mMTC) [1]. The mMTC use-case has high device density, moderate data rate, modest reaction time requirement and low transmit power. These devices are low cost and deployed in huge number to serve a variety of applications like factory assembly lines, traffic monitoring, smart buildings, weather forecast networks and so on. mMTC devices belong to a bigger class of devices generally referred to as machine type communication devices (MTCDs), which

S. Mishra is with the Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai, 600036 India. He was with Bell Labs, Nokia Paris-Saclay. E-mail: shashwat@tenet.res.in

L. Salaün and C. S. Chen are with Bell Labs, Nokia Paris-Saclay, 91620 Nozay, France. E-mail: {lou.salaun, chung\_shue.chen}@nokia-bell-labs.com

C. W. Sung is with the Department of Electrical Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong. E-mail: albert.sung@cityu.edu.hk

This work was supported in part by the Hong Kong Innovation and Technology Fund (ITS/066/17FP) under the HKUST-MIT Research Alliance Consortium. Corresponding author: Chung Shue Chen

are communicating entities that do not necessarily need human interaction. The expected device density for NB-IoT use-case is about 52500 devices sending small payload in an area of 0.86 km<sup>2</sup> using 200 kHz of bandwidth with OFDMA [2].

It is often not possible to deploy dedicated bandwidth for all the application scenarios of NB-IoT due to explosive overall growth of communication systems, so we must work with available bandwidths to achieve their service requirements. For example, current NB-IoT standards support in-band LTE deployment along with standalone deployment mode to conserve bandwidth. NB-IoT is a cellular low-power wide-area (LPWA) network technology that typically operates using a bandwidth of 180 kHz. One physical resource block (PRB) of 180 kHz bandwidth consists of 12 sub-carriers with a 15 kHz sub-carrier spacing. NB-IoT uses coverage enhancement techniques such as power boosting in downlink or sub-frame repetition in both uplink and downlink to meet the additional 20 dB maximum coupling loss (MCL) requirement. Other 3GPP IoT technologies include eMTC (enhanced Machine-Type Communication) and EC-GSM-IoT (Extended Coverage GSM-IoT). In downlink, NB-IoT systems use OFDM for communication while in uplink SC-FDMA is used because of the high peak to average power ratio (PAPR) issue of OFDM that makes it unattractive for low transmit power devices that are typical for NB-IoT deployments. mMTC has already been developed as part of 3GPP Release 13 LPWA technologies, which also includes NB-IoT. 3GPP specifies Cat-NB1 and Cat-NB2 [3] for exceptionally deep coverage and extremely low power applications. NB-IoT is a suitable key technology for 5G mMTC and can operate inside the 5G NR frequency bands similar to LTE today. mMTC focuses on providing connectivity to a large number of devices that transmit a small amount of data at irregular intervals. The standards expect a connection density of 10<sup>6</sup> devices/km<sup>2</sup> and battery life of up to 10 years for low data rates [1].

NOMA has emerged as a promising technical advancement which can accommodate a large number of devices through non-orthogonal resource allocation. This enables NOMA to improve the spectral efficiency of a communication system while allowing some amount of multiple access interference at the receiver. Superposition coding at the transmitter and successive interference cancellation (SIC) at the receiver makes it possible for NOMA to simultaneously allocate the same bandwidth to multiple devices. The basic signal waveform for NOMA is based on OFDM or discrete Fourier transform

spread OFDM (DFT-s-OFDM), usually following LTE baseline. With rapid advancements in semiconductor fabrication technology, it is very likely that SIC will be implementable even on low end devices that fall under NB-IoT or mMTC use-case. We, therefore, see an urgent requirement to develop solutions that maximize the number of connected devices per unit area using NOMA transmissions.

### A. Related Works

The work presented in [4], [5] illustrates the potential of NOMA for increasing the spectral efficiency of wireless networks. Methods for efficient power allocation for NOMA systems are also proposed in [6]–[8] for different performance metrics such as user fairness and quality of service (QoS) guarantee. In particular, the solution developed in [9] proposes fast orthogonal frequency division multiplexing (FastOFDM), to potentially double the number of connected devices by bandwidth compression. The authors in [10], [11] show the effectiveness of power domain NOMA for serving a greater number of devices. Note that in [11], the authors have relaxed the SIC limitations, thus allowing any number of superposed signals from devices to be successfully decoded, and shown that the system can support significantly more connected devices than traditional schemes. However, superposing a huge number of devices in any NOMA system is infeasible in practice due to error propagation issues in SIC and decoding complexity.

User signal superposition is crucial in NOMA and power allocation is inherently coupled with the way that devices are superposed. The work in [12] provides an optimal power control policy for SIC with imperfect channel estimation. In [13], a joint user pairing and power allocation problem is considered to optimize the achievable sum rate under each user's minimum rate requirement, which is a mixed integer programming problem. Reference [14] proposes a strategy to efficiently utilize the spectrum of unpaired devices in NOMA systems by pairing multiple similar link gain devices to a single distant user in a non-overlapping frequency band pairing fashion. In [15], the authors study the power consumption minimization problem for a generic multi-cell multiple input and single output NOMA system. The focus of this work is to develop user clustering and power allocation algorithms that maximize the energy efficiency of the system. The joint user grouping, beamforming and power control problem is formulated as a mixed integer non-convex programming problem and solved by an iterative distributed methodology.

Authors in [16] compare the performance of popular LPWA network technologies for IoT and show the power efficiency and deep coverage capabilities of NB-IoT. A popular paradigm is to optimize the energy consumption of NB-IoT networks. In [17], this problem is formulated as a stochastic optimization problem with the objective to minimize the total power consumption of the whole network under the constraint of all devices' long-term rate requirements and a solution based on branch-and-bound technique is proposed. In [18], the power allocation problem of NB-IoT devices on the same sub-carrier is converted into a difference of convex functions

(DC) problem and a sub-optimal solution is obtained which achieves better performance compared with existing schemes in the system throughput.

In [19], the authors use game theoretic methods to derive a many-to-many two-sided user to sub-channel matching algorithm for throughput maximization in a NOMA system. The work in [20] suggests a user clustering based approach for an NB-IoT network through bunching devices with different service rate requirements together and considering these clusters for allocation instead of individual devices, however, the emphasis of this work is to maximize the total throughput of the network by optimizing the resource allocation of MTC devices and NOMA clustering.

The literature that addresses the problem of connectivity maximization for NB-IoT networks using NOMA, where the number of connected devices is chosen as a metric for evaluation of NB-IoT deployments, is scarce. The authors in [21] formulate a joint sub-carrier and power allocation problem to maximize the connection density while taking the quality of service requirements and the transmit power constraints of IoT devices into account and solve it through a mixed integer linear programming method.

The work presented in [22] proposes a millimeter-wave NOMA transmission scheme for cellular machine-to-machine (M2M) communication systems for IoT applications and evaluates its downlink performance in terms of device outage probability. The authors develop an MTC pairing scheme based on the distance between the BS and the MTC devices.

Authors in [23] address the problem of uplink connectivity maximization in an NB-IoT system, providing heuristic solutions employing combinatorial optimization. This work solves the connectivity maximization problem under interference, rate and sub-carrier availability constraints and the solutions developed are applicable to any number of superposed devices per sub-carrier. Additionally, the result provides an analysis of non-contiguous sub-carrier allocation for the multi-tone uplink assignment problem.

In practice, it is often not possible to acquire perfect CSI at the BS without incurring severe overheads. However, we can obtain a partial CSI consisting only of the distance-dependent path-loss for devices, that varies slowly with time. Using this partial CSI instead of the actual instantaneous CSI has a significant impact on the number of connected devices due to its impact on device power assignment and subsequently the interference caused by it. To fully achieve the connectivity potential of NOMA, the power control and sub-carrier allocation schemes have to be specifically designed taking the nature of partial CSI probabilistic distribution into account. However, to the best of our knowledge, no connectivity maximization problem has been studied in this setting. Hence, in this paper, we study the impact of perfect and partial CSI on the downlink connectivity of NB-IoT systems, as well as develop specific optimization algorithms. The detailed contributions are presented in the next sub-section.

### B. Contributions

In this work, we propose a novel downlink connection density maximization framework for NB-IoT using NOMA

with the consideration of perfect and partial CSI. The main contributions of the paper can be summarized as follows:

- 1) We tackle the joint sub-carrier and power allocation optimization problem for maximizing the number of connected devices that achieve their QoS requirement. To solve this problem, we propose a novel power filling based technique assuming perfect CSI called the Stratified Device Allocation (SDA) strategy. We prove that SDA is optimal. Additionally, SDA is computationally efficient and can accommodate any system-imposed SIC constraint of maximum number of multiplexed users per sub-carrier.
- 2) Assuming that only partial CSI is available, we propose a heuristic improvement to SDA, called the SDA-EP method, where we provide additional power to devices in excess of the minimum power required by them for their service rate. We investigate two different excess power schemes and analyze their performance and effectiveness.
- 3) Given partial CSI and considering Rayleigh fading, we transform the connectivity problem to a probabilistic optimization problem and develop the Stochastic Connectivity Optimization (SCO) framework to solve it. The SCO is a mixed integer and non-convex optimization problem. We leverage the knowledge of the probability distribution of the fading component of the channel gain. To the best of our knowledge, no connectivity maximization algorithm has been developed in this setting.
- 4) We derive a closed-form expression of the objective function's expected value in the SCO framework. Based on this, we develop a concave approximation algorithm called SCO-CA to solve the problem. We prove that its performance loss (i.e., the average number of connected devices) compared to any possible optimal solution given the same amount of CSI is at most  $e^{-4} \times S \approx 0.0183 \times S$ , where  $S$  is the number of sub-carriers in the system. SCO-CA can be readily implemented using any standard convex solver.

The rest of this paper is organized as follows. Section II describes the system model and NOMA framework used in this work. In Section III, we define the connectivity maximization problem given perfect CSI and then derive the optimal SDA algorithm. In Section IV, we formulate the connectivity maximization problem given partial CSI. We first introduce SDA-EP which is a heuristic improvement to SDA to work under partial CSI. Then, we propose the SCO-CA algorithm and derive an upper bound on its performance loss. We evaluate the effectiveness of the proposed algorithms through computer simulations in Section V. Finally, Section VI concludes the paper.

## II. SYSTEM MODEL

Consider a set of  $D$  devices denoted by  $\mathcal{D} \triangleq \{1, \dots, D\}$ , dropped randomly in an area served by one base station (BS). We wish to maximize the number of connected devices in the downlink under a given power budget and user QoS requirement. The system can support up to  $M$  devices superposed per sub-carrier, which reflects the SIC limitation of NOMA [24]. We assume that perfect SIC decoding can be carried

out at the receiver. The total transmit power for all devices in one PRB is limited to  $P_{\max}$ . Each PRB is divided into  $S$  sub-carriers, the set of which is denoted by  $\mathcal{S} \triangleq \{1, \dots, S\}$ . The system has a bandwidth of  $B_s$  Hz with each sub-carrier of equal bandwidth  $B$  Hz. Each device can be allocated only one sub-carrier in the downlink. The numerology for sub-carrier size and number of allocated sub-carriers to each device used in this work follows 3GPP NB-IoT [2] to ensure compatibility with LTE deployments and is widely used for today's NB-IoT. The thermal noise considered in this work is additive white Gaussian noise, denoted as  $N \triangleq N_0 \cdot B \cdot 10^{\frac{F}{10}}$ , where  $B$  is the bandwidth under consideration,  $N_0$  is noise spectral density measured in W/Hz and  $F$  is the noise figure expressed in dB. Although the solutions developed in this work are valid even for different values of noise power for different devices, for the sake of simplicity in the following discussions, we will assume that all devices have the same noise power  $N$ .

### A. NOMA Framework

We consider a power domain NOMA based framework for connectivity maximization, where each sub-carrier can accommodate a maximum of  $M$  devices. Assume we have devices  $\{1, \dots, i, \dots, k\}$  where  $k \leq M$  with messages  $x_1, \dots, x_i, \dots, x_k$  respectively on the same sub-carrier. The downlink received signal at device 1 can be written as:

$$y_1 = h_1 \left\{ \sqrt{p_1} x_1 + \dots + \sqrt{p_i} x_i + \dots + \sqrt{p_k} x_k \right\} + N,$$

where  $p_i$  is the transmit power for device  $i$  and  $h_i$  represents its composite channel gain. This channel gain takes into account both path-loss and small-scale Rayleigh fading for device  $i$  such that  $h_i = b_i l_i^{1/2}$ , where  $b_i \sim \mathcal{CN}(0, 1)$  and  $l_i$  is the distance-dependent path-loss for device  $i$ . The channel is assumed to be flat-fading because of the narrow bandwidth. We assume, without loss of generality, that the devices are sorted in decreasing order of their channel gains. That is,

$$g_1 \geq \dots \geq g_D, \quad (1)$$

where  $g_i \triangleq |h_i|^2$  is the amplitude of the channel gain for device  $i$ . We denote the set of devices allocated to sub-carrier  $n$  by  $\mathcal{D}_n$  and define vector  $\mathbf{d}_n \triangleq (d_n(1), \dots, d_n(D_n))$  to represent these devices, where  $D_n \triangleq |\mathcal{D}_n| \leq M$ , for all  $n \in \mathcal{S}$ , due to the SIC constraint. In downlink NOMA, the optimal decoding order on sub-carrier  $n$  for SIC is to follow the increasing order of the channel gains  $g_i$ , since a strong device can decode both its own signal and the signal for the interferer while a weaker device treats the other devices' signals as noise and can decode only its own signal [24]. Following this, we assume that the  $d_n(i)$ 's are sorted in the decreasing order of their channel gains, i.e.,  $g_{d_n(1)} \geq \dots \geq g_{d_n(D_n)}$ . As a result,  $d_n(i)$  would be the  $(D_n - i + 1)$ -th decoded device on sub-carrier  $n$ . The SINR for device  $d_n(i)$  can be expressed as:

$$\gamma_{d_n(i)} = \frac{g_{d_n(i)} p_{d_n(i)}}{I_{d_n(i)} + N},$$

where  $I_{d_n(i)}$  is the interference to device  $d_n(i)$  caused by other devices on the same sub-carrier, which is expressible as:

$$I_{d_n(i)} \triangleq g_{d_n(i)} \sum_{j=1}^{i-1} p_{d_n(j)}.$$

### B. Constraints and Data Rate Requirements

There are the following constraints to be considered in our system. First, we have the power budget:

$$0 \leq \sum_{n=1}^S \sum_{i=1}^{D_n} p_{d_n(i)} \leq P_{\max}. \quad (2)$$

Here,  $P_{\max}$  is the maximum total transmit power. Additionally, each device must achieve a minimum critical data rate to remain viable in the network. The achievable data rate for device  $d_n(i)$  can be expressed as:

$$r_{d_n(i)} = B \log_2 \left( 1 + \frac{g_{d_n(i)} p_{d_n(i)}}{I_{d_n(i)} + N} \right). \quad (3)$$

We denote the aforementioned minimum service rate by  $R$  such that:

$$r_{d_n(i)} \geq R, \quad \forall n \in \mathcal{S}, \forall i \in \mathcal{d}_n. \quad (4)$$

The power required for device  $d_n(i)$  to achieve its minimum rate  $R$  can be obtained from (3) as:

$$\begin{aligned} p_{d_n(i)} &= \frac{\zeta}{g_{d_n(i)}} (I_{d_n(i)} + N) \\ &= \zeta \left( \sum_{j=1}^{i-1} p_{d_n(j)} + \frac{N}{g_{d_n(i)}} \right), \end{aligned} \quad (5)$$

where  $\zeta = 2^{\frac{R}{B}} - 1$  is the target SINR to achieve rate  $R$ .

## III. DOWNLINK CONNECTIVITY MAXIMIZATION WITH PERFECT CSI

In this section, we assume that the channel state information  $g_i$ , for any device  $i \in \mathcal{D}$ , is known perfectly at the BS.

### A. Problem Formulation

Let us define vector  $\mathbf{p}_n \triangleq (p_{d_n(1)}, \dots, p_{d_n(D_n)})$  to denote the powers of all devices on sub-carrier  $n$ , and vector  $\mathbf{p} \triangleq (\mathbf{p}_n)_{n \in \mathcal{S}}$  to represent the list of the power vectors  $\mathbf{p}_n$ . We define  $Z_n(\mathbf{p}_n, \mathbf{d}_n)$  as the number of connected devices on sub-carrier  $n$ :

$$Z_n(\mathbf{p}_n, \mathbf{d}_n) \triangleq \sum_{i=1}^{D_n} \mathbb{1}(r_{d_n(i)} \geq R). \quad (6)$$

Here,  $\mathbb{1}(r_{d_n(i)} \geq R)$  is the indicator function that takes value 1 if the rate for device  $d_n(i)$  is at least equal to the required service rate  $R$ , and its value is 0 otherwise. Thus,  $Z_n(\mathbf{p}_n, \mathbf{d}_n)$  has the QoS constraint (4) implicitly enforced in its definition.

Using (2), (4) and (6), the downlink connectivity maximization problem given perfect CSI can be stated as follows:

$$\begin{aligned} &\underset{\forall n \in \mathcal{S}, \mathbf{p}_n, \mathbf{d}_n}{\text{maximize}} \quad \sum_{n=1}^S Z_n(\mathbf{p}_n, \mathbf{d}_n) \\ &\text{subject to} \quad C1 : 0 \leq \sum_{n=1}^S \sum_{i=1}^{D_n} p_{d_n(i)} \leq P_{\max}, \\ &\quad C2 : \sum_{n \in \mathcal{S}} |\{k\} \cap \mathcal{D}_n| \leq 1, \forall k \in \mathcal{D}, \\ &\quad C3 : D_n \leq M, \forall n \in \mathcal{S}. \end{aligned} \quad (\mathcal{P})$$

The objective function in  $\mathcal{P}$  aims to maximize the number of connected devices over all sub-carriers in  $\mathcal{S}$  that satisfy their QoS requirement. Constraint C1 refers to the maximum total transmit power. Constraint C2 implies that each device can be allocated one sub-carrier according to 3GPP NB-IoT [2]. Constraint C3 states the SIC limit of the system of supporting at most  $M$  devices superposed per sub-carrier. Note that the total bandwidth taken up by all the connected devices must be less than  $M \cdot B_s$  and this constraint is inherently enforced by C3 and our definition of the sub-carrier bandwidth  $B$ , such that the system bandwidth is  $B_s$ .

### B. Stratified Device Allocation (SDA)

We propose an optimal solution for the connectivity maximization problem  $\mathcal{P}$  under perfect CSI. We call this solution the Stratified Device Allocation (SDA) and its pseudo-code is given in Algorithm 1. Devices are to be allocated a sub-carrier from the strongest to the weakest i.e., from  $i = 1$  to  $D$ , given that the devices are sorted in the decreasing order of their channel gains according to (1). In Algorithm 1, line 2 loops over variable “layer” from 1 to  $M$ . At each iteration, the sub-carriers are iteratively allocated to the strongest remaining device, which is represented by  $i$  at line 6. Algorithm 1 terminates when:

- We run out the power budget and there is not enough power to serve an additional device (see the first condition of line 5).
- All devices are already served (see the second condition of line 5).
- All sub-carriers have been allocated  $M$  devices and adding any more devices to a sub-carrier will violate the system SIC constraint (happens when the two for loops at lines 2-3 terminates).

As shown in Theorem 1, the SDA allocation is optimal, i.e., it connects the greatest number of devices among all possible sub-carrier and power allocation strategies under perfect CSI.

**Theorem 1** (Optimality of SDA).

SDA solves optimally the downlink connectivity maximization problem  $\mathcal{P}$  with perfect CSI.

*Proof.* The proof of this theorem along with the supporting Lemmas 1 and 2 is given in Appendix A.  $\square$

Not only is the solution obtained through SDA optimal for perfect CSI, this technique as highlighted in Algorithm 1 is also computationally efficient.

**Algorithm 1** Stratified Downlink Allocation (SDA)

**Input:**  $B, M, N, P_{\max}$  and  $g_d, \forall d \in \mathcal{D}$ 

- 1: **Initialization:**  $\forall n \in \mathcal{S}, \mathbf{d}_n \leftarrow \emptyset, i \leftarrow 1, P \leftarrow 0$
  - 2: **for** layer = 1 to  $M$  **do**
  - 3:   **for**  $n = 1$  to  $S$  **do**
  - 4:      $p \leftarrow$  the required power of device  $i$  on sub-carrier  $n$  is calculated using (5)
  - 5:     **if**  $P + p \leq P_{\max}$  **and**  $i \leq D$  **then**
  - 6:        $\mathbf{d}_n(\text{layer}) \leftarrow i$
  - 7:        $p_i \leftarrow p$
  - 8:        $P \leftarrow P + p$
  - 9:        $i \leftarrow i + 1$
  - 10:      **else**
  - 11:       **terminate**
  - 12:      **end if**
  - 13:     **end for**
  - 14: **end for**
- Output:**  $\mathbf{p}$  and  $\mathbf{d}_n, \forall n \in \mathcal{S}$

**IV. DOWNLINK CONNECTIVITY MAXIMIZATION WITH PARTIAL CSI**

It is not feasible to get the exact instantaneous CSI for devices in a practical system without incurring significant overhead. The timescale of fast channel fading variation such as Rayleigh fading is much faster than the timescale of change of the path-loss component that often changes significantly only in the order of seconds or more slowly under some environments and depending on the device mobility. The latter is easier to estimate in contrast to the much faster Rayleigh fading channel gain. We therefore consider the slowly changing path-loss component is given while the knowledge of instantaneous Rayleigh fading channel gain is unavailable. In practice, we can estimate the slow varying component of the channel gain with high accuracy. In the following, we consider that the instantaneous channel gain  $g_i$  for device  $i$  is unknown at the BS but its distance-dependent path-loss value  $l_i$  and the distribution of the channel fading are known. We refer to this knowledge as the partial CSI.

### A. Problem Formulation

We now present a probabilistic framework for maximizing the number of connected devices with partial CSI while assuming the channel variation follows Rayleigh fading. This framework is termed Stochastic Connectivity Optimization (SCO). The devices on sub-carrier  $n$  are sorted in the decreasing order of their path-loss  $l_i$  here instead of  $g_i$  as in the case of perfect CSI formulation, i.e., we have  $l_{d_n(1)} \geq \dots \geq l_{d_n(D_n)}$  on sub-carrier  $n$ .

Note that although we do not have the full CSI, under our system model  $g_i$  is an exponential random variable with scale parameter  $l_i$  [24]. The probability that a device can be connected consists of the probability of successfully performing SIC on all the weaker devices and the probability of decoding its own data. This probability of connection for device  $i$  on sub-carrier  $n$  can be obtained by comparison with its critical SINR as follows:

$$\Pr(i) = \Pr\left(\bigcap_{j=i}^{D_n} \left\{ \frac{g_{d_n(i)} p_{d_n(j)}}{N + g_{d_n(i)} \sum_{k=1}^{j-1} p_{d_n(k)}} \geq \zeta \right\}\right),$$

$$= \Pr\left(\bigcap_{j=i}^{D_n} \left\{ g_{d_n(i)} \geq \frac{\zeta N}{p_{d_n(j)} - \zeta \sum_{k=1}^{j-1} p_{d_n(k)}} \right\}\right), \quad (7)$$

$$= \Pr\left(g_{d_n(i)} \geq \max_{j=i, \dots, D_n} \left\{ \frac{\zeta N}{p_{d_n(j)} - \zeta \sum_{k=1}^{j-1} p_{d_n(k)}} \right\}\right), \quad (8)$$

$$= \Pr\left(g_{d_n(i)} \geq \frac{\zeta N}{\min_{j=i, \dots, D_n} \left\{ p_{d_n(j)} - \zeta \sum_{k=1}^{j-1} p_{d_n(k)} \right\}}\right),$$

$$= \exp\left(-\frac{\zeta \beta_{d_n(i)}}{\min_{j=i, \dots, D_n} \left\{ p_{d_n(j)} - \zeta \sum_{k=1}^{j-1} p_{d_n(k)} \right\}}\right), \quad (9)$$

$$= \min_{j=i, \dots, D_n} \left\{ \exp\left(-\frac{\zeta \beta_{d_n(i)}}{p_{d_n(j)} - \zeta \sum_{k=1}^{j-1} p_{d_n(k)}}\right)\right\}, \quad (10)$$

where  $\beta_{d_n(i)} = N/l_{d_n(i)}$ . We obtain the expression in (8) since the event that  $g_{d_n(i)}$  is greater than each of the terms on the right hand side of the inequality in (7) is equivalent to the event that  $g_{d_n(i)}$  be greater than the maximum of all those terms. The expression in (9) is obtained by evaluating the complementary cumulative distribution function of the exponential random variable  $g_{d_n(i)}$ , and (10) follows from there through simple algebraic manipulation.

We see that due to the stochastic nature of the Rayleigh fading, the number of connected devices on a sub-carrier is a random variable. Hence,  $\Pr(i)$  represents not only the probability of successful connection but also the expected number of connected devices achieved by serving device  $i$ . For example,  $\Pr(i) = 0.5$  implies that we connect device  $i$  with a probability of 0.5, or alternatively we connect 0.5 devices on an average by serving device  $i$ . In general, the expected number of connected devices on sub-carrier  $n$ , denoted by  $\hat{Z}_n$  below, can be obtained in a closed-form expression as follows:

$$\hat{Z}_n(\mathbf{p}_n, \mathbf{d}_n) = \sum_{i=1}^{D_n} \Pr(i). \quad (11)$$

For notational simplicity to express  $\hat{Z}_n$ , we make the following change of variables for  $\mathbf{p}_n$  and  $\beta_{d_n}$ :

$$x_{d_n(j)} = p_{d_n(j)} - \zeta \sum_{k=1}^{j-1} p_{d_n(k)}, \quad (12)$$

$$a_{d_n(i)} = \zeta \beta_{d_n(i)}. \quad (13)$$

Thus, by the substitution of (12)–(13) into (10), we can express (11) as:

$$\hat{Z}_n(\mathbf{x}_n, \mathbf{d}_n) = \sum_{i=1}^{D_n} \min_{j=i, \dots, D_n} \left\{ \exp\left(-\frac{a_{d_n(i)}}{x_{d_n(j)}}\right)\right\}, \quad (14)$$

where  $\mathbf{x}_n \triangleq (x_{d_n(1)}, \dots, x_{d_n(D_n)})$ .

Based on (14), the downlink connectivity maximization problem under partial CSI is given as follows:

$$\begin{aligned} & \underset{\forall n \in \mathcal{S}, \mathbf{x}_n, \mathbf{d}_n}{\text{maximize}} \quad \sum_{n=1}^S \hat{Z}_n(\mathbf{x}_n, \mathbf{d}_n) \\ & \text{subject to} \quad C1': 0 \leq \sum_{n=1}^S \sum_{i=1}^{D_n} x_{d_n(i)} (\zeta + 1)^{D_n-i} \leq P_{\max}, \\ & \quad C2, C3. \end{aligned} \quad (\mathcal{P}_{SCO})$$

Constraint  $C1'$  represents the system power budget. Constraints  $C2$  and  $C3$  are the same as in problem  $\mathcal{P}$ . In contrast with  $\mathcal{P}$  which maximizes a simple counting function with discrete values, here we maximize the expected value of a probabilistic function, which is a continuous function of  $\mathbf{x}_n$ . Furthermore, Theorem 2 shows that it is non-concave, therefore solving it using convex programming would not always give optimal solution.

**Theorem 2** (Convexity Analysis of SCO).

$\hat{Z}_n(\mathbf{x}_n, \mathbf{d}_n)$  is a non-concave function of  $\mathbf{x}_n$ .

*Proof.* See Appendix B.  $\square$

### B. SDA with Excess Power (SDA-EP)

We propose a simple heuristic improvement to SDA to adapt to the partial CSI in performing the device allocation. Instead of allocating the exact power required by each device to achieve its minimum data rate as determined in (5), we provide it some additional power to make up for the imperfect CSI. This approach is called SDA with Excess Power (SDA-EP). A first idea can be to have an additive excess power term  $\delta > 0$ , which is the same for all devices, as follows:

$$p_{d_n(i)} = \zeta \left( \sum_{m=1}^{i-1} p_{d_n(m)} + \frac{N}{l_{d_n(i)}} \right) + \delta. \quad (15)$$

Note that this additive excess power heuristic can be implemented in the same way as Algorithm 1 with the following changes:

- $g_d$  is replaced by  $l_d$ , for all device  $d \in \mathcal{D}$ ,
- the power allocation at line 7 of Algorithm 1 is computed using (15).

The effect of  $\delta$  on the connectivity performance is shown in Fig. 1. The  $x$ -axis shows different values of  $\delta$  and the  $y$ -axis shows the average number of connected devices, which is computed over 1000 independent runs with the service rate  $R = 70$  kbps. We see that, for well chosen values of  $\delta$ , the additive power SDA-EP approach performs better compared to SDA with partial CSI. But as  $\delta$  increases, the performance of SDA-EP degrades rapidly. This is because, as we can deduce from (15), the additional power allocated to a device would increase the interference considerably for subsequent devices. The issue becomes increasingly severe as the number of interferers of a device increases. This is illustrated by the connectivity performance of the algorithm for the case of  $M = 2$  and  $M = 3$ . Here despite giving the same excess power for both cases, the number of connected devices do not increase significantly, even though there are more devices

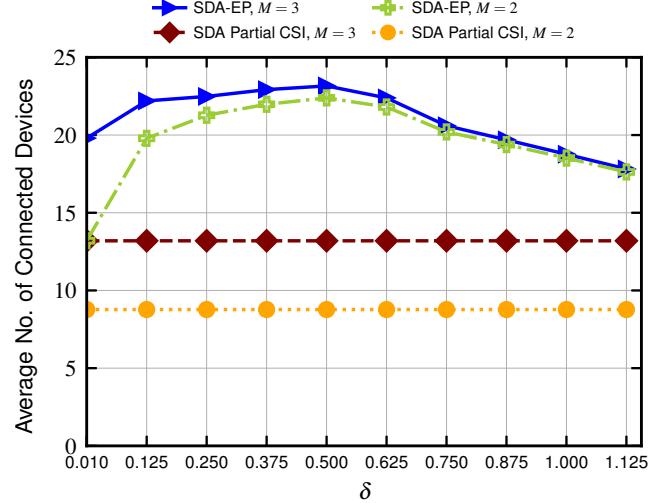


Fig. 1. Effect of  $\delta$  on the number of connected devices through SDA-EP based on (15).

available for connection in the case of  $M = 3$ . Furthermore, the power requirement of devices on different layers of allocation varies drastically: a strong device requires much less power for being connected compared to a weak device on the same sub-carrier, so having the same value of  $\delta$  for all devices is not well suited.

A better way of providing additional power is to assign it as follows:

$$p_{d_n(i)} = c\zeta \left( \sum_{m=1}^{i-1} p_{d_n(m)} + \frac{N}{l_{d_n(i)}} \right), \quad (16)$$

where  $c > 1$  is a multiplicative factor deciding the excess power and it is the same for all devices in the scaling up. To implement this method, we use Algorithm 1 with the following changes:

- $g_d$  is replaced by  $l_d$ , for all device  $d \in \mathcal{D}$ ,
- the power allocation at line 7 of Algorithm 1 is computed using (16), with a suitable value of  $c$  tuned empirically.

This method resolves the issue of adding the same amount of excess power to devices with different scales of power consumption as discussed in the context of the additive excess power approach. The excess power provided through the multiplicative scheme is proportional to the target SINR (required to achieve the desired service rate given perfect CSI), thus preventing huge interference increment for weaker devices due to the excess power assigned to stronger devices on the same sub-carrier.

An alternate view of the above strategy is that it artificially increases the target SINR required by all devices to achieve their data rates. Instead of requiring the devices to achieve a target SINR of  $\zeta$ , this approach forces the devices to achieve a target SINR of  $c\zeta (> \zeta)$  which in turn increases the power consumption of devices.

The effect of  $c$  on the connectivity performance of the algorithm is shown in Fig. 2. We can see that increasing the value of  $c$  up to a certain threshold progressively connects

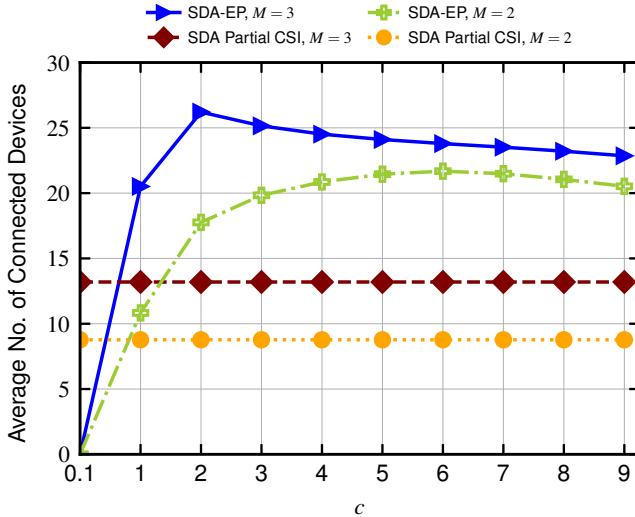


Fig. 2. Effect of  $c$  on the number of connected devices through SDA-EP based on (16).

more devices since the excess power assigned compensates for the imperfect CSI. Moreover, the additional interference incurred in doing so does not cause a significant increase in the power consumption of devices. After this threshold, the additional power would cause a significant increase in interference for devices and hence their power consumption. This causes the power budget to get exhausted with relatively fewer connected devices. Therefore, a proper choice of  $c$  is definitely important for maximizing the benefit in this method. The optimal value of  $c$  is sensitive to the service rate, number of devices superimposed on a sub-carrier and the cell radius. Note that since this multiplicative approach is superior to the additive approach, henceforth we will refer to SDA-EP as the multiplicative approach following (16).

### C. Device Allocation Optimization Through the SCO Framework

In the following, we will define a two-stage optimization scheme for solving problem  $\mathcal{P}_{SCO}$ . The first stage takes a subset of devices  $\mathcal{D}' \subset \mathcal{D}$  as input and outputs the device allocation on each-sub-carrier,  $\mathbf{d}_n$ , for all  $n \in \mathcal{S}$ . This procedure is illustrated in Algorithm 2. Here, we proceed in a manner similar to SDA by allocating devices with successively decreasing partial channel gains as shown in line 5, starting with the device having the highest partial channel gain. The allocation of devices to sub-carrier  $n$  will be of the form  $\mathbf{d}_n = \{n, n+S, \dots, n+(D_n-1)S\}$ , with their path-loss values  $l_i$  having the following order:

$$l_{d_n(1)} > l_{d_n(2)} > \dots > l_{d_n(D_n)}. \quad (17)$$

The second stage of the procedure consists of allocating power to the devices. However, since  $\mathcal{P}_{SCO}$  is a non-concave maximization problem, solving it using convex programming tools may not give us optimal results. Therefore, we develop an approximation to (14), called Concave Approximation under

SCO (SCO-CA). The technical details will be derived and explained in the coming sub-section IV-D.

### Algorithm 2 SCO Device Allocation

**Input:**  $B, M, N$ , a set of devices  $\mathcal{D}'$  and  $l_d, \forall d \in \mathcal{D}'$

```

1: Initialization:  $\forall n \in \mathcal{S}, \mathbf{d}_n \leftarrow \emptyset, i \leftarrow 1$ 
2: for layer = 1 to  $M$  do
3:   for  $n = 1$  to  $S$  do
4:     if  $i \leq |\mathcal{D}'|$  then
5:        $\mathbf{d}_n(\text{layer}) \leftarrow i$ 
6:        $i \leftarrow i + 1$ 
7:     else
8:       terminate
9:     end if
10:   end for
11: end for
Output:  $\forall n \in \mathcal{S}, \mathbf{d}_n$ 
```

### D. Concave Approximation for Power Control under SCO Framework (SCO-CA)

As an illustration, let's consider the simple case of the problem and its formulation for one sub-carrier with a single device. The expression (14) can be simply reduced as:

$$\hat{Z}_n(x, \mathbf{d}_n) = \exp\left(-\frac{a}{x}\right), \quad (18)$$

where  $\mathbf{d}_n$  refers to the single device. One can find that there is a change in convexity generally at  $x = \frac{a}{2}$ . The function  $\hat{Z}_n(x, \mathbf{d}_n)$  is concave only when  $x > \frac{a}{2}$ . We therefore approximate the non-concave part of  $\hat{Z}_n(x, \mathbf{d}_n)$  by its concave hull, denoted by  $\mathcal{L}_a(x)$ . In other words,  $\mathcal{L}_a(x)$  is the tightest concave lower bound of  $\hat{Z}_n(x, \mathbf{d}_n)$  for  $x \leq \frac{a}{2}$ .  $\mathcal{L}_a(x)$  can be expressed as follows:

$$\mathcal{L}_a(x) = \frac{4e^{-2}}{a} \left(x - \frac{a}{4}\right).$$

The function  $\hat{Z}_n(x, \mathbf{d}_n)$  and its concave approximation is illustrated in Fig. 3, where the approximated function is denoted by  $\tilde{Z}_n(x)$  and defined as follows:

$$\tilde{Z}_n(x, \mathbf{d}_n) = \begin{cases} \hat{Z}_n(x, \mathbf{d}_n) & x > \frac{a}{2}, \\ \mathcal{L}_a(x) & \frac{a}{4} \leq x \leq \frac{a}{2}, \\ 0 & x < \frac{a}{4}. \end{cases} \quad (19)$$

For simplicity, we set  $\tilde{Z}_n(x, \mathbf{d}_n) = 0$  when  $x < \frac{a}{4}$ , since the original objective function cannot be negative. Even though the approximated objective  $\sum_{n \in \mathcal{S}} \tilde{Z}_n$  is not globally concave due to this “zero region”, this pseudo-concave behavior can still be tackled by standard convex programming tools easily [25].

Similar to (18)–(19), we propose the following piece-wise approximation to  $\hat{Z}_n(x, \mathbf{d}_n)$ :

$$\tilde{Z}_n(x, \mathbf{d}_n) = \sum_{i=1}^{D_n} \min_{j=i, \dots, D_n} f(a_{d_n(i)}, x_{d_n(j)}),$$

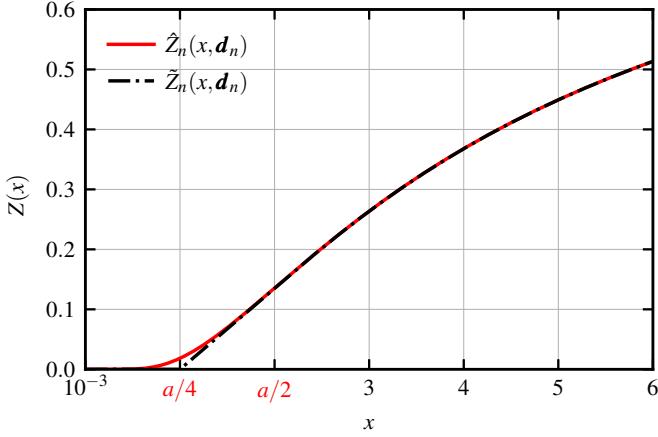


Fig. 3. Approximation of the original function by a concave function.

$$f(a_{d_n(i)}, x_{d_n(j)}) = \begin{cases} \exp\left(\frac{-a_{d_n(i)}}{x_{d_n(j)}}\right) & x_{d_n(j)} > \frac{a_{d_n(i)}}{2}, \\ \mathcal{L}_a(x_{d_n(j)}) & \frac{a_{d_n(i)}}{4} \leq x_{d_n(j)} \leq \frac{a_{d_n(i)}}{2}, \\ 0 & x_{d_n(j)} < \frac{a_{d_n(i)}}{4}, \end{cases} \quad (20)$$

where

$$\mathcal{L}_a(x_{d_n(j)}) \triangleq \frac{4e^{-2}}{a_{d_n(i)}} \left( x_{d_n(j)} - \frac{a_{d_n(i)}}{4} \right),$$

such that  $\tilde{Z}_n(\mathbf{x}_n, \mathbf{d}_n)$  is a concave function of  $\mathbf{x}_n$ . Using (20), the connectivity maximization problem in the concave approximation can be expressed similar to  $\mathcal{P}_{SCO}$  as follows:

$$\begin{aligned} & \underset{\forall n \in \mathcal{S}, \mathbf{x}_n, \mathbf{d}_n}{\text{maximize}} \quad \sum_{n=1}^S \tilde{Z}_n(\mathbf{x}_n, \mathbf{d}_n), \quad (\mathcal{P}_{CA}) \\ & \text{subject to} \quad C1 : 0 \leq \sum_{n=1}^S \sum_{i=1}^{D_n} x_{d_n(i)} (\zeta + 1)^{D_n - i} \leq P_{\max}, \\ & \quad C2, C3. \end{aligned}$$

Let us evaluate the approximation error incurred by the above formulation. For the case of one sub-carrier, by the definition of concave hull, this approximation is the best possible among all concave approximations and has the smallest approximation error. We see that the largest difference between the original function  $\hat{Z}_n(x, \mathbf{d}_n)$  and the approximated function  $\tilde{Z}_n(x, \mathbf{d}_n)$  is  $e^{-4} \approx 0.01832$ , which occurs at  $x = \frac{a}{4}$ . By a similar argument, we can provide an upper bound on the performance gap between the optimal solution of  $\mathcal{P}_{SCO}$  and that of  $\mathcal{P}_{CA}$  in Theorem 3, which shows that the performance loss (i.e., the average number of connected devices) of SCO-CA solution compared to any possible optimal solution given the same amount of CSI is at most  $e^{-4} \times S \approx 0.0183 \times S$ , where  $S$  is the number of sub-carriers in the system.

**Theorem 3** (Approximation Error Bound for SCO-CA).  
The performance gap between the optimal solutions of  $\mathcal{P}_{SCO}$  and  $\mathcal{P}_{CA}$  is at most  $e^{-4} \times S$ .

*Proof.* See Appendix C.  $\square$

As shown in Algorithm 3, we obtain the solution to  $\mathcal{P}_{CA}$  following a recursive approach. Variable  $\mathcal{D}'$  is the subset

---

### Algorithm 3 SCO-CA Joint Device and Power Allocation

---

**Input:**  $B, M, N, P_{\max}$  and  $l_d, \forall d \in \mathcal{D}$

- 1: **Initialization:**  $Z^* \leftarrow 0, \hat{Z} \leftarrow 0$
  - 2:  $\forall n \in \mathcal{S}, \mathbf{d}_n \leftarrow$  the optimal allocation of devices  $\mathcal{D}$  using Algorithm 2, and initialize  $\mathcal{D}'$  accordingly as the set of allocated devices of size  $|\mathcal{D}'| = \min(D, MS)$
  - 3: **while**  $\mathcal{D}' \neq \emptyset$  **do**
  - 4:      $\hat{Z} \leftarrow$  optimal value of  $\underset{\forall n \in \mathcal{S}, \mathbf{x}_n}{\text{maximize}} \sum_{n=1}^S \tilde{Z}_n(\mathbf{x}_n, \mathbf{d}_n)$  satisfying constraint  $C1$  computed by a convex solver
  - 5:     **if**  $\hat{Z} > Z^*$  **then**
  - 6:          $Z^* \leftarrow \hat{Z}$
  - 7:          $\mathbf{p}^* \leftarrow$  power allocation computed in line 4
  - 8:          $\forall n \in \mathcal{S}, \mathbf{d}_n^* \leftarrow \mathbf{d}_n$
  - 9:     **end if**
  - 10:    Remove  $|\mathcal{D}'|$  from the device allocation  $\mathbf{d}_n, n \in \mathcal{S}$
  - 11:     $\mathcal{D}' \leftarrow \mathcal{D}' \setminus \{|\mathcal{D}'|\}$
  - 12: **end while**
- Output:**  $Z^*$  and its corresponding allocation:  $\forall n \in \mathcal{S}, \mathbf{d}_n^*, \mathbf{p}^*$
- 

of devices considered for allocation at each iteration. At initialization at line 2, it contains all  $D$  devices if they can all be served on the  $S$  sub-carriers (i.e.  $D \leq MS$ ) or it contains the best  $MS$  devices if  $D > MS$ , thus  $|\mathcal{D}'| = \min(D, MS)$ . At line 4,  $\hat{Z}$  is obtained using any standard convex solver [26] for finding the optimal solution of the pseudo-concave function (20) assuming the optimal device allocation  $\mathbf{d}_n$  computed at line 2. Initially, this solution is the best solution found to the problem and it is stored in variable  $Z^*$ , along with the corresponding power allocation in  $\mathbf{p}^*$  and the device allocation stored in  $\mathbf{d}_n^*$ , for all  $n \in \mathcal{S}$  (see lines 6-8). Then, we remove from  $\mathcal{D}'$  the weakest device sorted according to their path-loss values  $l_{i \in \mathcal{D}'}$  (due to the available partial CSI) from the last allocated sub-carrier (see lines 10 and 11). By (17), this device is the one with index  $|\mathcal{D}'|$ . We solve the problem again considering the subset of devices  $\mathcal{D}'$  and check if the expected number of connected devices  $\hat{Z}$  is greater than the previous best solution  $Z^*$  (see lines 5-9). At each iteration, the above re-distributes the available power among devices having successively better decoding probabilities. We continue this procedure of removing the subsequent weakest device and updating  $Z^*$  until we run out of devices (see line 3). At termination, the algorithm returns the device allocation  $\mathbf{d}_n^*$ , for all  $n \in \mathcal{S}$ , and power allocation  $\mathbf{p}^*$  that achieves the best connectivity  $Z^*$ .

By construction, Algorithm 3 outputs an optimal solution of problem  $\mathcal{P}_{CA}$ . It follows from Theorem 3 that SCO-CA computes a solution of  $\mathcal{P}_{SCO}$  with an absolute error of at most  $e^{-4} \times S$ . For 1 PRB with  $S = 12$  sub-carriers, this corresponds to an error of less than 0.2198 connected devices in expectation. Assuming that 24 devices could potentially be connected for  $M = 2$ , SCO-CA guarantees an approximation ratio of at least 99% in this setting.

We now analyse the computational complexity of the developed algorithms. The SCO framework first requires sorting the path-loss values in decreasing order, as shown in (17). This can be done with complexity  $O(D \log(D))$  using quicksort

or mergesort. Furthermore, the complexity of Algorithm 3 is detailed in Theorem 4 shown below. However, this complexity can be improved in practice. Indeed,  $\tilde{Z}$  is uni-modal over the iterations on  $\mathcal{D}'$  (i.e., it first increases, then decreases), therefore it is possible to terminate the procedure once we start observing no increase in  $Z^*$  upon removing weaker devices.

**Theorem 4** (Computational Complexity of SCO-CA).

Assuming that the convex solver used at line 4 of Algorithm 3 is either gradient descent or interior point method, the computational complexity of SCO-CA is  $O(\min(D, MS)^2 M \log(1/\epsilon))$ , where  $\epsilon$  is the desired accuracy.

*Proof.* See Appendix D.  $\square$

## V. SIMULATION RESULTS AND PERFORMANCE EVALUATION

We evaluate the performance of our solutions through extensive simulations in a typical NB-IoT setting. Key system parameters are highlighted in Table I, which are based on 3GPP standards for NB-IoT [1]. We assume Rayleigh fading and frequency flat over the system bandwidth due to narrow-band consideration. We consider that perfect SIC for NOMA can be carried out by the receiver. We use CVXOPT [26] which is a free, open-source software package for convex optimization. Devices are dropped randomly following a uniform distribution in a hexagonal cell of radius 500 meters. All devices try to get connected with the same data rate requirement of  $R$  kbps. Note that in many of the presented results, we have chosen  $R$  to be a value that is substantially higher than the service rate requirement stated by 3GPP standards [1]. Not only do our algorithms achieve the current QoS requirement with ease, we will show they can also achieve good connectivity performance even at higher rates that may be viable in future use cases. Besides, we present the performance of our algorithms under both  $M = 2$  and  $M = 3$ . Each PRB of 180 kHz bandwidth has 12 sub-carriers which can accommodate upto 24 devices with  $M = 2$  and 36 devices with  $M = 3$ , respectively.

TABLE I  
KEY SYSTEM SIMULATION PARAMETERS

|                       |                                  |
|-----------------------|----------------------------------|
| Carrier Frequency     | 900 MHz                          |
| System Bandwidth      | 180 kHz / PRB                    |
| Sub-Carrier Bandwidth | 15 kHz                           |
| Path-Loss             | $120.9 + 37.6 \log_{1000} D + G$ |
| Antenna Gain (G)      | -4 dB                            |
| $P_{\max}$            | 23 dBm / PRB                     |
| AWGN Power            | -174 dBm/Hz                      |
| Noise Figure          | 5 dB                             |

Fig. 4 shows the performance comparison between different algorithms given perfect CSI in 1 PRB with  $R = 100$  kbps. Current NB-IoT deployments use OFDMA in the downlink for device assignment, so we consider OMA as the baseline, i.e., when  $M = 1$ . We see that SDA with  $M = 3$  and SDA with  $M = 2$  connects 2.27 and 2 times the devices when compared to the case of  $M = 1$  (OMA) respectively, when a total number of 36 devices are dropped. We can see that the growth in the number of connected devices will slow down as expected due

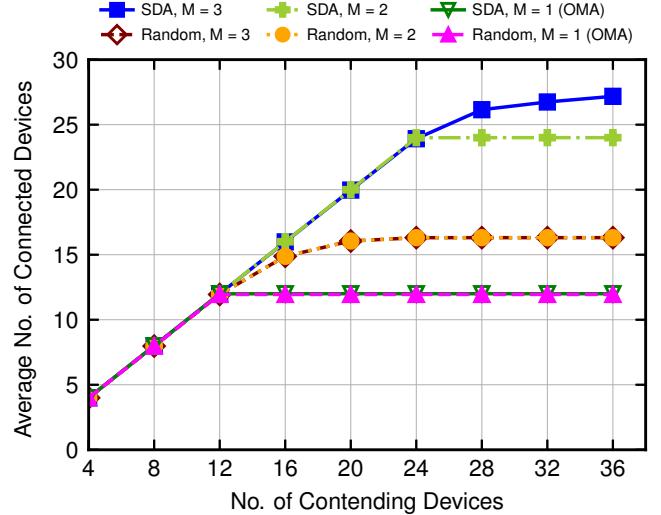


Fig. 4. Connectivity performance comparison with perfect CSI for 1 PRB.

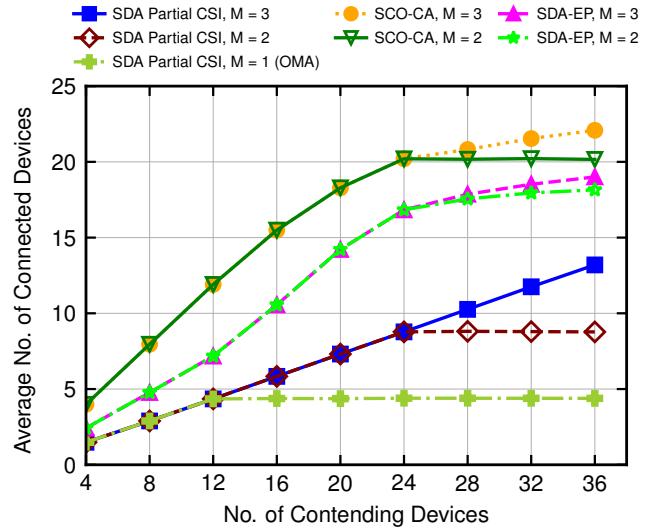
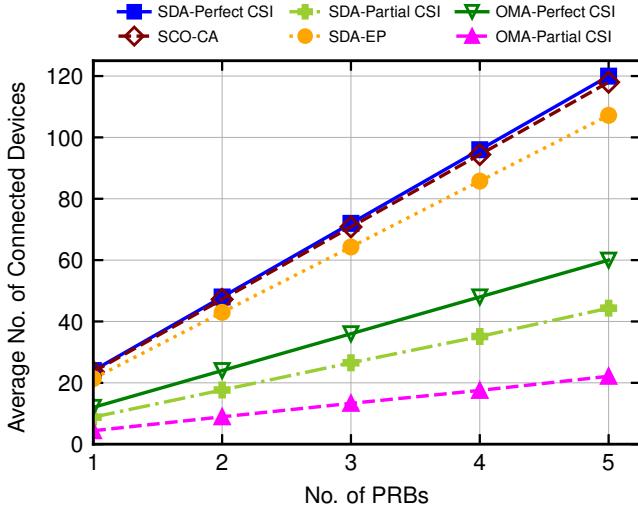
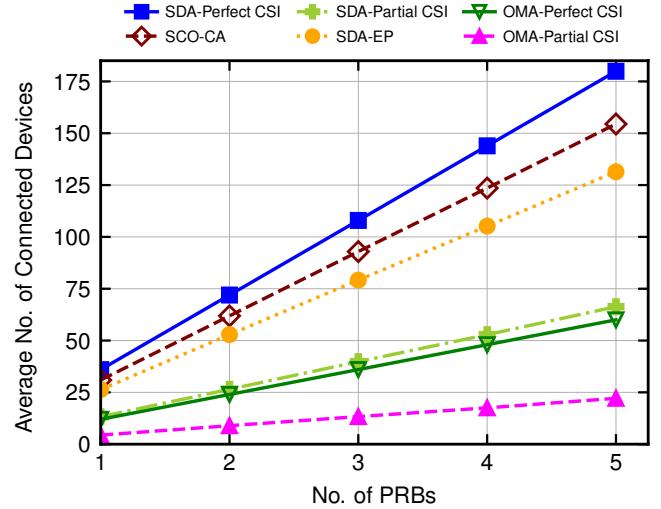


Fig. 5. Connectivity performance comparison with partial CSI for 1 PRB.

to greater interference with more connected devices. SDA with  $M = 3$  can connect 75.49% of the 36 devices on average. Meanwhile, we show the performance when the devices are assigned to each sub-carrier in a random order (labeled as “Random” in Fig. 4) to highlight the importance of proper device pairing for connectivity maximization. Note that the effect of device ordering is not pronounced for the case of  $M = 1$  (also because we assume frequency flat over the system bandwidth). However, for the cases of  $M = 2$  and  $M = 3$ , SDA connects up to 47.1% and 66.6% more devices, respectively. This gain becomes increasingly significant as the service rate and the number of devices superimposed on each sub-carrier increases. The comparison shows the effectiveness of SDA.

Next, we evaluate the performance of different algorithms when only partial CSI is known. Fig. 5 shows the number of connected devices in 1 PRB with  $R = 100$  kbps. For the

Fig. 6. Connectivity performance with  $R = 70$  kbps and  $M = 2$  for 5 PRBs.Fig. 7. Connectivity performance with  $R = 70$  kbps and  $M = 3$  for 5 PRBs.

case of  $M = 2$ , we can solve the problem only for at most  $2S$  devices. Even when more than  $2S$  devices are available, in the case of  $M = 2$  we must choose only a maximum of  $2S$  devices from the total number of available devices and solve the joint power and device allocation problem for this selected subset of devices. This is due to the SIC limitation of the system that allows a maximum of only  $M$  devices per sub-carrier. Thus, up to  $2S$  devices contending for connection, the performance of  $M = 2$  and  $M = 3$  for the same devices is the same because we are in fact solving the same optimization problem in both cases. For example, when we solve for, say 16 devices, Algorithm 2 for SCO Device allocation outputs the same  $\mathbf{d}_n$  for the case of  $M = 2$  and 3. We can see in Fig. 4 that with perfect CSI, SDA  $M = 2$  and SDA  $M = 3$  have the same performance for up to 24 dropped devices.

When the number of dropped devices increases, the problem of SDA with  $M = 2$  is solved with only 24 devices out of the total number of dropped devices while the problem of SDA with  $M = 3$  is solved with all the additional devices as well. The same behaviour applies to Fig. 5 but with a lower average number of connected devices due to the partial CSI. Since we only have partial CSI (also the presented result is for a high data rate requirement), the algorithms are unable to connect all the devices. As the number of devices per sub-carrier increases, the partial CSI problem becomes increasingly difficult to solve. The connected number of devices decreases even though there are more candidate devices since the improper power allocation due to partial availability of CSI causes increased interference for all the devices, causing the power budget to be exhausted with fewer connected devices and hence degrading the connectivity.

We see that SCO-CA is the most effective algorithm for the connectivity maximization in the case of partial CSI: SCO-CA with  $M = 2$  and SCO-CA with  $M = 3$  can respectively connect up to 4.59 and 5.03 times the devices connected by OMA with SDA given only partial CSI when 36 devices are competing for connection. SDA with  $M = 2$  and SDA with

$M = 3$  with partial CSI connect 2 and 3 times the devices when compared to SDA with  $M = 1$  respectively. SDA-EP with  $M = 2$  and SDA-EP with  $M = 3$  can connect up to 4.13 and 4.33 time the devices connected by SDA with partial CSI for  $M = 2$  and  $M = 3$ , respectively. We see that SCO-CA connects 61.3% of the total number of devices on average whereas SDA with partial CSI can only connect 36.5%. With excess power through SDA-EP, the connectivity increases to 52.7%. The performance of SCO-CA with partial CSI is close to that of SDA with perfect CSI since the former connects about 23% lesser devices than the latter on an average.

Fig. 6 shows a comparison between different algorithms with  $M = 2$  and  $R = 70$  kbps. We consider the availability of 5 PRBs, which can support a maximum of 120 devices. As expected, SDA with perfect CSI achieves the highest connectivity by successfully connecting 99.9% of the 120 devices. This is followed by SCO-CA, which connects 98.3% of the devices and can support 1.96 times the number of devices when compared to SDA with perfect CSI under OMA ( $M = 1$ ). We can see the effectiveness of SCO-CA. Note that SDA with partial CSI can connect only 36.7% of the devices. However, with excess power through SDA-EP, the connectivity reaches up to 89.3%.

Fig. 7 presents the performance of various algorithms considering  $M = 3$ . Here again the best algorithm in the case of partial CSI is SCO-CA which connects 85.9% of the devices. Meanwhile, it connects 2.57 times the devices compared to SDA with perfect CSI under OMA ( $M = 1$ ). SDA with perfect CSI connects 99.9% devices. SDA-EP gives a decent gain for SDA with partial CSI, connecting 73.1% of the total number of devices, which is 2.19 times the devices connected by SDA with perfect CSI under OMA ( $M = 1$ ).

Fig. 8 and 9 show the performance of algorithms for different service rate  $R$ . At lower service rates, SCO-CA connects almost the same number of devices as SDA with perfect CSI, connecting on an average 89.1% of the total number of devices with  $M = 2$  and 78.6% of the devices with

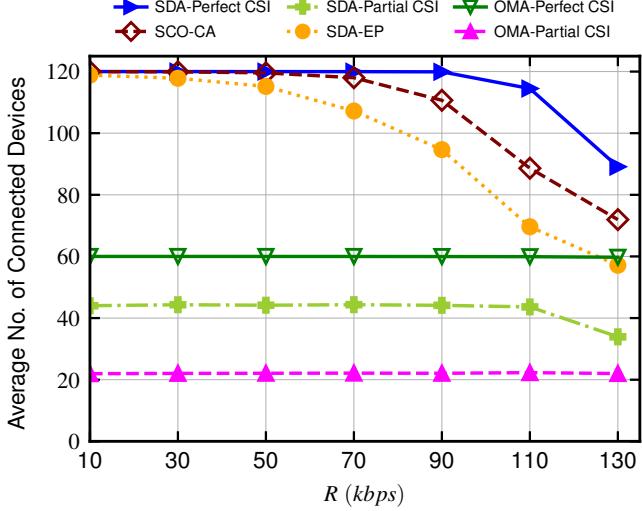


Fig. 8. Connectivity performance with varying service rate and  $M = 2$  for 5 PRBs.

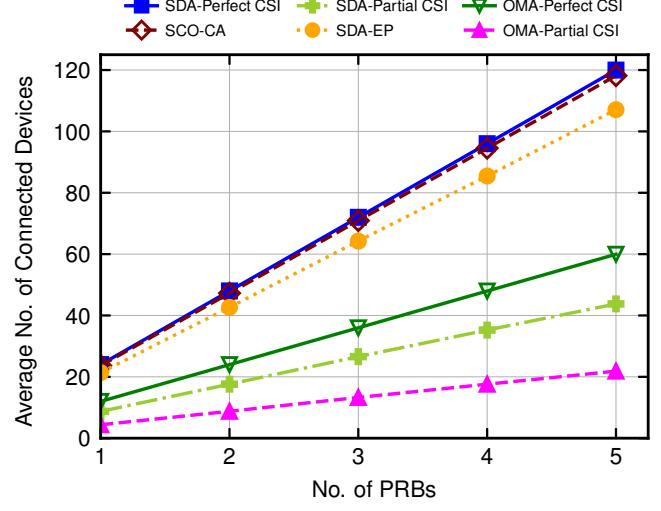


Fig. 10. Efficiency of various algorithms with 2 device classes and  $M = 2$ .

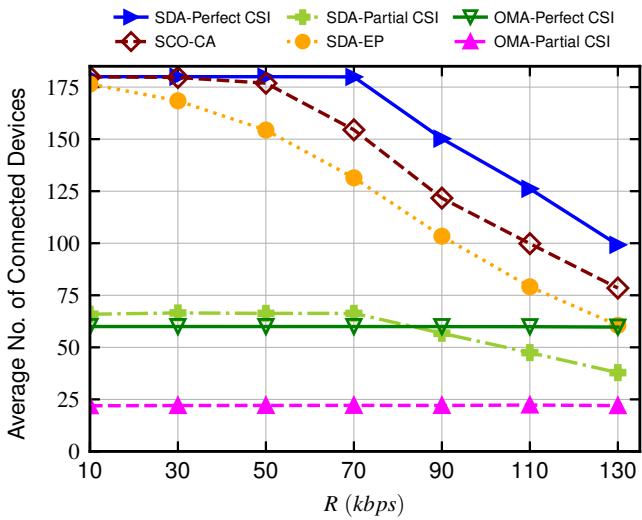


Fig. 9. Connectivity performance with varying service rate and  $M = 3$  for 5 PRBs.

$M = 3$ . The average number of devices connected by SCO-CA compared to OMA with perfect CSI is 1.78 times for  $M = 2$  and 2.36 times for  $M = 3$ . As expected, the connectivity for all algorithms decreases with the increase of  $R$ . Meanwhile, SCO-CA has steady superior performance compared to the other algorithms. Note that SDA does not perform well with partial CSI unless perfect CSI is provided. SDA-EP can connect 81% of the devices with 1.62 times the connected devices of OMA with perfect CSI when  $M = 2$  and 69.3% of the devices with 2.08 times the connected devices of OMA with perfect CSI when  $M = 3$ .

Furthermore, we perform an analysis when the devices have different QoS requirements. We divide the devices into different classes based on their service rate requirement. We consider two service classes with  $M = 2$ . Class-1 devices

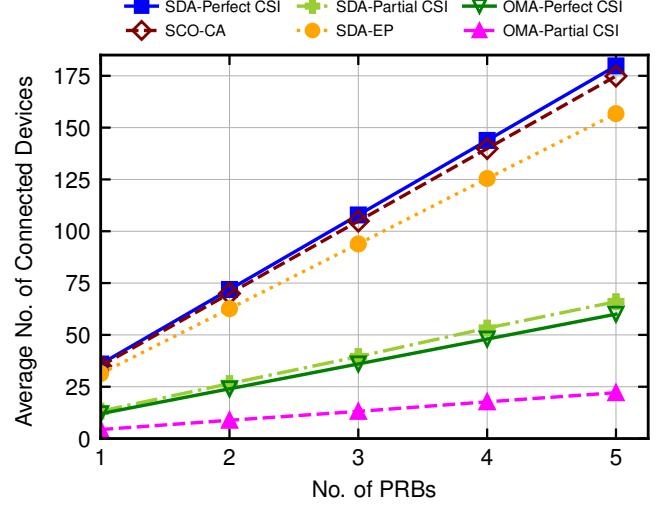


Fig. 11. Efficiency of various algorithms with 3 device classes and  $M = 3$ .

refer to devices with lower QoS where  $R$  is set to 50 kbps, whereas Class-2 devices correspond to devices with higher QoS where  $R$  is set to 100 kbps. Fig. 10 shows the connectivity performance. SCO-CA connects 98.4% of the devices while SDA-EP connects 89%. SDA with perfect CSI connects 99.9% devices.

We extend the above analysis with  $M = 3$  and consider three service classes. We define Class-1, Class-2 and Class-3 devices with  $R$  set to 25, 50 and 100 kbps, respectively. Fig. 11 shows the connectivity performance. SCO-CA can connect 97.1% of the devices, giving 2.91 times the connected devices as that of OMA with perfect CSI. SDA-EP connects 87% devices, while SDA with partial CSI connects only 36.7% devices.

A key characteristic of 5G is deployment flexibility, where connectivity must be provided over very small coverage regions like femtocells and also over very large regions like rural macro cells. Fig. 12 shows the performance for different cell

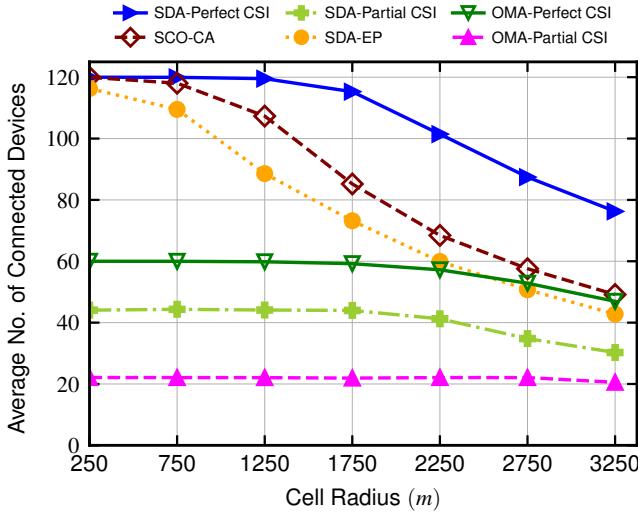


Fig. 12. Effect of cell radius on the connectivity performance of various algorithms with  $M = 2$ .

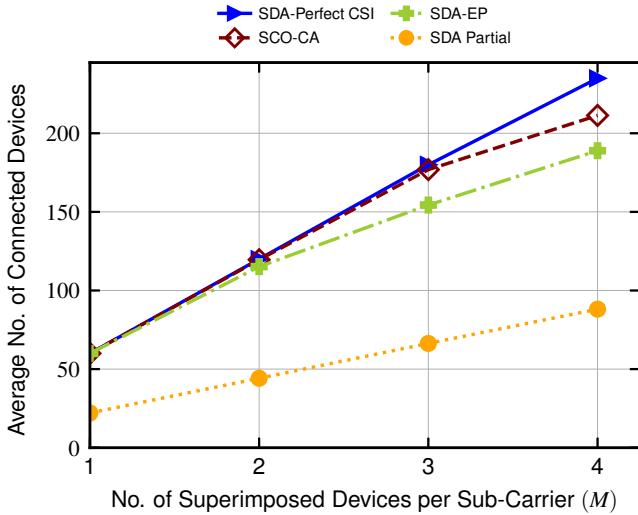


Fig. 13. Performance of algorithms with variation in the system SIC constraint  $M$ .

size. We deploy the same number of devices, i.e., 120 when considering  $M = 2$ , for each cell radius with the service rate  $R = 50$  kbps. The number of connected devices decrease with the increase of cell radius due to the corresponding increase in the distance-dependent path loss. SCO-CA connects 0.79 times the devices as that of SDA (with perfect CSI) on average over the different cell radius, whereas SDA-EP connects 0.71 times the devices in comparison. SCO-CA connects 2.09 times the number of devices compared to SDA with only partial CSI.

The next result, shown in Fig. 13, presents the performance of all the algorithms as we increase the number of superimposed devices per sub-carrier, i.e., the system SIC constraint  $M$ . The service rate is set to  $R = 50$  kbps and we consider device allocation for 5 PRBs.  $M = 1$  represents the case of OMA, where 5 PRBs imply a possible maximum of 60

connected devices. We see that the increase in the number of connected devices for all algorithms gradually decreases as  $M$  increases. This is so because each additional layer of devices faces significant increase in interference due to the previously allocated devices, especially at higher layers of SIC, which in turn increases their power requirement. SDA with perfect CSI connects 99.4% of devices, averaged over all the values of  $M$ . This is followed by SCO-CA connecting 96.5% devices and SDA-EP connecting 90.1% devices. SDA with partial CSI connects only 36.8% devices on an average. This enforces the fact that a proper sub-carrier and power allocation strategy, like SCO-CA, is crucial to ensure connectivity in a practical scenario when only partial CSI is available.

## VI. CONCLUSION

In this work, we study the problem of connection density maximization in the downlink for NB-IoT networks. First, we analyze this problem assuming that perfect CSI is available at the BS and present the SDA algorithm. We prove that the sub-carrier and power allocation through SDA is optimal for maximizing the device connectivity when perfect CSI for each device is known. Since it is not possible to obtain perfect CSI without incurring severe overheads in practice, we formulate the connectivity maximization problem given partial CSI and develop the stochastic connectivity optimization (SCO) framework. We show that SCO is a non-convex mixed integer problem, hence it is difficult to solve through standard optimization methods. We first develop the SDA-EP method, which is a heuristic improvement to the SDA algorithm when only partial CSI is available. We analyze how the excess power parameters of SDA-EP ( $c$  and  $\delta$ ) affect the connectivity performance. Secondly, we propose a concave approximation of the SCO framework, called SCO-CA and solve it through a recursive approach using standard convex optimization solver. We prove that SCO-CA has a performance loss of at most  $e^{-4} \times S$  compared to any possible optimal solution, where  $S$  is the number of sub-carriers in the system. Through computer simulations, we compare the performance of all the developed algorithms under different service rate, cell size and QoS requirement. In all the scenarios, SCO-CA steadily outperforms the other algorithms when only partial CSI is available. Moreover, the performance achieved by SCO-CA under partial CSI is close to the optimal connectivity by SDA assuming perfect CSI especially at low service rates that are typical in NB-IoT deployments.

## ACKNOWLEDGMENT

The authors would like to thank Shane He (Nokia Bell Lab) and Prof K. Giridhar (IIT Madras) for their continuous support to this work and thank the editor and reviewers for their invaluable comments and suggestions, which have been of significant help for improving the quality of this paper. A part of the work has been carried out at LINCS ([www.lincs.fr](http://www.lincs.fr)).

## APPENDIX A PROOF OF THEOREM 1

The devices powers on sub-carrier  $n$  can be computed iteratively from  $i = 1$  to  $D_n$  using (5). For all  $i \in \{1, \dots, D_n\}$ , we have:

$$p_{d_n(i)} = \zeta N \left\{ \sum_{j=1}^{i-1} \frac{\zeta}{g_{d_n(j)}} (\zeta + 1)^{i-j-1} + \frac{1}{g_{d_n(i)}} \right\}.$$

By the above, the power consumed by all the devices on sub-carrier  $n$  can be evaluated as follows:

$$\begin{aligned} p_{\mathbf{d}_n} &= \sum_{l=1}^{D_n} p_{d_n(l)} \\ &= \sum_{l=1}^{D_n} \frac{\zeta N}{g_{d_n(l)}} (\zeta + 1)^{D_n-l}. \end{aligned} \quad (21)$$

Now, the total power taken up by all the devices is simply the sum of the power on each of the sub-carriers, i.e.,

$$\mathbf{p} = \sum_{n=1}^S p_{\mathbf{d}_n}. \quad (22)$$

### Remark 1.

*From (21), we see that the fraction of power associated with a particular device on sub-carrier  $n$  depends only on the layer of SIC  $l$  to which that device is assigned. In particular, it does not depend on the other devices assigned to the same sub-carrier  $n$ . This can be seen from the fact that the  $l$ -th term of the sum only depends on  $g_{d_n(l)}$ , for any  $l \in \{1, \dots, D_n\}$ . In other words, a device can be moved from one sub-carrier to another without changing the minimum required total power as computed in (22), as long as it stays on the same layer  $l$ .*

### Lemma 1 (Optimal Allocation Characterization).

*Assuming that there is no power constraint ( $P_{\max} \rightarrow \infty$  in C1) and SIC limitation ( $M \rightarrow \infty$  in C3). Let  $\mathcal{D}_1, \dots, \mathcal{D}_S$  be a valid allocation of devices  $\mathcal{D}$  on sub-carriers  $S$ . The following two statements are equivalent:*

- (A)  $\mathcal{D}_1, \dots, \mathcal{D}_S$  is an optimal allocation, i.e., it consumes the least power among all valid allocations.
- (B) For any  $l \in \{1, \dots, \lceil D/N \rceil\}$  and any device  $i \in \mathcal{D} \cap \{(l-1)S+1, \dots, (l-1)S+S\}$ ,  $i$  is allocated to the  $l$ -th layer of a sub-carrier, i.e., there exists  $n \in S$  such that  $d_n(l) = i$ .

*Proof.* For a given device allocation, the minimum required total power can be computed using (21) and (22). Let us assume by contradiction that there exists a device allocation for which statement (A) is true and statement (B) is false. Then, one of the following statements is true for this device configuration:

- (1) There exists a device  $i$  in  $\mathcal{D} \cap \{(l-1)S+1, \dots, (l-1)S+S\}$  which is assigned to layer  $l' < l$ , or
- (2) There exists a device  $i$  in  $\mathcal{D} \cap \{(l-1)S+1, \dots, (l-1)S+S\}$  which is assigned to layer  $l' > l$ .

For the first case (1), since device  $i$  is taking one slot in the first  $l'$  layers, there is necessarily one device  $j \in \mathcal{D} \cap \{(k-1)S+1, \dots, (k-1)S+S\}$ , where  $k \leq l'$ , that is allocated to a higher layer  $k' > l'$ . As stated in Remark 1, we can

assume, without loss of generality, that  $i$  and  $j$  are assigned to the same sub-carrier  $n$ . Let  $p_{\mathbf{d}_n}$  be the minimum required power on sub-carrier  $n$  as shown in (21). Now, we swap these two devices, i.e., put device  $i$  at layer  $k'$  and  $j$  at layer  $l'$ , on sub-carrier  $n$ . The power on sub-carrier  $n$  after this device exchange is denoted by  $\hat{p}_{\mathbf{d}_n}$ . The difference between the two configurations is:

$$p_{\mathbf{d}_n} - \hat{p}_{\mathbf{d}_n} = p_{d_n(i)} + p_{d_n(j)} - \hat{p}_{d_n(i)} - \hat{p}_{d_n(j)},$$

where

$$\begin{aligned} p_{d_n(i)} &= \frac{\zeta N}{g_{d_n(i)}} (\zeta + 1)^{D_n-l'}, & p_{d_n(j)} &= \frac{\zeta N}{g_{d_n(j)}} (\zeta + 1)^{D_n-k'}, \\ \hat{p}_{d_n(i)} &= \frac{\zeta N}{g_{d_n(i)}} (\zeta + 1)^{D_n-k'}, & \hat{p}_{d_n(j)} &= \frac{\zeta N}{g_{d_n(j)}} (\zeta + 1)^{D_n-l'}. \end{aligned}$$

Since by construction,  $g_{d_n(i)} < g_{d_n(j)}$  and  $l' < k'$ , the following holds true:

$$p_{\mathbf{d}_n} - \hat{p}_{\mathbf{d}_n} = \frac{\hat{t}}{g_{d_n(i)}} - \frac{\hat{t}}{g_{d_n(j)}} > 0, \quad (23)$$

where  $\hat{t} = \zeta N ((\zeta + 1)^{D_n-l'} - (\zeta + 1)^{D_n-k'})$  is a constant greater than zero. We infer that  $\hat{p}_{\mathbf{d}_n} < p_{\mathbf{d}_n}$ . Since only sub-carrier  $n$  is affected by this change, we deduce that  $\hat{p} < p$ . The same result holds true for the second case (2) and this can be shown following similar reasoning as the previous case with the exception that here  $g_{d_n(i)} > g_{d_n(j)}$ .

Hence, by (23), the new allocation that we constructed by swapping devices  $i$  and  $j$  consumes lesser power than the optimal allocation. This is a contradiction, thus we deduce that statement (A) implies statement (B). In addition, we observe from Remark 1 that all allocations satisfying (B) consume the same amount of power. As a consequence, all allocations satisfying (B) are also optimal, i.e., (B) implies (A), which concludes the proof.  $\square$

## LEMMA 2 (Connectivity Maximization through SDA).

*SDA solves optimally the downlink connectivity maximization problem  $\mathcal{P}$  with perfect CSI.*

*Proof.* Let  $\mathcal{D}^* \subseteq \mathcal{D}$  be the subset of devices allocated at termination of SDA. By construction,  $\mathcal{D}^*$  contains the  $D^*$  devices with the highest channel gains. Due to the three termination conditions (run out of power budget, all devices are already served, all sub-carriers have been allocated  $M$  devices),  $\mathcal{D}^*$  is the largest set of allocated devices that does not violate C1 or C3. In addition, we see that at each iteration of SDA, the subset of devices already allocated satisfies statement (B) in Lemma 1. Hence, according to Lemma 1, SDA is the optimal solution of  $\mathcal{P}$ .  $\square$

## APPENDIX B PROOF OF THEOREM 2

For  $\hat{Z}_n(\mathbf{x}, \mathbf{d}_n)$  in (14) to be concave, the following expression should hold true for all  $\mathbf{x}$  and  $\mathbf{x}'$  in its domain:

$$\hat{Z}_n((1-\lambda)\mathbf{x} + \lambda\mathbf{x}', \mathbf{d}_n) \geq (1-\lambda)\hat{Z}_n(\mathbf{x}, \mathbf{d}_n) + \lambda\hat{Z}_n(\mathbf{x}', \mathbf{d}_n), \quad (24)$$

where  $0 \leq \lambda \leq 1$ . We will prove that the following values, with  $\lambda = 0.5$ , violate the above concavity condition:

$$\mathbf{x} = \{a_{d_n(1)}/4, a_{d_n(2)}/4, \dots, a_{d_n(D_n)}/4\}, \quad (25)$$

$$\mathbf{x}' = \{a_{d_n(1)}/2, a_{d_n(2)}/2, \dots, a_{d_n(D_n)}/2\}. \quad (26)$$

Since the path-loss values are sorted in decreasing order according to (17), we deduce that  $a_{d_n(1)} \leq a_{d_n(2)} \leq \dots \leq a_{d_n(D_n)}$ . As a consequence, the expression in (14) reduces to the following:

$$\begin{aligned} \hat{Z}_n(\mathbf{x}, \mathbf{d}_n) &= \sum_{i=1}^{D_n} \exp\left(-\frac{a_{d_n(i)}}{x_{d_n(i)}}\right), \\ \hat{Z}_n(\mathbf{x}', \mathbf{d}_n) &= \sum_{i=1}^{D_n} \exp\left(-\frac{a_{d_n(i)}}{x'_{d_n(i)}}\right). \end{aligned} \quad (27)$$

Substituting (25)-(27) to (24), we get the following:

$$\begin{aligned} \hat{Z}_n((1-\lambda)\mathbf{x} + \lambda\mathbf{x}', \mathbf{d}_n) &= e^{-8/3} D_n \approx 0.069 \times D_n, \\ (1-\lambda)\hat{Z}_n(\mathbf{x}, \mathbf{d}_n) + \lambda\hat{Z}_n(\mathbf{x}', \mathbf{d}_n) &= \frac{(e^{-2} + e^{-4})D_n}{2} \approx 0.077 \times D_n. \end{aligned}$$

Hence, condition (24) does not hold, and we conclude that  $\hat{Z}_n(\mathbf{x}_n, \mathbf{d}_n)$  is not a globally concave function of  $\mathbf{x}_n$ .

## APPENDIX C PROOF OF THEOREM 3

We define the difference between the original function  $\hat{Z}$  and its approximation  $\tilde{Z}$  on sub-carrier  $n$  to be  $\delta$ , that is:

$$\delta_{\mathbf{d}_n} = \hat{Z}_n(\mathbf{x}_n, \mathbf{d}_n) - \tilde{Z}_n(\mathbf{x}_n, \mathbf{d}_n).$$

From (14), we can simplify  $\delta$  as follows:

$$\delta_{\mathbf{d}_n} = \begin{cases} 0 & x_{d_n(j)} > \frac{a_{d_n(i)}}{2}, \\ \exp\left(-\frac{a_{d_n(i)}}{x_{d_n(j)}}\right) - \mathcal{L}_a(x_{d_n(j)}) & \frac{a_{d_n(i)}}{4} \leq x_{d_n(j)} \leq \frac{a_{d_n(i)}}{2}, \\ \exp\left(-\frac{a_{d_n(i)}}{x_{d_n(j)}}\right) & x_{d_n(j)} < \frac{a_{d_n(i)}}{4}. \end{cases}$$

By evaluating the first derivative of  $\delta_{\mathbf{d}_n}$ , we see that it is a decreasing function in the range  $\frac{a_{d_n(i)}}{4} \leq x_{d_n(j)} \leq \frac{a_{d_n(i)}}{2}$ , so the local maximum occurs at  $x_{d_n(j)} = \frac{a_{d_n(i)}}{4}$ .

$$\frac{d\delta_{\mathbf{d}_n}}{dx_{d_n(j)}} = \frac{a_{d_n(i)}}{x_{d_n(j)}^2} \exp\left(-\frac{a_{d_n(i)}}{x_{d_n(j)}}\right) - \frac{4e^{-2}}{a_{d_n(i)}}$$

In the range  $0 \leq x_{d_n(j)} \leq \frac{a_{d_n(i)}}{4}$ ,  $\delta$  is an increasing function and the local maximum occurs at  $x_{d_n(j)} = \lim_{\epsilon \rightarrow 0} \frac{a_{d_n(i)}}{4} - \epsilon$ .

$$\frac{d\delta_{\mathbf{d}_n}}{dx_{d_n(j)}} = \frac{a_{d_n(i)}}{x_{d_n(j)}^2} \exp\left(-\frac{a_{d_n(i)}}{x_{d_n(j)}}\right)$$

Thus, we conclude that the maximum value of  $\delta_{\mathbf{d}_n}$  occurs at  $x_{d_n(j)} = \frac{a_{d_n(i)}}{4}$ . At this point, the difference between the original function and the approximated function is the largest and has a value  $\delta_{\mathbf{d}_n} = e^{-4}$ . This shows the maximum difference in the expected number of connected devices for one sub-carrier, hence over all sub-carriers we see that this difference is upper bounded by  $\sum_{n=1}^S \delta_{\mathbf{d}_n} = e^{-4} \times S$ .

## APPENDIX D PROOF OF THEOREM 4

Algorithm 3 first calls Algorithm 2 at line 2 to perform SCO device allocation. Therefore, we first study the complexity of Algorithm 2. The device allocation happening at line 5 is inside two nested for loops at lines 2-3, which run for a maximum of  $M$  and  $S$  times respectively. The device allocation also lies nested inside the if condition at line 4 that terminates if the number of allocated devices exceeds  $|\mathcal{D}'|$ , which takes a maximum value of  $D$ . Thus, the computational complexity of Algorithm 2 is  $O(\min(D, MS))$ .

Now let us evaluate the complexity of Algorithm 3. The algorithm consists in iterating over the set of devices  $\mathcal{D}'$ , see the while loop at line 3. Initially,  $\mathcal{D}'$  starts with  $\min(D, MS)$  devices, then a device is removed at each iteration as shown at line 10, until there is no device left. At each iteration, a solution  $Z^*$  is computed at line 4 using either the gradient descent or interior point method. According to [25, Chapters 9 and 11] and given that the objective function  $\tilde{Z}$  is strongly concave, the aforementioned convex solvers converge in  $O(\log(1/\epsilon))$  steps, where  $\epsilon$  is the desired solution accuracy. Each step requires evaluating the objective function, its gradient and Hessian with complexity  $O(|\mathcal{D}'|M)$ .

To sum up, the complexity of Algorithm 3 is:

$$O\left(\min(D, MS) + \sum_{D'=1}^{\min(D, MS)} D'M \log(1/\epsilon)\right). \quad (28)$$

The first term  $O(\min(D, MS))$  comes from the call of Algorithm 2 at line 2. The second term contains the number of convex solver's steps  $O(\log(1/\epsilon))$  multiplied by the complexity of the function evaluations  $O(D'M)$ . This complexity is then summed over all the explored  $D' = 1, \dots, \min(D, MS)$  in the while loop at line 3. The asymptotic complexity (28) can be further simplified as  $O(\min(D, MS)^2 M \log(1/\epsilon))$ .

## REFERENCES

- [1] 3GPP, "Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 45.820, Dec. 2015, version 13.1.0. [Online]. Available: <http://www.3gpp.org/DynaReport/45820.htm>
- [2] Y. 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, 2017.
- [3] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.101, Oct. 2019, version 16.3.0. [Online]. Available: <http://www.3gpp.org/DynaReport/36101.htm>
- [4] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, vol. E98.B, no. 3, pp. 403–414, 2015.
- [5] Y. Fu, L. Salaiün, C. W. Sung, and C. S. Chen, "Subcarrier and power allocation for the downlink of multicarrier NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11 833–11 847, 2018.
- [6] 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, 2016.
- [7] P. Liu, K. Wu, J. Liang, J. Chen, and Y. Tseng, "Energy-efficient uplink scheduling for ultra-reliable communications in NB-IoT networks," in *IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2018.

- [8] L. Salaün, M. Coupechoux, and C. S. Chen, "Joint subcarrier and power allocation in NOMA: Optimal and approximate algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 2215–2230, 2020.
- [9] 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 Things J.*, vol. 5, no. 3, pp. 2120–2129, 2018.
- [10] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance of downlink non-orthogonal multiple access (NOMA) under various environments," in *IEEE VTC-Spring*, 2015.
- [11] Z. Zhang, Y. Hou, Q. Wang, and X. Tao, "Joint sub-carrier and transmission power allocation for MTC under power-domain NOMA," in *IEEE ICC Workshops*, 2018.
- [12] J. G. Andrews and T. H. Meng, "Optimum power control for successive interference cancellation with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 375–383, 2003.
- [13] L. Zhu, J. Zhang, Z. Xiao, X. Cao, and D. O. Wu, "Optimal user pairing for downlink non-orthogonal multiple access (NOMA)," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 328–331, 2019.
- [14] M. B. Shahab, M. F. Kader, and S. Y. Shin, "A virtual user pairing scheme to optimally utilize the spectrum of unpaired users in non-orthogonal multiple access," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1766–1770, 2016.
- [15] Y. Fu, M. Zhang, L. Salaün, C. W. Sung, and C. S. Chen, "Zero-forcing oriented power minimization for multi-cell miso-noma systems: A joint user grouping, beamforming, and power control perspective," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1925–1940, 2020.
- [16] A. Ikpehai, B. Adebisi, K. M. Rabie, K. Anoh, R. E. Ande, M. Hammoudeh, H. Gacanin, and U. M. Mbanaso, "Low-power wide area network technologies for Internet-of-Things: A comparative review," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2225–2240, 2019.
- [17] D. Zhai, R. Zhang, L. Cai, B. Li, and Y. Jiang, "Energy-efficient user scheduling and power allocation for NOMA-based wireless networks with massive IoT devices," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1857–1868, 2018.
- [18] W. Chen, H. Zhang, H. Ji, and X. Li, "Joint QoS-aware downlink and resource allocation for throughput maximization in narrow-band IoT with NOMA," in *Commun. and Netw.*, X. Liu, D. Cheng, and L. Jinfeng, Eds. Springer International Publishing, 2019, pp. 3–13.
- [19] B. Di, S. Bayat, L. Song, and Y. Li, "Radio resource allocation for downlink non-orthogonal multiple access (NOMA) networks using matching theory," in *IEEE GLOBECOM*, 2015, pp. 1–6.
- [20] A. Shahini and N. Ansari, "NOMA aided narrowband IoT for machine type communications with user clustering," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7183–7191, 2019.
- [21] 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, 2019.
- [22] T. Lv, Y. Ma, J. Zeng, and P. T. Mathiopoulos, "Millimeter-wave NOMA transmission in cellular M2M communications for Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1989–2000, 2018.
- [23] S. Mishra, L. Salaün, and C. S. Chen, "Maximizing connection density in NB-IoT networks with NOMA," in *IEEE VTC-Spring*, 2020, pp. 1–6.
- [24] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. USA: Cambridge University Press, 2005.
- [25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [26] M. S. Andersen, J. Dahl, and L. Vandenberghe, "CVXOPT: A python package for convex optimization," Available at <https://cvxopt.org/>, 2013, version 1.1.6.