Dynamic Resource Allocation for Uplink MIMO NOMA VWN with Imperfect SIC

Daniel Tweed and Tho Le-Ngoc
Department of Electrical & Computer Engineering, McGill University, Montreal, Canada
daniel.tweed@mail.mcgill.ca; tho.le-ngoc@mcgill.ca

***Abstract*—We investigate the uplink resource allocation problem for virtualized wireless networks (VWNs) supported by multiple-input multiple-output (MIMO) non-orthogonal multiple access (NOMA) and present a sensitivity analysis to imperfect successive interference cancellation (SIC) and various system parameters. The proposed algorithm for power and sub-carrier allocation is derived from the non-convex power minimization subject to rate and sub-carrier reservations, for which an optimal solution is NP-hard. To develop an efficient solution, the resource allocation is decomposed into separate power and sub-carrier allocation problems and an iterative algorithm based on successive convex approximation and complementary geometric programming is proposed. Simulation results demonstrate that compared to orthogonal multiple access, the proposed algorithm for MIMO NOMA can offer significant improvement in spectrum and power efficiency.**

***Index Terms*—Non-orthogonal multiple access, wireless virtualization, dynamic resource allocation, MIMO.**

I. Introduction

Supporting the exponential growth rate in mobile network traffic and diverse use-cases expected for fifth generation (5G) cellular networks requires significant increases in spectral and power efficiency, traffic density, and overall network utility [1]. Non-orthogonal multiple access (NOMA) and virtualized wireless networks (VWN) are two promising techniques positioned to meet the targets set for 5G and beyond by maximizing spectrum and network utility through resource sharing and reuse across service providers and users [2], [3]. For both NOMA and VWN, the resource allocation problem is inherently complex and dynamic resource allocation is essential to achieve the potential gains from these techniques.

In contrast to orthogonal multiple access (OMA), under NOMA sub-carriers can be shared by several users concurrently with successive interference cancellation (SIC) applied at the receiver [2]. When proposed for downlink (DL) transmissions in [4], NOMA demonstrated gains in spectral and power efficiency compared to OMA, and also maintained those gains as the number of users was increased. However, the complexity of performing SIC was seen as a limitation, especially for power and processing limited user equipment (UE). In uplink (UL) NOMA, the complexity of performing SIC is with the base station (BS), which has resources to perform SIC for more users. As a result, recently UL NOMA has received more attention. In [5], a dynamic power allocation for two-user UL and DL NOMA is derived for a fixed sub-carrier assignment such that achieved user rates are improved compared to OMA. [6] proposes an UL power control scheme for NOMA where power is allocated to users in discrete steps based on rank in the channel. In [7], a linear beamforming technique for DL multiple-input multiple-output (MIMO) NOMA is proposed with user grouping based on spatial proximity. In [8], proof that there exists a power allocation for which the rate gap between MIMO NOMA and MIMO OMA exists is presented. [9] proposes new precoding and detection matrices and examines user pairing, cognitive radio inspired power allocations, and derives closed-form expressions of the rate gap between NOMA and OMA in MIMO systems.

The use of VWN in cellular networks has been of increasing interest and virtualization is seen as the preferred end-to-end network architecture for 5G networks [10]. In VWN, groups of users called slices access a share of network resources to meet the needs of their users. Meeting these needs without affecting the quality of service (QoS) received by other slices is particularly challenging over time-varying wireless channels with mobile dynamic users. Many dynamic resource allocation algorithm have been proposed for VWN in LTE-A networks. For example, [11] proposes a framework for VWN based on a Vickey-Clarke-Groves auction. In [12], a resource allocation algorithm for software defined heterogeneous cellular networks based on quantum-behaved particle swarm optimization is presented. An asymptotically optimal solution to bandwidth provisioning in VWN is derived in [13]. [14] proposes a joint user association and resource allocation in multi-cell VWN. The application of DL NOMA to support VWNs is investigated in [15] with the assumption of perfect SIC and no restriction on the number of users per sub-carrier. In [16], we proposed an iterative resource allocation algorithm for UL NOMA over SISO channels with imperfect SIC based on successive convex approximation (SCA) and complementary geometric programming (CGP).

In this work, we extend the work in [16] to UL MIMO NOMA systems supporting VWN networks with power-limited devices requiring strict QoS. We first define an equivalent channel gain which can be used for user grouping and decoding order in MIMO NOMA systems and then formulate the resource allocation problem as minimizing UE transmit power under slice and network constraints to derive an efficient iterative joint resource allocation algorithm based on SCA and CGP. Simulation results validate the proposed algorithm provides better performance for MIMO NOMA compared to MIMO OMA in terms of both required user transmit power and the ability to support more users over fewer sub-carriers.

The remainder of this paper is organized as follows: Section II presents the system model and problem formulation. The derivation and definition of the proposed algorithm is provided in Section III. Section IV presents simulation results and analysis, followed by concluding remarks in Section V.

This work was partially supported by Huawei Technologies Canada and the Natural Sciences and Engineering Research Council of Canada (NSERC).

978-1-5386-3180-5/18/$31.00 ©2018 IEEE

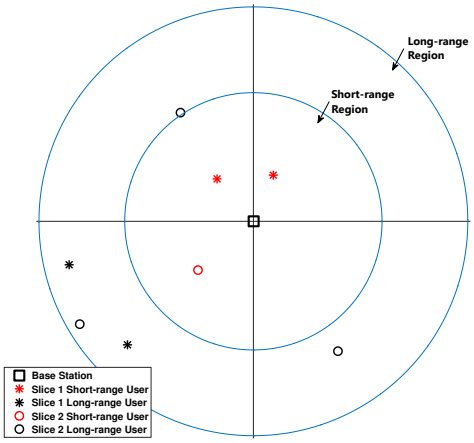


Fig. 1: Illustration of base station coverage area

II. System and Signal Models

Consider a BS equipped with M_r receive antennas serving a set of slices $\mathcal{S}$ where each slice $s \in \mathcal{S}$ has negotiated QoS as a minimum reserved slice rate $R_s^{\mathbf{rsv}}$ and a minimum reserved number of sub-carriers $N_s^{\mathbf{rsv}}$ to meet the needs of its users. For each slice, the set of users is $\mathcal{K}_s = \{1, \ldots, K_s\}$ and each user k_s is equipped with M_{t,k_s} transmit antennas and are randomly placed within the BS coverage area, as shown in Fig. 1 for two slices. The total number of users in the system is $K = \sum_{s \in \mathcal{S}} K_s$ and each of the available sub-carriers, $n \in \mathcal{N} = \{1, \ldots, N\}$, can be shared by at most $N_{\mathbf{max}}$ users.

The signals transmitted by each user can be expressed

$$\mathbf{x}_{k_s,n} = \mathbf{P}_{k_s} \tilde{\mathbf{x}}_{k_s,n}, \tag{1}$$

where $\tilde{\mathbf{x}}_{k_s,n} = \sqrt{\boldsymbol{\beta}_{k_s,n}} \mathbf{s}_{k_s,n} \in \mathbb{C}^{M_{t,k_s} \times 1}$ is the transmitted data vector, $\mathbf{s}_{k_s,n}$ the information bearing signal transmitted by user k_s on sub-carrier n with transmission power coefficients $\boldsymbol{\beta}_{k_s,n} = [\beta_{k_s,n,1}, \beta_{k_s,n,2}, \ldots \beta_{k_s,n,M_{t,k_s}}]^T$, and $\mathbf{P}_{k_s} \in \mathbb{C}^{M_{t,k_s} \times M_{t,k_s}}$ used for precoding at the transmitter.

For UL NOMA, users are grouped on sub-carriers and SIC is performed to resolve each of the superimposed transmissions received at the BS. To facilitate this, in each sub-carrier n shared by K users, they can be ranked according to their experienced channel gains to determine the decoding order from strongest $i = 1$ to weakest $i = K$ received signal strength. For MIMO systems, in general, user k_s with its $M_{t,k_s} \geq 1$ antennas can transmit $x \geq 1$ parallel independent data streams, each data stream using n_y antennas, where $\sum_{y=1}^{x} n_y \leq M_{t,k_s}$. Unfortunately, such a general approach will require impractically complex user grouping and decoding ordering operations. For simplicity, we consider the approach in which user k_s uses its $M_{t,k_s} \geq 1$ antennas to transmit either: (i) Just 1 data stream to achieve the maximum diversity gain when it is in the long-range region, as shown in Fig. 1, with relatively low average channel power gain; or (ii) M_{t,k_s} independent parallel data streams to achieve the maximum multiplexing gain, when it is in the short-range region with relatively higher average channel power gain. The multiple parallel independent data streams generated by short-range users must be treated independently in resource allocation, subject to the constraint that total power across all antennas is within the maximum power available to the user and that sub-carrier assignment groups these data streams appropriately.

Without loss of generality, let us consider the user ranked i in the decoding process. Signals from users ranked lower than i are removed by SIC and signals from users ranked higher than i are treated as unresolvable interference. Let $\mathbf{G}_{i,n} \in \mathbb{C}^{M_r \times M_{t,i}}$ be the MIMO channel matrix for user i and $\mathbf{V}_{i,n} \in \mathbb{C}^{M_r \times M_r}$ be the matrix used for receiver shaping and detection for this user at the BS. Thus, the observation at the BS for user i can be expressed

$$\mathbf{y}_{i,n} = \mathbf{V}_{i,n}\mathbf{G}_{i,n}\mathbf{x}_{i,n} + I^r_{i,n} + I^e_{i,n} + \mathbf{V}_{i,n}\mathbf{z}_{i,n}, \tag{2}$$

where $\mathbf{z}_{i,n}$ is an additive white Gaussian noise (AWGN) vector, $I^r_{i,n}$ is the interference from users which are ranked higher than i and not removed by SIC, defined as

$$I^r_{i,n} = \sum_{j=i+1}^{K} \mathbf{V}_{i,n}\mathbf{G}_{j,n}\mathbf{x}_{j,n}, \tag{3}$$

and $I^e_{i,n}$ is residual interference from cancellation errors in performing SIC to remove signals from users ranked lower then i, which we define as

$$I^e_{i,n} = \sum_{j=1}^{i-1} ||\mathbf{x}_{j,n} - \hat{\mathbf{x}}_{j,n}|| \mathbf{V}_{i,n}\mathbf{G}_{j,n}\mathbf{x}_{j,n}. \tag{4}$$

Here $||\mathbf{x}_{j,n} - \hat{\mathbf{x}}_{j,n}||$ is residual interfering signal power due to differences between the actual and estimated signal for user j. In the case of perfect SIC, $\hat{\mathbf{x}}_{j,n} = \mathbf{x}_{j,n}$ and then $I^e_{i,n} = 0$ for all users. Otherwise, some portion of the received power remains as interference. The magnitude of SIC error is dependent on the type of SIC employed, the number of signals being cancelled, and channel and user mobility conditions. To account for all sources of error, we define the expected level of cancellation achieved by SIC as $\sigma_e^2 = \mathbf{E}\left[||\mathbf{x}_{j,n} - \hat{\mathbf{x}}_{j,n}||\right]$, e.g. 20 dB cancellation is equivalent to $\sigma_e^2 = 0.01$ [17]. Thus, we have

$$I^e_{i,n} = \sigma_e^2 \sum_{j=1}^{i-1} \mathbf{V}_{i,n}\mathbf{G}_{j,n}\mathbf{P}_j\tilde{\mathbf{x}}_{j,n}. \tag{5}$$

To simplify and facilitate user grouping and decoding, we consider the effective channel gain after detection and combining at the receiver. With careful design of precoding and detection, the MIMO channel can be decomposed into independent parallel channels, reducing the MIMO NOMA system to SISO channels [7], [9], [18]. The design of precoding and detection is beyond the scope of this paper, however, one method to design the required precoding and detection is to take the singular value decomposition of the MIMO channel, from which we obtain

$$\mathbf{G}_{i,n} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}, \tag{6}$$

where the $M_r \times M_r$ matrix $\mathbf{U}$ and the $M_{t,i} \times M_{t,i}$ matrix $\mathbf{V}$ are unitary matrices used for detection and transmit precoding, respectively. [19]. $\boldsymbol{\Sigma}$ is an $M_r \times M_{t,i}$ diagonal matrix of the singular values of $\mathbf{G}_{i,n}$, of which $R_{G_{i,n}} \leq \min\{M_r, M_{t,i}\}$ are non-zero, $\sqrt{\sigma_{r,i,n}}$, $r = 1, 2, .., R_{G_{i,n}}$ and $\sigma_{r,i,n}$'s are the eigenvalues of the matrix $\mathbf{G}^{\mathrm{H}}_{i,n}\mathbf{G}_{i,n}$, representing the power gains of $R_{G_{i,n}}$ parallel virtual SISO channels after transmit precoding and receive separating, and $R_{G_{i,n}}$ is the rank of $\mathbf{G}_{i,n}$. Therefore, user i has either 1 virtual channel with power gain of $\tilde{h}_{i,n} = \sum_{r=1}^{R_{G_{i,n}}} \sigma_{r,i,n}$, if it is in the long-range region, or $R_{G_{i,n}}$ virtual channels with power gains of $\tilde{h}_{i_r,n} = \sigma_{r,i,n}, r = 1, 2, .., R_{G_{i,n}}$, if it is in the short-range

region, as depicted in Fig. 1.

Let K' be the total number of virtual channels of all users in the system for each sub-carrier n. The K' virtual channels gains are sorted in the descending order as follows

$$\tilde{h}_{1,n} > \tilde{h}_{2,n} > \cdots > \tilde{h}_{K',n}.$$

With precoding and detection as defined for user grouping and decoding order purposes, we must consider power allocation per data stream. Letting $\tilde{\beta}_{k_s,n}$ denote the power allocation coefficient for data stream k_s on sub-carrier n we can define power for each data stream as

$$\tilde{\beta}_{k_s,n} = \begin{cases} \beta_{k_s,n,r}, r = 1, 2, .., R_{G_{i,n}} & \text{Short-range users,} \\ \sum_{r=1}^{R_{G_{i,n}}} \beta_{k_s,n,r}, & \text{Long-range users.} \end{cases} \tag{7}$$

Then SINR for each data stream can be written as

$$\tilde{\gamma}_{i,n} = \frac{\tilde{h}_{i,n}\tilde{\beta}_{i,n}}{\mathrm{z}_{i,n} + I^r_{i,n} + I^e_{i,n}}, \tag{8}$$

and the achieved rate per data streams can be expressed

$$R_{i,n} = \log_2\Big(1 + \tilde{\gamma}_{i,n}\Big). \tag{9}$$

Let us define $\alpha_{k_s,n} \in \{0, 1\}$ to be the sub-carrier allocation indicator, where $\alpha_{k_s,n} = 1$ means user k_s is allocated to sub-carrier n and is applied to all users' data streams. Then, defining $\mathcal{K}'_s$ to be the total number of virtual channels over all users of slice s, the sum rate achieved by each slice s is

$$R_s = \sum\nolimits_{k_s \in \mathcal{K}'_s} \sum\nolimits_{n \in \mathcal{N}} \alpha_{k_s,n} R_{i,n}. \tag{10}$$

Finally, we can formalize the constraint on slice rate reservation to ensure QoS and isolation as

$$\text{C1: } R_s \geq R_s^{\mathbf{rsv}}, \qquad \forall s \in \mathcal{S}, \tag{11}$$

and the constraint on slice sub-carrier reservations can be expressed

$$\text{C2: } \sum\nolimits_{n \in \mathcal{N}} \sum\nolimits_{k_s \in \mathcal{K}_s} \alpha_{k_s,n} \geq N_s^{\mathbf{rsv}}, \quad \forall s \in \mathcal{S}. \tag{12}$$

To respect UE transmission power restrictions and limit the complexity of performing SIC at the BS we add two additional constraints

$$\text{C3: } \sum\nolimits_{n \in \mathcal{N}} \sum\nolimits_{m_t=1}^{M_{t,k_s}} \beta_{k_s,n,m_t} \leq P_{\mathbf{max}}, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s, \tag{13}$$

$$\text{C4: } \sum\nolimits_{s \in \mathcal{S}} \sum\nolimits_{k_s \in \mathcal{K}_s} \alpha_{k_s,n} \leq N_{\mathbf{max}}, \forall n \in \mathcal{N}. \tag{14}$$

Finally, to ensure that no power is allocated to sub-carriers which the UE is not assigned, we force the power allocation to zero when the sub-carrier allocation indicator is zero

$$\text{C5: } \sum\nolimits_{m_t=1}^{M_{t,k_s}} \beta_{k_s,n,m_t} - \alpha_{k_s,n} \times P_{\mathbf{max}} \leq 0, \quad \forall n \in \mathcal{N}, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s. \tag{15}$$

Then, the minimum power needed to meet slice reservations within practical system limitations is obtained by solving the following optimization problem

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}'_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n} \tilde{\beta}_{k_s,n} \tag{16}$$

Subject to: C1–5 .

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the $K' \times N$ matrices of $\alpha_{k_s,n}$ sub-carrier allocation indicators and $\tilde{\beta}_{k_s,n}$ user transmit power factors.

III. Proposed Iterative Algorithm

All of C1, C2, and C5 in (16) are non-convex constraints and the complexity induced by the binary variable yields a problem which also NP-hard. To tackle this complexity and efficiently solve the problem, we decompose (16) into separate sub-carrier and power allocation problems and adopt an iterative approach to solve each sub-problem until the overall solution converges, as described in Algorithm 1. In Step 1, for a given power allocation, the sub-carrier allocation problem is solved according to Algorithm 1.1 described in Section III-A. This sub-carrier allocation is then used in Step 2 to solve the power allocation problem according to Algorithm 1.2 discussed in detail in Section III-B.

Algorithm 1 Iterative Power and Sub-carrier Allocation

Initialize: Set $t = 0$, $\boldsymbol{\alpha}^*(0) = [1]_{K \times N}$ and $\boldsymbol{\beta}^*(0) = [{}^{P_{\mathbf{max}}}/_N]_{K \times N}$, $0 < \varepsilon_1, \varepsilon_2 \ll 1$
repeat
$t = t + 1$
Step 1: Derive sub-carrier allocation $\boldsymbol{\alpha}^*(t)$ according to Algorithm 1.1 with input $\boldsymbol{\beta}^*(t-1)$, $\boldsymbol{\alpha}^*(t-1)$
Step 2: Derive power allocation $\boldsymbol{\beta}^*(t)$ according to Algorithm 1.2 with input $\boldsymbol{\alpha}^*(t)$, $\boldsymbol{\beta}^*(t-1)$
until $\|\boldsymbol{\beta}^*(t) - \boldsymbol{\beta}^*(t-1)\| \leq \varepsilon_1$ and $\|\boldsymbol{\alpha}^*(t) - \boldsymbol{\alpha}^*(t-1)\| \leq \varepsilon_2$,

Algorithm 1.1 Sub-carrier Allocation

Require: Power allocation $\boldsymbol{\beta}^*(t-1)$, $\boldsymbol{\alpha}^*(t-1)$
Initialize: Set $t_1 = 1$, $\boldsymbol{\alpha}(0) = \boldsymbol{\alpha}^*(t-1)$
repeat
Step 1: Update $\eta_{k_s,n}(t_1)$ and $\kappa_{k_s,n}(t_1)$ from (20), (22)
Step 2: Derive $\boldsymbol{\alpha}(t_1)$, according to (23) using CVX [22]
until $\|\boldsymbol{\alpha}(t_1) - \boldsymbol{\alpha}(t_1 - 1)\| \leq \epsilon_1$, otherwise set $t_1 = t_1 + 1$
return $\boldsymbol{\alpha}^*(t) = \boldsymbol{\alpha}(t_1)$

Each of the sub-problems remains computationally intractable and non-convex. To solve them efficiently, we use variable relaxation and the technique of SCA. We apply CGP to convert the problems into the geometric programming (GP) form by approximating non-convex constraints with the arithmetic-geometric mean approximation (AGMA) on each iteration until the solution converges. Interested readers on AGMA and CGP are referred to [20], [21], where it has been shown that the output of algorithms of this form converge to a local optimum of the original problem and have very close performance to the optimal solution.

A. Sub-carrier Allocation Sub-problem

Given a fixed power allocation, the residual sub-carrier allocation minimizing user transmit power can be expressed

$$\min_{\boldsymbol{\alpha}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}'_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n} \tilde{\beta}_{k_s,n} \tag{17}$$

Subject to: C1, C2, C4.

While (17) is simpler than (16), each of C1 and C2 remains non-convex. Additionally, the binary variable $\boldsymbol{\alpha}$ leads to a problem which is NP-hard. First, to reduce this complexity we relax the binary variable so that α_k is continuous on the interval $[0, 1]$. Then, for iteration index t_1, we apply CGP and approximate the non-convex constraints via AGMA.

We can re-write C1 as

$$\frac{R_s^{\mathbf{rsv}}}{\sum_{k_s \in \mathcal{K}'_s} \sum_{n \in \mathcal{N}} R_{k_s,n}} \leq 1, \tag{18}$$

Algorithm 1.2 Power Allocation

Require: Sub-carrier allocation $\boldsymbol{\alpha}^*(t)$, $\boldsymbol{\beta}^*(t-1)$
Initialize: Set $t_2 = 1$, $\boldsymbol{\beta}(0) = \boldsymbol{\beta}^*(t-1)$
repeat
Step 1.1: Update $\phi_{i,n}(t_2)$, $\gamma_{j,n}(t_2)$, $\mu_{j,n}(t_2)$, $\rho_{i,n}(t_2)$, and $\zeta_{i,n}(t_2)$ from (28-32)
Step 1.2: Derive $\boldsymbol{\beta}(t_2)$ according to (33) using CVX [22]
until $\|\boldsymbol{\beta}(t_2) - \boldsymbol{\beta}(t_2-1)\| \leq \epsilon_2$, otherwise set $t_2 = t_2 + 1$
return $\boldsymbol{\beta}^*(t) = \boldsymbol{\beta}(t_2)$

which can be approximated for each slice $s \in \mathcal{S}$ by the following convex constraint

$$\widetilde{\mathrm{C1}} = R_s^{\mathbf{rsv}} \times \prod_{k_s \in \mathcal{K}'_s} \prod_{n \in \mathcal{N}} \left(\frac{\alpha_{k_s,n}(t_1) R_{k_s,n}}{\eta_{k_s,n}(t_1)} \right)^{-\eta_{k_s,n}(t_1)} \leq 1, \tag{19}$$

where

$$\eta_{k_s,n}(t_1) = \frac{\alpha_{k_s,n}(t_1-1) R_{k_s,n}}{\sum_{k_s \in \mathcal{K}'_s} \sum_{n \in \mathcal{N}} \alpha_{k_s,n}(t_1-1) R_{k_s,n}} \tag{20}$$

Similarly, we can transform C2 and define the convex constraint

$$\widetilde{\mathrm{C2}} = N_s^{\mathbf{rsv}} \times \prod_{k_s \in \mathcal{K}'_s} \prod_{n \in \mathcal{N}} \left(\frac{\alpha_{k_s,n}(t_1)}{\kappa_{k_s,n}(t_1)} \right)^{-\kappa_{k_s,n}(t_1)} \leq 1, \tag{21}$$

where

$$\kappa_{k_s,n}(t_1) = \frac{\alpha_{k_s,n}(t_1-1)}{\sum_{k_s \in \mathcal{K}'_s} \sum_{n \in \mathcal{N}} \alpha_{k_s,n}(t_1-1)}. \tag{22}$$

Thus, at each iteration t_1 we can transform (17) into the following GP form

$$\min_{\boldsymbol{\alpha}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}'_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n}(t_1) \tilde{\beta}_{k_s,n} \tag{23}$$

$$\text{Subject to: } \widetilde{\mathrm{C1}}, \widetilde{\mathrm{C2}}, \mathrm{C4},$$

as described in Algorithm 1.1.

B. Power Allocation Sub-problem

Given a fixed sub-carrier allocation, the optimization reduces to

$$\min_{\boldsymbol{\beta}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}'_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n} \tilde{\beta}_{k_s,n} \tag{24}$$

$$\text{Subject to: C1, C3, C5.}$$

Here, C3 is in the proper form and C5 is a convex constraint for fixed sub-carrier assignment but C1 remains non-convex due to the presence of the interference terms in the expression for $R_{k_s,n}$.

We can rewrite the rate expression as

$$R_{i,n} = \log_2 \Big(\frac{z_{i,n} + I^r_{i,n} + I^e_{i,n} + \tilde{h}_{i,n} \tilde{\beta}_{i,n}}{z_{i,n} + I^r_{i,n} + I^e_{i,n}} \Big). \tag{25}$$

Removing the logarithm, C1 can then be expressed $\forall s \in \mathcal{S}$ as

$$\prod_{k_s \in \mathcal{K}'_s} \prod_{n \in \mathcal{N}} \Big(\frac{z_{i,n} + I^r_{i,n} + I^e_{i,n}}{z_{i,n} + I^r_{i,n} + I^e_{i,n} + \tilde{h}_{i,n} \tilde{\beta}_{i,n}} \Big)^{\alpha_{k_s,n}} \leq 2^{-R_s^{\mathbf{rsv}}}. \tag{26}$$

Expanding matrix operations and applying AGMA, at iteration index t_2 we can approximate the product terms in (26) with the following convex function

$$\begin{aligned} x_{i,n}(t_2) = {} & \big(z_{i,n} + I^r_{i,n}(t_2) + I^e_{i,n}(t_2) \big) \times \Big(\frac{z_{i,n}}{\phi_{i,n}(t_2)} \Big)^{-\phi_{i,n}(t_2)} \\ & \times \prod_{j=1}^{i-1} \Big(\frac{\tilde{\beta}_{j,n}(t_2) \tilde{h}_{j,n} \sigma_e^2}{\gamma_{j,n}(t_2)} \Big)^{-\gamma_{j,n}(t_2)} \\ & \times \prod_{j=i+1}^{K'} \Big(\frac{\tilde{\beta}_{j,n}(t_2) \tilde{h}_{j,n}}{\mu_{j,n}(t_2)} \Big)^{-\mu_{j,n}(t_2)} \times \Big(\frac{\tilde{\beta}_{i,n}(t_2) \tilde{h}_{i,n}}{\rho_{i,n}(t_2)} \Big)^{-\rho_{i,n}(t_2)}, \end{aligned} \tag{27}$$

where we have

$$\phi_{i,n}(t_2) = z_{i,n} / \zeta_{i,n}(t_2) \tag{28}$$

$$\gamma_{j,n}(t_2) = \tilde{\beta}_{j,n}(t_2 - 1) \tilde{h}_{j,n} \sigma_e^2 / \zeta_{i,n}(t_2), \tag{29}$$

$$\mu_{j,n}(t_2) = \tilde{\beta}_{j,n}(t_2 - 1) \tilde{h}_{j,n} / \zeta_{i,n}(t_2), \tag{30}$$

$$\rho_{i,n}(t_2) = \tilde{\beta}_{i,n}(t_2 - 1) \tilde{h}_{i,n} / \zeta_{i,n}(t_2), \tag{31}$$

$$\zeta_{i,n}(t_2) = z_{i,n} + I^r_{j,n}(t_2 - 1) + I^e_{j,n}(t_2 - 1) + \tilde{h}_{i,n} \tilde{\beta}_{i,n}. \tag{32}$$

Then at each iteration, t_2, we solve

$$\min_{\boldsymbol{\beta}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}'_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n} \tilde{\beta}_{k_s,n}(t_2)$$

$$\text{Subject to: C3, C5}$$

$$\prod_{k_s \in \mathcal{K}'_s} \prod_{n \in \mathcal{N}} [x_{i,n}(t_2)]^{\alpha_{k_s,n}} \leq 2^{-R_s^{\mathbf{rsv}}}, \forall s \in \mathcal{S}, \tag{33}$$

as described in Algorithm 1.2.

C. Convergence Analysis

When CVX is applied to the problems in (23) and (33) an interior point method is used. The number of required iterations to solve by this method is equal to $\frac{\log c/t^0 \delta}{\log(\xi)}$, where c is the total number of constraints, t^0 is the initial point, $0 < \delta \ll 1$ is the stopping criterion, and ξ is used for updating the accuracy of the method [23].

For our formulation, in (23) the total number of constraints is $c_1 = 2S + N$ and for (33) we have $c_2 = S + K + K'N$, where K' is the total number of virtual channels, as defined in Section II. The worst-case number of computations required to convert to the GP form using AGMA in Algorithm 1.1 is $i_1 = 4K'N$. For (33), due to the expressions for $I^r_{i,n}$ and $I^e_{i,n}$, we need to consider the user with the most transmit antenna so we define $M_t^* = \max\{M_{t,k_s} : k_s \in \mathcal{K}_s, s \in \mathcal{S}\}$. Then the worst-case number of computations required for Algorithm 1.2 is $i_2 = (K')^2 N M_r M_t^* + 2K'N + 3K'N M_r M_t^*$.

Therefore, each algorithm has complexity on the order

$$\text{Algorithm 1.j:} \qquad i_j \times \frac{\log(c_j/(t^0\delta))}{\log(\xi)}, j = 1, 2 \tag{34}$$

From which we see that the Alg. 1.2 has higher complexity than Alg. 1.1 and is more sensitive to both the number of antennas and K and N. The complexity of these algorithms and the number of iterations required to achieve convergence is studied further in Section IV.

IV. Numerical Results

To evaluate the proposed algorithm for UL MIMO NOMA, we simulate a single cell VWN serving two slices, each with $K_s = 4$ users, except where otherwise noted. The BS is equipped with $M_r = 4$ receive antennas and the $N = 16$

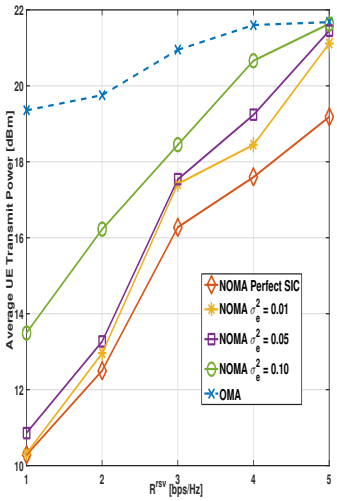


Fig. 2: Average UE transmit power versus $R^{\mathbf{rsv}}$

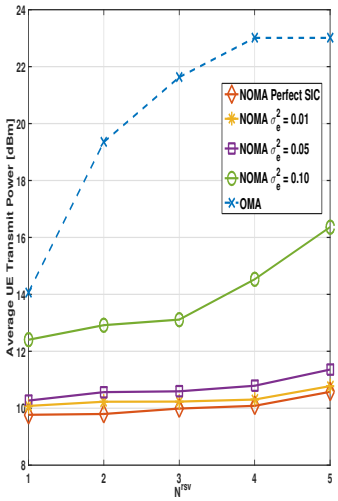


Fig. 3: Average UE transmit power versus $N^{\mathbf{rsv}}$

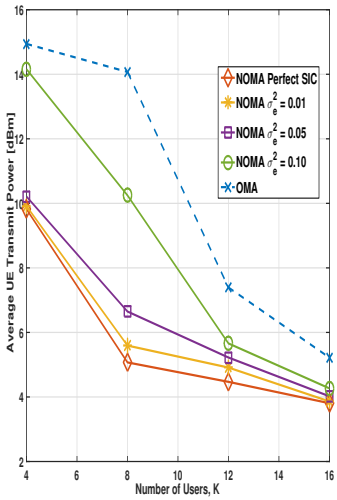


Fig. 4: Average UE transmit power versus K

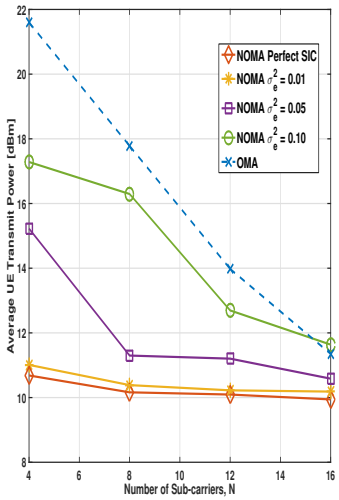


Fig. 5: Average UE transmit power versus N

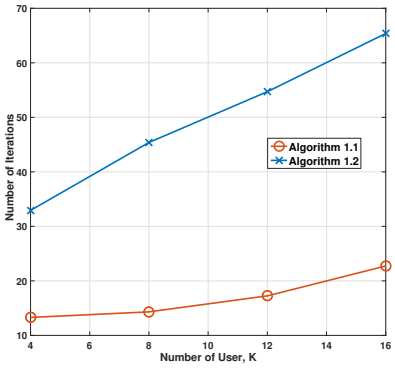
sub-carriers can each be shared by at most $N_{\mathbf{max}} = 4$ users each of which is equipped with $M_{t,k_s} = 2$, except for one random user in each slice for which $M_{t,k_s} = 1$. In all trials we have set $P_{\mathbf{max}} = 23$ dBm, and $R_s^{\mathbf{rsv}} = R^{\mathbf{rsv}}$ and $N_s^{\mathbf{rsv}} = N^{\mathbf{rsv}}$ for all slices. The users are placed randomly within the BS coverage area following a uniform distribution with distances d_{k_s} normalized to the radius of the coverage area, $\mathcal{R}$. The coverage area is divided at radius $r' = 0.5 \times \mathcal{R}$ with short-range users having $d_{k_s} <= r'$ and long-range users having $d_{k_s} > r'$, as depicted in Fig. 1. We assume a rich scattering environment and uncorrelated antennas and thus have $R_{G_{k_s,n}} = M_{t,k_s}$ independent channels from the SVD of the MIMO channel matrix. Short-range users transmit M_{t,k_s} data-streams and long-range users transmit a single data stream over all M_{t,k_s} antenna.

With precoding and detection as defined in Section II, the channel is modelled as a Rayleigh fading channel such that $||\mathbf{V}_{k_s,n}\mathbf{G}_{k_s,n}\mathbf{P}_{k_s}||^2 = \chi_{k_s,n} d_{k_s}^{-\lambda}$, where $\chi \sim \text{Exp}(1)$ and $\lambda = 3$ is the path loss exponent. Errors in performing SIC are introduced as residual interfering signal power with $\sigma_e^2 \in \{0.01, 0.05, 0.1\}$, i.e., up to 10% residual power remaining as inter-user interference after performing SIC at the BS. For comparison, we also present results for the perfect SIC case, where we set $\sigma_e^2 = 0$, as well as for MIMO OMA, where we set $N_{\mathbf{max}} = 1$ and only a single user can be assigned to each sub-carrier. The results shown are taken over the average of 100 channel realizations. Where no feasible solution exists, we set user transmit power to $P_{\mathbf{max}}$ and achieved rate to zero.

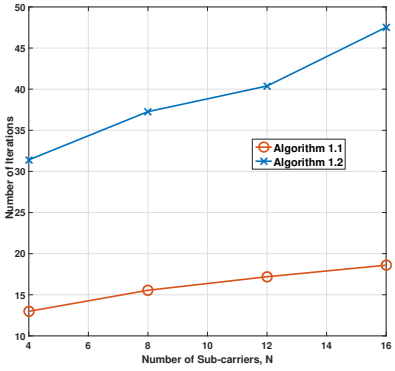
In Fig. 2 the average transmit power per UE is plotted versus $R^{\mathbf{rsv}}$ with $N^{\mathbf{rsv}} = 1$ for OMA and NOMA and varied levels of SIC error. Power increases with increasing $R^{\mathbf{rsv}}$ and SIC error levels since higher power is required to meet the slice rate reservation and overcome the increased inter-user interference from SIC imperfections. In all cases, NOMA is more power efficient than OMA, though for $R^{\mathbf{rsv}} = 3$ bps/Hz and $\sigma_e^2 > 0$ the power savings are significantly reduced. For example, for $R^{\mathbf{rsv}} = 1$ bps/Hz, NOMA requires an average UE transmit power of 6.32 dB, 6.28 dB, 5.74 dB, and 3.11 dB lower than OMA for $\sigma_e^2 = 0, 0.01, 0.05, 0.10$, respectively. For $R^{\mathbf{rsv}} = 5$ bps/Hz, power savings over OMA fall to 2.15 dB, 0.56 dB, 0.23 dB, and 0.05 dB for $\sigma_e^2 = 0, 0.01, 0.05, 0.10$, respectively.

The relationship between average transmit power and slice sub-carrier reservations, $N^{\mathbf{rsv}}$ is plotted in Fig. 3 for $R^{\mathbf{rsv}} = 1$ bps/Hz. As with increasing rate reservations, higher transmit power is required for higher sub-carrier reservations. The effect of varying $N^{\mathbf{rsv}}$ is particularly impactful for OMA which cannot avoid weak sub-carriers in favour of shared strong sub-carriers. For $\sigma_e^2 = 0.10$ we see a similarly sharp increase in transmit power for $N^{\mathbf{rsv}} > 3$ as either high levels of interference or the use of weaker sub-carriers require higher transmit power. However, even for high levels of SIC error, NOMA remains more power efficient than OMA. For example, with $N^{\mathbf{rsv}} = 1$, NOMA requires an average UE transmit power of 6.83 dB, 6.52 dB, 6.33 dB, and 4.20 dB lower than OMA for $\sigma_e^2 = 0, 0.01, 0.05, 0.10$, respectively. At $N^{\mathbf{rsv}} = 4$, NOMA is able to meet rate reservations with 11.93 dB, 11.71 dB, 11.22 dB, and 7.48 dB lower power than OMA for $\sigma_e^2 = 0, 0.01, 0.05, 0.10$, respectively. In comparison to increased rate reservations, where the optimal solution can allocate sufficient transmit power to a few users on strong sub-carriers, for OMA the results for $N^{\mathbf{rsv}} \geq 3$ are significantly impacted by reduced feasibility of finding a suitable solution as the sub-carrier reservation forces the use of weaker sub-carriers.

The performance impact of system density, i.e., the ratio of users to sub-carriers in the system, is depicted in Fig. 4 and 5. In Fig. 4, the average transmit power versus the number of users K is plotted for $N = 16$ with $N^{\mathbf{rsv}} = 1$ and $R^{\mathbf{rsv}} = 1$ bps/Hz. As expected, with increasing SIC error levels required power increases as there will be cases where shared sub-carriers are needed to meet the slice rate reservation. Since the system will allow the users to be pushed into outage as long as the reserved slice rate is met, as the number of users increases each slice can leverage users with stronger sub-carriers or spread users over more sub-carriers at lower power. In general, this causes the required power to



(a)



(b)

Fig. 6: Number of required iterations for each Algorithm versus (a) users, K, and (b) sub-carriers, N

meet slice reservation decreases with more users per slice and the difference in power between the different levels of SIC error is reduced as sub-carrier sharing becomes less beneficial. The effect of varying the total number of sub-carriers, N, in the system for fixed number of users is depicted in Fig. 5, for $K = 4$ with $N^{\mathbf{rsv}} = 1$ and $R^{\mathbf{rsv}} = 1$ bps/Hz. As before, we see that with decreasing system density, in this case for increasing number of sub-carriers, required average required power for both OMA and NOMA goes down due to the increased flexibility in ignoring weaker sub-carriers and decreased reliance on power-domain multiplexing. Required power under NOMA is lower than for OMA in general, but we note that, as the system density increases, for high levels of SIC error the performance gains over OMA may disappear as the optimal solution becomes one single user per sub-carrier.

To analyse the convergence and complexity of the proposed algorithms, Fig. 6 shows the number of iterations for convergence of Algorithms 1.1 and 1.2 with $N^{\mathbf{rsv}} = 1$ and $R^{\mathbf{rsv}} = 1$ bps/Hz. In Fig. 6 (a) we see the required iterations as a function of the number of users, K, and in Fig. 6 (b) as a function of the number of sub-carriers, N. In the analysis in Section III-C, it was shown that the complexity of Algorithm 1.2 was higher than that of Algorithm 1.1, with Algorithm 1.2 more sensitive to system parameters K and N, and particularly to increasing K. These relations are confirmed in the required number of iterations for each algorithm.

V. Conclusions

In this work, we have investigated the performance of UL MIMO NOMA in supporting VWNs in the presence of increased inter-user interference from errors in performing SIC. With the goal of minimizing required transmit power for battery dependent devices, while ensuring slice isolation and minimum system performance, we proposed an iterative algorithm based on SCA and CGP to solve the resulting non-convex and computationally intractable resource allocation problem. Simulation results were presented validating the proposed algorithm and demonstrating that, even for high level of SIC error, MIMO NOMA outperforms MIMO OMA in terms of required user transmit power and overall system density.

References

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

[2] H. Tabassum, et al., "Non-orthogonal multiple access (NOMA) in cellular uplink and downlink: Challenges and enabling techniques," arXiv preprint arXiv:1608.05783, 2016.

[3] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 358–380, First quarter 2015.

[4] Y. Saito, et al., "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE Veh. Tech. Conf. (VTC)*, June 2013, pp. 1–5.

[5] Z. Yang, et al., "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.

[6] N. Zhang, et al., "Uplink nonorthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 458–461, Mar 2016.

[7] S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565–577, 2017.

[8] Y. Liu, et al., "On the capacity comparison between MIMO-NOMA and MIMO-OMA," *IEEE Access*, vol. 4, pp. 2123–2129, 2016.

[9] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan 2016.

[10] Z. Feng, et al., "An effective approach to 5G: Wireless network virtualization," *IEEE Commun. Mag.*, vol. 53, no. 12, pp. 53–59, Dec 2015.

[11] L. Gao, et al., "Virtualization framework and VCG based resource block allocation scheme for LTE virtualization," in *Proc. IEEE Veh. Tech. Conf. (VTC)*, May 2016, pp. 1–6.

[12] C. Bao, et al., "Energy-efficient user association and downlink power allocation in software defined hetnet," in *Proc. IEEE Veh. Tech. Conf. (VTC)*, May 2016, pp. 1–5.

[13] T. D. Tran and L. B. Le, "Resource allocation for efficient bandwidth provisioning in virtualized wireless networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar 2017, pp. 1–6.

[14] S. Parsaeefard, et al., "Joint user-association and resource-allocation in virtualized wireless networks," *IEEE Access*, vol. 4, pp. 2738–2750, 2016.

[15] R. Dawadi, et al., "Power-efficient resource allocation in NOMA virtualized wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec 2016.

[16] D. Tweed, et al., "Dynamic resource allocation for MC-NOMA virtualized wireless networks with imperfect SIC," in *Proc. IEEE Intl. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Oct. 2017.

[17] S. Lim and K. Ko, "Non-orthogonal multiple access (NOMA) to enhance capacity in 5G," *Intl. J. of Contents*, vol. 11, no. 4, pp. 38–43, Dec 2015.

[18] 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, June 2016.

[19] A. Goldsmith, *Wireless Comunications*. Cambridge University Press, 2004.

[20] M. Chiang, et al., "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, Jul 2007.

[21] G. Xu, "Global optimization of signomial geometric programming problems," *European Journal of Operational Research*, vol. 233, no. 3, pp. 500–510, 2014.

[22] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1, build 1116," http://cvxr.com/cvx, Mar 2017.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.