Semi-Static Modulation Compression Optimization for Next Generation RANs

Sandra Lagén*, Xavier Gelabert†, Lorenza Giupponi*, Andreas Hansson†

*Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Castelldefels, Barcelona, Spain

†Huawei Technologies Sweden AB, Kista, Sweden

{sandra.lagen, lorenza.giupponi}@cttc.es, {xavier.gelabert, andreas.hansson}@huawei.com

Abstract—Next generation Radio Access Networks (RANs) consider virtualized architectures in which base station functions are distributed in different logical nodes that are connected through fronthaul links. To reduce the required fronthaul capacity, modulation compression is considered as a key enabler. Modulation compression achieves fronthaul capacity reduction at the cost of reducing the maximum modulation order that can be used over the air interface, thus creating a cell-fronthaul trade-off. The trade-off is further accentuated and needs to be optimized appropriately when multiple cells share the same fronthaul link. In this paper, a multi-cell scenario with a shared fronthaul link across multiple cells is considered, and we focus on optimizing the modulation compression of each cell. We propose semi-static optimization procedures that aim at maximizing the air interface performance subject to a shared fronthaul capacity constraint, by taking into account the average traffic load and system configuration of every cell. The problem is formulated as a convex optimization problem, which allows deriving the optimal maximum modulation order that is permitted per cell, under two different optimization criteria. Then, we use a dynamic multi-cell 5G NR system-level simulator based on ns-3 to evaluate the proposed optimized solutions.

Index Terms—next generation RAN architectures, C-RAN, fronthaul compression, modulation compression, optimization.

I. INTRODUCTION

5G networks are anticipated to enable diverse applications with various requirements on data rates and latency that pose significant challenges for the RAN architecture [1]. To effectively cope with such challenges, the Centralized RAN (C-RAN) is widely seen as the most promising technology to provide high spectral efficiency, low power consumption, resource pooling, scalability, and layer interworking [2]. Unlike legacy distributed RAN architectures, prevalent in early releases of 4G LTE, C-RAN provides the means for the migration of the Baseband (BB) processing of multiple cells towards a centralized location [3]. In particular, 3GPP considers a hybrid C-RAN architecture for 5G NR [4]. A 5G base station consists of a Centralized Unit (CU) and one or more Distributed Units (DUs) connected to the CU via a midhaul interface. In turn, each DU connects to one or more Radio Units (RUs) through a fronthaul interface. The BB processing of a cell can then be split between different BB entities, located at the CU, the DU, and the RU. Studies about different functional split options and architectures for C-RAN were conducted in 3GPP Release-14/15 [5], [6], while now are progressing in the context of the Open-RAN (O-RAN) Alliance [7], [8].

The main obstacles in deploying C-RAN architectures are tight fronthaul capacity and latency requirements [9]. Such requirements are further accentuated in 5G NR by the use of wider channel bandwidths, massive antennas, and higher modulation orders. Given a particular C-RAN architecture, with specific functional splits and a fixed fronthaul topology, it is of utmost importance to provide efficient methods to control and optimize the utilization of fronthaul resources with limited (or no) impact on the air interface performance. In particular, under low layer splits [6], it is critical to use some kind of fronthaul compression technique to reduce the data rate across one or multiple RUs for all the served users in an adaptative manner, so as to conform to the available fronthaul capacity. Fronthaul compression techniques have been widely studied in the literature since the introduction of C-RAN in 4G LTE, from the points of view of signal processing and information theory (see a comprehensive survey in [2]). Recently, more practical schemes have been defined in [8]. Among the different compression techniques, modulation compression is seen as a promising technique because it allows a dramatic reduction of the required downlink fronthaul capacity without degradation of the signal quality sent over the fronthaul interface and without the need for complex algorithms/schemes [8, §A.5]. Basically, the IQ bitwidth sent over the fronthaul interface is reduced by means of reducing the maximum modulation order that is permitted over the air interface.

Besides, scenarios with multiple cells sharing the same fixed fronthaul (i.e., multiple RUs connected by using a switch as a hub that connects to the DUs through a single fronthaul link) are particularly challenging, as they result in a shared fronthaul capacity utilization. This calls for system wide designs and optimization solutions, whose evaluations need to be based on dynamic, end-to-end, multi-cell simulations. In modulation compression, cells with unused spectrum resources can limit their maximum allowed modulation orders to cater to particular fronthaul congestion situations. This would benefit other jeopardized cells that could increase their maximum allowed modulation order, or allow for the deployment of additional RUs in challenging areas. To the best of our knowledge, modulation compression techniques in multi-cell scenarios with shared fronthaul capacity have not been previously studied in the literature.

In this paper, we consider a C-RAN architecture composed of multiple cells, using intra-PHY functional split 7.2x and

modulation compression for the downlink (as adopted in O-RAN [8]), with a shared fronthaul link towards the cells (which is present in ring and star fronthaul topologies [10]). In this context, we derive semi-static modulation compression methods, in which the maximum allowed modulation order per cell is adjusted in a semi-static manner, considering the statistics of the per-cell traffic loads, the system configuration of every cell, and the total available fronthaul capacity. Then, we evaluate the proposed methods in a dynamic multi-cell 5G system-level simulator, using the NR module of the open-source, full-stack, end-to-end ns-3 network simulator [11].

The paper is organized as follows. Section II describes the system model and modeling assumptions. Section III presents the semi-static optimization of modulation compression in multi-cell scenarios with shared fronthaul. Then, we present the ns-3 based simulation scenario and assess the semi-static modulation compression performance in Section IV. Finally, Section V concludes the paper.

II. SYSTEM MODEL

Consider a system model that is composed of N cells, in which each cell ($n = 1, \ldots, N$) serves K_n user equipments (UEs). So, in the system there is a total of $K = \sum_{n=1}^{N} K_n$ UEs. Assume a hybrid C-RAN architecture with centralized, distributed, and radio units, using functional split Option 2 for the CU-DU split and intra-PHY functional split 7.2x for the DU-RU split, as standardized by 3GPP [4] and O-RAN [8]. For the CU/DU/RU deployment scenario, we consider that the CUs and DUs of all cells are placed together in a centralized entity (in an edge or regional cloud), while the RUs are located in proprietary cell sites (which is in line with Scenario B of O-RAN [7]). Therefore, with the mentioned functional splits and the deployment scenario under consideration, the high-PHY functionalities and above processing of all the cells are deployed together in the centralized cloud, while low-PHY functionalities and RF processing of each cell are placed in the RU of each site [5]. Without loss of generality, we focus on downlink data transmissions. Specifically, in split 7.2x for the downlink, the inverse Fast Fourier Transform (IFFT), cyclic prefix addition, resource mapping and precoding functions reside in the RU, while the remaining PHY functions (high-PHY) reside in the DU [8]. A low layer fronthaul interface is used to interconnect RUs with DUs [8]. In this regard, we assume a star fronthaul topology, so that the N RUs share the same fronthaul link that is characterized by a limited capacity (of C bps). The system model is shown in Fig. 1.

To reduce the required fronthaul capacity in multi-cell scenarios with shared fronthaul interface, we adopt the so-called modulation compression technique. Modulation compression is a lossless IQ data compression (see definition in [8, §A.5]). Basically, it assumes that modulated data symbols can be represented by a limited number of IQ bits provided some alterations in the modulation constellation points are given. Specifically, to represent the constellation points as IQ values that overlap over different constellation sizes, the constellations are shifted to allow any IQ value to represent any constellation
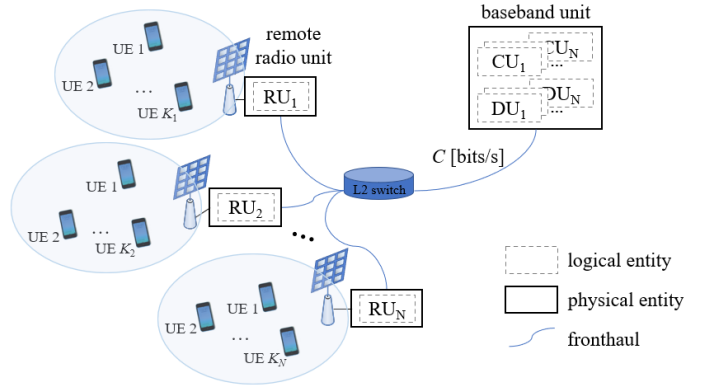

Fig. 1: Deployment scenario composed of N cells, each serving K_n UEs, with shared fronthaul interfacing between distributed and radio units (DUs and RUs) in a star topology. CUs and DUs are placed together in a centralized cloud, and each RU is located in a cell site.

point. Once the constellations are shifted, the IQ values may be encoded in a limited number of bits, which corresponds to the modulation order needed to represent the largest constellation in the compression block (e.g., 6 bits for a multiplexed 64QAM and QPSK block).

In particular, we focus on semi-static optimizations of the modulation compression per cell. Such a semi-static approach can, for example, be supported by the O-RAN management plane [10]. The semi-static adjustment is feasible in terms of signaling overhead, since the maximum allowed modulation order per cell is communicated to every RU after certain periods that can reach the order of minutes.

A. Air interface: Cell load modeling

For the air interface, we consider the *normalized load* (NL) as a metric of interest. The NL of a cell n (ρ_n) is defined as the offered traffic load over the maximum served traffic (i.e., the capacity) at the n-th cell. It is given by:

$$\rho_n = \frac{\sum_{k=1}^{K_n} L_{n,k}\lambda_{n,k}}{N_{symb,n}N_{sc,n}N_{L,n}M_n 1000} \triangleq \frac{\alpha_n}{\beta_n M_n}, \quad (1)$$

where $\lambda_{n,k}$ and $L_{n,k}$ are the packet arrival rate (packets/s) and the packet length (bits) of the k-th UE in the n-th cell, M_n is the maximum modulation order that is permitted in the n-th cell, $N_{symb,n}$ is the number of OFDM symbols in 1ms, $N_{sc,n}$ is the number of subcarriers within the channel bandwidth used by the n-th cell, and $N_{L,n}$ is the number of MIMO layers supported by the n-th cell. In Eq. (1), $\alpha_n = \sum_{k=1}^{K_n} L_{n,k}\lambda_{n,k}$ can be estimated based on the average traffic load of the n-th cell and $\beta_n = 1000N_{symb,n}N_{sc,n}N_{L,n}$ can be obtained from the air frame structure and MIMO capabilities of the n-th cell.

Note that the NL of a cell (ρ_n) could be larger than 1 by definition, if the offered traffic load is very high. However, systems should be properly designed so as not to congest the users' queues, i.e., should work under stable conditions with $\rho_n < 1$ [12], [13]. The motivation to consider the NL as a metric of interest is that we do not want Radio Link Control (RLC) buffers to grow exponentially and we would like to have the cells working in non-saturated air interface conditions. In addition, as shown in [12], ρ_n in (1) is related to the average

number of bits in the RLC buffers at the n-th cell, in such a way that, in a single-cell system, minimizing the NL is equivalent to minimizing the average number of bits in the RLC queues.

From Eq. (1), it can be observed that increasing the maximum allowed modulation order (M_n) leads to a low NL (ρ_n), because the cell is able to send the same amount of bits over fewer subcarriers/symbols, i.e., the cell has higher spectral efficiency (in bits/s/Hz). On the contrary, reducing M_n leads to a high ρ_n, due to the lower spectral efficiency, which may increase the buffering delay.

B. Fronthaul interface: Capacity modeling

Regarding the fronthaul, the total required capacity (C_{req}) that is needed through all the cells ($n = 1, \ldots, N$) using functional split 7.2x is:

$$C_{\text{req}} = \sum_{n=1}^{N} \left(N_{symb,n} N_{sc,n} N_{L,n} M_n 1000 + K_n \mathcal{O}_{\text{mac}} \right)$$

$$= \sum_{n=1}^{N} \beta_n M_n + K \mathcal{O}_{\text{mac}}, \quad (2)$$

where K is the total number of UEs associated to the cells that share the fronthaul, and $\mathcal{O}_{\text{mac}}$ is the MAC overhead per UE (in bps) needed to implement the intra-PHY 7.2x split (including PRB assignment information, antenna configuration, beamforming vectors, among others). Note that here we do not assume dynamic adaptation, but just consider the required fronthaul capacity assuming that, at a certain slot, all PRBs/symbols of the air interface may be occupied by a RU, as considered in the literature [5], [9]. That is, we consider a worst-case design scenario.

From Eq. (2), it can be observed that increasing M_n leads to a higher required fronthaul capacity (C_{req}). On the contrary, reducing M_n reduces the amount of data that needs to be sent through the fronthaul.

C. Cell-fronthaul trade-off

From previous expressions in Eq. (1) and Eq. (2), it is evident that the selection of the maximum allowed modulation order per cell creates an air interface-fronthaul trade-off that needs to be optimized appropriately. Reducing the maximum allowed modulation order of a cell reduces the required fronthaul capacity and increases the NL of the cell. The trade-off is further accentuated in multi-cell scenarios, when multiple RUs share the fronthaul capacity (as reflected in Fig. 1) and potentially the air interface channel. In those multi-cell settings, not selecting the most adequate maximum modulation order for each of the cells, from a system wide perspective, may lead to situations in which certain cells experience congestion (i.e., high NL) while others have high spectral efficiency and unused spectrum resources (i.e., low NL). Cells in the latter condition could thus increase their spectrum resource usage (increase NL) by reducing the maximum modulation order, which would spare some fronthaul capacity in advantage to cells with more challenging conditions.

An efficient fronthaul compression control mechanism should be designed to exploit the shared fronthaul capacity while avoiding situations of high air interface NL in all the managed cells (which would lead to high packet delays). In this sense, a suitable optimization criterion is the minimization of the maximum NL among all the cells [12], so that the fronthaul capacity is fairly distributed among the cells and more capacity is given to cells with higher offered traffic loads and/or lower number of available spectrum resources. An alternative criterion is the minimization of the weighted sum of NLs [13], so that different priorities can be assigned to every cell.

III. MODULATION COMPRESSION OPTIMIZATION

Following the discussion from the previous section, in this section we set two optimization problems: 1) minimization of the maximum NL; and 2) minimization of the weighted sum of NLs. The objective is to semi-statically configure the maximum modulation order allowed at each RU (M_n) as a function of the system parameters and cell's traffic loads, in such a way that the air interface performance is optimized (by means of the NL) and the maximum capacity of the shared fronthaul is met (i.e., $C_{\text{tot}} \leq C$).

A. Minimization of the maximum NL

If we focus on minimizing the maximum of the NLs (ρ_n) among all the cells, subject to a shared fronthaul capacity constraint, the optimization problem results:

$$\text{minimize}_{\{M_n\}} \quad \max_n \left(\frac{\alpha_n}{\beta_n M_n} \right) \quad (3)$$

$$\text{subject to} \quad \sum_{n=1}^{N} \beta_n M_n + K \mathcal{O}_{\text{mac}} \leq C$$

where C is the available fronthaul capacity (in bps).

Assuming that the maximum allowed modulation order (M_n) is a continuous optimization variable, then the optimization problem in (3) is convex on $\{M_n\}$. Furthermore, the optimal solution can be found in closed form by reformulating the problem, as detailed next. Problem in (3) can be equivalently written through the following optimization problem:

$$\text{minimize}_{\{M_n\},t} \quad t \quad (4)$$

$$\text{subject to} \quad \sum_{n=1}^{N} \beta_n M_n + K \mathcal{O}_{\text{mac}} \leq C$$

$$\frac{\alpha_n}{\beta_n M_n} \leq t \quad \forall n$$

The optimization problem in (4) is jointly convex with respect to $\{M_n\}$ and t. Interestingly, the optimal solution is such that the resulting NL of all cells is the same and the shared fronthaul capacity is fully distributed among all the cells. Therefore, the optimal solution has the following structure: $M_n = \frac{\alpha_n}{t \beta_n}$ for all cells. By including such structure into the fronthaul constraint, we can obtain t as: $t = \sum_{n=1}^{N} \frac{\alpha_n}{C - K \mathcal{O}_{\text{mac}}}$. Finally, by combining both expressions, the optimal solution

results:

$$M_n^* = \frac{\alpha_n}{\beta_n} \frac{C - K\mathcal{O}_{\text{mac}}}{\sum_{n=1}^{N} \alpha_n}. \tag{5}$$

Note that those cells having a larger traffic load will have a larger maximum allowed modulation compression. Also, those cells having more available resources for the air interface (PRBs/symbols) will result in a lower maximum allowed modulation order.

B. Minimization of the weighted sum of NLs

Another option is to consider as objective function the minimization of the weighted sum of NLs subject to the shared fronthaul capacity constraint, i.e.,

$$\text{minimize}_{\{M_n\}} \quad \sum_{n=1}^{N} \mu_n \left(\frac{\alpha_n}{\beta_n M_n} \right) \tag{6}$$

$$\text{subject to} \quad \sum_{n=1}^{N} \beta_n M_n + K\mathcal{O}_{\text{mac}} \leq C$$

where μ_n is a weighting coefficient associated to the priority of the n-th cell.

For continuous $\{M_n\}$, the optimization problem in (6) is convex on $\{M_n\}$. In addition, the optimal solution can be found in closed form by using the Lagrangian function. Let us denote by γ to the Lagrange multiplier associated to the constraint in (6). Then, the Lagrangian function is:

$$L = \sum_{n=1}^{N} \mu_n \left(\frac{\alpha_n}{\beta_n M_n} \right) + \gamma \left(\sum_{n=1}^{N} \beta_n M_n + K\mathcal{O}_{\text{mac}} - C \right). \tag{7}$$

The derivative of L equals to 0 leads to: $\frac{-\mu_n \alpha_n}{\beta_n (M_n)^2} + \gamma \beta_n = 0$, so that the optimal solution has the form: $M_n = \sqrt{\frac{\mu_n \alpha_n}{\gamma (\beta_n)^2}}$. By including it into the constraint in (6), we can isolate the Lagrange multiplier as: $\gamma = \left(\frac{\sum_{n=1}^{N} \sqrt{\mu_n \alpha_n}}{C - K\mathcal{O}_{\text{mac}}} \right)^2$. Finally, by combining the expressions, the optimal solution is:

$$M_n^* = \frac{\sqrt{\mu_n \alpha_n}}{\beta_n} \frac{C - K\mathcal{O}_{\text{mac}}}{\sum_{n=1}^{N} \sqrt{\mu_n \alpha_n}}. \tag{8}$$

Similar to the min-max NL problem, the solution gives a larger modulation order to those cells having a larger traffic load, as well as to those cells having less resources available for the air interface (PRBs/symbols).

Note that both optimization problems (min-max NL and min-sum NL) have been solved assuming that the optimization variables ($\{M_n\}$) are continuous. Thus, there is the need to quantize them into practical modulation orders, by considering the modulation orders available at the cell for the air interface. In a typical NR system that supports up to 256QAM, M_n can take values of 2, 4, 6, or 8. Accordingly, a simple quantization process is to take the floor over even numbers.

IV. SIMULATION RESULTS

For the performance assessment, we use the NR module of ns-3 [11] with the 3GPP spatial channel model developed in [14], compliant with TR 38.901, and new NR-based PHY abstraction models [15].

A. Scenario

We use a typical hexagonal site deployment. Each site has 3 cells, with 3 uniform planar antenna arrays that cover 120° in azimuth each. We simulate two different scenarios: a) $N = 3$ RUs, in which only the central site with 3 cells is deployed; b) $N = 21$ RUs, in which the central site plus six additional sites belonging to the first outer ring are deployed, leading to a total of 21 cells. A 5G scenario with three orthogonal bandwidth parts is assumed; each cell of a site transmits in a separate bandwidth part, so that sectors of the same site do not interfere. We use an Urban Micro scenario, characterized by an inter-site distance of 200 m, RU antenna height of 10 m, and 30 dBm transmit power. $K_n = 2$ or $K_n = 4$ UEs are randomly deployed per cell. The operational bandwidth parts are in the 2 GHz region, with 100 MHz channel bandwidth per cell and a PRB overhead of 0.04 (typical of NR). Numerology 2 is used, i.e., 60 KHz subcarrier spacing. The duplexing mode is TDD, with dynamic downlink-uplink slot structure. Therefore, the number of subcarriers and symbols in 1 ms is: $N_{sc,n} = 1596$ and $N_{symb,n} = 56$. The antenna array configuration consists of 64 directional antenna elements at RUs and 1 isotropic antenna element at UEs. Single MIMO layer is assumed, i.e., $N_{L,n} = 1$. We use Modulation and Coding Scheme (MCS) Table2 of NR that includes up to 256QAM (see [16, Table 5.1.3.1-2]). For HARQ, we use Incremental Redundancy. RLC Unacknowledged Mode (UM) is used.

For each UE, we generate a downlink UDP flow at constant bit rate, with $L_{n,k} = 600$ bytes packet size. We consider that the traffic load varies per cell according to the following expression: $\lambda_{n,k} = 1000 + (n-1) \times 2000$ packets/s in case of 3 RUs ($n = 1, 2, 3$) and according to $\lambda_{n,k} = 1000 + (n-1) \times 200$ packets/s in case of 21 RUs ($n = 1, \ldots, 21$). All UEs in a cell have the same packet arrival rate. This represents a load that varies from 4.8 to 24 Mbps per UE at application layer, which, when adding headers of the higher layers, results roughly from 5 to 25 Mbps traffic load at IP layer per UE.

Regarding the fronthaul, as previously mentioned, we assume the scenario shown in Fig. 1. In this way, CUs/DUs are centralized in the cloud, giving service to all the cells (RUs) of the deployment scenario under consideration, following functional split Option 7-2x (DU-RU). Due to the shared fronthaul topology, the fronthaul capacity is shared among 3 RUs or 21 RUs, depending on the scenario. We assume an available fronthaul capacity of $C = 2$ Gbps and $C = 15$ Gbps for the case of 3 RUs and 21 RUs, respectively. For the required fronthaul capacity, C_{req}, we use the expression in (2) with $\mathcal{O}_{\text{mac}} = 100$ Mbps.

B. Analysis of semi-static solutions

Fig. 2 shows the optimal maximum allowed modulation order (M_n^*) that is obtained for each RU, as a result of the proposed optimization problems (min-max NL and min-sum NL), for the cases of (a) 3 RUs and (b) 21 RUs, under deployments
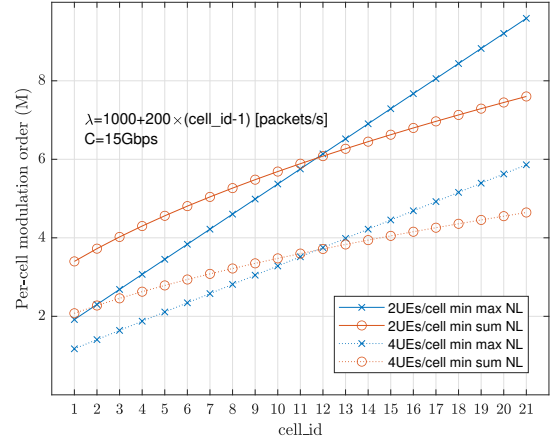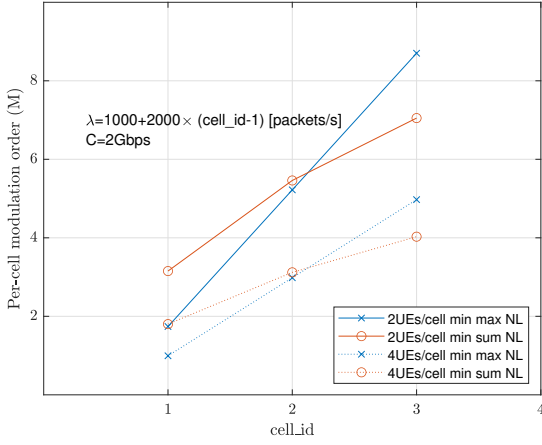
Fig. 2: Optimal modulation order per RU, for different optimization criteria (min-max NL and min-sum NL) and number of UEs/cell (2 and 4). (a) 3 RUs. (b) 21 RUs.

of 2 UEs/cell and 4 UEs/cell. As expected, cells with a larger traffic load (larger $\lambda_{n,k}$), get a larger maximum modulation compression permitted. Also, increasing the number of UEs in the scenario, reduces the maximum modulation order that is allowed on each RU, to meet the shared fronthaul capacity C. As it can be observed, "min-max NL" exhibits a linear trend in the resulting maximum allowed modulation order with respect to the traffic load ($\lambda_{n,k}$), so that all cells tend to have the same NL after the optimization (note that the NL of a cell is dependent on $\lambda_{n,k}/M_n$). On the other hand, the "min-sum NL" tends to determine the maximum modulation order in a more equally distributed fashion, thus leading to different NLs among the cells. In this case, cells with a larger $\lambda_{n,k}$ get a larger NL.

C. End-to-end results

In the end-to-end evaluation, for each deployment scenario (3 or 21 RUs, and 2 or 4 UEs/cell), we assess the impact of using modulation compression by using different strategies:

- static and uniform maximum allowed modulation order across all the RUs, limited to 256QAM, 64QAM, 16QAM, and QPSK. In this case, the required fronthaul capacity may overpass the available fronthaul capacity.
- semi-static solutions that optimize the maximum allowed modulation order per RU (min-max NL and min-sum NL) based on the system configuration, traffic load, and available fronthaul capacity.

As key performance indicators of the air interface, we consider the *mean E2E throughput* (measured at the IP layer) and the *mean E2E latency* of packets that arrive at the IP layer of the UE. Fig. 3 shows the required fronthaul capacity (top) and the obtained air interface performance (middle and bottom), for the cases of (a) 3 RUs and (b) 21 RUs. Results are depicted as a function of the maximum permitted modulation order in case of fixed modulation compression, and for the two considered optimization problems (min-max NL, min-sum NL), for 2 and 4 UEs/cell. In the top subfigure, the available fronthaul capacity is shown with a solid black line.

As expected, when using the same modulation compression across all the RUs, the required fronthaul capacity linearly increases with the maximum allowed modulation order, as well as with the number of UEs/cell. Note that the required fronthaul capacity surpasses the maximum fronthaul capacity for modulation orders of 64QAM and above in case of 2 UEs/cell and for 16QAM and above for 4 UEs/cell, for both the cases of 3 and 21 RUs. Instead, the proposed semi-static solutions allow meeting the available fronthaul capacity in all cases, because the maximum allowed modulation order is adjusted per RU, based on the system configuration (including number of UEs), traffic load, and maximum fronthaul capacity.

For the case of 3 RUs, it can be observed that there is no difference between the air interface performance (delay and throughput) with 2 UEs/cell, since the total traffic load is low enough and the air interface is not appreciably affected by reducing the modulation order. Instead, for the case of 4 UEs/cell, the only static/uniform modulation compression across the RUs that allows meeting the fronthaul capacity requirement (i.e., QPSK) does lead to very high packet delays, and so, to a significant air interface degradation. However, thanks to the proposed algorithms (min-max NL and min-sum NL), we are able to meet the fronthaul capacity, while maintaining a good air interface performance. Among the two proposed solutions, we can observe that there is not an appreciable difference in the obtained air interface performance.

For the case of 21 RUs, we reach similar conclusions, as for the case of 3 RUs: when there are 4 UEs/cell (i.e., the total traffic load of the cell increases), the proposed algorithms allow meeting the fronthaul-air interface trade-off effectively, as can be observed in Fig. 3(b), by comparing the case of QPSK versus the cases of semi-static modulation compression. Remarkably, in the case of 21 RUs, we observe a non-linear effect in the delay performance when increasing the maximum allowed modulation order. This is because the fast fading and dynamic interference conditions generate a higher percentage of losses at the PHY layer in case that higher MCSs are used. Higher MCSs (and so, higher modulation orders) are more sensitive to errors. Instead, if lower MCSs are used,
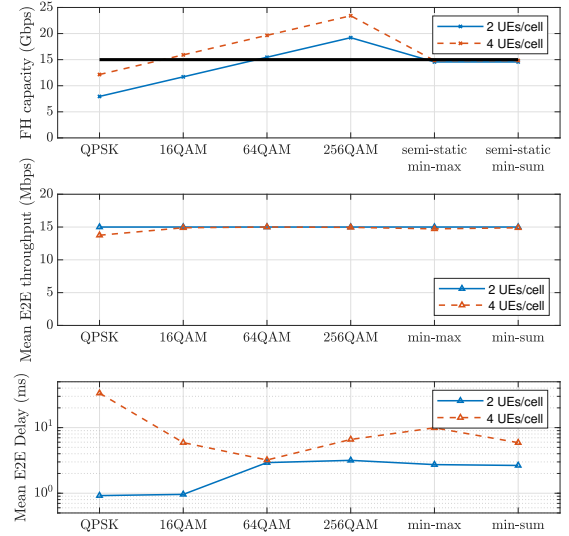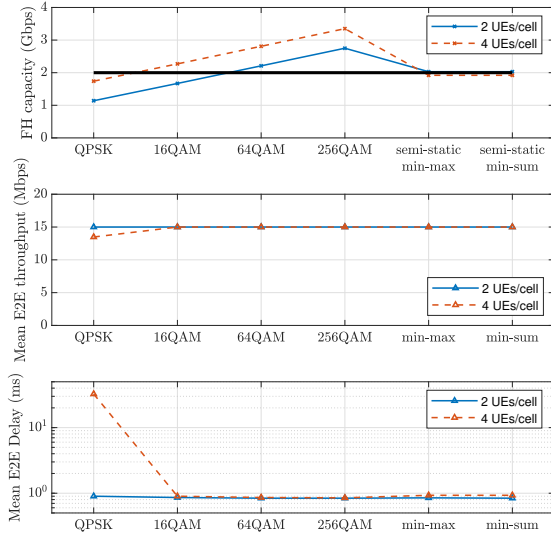
Fig. 3: Fronthaul capacity requirement vs. air interface performance (E2E throughput and delay) against different statically uniform (labelled QPSK, 16QAM, 64QAM and 256QAM) and semi-statically optimized modulation compressions (abscissa) and number of UEs/cell (2 and 4). (a) 3 RUs. (b) 21 RUs.

the system is more robust to channel impairments, allowing to receive without the need for retransmission combining at MAC and without reordering at RLC. This justifies the higher delay with 256QAM. Also, and as a consequence of such an effect, we observe some small difference in the delay of the two proposed solutions: "min-sum NL" leads to a smaller delay. This is because "min-sum NL" provides more equally distributed maximum allowed modulation orders (see Fig. 2) and so it is more robust than "min-max NL" to the varying interference and fast fading conditions.

V. CONCLUSIONS

In this paper, we have proposed two semi-static optimization solutions to configure the maximum modulation order that is permitted per cell, in a C-RAN architecture with multiple cells sharing a fronthaul link, assuming that modulation compression is used to reduce the required fronthaul capacity. The proposed solutions allow controlling the fronthaul resources' utilization without exceeding its maximum capacity while optimizing the air interface performance based on each cell's average traffic load and air interface configuration. We have evaluated the air interface performance and the related fronthaul capacity requirement of each solution in an end-to-end, dynamic, multi-cell simulation built on the ns-3 5G NR module. Simulation results in an Urban Micro scenario with multiple cells and various users per cell have shown that the proposed solutions can meet the available fronthaul capacity without appreciable air interface degradation.

VI. ACKNOWLEDGMENTS

REFERENCES

[1] S. Parkvall *et al.*, "NR: the new 5G radio access technology," *IEEE Commun. Standards Mag.*, vol. 1, pp. 24–30, Dec. 2017.

[2] M. Peng *et al.*, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 2282–2308, 2016.

[3] A. de la Oliva *et al.*, "An overview of the CPRI specification and its application to C-RAN-based LTE scenarios," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 152–159, 2016.

[4] 3GPP TS 38.401, *TSG RAN; NG-RAN; Architecture description*, Release 16, v16.1.0, Mar. 2020.

[5] 3GPP TR 38.801, *TSG RAN; Study on new radio access technology: Radio access architecture and interfaces*, Release 14, v14.0.0, Mar. 2017.

[6] 3GPP TR 38.816, *TSG RAN; Study on CU-DU lower layer split for NR*, Release 15, v15.0.0, Dec. 2017.

[7] O-RAN Alliance White Paper, *O-RAN use cases and deployment scenarios*, Feb. 2020.

[8] O-RAN Fronthaul Working Group, *Technical Specification; Control, User and Synchronization Plane Specification*, v03.00, Apr. 2020.

[9] L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A survey of the functional splits proposed for 5G mobile crosshaul networks," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 146–172, 2019.

[10] A. Umesh, T. Yahima, T. Uchino, and S. Okuyama, "Overview of O-RAN Fronthaul Specifications," *Special Articles on Standardization Trends towards Open and Intelligent Radio Access Networks, NTT Docomo technical journal*, vol. 21, July 2019.

[11] N. Patriciello, S. Lagen, B. Bojovic, and L. Giupponi, "An E2E simulator for 5G NR networks," *Elsevier Simulation Modelling Practice and Theory*, vol. 96, p. 101933, 2019.

[12] S. Lagen *et al.*, "Long-term provisioning of radio resources based on their utilization in dense OFDMA networks," *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, Sep. 2016.

[13] S. Lagen *et al.*, "Subband configuration optimization for multiplexing of numerologies in 5G TDD New Radio," *IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun.*, Sep. 2018.

[14] T. Zugno *et al.*, "Implementation of a spatial channel model for ns-3," *ns-3 Workshop 2020*, June 2020.

[15] S. Lagen *et al.*, "New Radio Physical Layer Abstraction for System-Level Simulations of 5G Networks," in *2018 IEEE Int. Conf. on Commun.*, June 2020.

[16] 3GPP TS 38.214, *TSG RAN; NR; Physical layer procedures for data (Release 16), v16.2.0*, June 2020.