IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 30, 2017, accepted July 16, 2017, date of publication July 20, 2017, date of current version August 8, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2729599

Downlink Scheduling and Resource Allocation for 5G MIMO-Multicarrier: OFDM vs FBMC/OQAM

GUILLEM FEMENIAS[1], (Senior Member, IEEE), FELIP RIERA-PALOU[1], (Senior Member, IEEE), XAVIER MESTRE[2], (Senior Member, IEEE), AND JUAN J. OLMOS[3], (Member, IEEE)

[1]Mobile Communications Group, University of the Balearic Islands, 07122 Illes Balears, Spain
[2]Centre Tecnològic de Telecomunicacions de Catalunya, 08860 Barcelona, Spain
[3]Wireless Communication Technologies Group, Polytechnic University of Catalonia, 08860 Barcelona, Spain

Corresponding author: Guillem Femenias (guillem.femenias@uib.es)

This work was supported in part by the Agencia Estatal de Investigación and Fondo Europeo de Desarrollo Regional through the projects ELISA and DEFINE-5G under Grant TEC2014-59255-C3-1-R, Grant TEC2014-59255-C3-2-R, and Grant TEC2014-60258-C2-2-R.

ABSTRACT The definition of the next generation of wireless communications, so-called 5G networks, is currently underway. Among many technical decisions, one that is particularly fundamental is the choice of the physical layer modulation format and waveform, an issue for which several alternatives have been proposed. Two of the most promising candidates are: 1) orthogonal frequency division multiple (OFDM), a conservative proposal that builds upon the huge legacy of 4G networks and 2) filter bank multicarrier/offset quadrature amplitude modulation (FBMC/OQAM), a progressive approach that in frequency selective channels sacrifices subcarrier orthogonality in lieu of an increased spectral efficiency. The comparative merits of OFDM and FBMC/OQAM have been well investigated over the last few years but mostly, from a purely physical layer point of view and largely neglecting how the physical layer performance translates into user-relevant metrics at the upper-layers. This paper aims at presenting a comprehensive comparison of both modulation formats in terms of practical network indicators, such as goodput, delay, fairness, and service coverage, and under operational conditions that can be envisaged to be realistic in 5G deployments. To this end, a unifying cross-layer framework is proposed that encompasses the downlink scheduling and resource allocation procedures and that builds upon a model of the queueing process at the data-link control layer and a physical layer abstraction that can be chosen to model either OFDM or FBMC/OQAM. Extensive numerical results conclusively demonstrate that most of the *a priori* advantages of FBMC/OQAM over OFDM do indeed translate into improved network indicators, that is, the increase in spectral efficiency achieved by FBMC/OQAM makes up for the distortion caused by the loss of orthogonality.

INDEX TERMS 5G, OFDM, FBMC/OQAM, MIMO, scheduling, resource allocation.

I. INTRODUCTION

A. MOTIVATION AND PREVIOUS WORK

Standardization bodies, industry and academia are still trying to devise and define what the fifth generation (5G) of wireless and mobile networks will be [1] and [2]. What can be ensured, however, is that 5G networks will have more stringent QoS requirements than fourth generation (4G) ones and will feature innovative strategies at the physical (PHY) and data link control (DLC) layers that will be impacted by the adopted modulation and multiple access strategies. Orthogonal frequency division multiplexing (OFDM) and orthogonal frequency division multiple access (OFDMA) have been the modulation format and multiple access strategy of choice in the downlink of LTE/LTE-Advanced cellular network standards because of a number of advantages they offer [3]. These advantages include, amongst others, the orthogonality between subcarriers, which allow (i) an easy implementation of both the transmitter and the receiver thanks to the use of fast Fourier transform (FFT) and inverse FFT (IFFT) blocks, (ii) the simplification of the channel equalization process through the use of a single-tap equalizer per subcarrier, and (iii) the trivial integration with MIMO techniques [4], [5]. Such a pool of advantages, however, come at the cost of some weaknesses. Orthogonality between subcarriers, which allows controlling the inter-carrier interference (ICI), can only be obtained by ensuring time and frequency synchronization at the receiver and using per-subcarrier rectangular shaping filters. Synchronization at the receiver side can only be easily implemented in the downlink of OFDMA-based cellular systems, thus limiting the range of

scenarios in which OFDM-based systems can be applied. Furthermore, characterized by very large sidelobes, rectangular shaping filters exhibit poor stopband attenuation and thus require the allocation of null guard band subcarriers at the spectrum edges in order to control out-of-band (OOB) interference. This effect, jointly with the fact that OFDM-based systems use cyclic prefixing (CP) to control inter-symbol interference (ISI), limits the maximum spectral efficiency that PHY layers based on OFDM can offer.

The aforementioned drawbacks exhibited by OFDM/OFDMA-based systems open the door to a debate on which are the most appropriate modulation format and multiple access strategy to be implemented in 5G networks [2], [6], [7]. One of the alternatives to OFDM is the well-known filter bank multicarrier (FBMC) format based on the offset quadrature amplitude modulation (OQAM), referred to as FBMC/OQAM [8]. This multicarrier modulation format does not transmit a CP and shapes subcarriers using well frequency-localized waveforms that suppress signals' sidelobes, thus providing larger spectral efficiencies than OFDM. Furthermore, the transmitter and receiver may still be implemented through FFT/IFFT blocks or polyphase filter structures, and the receiver, thanks to signal bandlimitedness, is able to cope with time and frequency asynchronism and is robust against frequency misalignments among users [7]. The price paid for these advantages is a weak orthogonality among subcarriers that is only satisfied in the real domain and that, in most practical implementations, translates into an ISI/ICI term at the output of the receive filters, specially relevant under strong channel frequency selectivity [9]–[12].

Although quite a few research works have been conducted with the aim of comparing the performance of OFDM and FBMC modulation formats at the PHY layer level (see, for instance, [6], [13]–[16] and references therein), much less investigations have been devoted to compare the performance attained by these multicarrier waveforms when considering the cross-layer optimization of the scheduling and resource allocation (SRA) algorithms at both the PHY and DLC layers. In fact, how the spectral efficiency advantage of FBMC/OQAM-based PHY layers can be exploited by the cross-layer SRA algorithms and the effects the additional ISI/ICI term really has on these systems when compared to OFDM-based ones still remains to be shown, particularly when transmitting on channels showing strong frequency selectivity.

One of the few works comparing the performance of SRA strategies in the downlink of OFDM- and FBMC-based multicarrier systems was presented by Bai *et al.* [17]. In this work the authors proposed a cross-layer approach where the scheduler established packet priorities according to the quality-of-service (QoS) requirements of heterogeneous service flows, and the resource allocation algorithms were in charge of allocating subchannels, data rates and transmit powers given the prioritized packet lists. The main conclusion of this work was that thanks to the better frequency containment of FBMC systems, they are able to fulfill the QoS requirements of a larger quantity of users than OFDM-based schemes. The authors of this paper, however, sought the minimization of the transmitted power on single-input single-output (SISO) channels that were considered to be frequency-flat over each subcarrier, thus neglecting the ISI/ICI term caused by strong channel frequency selectivity. Payaró *et al.* [18] addressed the resource allocation problem in the uplink multiple access channel of cellular networks using either MIMO-OFDM or MIMO-FBMC/OQAM. Again, the better spectral properties exhibited by the FBMC/OQAM waveform were explicitly exploited in this paper to reduce the frequency guard bands between groups of subcarriers allocated to different users and to improve the performance provided by rate adaptive SRA algorithms given power and minimum rate constraints. Unfortunately, the conclusions that were drawn from the analysis of an asynchronous uplink multiple access channel cannot be directly translated to a synchronous downlink multicast scenario. Shaat and Bader [19] addressed a similar resource allocation problem as the one posed in [17] but in the context of multicarrier-based cognitive radio networks. The objective in this work was the maximization of the downlink network capacity with constraints on both the available transmit power and the amount of interference the secondary users were allowed to generate on the primary users. In contrast, based on a modification of the SRA algorithm proposed in [19], Caus *et al.* presented a margin adaptive SRA framework aiming at the minimization of the transmit power in the downlink of a FBMC/OQAM-based cellular network subject to minimum rate constraints [9]. However, similar to the approaches followed by most previously mentioned works, the proposed SRA framework was based on the use of unrealistic per-subcarrier Shannon capacity formulas to model the transmission data rates allocated to each of the active users in the system.

B. CONTRIBUTIONS OF THE PAPER

In order to shed some more light on this problem, our aim in this paper is to present a cross-layer SRA framework allowing a fair and realistic comparison between downlink cellular systems based on either OFDM or FBMC/OQAM. Our main contributions can be summarized as follows:

- Based on an equivalent frequency domain approach, we present analytical closed-form expressions characterizing the PHY layer of FBMC/OQAM- and OFDM-based MIMO architectures in terms of the signal-to-interference-plus-noise ratio (SINR) experienced at the output of the detector when considering the transmission on channels showing strong frequency selectivity. The analytical formulation is general enough to encompass, depending on the amount of channel state information available at the transmitter (CSIT) and/or the receiver (CSIR) sides, different MIMO techniques that serve to exploit the spatial diversity/multiplexing capabilities provided by the MIMO channel.

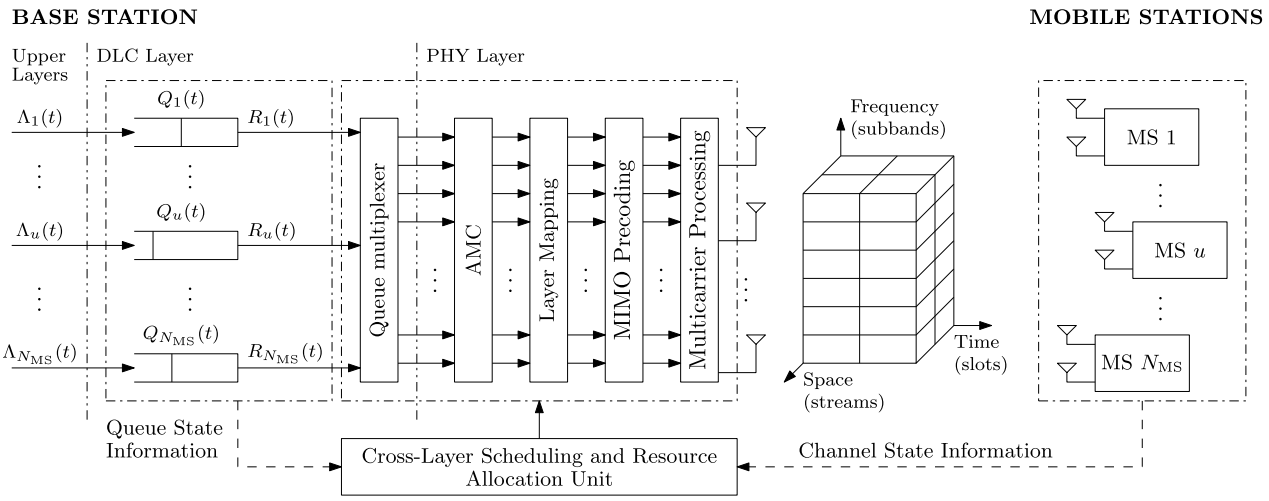

FIGURE 1. Downlink time-slotted MIMO-multicarrier architecture.

- Based on the previous PHY layer characterization of both OFDM- and FBMC/OQAM-based systems, a cross-layer SRA optimization problem is proposed whose aim is to determine, for each scheduling interval, which is the optimal allocation of resources to users in order to maximize prescribed utility functions with constraints enforcing the exclusive allocation of subbands, a transmission capacity adjusted to the contents of data queues at the DLC and a guaranteed minimum quality of transmission. To tackle this combinatorial optimization problem, infeasible in practice for most 4G/5G network configurations, the optimal resource allocation problem is decoupled in two suboptimal problems whose global complexity increases linearly with the number of active users in the system and the number of resources to allocate.
- A comprehensive performance evaluation of the proposed cross-layer SRA algorithms is conducted using the abstraction of an LTE-like link level model. Obtained results illustrate how the proposed SRA procedures are affected by and cope with the particularities of OFDM- and FBMC/OQAM-based systems. The impacts of the number of served users, traffic classes, traffic arrival rates, scheduling rules, MIMO configuration, LTE channel models and spatial correlation profiles on the proposed SRA algorithms are considered.

C. PAPER ORGANIZATION AND NOTATIONAL REMARKS

This paper is organized as follows. The system model is presented in Section II where analytical expressions used to characterize both the queue behavior at the DLC layer and the instantaneous SINR at the output of the MIMO detector at the PHY layer are derived. Section III presents the cross-layer SRA framework proposed in this paper. Comprehensive numerical results are provided in Section IV. Finally, the main outcomes of this paper are recapped in Section V.

This introduction ends with some notational remarks. Vectors and matrices are denoted by lower-case and upper-case boldface symbols. The superscripts $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ denote the transpose, the conjugate and the conjugate transpose (also known as the Hermitian), respectively. The M-dimensional identity matrix is represented by $\boldsymbol{I}_M$. The symbols $\mathbb{R}$ and $\mathbb{C}$ serve to denote the sets of real and complex numbers, respectively. The operator $\mathrm{Tr}\{\cdot\}$ denotes the trace of a matrix and the operator $\mathrm{diag}(\boldsymbol{x})$ is used to denote a rectangular diagonal matrix containing the vector $\boldsymbol{x}$ on its main diagonal.

II. SYSTEM MODEL AND ASSUMPTIONS

Let us consider the downlink[1] of a time-slotted MIMO-multicarrier wireless access system, such as the one schematically depicted in Fig. 1, where a base station (BS) transmitting with a total power P_T and equipped with N_T transmit antennas is serving N_{MS} mobile stations (MSs). In order to simplify the mathematical notation of the problem, all the MSs are assumed to be equipped with the same number N_R of receive antennas and only one service data flow (also known as a connection or session) per active MS is considered. Depending on the traffic type, three classes of service and the associated QoS requirements will be accounted for: *Best Effort* (BE), *Non-Real-Time* (nRT), and *Real-Time* (RT).

At the DLC layer of the BS, traffic flows arriving from higher layers are buffered into the corresponding N_{MS} first-in first-out (FIFO) queues. At the beginning of each scheduling time interval, based on the available joint channel state

[1] An analogous approach could be applied to the uplink by considering that, on one hand, OFDM is much more sensitive than FBMC/OQAM to the loss of orthogonality caused by imperfect synchronization and, on the other hand, although both multicarrier-based PHY layers need to place a guard frequency band between the RBs allocated to different users in order to control the interference among users, this required guard band turns out to be broader in OFDM than in FBMC/OQAM [18].

and queue state information (CSI and QSI), the cross-layer SRA unit selects which users are to be served and part of (or all) information bits in the corresponding queues are then forwarded to the PHY layer for transmission. At the PHY layer, the data bits of a selected user go through an interleaver to reduce burst errors and are adaptively modulated and channel encoded using one of the available modulation and coding schemes (MCSs). Then, depending on the CSI delivered by the MS, the BS selects the number of spatial layers (or streams) over which the coded symbols will be transmitted using MIMO precoding. The cross-layer SRA unit is also in charge of allocating resources (e.g., power, frequency subbands and/or time-slots) to each of the selected users in such a way that the precoded symbol streams, jointly with added signaling information, can be mapped onto the time-frequency plane by using appropriate multicarrier processing strategies.

The time-frequency plane characterizing the downlink of a time-slotted MIMO-multicarrier system is organized in resource allocation units, also known as resource blocks (RB). Each RB is formed by a slot in the time axis and a subband in the frequency axis:

- In the time axis, each RB occupies a time slot of duration $T_{\text{slot}}^{\text{PHY}}$, where the tag PHY is used as a token to represent either OFDM or FBMC/OQAM depending on the PHY layer in use. Each slot consists of a fixed number N_s of OFDM/FBMC symbols (N_i information symbols plus $N_s - N_i$ signaling symbols) whose payload part has a duration T_p, irrespective of the PHY layer being used. The subcarrier bandwidth, for both OFDM- and FBMC/OQAM-based systems, is thus given by $\Delta f = 1/T_p$. The OFDM-based system, in contrast to what happens with the FBMC/OQAM-based one, uses cyclic prefixing to eliminate ISI. In the time-slot of an OFDM-based system there are, typically, $N_{s\text{long}}$ OFDM symbols that are prefixed using a long cyclic prefix of duration $T_{CP\text{long}}$ and the remaining $N_{s\text{short}} = N_s - N_{s\text{long}}$ ones are prefixed using a short cyclic prefix of duration $T_{CP\text{short}}$. Thus,

$$T_{\text{slot}}^{\text{FBMC}} = N_s T_p, \tag{1a}$$

$$T_{\text{slot}}^{\text{OFDM}} = N_s T_p + N_{s\text{long}} T_{CP\text{long}} + N_{s\text{short}} T_{CP\text{short}}. \tag{1b}$$

- Slotted transmissions employ a bandwidth B that is split into N_f orthogonal subcarriers, out of which, on average, N_d^{PHY} are used to transmit data and N_p^{PHY} are used to transmit pilots, synchronization and control channels and to set guard frequency bands. The N_d^{PHY} data subcarriers are grouped into N_B^{PHY} orthogonal subbands, which are indexed by the set $\mathcal{N}_B = \{1, \ldots, N_B^{\text{PHY}}\}$, each consisting of N_{sc} adjacent subcarriers and with a bandwidth $B_B = N_{sc}\Delta f$. Due to the virtual absence of guard bands in the frequency domain of FBMC/OQAM-based systems, then $N_p^{\text{FBMC}} < N_p^{\text{OFDM}}$ and thus, $N_B^{\text{FBMC}} > N_B^{\text{OFDM}}$.

Although the basic resource unit is an RB (one subband and one time slot), and following the LTE/LTE-Advanced SRA procedure, it will be assumed in this work that the SRA process takes place over a transmission time interval (TTI) of two consecutive time slots (RB pair).

A. MODELING QUEUE BEHAVIOR AT THE DLC LAYER

Let us assume that at the beginning of TTI t the BS has $Q_u(t)$ bits in the queue of MS u. Let us also assume that, according to CSI/QSI, the SRA unit allocates L_u spatial streams to user u and a transmission capacity of $r_{u,l}(t, \mathcal{N}_{Bu})$ bits to spatial stream $l \in \mathcal{L}_u = \{1, \ldots, L_u\}$, with $\mathcal{N}_{Bu}$ denoting the set of RBs allocated to user u. In this case, assuming queues of infinite capacity,[2] the queue length at the end of this TTI can be expressed as

$$Q_u(t+1) = Q_u(t) + \Lambda_u(t) - R_u(t), \tag{2}$$

where $\Lambda_u(t)$ is used to denote the number of information bits arriving during TTI t and $R_u(t)$ represents the number of bits successfully transmitted to user u, which can be calculated as

$$R_u(t) = \sum_{l=1}^{L_u} \vartheta_{u,l}(t) r_{u,l}(t, \mathcal{N}_{Bu}), \tag{3}$$

with $\vartheta_{u,l}(t) = 1$ in case of a successful transmission of stream l and $\vartheta_{u,l}(t) = 0$, otherwise. In order to avoid serving empty queues (frugality constraint), the SRA unit should control the service rate in such a way that the transmission capacity allocated to user u does not exceed the number of bits buffered in the corresponding queue, that is, $\sum_{l=1}^{L_u} r_{u,l}(t, \mathcal{N}_{Bu}) \leq Q_u(t)$.

B. TRANSMITTER/RECEIVER

Our aim in this subsection is to present a formulation for a MIMO multicarrier modulation scheme that provides a description of the PHY layer of FBMC/OQAM- and OFDM-based architectures in terms of the SINR experienced at the output of the detector. This approach will be based on the equivalent frequency domain matrix processing sketched in the block diagrams of Fig. 2. MIMO technology comprises a great variety of techniques that can be used to exploit the propagation paths between the N_T transmit antennas at the BS and the N_R receive antennas at its served MS u (see [21] for a review). Let us assume that the BS simultaneously transmits $L_u \leq \min(N_T, N_R)$ data streams to its served MS u using MIMO linear precoding.[3] In this case, assuming that subband b has been allocated to MS u during TTI t, we can use $\boldsymbol{s}_{u,b,s}(t,q) \in \mathbb{C}^{L_u \times 1}$ to denote

[2] At the cost of increasing the notational complexity, the proposed framework could be readily extended to systems with finite capacity queues [20, Sec. IV.A].

[3] We note that the linear precoding framework is able to encompass both the presence or absence of CSI at the transmitter side. In the latter case, the precoding matrix will be constant and independent of the channel (typically, an identity matrix). Furthermore, in this work only single-user spatial multiplexing techniques will be considered and the extension to multi-user MIMO is left for future work.

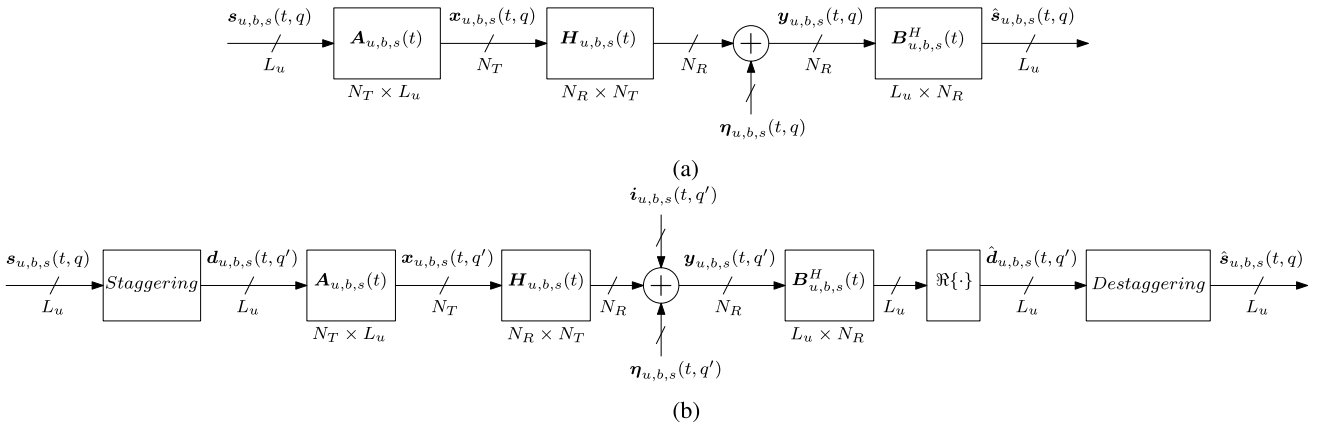

FIGURE 2. Per-subcarrier matrix processing interpretation of the multicarrier MIMO communication system. (a) OFDM. (b) FBMC/OQAM.

the vector of L_u complex-valued symbols transmitted during multicarrier symbol q on the corresponding subcarrier s (we assume zero-mean unit-energy uncorrelated symbols, i.e., $\mathbb{E}\left\{\boldsymbol{s}_{u,b,s}(t,q)\boldsymbol{s}^H_{u,b,s}(t,q)\right\} = \boldsymbol{I}_{L_u}$). The signal processing applied to this vector of transmitted symbols differs for OFDM and FBMC/OQAM schemes.

1) OFDM

In an OFDM system, as it can be observed in Fig. 2a, the vector $\boldsymbol{s}_{u,b,s}(t,q)$ is first linearly precoded to generate the complex-valued $N_T \times 1$ vector

$$\boldsymbol{x}_{u,b,s}(t,q) = \boldsymbol{A}_{u,b,s}(t)\boldsymbol{s}_{u,b,s}(t,q), \tag{4}$$

where $\boldsymbol{A}_{u,b,s}(t) = \left[\boldsymbol{a}_{u,b,s,1}(t) \ldots \boldsymbol{a}_{u,b,s,L_u}(t)\right] \in \mathbb{C}^{N_T \times L_u}$ is the transmit precoding matrix. Although the framework described in this paper could be extended to adaptive power allocation (APA) strategies (see, for instance, [22]–[24]), for the sake of affordable complexity, a uniform power allocation (UPA) scheme will be adopted in which the transmit precoding matrix must satisfy the constraint $\mathrm{Tr}\left(\boldsymbol{A}^H_{u,b,s}(t)\boldsymbol{A}_{u,b,s}(t)\right) = P_s$, with P_s denoting the power allocated per data subcarrier. Notice that, as stated in [22], even though adaptive power allocation schemes improve the performance of UPA-based ones, using AMC schemes with a large set of modulation and coding formats can make completely superfluous the application of power allocation strategies.

Precoded symbols in the frequency domain are then converted to the time domain by using an IFFT and furthermore, a cyclic prefix is added to each OFDM symbol. After propagation through the wireless MIMO channel, the cyclic prefix part is removed from the time-domain received signal and an FFT is used to obtain the frequency domain samples. Very common simplifying assumptions used in the analysis of OFDM-based systems are the consideration of ideal synchronization and sampling processes at the receiver side and the use of cyclic prefixes of duration greater than the maximum delay spread of the channel impulse response. Under these assumptions, the system is not affected neither by ISI nor by ICI and hence, for the subcarrier s in subband b, the received samples at the output of the N_R processing stages of MS u can be expressed by the $N_R \times 1$ complex-valued vector (see Fig. 2a)

$$\boldsymbol{y}_{u,b,s}(t,q) = \boldsymbol{H}_{u,b,s}(t)\boldsymbol{x}_{u,b,s}(t,q) + \boldsymbol{\eta}_{u,b,s}(t,q), \tag{5}$$

where the noise vector $\boldsymbol{\eta}_{u,b,s}(t,q) \sim \mathcal{CN}_{N_R,1}\left(\boldsymbol{0}, \sigma^2_\eta \boldsymbol{I}_{N_R}\right)$ and $\boldsymbol{H}_{u,b,s}(t) \in \mathbb{C}^{N_R \times N_T}$ is the MIMO channel matrix characterizing the propagation loss, large-scale shadow fading and small-scale time/frequency/space selective fading between the BS and MS u.

Finally, the equalization matrix $\boldsymbol{B}^H_{u,b,s}(t) = [\boldsymbol{b}_{u,b,s,1}(t) \ldots \boldsymbol{b}_{u,b,s,L_u}(t)]^H \in \mathbb{C}^{L_u \times N_R}$ is used to process the received vector $\boldsymbol{y}_{u,b,s}(t,q)$ and produce the $L_u \times 1$ complex-valued vector

$$\begin{aligned}\hat{\boldsymbol{s}}_{u,b,s}(t,q) &= \boldsymbol{B}^H_{u,b,s}(t)\boldsymbol{y}_{u,b,s}(t,q) \\ &= \mathrm{diag}\left(\boldsymbol{\alpha}_{u,b,s}(t)\right)\boldsymbol{s}_{u,b,s}(t) \\ &\quad + \boldsymbol{D}_{u,b,s}(t)\boldsymbol{s}_{u,b,s}(t) + \boldsymbol{B}^H_{u,b,s}(t)\boldsymbol{\eta}_{u,b,s}(t),\end{aligned} \tag{6}$$

where we can distinguish between the useful term, the multistream interference (MSI) term and the AWGN term. The variables $\boldsymbol{\alpha}_{u,b,s}(t)$ and $\boldsymbol{D}_{u,b,s}(t)$ have been defined, respectively, as

$$\boldsymbol{\alpha}_{u,b,s}(t) = \left[\alpha_{u,b,s,1}(t) \ldots \alpha_{u,b,s,L_u}(t)\right]^T, \tag{7}$$

with

$$\alpha_{u,b,s,l}(t) = \boldsymbol{b}^H_{u,b,s,l}(t)\boldsymbol{H}_{u,b,s}(t)\boldsymbol{a}_{u,b,s,l}(t), \tag{8}$$

and

$$\boldsymbol{D}_{u,b,s}(t) = \boldsymbol{B}^H_{u,b,s}(t)\boldsymbol{H}_{u,b,s}(t)\boldsymbol{A}_{u,b,s}(t) - \mathrm{diag}\left(\boldsymbol{\alpha}_{u,b,s}(t)\right). \tag{9}$$

Note that in a fully diagonalized MIMO system the MSI term vanishes [21].

Since we assume a block fading channel, all symbols of data stream $l \in \mathcal{L}_u = \{1, \ldots, L_u\}$ transmitted on subcarrier s

of subband b and during TTI t are characterized by the same SINR $\gamma_{u,b,s,l}(t)$ that can be expressed as

$$\gamma_{u,b,s,l}(t) = \frac{|\alpha_{u,b,s,l}(t)|^2}{\left[\boldsymbol{R}_{u,b,s}^{\mathrm{OFDM}}(t)\right]_{l,l} + \sigma_\eta^2 \boldsymbol{b}_{u,b,s,l}^H(t)\boldsymbol{b}_{u,b,s,l}(t)}, \quad (10)$$

where the covariance matrix of the MSI term can be obtained as

$$\boldsymbol{R}_{u,b,s}^{\mathrm{OFDM}}(t) = \boldsymbol{D}_{u,b,s}(t)\boldsymbol{D}_{u,b,s}^H(t). \quad (11)$$

2) FBMC/OQAM

In a FBMC/OQAM system (see Fig. 2b), a time-offset (staggering) equal to half the OFDM symbol period on the real or imaginary part (depending on the subcarrier parity) of the complex-valued information symbols $\boldsymbol{s}_{u,b,s}(t,q)$ is first introduced. This staggering process produces two $L_u \times 1$ real-valued vectors $\boldsymbol{d}_{u,b,s}(t,q')$, with $q' = 2q$ and $q' = 2q+1$, that are eventually transmitted using OQAM. Each of the real-valued vectors $\boldsymbol{d}_{u,b,s}(t,q')$ is linearly precoded to generate the complex-valued $N_T \times 1$ transmitted vector

$$\boldsymbol{x}_{u,b,s}(t,q') = \boldsymbol{A}_{u,b,s}(t)\boldsymbol{d}_{u,b,s}(t,q'), \quad (12)$$

where $\boldsymbol{A}_{u,b,s}(t) \in \mathbb{C}^{N_T \times L_u}$ is a transmit precoding matrix that, as in the OFDM case, is assumed to satisfy the UPA constraint.

Precoded symbols in the frequency domain are then processed using time-frequency shifted versions of a prototype filter (synthesis filter bank) with better frequency localization properties than those used in OFDM-based systems. After propagation through the wireless MIMO channel, the time-domain received samples are processed again using time-frequency shifted versions of a prototype filter (analysis filter bank) to obtain the frequency-domain received signal vector $\boldsymbol{y}_{u,b,s}(t,q')$. In contrast to what happens in OFDM systems, channel frequency selectivity causes loss of system orthogonality and the consequent appearance of ISI and ICI, even in systems with ideal synchronization and sampling processes. The received samples at the output of the N_R processing stages of MS u can be expressed by the $N_R \times 1$ complex-valued vector (see Fig. 2b)

$$\boldsymbol{y}_{u,b,s}(t,q') = \boldsymbol{H}_{u,b,s}(t)\boldsymbol{x}_{u,b,s}(t,q') + \boldsymbol{i}_{u,b,s}(t,q') + \boldsymbol{\eta}_{u,b,s}(t,q'), \quad (13)$$

where $\boldsymbol{i}_{u,b,s}(t,q')$ denotes the ISI/ICI term and again, the noise vector $\boldsymbol{\eta}_{u,b,s}(t,q) \sim \mathcal{CN}_{N_R,1}\left(\boldsymbol{0}, \sigma_\eta^2 \boldsymbol{I}_{N_R}\right)$.

The complex-valued vector $\boldsymbol{y}_{u,b,s}(t,q')$ is processed by the equalization matrix $\boldsymbol{B}_{u,b,s}^H(t) \in \mathbb{C}^{L_u \times N_R}$ and only the real part of the equalized samples is kept producing the $L_u \times 1$ real-valued vector

$$\hat{\boldsymbol{d}}_{u,b,s}(t,q') = \Re\left\{\boldsymbol{B}_{u,b,s}^H(t)\boldsymbol{y}_{u,b,s}(t,q')\right\}. \quad (14)$$

After de-staggering, the estimation of the complex-valued symbol vector transmitted to MS u on subcarrier s of subband b and during the qth multicarrier symbol period of TTI t can thus be expressed as

$$\begin{aligned}\hat{\boldsymbol{s}}_{u,b,s}(t,q) = {} & \Re\left\{\mathrm{diag}\left(\boldsymbol{\alpha}_{u,b,s}(t)\right)\right\}\boldsymbol{s}_{u,b,s}(t,q) \\ & + \Re\left\{\boldsymbol{D}_{u,b,s}(t) - \mathrm{diag}\left(\boldsymbol{\alpha}_{u,b,s}(t)\right)\right\}\boldsymbol{s}_{u,b,s}(t,q) \\ & + \hat{\boldsymbol{i}}_{u,b,s}(t,q) + \hat{\boldsymbol{\eta}}_{u,b,s}(t,q), \end{aligned} \quad (15)$$

where the filtered ISI/ICI term is given by

$$\begin{aligned}\hat{\boldsymbol{i}}_{u,b,s}(t,q) = {} & \Re\left\{\boldsymbol{B}_{u,b,s}^H(t)\boldsymbol{i}_{u,b,s}(t,2q)\right\} \\ & + j\Re\left\{\boldsymbol{B}_{u,b,s}^H(t)\boldsymbol{i}_{u,b,s}(t,2q+1)\right\}, \end{aligned} \quad (16)$$

and the filtered noise term can be expressed as

$$\begin{aligned}\hat{\boldsymbol{\eta}}_{u,b,s}(t,q) = {} & \Re\left\{\boldsymbol{B}_{u,b,s}^H(t)\boldsymbol{\eta}_{u,b,s}(t,2q)\right\} \\ & + j\Re\left\{\boldsymbol{B}_{u,b,s}^H(t)\boldsymbol{\eta}_{u,b,s}(t,2q+1)\right\}. \end{aligned} \quad (17)$$

Hence, the SINR $\gamma_{u,b,s,l}(t)$ experienced by any of the symbols of data stream $l \in \mathcal{L}_u = \{1,\ldots,L_u\}$ transmitted on subcarrier s of subband b and during TTI t can be obtained as

$$\gamma_{u,b,s,l}(t) = \frac{\left(\Re\left\{\alpha_{u,b,s,l}(t)\right\}\right)^2}{\left[\boldsymbol{R}_{u,b,s}^{\mathrm{FBMC}}(t)\right]_{l,l} + \sigma_\eta^2 \boldsymbol{b}_{u,b,s,l}^H(t)\boldsymbol{b}_{u,b,s,l}(t)}, \quad (18)$$

where, for large values of N_f and under the same assumptions that were used in [12, Sec. II.B], the covariance matrix of the MSI plus ISI/ICI terms, can be approximated as (see the proof of [12, Th. II.1])

$$\begin{aligned}\boldsymbol{R}_{u,b,s}^{\mathrm{FBMC}}(t) = {} & \boldsymbol{D}_{u,b,s}(t)\boldsymbol{D}_{u,b,s}^H(t) \\ & + \alpha_u^{(1)}\left(\boldsymbol{B}_{u,s,b}^H(t)\boldsymbol{H}'_{u,s,b}(t)\boldsymbol{A}_{u,s,b}(t)\right) \\ & \times \left(\boldsymbol{B}_{u,s,b}^H(t)\boldsymbol{H}'_{u,s,b}(t)\boldsymbol{A}_{u,s,b}(t)\right)^H \\ & + \alpha_u^{(1)}\Re\left\{\boldsymbol{D}_{u,b,s}(t)\right\} \\ & \times \Re^T\left\{\left(\boldsymbol{B}_{u,s,b}^H(t)\boldsymbol{H}_{u,s,b}(t)''\boldsymbol{A}_{u,s,b}(t)\right)^H\right\} \\ & + \alpha_u^{(1)}\Im\left\{\boldsymbol{D}_{u,b,s}(t)\right\} \\ & \times \Im^T\left\{\left(\boldsymbol{B}_{u,s,b}^H(t)\boldsymbol{H}_{u,s,b}(t)''\boldsymbol{A}_{u,s,b}(t)\right)^H\right\} \\ & + \alpha_u^{(2)}\Im\left\{\boldsymbol{D}_{u,b,s}(t)\right\} \\ & \times \Im^T\left\{\boldsymbol{B}_{u,s,b}^H(t)\left(\boldsymbol{H}_{u,s,b}(t)\boldsymbol{A}'_{u,s,b}(t)\right)'\right\} \\ & + \alpha_u^{(2)}\Im\left\{\boldsymbol{B}_{u,s,b}^H(t)\boldsymbol{H}_{u,s,b}(t)\boldsymbol{A}'_{u,s,b}(t)\right\} \\ & \times \Im^T\left\{\boldsymbol{B}_{u,s,b}^H(t)\left(\boldsymbol{H}_{u,s,b}(t)\boldsymbol{A}_{u,s,b}(t)\right)'\right\} \end{aligned} \quad (19)$$

using $\alpha_u^{(1)} = \frac{2\eta_{1010}^{(+,-)}}{N_d^2}$ and $\alpha_u^{(2)} = \frac{4\left(\eta_{1010}^{(+,-)} + \eta_{0011}^{(-,u)}\right)}{N_d^2}$, with $\eta_{1010}^{(+,-)}$ and $\eta_{0011}^{(+,-)}$ denoting pulse-related quantities defined in [12, Appendix A]. Furthermore, $\boldsymbol{X}'$ and $\boldsymbol{X}''$ refer to the first and second derivatives of matrix $\boldsymbol{X}$ with respect to the frequency.

III. CROSS-LAYER SCHEDULING AND RESOURCE ALLOCATION

Let us define the RB allocation set as $\boldsymbol{\mathcal{N}}_B = \{\mathcal{N}_{B1}, \ldots, \mathcal{N}_{BN_{\rm MS}}\}$ and the MCS allocation set as $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{N_{\rm MS}}\}$, with $\boldsymbol{\mu}_u = \{\mu_{u,1}, \ldots, \mu_{u,L_u}\}$ denoting the set of MCSs allocated to each of the spatial streams of user u. Using these definitions, the cross-layer SRA problem at hand can be formally posed as[4]

$$\begin{aligned}
&\max_{\boldsymbol{\mathcal{N}}_B, \boldsymbol{\mu}} \sum_{u=1}^{N_{\rm MS}} \sum_{l=1}^{L_u} w_u r_{u,l}\left(\mathcal{N}_{Bu}\right)\left[1 - \mathrm{BLER}_{u,l}^{(\mu_{u,l})}\left(\mathcal{N}_{Bu}\right)\right] \\
&\text{subject to } \mathcal{N}_{Bk} \cap \mathcal{N}_{Bj} = \emptyset \quad \forall k \neq j \\
&\qquad \sum_{l=1}^{L_u} r_{u,l}\left(\mathcal{N}_{Bu}\right) \leq Q_u \quad \forall\,(u,l) \\
&\qquad \mathrm{BLER}_{u,l}^{(\mu_{u,l})}\left(\mathcal{N}_{Bu}\right) \leq \mathrm{BLER}_0 \quad \forall (u,l) \qquad (20)
\end{aligned}$$

where w_u is the weighing coefficient of MS u used to implement different schedulers, $\mathrm{BLER}_{u,l}^{(\mu_{u,l})}\left(\mathcal{N}_{Bu}\right)$ is the block error rate (BLER) experienced by user u on spatial stream l when using MCS $\mu_{u,l}$ to transmit on the set of RBs $\mathcal{N}_{Bu}$ and BLER_0 is the maximum target BLER that can be supported by the system. The first constraint enforces a given RB to be allocated to a single user, the second one is the previously introduced frugality constraint and the third one is used to ensure that the average BLER experienced by a given user is always upper bounded by the target value BLER_0. In words, the optimization problem consists in determining, for each scheduling interval, which are the subsets of RBs and MCSs that have to be allocated to each MS in the system in order to maximize the weighted system goodput with constraints enforcing the exclusive allocation of RBs, frugality in the allocation of resources and a guaranteed minimum quality of transmission. Details on how to set the weighing coefficients to follow prescribed scheduling rules can be found in [22]. For completeness, weights for some popular schedulers, such as *Max-sum-rate* (MSR) [25], *Proportional fair* (PF) [26], *Modified largest weighted delay first* (MLWDF) [27], and *Exponential* (EXP) [28], are:

$$w_u = \begin{cases} 1 & \text{MSR} \\ 1/\overline{r}_u & \text{PF} \\ \phi_u W_{\mathrm{HOL},u}/\overline{r}_u & \text{MLWDF} \\ \dfrac{\phi_u}{\overline{r}_u} \exp\left(\dfrac{\phi_u W_{\mathrm{HOL},u} - \overline{\phi W}}{1 + \sqrt{\overline{\phi W}}}\right) & \text{EXP,} \end{cases} \qquad (21)$$

where $\overline{r}_u$ is the average effective data rate actually allocated to user u, which is calculated using a moving average over a relatively long sliding window [29], $W_{\mathrm{HOL},u}$ is the head-of-line delay[5] (HOL) of user u, ϕ_u is a positive constant that can be used to set different priority levels between traffic flows, and $\overline{\phi W} = \frac{1}{N_{\rm MS}} \sum_{u=1}^{N_{\rm MS}} \phi_u W_{\mathrm{HOL},u}$. In order to guarantee that users with absolute delay requirement $\check{D}_u$ and maximum outage delay probability requirement $\check{\xi}_u$ will be satisfied, the authors of [27] propose to *properly* set $\phi_u = -\log(\check{\xi}_u)/\check{D}_u$, $\forall\, u$, providing in this way QoS differentiation among flows.

Irrespective of the scheduling policy implemented at the DLC layer, the only way to solve problem (20) is to carry out an exhaustive search over all possible allocations of RBs and MCSs to users in order to find the constrained allocation maximizing the global weighted goodput. Such an exhaustive search happens to be infeasible for most system configurations that can be found in 4G/5G networks. Fortunately, the computational complexity of the optimization problem can be reduced by, on one hand, realizing that for a given allocation of RBs it is relatively straightforward to obtain the optimal allocation of MCSs and, on the other hand, resorting to a greedy suboptimal strategy to allocate the RBs to users. Both steps are described in detail in the following subsections.

A. MCS SELECTION. LINK ADAPTATION

Aiming at adapting the information data rate of user u to its instantaneous channel capacity, the link adaptation process at the BS, based on the CSI received on a feedback channel from the MS, selects the MCS to be used in the next TTI. In this work it will be assumed that the BS can choose from a set of $M = 15$ MCSs whose main characteristics, very similar to those of some of the MCSs used by LTE/LTE-Advanced systems, are listed in Table 1 (results presented in this table have been obtained from [30, Ch. 5]). Furthermore, as it is done in LTE/LTE-Advanced networks, an scheduled user u will be enforced to transmit the codeword of spatial stream l using the same MCS on all allocated RBs.

TABLE 1. Modulation scheme, effective code rate, information bits per modulated symbol and SINR boundaries for a $\mathrm{BLER}_0 = 0.1$ when using $N_{Bu} = 1$, 4 or 25 RBs.

MCS	Modulation	R_{cm}	bits/symbol	$\Gamma_m(1)$	$\Gamma_m(4)$	$\Gamma_m(25)$
0	–	–	–	$-\infty$	$-\infty$	$-\infty$
1	4QAM	0.076	0.152	-6.76	-6.93	-7.58
2	4QAM	0.117	0.234	-4.85	-5.21	-5.78
3	4QAM	0.189	0.396	-2.66	-3.31	-3.76
4	4QAM	0.301	0.602	-0.92	-1.45	-1.77
5	4QAM	0.439	0.878	1.00	0.48	0.17
6	4QAM	0.588	1.176	2.85	2.35	2.07
7	16QAM	0.369	1.476	4.69	4.23	3.99
8	16QAM	0.478	1.912	6.48	6.06	5.83
9	16QAM	0.602	2.408	8.43	8.03	7.91
10	64QAM	0.455	2.730	10.31	9.86	9.76
11	64QAM	0.554	3.324	12.13	11.75	11.64
12	64QAM	0.650	3.900	14.03	13.64	13.65
13	64QAM	0.754	4.524	16.01	15.54	15.58
14	64QAM	0.853	5.118	17.93	17.47	17.48
15	64QAM	0.926	5.556	19.87	19.39	19.46

Using the notation introduced in previous sections, let us assume that user u has been scheduled to transmit on the set of RBs $\mathcal{N}_{Bu} \subseteq \mathcal{N}_B$. For a given spatial stream l,

[4]Note that since optimization is performed on a TTI-by-TTI basis, from this point onwards the time dependence (i.e., (t)) of all the variables will be dropped.

[5]The head-of-line delay of user u is defined as the delay of the oldest bits in the traffic queue of this user.

each subcarrier may experience a different SINR $\gamma_{u,b,s,l}$ and hence, the corresponding transmitted codeword is affected by a *multistate* channel[6] characterized by the vector of SINRs

$$\boldsymbol{\gamma}_{u,l}\left(\mathcal{N}_{Bu}\right) \triangleq \left[\boldsymbol{\gamma}_{u,b,l}\right]_{\forall\, b\in\mathcal{N}_{Bu}}, \tag{22}$$

with

$$\boldsymbol{\gamma}_{u,b,l} = \left[\gamma_{u,b,1,l}\ldots\gamma_{u,b,N_{sc},l}\right]. \tag{23}$$

Based on the multistate CSI $\boldsymbol{\gamma}_{u,l}\left(\mathcal{N}_{Bu}\right)$, adaptive modulation and coding (AMC) strategies are used to select the MCS with the objective of maximizing the spectral efficiency subject to a frugality constraint and a maximum target BLER. Mathematically, denoting by $\mathcal{M}$ the set of available MCSs in the system, the *instantaneous* BLER experienced by user u when transmitting a transport block on the set of RBs $\mathcal{N}_{Bu}$ using MCS $m\in\mathcal{M}$, can be expressed as

$$\mathrm{BLER}_{u,l}^{(m)}\left(\mathcal{N}_{Bu}\right) = \mathcal{F}_m\left(\boldsymbol{\gamma}_{u,l}\left(\mathcal{N}_{Bu}\right)\right), \tag{24}$$

where $\mathcal{F}_m(\cdot)$ represents an MCS-dependent mapping function and N_{Bu} is the cardinality of the set $\mathcal{N}_{Bu}$. The optimum MCS selection process can then be formulated as

$$\begin{aligned}\mu_{u,l} = \arg\max_{m\in\mathcal{M}}\ & R_m\left(N_{Bu}\right)\left[1-\mathcal{F}_m\left(\boldsymbol{\gamma}_{u,l}\left(\mathcal{N}_{Bu}\right)\right)\right]\\ \text{subject to } & R_m\left(N_{Bu}\right)\le Q_{u,l}\\ & \mathcal{F}_m\left(\boldsymbol{\gamma}_{u,l}\left(\mathcal{N}_{Bu}\right)\right)\le \mathrm{BLER}_0,\end{aligned} \tag{25}$$

where $R_m\left(N_{Bu}\right)$ denotes the amount of transmitted information bits when using MCS m on a particular set of N_{Bu} RBs. Furthermore, assuming that when allocating bits to spatial stream l, spatial streams with indexes ranging from 1 to $l-1$ have already been allocated an amount of bits $\sum_{k=1}^{l-1} R_{\mu_{u,l}}\left(N_{Bu}\right) < Q_u$, then the frugality constraint enforces $Q_{u,l} = Q_u - \sum_{k=1}^{l-1} R_{\mu_{u,k}}\left(N_{Bu}\right)$.

Unfortunately, it is very difficult, if at all possible, to find closed-form mapping functions $\mathcal{F}_m(\cdot)$ that accurately relate the vector of SINRs $\boldsymbol{\gamma}_{u,l}\left(\mathcal{N}_{Bu}\right)$ to a single BLER value for an arbitrary RB allocation and a given MCS $m\in\mathcal{M}$. In order to overcome this problem, several link abstraction techniques have been developed that, taking as input the vector of SINRs, obtain a single scalar quality value $\gamma_{\mathrm{eff}_{u,l}}^{(m)}\left(\mathcal{N}_{Bu}\right)$, denoted as *effective* SINR, that can be easily associated to $\mathrm{BLER}_{u,l}^{(m)}\left(\mathcal{N}_{Bu}\right)$ using a set of reference look-up tables (LUTs) containing the BLER values for the different MCSs and transport block lengths in AWGN channels [31]. That is, the effective SINR can be defined as the SINR that would be required by a given MCS on an AWGN channel to attain the same BLER value experienced on the frequency selective fading channel characterized by the vector $\boldsymbol{\gamma}_{u,l}\left(\mathcal{N}_{Bu}\right)$. As an example, Fig. 3 shows the AWGN BLER curves of the transmission modes representing the 15 LTE/LTE-Advanced CQIs (without HARQ) for $N_{Bu}=1$ RB (dashed lines) and $N_{Bu}=25$ RBs (solid lines). As it can be observed, the higher the value of N_{Bu} the higher is the turbo code block size that can be applied and the steeper is the slope of the waterfall region of the corresponding BLER curves.

[6]In systems using hybrid automatic repeat request (HARQ), the combination of signals (typically in the form of log-likelihood ratios (LLRs)) received in different time-slots, and even the unequal error protection of bits transmitted on high order modulation symbols, can exacerbate the multistate nature of the link model.

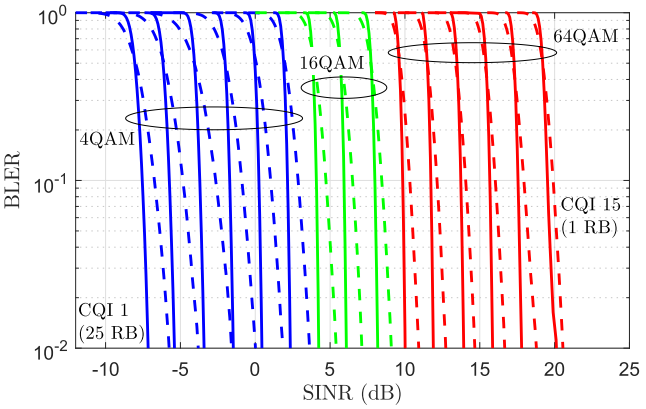

FIGURE 3. AWGN BLER curves versus SINR for the 15 LTE CQIs without HARQ (dashed and solid lines correspond, respectively, to $N_{Bu}=1$ RB and $N_{Bu}=25$ RBs) [30].

One of the most popular choices for link abstraction is the mutual information effective SINR mapping (MIESM), which has the advantage over alternative methods, such as exponential effective SINR mapping (EESM), that it does not require an empirical off-line calibration step as long as codes that perform close to capacity are employed (e.g., Turbo codes or low-density parity check codes) [32]. Using the MIESM-based approach, the effective SINR experienced by user u when using MCS mode m to transmit spatial stream l on a multistate channel characterized by $\boldsymbol{\gamma}_{u,l}\left(\mathcal{N}_{Bu}\right)$ can be obtained as

$$\gamma_{\mathrm{eff}_{u,l}}^{(m)}\left(\mathcal{N}_{Bu}\right) = J_{\mathfrak{m}(m)}^{-1}\left(\frac{1}{N_{sc}N_{Bu}}\sum_{\forall\, b\in\mathcal{N}_{Bu}}\sum_{s=1}^{N_{sc}} J_{\mathfrak{m}(m)}\left(\gamma_{u,b,s,l}\right)\right), \tag{26}$$

where $\mathfrak{m}(m)$ denotes the modulation format used in MCS m and $J_{\mathfrak{m}(m)}(\cdot)$ denotes the bit-interleaved coded modulation (BICM) capacity for the modulation format $\mathfrak{m}(m)$, which can be obtained by numerically evaluating [33, eq. (14)] via the Monte Carlo method. Figure 4 shows the BICM capacity versus SINR for 4QAM, 16QAM and 64QAM over the AWGN channel. Notice that all MCSs sharing the same modulation format experience the same effective SINR.

Using the previously described strategy, the optimization problem posed in (25) simplifies to

$$\begin{aligned}\mu_{u,l} = \arg\max_{m\in\mathcal{M}} \min & \left(Q_{u,l},\, \rho_{u,l}^{(m)}\left(\mathcal{N}_{Bu}\right)\right)\\ & \times\left[1-\psi_{N_{Bu}}^{(m)}\left(\gamma_{\mathrm{eff}_{u,l}}^{(m)}\left(\mathcal{N}_{Bu}\right)\right)\right]\end{aligned} \tag{27}$$

where $\psi_{N_{Bu}}^{(m)}\left(\gamma_{\mathrm{eff}_{u,l}}^{(m)}\left(\mathcal{N}_{Bu}\right)\right)$ represents the curve used to map the effective SINR $\gamma_{\mathrm{eff}_{u,l}}^{(m)}\left(\mathcal{N}_{Bu}\right)$ onto a predicted BLER value

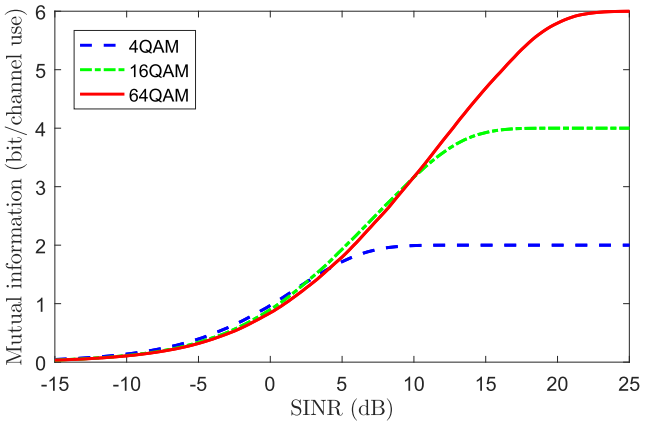

FIGURE 4. BICM capacity as a function of the SINR for the 4QAM, 16QAM and 64QAM modulation formats.

for a given MCS m and a number of allocated RBs N_{Bu} (see the curves plotted in Fig. 3). Furthermore, $\rho_{u,l}^{(m)}(\mathcal{N}_{Bu})$ denotes the amount of information bits that can be potentially transmitted by the BS on the subset of RBs $\mathcal{N}_{Bu}$, using spatial stream l and with a guaranteed BLER less or equal than BLER_0, which can be expressed as

$$\rho_{u,l}^{(m)}(\mathcal{N}_{Bu}) = \begin{cases} N_{Bu}\varrho_m, & \gamma_{\text{eff}_{u,l}}^{(m)}(\mathcal{N}_{Bu}) \geq \Gamma_m(N_{Bu}) \\ 0, & \text{otherwise,} \end{cases} \tag{28}$$

where $\varrho_m = N_{sc}N_iR_{cm}\log_2 M_m$ is the number of information bits per RB when using MCS m, with R_{cm} and M_m denoting the coding rate and the modulation order of MCS m. Moreover, $\{\Gamma_m(N_{Bu})\}_{m=1}^{M}$, with $\Gamma_m(N_{Bu}) \leq \Gamma_{m+1}(N_{Bu})$, are the boundaries defining the intervals of effective SINR over which a given MCS provides a BLER upperbounded by BLER_0 when using N_{Bu} RBs. As an example, the values of the coding rate and SINR boundaries for the 15 MCSs that will be used in this work are listed in Table 1 for $N_{Bu} = 1, 4$ and 25 RBs and $\text{BLER}_0 = 0.1$ (see also the results presented in Fig. 3).

B. GREEDY ALLOCATION OF RBs

As previously stated, the constrained optimization problem posed in (20) is in charge of determining, for each scheduling period, the RB allocation set $\boldsymbol{\mathcal{N}}_B = \{\mathcal{N}_{B1}, \ldots, \mathcal{N}_{BN_{\text{MS}}}\}$ and the MCS allocation set $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{N_{\text{MS}}}\}$ optimizing the weighted system goodput taking into account that: i) a given RB has to be exclusively allocated to a single user, ii) the resources have to be allocated frugally and iii) the selected MCS has to guarantee a transmission with a BLER lower than a prescribed target BLER_0. In spite of using the rather simplified link adaptation procedure described in Subsection III-A, a brute-force exhaustive search over all possible combinations of exclusive allocations of RBs is computationally prohibitive even for modest values of N_{MS} and N_B^{PHY}. Thus, in order to reduce the complexity of the search, a greedy suboptimal approach is presented in this section whose global complexity increases linearly with N_{MS} and N_B^{PHY}.

TABLE 2. Greedy algorithm for RB and MCS allocation.

Inputs:
QSI: $Q_u \ \forall\ u \in \mathcal{N}_u$
CSI: $\boldsymbol{\gamma}_{u,l}(\mathcal{N}_B) \ \forall\ u \in \mathcal{N}_u,\ l \in \{1, \ldots, L_u\}$
Scheduler weights: $w_u \ \forall\ u \in \mathcal{N}_u$

Initialization:
Set of active users: $\mathcal{N}_u^{(A)} = \{u \in \mathcal{N}_u \ :\ Q_u \neq \emptyset\}$
Set of non-allocated RBs: $\mathcal{N}_B^{(\text{free})} = \mathcal{N}_B$
Sets of allocated RBs (per user): $\mathcal{N}_{Bu} = \emptyset\ \forall\ u \in \mathcal{N}_u^{(A)}$
Sets of allocated MCSs (per user): $\boldsymbol{\mu}_u = \emptyset\ \forall\ u \in \mathcal{N}_u^{(A)}$
Utility (per user): $U_u = 0\ \forall\ u \in \mathcal{N}_u^{(A)}$

Allocation of the first RB:
for $u \in \mathcal{N}_u^{(A)}$
for $b \in \mathcal{N}_B$
$\Delta U_{u,b} = w_u \sum_{l=1}^{L_u} \rho_{u,l}^{(\upsilon_{u,l})}(\{b\}) \left[1 - \psi_1^{(\upsilon_{u,l})}\left(\gamma_{\text{eff}_{u,l}}^{(\upsilon_{u,l})}(\{b\})\right)\right]$
end
end
$(u^*, b^*) = \arg \max_{\substack{u \in \mathcal{N}_u^{(A)} \\ b \in \mathcal{N}_B}} \Delta U_{u,b}$
if $\sum_{l=1}^{L_u^*} \rho_{u^*,l}^{(\upsilon_{u^*,l})}(\{b^*\}) = Q_{u^*}$
$\mathcal{N}_u^{(A)} = \mathcal{N}_u^{(A)} \backslash u^*$
end
$\mathcal{N}_B^{(\text{free})} = \mathcal{N}_B \backslash b^*, \quad \mathcal{N}_{Bu^*} = \{b^*\}, \quad U_{u^*} = \Delta_{u^*,b^*}, \quad \boldsymbol{\mu}_{u^*} = \boldsymbol{\upsilon}_{u^*}$

Allocation of the remaining RBs:
while $\mathcal{N}_B^{(\text{free})} \neq \emptyset$ and $\mathcal{N}_u^{(A)} \neq \emptyset$
if $u^* \in \mathcal{N}_u^{(A)}$
for $b \in \mathcal{N}_B^{(\text{free})}$
$U_{u,b}^{\text{aux}} = w_{u^*} \sum_{l=1}^{L_{u^*}} \rho_{u^*,l}^{(\upsilon_{u^*,l})}(\mathcal{N}_{Bu^*} \cup \{b\}) \times \left[1 - \psi_{N_{Bu^*}+1}^{(\upsilon_{u^*,l})}\left(\gamma_{\text{eff}_{u^*,l}}^{(\upsilon_{u^*,l})}(\mathcal{N}_{Bu^*} \cup \{b\})\right)\right]$
$\Delta U_{u^*,b} = U_{u,b}^{\text{aux}} - U_{u^*}$
end
end
$(u^*, b^*) = \arg \max_{\substack{u \in \mathcal{N}_u^{(A)} \\ b \in \mathcal{N}_B^{(\text{free})}}} \Delta U_{u,b}$
if $\sum_{l=1}^{L_{u^*}} \rho_{u^*,l}^{(\upsilon_{u^*,l})}(\mathcal{N}_{Bu^*} \cup \{b\}) = Q_{u^*}$
$\mathcal{N}_u^{(A)} = \mathcal{N}_u^{(A)} \backslash u^*$
end
$\mathcal{N}_B^{(\text{free})} = \mathcal{N}_B \backslash b^*, \quad \mathcal{N}_{Bu^*} = \{b^*\}, \quad U_{u^*} = U_{u,b}^{\text{aux}}, \quad \boldsymbol{\mu}_{u^*} = \boldsymbol{\upsilon}_{u^*}$
end

Outputs:
RBs allocation set: $\boldsymbol{\mathcal{N}}_B = \{\mathcal{N}_{B1}, \ldots, \mathcal{N}_{BN_{\text{MS}}}\}$
MCSs allocation set: $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{N_{\text{MS}}}\}$

The utility-based greedy suboptimal RB (and, implicitly, MCS) allocation algorithm is detailed in Table 2. In words, using the QSI obtained from the DLC, the CSI received from the MSs and the scheduler weights, the algorithm first initializes the set of active users (set of users with a non empty queue), the set of non allocated RBs and, for each active user in the system, the set of allocated RBs, the set of allocated MCSs and the utilities. After initialization, the algorithm first obtains the utility increase that would potentially produce the allocation of each of the non allocated RBs to each of

the active users in the system and then the first RB b^* is allocated to the active user u^* for which the corresponding utility increase is maximized. Note that the allocation of this RB leads to a provisional MCS allocation $\mu_{u^*,l}$, for all $l \in \{1, \ldots, L_u\}$, that will be permanent if at the end of the algorithm user u^* has not been allocated any other RB. If the data transmission capacity (measured in bits) of this RB, when used to transmit to user u^*, is higher than the corresponding queue length, then user u^* is removed from the set of active users. Prior to the next step in the algorithm, the allocated RB b^* is removed from the set of non allocated RBs and put in the set of RBs allocated to user u^*. Furthermore, the provisional utility of user u^* is set to the utility increase produced by the corresponding allocation of RB b^*. After the allocation of the first RB the greedy algorithm enters a loop that is executed while the sets of non allocated RBs and active users are not empty. In each loop cycle a new RB (and, implicitly, a new MCS) is allocated to the active user making the most of it in terms of utility increase. To do that, if the user u^* that was selected in the previous step is still active, the corresponding utility increase for each of the non allocated RBs is updated taking into account all the RBs previously allocated to that particular user. Next, RB b^* is allocated to the active user u^* for which the corresponding utility increase is maximized. Again, if the joint capacity of all RBs allocated to user u^* is higher than the corresponding queue, then user u^* is removed from the set of active users. Furthermore, prior to the next cycle, RB b^* is added to the set of RBs allocated to user u^* an removed from the set of non allocated RBs, and the provisional utility of user u^* is set to the utility produced by the corresponding allocation of RBs. After exiting the loop, the algorithm provides suboptimal RB and MCS allocation sets.

IV. NUMERICAL RESULTS

In this section, a circular single cell downlink MIMO-multicarrier wireless system with a central BS serving a set of MSs uniformly distributed over the whole coverage area is considered. Whenever possible, system parameters have been defined based on current LTE/LTE-Advanced specifications [34], with some specific modifications mainly affecting, for obvious reasons, the FBMC/OQAM-based scheme. In particular, Table 3 summarizes the most relevant parameters considered in the simulations.[7] The PHYDYAS prototype pulse with overlapping factor $\kappa = 4$ has been used in the simulations of the FBMC/OQAM system [35]. Remarkably, because of the higher spectral confinement shown by the FBMC/OQAM PHY layer, allowing for a lower out-of-band radiation, it requires of much smaller guard bands than those needed by the OFDM-based scheme. Thus, a system bandwidth of $B = 10$ MHz, which is divided into 50 subbands when using an OFDM-based LTE/LTE-Advanced system, can be split into 55 subbands if the PHY layer is based on a FBMC/OQAM strategy. Furthermore, as the FBMC/OQAM-based scheme does not use cyclic prefixing, the TTI duration of 1 ms specified by the OFDM-based LTE/LTE-Advanced system can be reduced to $0.9\dot{3}$ ms when using an FBMC/OQAM-based PHY layer.[8] The combined multiplicative effects of a higher spectral confinement and absence of cyclic prefixing provide the FBMC/OQAM-based PHY layer a brute spectral efficiency advantage of roughly 18% when compared to the OFDM-based one. Furthermore, the power that the OFDM-based system dedicates to cyclic prefixing can be fully devoted to payload transmission in a FBMC/OQAM-based system. Hence, a PHY layer based on FBMC/OQAM presents a non-negligible energy efficiency advantage with respect to a PHY layer based on OFDM. It remains to be shown how this brute spectral/energy efficiency advantage can be exploited by the cross-layer SRA algorithms and the effects the additional ISI plus ICI term has on the FBMC/OQAM-based system.

TABLE 3. Summary of simulation parameters.

Common Parameters	Value	
BS transmit power (P_T)	47 dBm	
BS/UE antenna gain (G_T/G_R)	6/2 dBi	
UE noise figure (NF_{UE})	7 dB	
Thermal noise (N_0)	-174 dBm/Hz	
Cell radius (R)	500 m	
Shadowing deviation (σ_s)	6 dB	
Default MIMO configuration ($N_T \times N_R$)	2×2	
Carrier frequency (f_0)	2.1 GHz	
System bandwidth (B)	10 MHz	
Subcarrier spacing (Δf)	15 kHz	
Subcarriers per subband (N_{sc})	12	
Multicarrier symbols per TTI (N_s)	14	
Multicarrier information symbols per TTI (N_i)	11	
Specific Parameters	**OFDM**	**FBMC**
Orthogonal subbands (N_B^{PHY})	50	55
Slot duration ($T_{\text{slot}}^{\text{PHY}}$)	0.5 ms	0.45 ms
TTI duration ($T_{\text{TTI}}^{\text{PHY}}$)	1 ms	0.9 ms
Symbols/slot with long cyclic prefix ($N_{s\,\text{long}}$)	1	N/A
Symbols/slot with short cyclic prefix ($N_{s\,\text{short}}$)	6	N/A
Duration of long cyclic prefix ($T_{\text{CP}\,\text{long}}$)	5.2 μs	N/A
Duration of short cyclic prefix ($T_{\text{CP}\,\text{short}}$)	4.7 μs	N/A

On its way from the BS to the MSs, the transmitted signal experiences path-losses, large-scale shadow fading and small-scale frequency-, time- and space-selective fading. The path-losses have been simulated using the macro cell propagation model for urban area described in [36, Sec. 4.5.2], assuming a BS antenna height of 30 m and a log-normally distributed shadow fading with a standard deviation of 10 dB.

[7] Note that we assume the use of normal cyclic prefixing. In a fair comparison between OFDM and FBMC/OQAM in a system using extended cyclic prefixing the potential advantages shown by the FBMC/OQAM-based system would be even more exacerbated.

[8] An equivalent spectral efficiency advantage can be obtained by maintaining the TTI duration of 1 ms and increasing the number of multicarrier symbols per TTI from 14 to 15.

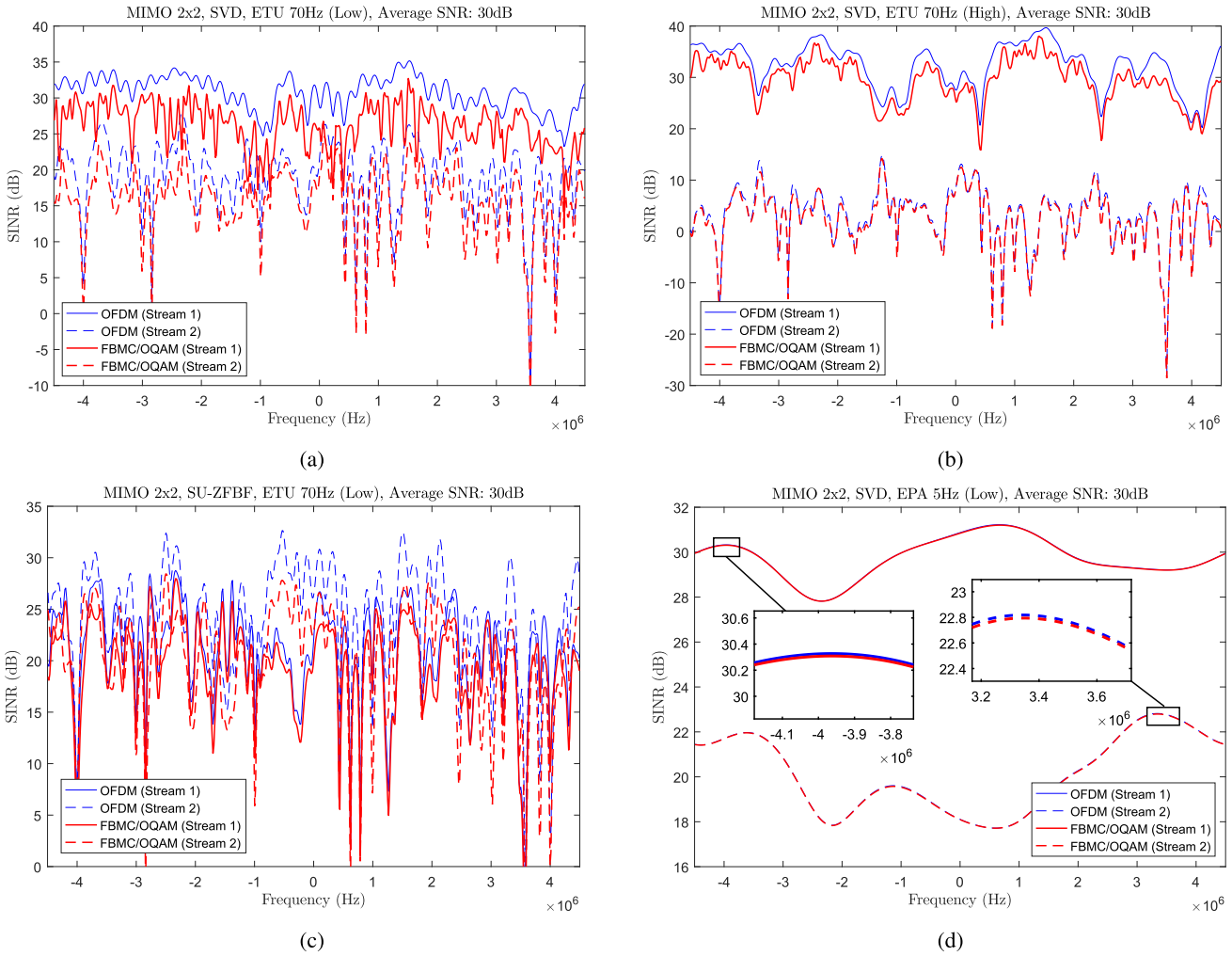

FIGURE 5. A realization of the instantaneous SINR experienced at the output of the MIMO equalizer under different system and channel conditions. (a) SVD over ETU channel (70 Hz, Low correlation). (b) SVD over ETU channel (70 Hz, High correlation). (c) SU-ZFBF over ETU channel (70 Hz, Low correlation). (d) SVD over EPA channel (5 Hz, Low correlation).

Small-scale fading has been generated using channel power delay profiles conforming to either the 3GPP Extended Typical Urban (ETU) or the 3GPP Extended Pedestrian A (EPA) channel models defined within LTE/LTE-Advanced [37], showing root mean square delay spreads of 991 ns and 45 ns, respectively. Furthermore, the ETU and EPA channel models have been simulated using maximum Doppler frequencies of $f_d = 70$ Hz and $f_d = 5$ Hz, respectively. That is, the ETU channel with $f_d = 70$ Hz represents an environment showing high time and frequency selectivity, whereas the EPA channel with $f_d = 5$ Hz is a good example of a low-mobility scenario characterized by very low time and frequency selectivity. Moreover, for all scenarios, a default 2×2 MIMO configuration has been considered with transmit and receive correlation matrices conforming to one of the *low*, *medium* or *high* correlation profiles defined in [37].

A. SINR AT THE OUTPUT OF THE MIMO EQUALIZER

Our aim in this subsection is to analyze the behaviour of the instantaneous SINR at the output of the MIMO equalizer under different system and channel conditions. Figure 5a shows a realization of the per-subcarrier SINR experienced by the two spatial streams at the output of a singular value decomposition (SVD)-based scheme with uniform power allocation per stream [21] when using both the FBMC/OQAM- and OFDM-based PHY layers to transmit on an ETU channel model characterized by a *low* MIMO correlation profile and an average SNR of 30 dB. As it can be observed, the ETU channel model is characterized by a very high frequency selectivity. Hence, as the power of the ISI/ICI term in (19) is directly related to the first and second derivatives (with respect to the frequency) of both the MIMO channel matrix $\boldsymbol{H}_{u,s,b}$ and the MIMO precoding matrix $\boldsymbol{A}_{u,b,s}$, the SINR values experienced at the

output of the FBMC/OQAM-based PHY layer are noticeably worse than those observed at the output of the OFDM-based one. Note also that the difference in SINR performance between FBMC/OQAM and OFDM is more apparent for the first spatial stream than for the second one. This can explained by the fact that the first spatial stream at the output of the SVD equalizer experiences an average SINR regime much higher than that experienced by the second spatial stream and consequently, the corresponding instantaneous SINR presents a higher tendency to be limited by the ISI/ICI distortion induced by the channel frequency selectivity.[9] This effect is exacerbated by the transmit/receive antenna correlation as it can be observed in Fig. 5b where, again, a realization of the SINR experienced by the two spatial streams at the output of an SVD-based scheme is presented but, in this case, assuming an ETU channel model characterized by a *high* MIMO correlation profile. Another interesting and well-known effect that can be corroborated when analyzing graphs presented in Figs. 5a and 5b is that increasing the spatial correlation widens the difference between the SINR values experienced by the two spatial streams.

The effects of using different MIMO linear precoding/equalization strategies can be inferred from the comparison of the instantaneous SINR results presented in Figs. 5a and 5c, where the only difference is that the graphs presented in the former correspond to an SVD-based scheme while those presented in the latter correspond to a single-user zero forcing beamforming (SU-ZFBF)-based scheme with vector normalization [21], [38]. As stated in the previous paragraph, these results have been obtained assuming that the transmission takes place on an highly frequency selective ETU channel model characterized by a *low* MIMO correlation profile. Similar to the SVD-based scheme, when transmitting on a PHY layer based on FBMC/OQAM, the SU-ZFBF precoder is considerably affected by the ISI/ICI term induced by the MIMO channel frequency selectivity. The only difference between these precoding/equalization strategies is that the SU-ZFBF scheme produces two spatial streams with exactly the same statistical distribution while the SVD generates two spatial streams with very dissimilar statistical behaviors. Although not shown in this figure because of space limitations, the zero forcing (ZF) equalizer provides exactly the same performance results as the SU-ZFBF precoder. Furthermore, the conventional minimum mean square error (MMSE)[10] equalizer only provides a negligible performance improvement when compared to the SU-ZFBF precoder (or, equivalently, the ZF equalizer).

Finally, Fig. 5d shows a realization of the per-subcarrier SINR experienced by the two spatial streams at the output of an SVD-based scheme when using both the FBMC/OQAM- and OFDM-based PHY layers to transmit on an EPA channel model characterized by a *low* MIMO correlation profile. The EPA channel is characterized by a frequency response that is almost flat at the subcarrier level. Consequently, when transmitting on a MIMO channel characterized by the EPA profile, the ISI/ICI term in (19) almost vanishes and the instantaneous SINR values experienced at the output of the FBMC/OQAM-based PHY layer tend to those experienced at the output of the OFDM-based PHY layer.

B. HOMOGENEOUS TRAFFIC. CHANNEL-AWARE SRA

In this subsection we compare the performance of OFDM- and FBMC/OQAM-based PHY layers when dealing with homogeneous (single class) users with saturated queues (full-buffer) and assuming the use of a 2×2 MIMO arrangement with a ZF equalizer at the receiver side. Under homogeneous traffic saturation conditions, the availability of QSI does not provide any advantage to channel- and queue-aware SRA algorithms and consequently, only channel-aware SRA units will be considered in this scenario.

The average system goodputs for both FBMC/OQAM- and OFDM-based PHY layers and assuming the use of either an MSR or a PF scheduler are compared in Fig. 6a. Results presented in this figure have been plotted against the number of active users in the system and assuming an ETU channel model with $f_d = 70$ Hz and a low correlation profile. As expected, the MSR scheduler provides the best goodput performance as it is specifically designed to exploit the multiuser (and multicarrier) diversity. The PF scheduling algorithm, instead, sacrifices transmission rate in return for goodput fairness among users. Irrespective of the scheduling rule, the FBMC/OQAM-based PHY layer provides a clear spectral efficiency advantage when compared to the OFDM-based one. Remarkably, under heavy user loads this spectral efficiency advantage approaches the promised 18%, thus proving that the combination of an FBMC/OQAM-based PHY layer with a properly designed cross-layer SRA algorithm can make use of the better spectral containment of FBMC filters without barely sacrificing performance due to increased ISI/ICI distortion. It is interesting to notice that when the number of users is large the MSR scheduler can allocate, with a very high probability, all the RBs to users (and subbands) experiencing very favorable channel propagation conditions and thus can transmit using the highest MCS on both spatial streams. Under these circumstances, the saturation goodput of the OFDM- and FBMC/OQAM-based system approach the maximum of 72.6 Mbit/s and 85.56 Mbit/s, respectively.

Results presented in Fig. 6b, comparing the average system goodputs for both FBMC/OQAM- and OFDM-based PHY layers when transmitting on either an ETU ($f_d = 70$ Hz) or an

[9] Although not shown in these figures, the instantaneous SINR values at the output of both PHY layer schemes are almost indistinguishable when considering a MS experiencing a fixed average SNR of 0 dB. That is, the ISI/ICI distortion caused by the channel frequency selectivity becomes irrelevant for a MS working in a noise-limited regime.

[10] The conventional MMSE equalizer is designed by only taking into account the MSI and the AWGN. It is optimal for the ideal OFDM-based system but it is suboptimal for the FBMC/OQAM-based scheme. Improved, yet more complex, MMSE designs for the FBMC/OQAM-based system can be obtained by considering also the ISI/ICI terms as was done, for instance, in [12, Sec. III].

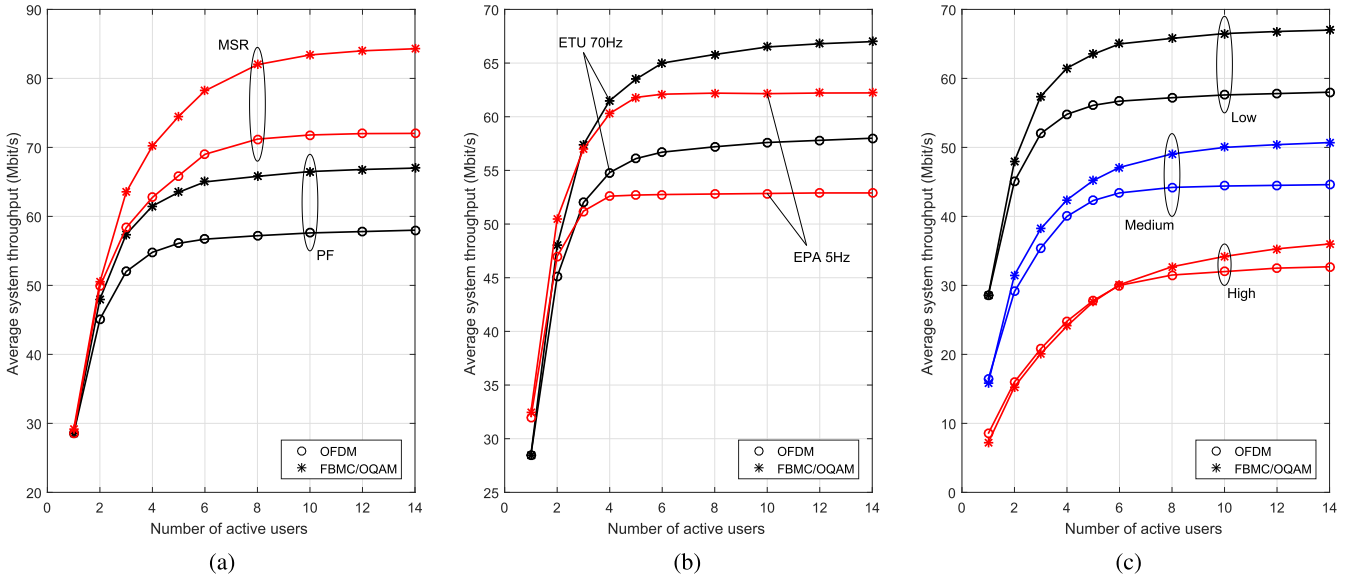

FIGURE 6. Average system goodput for both FBMC/OQAM- and OFDM-based PHY layers under different system and channel configurations (MIMO 2 × 2 with a ZF equalizer). (a) Goodput *vs* scheduling rules. (b) Goodput *vs* channel model. (c) Goodput *vs* correlation profile.

EPA ($f_d = 5$ Hz) channel model, have been plotted against the number of active users in the system and assuming the use of a PF scheduler. As it can be observed, the spectral efficiency advantage provided by the FBMC/OQAM-based PHY layer over the OFDM-based one when transmitting on an ETU channel model is also maintained when transmitting on an EPA channel model. However, the lower frequency selectivity shown by the EPA channel seems to play a different role when transmitting on scenarios with either a low or a high user load. In scenarios in which the number of active users is low, the lower frequency selectivity of the EPA channel favours the special characteristics of the joint use of a greedy resource allocation algorithm and a discrete rate allocation scheme that provide a goodput performance slightly higher than the one obtained when transmitting on an ETU channel. In highly loaded scenarios, in contrast, the scheduler is able to exploit the higher frequency selectivity provided by the ETU channel model in order to increase the multiuser diversity and this directly results in a clear system goodput performance advantage when compared to systems transmitting on scenarios characterized by the EPA channel model.

Finally, Fig. 6c shows results comparing the average system goodputs for both FBMC/OQAM- and OFDM-based PHY layers when transmitting on an ETU channel model with different correlation profiles and assuming the use of a PF scheduler. As the spatial correlation increases, the capacity of the MIMO channel decreases thus leading to a reduction of the average system goodput. Furthermore, in cases where the MIMO channel is strongly correlated at the transmitter and/or receiver sides, the channel matrix tends to be ill-conditioned and the two spatial streams tend to show a very dissimilar statistical behaviour, with one of the streams experiencing very high SINR values and the other stream being affected by very low SINRs. This behaviour has significant consequences in an LTE-like system using a finite set of transmission modes. In particular, even though the spatial stream experiencing a very high SINR is often characterized by a very high capacity, it can only be allocated the maximum MCS provided by the AMC schemes of LTE/LTE-Advanced (see Table 1). Thus, although the FBMC/OQAM-based systems experience higher SINR values than the OFDM-based ones, they cannot fully exploit this advantage because in most cases the performance becomes limited by the largest available MCS, that is, the AMC scheme is not able to allocate an MCS matching the capacity of the corresponding spatial streams.

C. HETEROGENEOUS TRAFFIC. CHANNEL- AND QUEUE-AWARE SRA

The proposed framework is able to schedule and allocate resources to service flows with heterogeneous QoS requirements, including RT, nRT and BE traffic classes. In order to show how the cross-layer SRA unit behaves when dealing with heterogeneous traffic on PHY layers based either on FBMC/OQAM or OFDM, a multi-class scenario has been set-up in which $N_{\text{MS}} = 15$ MSs are partitioned into three equal-sized groups (5 MSs per group) with each group representing a different traffic class. Irrespective of the user class, traffic arrivals have been modeled as Poisson random variables, with a mean that depends on the average arrival rate per flow (measured in bits per second). Furthermore, conforming to LTE/LTE-A specifications (see [3, Table 13.1]), RT, nRT and BE traffic flows have been characterized by maximum allowable delays of 50, 100 and 300 ms, and delay outage probabilities of 0.01, 0.1 and 0.1, respectively.

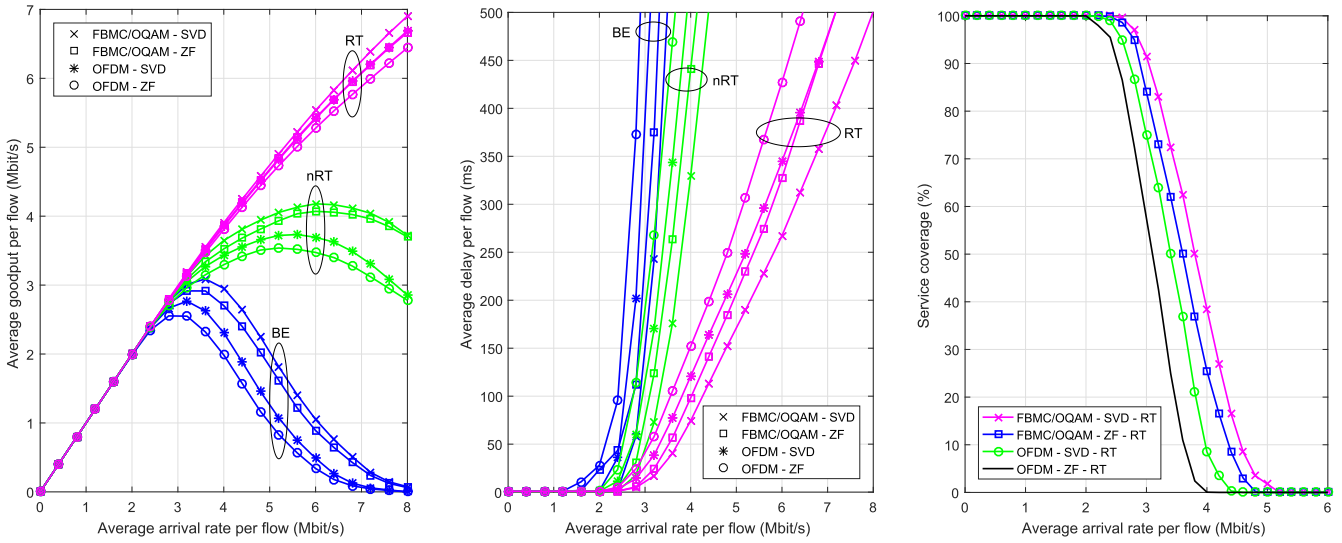

FIGURE 7. Performance metrics as a function of the incoming average traffic for both FBMC/OQAM- and OFDM-based PHY layers and assuming the use of both ZF and SVD schemes (ETU 70 Hz channel model with low correlation). (a) Average goodput per flow. (b) Average delay per flow. (c) Service coverage.

Figures 7a-7c show the per-user average goodput, per-user average delay and service coverage, respectively, as a function of the corresponding incoming average traffic (measured in Mbit/second) for both FBMC/OQAM- and OFDM-based PHY layers and assuming the use of both ZF and SVD precoding/equalization schemes. Although not shown in the graphs, for the sake of clarity, obtained results using SU-ZFBF coincide with those attained using a ZF equalizer and the MMSE equalizer provides marginally higher goodput, lower delay and better service coverage than the ZF counterpart. Moreover, despite the fact that similar conclusions could be derived assuming the use of any properly designed queue- and channel-aware scheduler, all the results presented in these figures have been obtained using an EXP scheduling rule. The first interesting result to highlight is that, without any doubt, the spectral/energy advantage of FBMC/OQAM over OFDM transforms into QoS performance gains that, although they could be considered rather marginal for RT traffic, they are very remarkable for nRT and BE services. Furthermore, irrespective of the PHY layer in use, the SVD-based strategy, as expected, outperforms the simpler ZF scheme although, again, the performance differences between both precoding/equalization approaches become more evident for nRT and BE traffic classes.

For low arrival rates and irrespective of the traffic class, multicarrier processing strategy or precoding/equalization scheme in use, the goodput and delay graphs behave as unit-slope and constant lines, respectively, indicating that any of the PHY layer settings under study is able to serve all incoming traffic with a negligible average delay. As traffic arrivals increase, the goodput and delay graphs corresponding to different traffic classes, different PHY layers and different MIMO precoding/equalization schemes eventually diverge from the unit-slope and constant lines. The point at which divergence starts, which is dependent on the traffic class, the PHY layer and the MIMO precoding/equalization strategy, reveals at which average arrival rate the incoming traffic is not fully served and there are some users for which part of the incoming traffic accumulates in the queues thus increasing the average delay. Irrespective of the PHY layer and MIMO precoding/equalization scheme in use, the queue- and channel-aware scheduler works as expected when dealing with the different traffic classes. As incoming traffic increases, the first users suffering service restrictions are the ones using BE applications, in an attempt to fulfil the demands of users running nRT and RT applications. Then, as the arrival traffic keeps increasing and the system is unable to fully serve nRT and RT users, the nRT service also begins to suffer restrictions for the benefit of RT users which have the most stringent QoS requirements. Eventually, for very high arrival rates, even the RT users experience service restraints that translate into a goodput saturation and an increase of delay. Regarding the delay performance, it is interesting to note the differences in the behavior of the graphs for different traffic classes. On one hand, it can be observed how the delay experienced by BE users, and to a lesser extent by nRT users, increases at a very high pace, effectively implying that once the traffic load reaches the point for which the system is not able to fully serve all the traffic demands, the scheduler quickly decreases the weighing factors of the BE, and to a lesser extent the nRT services, in order to prioritize the RT services. On the other hand, because there is no traffic class with a higher priority that the RT one, the average delay graphs corresponding to RT users rise at a rather moderate

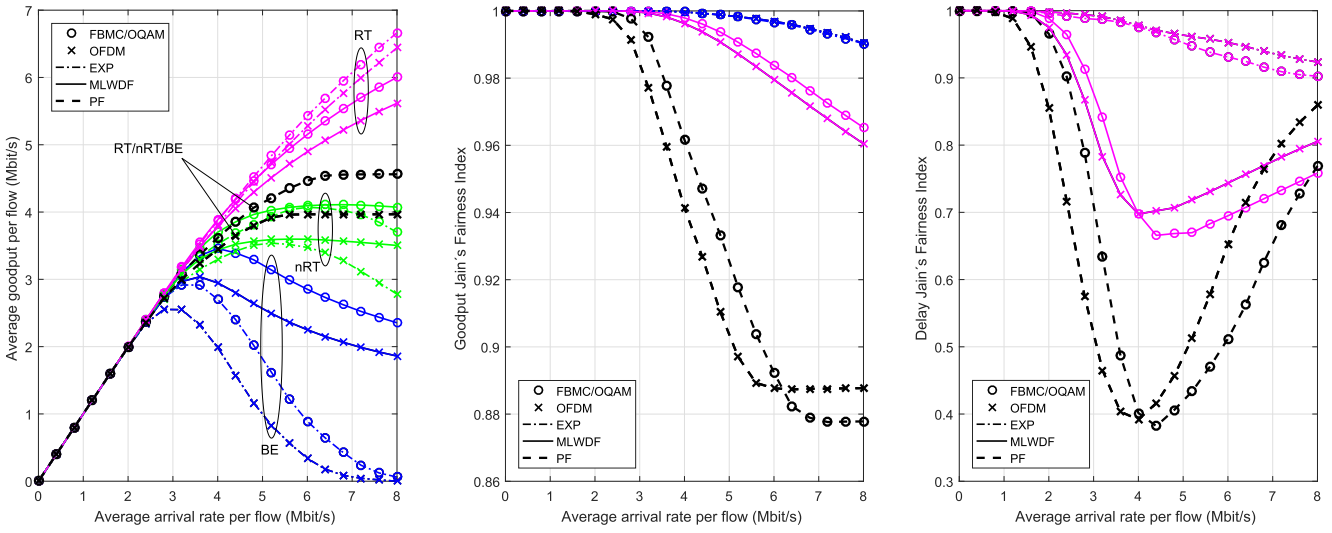

FIGURE 8. Performance metrics as a function of the incoming average traffic for FBMC/OQAM- and OFDM-based PHY layers and assuming the use of PF, MLWDF or EXP schedulers (ZF equalizer, ETU 70 Hz channel model with low correlation). (a) Average goodput per flow. (b) Goodput JFI (RT). (c) Delay JFI (RT).

pace as the scheduler keeps providing them with almost all of the available resources in the system.

The advantage of FBMC/OQAM over OFDM can be fully appreciated when examining the service coverage performance of both PHY layers as shown in Fig. 7c. For the sake of presentation clarity, only results for the RT traffic class are presented notwithstanding the fact that for the nRT and BE traffic classes similar qualitative trends have been observed. In particular, it can be observed how the FBMC/OQAM-based system is able to sustain average traffic arrivals per-user of 90% of the RT users for up to 3 and 2.9 Mbps when using either SVD or ZF, respectively, whereas the corresponding arrival rates supported by an OFDM-based PHY layer using either SVD or ZF barely reach 2.7 and 2.5 Mbps, respectively.

Just to show that the results obtained using the EXP scheduling rule are qualitatively generalizable to any queue- and channel-aware scheduler but not to an only channel-aware scheduler, Figs. 8a-8c compare, respectively, the average goodput per flow, the well-know goodput Jain's fairness index (JFI) and delay JFI [39] as a function of the average arrival rate per flow for the PF, EXP and MLWDF scheduling rules. To obtain the results plotted in these figures, it has been assumed the use of both FBMC/OQAM- and OFDM-based PHY layers to transmit on a MIMO ETU channel model characterized by a low correlation profile employing a ZF equalizer (although not shown in the graphs, similar results have been obtained using other precoding/equalization schemes). Focusing first on the average goodput results (see Fig. 8a), note that, in contrast to the EXP and MLWDF schedulers, the PF scheduling rule is unable to differentiate between traffic classes as its SRA decisions are taken only on the basis of CSI. Qualitatively, the MLWDF scheduler performs exactly the same way as the EXP scheduling rule although, interestingly enough, the EXP scheduler is more expeditious than MLWDF when restraining or even denying service to the BE and nRT users while trying to satisfy the tighter QoS constraints of RT users. In fact, as it can be clearly observed in Fig. 8a, contrasting with the policy of the MLWDF scheduler, when falling short of resources the EXP scheduling rule favors the goodput performance of RT users by allowing the BE users and then the nRT users to starve. Concerning the multicarrier techniques under consideration and irrespective of the scheduling rule, FBMC/OQAM-based PHY layers are able to exploit their spectral/energy advantage over OFDM, which is transformed into significant performance gains in terms of average goodput per flow that are particularly noticeable for BE and nRT traffic classes.

The graphs in Figs. 8b and 8c, showing the goodput and delay JFIs, respectively, serve to complement the goodput outcomes. Again, for clarity of presentation, only the results for the RT traffic class are shown here. An important effect to notice is that, in this particular scenario, the EXP scheduling rule provides the best goodput/delay JFI results at the cost of sacrificing the average goodput of both nRT and BE traffic flows. The PF scheduling rule, being unable to consider the QSI, is the one providing the worst performance in terms of JFI. The performance provided by the MLWDF scheduler lies in the middle ground between EXP and PF, with nRT and BE traffic flows showing a better average goodput performance than that obtained using an EXP scheduling rule to the detriment of the goodput and delay JFIs.

V. CONCLUSIONS

This paper has introduced a unified cross-layer framework for downlink scheduling and resource allocation able to embody the particularities of two of the most promising waveform

technologies under consideration for 5G networks, namely, OFDM and FBMC/OQAM. Unlike previous comparative publications between these two technologies, that mostly focused on purely PHY metrics, this work allows the extraction of user-relevant metrics like throughput, average delay, Jain's fairness index and service coverage in a multi-class traffic environment. Notably, the framework incorporates many of the operational constraints now enforced in current 4G networks (LTE/LTE-Advanced) and likely to remain in 5G standards such as, for example, the availability of a discrete set of MCSs or the selection of the same MCS to transmit on all subbands allocated to a given user. Scheduling has proved to be a fundamental step when maximizing utility functions based on any of the aforementioned metrics. To this end, a new user selection technique, valid for both OFDM and FBMC/OQAM, has been proposed that, based on greedy principles and combined with different schedulers, has led to a variety of tradeoffs between network performance and user fairness. Numerical results have shown that despite the FBMC/OQAM-based PHY layer must account for new sources of interference caused by the orthogonality loss, which are typically not present in OFDM-based architectures, the benefits of not having to transmit a CP per OFDM symbol and the ability to operate without large guard bands clearly outperform the appearance of new interfering terms. Remarkably, the theoretically predicted 18% gain in throughput advantage FBMC/OQAM is expected to provide over OFDM, is indeed almost fully realized in practice. This throughput advantage becomes especially notorious in multi-class traffic scenarios, where FBMC/OQAM-based systems are able to capitalize on this extra resources to very significantly outperform OFDM-based counterparts.

REFERENCES

[1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] S. Buzzi, A. Ugolini, A. Zappone, and G. Colavolpe, "An introduction to modulations and waveforms for 5G networks," in *Signal Processing for 5G: Algorithms and Implementations*. Hoboken, NJ, USA: Wiley, 2016, pp. 616–688.

[3] C. Cox, *An Introduction to LTE: LTE, LTE-Advanced, SAE and 4G Mobile Communications*. Hoboken, NJ, USA: Wiley, 2012.

[4] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. Norwood, MA, USA: Artech House, 2000.

[5] G. Y. Li and G. L. Stuber, *Orthogonal Frequency Division Multiplexing for Wireless Communications* (Signals and Communication Technology). Secaucus, NJ, USA: Springer-Verlag, 2006.

[6] B. Farhang-Boroujeny, "OFDM versus filter bank multicarrier," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 92–112, May 2011.

[7] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, and A. Ugolini, "Modulation formats and waveforms for 5G networks: Who will be the heir of OFDM?: An overview of alternative modulation schemes for improved spectral efficiency," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 80–93, Nov. 2014.

[8] P. Siohan, C. Siclet, and N. Lacaille, "Analysis and design of OFDM/OQAM systems based on filterbank theory," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1170–1183, May 2002.

[9] M. Caus and A. I. Pérez-Neira, "Multi-stream transmission for highly frequency selective channels in MIMO-FBMC/OQAM systems," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 786–796, Feb. 2014.

[10] A. I. Pérez-Neira *et al.*, "MIMO signal processing in offset-QAM based filter bank multicarrier systems," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5733–5762, Nov. 2016.

[11] X. Mestre and D. Gregoratti, "Parallelized structures for MIMO FBMC under strong channel frequency selectivity," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1200–1215, Mar. 2016.

[12] F. Rottenberg, X. Mestre, F. Horlin, and J. Louveaux, "Single-tap precoders and decoders for multiuser MIMO FBMC-OQAM under strong channel frequency selectivity," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 587–600, Feb. 2017.

[13] I. Estella, A. Pascual-Iserte, and M. Payarö, "OFDM and FBMC performance comparison for multistream MIMO systems," in *Proc. Future Netw. Mobile Summit*, Jun. 2010, pp. 1–8.

[14] H. Saeedi-Sourck, Y. Wu, J. W. M. Bergmans, S. Sadri, and B. Farhang-Boroujeny, "Complexity and performance comparison of filter bank multicarrier and OFDM in uplink of multicarrier multiple access networks," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1907–1912, Apr. 2011.

[15] W. Jiang and T. Kaiser, "From OFDM to FBMC: Principles and comparisons," in *Signal Processing for 5G: Algorithms and Implementations*. Hoboken, NJ, USA: Wiley, 2016, pp. 616–688.

[16] D. Zhang, M. Matthé, L. L. Mendes, and G. Fettweis, "A study on the link level performance of advanced multicarrier waveforms under MIMO wireless communication channels," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2350–2365, Apr. 2017.

[17] Q. Bai, N. Passas, and J. A. Nossek, "Scheduling and resource allocation in OFDM and FBMC systems: An interactive approach and performance comparison," in *Proc. Eur. Wireless Conf. (EW)*, Apr. 2010, pp. 1042–1050.

[18] M. Payarö, A. Pascual-Iserte, A. García-Armada, and M. Sanchez-Fernández, "Resource allocation in multi-antenna MAC networks: FBMC vs OFDM," in *Proc. IEEE 73rd Veh. Technol. Conf. (VTC Spring)*, May 2011, pp. 1–5.

[19] M. Shaat and F. Bader, "Computationally efficient power allocation algorithm in multicarrier-based cognitive radio networks: OFDM and FBMC systems," *EURASIP J. Adv. Signal Process.*, vol. 2010, Feb. 2010, Art. no. 528378.

[20] Z. Kong, Y. K. Kwok, and J. Wang, "A low-complexity QoS-aware proportional fair multicarrier scheduling algorithm for OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 58, no. 5, pp. 2225–2235, Jun. 2009.

[21] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications—A key to gigabit wireless," *Proc. IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[22] G. Femenias, B. Dañobeitia, and F. Riera-Palou, "Unified approach to cross-layer scheduling and resource allocation in OFDMA wireless networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, pp. 1–19, 2012.

[23] G. Femenias, F. Riera-Palou, and J. S. Thompson, "Robust scheduling and resource allocation in the downlink of spatially correlated MIMO-OFDMA wireless systems with imperfect CSIT," *IEEE Trans. Veh. Technol.*, vol. 65, no. 2, pp. 614–629, Feb. 2016.

[24] G. Femenias and F. Riera-Palou, "Scheduling and resource allocation in downlink multiuser MIMO-OFDMA systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2019–2034, May 2016.

[25] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1. Sep. 1995, pp. 331–335.

[26] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, Mar. 1998.

[27] M. Andrews *et al.*, "Dynamic bandwidth allocation algorithms for high-speed data wireless networks," *Bell Labs Tech. J.*, vol. 3, no. 3, pp. 30–49, Jul. 1998.

[28] S. Shakkottai and A. Stolyar, "Scheduling for multiple flows sharing a time varying channel: The exponential rule," Bell Labs, Murray Hill, NJ, USA, Tech. Rep., 2000.

[29] S. Shakkottai and A. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR," Bell Labs, Lucent Technol., New Providence, NJ, USA, Tech. Rep., 2000.

[30] A. Serra-Pagés, "Link level performance evaluation and link abstraction for LTE/LTE-advanced downlink," Ph.D. dissertation, Dept. Signal Theory Commun., Universitat Politécnica de Catalunya, Barcelona, Spain, Nov. 2015.

[31] J. Colom-Ikuno, "System level modeling and optimization of the LTE downlink," Ph.D. dissertation, Inst. Telecommun., Technische Universität Wien, Wien, Austria, 2012.

[32] K. Brueninghaus *et al.*, "Link performance models for system level simulations of broadband radio access systems," in *Proc. IEEE 16th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, vol. 4. Sep. 2005, pp. 2306–2311.
[33] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
[34] *LTE—Evolved Universal Terrestrial Radio Access (E-UTRA) Physical Channels and Modulation—Version 11.1.0 Release 11)*, document 3GPP TS 36.211, 3GPP, 2011.
[35] M. G. Bellanger, "Specification and design of a prototype filter for filter bank based multicarrier transmission," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4. Apr. 2001, pp. 2417–2420.
[36] *LTE—Evolved Universal Terrestrial Radio Access (E-UTRA) Radio Frequency (RF) System Scenarios—Version 10.2.0 Release 10)*, document 3GPP TR 36.942, 3GPP, 2011.
[37] *LTE—Evolved Universal Terrestrial Radio Access (E-UTRA) Base Station (BS) Radio Transmission and Reception—Version 11.2.0 Release 11)*, document 3GPP TS 36.104, 3GPP, 2012.
[38] Y. G. Lim, C. B. Chae, and G. Caire, "Performance analysis of massive MIMO for cell-boundary users," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6827–6842, Dec. 2015.
[39] R. Jain, D. M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared systems," DEC Res., Palo Alto, CA, USA, Tech. Rep. TR-301, 1984.

GUILLEM FEMENIAS (SM'11) received the degree in telecommunication engineering and the Ph.D. degree in electrical engineering from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1987 and 1991, respectively. From 1987 to 1994, he was a Researcher with UPC, where he became an Associate Professor in 1992. In 1995, he joined the Department of Mathematics and Informatics, University of the Balearic Islands (UIB), Mallorca, Spain, where he became a Full Professor in 2010. He is currently leading the Mobile Communications Group at UIB, where he has been the Project Manager of projects ARAMIS, DREAMS, DARWIN, MARIMBA, COSMOS, and ELISA, all of them funded by the Spanish and Balearic Islands Governments. He was also involved with several European projects (ATDMA, CODIT, and COST). He has authored or co-authored over 90 journal and conference papers, as well as some book chapters in his research topics. His current research interests and activities span the fields of digital communications theory and wireless communication systems, with particular emphasis on cross-layer transceiver design, resource management, and scheduling strategies applied to fourth- and fifth-generation wireless networks. He has served for various IEEE conferences as a Technical Program Committee Member, as the Publications Chair at the IEEE 69th Vehicular Technology Conference (VTC-Spring 2009) and as Local Organizing Committee Member of the IEEE Statistical Signal Processing (SSP 2016). He was a recipient of the Best Paper Awards at the 2007 IFIP International Conference on Personal Wireless Communications and at the 2009 IEEE Vehicular Technology Conference - Spring.

FELIP RIERA-PALOU (SM'11) received the B.S. and M.S. degrees in computer engineering from the University of the Balearic Islands (UIB), Mallorca, Spain, in 1997, the M.Sc. and Ph.D. degrees in communication engineering from the University of Bradford, U.K., in 1998 and 2002, respectively, and the M.Sc. degree in statistics from The University of Sheffield, U.K., in 2006. From 2002 to 2005, he was with the Philips Research Laboratories, Eindhoven, The Netherlands, first as a Marie Curie Post-Doctoral Fellow (European Union) and later as a Member of Technical Staff. While at Philips, he was involved in research programs related to wideband speech/audio compression and speech enhancement for mobile telephony. From 2005 to 2009, he was a Research Associate with the Ramon y Cajal Program, Spanish Ministry of Science, Mobile Communications Group, Department of Mathematics and Informatics, UIB. Since 2010, he has been an Associate Research Professor with the I3 Program, Spanish Ministry of Education, UIB. His current research interests are in the general areas of signal processing and wireless communications.

XAVIER MESTRE (SM'09) received the M.S. and Ph.D. degrees in electrical engineering from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1997 and 2002, respectively, and the Licenciate degree in mathematics in 2011. From 1998 to 2002, he was with the UPC's Communications Signal Processing Group, where he was a Research Assistant. In 2003, he joined the Telecommunications Technological Center of Catalonia, where he currently a Senior Research Associate in the area of radio communications. During this time, he has actively participated in multiple European projects and several contracts with the local industry. He is currently the Head of the Advanced Signal and Information Processing Department. He has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, 2007–2011, 2015–present, and an Elected Member of the IEEE Sensor Array and Multichannel Signal Processing Technical Committee. During the Ph.D. degree, he received the 1998-2001 Ph.D. scholarship (granted by the Catalan Government) and the 2002 Rosina Ribalta Second Prize for the Best Doctoral Thesis Project within areas of Information Technologies and Communications by the Epson Iberica Foundation.

JUAN J. OLMOS (M'91) received the M.Sc. degree in telecommunications engineering and the Ph.D. degree from the Universitat Politècnica de Catalunya in 1983 and 1987, respectively. In 1983, he joined the Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona, where he became an Associate Professor in 1991. Since 1983, his research field has been digital radio communications, including fixed radio links and mobile communications. He has actively participated in several research projects in the frame of European programs, such as RACE, ACTS, IST, and COST, all of them focused on the evolution of mobile communication systems. He has authored or co-authored many papers in his research field and is co-author of two books on LTE. He currently belongs to the WiComTec Research Group, Escola d'Enginyeria de Telecomunicació i Aeroespacial de Castelldefels, Barcelona, where his current research interest is on 5G mobile communications.