Advanced Resource Allocation Technique for a Fair Downlink Non-Orthogonal Multiple Access System

Marie-Rita Hojeij, Charbel Abdel Nour, Joumana Farah, Catherine Douillard

Advanced Resource Allocation Technique for a Fair Downlink Non-Orthogonal Multiple Access System

Marie-Rita Hojeij, *Student Member, IEEE,* Charbel Abdel Nour, *Member, IEEE,* Joumana Farah, *Member, IEEE,*
and Catherine Douillard, *Senior Member, IEEE*

Abstract—In this paper, a novel resource allocation technique is proposed for a power-domain non-orthogonal multiple access scheme using successive interference cancellation in a downlink cellular system. It aims at providing a flexible balancing between throughput and fairness maximization and incorporates unequal power repartition among allocated subbands using waterfilling. Its main strength is the ability to achieve a high level of fairness in every scheduling slot, therefore improving the quality of experience for all users. Fairness is achieved by the introduction of a new efficient scheduling metric within the allocation process. Extensive simulation results show that the proposed metric enhances both long-term and short-term fairness, and improves system capacity, when compared to the conventional proportional fairness scheduler.

Index Terms—Non-orthogonal multiple access, power allocation, scheduling, proportional fairness, long-term and short-term fairness.

I. Introduction

Recent radio access technologies are characterized by multiple access schemes which provide the mean for multiple users to access and share resources at the same time. In the 3.9 and 4th generations of mobile communication systems, such as Long-Term Evolution (LTE) and LTE-Advanced, Orthogonal Multiple Access (OMA) based on OFDM (Orthogonal Frequency Division Multiplexing) and Single Carrier Frequency Division Multiple Access (SC-FDMA) were adopted, respectively for the downlink and uplink transmissions. OMA techniques have gained their success from their ability to achieve good system-level throughput performance in packet-domain services and their viability in practical implementations with single-user detection.

However, with the proliferation of smart and machine-to-machine devices, it is expected that the mobile traffic volume supported by future communication networks will be, by 2020, almost 10 times larger than today's [1]. To respond favorably to such constraints while keeping a high level of user Quality of Experience (QoE), system capacity and user fairness should be largely improved for the future 5th generation (5G) of mobile communication systems. To this end, Non-Orthogonal Multiple Access (NOMA) has recently emerged as a promising candidate for future radio access. By exploiting an additional

M. Hojeij, C. Abdel Nour and C. Douillard are with IMT Atlantique, Lab-STICC, CNRS UMR 6285, CS 83 818-29238 Brest Cedex 3, France (marie.hojeij, charbel.abdelnour, catherine.douillard@imt-atlantique.fr) and M. Hojeij is also with the Department of Telecommunications, Faculty of Engineering, Holy-Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon. J. Farah is with the Department of Electricity and Electronics, Faculty of Engineering, Lebanese University, Roumieh, Lebanon (joumana.farah@ul.edu.lb).

multiplexing domain, Power Domain NOMA (PD-NOMA) allows the cohabitation of two or more users per subband at the transmitter side on top of the OFDM layer. Multi-user signal separation takes place at the receiver side, using Successive Interference Cancellation (SIC) [2], [3].

The majority of existing works dealing with PD-NOMA have investigated the system-level performance in terms of system capacity and cell-edge user throughput [3]–[6]. They show that, by solving the user pairing, subband assignment and power allocation problems appropriately, PD-NOMA can provide a significant improvement compared to OMA.

In [2], system efficiency and user fairness of PD-NOMA are investigated and compared to OMA. In this sense, universal frequency reuse and proportional fairness (PF) scheduling [7] are adopted. A large throughput gain is observed for users located near the base station, but the gain achieved by cell-edge user throughput is shown to be rather limited.

Aiming at further enhancing user fairness and system capacity, we propose in this paper a resource allocation technique where fairness is targeted at each channel realization. By doing so, not only short-term fairness is achieved among users, but also system capacity and long-term fairness are consequently improved.

This paper is organized as follows: In Section II, we introduce the general system model and give a brief description of the NOMA-based PF scheduler. Section III details the proposed resource allocation schemes. Simulation results are given in Section IV, while Section V concludes the paper.

II. System Description

A. Description of the basic PD-NOMA scheme with SIC

In this section, we describe the basic concept of PD-NOMA including user multiplexing at the transmitter of base station (BS) and signal separation at user terminal.

In this paper, a downlink air interface with a single transmitter and a single antenna receiver is considered. The system consists of K users per cell, and a total bandwidth B divided into S subbands. Among the K users, a set of users $U_s = \{k_1, k_2, ..., k_n, ..., k_{n(s)}\}$ is selected to be scheduled on each frequency subband s, $(1 \leq s \leq S)$. The n-th $(1 \leq n \leq n(s))$ user scheduled at subband s is denoted by k_n, and $n(s)$ indicates the number of users non-orthogonally scheduled at subband s. At the BS transmitter side, the information sequence of each scheduled user at subband s is independently coded and modulated, resulting into symbol x_{s,k_n} for the n-th scheduled user. Therefore, the signal x_s, transmitted by the

BS on a certain subband s, represents the sum of the coded and modulated symbols of the $n(s)$ scheduled users:

$$x_s = \sum_{n=1}^{n(s)} x_{s,k_n}, \; with \; \mathrm{E}\left[\left|x_{s,k_n}\right|^2\right] = P_{s,k_n} \tag{1}$$

where P_{s,k_n} is the power allocated to user k_n at subband s. The received signal of user k_n at subband s, y_{s,k_n}, is represented by:

$$y_{s,k_n} = h_{s,k_n} x_{s,k_n} + w_{s,k_n} \tag{2}$$

where h_{s,k_n} is the channel coefficient between the BS and user k_n, at subband s. w_{s,k_n} indicates the Gaussian noise plus inter-cell interference received by user k_n on subband s. Let P_{max} be the BS maximum transmit power. Hence, the sum power constraint is formulated as:

$$\sum_{s=1}^{S} \sum_{n=1}^{n(s)} P_{s,k_n} = P_{\max} \tag{3}$$

At the receiver side, we consider SIC processing [8]. The optimal order for user decoding is in increasing order of the channel gains observed by the users, normalized by the noise and inter-cell interference $h_{s,k_n}^2/n_{s,k_n}$, where h_{s,k_n}^2 is the equivalent channel gain and n_{s,k_n} the average power of w_{s,k_n}. Therefore, user k_n can remove the inter-user interference from user k_j at subband s, provided that $h_{s,k_j}^2/n_{s,k_j}$ is lower than $h_{s,k_n}^2/n_{s,k_n}$, and the received signals from other users with higher $h_{s,k_j}^2/n_{s,k_j}$ are treated as noise [9].

Assuming successful decoding and no error propagation, the throughput of user k_n, at subband s, R_{s,k_n}, is given by:

$$R_{s,k_n} = \frac{B}{S} \log_2 \left(1 + \frac{h_{s,k_n}^2 P_{s,k_n}}{\displaystyle\sum_{j \in \mathrm{U}_s,\, \frac{h_{s,k_n}^2}{n_{s,k_n}} < \frac{h_{s,k_j}^2}{n_{s,k_j}}} h_{s,k_n}^2 P_{s,k_j} + n_{s,k_n}} \right) \tag{4}$$

B. Proportional fairness scheduler for PD-NOMA

In [10], the operation of the PF scheduler is detailed: at the beginning of each scheduling slot, each user provides the base station with its channel state or its feasible rate. The PF algorithm keeps track of the historical rate $T_k(t)$, i.e., the average throughput of each user in a sliding window of length t_c. In scheduling slot t, a user is selected to be served based on:

$$\arg\max_k \frac{R_k(t)}{T_k(t)} \tag{5}$$

where $R_k(t)$ is the feasible rate of user k in slot t.

An approximated version of the PF scheduler for multiple users transmission has been adopted in the majority of the works dealing with NOMA [9], [11], [12] for selecting users to be non-orthogonally scheduled on available resources. For each subband s, all possible candidate user sets are searched, and the set of scheduled users U_s is chosen in such a way to maximize the PF scheduling metric :

$$U_s = \arg\max_U \sum_{k \in U} \frac{R_{s,k|U}(t)}{T_k(t)} \tag{6}$$

U is a possible candidate user set, $R_{s,k|U}(t)$ is the instantaneous achievable throughput of user k on subband s at time instance t, the time index of a subframe. $R_{s,k|U}(t)$ is calculated based on Eq. 4, whereas $T_k(t)$ is updated according to:

$$T_k(t+1) = \left(1 - \frac{1}{t_c}\right) T_k(t) + \frac{1}{t_c} \sum_{s=1}^{S} R_{s,k}(t) \tag{7}$$

The throughput averaging time window, t_c, is chosen such that a tradeoff between system performance in terms of fairness and system capacity is guaranteed.

III. PROPOSED RESOURCE ALLOCATION TECHNIQUE FOR A FAIR PD-NOMA SYSTEM

The objective of the PF is to maximize the long-term averaged user rates, thus targeting long-term fairness. However, short-term fairness and fast convergence towards required performance are an important issue to be addressed in upcoming mobile standards [1]. Since the PF scheduler gives each user a priority inversely proportional to its average throughput, and since users instantaneous rates are not considered in the scheduling metric, it happens that some users do not gain access to available resources whenever their historical rates are high. This can be problematic for applications requiring a quasi-constant QoE, such as multimedia transmissions. In addition, the PF scheduler does not take advantage of the fact that the sum rate of NOMA systems increases when increasing the difference in channel gains of paired users. We propose a resource allocation technique that aims at addressing these issues.
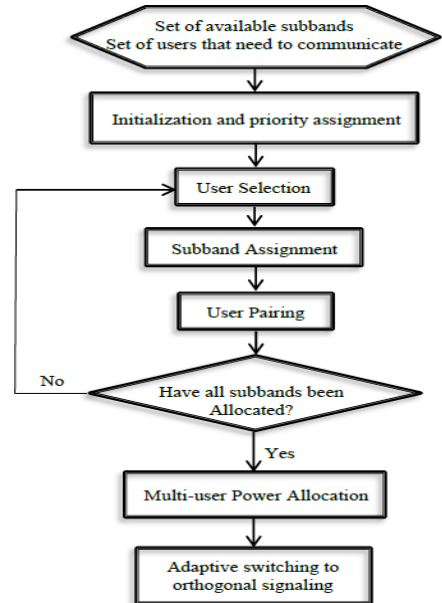

Fig. 1. Proposed resource allocation algorithm

The design steps of the proposed resource allocation technique are depicted in Fig. 1 and detailed in the following subsections.

A. Initialization and Priority Assignment

At the beginning of the process, allocated powers and rates are all set to zero. Therefore, users are first chosen based on the priority constraint proposed in a previous work [6]. Priorities are assigned in the BS based on the channel coefficients experienced by users on available subbands. For each user k, its highest channel coefficient, $h_{s_{max},k}$, is selected among the channel gains experienced over all subbands. Then, the users are processed in the increasing order of $h_{s_{max},k}$ values: user k_{min} with the lowest value of $h_{s_{max},k}$ is assigned first while user k_{max} with the highest value $h_{s_{max},k}$ is assigned last.

B. User Selection and Subband Assignment

After the initialization phase, unlike for PF scheduling, user rates already granted during the current allocation step are also taken into consideration. User k_1 is selected among the set of users that need to communicate as the user experiencing the lowest total achieved throughput. User selection can be written as follows:

$$k_1 = \min_k \left(\left(1 - \frac{1}{t_c}\right) T_k(t) + \frac{1}{t_c} \sum_{s=1}^{S} R_{s,k}(t) \right) \quad (8)$$

$R_{s,k}(t)$ is calculated based on Eq. 4, and can be equal to zero if user k is not scheduled on subband s.

User k_1 is then attributed the most favorable subband, i.e., the one with the highest channel coefficient, denoted by s_f.

C. User Pairing

We have limited our study to the case where the number of scheduled users per subband is equal to 2 ($n(s) = 2$), since it was shown in [13] that the improvement in cell throughput is 1% when the number of scheduled users per subband is equal to 3, compared to the case where it is equal to 2. However, the proposed pairing technique described in the following has the ability to consider more than 2 users per subband if needed.

Having determined the subband s_f attributed to selected user k_1, the following step consists in selecting the best user k_2 to be paired with k_1. The selection of this second user is based on metric calculations requiring rate estimates, which in turn include power estimates. In the majority of papers dealing with PD-NOMA, such as [2], [12], an equal repartition of power among subbands is considered to simplify the scheduling problem. Quite the contrary, in our work, power is assigned to the users on the current subband s_f based on the waterfilling principle, since it allows the achieved system throughput to be maximized [3]. The implementation of an optimal Power Allocation (PA) scheme based on a waterfilling process jointly performing intra and inter-subband PA is quite complex since it requires the resolution of a non-convex optimization problem. For this purpose, we have proposed a sub-optimal PA technique using an iterative waterfilling process [14], where only the user with the highest channel coefficient among

scheduled users is considered in the inter-subband PA step. This technique was applied in the current work, combined with the Fractional Transmit Power Allocation (FTPA) technique [15] for intra-subband PA between the two paired users.

We have also introduced a new metric called Flexible Throughput vs. Fairness Maximization Metric (FTFMM) in order to cope with applications where throughput versus fairness maximization is essential. FTFMM aims at striking a flexible balancing between throughput and fairness, since it is based on the instantaneous rate on the considered subband s_f, the current user rates on the so far attributed subbands within the current allocation slot and on the historical rates of each possible candidate. A user is selected to be paired with user k_1 on the considered subband s_f based on:

$$k_2 = \max_k \frac{R_{s_f,k} + R_{s_f,k_1}}{aR_{k,tot} + bR_{k_1,tot}} \quad (9)$$

where $R_{s_f,k}$ is the data rate achieved by user k over subband s_f, calculated from Eq. 4.

Parameters a and b can take values between 0 and 1 and are used to provide flexibility between throughput and fairness maximization. If a is set to 1 and b to 0 in Eq. 9, user k_2 is selected in a fair manner by taking the instantaneous rates as well as the total user rates of the possible candidates into consideration. On the contrary, if a is set to 0 and b to 1, user k_2 is selected to be paired with k_1 according to its instantaneous rate achieved over s_f, and therefore the choice is only based on the rate maximization on the current subband. Neither his historical rate nor his current rate have been taken into account. In such a situation, the notion of fairness is completely flawed. Between these two extreme cases, and depending on the requirement of the application, parameters a and b can be varied in such a way to favor either the fairness or the throughput or to ensure balance between them.

D. Adaptive switching to Orthogonal Signaling

The estimation of power conducted within each stage of the allocation process is considered as temporary until the final stage is reached. At the final stage, once all subbands have been allocated, power is eventually distributed on each subband and data rates achieved by users are computed accordingly. However, the increase in spectral efficiency incurred by the use of NOMA compared to OMA is not systematic. Actually, the loss in data rate experienced by user k_1 when sharing its subband with user k_2 is sometimes significantly greater than the data rate gain achieved by k_2 on this subband. In this case, NOMA is not the appropriate solution, and the allocation of this subband to user k_1 alone turns out to be more beneficial. In a previous work [6], we have proposed to adaptively switch to orthogonal signaling if:

$$\gamma \left(R_s - R_{s,k_1} \right) > R_{s,k_2} \left(\text{or } R_s > R_{s,k_1} + \frac{R_{s,k_2}}{\gamma} \right) \quad (10)$$

where R_s is the data rate achieved on subband s by solely attributing the subband to a unique user. γ is a control parameter to be determined *a priori* via simulations such as to maximize the system performance.

IV. NUMERICAL RESULTS

A. System Model Parameters

Extensive system level simulations were carried out in order to evaluate the proposed allocation technique. The parameters considered in this work are based on existing LTE/LTE-Advanced specifications. The maximum BS transmission power is 46 dBm. The system bandwidth is 10 MHz divided into 8, 16, 32, 64, or 128 subbands, with a carrier frequency of 2 GHz. The noise power spectral density is 4.10^{-18} mW/Hz. Users are randomly deployed in a cell of radius 500 meters, with a minimum distance of 35 meters between users and BS. Distance-dependent path loss is considered with a decay factor of 3.76. The Extended Typical Urban (ETU) channel model is used, with a mobile velocity of 50 km/h. We assume a t_c window of 100 ms (100 subframes of 1 ms). Perfect channel estimation is assumed. The maximum number of scheduled users per subband is 2 ($n(s) = 2$). The FTPA technique is applied for intra-subband PA with coefficient α equal to 0.7 [15].

B. Performance Evaluation

In this study, we mainly consider three system-level performance indicators: achieved system capacity, long-term fairness, and short-term fairness.

In order to evaluate the achieved fairness performance, we use the Gini fairness index [16], defined as follows:

$$G = \frac{1}{2K^2 \bar{r}} \sum_{k'=1}^{K} \sum_{k''=1}^{K} |r_{k'} - r_{k''}| \quad (11)$$

$$\text{with} \qquad \bar{r} = \frac{\sum_{k=1}^{K} r_k}{K} \quad (12)$$

Gini fairness index takes values between 0 and 1, where $G = 0$ corresponds to the maximum level of fairness among users, while a high value value of G indicates that the resource allocation scenario is highly unfair. When fairness among users is to be evaluated within each scheduling slot, short-term fairness is considered and r_k is equal to $R_k(t)$, the actual achieved throughput of user k during scheduling slot t.

Fig. 2 and Fig. 3 compare the capacity of a NOMA system using the proposed approach with conventional OMA and NOMA systems based on PF scheduling, as a function of the number of users per cell and subbands, respectively. When different couples of parameters a and b values are evaluated, one can observe that the capacity is maximized for $a = 0$ and $b = 1$ (upper throughput bound), decreases progressively when a increases and b decreases, and reaches its lowest value for $a = 1$ and $b = 0$. This is due to the fact that when parameter a is set to 0, Eq. 9 targets the maximization of the sum rates of paired users. As expected, a gain in performance is observed with the proposed FTFMM metric, when compared to a NOMA-based classical PF, since the proposed metric tries to provide rates to all users, even those experiencing bad channel conditions. For all the simulated systems, the throughput increases with the number of users per cell. This is
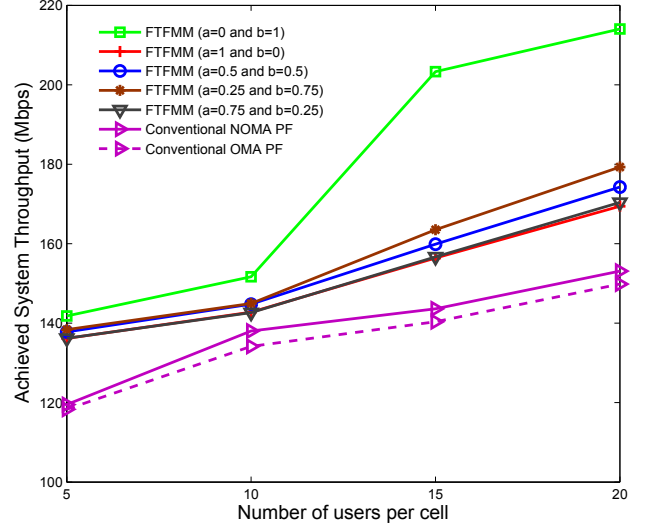

Fig. 2. Achieved system throughput of the proposed scheduling schemes, as a function of the number of users per cell
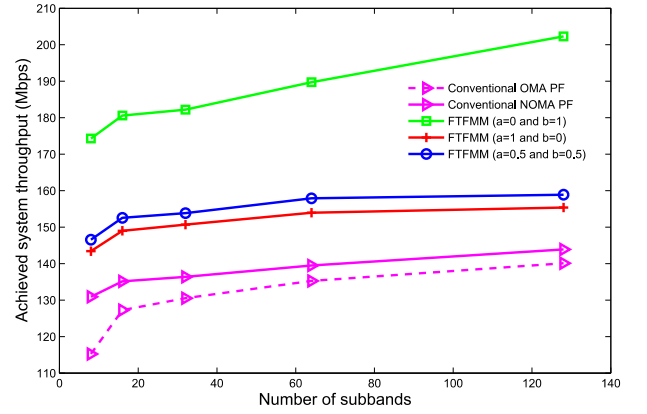

Fig. 3. Achieved system throughput of the proposed schemes, as a function of the number of subbands, for 15 users per cell

due to the fact that the higher the number of users per cell, the better the multi-user diversity is exploited by the scheduling scheme. For 15 users per cell, FTFMM with $a = 0$ and $b = 1$ allows the system throughput to be increased by 35% to 40% with respect to NOMA-based classical PF, depending on the number of subbands.

Fig. 4 and Fig. 5 compare the long-term fairness of a NOMA system using the proposed approach with conventional OMA and NOMA systems based on PF scheduling, as a function of the number of users per cell and subbands, respectively. One can observe that the highest level of fairness is achieved when $a = 1$ and $b = 0$ (upper fairness bound). FTFMM always shows better performance than the NOMA-based classical PF, whatever the number of users and subbands. Since it tends to achieve fairness in every scheduling slot by taking into account current rates, long-term fairness is enhanced consequently.
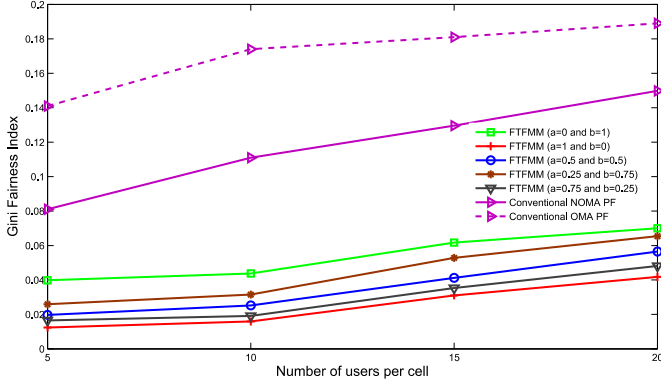
Fig. 4. Gini fairness index of the proposed scheduling schemes, as a function of the number of users per cell
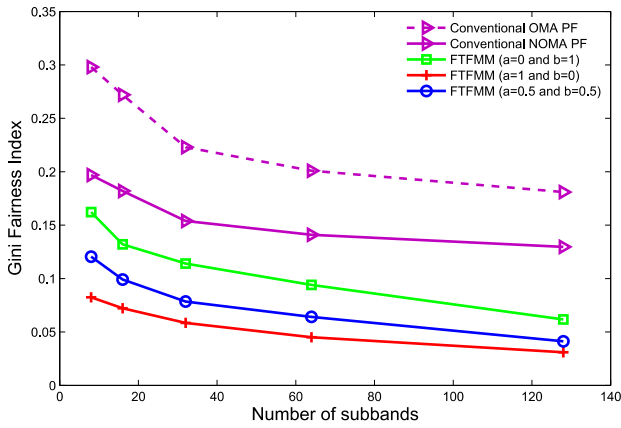

Fig. 6. Gini fairness index versus time for the proposed schemes

Fig. 5. Gini fairness index of the proposed schemes, as a function of the number of subbands, for 15 users per cell

Fig. 7. User throughput versus time

This is in compliance with Eq. 9, where total achieved rates and current rates are both considered for user pairing, thus providing a high level of fairness.

Fig. 6 evaluates the evolution over time of the short-term fairness achieved by the proposed allocation schemes. The studied techniques are compared based on the time required to converge to a Gini fairness index close to zero. FTFMM always gives better fairness performance compared to classical PF, for all simulated values of a and b. Especially when $a = 1$ and $b = 0$, a high level of fairness is achieved in a limited number of allocation steps (less than 10 subframes for a Fairness index less than 0.1).

In order to assess the QoE achieved by users with the proposed scheduling schemes, we have also measured the rate latency, i.e., the waiting time before being served for the first time, and observed the variation of the user achieved rate versus the scheduling index t. For this purpose, Fig. 7 shows the achieved data rate of the user experiencing the largest rate latency versus time, for the different scheduling schemes under study. We can observe that no rate is provided to such users for the first five scheduling slots when classical PF is used. In addition, severe rate fluctuations are observed through
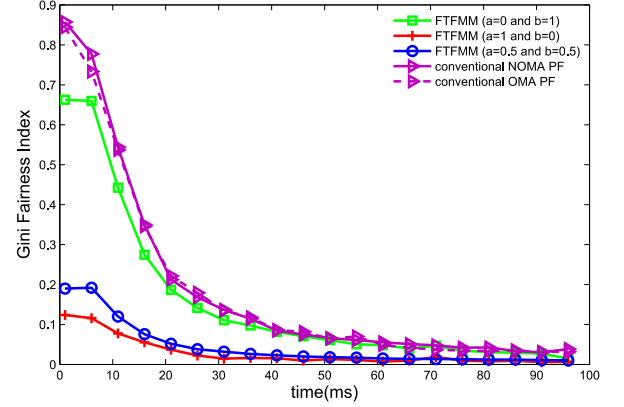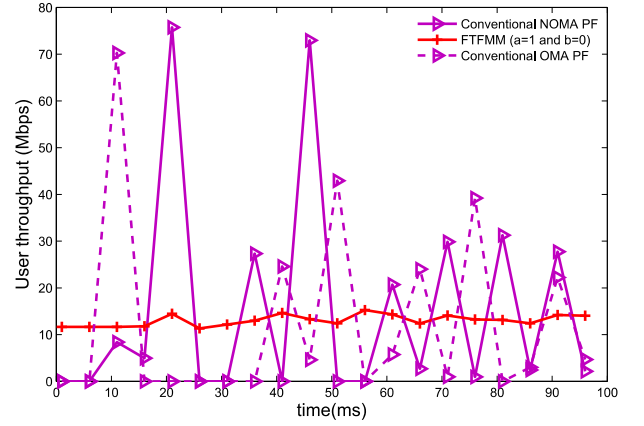
time. On the contrary, with the proposed approach, a non-zero rate is allocated to the least privileged users from the first scheduling slot, and remains stable for all the following slots. With the conventional PF scheduler, users experiencing bad channel conditions are unlikely to be chosen and their achieved data rates is then equal to zero at the beginning of the scheduling process, while, using the proposed metric, zero rates are avoided thus suppressing the rate latency.

V. CONCLUSION

In this paper, we have proposed a novel resource allocation scheme for improving capacity and fairness of NOMA-based systems. For this purpose, a new pairing metric was designed, incorporating a non-uniform inter-subband power allocation scheme. This FTFMM metric targets a flexible balancing between throughput and fairness maximization. Simulation results show that the proposed scheme allow a significant increase in the system capacity and long-term fairness, when compared to a conventional NOMA-based PF scheduler. Furthermore, the proposed technique achieves a higher level of

fairness within each scheduling slot, which improves the QoE of each user.

REFERENCES

[1] http://www.ericsson.com/res/docs/2015/ericsson-mobility-report-june-2015.pdf

[2] Otao, N., Kishiyama, Y., and Higuchi, K., "Performance of non-orthogonal access with SIC in cellular downlink using proportional fair-based resource allocation", in *International Symposium on Wireless Communication Systems (ISWCS 2012)*, 2012.

[3] Saito, Y., Kishiyama, Y., Benjebbour, A., Nakamura, T., Li, A., and Higuchi, K., "Non-orthogonal multiple access (NOMA) for Cellular future radio access", in *Proc. IEEE Vehicular Technology Conference (VTC 2013-Spring)*, 2013.

[4] Schaepperle, J., "Throughput of a wireless cell using superposition based multiple-access with optimized scheduling", in *Proc. IEEE Personal Indoor and mobile radio Communications (PIMRC 2010)*, 2010.

[5] Ding, Z., Yang, Z., Fan, P., and Poor, H., "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users", *IEEE Signal Processing Letters*, 21(12),2014, pp. 1501-1505.

[6] Hojeij, M.R., Farah, J., AbdelNour, C., and Douillard, C., "New Optimal and Suboptimal Resource Allocation Techniques for Downlink Non-orthogonal Multiple Access", *Wireless Personal Communications*, vol 87(3), 2016, pp. 837-867.

[7] Kelly, F.P., Maulloo, A.K., Tan, and D.K.H., "Rate control for communication networks: shadow prices, proportional fairness and stability", *Journal of the Operational Research Society*, vol 49(3), 1998, pp. 237-252.

[8] Tse, D., and Viswanath, P., Fundamentals of Wireless Communication, Cambridge University Press, 2005.

[9] Tomida, S., and Higuchi, K., "Non-orthogonal Access with SIC in Cellular Downlink for User Fairness Enhancement", in *Proc. International Symposium on Intelligent Signal Processing and Comm. Systems (ISPACS 2011)*, 2011.

[10] Viswanath, P., Tse, D., and Laroia, R., "Opportunistic Beamforming using Dumb Antennas", *IEEE Transactions on Information Theory*, vol 48(6), 2002, pp. 1277-1294.

[11] Saito, Y., Benjebbour, A., Kishiyama, Y., and Nakamura, T., "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)", in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2013)*, 2013.

[12] Benjebbour, A., Saito, Y., Kishiyama, Y., Li, A., Harada, A, and Nakamura, T., "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access", in *Proc. International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2013.

[13] Y. Saito, Y., Benjebbour, A., Kishiyama,Y., and Nakamura, T.,"System-level performance of downlink non-orthogonal multiple access (NOMA) under various environments" in *proc. IEEE Vehicular Technology Conference (VTC 2015-Spring)*, 2015.

[14] Hojeij, M.-R., Abdel Nour, C., Farah, J., Douillard, C., "Waterfilling-based proportional fairness scheduler for downlink non-orthogonal multiple access", *IEEE Wireless Communication Letters*, 6(2), 2017, pp. 230-233.

[15] Benjebbour, A., Li, A., Saito, Y., Kishiyama, Y., Harada, A., and Nakamura, T., "System-level performance of downlink NOMA for future LTE enhancements" in *Proc. IEEE Global Communications Conference (Globecom 2013)*, 2013.

[16] Dianati, M., Shen, X., and Naik, S., "A new fairness index for radio resource allocation in wireless networks", in *Proc. IEEE Wireless Communication and Networking Conference*, 2005.