

Beamforming and Scheduling for mmWave Downlink Sparse Virtual Channels With Non-Orthogonal and Orthogonal Multiple Access

Alessandro Brighente and Stefano Tomasin

Department of Information Engineering, University of Padova
via Gradenigo 6/B, 35131 Padova, Italy.

Email: brighent@dei.unipd.it, tomasin@dei.unipd.it

Abstract—We consider the problem of scheduling and power allocation for the downlink of a 5G cellular system operating in the millimeter wave (mmWave) band and serving two sets of users: fix-rate (FR) users typically seen in device-to-device (D2D) communications, and variable-rate (VR) users, or high data rate services. The scheduling objective is the weighted sum-rate of both FR and VR users, and the constraints ensure that active FR users get the required rate. The weights of the objective function provide a trade-off between the number of served FR users and the resources allocated to VR users. For mmWave channels the virtual channel matrix obtained by applying fixed discrete-Fourier transform (DFT) beamformers at both the transmitter and the receiver is sparse. This results into a sparsity of the resulting multiple access channel, which is exploited to simplify scheduling, first establishing an interference graph among users and then grouping users according to their orthogonality. The original scheduling problem is solved using a graph-coloring algorithm on the interference graph in order to select sub-sets of orthogonal VR users. Two options are considered for FR users: either they are chosen orthogonal to VR users or non-orthogonal. A waterfilling algorithm is then used to allocate power to the FR users.

Index Terms—5G; Beamforming; mmWave; Non-orthogonal Multiple Access; Scheduling.

I. INTRODUCTION

MmWave transmission systems operating with multiple users will pose many design challenges, related to the need of precise beamforming to overcome the strong attenuation at GHz frequencies [6]. On the other hand, the use of mmWave has been advocated for fifth-generation (5G) systems in order to satisfy on ten-fold achievable data rate with respect to the previous mobile communication generation: in this scenario the use of extremely high-frequency bands becomes even more problematic due to the presence of multiple users and the need to design suitable multiple access schemes. From a paradigm based on orthogonality among users to maximize the sum rate – mainly pursued in 4G systems – the 5G trend is now shifting back to non-orthogonal multiple access (NOMA) schemes, better suited to heterogeneous devices envisioned in the 5G network, e.g., sensors and actuators of the Internet of Things (IoT) [7]. All these problems are further complicated by the use of a huge number of antennas, the massive multiple input multiple output (MIMO) regime, again proposed for 5G

systems in order to propel data rate and cell user density. In this context, a radical review of both beam-forming and scheduling is needed, taking into account also the need of keeping both the signaling exchange on air at the minimum (for a higher data-rate) and the signal processing complexity under control, especially on the terminal side.

Focusing on a downlink scenario where a base station (BS) equipped with many antennas aims at serving multiple mobile terminals (MTs) on the mmWave band, many solutions are available in the literature for beam-forming and scheduling. For example, orthogonal user transmission can be considered in order to maximize the rate, thus (regularized) zero-forcing beamforming can be used at the BS, together with water-filling power allocation and scheduling [9], [10]. Similar approaches include diagonalization and block-diagonalization of the channel [8]. Indeed, it has been shown that in a massive MIMO regime, eigenbeamforming and matched filter precoders are optimal, and the transmit power can be made arbitrarily small [11]. However, these results leverage orthogonality among users and have as target the maximization of the sum rate. These assumptions must be revised in 5G systems used also for device-to-device (D2D) communication in the IoT context, typically having fix-rate (FR) requirements: for these services the target becomes serving the largest number of users, providing them the required fixed data rate. In the NOMA context, two approaches have been proposed in the literature: superposition coding and successive interference cancellation. With the first approach users suffer from the interference but can still decode the intended signal as long as the signal to noise plus interference ratio (SNIR) is above a given threshold; with the latter approach the interfering signal is decoded and then canceled before the useful signal is decoded. For superposition coding a large literature is available, for the maximization of the weighted sum-rate and energy-efficient solutions (see [12], [15] for surveys). In [14] beamforming and power allocation have been designed using as metrics the sum rate and outage probability of MIMO-NOMA systems. In [13] optimal beamforming and power allocation solutions have been derived for the weighted sum-rate maximization under a total power constraint and a max-min balancing of user rate under a total power constraint: in this work however no FR user is present. A simple approach has been proposed in

This work has been supported in part by Huawei Technologies, Milan, Italy.

[16], [17], where FR and variable-rate (VR) users are allocated orthogonal resources, then using further orthogonal resources for the VR users and non-orthogonal resources for the FR users. In [18] an scheduling considering also the pilot overhead needed for channel estimation, adapted to the user channels characteristics has been considered.

In this report we focus on the problem of scheduling and power allocation for the mixed user characteristics of [16]: first, we formulate the problem of maximizing the weighted sum-rate of both FR and VR users, with constraints ensuring that active FR users get the required rate. This turns out to be a mixed integer programming problem, where the integer variables dictate which FR users are active, and continuous variables provide the power allocation for all users. The weights of the objective function provide a trade-off between the number of served FR users and the resources allocated to variable users. We then exploit the peculiarity of mmWave channels which can be suitably represented as a sparse matrix in a dual domain [4]. The sparse channel is obtained by applying fixed discrete Fourier transform (DFT) beamformers at both the transmitter and the receiver, thus providing a low-complexity implementation. On the other hand, the sparsity of the resulting multiple access channel is exploited to simplify scheduling, first establishing an interference graph among users and then grouping users according to their orthogonality. In particular, the original scheduling problem is solved using a graph-coloring algorithm on the interference graph in order to select sub-sets of orthogonal VR users. Two options are considered for FR users: either they are chosen orthogonal to VR users or they are allowed to interfere with the other set. A waterfilling algorithm is then used to allocate power to the FR users, while the power needed to achieve the required rate is allocated to FR users. The rest of the report is organized as follows. Firstly we derive the system model for the described scenario and give a mathematical description of the power allocation and scheduling problem. Secondly we propose three solutions: the direct power allocation and scheduling, a solution based on graph coloring (GC), where both VR and FR users are clustered in different sets in order to eliminate interference, and a partial GC (PGC) solution where only VR users are clustered to eliminate interference. Lastly we analyse the performance of the proposed solutions in terms of achievable sum-rate when the optimization target moves from VR users to FR users and show how the number of antennas at the BS impacts on the proposed solutions.

II. SYSTEM MODEL

We consider a massive MIMO downlink transmission from a BS with a large number N_a of antennas organized in a uniform planar array (UPA) to a set of N_u single-antenna users, with $N_u \ll N_a$. Assuming an additive white Gaussian noise (AWGN) flat-fading channel model among each antenna couple, the resulting $N_u \times N_a$ channel matrix is denoted as \mathbf{H} . For a transmission in the mmWave band, each BS-user link is described by L paths, with L typically being small. Moreover, considering the UPA configuration at the BS and

the single antenna MT, as derived in [1] the row channel vector \mathbf{h}_k of user k can be transformed into a virtual channel vector $\mathbf{h}_{v,k}$ by the transformation

$$\mathbf{h}_k = \mathbf{h}_{v,k} \mathbf{U}_t^*, \quad (1)$$

where \mathbf{U}_t is the $N_a \times N_a$ unitary DFT matrix. By stacking the rows $\mathbf{h}_{v,k}$ into the matrix \mathbf{H}_v we can write the channel matrix as

$$\mathbf{H} = \mathbf{H}_v \mathbf{U}_t^*. \quad (2)$$

When the number of antennas at the BS tends to infinity, the virtual channel matrix \mathbf{H}_v becomes sparse, asymptotically having $L \cdot N_u$ non-zero entries associated to the L paths of each one of the N_u users. Also, when angles of departures are aligned with DFT direction [4] again \mathbf{H}_v becomes sparse with $L \cdot N_u$ non-zero complex entries, with probability density function

$$p_{[\mathbf{H}_v]_{j,k}}(a) = (1 - \eta) \delta(a) + \frac{\eta}{\pi \sigma_L^2} \exp\left(-\frac{|a|^2}{\sigma_L^2}\right), \quad (3)$$

where $\eta = L/N_a$ is the probability that a virtual channel matrix entry is different from zero, $\delta(\cdot)$ is the delta function and σ_L^2 is the average channel gain.

Received signal is affected by complex AWGN, independent at each user, with zero mean and variance σ^2 .

We consider two different types of users:

- K VR users having as target the maximization of their achievable sum-rate (ASR);
- N FR users aiming at being active, i.e., transmitting with a fixed data rate R_{fix} , corresponding to fixed signal to interference plus noise ratio (SINR) γ_{fix} with

$$R_{\text{fix}} = \log_2(1 + \gamma_{\text{fix}}). \quad (4)$$

The total number of users is $K + N = N_u$. A FR user assumes only two possible states: transmission with a FR imposed by the SINR requirement or turn off. In particular, the first K users are the VR users, while next N users are the FR users.

We denote by p_j , $j = 1, \dots, N_u$ the power assigned to user j . We also collect into the N_u column vector \mathbf{p} all user powers. We consider a total power constraint, i.e.,

$$p_k \geq 0, \quad \sum_{k=1}^{N_u} p_k \leq 1. \quad (5)$$

According to the beamformer adopted by the BS, as described in the following section, the resulting SINR for user $k = 1, \dots, N_u$ is denoted as $\gamma_k(\mathbf{p})$, depending on the power allocation. For the beamforming and scheduling problem we aim at maximizing the weighted ASR

$$\mathcal{R}(\mathbf{p}) = \sum_{k=1}^K w_k \log_2(1 + \gamma_k(\mathbf{p})) + \sum_{k=K+1}^{K+N} \rho_k x_k R_{\text{fix}}, \quad (6)$$

where ρ is the weight of FR users ASR, w_k is the weight of the ASR of VR user k and x_k is a binary variable which assumes value 1 if the considered FR user is active and 0 otherwise.

III. BEAMFORMING AND SCHEDULING

Considering the properties of the virtual channel, and that it can be obtained by the inverse discrete Fourier transform (DFT) matrix applied at the BS, we consider as BS transmit beamformer matrix \mathbf{U}_t , therefore the equivalent channel seen by the users is the sparse virtual channel \mathbf{H}_v having N_a virtual transmit antennas and N_u virtual receive antennas. We will use one transmit virtual antenna per user, and in particular for user k we select in $\mathbf{h}_{v,k}$ the entry with largest gain, i.e.,

$$\ell_k = \operatorname{argmax}_i |[\mathbf{h}_{v,k}]_i|^2, \quad (7)$$

and we indicate with

$$G_{k,k} = |[\mathbf{h}_{v,k}]_{\ell_k}|^2 \quad (8)$$

the corresponding gain. With this choice, user k will suffer from the the interference of all other user signals, based on his virtual channel $\mathbf{h}_{v,k}$, i.e. the gain of the interference channel from user j to user k is

$$G_{j,k} = |[\mathbf{h}_{v,k}]_{\ell_j}|^2, \quad (9)$$

corresponding to the use of virtual antenna ℓ_k by user j through the channel of user k . Therefore $N_u \times N_u$ matrix \mathbf{G} having entries $G_{j,k}$ defines the useful gain and interference gain for all users. With this choice, user k experiences a SINR

$$\gamma_k(\mathbf{p}) = \frac{G_{k,k}p_k}{\sum_{j \neq k} G_{j,k}p_j + \sigma^2}. \quad (10)$$

Let us consider the weighted ASR maximization problem with variables p_k and x_k . Our aim is to select a subset of FR users and to allocate power to both VR users and active FR users in order to maximize the weighted ASR. Let $\mathbf{x} = [x_1, \dots, x_{K+N}]$. The problem can be modeled as

$$\max_{\mathbf{p}, \mathbf{x}} \sum_{k=1}^K w_k \log_2 \left(1 + \frac{G_{k,k}p_k}{\sum_{j=1, j \neq k}^{K+N} G_{j,k}p_j x_j + \sigma^2} \right) + \sum_{k=K+1}^{K+N} \rho_k R_{\text{fix}} x_k \quad (11a)$$

s.t. (5) and

$$\frac{G_{k,k}p_k}{\sum_{j=1, j \neq k}^{K+N} G_{j,k}p_j x_j + \sigma^2} \geq \gamma_{\text{fix}} \quad k = K+1, \dots, K+N \quad (11b)$$

$$x_k = 1 \quad k = 1, \dots, K \quad (11c)$$

$$x_k \in \{0, 1\} \quad k = K+1, \dots, K+N \quad (11d)$$

This problem of joint power allocation and scheduling is modelled as a mixed-integer optimization problem, which belongs to the NP-hard class of problems. We will hence first analyse the solution and then discuss two different methods that can reduce the computational complexity of the solution while ensuring the same weighted ASR value.

A. Direct solution

In order to solve the mixed-integer programming problem we divide it into two problems: the problem of finding the set of active users (thus solving on \mathbf{x}) and the problem of allocating the power (thus solving on \mathbf{p}). In particular we exhaustively consider all possible sets of active FR users, and for each set we solve the power allocation problem (if feasible) maximizing the weighted ASR. For each set of active FR users \mathbf{x} the resulting power allocation problem is hence

$$\max_{\mathbf{p}} \sum_{k=1}^K w_k \log_2 \left(1 + \frac{G_{k,k}p_k}{\sum_{j=1, j \neq k}^{K+N} G_{j,k}p_j x_j + \sigma^2} \right) + \sum_{k=K+1}^{K+N} \rho_k R_{\text{fix}} x_k \quad (12)$$

s.t. (5), (13c) and (11d).

We notice that the resulting optimization problem is non-convex due to the argument of the the log function of (12) and we approximate the target function to make it convex. In particular, we consider a high SINR regime and perform the change of variables $q_k = \log_2 p_k$, obtaining the convex maximization problem

$$\max_{\mathbf{p}} \sum_{k=1}^K w_k \log_2 \left(\frac{G_{k,k}e^{q_k}}{\sum_{j=1, j \neq k}^{K+N} G_{j,k}e^{q_j} x_j + \sigma^2} \right) + \sum_{k=K+1}^{K+N} \rho_k R_{\text{fix}} x_k \quad (13a)$$

$$\text{s.t. } e^{q_k} \geq 0, \quad \sum_{k=1}^{K+N} e^{q_k} \leq 1 \quad (13b)$$

$$\frac{G_{k,k}e^{q_k}}{\sum_{j=1, j \neq k}^{K+N} G_{j,k}e^{q_j} x_j + \sigma^2} \geq \gamma_{\text{fix}} \quad k = K+1, \dots, K+N. \quad (13c)$$

The resulting maximization problem is hence convex and can be solved with standard optimization solutions. We then pick the set of active users for which the problem is feasible and provides that maximum weighted ASR.

B. GC solution

The second solution aims at reducing the computational complexity of the optimization problem by clustering the users in different groups, in which users transmit without interference. Then we transmit only to one set of non-interfering users, allocating powers in order to maximize the weighted ASR. The set of active users is the one that maximizes the weighted ASR among all sets.

In order to perform the clustering operation we use the interference matrix \mathbf{G} . We notice that, since the mmWave virtual channel results sparse, the resulting interference matrix computed over the matrix of the gains of the virtual channel \mathbf{G} will have mostly small or zero off-diagonal entries. In particular, as the ratio N_a/N_u of transmit and receive antennas tends to infinity, the interference matrix itself will become

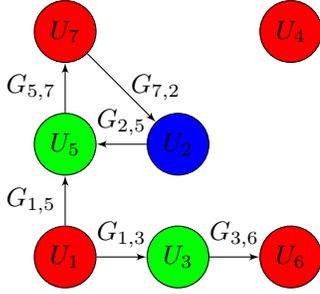


Fig. 1. Interference graph obtained from a particular realization of the interference matrix \mathbf{G} .

sparse. Indeed, in sparse massive-MIMO channels the channel vectors associated to different users become orthogonal as the number of antennas goes to infinity [5]. Our system behaves in a similar manner. Considering the virtual channel model with L non-zero paths and a growing number of antennas at the BS, we have that, asymptotically, users do not interfere. Therefore the interference matrix becomes a diagonal matrix and the graph colouring leads to all users labelled with the same colour.

From the interference matrix we build the interference graph, i.e. a directed graph composed by a node for each user and with edges connecting interfering users: edges are directed from the interfering user toward the user suffering the interference. No edge between nodes k and n is present if and only if $G_{k,n} = 0$. Nodes are then clustered into groups of non-interfering users: this corresponds to color the graph, where each color represents a group and nodes with the same colors must not be connected by edges. We aim at finding the minimum number of colors needed for the graph, and we resort to the solution proposed in [3]. Fig. 1 shows an example of interference graph obtained for a particular realization of the interference matrix, where, after graph colouring, each node (corresponding to user U_k) has been assigned a colour red or green. Since users having the same color do not interfere, assuming to transmit only to users of the same color the SINR of user k can be written as

$$\gamma_k = \frac{G_{k,k}p_k}{\sigma^2}, \quad (14)$$

which coincides with the signal to noise ratio (SNR).

Let \mathcal{C} the set of colors, \mathcal{F}_c be the set of indices of FR users with color $c \in \mathcal{C}$, and let \mathcal{V}_c the set of VR users with color c . To select the subset of active FR users (with the same color) we resort to the exhaustive search approach, thus computing the weighted ASR corresponding to each active set. For the FR users the allocated power is

$$p_k = \gamma_{\text{fix}} \frac{\sigma^2}{G_{k,k}} x_k, \quad k \in \mathcal{F}_c \quad (15)$$

For VR users since no interference is present, the optimal power allocation strategy is provided by the waterfilling algo-

rithm

$$\max_{\{p_k, k \in \mathcal{V}_c\}} \sum_{k \in \mathcal{V}_c} w_k \log_2 \left(1 + \frac{G_{k,k}p_k}{\sigma^2} \right) \quad (16a)$$

$$\text{s.t. } p_k \geq 0, \quad \sum_{k \in \mathcal{V}_c} p_k \leq 1 - \sum_{k \in \mathcal{F}_c} x_k p_k. \quad (16b)$$

The graph-coloring solution is summarized in Algorithm 1.

Algorithm 1: GC Power allocation and Scheduling Algorithm

Data: \mathbf{G}, σ^2 , color set \mathcal{C} , indices sets \mathcal{V}_c and \mathcal{F}_c , $c \in \mathcal{C}$.
Result: \mathbf{p}, \mathbf{x}

```

1 for  $c \in \mathcal{C}$  do
2   for all possible sets of FR users  $\mathbf{x}$  do
3      $\forall k \in \mathcal{F}_k$  compute  $p_k$  | (11b) is satisfied
4     for  $k \in \mathcal{F}_c$  do
5       compute water-filling level  $\lambda$  such that
          
$$\sum_{k \in \mathcal{V}_c} \left( \lambda - \frac{\sigma^2}{G_{k,k}} \right)^+ = 1 - \sum_{k \in \mathcal{F}_c} x_k p_k \quad (17)$$

           $\mathcal{R}_c(\mathbf{x}) = \text{weighted ASR for } \mathbf{x}$ 
6     end
7   end
8    $\mathbf{x}_c = \text{argmax}_{\mathbf{x}} \mathcal{R}_c(\mathbf{x})$ 
9 end
10  $c^* = \text{argmax}_c \mathcal{R}_c$ 
11  $\mathbf{p} = \mathbf{p}_{c^*}$ 
12  $\mathbf{x} = \mathbf{x}_{c^*}$ 

```

C. PGC solution

The third solution applies grouping (i.e., GC) only on VR users, while the FR users are allowed to interfere with the VR users and among themselves. This allows a more flexible allocation of FR users, while at the same time due to their low-rate request we expect that the resulting interference on the VR users is reduced.

Therefore, in this case we apply the GC only on the interference graph of the VR users, while we must perform the exhaustive search of the the active FR users among all FR users. Therefore, for each color $c \in \mathcal{C}$ we must solve the optimal scheduling and power allocation problem for the set of VR users \mathcal{V}_c and for all FR users. With respect to the global direct approach we have reduced the set of VR users, but on the other hand we must run the optimization algorithm for each color.

IV. NUMERICAL RESULTS

We present now numerical results obtained with the different solutions described in previous sections. We consider $N_u = 7$ users where $K = 4$ are VR users and $N = 3$ are FR users. We assume that the virtual channel matrix \mathbf{H}_v is composed by $L = 100$ non-zero entries and that they are

distributed as in (3) and the SINR requirement for FR users is 5 dB while AWGN leads to an average SNR of 17 dB.

We start verifying the assumption that the interference matrix is sparse, due to the large number of BS antennas. Let us consider matrix \mathbf{D} , defined as the matrix containing off-diagonal elements of \mathbf{G} , i.e., $\mathbf{D} = \mathbf{G} - \text{diag}(\mathbf{G})$. Fig. 2 shows the average number of non-zero entries of \mathbf{D} vs the number of antennas at the BS. We observe that the average number of non-zero entries sharply decreases as the number of antennas increases, and at $N_a = 200$ we have on average about 3 non-zero off-diagonal entries. In the following we consider $N_a = 200$.

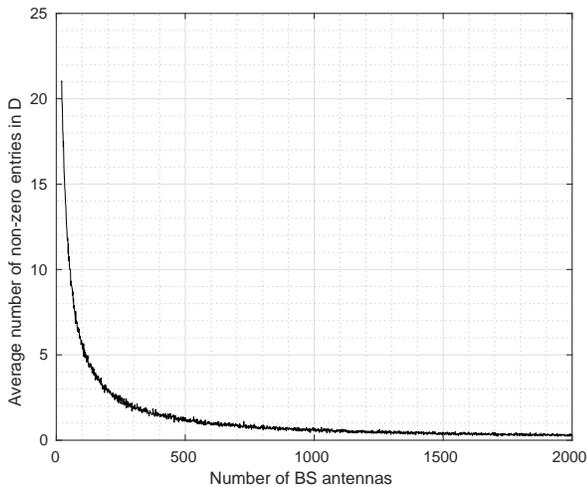


Fig. 2. Average number of non-zero entries in \mathbf{D} vs. number of antennas at the BS.

We now consider the scheduling and power allocation algorithms described in the paper. The weight w_k has been chosen to be equal to $1 - \rho$ for each VR user, i.e. $w_k = 1 - \rho \quad \forall k = \{1, \dots, K\}$ and $\rho \in \{0, 1\}$. With this choice we do not distinguish among the users in the two sets, but only use ρ to balance the resources uses for VR and FR users. In particular, when $\rho \approx 0$, we maximize the ASR of VR users, whereas for $\rho \approx 1$ we maximize the ASR of FR users. Therefore we consider as performance metric the average ASR for both VR and FR users. Fig. 3 shows the average (over various channel realizations) ASR vs. ρ , for the different scheduling and power allocation solutions. The direct algorithm is denoted as `direct`, and then we have the GC and PGC algorithms. As we expect, with increasing values of ρ , VR users ASR decreases while FR users ASR increases. Note also that even if $\rho = 1$ we still have a non-zero ASR for VR users, since all FR are served and the remaining power is allocated to VR users.

We notice that the direct solution to power allocation and scheduling is well approximated by both GC and PGC solutions for VR users, whereas, for FR users, the PGC solution is the best performing. Worse ASR performance for FR users in GC solution are due to the fact that VR users

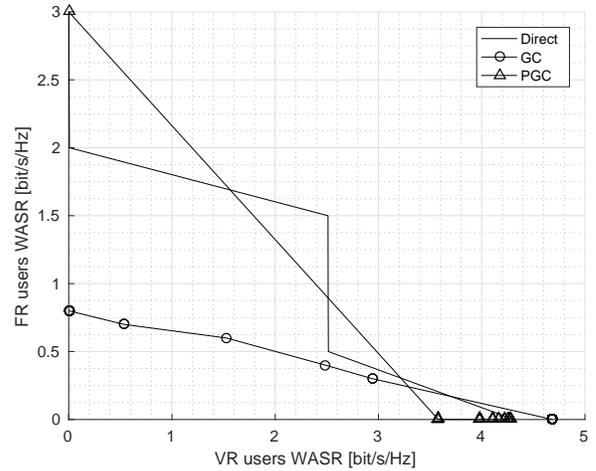


Fig. 3. FR vs. VR users ASR for different ρ values.

could be labelled with different colours and hence can not transmit together. Notice that the GC solution for VR users outperforms the direct solution. This is due to the fact that even if the maximization problem is focused on FR users ASR, they could be labelled with different colours and hence in GC solution they can not transmit all together, resulting in a power allocation for VR users with higher values than direct solution. This results in higher ASR values for VR users when ρ tends to 1.

V. CONCLUSIONS

In this work we considered a mmWave massive MIMO system and we solved the problem of power allocation and users scheduling when two different sets of users transmit: VR users and FR users with a required SINR level for transmission. We analysed the direct solution and proposed two algorithms, GC and PGC, that present similar ASR performance for both VR users and FR users while reducing the computational complexity. Then we analysed the performance of all the proposed solutions in terms of ASR when the weight ρ (and hence w_k) assigned to VR users changes and showed that the proposed algorithms present the same performance obtained with the direct solution. We also showed that with an increasing number of antennas at the BS the interference matrix becomes diagonal and hence that the low-complexity GC solution achieves optimal results.

REFERENCES

- [1] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563-2579, Oct 2002.
- [2] J. Papandriopoulos and J. S. Evans, "Low-Complexity Distributed Algorithms for Spectrum Balancing in Multi-User DSL Networks", in Proc. *IEEE International Conference on Communications*, vol.7, pp. 3270-3275, June 2006.
- [3] P. S. Segundo "A new DSATUR-based algorithm for exact vertex coloring", *Computers and Operations Research*, no.39, pp. 1724-1733, 2012.

- [4] J. Mo, P. Schniter, N. G. Prelcic and R. W. Heath Jr. "Channel Estimation in Millimeter Wave MIMO Systems with One-Bit Quantization", in Proc. *48th Asilomar Conference on Signals, Systems and Computers*, pp. 957-961, Nov. 2014.
- [5] Z. Gao, L. Dai, C. Yuen, Z. Wang, "Asymptotic Orthogonality Analysis of Time-Domain Sparse Massive MIMO Channels", *IEEE Communications Letters*, vol. 19, no. 10, Oct. 2015
- [6] S. Rangan, T. S. Rappaport and E. Erkip, "Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366-385, Mar. 2014.
- [7] A. Gupta and R. K. Jha, "A Survey of 5G Network: Architecture and Emerging Technologies," *IEEE Access*, vol. 3, pp. 1206-1232, 2015.
- [8] K. K. Wong, R. D. Murch, and K. B. Letaief, "A joint-channel diagonalization for multiuser MIMO antenna system," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 773-786, Jul. 2003.
- [9] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461-471, Feb. 2004.
- [10] H. Yang and T. L. Marzetta, "Performance of Conjugate and Zero-Forcing Beamforming in Large-Scale Antenna Systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 172-179, Feb. 2013.
- [11] J. Hoydis, S. ten Brink and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160-171, Feb. 2013.
- [12] F. Fang, H. Zhang, J. Cheng and V. C. M. Leung, "Energy-Efficient Resource Allocation for Downlink Non-Orthogonal Multiple Access Network," in *IEEE Transactions on Communications*, vol. 64, no. 9, pp. 3722-3732, Sept. 2016.
- [13] S. Shi, M. Schubert and H. Boche, "Rate Optimization for Multiuser MIMO Systems With Linear Processing," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 4020-4030, Aug. 2008.
- [14] 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.
- [15] Wei Zhiqiang, Yuan Jinhong, Derrick Wing Kwan Ng, Maged ElKashlan, and Ding Zhiguo, "A Survey of Downlink Non-Orthogonal Multiple Access for 5G Wireless Communication Networks", *ZTE Communications*, Vol.14 No. 4, Oct. 2016.
- [16] N. Ksairi, S. Tomasin and M. Debbah, "A Multi-Service Oriented Multiple Access Scheme for M2M Support in Future LTE," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 218-224, Jan. 2017.
- [17] N. Ksairi, S. Tomasin and M. Debbah, "A multi-service oriented multiple-access scheme for next-generation mobile networks," in Proc. *2016 European Conference on Networks and Communications (EuCNC)*, Athens, 2016, pp. 355-359.
- [18] N. Ksairi, S. Tomasin, and B. Tomasi, "Systems and Methods for Scheduling of Resources and Pilot Patterns to User Terminals in a Multi-User Wireless Network," Patent PCT/EP2016/052202, Filed on February 2, 2016.