Performance of NOMA with Massive MIMO for 5G

Mário Marques da Silva
marques.silva@ieee.org

Rui Dinis
rdinis@fct.unl.pt

***Abstract*—This paper studies the performance of Non-Orthogonal Multiple Access (NOMA), for both conventional and cooperative NOMA, associated to Single Carrier with Frequency Domain Equalization (SC-FDE) block transmission technique, and to massive Multiple-Input Multiple-Output (MIMO). 5G makes use of millimeter wave communications (mm-Wave) and massive MIMO (m-MIMO). Moreover, NOMA is a potential solution, to be incorporated in future releases of 5G, to increase the capacity required to support extremely high number of Internet of Things (IoT). It is shown that combining NOMA with SC-FDE and m-MIMO only suffers a moderate degradation, while a capacity gain is achieved due to the sharing of the spectrum.**

Keywords- NOMA; 5G; Massive MIMO; SC-FDE; mm-Wave.

I. INTRODUCTION

The Fourth Industrial Revolution is a new change of paradigm in the way society and organizations are shaped, consisting of a massive replacement of humans by robots in a myriad of areas [1]. Robots, such as autonomous vehicles, use sensors and IoT to communicate with the environment, generating large amount of data (big data). Such data is processed by Artificial Intelligence (AI) algorithms to generated knowledge, which will then be used by robots to make decisions. Moreover, the 4th Industrial Revolution also comprises the massive translation of communications between humans into machine-to-machine communications (i.e., IoT).

5G is considered as disruptive relating to previous generations because the added value of this new generation is not only higher speeds, but also increased capacity to face the massive quantity of IoT devices and lower latencies to face the emergent services such as autonomous vehicles or remote surgeries. 5G Communications give a strong contribution to the implementation of the Fourth Industrial Revolution in a wide range of areas, such as in autonomous vehicles, smart cities, smart industries and agriculture, remote surgeries, etc. [2]. Besides the higher data rates and lower latencies, 5G allows direct device-to-device communications (without being through a base station), which is especially important in terms of IoT for the implementation of e.g., Smart Cities or Autonomous Vehicles. One important novelty of 5G relies on the implementation of three use cases to provide different services. As can be seen from Figure 1, 5G comprises three groups of use cases: (1) Enhanced Mobile Broadband (eMBB), consisting of the traditional cellular communications when mobility is a requirement, through a base station, but providing higher speeds and some additional special features as compared to the previous generation of cellular system. This group of use cases is specific of services that require high mobility and high data rate; (2) Massive Machine Type Communications (mMTC), being well suited for communications requiring low power and low duty cycle, typically for IoT devices used in smart cities, smart industries or smart logistics. This group of use cases is specific of services that require high capacity (high number of devices per square kilometer); (3) URLLC relies on communications which are delay sensitive, and sensitive to corrupted data, such as in autonomous vehicles or remote surgeries.

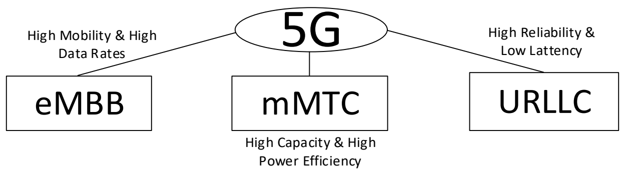

FIGURE 1 THE THREE GROUPS OF USE CASES OF 5G

The extremely high number of IoT devices required to support Smart Cities, autonomous vehicles or smart logistics bring important limitations in terms of spectrum availability. By allowing the spectrum by different users, NOMA is a potential solution for this limitation [3]. It is known that the sharing of spectrum with regular power-ordered NOMA tends to be limited in terms of performance [4]. Cooperative NOMA aims to mitigate such limitation by making lower power users relaying, over the air, the signals of higher power users, clean of interference.

Two important techniques that allow 5G communications achieving the initial requirements in terms of throughputs and network capacity are [3]: (1) Massive MIMO and (2) millimeter-wave communications. Employing carrier frequencies of around 60 GHz, mm-Wave presents the advantage of having much higher channel coherence bandwidth but experiencing much higher path loss. On the other hand, block transmission techniques, such as Orthogonal Frequency Division Multiplexing (OFDM) or Single-Carrier with Frequency-Domain Equalization (SC-FDE) are widely employed to mitigate Intersymbol Interference (ISI). SC-FDE tends to be a better solution due to its lower peak-to-average power ratio [5], especially in the uplink. ISI is also commonly mitigated by employing MIMO systems. A system that combines m-MIMO with mm-wave and with a block transmission technique, alongside with NOMA [6], results in an efficient use of the available spectrum, while keeping the interferences at a low level, and therefore, improving the performance. MIMO receivers are normally associated with a high level of complexity and processing requirements, being even more demanding in case of m-MIMO. Zero Forcing (ZF) receivers require the inversion of the channel matrix for each frequency component of the channel [3]. Maximum Ration

Combiner (MRC) is a receiver that can be employed to reduce the complexity, by avoiding the need to compute the inversion of the channel matrix, for each frequency component of the channel [3].

This paper is organized as follows: section 2 describes the system and signal characterization for m-MIMO using SC-FDE transmissions; section 3 deals with the receiver design for NOMA; section 4 analyses the performance results and section 5 concludes the article.

II. System and Signal Model

This paper considers SC-FDE signals, with multi-layer MIMO, associated to Quadrature Phase Shift Keying (QPSK) modulation. We consider T parallel data streams, with a diversity of order R, corresponding to the number of receiving antennas. A block transmission of length N is adopted. The symbols transmitted by the tth transmitting antenna is given by $\{x_n^{(t)}; n=0,1,...,N-1\} = IDFT\left\{X_k^{(t)}; k=0,1,...,N-1\right\}$, and the received block associated with the rth receiving antenna is $\{y_n^{(r)}; k=0,1,...,N-1\} = IDFT\left\{Y_k^{(r)}; k=0,1,...,N-1\right\}$ [7]. The frequency-domain block $\{Y_k^{(r)}; k=0,1,...,N-1\}$ satisfies

$$\mathbf{Y}_k = \left[Y_k^{(1)},...,Y_k^{(R)}\right]^T = \mathbf{H}_k\mathbf{X}_k + \mathbf{N}_k \tag{1}$$

where $\mathbf{X}_k = \left[X_k^{(1)},...,X_k^{(T)}\right]^T$, and where $\mathbf{H}_k$ denotes the $T \times R$ channel matrix for the kth subcarrier, with $(r,t)th$ element $\mathbf{H}_k^{(t,r)}$, and with $\left\{H_k^{(t,r)}; k=0,1,...,N-1\right\} = DFT\left\{h_n^{(t,r)}; n=0,1,...,N-1\right\}$. Finally, N_k represents the frequency-domain block channel noise for that subcarrier, with $\mathbf{N}_k = \left[N_k^{(1)},...,N_k^{(R)}\right]^T$ representing the $1 \times R$ channel matrix of the noise.

For a non-iterative receiver, the frequency domain estimated data symbols $\tilde{\mathbf{X}}_k = \left[\tilde{X}_k^{(1)},...,\tilde{X}_k^{(R)}\right]^T$ comes [3]:

$$\tilde{\mathbf{X}}_k = \mathbf{B}_k\mathbf{Y}_k \tag{2}$$

where $\mathbf{B}_k$ is defined in [7], separately, for the ZF and MRC, namely as $\mathbf{B}_k = \left(\mathbf{H}_k^H\mathbf{H}_k\right)^{-1}\mathbf{H}_k^H$ for ZF and $\mathbf{B}_k = \mathbf{H}_k^H$ for the MRC.

Different MIMO decoders can be employed, such as the ZF or the MRC. A limitation of the ZF relies on its high level of complexity, as it computes the pseudo-inverse of the channel matrix, for each frequency component. To make the receiver simples, the MRC can be employed. Nevertheless, is presents some level of residual interference generated in the decoding process for moderate values of T/R, which can be mitigated by employing an iterative receiver.

For moderate values of T/R, the level of interference can still be representative, which can be mitigated by implementing the iterative receiver [3][7]:

$$\tilde{\mathbf{X}}_k = \mathbf{B}_k^H\mathbf{Y}_k - \mathbf{C}_k\bar{\mathbf{X}}_k, \tag{3}$$

where $\tilde{\mathbf{X}}_k = \left[\tilde{X}_k^{(1)},...,\tilde{X}_k^{(R)}\right]^T$ stands for the frequency domain estimated data symbols. Moreover, the interference cancellation matrix $\mathbf{C}_k$ can be computed as defined in [7]:

$$\mathbf{C}_k = \mathbf{A}_k^H\mathbf{H}_k - \mathbf{I}, \tag{4}$$

where $\mathbf{I}$ is an $R \times R$ identity matrix.

III. NOMA Receiver

The extremely high number of IoT devices comprised by the 4th Industrial Revolution, such as used in autonomous vehicles of smart cities, need to be interconnected. A great limitation of 5G relies on the unavailability of spectrum to interconnect all such IoT devices. A possible solution for such limitation relies on using NOMA, as it comprises the sharing of spectrum by different users [4][6], while signals are separated in the power domain.

Using NOMA, users closer to the base station use lower transmit powers, while users further from the base station require higher transmit powers levels. Consequently, since the signals of different users that share the spectrum present different received power levels, NOMA uses such property to differentiate signals. A Successive Interference Cancellation (SIC) is employed at the receiver to detect the signals by their descending power levels, allowing the differentiation of different NOMA signals that share the spectrum.

A. Conventional NOMA

In the case of conventional NOMA there are two limits of scenarios [3][4]: (a) signals with power levels higher than the reference user are estimated and cancelled by the SIC, before the reference user's signal is detected; (b) signals with power levels lower than the reference user are not cancelled, as they represent low level of interference. Instead of conventional NOMA, cooperative NOMA can be employed. In this latter case, all interfering signals can be cancelled, including those with lower powers than the reference one.

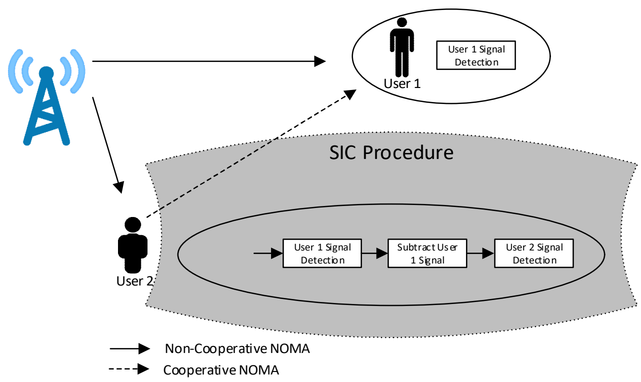

Figure 2 NOMA and Cooperative NOMA Concepts

Let us assume that there are a total of U users sharing the spectrum. The received signal at the rth receiving antenna, and nth data symbol, is the cumulative sum of the $\{U;(u=1...U)\}$ signals that share the spectrum using NOMA:

$$y_n^{(r)}(t)=\sum_{u=1}^{U} y_{u,n}^{(r)}(t) \quad (5)$$

Considering only NOMA users, (1) can be re-written as [3]:

$$\mathbf{Y}_{u,k}=\left[Y_{u,k}^{(1)},...,Y_{u,k}^{(R)}\right]^T=\mathbf{H}_{u,k}\mathbf{X}_{u,k}+\mathbf{N}_{u,k} \quad (6)$$

Due to the descending received power order of cancellation performed by the SIC, all weaker users benefit from the fact that stronger users are cancelled first. The signal y'_n at the output of the SIC, for the nth data symbol, after the cancellation of the higher power users U', comes [3]:

$$y'_n=y_{u,n}-\sum_{i=1}^{U'}\hat{y}_{i,n}\,, \quad (7)$$

where $\hat{y}_{i,n}$ stands for the estimate of the received signal of the ith interfering user and nth data symbol.

B. Cooperative NOMA

With conventional NOMA, in case the reference user presents higher power, the SIC is not able to cancel interference, and the lower power users represent residual interference. This limitation can be overcome by using Cooperative NOMA. As seen in Figure 2, Cooperative NOMA aims to mitigate this limitation by making lower power users relaying, over the air, the signals of higher power users, clean of interference. The higher power users can use any algorithm to combine such relayed signal(s) with the one received directly from the base station [4]. We assume the relay of type decode and forward.

As seen in Figure 2, assuming that user 1 is the reference one (higher power user), and that user 2 is an interfering user (lower power user), as in the case of conventional NOMA, user 2 employs the SIC for subtracting the signal of user 1 from the overall received signal (using the SIC), before user 2 is detected. Cooperative NOMA considers that interfering users (user 2, in the case of Figure 2) relay, over the air, the signals detected by the SIC (user 1, in the case of Figure 2). This allows that a user with higher received power level (user 1, in the case of Figure 2) receives several copies of its signal: (a) the signal received directly from the base station, where the interference created by lower power users cannot be cancelled by the SIC; (b) the signal relayed by the other interfering users (user 2, in the case of Figure 2), which can be clean of interference of lower power users. Let us focus on Figure 2. User 2 signal is assumed presenting lower power and, once detected, can be regenerated and cancelled from the overall signal of user 1, before it is relayed, as this is a matter of introducing an additional iteration. The signal of user 1 (higher power user) can then be relayed over the air. The several copies of signals of user 1 (higher power) can then be combined at the data symbol level $\hat{b}$, using any combining algorithm. [3] presents a combining algorithm based on the Mean Squared Error.

IV. Simulation Results

The performance was evaluated using Monte Carlo simulations, observing the Bit Error Rate (BER) as a function of E_b/N_0, where E_b is the energy of the transmitted bits and N_0 is the one-sided power spectral density of the noise. Ideal channel estimation and the SC-FDE was assumed, with Quaternary Phase Shift Keying (QPSK) modulation and with a block length of $N = 256$ symbols (similar results were observed for other values of N, provided that $N >> 1$). A Rayleigh fading channel was considered with 16 uncorrelated equal power paths. The duration of the useful part of the blocks (N symbols) is 1µs and the cyclic prefix has a duration of 0.125 µs. The Match Filter Bound (MFB) curve is a way to measure the channel modeled by the sum of delayed and independently Rayleigh-fading rays, which can be viewed as a lower bound.

Figure 3 shows the performance results for conventional NOMA (designated in figures as "NOMA") and Cooperative NOMA (designated in figures as "COOP NOMA"), with 4X32 MIMO, considering two receivers: ZF and MRC. Two NOMA users were considered in the simulation, with receive power levels [1 0.5], where the first value in the vector [1 0.5] (1, in this case) corresponds always to the power of the reference user, being the other value the power of the interfering user (0.5, in this case). In this scenario, the power of the interfering user is 3 dB lower than that of the reference user. Weak users are those that tends to be further from the base station (can also be due to e.g., fading), and therefore the propagation losses are higher. With NOMA, this is mitigated by employing higher transmit power. Therefore, it is assumed that the interfering user with transmit power 0.5 tends to be closer to the base station than the reference user, whose transmit power is 1.

As can be seen, the results of Figure 3 obtained with conventional NOMA are quite limited, due to the existence of residual interference. Noteworthy is that the SIC that is part of the receiver only detects, regenerates and cancels users' signals with powers higher than those of the reference user, which is not the case here (the interfering user has power 0.5, which is not cancelled). This explains the low performance achieved with conventional NOMA, for both ZF and MRC. It is also viewed that, with conventional NOMA, the ZF performs better than the MRC. However, the MRC consists of an iterative receiver that estimates the transmitted symbols and aims to improve such estimate in each iteration. With high level of interference associated with the non-cancelled interfering NOMA user, the symbol estimates performed in each iteration of the MRC receiver is poor, and therefore, it is not able to perform well. On the other hand, even with the noise enhancement typical of the ZF receiver [7], it performs better than that of the MRC.

Cooperative NOMA comprises the cancellation of the interfering signals associated with all users and exploits diversity. Cooperative NOMA considers that the lower power users retransmit the symbols detected by the SIC of higher power users (typically using decode and forward, in Time Division Multiplexing). These additional signals are utilized by higher power users to exploit diversity, as the same signals are also received directly from the base station (assuming downlink), and a copy of them (clean of interferences) is relayed by the lower power users. These signals are combined to improve performance. Figure 3 shows that such combination of signals performed with the Cooperative NOMA results in a good performance improvement, when compared with conventional NOMA, for both MRC and ZF, whose performances are close to that of the Match Filter Bound (MFB). Note that the MFB curve is a way to measure the channel modelled by the sum of delayed and independently Rayleigh-fading rays, which can be viewed as a lower bound. Figure 3 also evidences that, with Cooperative NOMA, the MRC performs better than the ZF. This derives from the fact that Cooperative NOMA allows the cancellation of all NOMA interference (not only of users with higher receive power levels) and, additionally, allows exploiting diversity. This lower level of interference allows the iterative MRC receiver to improve the symbol estimates in different iterations. On the other hand, the performance of the ZF is limited, as this receiver presents noise enhancement. Furthermore, the level of complexity of the ZF is substantially higher than that of the MRC, as it requires the computation of the pseudo-inverse of the channel matrix for each frequency component.

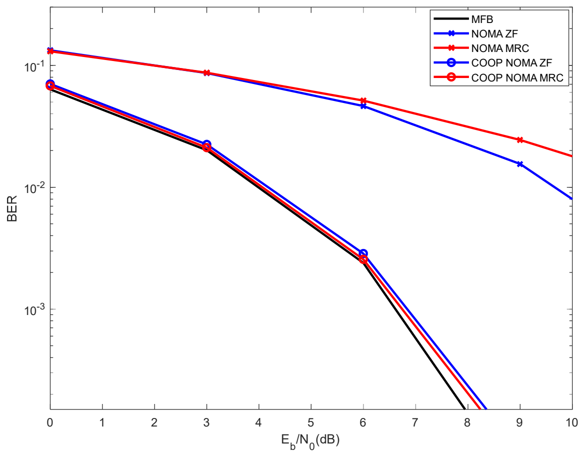

FIGURE 3 – RESULTS FOR 2 NOMA USERS WITH POWERS [1 0.5], 4X32 MIMO

Figure 4 shows the BER performance in the same scenario as that of Figure 3, with the difference that the power of users is [0.5 1]. In this scenario, the reference user tends to be closer to the base station (power 0.5), while the interfering user tends to be further from the base station (power 1). As previously mentioned, we refer to the distance from the base station only for explanation purposes, as the fading or the power control can also make the received power suffer variations. Noteworthy is that the conventional NOMA receiver comprises the detection, regeneration and cancellation of the users' signals by their descending order of powers (up to the power of the reference user), before the detection of the reference user takes place. Consequently, the detection of the reference user tends to be clean of interferences and the performance achieved with conventional NOMA is already good. In this scenario, since the reference user presents a power level (e.g., 0.5) which is 3 dB lower than the interfering user (e.g., 1), such signal detection is not carried out in the SIC receiver of the interfering user, and Cooperative NOMA is not implemented, as it does not bring any added value. Moreover, note due to its lower power, the SIC of the reference user, using the mode of conventional NOMA, is now able to cancel the interfering signal, as the power of the interfering user is higher, making the effectiveness of cooperative NOMA useless.

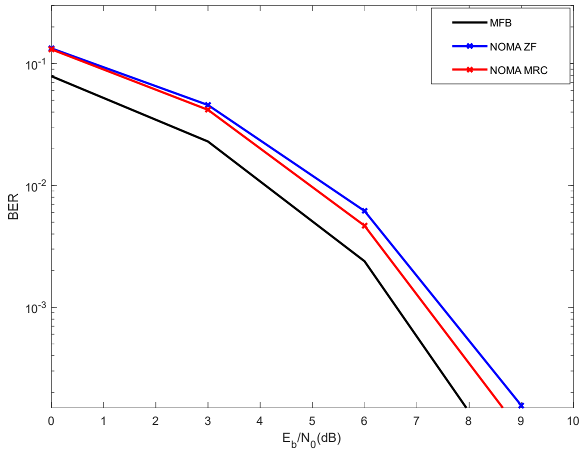

FIGURE 4 – RESULTS FOR 2 NOMA USERS WITH POWERS [0.5 1], 4X32 MIMO

V. CONCLUSIONS

The extremely high number of IoT devices comprised by the 4th Industrial Revolution, such as used in autonomous vehicles of smart cities, need to be interconnected. A great limitation of 5G relies on the unavailability of spectrum to interconnect all such IoT devices. A possible solution for such limitation relies on using NOMA, as it comprises the sharing of spectrum by different users, while signals are separated in the power domain. Nevertheless, as the number of NOMA users increases, a degradation of the BER performance tends to occur, especially for users that have higher receive power levels. Cooperative NOMA brings special added value for users with higher power levels. With conventional NOMA, the SIC only detects, regenerates and cancels users' signals with power levels higher than those of the reference one, and those with less received power levels are not cancelled, representing residual interference, and degrading the performance.

This paper considers NOMA and Cooperative NOMA, associated to m-MIMO and SC-FDE. Two receivers were employed: ZF and MRC. Due to the increased simplicity, the MRC tends to be a better solution, whose performance is very close to that obtained with the ZF. This can be considered as a good combination to achieve the requirements of future evolutions of 5G.

ACKNOWLEDGMENT

This work is funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/EEA/50008/2020.

REFERENCES

[1] 1 K. Zhou, T. Liu, L. Zhou, “Industry 4.0: Towards future industrial opportunities and challenges,” Proc. of 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjiajie, China, Aug. 2015.

[2] 2 M. Marques da Silva, J. Guerreiro, “On the 5G and Beyond”, MDPI Applied Sciences, 2020, 10, 7091, 26 October 2020 (https://www.mdpi.com/2076-3417/10/20/7091)

[3] 3 Marques da Silva, M., Dinis, R., “Power-Ordered NOMA with Massive MIMO for 5G Systems”, MDPI Applied Sciences, 11(8), 3541, 15 April 2021 (https://doi.org/10.3390/app11083541)

[4] 4 K. Higuchi and A. Benjebbour, "Non-Orthogonal Multiple Access (NOMA) with Successive Interference Cancellation for Future Radio Access", IEICE Trans. Commun., vol. E98.B, no. 3, pp. 403-14, 2015.

[5] 5 Marques da Silva, M.; Dinis, R. Iterative Frequency-Domain Detection and Channel Estimation for Space-Time Block Codes European Transactions on Telecommunications; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2011; Volume 22, pp. 339–351.

[6] 6 Z. Ding, M. Peng and H. V. Poor, "Cooperative Non-Orthogonal Multiple Access in 5G Systems", IEEE Commun. Lett., vol. 19, no. 8, pp. 1462-65, Aug. 2015.

[7] 7 Marques da Silva, M.; Dinis, R. A Simplified Massive MIMO Implemented with Pre or Post-Processing. Phys. Commun. 2017, 1, 1–12, doi:10.1016/j.phycom.2017.06.002.

[8] 8 L. Dai et al., "Non-Orthogonal Multiple Access for 5G: Solutions Challenges Opportunities and Future Research Trends", IEEE Commun. Mag., vol. 53, no. 9, pp. 74-81, Sept. 2015.