Spectral and Energy Spectral Efficiency Optimization of Joint Transmit and Receive Beamforming Based Multi-Relay MIMO-OFDMA Cellular Networks

Kent Tsz Kan Cheung, *Student Member, IEEE*, Shaoshi Yang, *Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—We first conceive a novel transmission protocol for a multi-relay multiple-input–multiple-output orthogonal frequency-division multiple-access (MIMO-OFDMA) cellular network based on joint transmit and receive beamforming. We then address the associated network-wide spectral efficiency (SE) and energy spectral efficiency (ESE) optimization problems. More specifically, the network's MIMO channels are mathematically decomposed into several effective multiple-input–single-output (MISO) channels, which are essentially spatially multiplexed for transmission. Hence, these effective MISO channels are referred to as spatial multiplexing components (SMCs). For the sake of improving the SE/ESE performance attained, the SMCs are grouped using a pair of proposed grouping algorithms. The first is optimal in the sense that it exhaustively evaluates all the possible combinations of SMCs satisfying both the semi-orthogonality criterion and other relevant system constraints, whereas the second is a lower-complexity alternative. Corresponding to each of the two grouping algorithms, the pair of SE and ESE maximization problems are formulated, thus the optimal SMC groups and optimal power control variables can be obtained for each subcarrier block. These optimization problems are proven to be concave, and the dual decomposition approach is employed for obtaining their solutions. Relying on these optimization solutions, the impact of various system parameters on both the attainable SE and ESE is characterized. In particular, we demonstrate that under certain conditions the lower-complexity SMC grouping algorithm achieves 90% of the SE/ESE attained by the exhaustive-search based optimal grouping algorithm, while imposing as little as 3.5% of the latter scheme's computational complexity.**

***Index Terms*—Green communications, spatial multiplexing, beamforming, multi-relay, multiple-input–multiple-output orthogonal frequency-division multiple-access (MIMO-OFDMA), fractional programming, dual decomposition, cross-layer design.**

I. Introduction

RECENT wireless mobile broadband standards optionally employ relay nodes (RNs) and multiple-input–multiple-output orthogonal frequency-division multiple-access (MIMO-OFDMA) systems [1], [2] for supporting the ever-growing wireless capacity demands. These systems benefit from a capacity gain increasing roughly linearly both with the number of available OFDMA subcarriers (each having the same bandwidth) as well as with the minimum of the number of transmit antennas (TAs) and receive antennas (RAs). However, this capacity-oriented approach conflicts with the increasing need to reduce the system's carbon footprint [3] as increasing the number of radio frequency (RF) chains and subcarriers will incur additional energy costs. *In light of these discussions, the goal of this paper is to formally optimize the spectral efficiency (SE) or energy spectral efficiency (ESE) of the downlink (DL) in a multi-relay MIMO-OFDMA cellular system by intelligently allocating the available power and frequency resources and employing joint transmit and receive beamforming (BF).*

It is widely acknowledged that under the idealized simplifying condition of having perfect channel state information (CSI) at the transmitter, the DL or broadcast channel (BC) capacity [4], [5] may be approached with the aid of dirty paper coding (DPC) [6]. However, the practical implementation of DPC is hampered by its excessive algorithmic complexity upon increasing the number of users. On the other hand, BF is an attractive suboptimal strategy for allowing multiple users to share the BC while resulting in reduced multi-user interference (MUI). A low-complexity transmit-BF technique is the zero-forcing based BF (ZFBF), which can asymptotically achieve the BC capacity as the number of users tends to infinity [7]. Furthermore, ZFBF may be readily applied to a system with multiple-antenna receivers through the use of the singular value decomposition (SVD). As a result, the associated MIMO channels may be mathematically decomposed into several *effective* multiple-input–single-output (MISO) channels, which are termed spatial multiplexing components (SMCs)[1] in this work. Furthermore, in [8], these SMCs are specifically grouped so that the optimal grouping as well as the optimal allocation of the power

Manuscript received November 1, 2013; revised March 16, 2014 and July 29, 2014; accepted August 9, 2014. This research has been funded by the Industrial Companies who are Members of the Mobile VCE, with additional financial support from the UK Government's Engineering & Physical Sciences Research Council (EPSRC). The financial support of the Research Councils UK (RCUK) under the India-UK Advanced Technology Center (IU-ATC), of the EU under the auspices of the Concerto project, and of the European Research Council's Senior Research Fellow Grant is also gratefully acknowledged. The associate editor coordinating the review of this paper and approving it for publication was L. K. Rasmussen.

The authors are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: ktkc106@ecs.soton.ac.uk; shaoshi.yang@ecs.soton.ac.uk; lh@ecs.soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2014.2348996

[1]Note that these effective MISO channels are different from the physical MISO channels directly composing the physical MIMO channel. For brevity, we coin the term SMC to emphasize that these effective MISOs will be used for the purpose of spatial multiplexing. A more in-depth discussion regarding the concept of SMCs will be provided in Section III.

may be found on each subcarrier block using convex optimization. In contrast to the channel-diagonalization methods of [9], [10], [11], the ZFBF approach does not enforce any specific relationship between the total numbers of TAs and RAs. Therefore, ZFBF is more suitable for practical systems, since the number of TAs at the BS is typically much lower than the total number of RAs of all the active user equipments (UEs). Compared to the random beamforming methods, such as that of [12], ZFBF is capable of completely avoiding the interference, allowing us to formulate our SE/ESE maximization (SEM/ESEM) problems as convex optimization problems. Due to its desirable performance versus complexity trade-off, in this paper we employ ZFBF in the context of multi-relay aided MIMO-OFDMA systems, where the direct link between the base station (BS) and the UE may be exploited in conjunction with the relaying link for further improving the system's performance.

We formally define the ESE as a counterpart of the area spectral efficiency (ASE) [13], where the latter has the units of [bits/sec/Hz/m^2], while the former is measured in [bits/sec/Hz/Joule]. The ESE metric has been justified, for example, in [14]–[17]. However, these contributions did not consider resource allocation in the context of a MIMO system, and only [17] incorporated relaying. On the other hand, although there are numerous contributions on optimal resource allocation in MIMO systems, they typically only focused on either the SEM (equivalently, the sum-rate maximization) or the power minimization [8], [18]–[21]. For example, the authors of [8] applied BF to a DL cellular system and aimed for minimizing the resultant total transmission power, while simultaneously satisfying the per-user rate requirements. The authors of [19] instead choose to minimize the per-antenna transmission powers, while satisfying both the maximum per-antenna power constraints as well as the per-user signal-to-noise-plus-interference (SINR) requirements. Although there exists some literature studying the ESE of relay-aided MIMO systems [22], [23], these contributions typically focus their attention on a simple three-node network consisting of the source, the destination and a single RN.

To summarize, there is a paucity of literature on the convex optimization approach to the ESEM problem associated with both resource allocation and joint transmit/receive beamforming in the context of multi-user multi-relay MIMO-OFDMA systems. Additionally, the Charnes-Cooper transformation [24] is employed in this paper for solving the associated ESEM problem, in contrast to the scalarization approach [25] that requires the weighting of multiple objectives. On the other hand, the Dinkelbach's method [14], [17], [26], [27] is avoided as it would require solving a series of parametric convex problems, rather than the resultant single convex problem of the Charnes-Cooper transformation. Although the latter approach does impose an additional linear constraint on the problem, in our experience, this only marginally increases the complexity of the solution algorithm. The authors of [28] employed the Charnes-Cooper variable transformation for the ESEM of a simple point-to-point link. However, as far as we are aware, the Charnes-Cooper transformation has rarely been used in other contexts for solving the ESEM problem.

Let us now summarize the above discussions and provide a concise list of the novel contributions of this paper:

- **We consider a generalized multi-user multi-relay assisted MIMO-OFDMA cellular system model for the SEM/ESEM problems**. To provide some justification, this system model accounts for both the direct links between the BS and the UEs, as well as the relaying links employing the decode-and-forward (DF) relaying protocol [29]. This system model is unlike that of [7], [8], which did not consider relaying, and it is also distinct from that of [22], [23], which only consider a single RN and a single UE. Additionally, we dispense with the constraint that the number of antennas at the BS needs to be greater or equal to the sum of the number of antennas at the UEs, which was assumed in [9]–[11]. Furthermore, this system model is built upon our previous work [17] as the network elements may now be equipped with an arbitrary number of antennas for improving the system's SE or ESE performance.
- **A sophisticated novel transmission protocol is proposed for improving the system's SE/ESE performance**. Since the multi-relay MIMO-OFDMA system model considered has not been studied in the context of the SEM/ESEM problems before, we develop a novel transmission protocol that exploits spatial multiplexing in both transmission phases while allowing both the direct and relaying links to be simultaneously active. Although this protocol does not benefit from a higher spatial degree of freedom than that of the conventional half-duplex relay based cooperative system, we glean more flexibility in choosing the best group of channels for each transmission phase, which leads to additional selection diversity. As a result, the achievable SE/ESE performance may be improved. Again, this protocol is distinct from that presented in [7], [8], since relaying is not considered in those works. Another benefit is that since spatial multiplexing is employed in conjunction with OFDMA, multiple data streams may be served using *the same* subcarrier block, while the transmit ZFBF is employed for avoiding the interference. Furthermore, the receive-BF matrices are designed with the aim of generating a number of SMCs that may be grouped for the purpose of increasing the attainable spatial multiplexing gain.
- **Two SMC grouping algorithms are proposed**. To elaborate, we present a pair of novel algorithms for grouping the SMCs transmissions. The possibility of relayed transmissions means that we have to partition each transmission period into two halves, one consisting of BS-to-UE and BS-to-RN links, and the other consisting of additional BS-to-UE as well as RN-to-UE links. As a result, the SMC-pairs of the two-hop relaying links are incomparable to the SMCs of the direct links in either the first or the second transmission phases. This is because, firstly the RNs are subject to their individual maximum transmission power constraints, and secondly they employ the DF protocol, which means that the information conveyed on the RN-to-UE link cannot be more than that conveyed on the BS-to-RN link. These challenging issues are resolved

by the proposed grouping algorithms. The first grouping algorithm is optimal in the sense that it is based on exhaustive search over all the SMC groupings that satisfy the semi-orthogonality criterion, while the second algorithm constitutes a lower-complexity alternative. This complexity-reduction is achieved by a multi-stage SMC group construction process. In each stage, we firstly compute the orthogonal components with respect to the vectors contained in the tentative SMC group to be constructed using all the residual legitimate SMC vectors, and then insert the particular SMC vector that results in the orthogonal component having the highest norm into the SMC group to be constructed. In principle, this method is similar to that of [7], [8], but it has been appropriately adapted for the multi-relay cellular network considered under the above-mentioned particular constraints.

- **The problems of choosing the SE- or ESE-optimal SMC groupings and their associated power control values are formulated and solved using convex optimization**. In contrast to [8], [18]–[21], the crucial objective of maximizing the ESE metric is employed, as motivated above. On the other hand, in contrast to [14]–[16], we consider a system that allows for *simultaneous* direct and relayed transmissions for the sake of increasing the attainable spatial multiplexing gain. Although there exist other methods of solving this ESEM problem [14], [17], [25], [27], we employ the Charnes-Cooper transformation [24] for obtaining the maximum ESE solution, as it exhibits a reduced complexity from having to solve only a single convex optimization problem.

The rest of this paper is organized as follows. Section II describes the multi-relay MIMO-OFDMA cellular network considered, while Section III characterizes our novel transmission protocol that allows for both direct and relaying links to be simultaneously and continuously activated. In Section IV, we elaborate on the aforementioned SMC grouping algorithms conceived for forming the sets of possible SMC transmission groups. The issue of finding the optimal SMC transmission groups and the optimal power control variables is then formulated as an optimization problem in Section V, which is then solved by using a number of variable transformations and relaxations. The performance of both our SMC grouping algorithms and of the SEM/ESEM solvers are characterized in Section VI. Finally, we present our conclusions and future research ideas in Section VII.

II. System Model

We focus our attention on the DL of a multi-relay MIMO-OFDMA cellular network, as shown in Fig. 1. The BS, M DF-assisted RNs and K UEs are each equipped with N_B, N_R and N_U antennas, respectively. The cellular system has access to N subcarrier blocks, each encompassing W Hertz of wireless bandwidth. The subcarrier blocks considered here are similar to the resource blocks in the LTE-nomenclature [30]. The BS is located at the cell-center, while the RNs are each located at a fixed distance from the BS and are evenly spaced around it. The

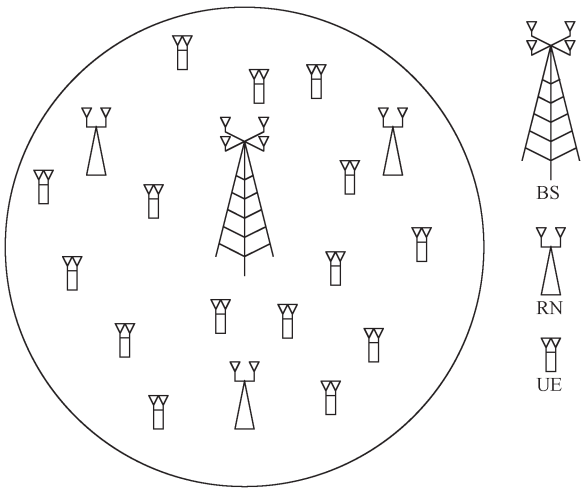

Fig. 1. An example of a multi-relay MIMO-OFDMA cellular network, containing a BS at the cell-center, 3 RNs and 15 UEs.

ratio of the distance between the BS and RNs to the cell radius is denoted by D_r. On the other hand, the UEs are uniformly distributed in the cell. The BS coordinates and synchronizes its own transmissions with that of the RNs, which employ the DF protocol and thus avoids the problem of noise amplification. As it will be shown in Section V-C-1, this strategy results in a simple algorithm for finding the optimal power control variables.

For the subcarrier block $n \in \{1, \ldots, N\}$, let us define the complex-valued wireless channel matrices between the BS and UE $k \in \{1, \ldots, K\}$, between the BS and RN $m \in \{1, \ldots, M\}$, and between RN m and UE k as $\mathbf{H}_{n,k}^{BU} \in \mathbb{C}^{N_U \times N_B}$, $\mathbf{H}_{n,m}^{BR} \in \mathbb{C}^{N_R \times N_B}$ and $\mathbf{H}_{n,m,k}^{RU} \in \mathbb{C}^{N_U \times N_R}$, respectively. These complex-valued channel matrices account for both the frequency-flat Rayleigh fading and the path-loss between the corresponding transceivers. The coherence bandwidth of each wireless link is assumed to be sufficiently high, so that each individual subcarrier block experiences frequency flat fading, although the level of fading may vary from one subcarrier block to another in each transmission period. Additionally, the transceivers are stationary or moving slowly enough so that the level of fading may be assumed to be fixed for the duration of a scheduled transmission period. Furthermore, the RAs are spaced sufficiently far apart, so that each TA/RA pair experiences independent and identically distributed (i.i.d.) fading. Since these channels are slowly varying, the system is capable of exploiting the benefits of channel reciprocity associated with time-division duplexing (TDD), so that the CSI becomes available at each BS- and RN-transmitter and at each possible RN- and UE-receiver. To elaborate, $\mathbf{H}_{n,k}^{BU}$ and $\mathbf{H}_{n,m}^{BR}$ are known at the BS, $\mathbf{H}_{n,m}^{BR}$ and $\mathbf{H}_{n,m,k}^{RU}$ are known at the RN m, while $\mathbf{H}_{n,k}^{BU}$ and $\mathbf{H}_{n,m,k}^{RU}$ are also known at UE k. Additionally, through the use of dedicated low-rate error-free feedback channels, $\mathbf{H}_{n,m,k}^{RU}$ is also assumed to be known at the BS so that the BS may perform network-wide scheduling.[2] These channel matrices are assumed to have

[2]In this paper, since our focus is on the resource allocation and the associated SE/ESE optimization problems, the idealized simplifying assumption of the availability of perfect CSI is employed. At the current stage, accounting for erroneous CSI using, for example, robust optimization [31] is beyond the scope of this paper and may be addressed in our future work.

full row rank, which may be achieved with a high probability for typical DL wireless channel matrices.

Furthermore, each receiver suffers from additive white Gaussian noise (AWGN) having a power spectral density of N_0. The maximum instantaneous transmission power available to the BS and to each RN due to regulatory and health-constraints is P_{max}^{B} and P_{max}^{R}, respectively. Since OFDMA modulation constitutes a linear operation, we focus our attention on a single subcarrier block and as usual, we employ the commonly-used equivalent baseband signal model.[3]

III. Transmission Protocol Design

The system can simultaneously use two transmission modes to convey information to the UEs, namely the BS-to-UE mode, and the relaying-based BS-to-RN and RN-to-UE mode. Note that although in classic OFDMA each data stream is orthogonal in frequency, for the sake of further improving the system's attainable SE or ESE performance, our system employs spatial multiplexing in conjunction with ZFBF *so that multiple data streams may be served using the same subcarrier block, without suffering from interference*. Additionally, since the relaying-based transmission can be split into two phases, the design philosophy of the BF matrices in each phase are described separately, although for simplicity we have assumed that the respective channel matrices remain unchanged in both phases. Firstly, the definition of the semi-orthogonality criterion is given as follows [7].

Definition 1: A pair of MISO channels, represented by the complex-valued column vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, are said to be semi-orthogonal to each other with parameter $\alpha \in [0, 1]$, when

$$\frac{\left|\Re\left(\mathbf{v}_1^{\mathrm{H}}\mathbf{v}_2\right)\right|}{\|\mathbf{v}_1\|\|\mathbf{v}_2\|} \leq \alpha. \tag{1}$$

To be more specific, a measure of the grade of orthogonality between $\mathbf{v}_1$ and $\mathbf{v}_2$ is given by the left-hand side of inequality (1), which ranges from 0 for orthogonal vectors to 1 for linearly dependent vectors.

The authors of [7] demonstrated that employing the ZFBF strategy for MISO channels that satisfy $\alpha \to 0$, while the number of users obeys $K \to \infty$, asymptotically achieves the DPC capacity, and it is therefore optimal for the BC channel. Similar principles are followed when maximizing the SE or ESE of the system considered in this paper.

A. BF Design for the First Transmission Phase

In the first transmission phase, only the BS is transmitting, while both the RNs and the UEs act as receivers. This is similar to the classic DL multi-user MIMO model. As described above, our aim is 1) to design a ZFBF matrix for the BS to avoid interference between data streams, and 2) to design receive BF matrices for the UEs and RNs so that the resultant *effective DL channel matrices* contain as many semi-orthogonal rows as possible that satisfy (1) for a given α. Ideally, all receivers (UEs and RNs) should jointly compute[4] their receive BF matrices to accomplish the second goal. However, this is generally impossible, since we cannot realistically assume that the channel matrices associated with each UE and RN are shared among them, due to the geographically-distributed nature of the UEs and RNs. As a compromise, we opt for guaranteeing that each individual effective DL channel matrix contains locally orthogonal rows by employing the SVD [7], [8]. Although these locally orthogonal rows may not remain orthogonal globally, they can be characterized using the semi-orthogonality metric of (1).

Bearing this in mind, the channel matrices of all DL transmissions originating from the BS are decomposed at the BS, UEs and RNs using the SVD [32] as $\mathbf{H}_{n,k}^{BU} = \mathbf{U}_{n,k}^{BU}\mathbf{S}_{n,k}^{BU}(\mathbf{V}_{n,k}^{BU})^{\mathrm{H}}$ and $\mathbf{H}_{n,m}^{BR} = \mathbf{U}_{n,m}^{BR}\mathbf{S}_{n,m}^{BR}(\mathbf{V}_{n,m}^{BR})^{\mathrm{H}}$, respectively. Thus, the receive-BF matrices for UE k and RN m are given by $\mathbf{R}_{n,k}^{BU,T_1} = (\mathbf{U}_{n,k}^{BU})^{\mathrm{H}} \in \mathbb{C}^{N_U \times N_U}$ and $\mathbf{R}_{n,m}^{BR,T_1} = (\mathbf{U}_{n,m}^{BR})^{\mathrm{H}} \in \mathbb{C}^{N_R \times N_R}$, and the effective DL channel matrices are then given[5] by $\underline{\mathbf{H}}_{n,k}^{BU,T_1} = \mathbf{R}_{n,k}^{BU,T_1}\mathbf{H}_{n,k}^{BU} = \mathbf{S}_{n,k}^{BU}(\mathbf{V}_{n,k}^{BU})^{\mathrm{H}} \in \mathbb{C}^{N_U \times N_B}$ and $\underline{\mathbf{H}}_{n,m}^{BR,T_1} = \mathbf{R}_{n,m}^{BR,T_1}\mathbf{H}_{n,m}^{BR} = \mathbf{S}_{n,m}^{BR}(\mathbf{V}_{n,m}^{BR})^{\mathrm{H}} \in \mathbb{C}^{N_R \times N_B}$, respectively. Since $\mathbf{V}_{n,k}^{BU}$ and $\mathbf{V}_{n,m}^{BR}$ are both unitary, while $\mathbf{S}_{n,k}^{BU}$ and $\mathbf{S}_{n,m}^{BR}$ are both real and diagonal, these effective DL channel matrices respectively consist of $\min(N_B, N_U)$ and $\min(N_B, N_R)$ orthogonal non-zero rows[6] with norms equal to their corresponding singular values. We refer to these non-zero orthogonal rows as the SMCs of their associated MIMO channel matrix.[7] The K BS-to-UE MIMO channel matrices and M BS-to-RN channel matrices generate a total of $[K \cdot \min(N_B, N_U) + M \cdot \min(N_B, N_R)]$ SMCs. Since these SMCs are generated from independent MIMO channel matrices associated with geographically distributed UEs and RNs, they are not all guaranteed to be orthogonal to each other. Furthermore, since each UE or RN has multiple antennas and N_B might not be sufficiently large to simultaneously support all UEs and RNs, we have to determine which specific SMCs should be served. As a result, for each two-phase transmission period, we opt for selecting a SMC group accounting for both phases from the set of available SMC groups. This selection process is achieved by jointly using the SMC grouping algorithm and solving the optimization problem detailed below. For the sake of clarity, the concepts of the SMC, of the SMC group and of the set of SMC groups are illustrated in Fig. 2.

To elaborate a little further, a set of SMC groups, $\mathcal{G}_n$, which is associated with subcarrier block n, may be obtained

[3] Since the specific signal model expressions of each link is dependent on the transmission protocol to be designed, they are not presented here but instead detailed in Section III.

[4] The joint computation is required only for attaining the highest number of semi-orthogonal rows globally.

[5] Note that T_1 is used for indicating the first transmission phase, and underline is used to denote the effective DL channel matrices.

[6] The reason why we use $\min(N_B, N_U)$ and $\min(N_B, N_R)$, instead of N_U and N_R, is so that the antenna configuration $N_B \leq N_U$ and/or $N_B \leq N_R$ is also covered.

[7] Note that only when $N_B \geq N_U$ and $N_B \geq N_R$, a single SMC is generated for each receive antenna.

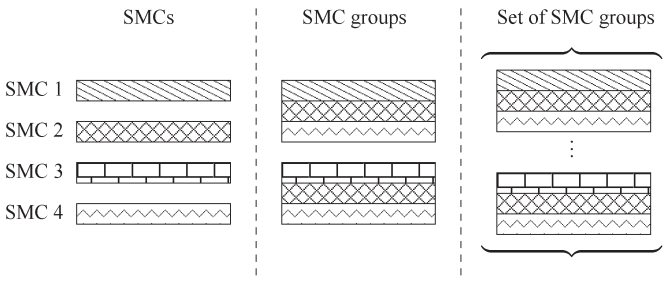

Fig. 2. A conceptual illustration of the differences between SMCs, SMC groups and a set of SMC groups.

using one of the grouping algorithms presented in Section IV. The BS selects a single group, $j \in \mathcal{G}_n$, containing (but not limited to)[8] $Q_j^{T_1}$ SMCs out of the $[K \cdot \min(N_B, N_U) + M \cdot \min(N_B, N_R)]$ available SMCs to be supported by using ZFBF. Thus, we have $Q_j^{T_1} \leq \min[N_B, K \cdot \min(N_B, N_U) + M \cdot \min(N_B, N_R)]$ and a multiplexing gain of $Q_j^{T_1}$ is achieved. Let us denote the refined effective DL channel matrix with rows being the $Q_j^{T_1}$ selected SMCs as $\underline{\mathbf{H}}_{n,j}^{T_1} \in \mathbb{C}^{Q_j^{T_1} \times N_B}$. The ZFBF transmit matrix applied at the BS to subcarrier block n is then given by the following right inverse $\mathbf{T}_{n,j}^{T_1} = (\underline{\mathbf{H}}_{n,j}^{T_1})^{\mathrm{H}} \cdot [\underline{\mathbf{H}}_{n,j}^{T_1} (\underline{\mathbf{H}}_{n,j}^{T_1})^{\mathrm{H}}]^{-1} \in \mathbb{C}^{N_B \times Q_j^{T_1}}$. Since $\underline{\mathbf{H}}_{n,j}^{T_1} \mathbf{T}_{n,j}^{T_1} = \mathbf{I}_{N_B}$, the potential interference between the $Q_j^{T_1}$ selected SMCs is completely avoided. Furthermore, the columns of $\mathbf{T}_{n,j}^{T_1}$ are normalized by multiplying the diagonal matrix $\mathbf{W}_{n,j}^{T_1}$ on the right-hand side of $\mathbf{T}_{n,j}^{T_1}$ to ensure that each SMC transmission is initially set to unit power.[9]

Then, $\mathbf{T}_{n,j}^{T_1} \mathbf{W}_{n,j}^{T_1}$ is used as the DL transmit-BF matrix for the BS in the first phase. Thus, the effective channel-to-noise ratios (CNRs) in the first transmission phase can be written as $G_{n,j,e_1}^{BU,T_1} = |w_{n,j,e_1}^{BU,T_1}|^2 / \Delta\gamma N_0 W$ and $G_{n,j,e}^{BR,T_1} = |w_{n,j,e}^{BR,T_1}|^2 / \Delta\gamma N_0 W$, respectively, where w_{n,j,e_1}^{BU,T_1} and $w_{n,j,e}^{BR,T_1}$ are the diagonal elements in $\mathbf{W}_{n,j}^{T_1}$. More specifically, these diagonal elements correspond to SMC group j and subcarrier block n, and they are associated with either a direct BS-to-UE SMC or a BS-to-RN SMC. The additional subscripts $e_1 \in \{0, \ldots, \min[N_B, K \cdot \min(N_B, N_U)]\}$ and $e \in \{0, \ldots, \min[N_B, M \cdot \min(N_B, N_R), K \cdot \min(N_R, N_U)]\}$ are used for distinguishing the multiple selected SMCs of the direct links (i.e. those related to UEs), from the multiple selected SMC-pairs[10] that may be associated with a particular RN $\mathcal{M}(e)$, respectively. Note that $\mathcal{M}(e)$ is a function of e, representing the RN index (similar to m used before) associated with the SMC-pair e, as further detailed in Section IV.

[8]The SMC group selection, as a part of the scheduling operation, is carried out at the BS before initiating the first transmission phase. Hence, the selected SMC group will also contain $Q_j^{T_2}$ SMCs selected by the BS for the second transmission phase, as detailed in Section III-B.

[9]Each diagonal element of $\mathbf{W}_{n,j}^{T_1}$ is equal to the reciprocal of the norm of the column vector to be normalized.

[10]A single SMC-pair consists of a SMC for the first phase and another for the second phase. Although these SMCs are generated separately in each phase, the SMC-pair associated with a common RN has to be considered as a single entity in the SMC grouping algorithms presented in Section IV.

At a given bit-error rate (BER) requirement, $\Delta\gamma$ is the signal-to-noise ratio (SNR) gap between the lower-bound SNR required for achieving the discrete-input–continuous-output memoryless channel (DCMC) capacity and the actual higher SNR required by the modulation/coding schemes of the practical physical layer transceivers employed. For example, making the simplifying assumption that idealized transceivers capable of achieving exactly the DCMC capacity are employed, then $\Delta\gamma = 0$ dB. Although, strictly speaking, so far it is not possible to operate exactly at the DCMC channel capacity, there does exist several physical layer transceiver designs that operate very close to it [33]. Furthermore, the noise power received on each subcarrier block is given by $N_0 W$.

B. BF Design in the Second Transmission Phase

The second transmission phase may be characterized by the *MIMO interference channel*. A similar methodology is employed in the second transmission phase, except that now both the BS and the RNs are transmitters, while a number of UEs are receiving. In this phase, our aim is 1) to design ZFBF matrices for the BS and RNs to avoid interference between data streams, 2) and to design a receive-BF matrix for each UE so that the effective channel matrices associated with each of its transmitters contain rows which satisfy the semi-orthogonal condition (1) for a given α. This means that more data streams may be served simultaneously, thus improving the attainable SE or ESE performance. Since there are multiple *distributed* transmitters/MIMO channel matrices associated with each UE, the SVD method described in Section III-A, which is performed in a centralized fashion, cannot be readily applied at the transmitter side. Instead, we aim for minimizing the resultant correlation between the generated SMCs, thus increasing the number of SMCs which satisfy (1) for a given α. To accomplish this goal, we begin by introducing the shorthand of $\underline{\mathbf{H}}_{n,k}^{BU,T_2} = \mathbf{R}_{n,k}^{U,T_2} \mathbf{H}_{n,k}^{BU} \in \mathbb{C}^{N_U \times N_B}$ and $\underline{\mathbf{H}}_{n,m,k}^{RU,T_2} = \mathbf{R}_{n,k}^{U,T_2} \mathbf{H}_{n,m,k}^{RU} \in \mathbb{C}^{N_U \times N_R}$ as the effective channel matrices between the BS and UE k, and between RN m and UE k, respectively, on subcarrier block n in the second transmission phase, where $\mathbf{R}_{n,k}^{U,T_2} \in \mathbb{C}^{N_U \times N_U}$ is the yet-to-be-determined UE k's receive-BF matrix. In light of the preceding discussions, one of our aims is to design $\mathbf{R}_{n,k}^{U,T_2}$ so that the off-diagonal values of the matrices given by $\mathbf{A}_0 = \underline{\mathbf{H}}_{n,k}^{BU,T_2} (\underline{\mathbf{H}}_{n,k}^{BU,T_2})^{\mathrm{H}} = \mathbf{R}_{n,k}^{U,T_2} \mathbf{H}_{n,k}^{BU} (\mathbf{H}_{n,k}^{BU})^{\mathrm{H}} (\mathbf{R}_{n,k}^{U,T_2})^{\mathrm{H}}$ and $\mathbf{A}_m = \underline{\mathbf{H}}_{n,m,k}^{RU,T_2} (\underline{\mathbf{H}}_{n,m,k}^{RU,T_2})^{\mathrm{H}} = \mathbf{R}_{n,k}^{U,T_2} \mathbf{H}_{n,m,k}^{RU} (\mathbf{H}_{n,m,k}^{RU})^{\mathrm{H}} (\mathbf{R}_{n,k}^{U,T_2})^{\mathrm{H}}$, $\forall m$ are as small as possible. This design goal may be formalized as

$$\min_{\mathbf{R}_{n,k}^{U,T_2}} \left\| \mathbf{H}_{n,k}^{BU} (\mathbf{H}_{n,k}^{BU})^{\mathrm{H}} - \left(\mathbf{R}_{n,k}^{U,T_2}\right)^{-1} \mathbf{\Lambda}_0 \left(\mathbf{R}_{n,k}^{U,T_2}\right)^{-\mathrm{H}} \right\|_{\mathrm{F}}^2 + \sum_{m=1}^{M} \left\| \mathbf{H}_{n,m,k}^{RU} (\mathbf{H}_{n,m,k}^{RU})^{\mathrm{H}} - \left(\mathbf{R}_{n,k}^{U,T_2}\right)^{-1} \mathbf{\Lambda}_m \left(\mathbf{R}_{n,k}^{U,T_2}\right)^{-\mathrm{H}} \right\|_{\mathrm{F}}^2, \quad (2)$$

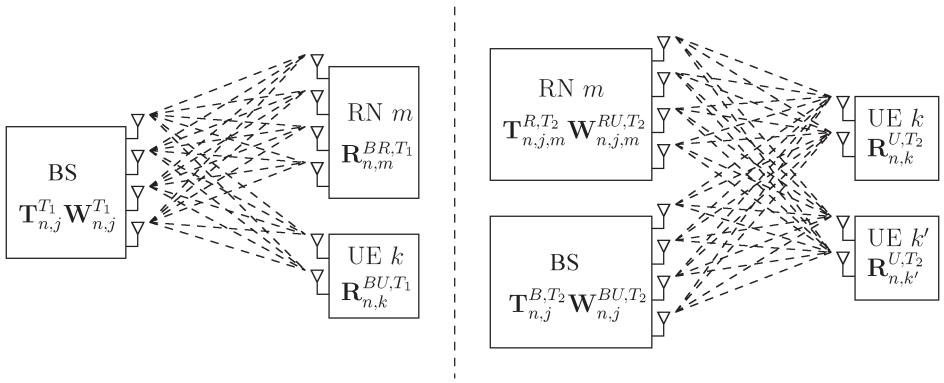

Fig. 3. A conceptual schematic of the transmit- and receive-BF matrices employed in the first and second transmission phases. In the first phase, the BS employs its ZFBF matrix to serve multiple data streams without interference. In the second phase, the BS and RNs employ separate ZFBF matrices to distributively avoid interference between the data streams being served.

where $\mathbf{\Lambda}_0$ and $\mathbf{\Lambda}_m$ are diagonal matrices containing the diagonal elements of $\mathbf{A}_0$ and $\mathbf{A}_m$, respectively. Therefore, $(\mathbf{R}_{n,k}^{U,T_2})^{-1}$ is the *jointly diagonalizing matrix* [34], while $\mathbf{H}_{n,k}^{BU}(\mathbf{H}_{n,k}^{BU})^{\mathrm{H}}$ and $\mathbf{H}_{n,m,k}^{RU}(\mathbf{H}_{n,m,k}^{RU})^{\mathrm{H}}$, $\forall m$ are the matrices to be diagonalized. Thus, the algorithm presented in [34] for solving[11] (2) may be invoked at UE k for obtaining $\mathbf{R}_{n,k}^{U,T_2}$, which may be further fed back to the BS and RNs. Hence, the BS and RNs do not have to share $\mathbf{H}_{n,k}^{BU}$ or $\mathbf{H}_{n,k}^{RU}$ via the wireless channel and do not have to solve (2) again. As a result, we accomplish the goal of creating effective channel matrices that contain rows aiming to satisfy (1). Additionally, the columns of $\mathbf{R}_{n,k}^{U,T_2}$ have been normalized so that the power assigned for each SMC remains unaffected.

After obtaining the receive-BF matrix, the SMCs of the transmissions to UE k on subcarrier block n are given by the non-zero rows of the effective channel matrices $\underline{\mathbf{H}}_{n,k}^{BU,T_2}$ and $\underline{\mathbf{H}}_{n,m,k}^{RU,T_2}$, $\forall m$. Since the BS and the RNs act as distributed broadcasters in the second phase, they are only capable of employing *separate* ZFBF transmit matrices to ensure that none of them imposes interference on the SMCs it does not explicitly intend to serve. By employing one of the grouping algorithms described in Section IV, the BS schedules[12] $Q_j^{T_2} \leq \min[\min(N_B, N_R), \sum_{i=1}^{K} L_i^B + L_i^R]$ SMCs to serve simultaneously in the second phase, where L_i^B and L_i^R represent the number of SMCs of UE i served by the BS and by RNs in this phase, respectively, where we have $L_i^B + L_i^R \leq N_U$, $L_i^B \leq \min(N_B, N_U)$, and $L_i^R \leq \min(N_R, N_U)$. Note that since UE i may be simultaneously served both by the BS and by a RN (each of them serves a fraction of UE i's SMCs), it is *possible* that the summation of the respective number of UEs served[13] by the BS and by RNs may be higher than K. Let us denote the *refined* effective DL channel matrices, from the perspectives of the BS and RN m, consisting of the $Q_j^{T_2}$ selected SMCs as $\underline{\mathbf{H}}_{n,j}^{B,T_2} \in \mathbb{C}^{Q_j^{T_2} \times N_B}$ and $\underline{\mathbf{H}}_{n,j,m}^{R,T_2} \in \mathbb{C}^{Q_j^{T_2} \times N_R}$, respectively. Since these are known to each transmitter, they may employ ZFBF transmit matrices in the second phase, given by the right inverses $\mathbf{T}_{n,j}^{B,T_2} = (\underline{\mathbf{H}}_{n,j}^{B,T_2})^{\mathrm{H}} \cdot [\underline{\mathbf{H}}_{n,j}^{B,T_2} (\underline{\mathbf{H}}_{n,j}^{B,T_2})^{\mathrm{H}}]^{-1} \in \mathbb{C}^{N_B \times Q_j^{T_2}}$ for the BS, and $\mathbf{T}_{n,j,m}^{R,T_2} = (\underline{\mathbf{H}}_{n,j,m}^{R,T_2})^{\mathrm{H}} \cdot [\underline{\mathbf{H}}_{n,j,m}^{R,T_2} (\underline{\mathbf{H}}_{n,j,m}^{R,T_2})^{\mathrm{H}}]^{-1} \in \mathbb{C}^{N_R \times Q_j^{T_2}}$ for RN m. Similar to the first transmission phase, these ZFBF transmit matrices are normalized by $\mathbf{W}_{n,j}^{BU,T_2}$ and $\mathbf{W}_{n,j,m}^{RU,T_2}$, respectively, to ensure that each SMC transmission is initially set to unit power. Upon obtaining the selected SMCs, we denote the effective CNRs in the second transmission phase as $G_{n,j,e_2}^{BU,T_2} = |w_{n,j,e_2}^{BU,T_2}|^2 / \Delta\gamma N_0 W$ and $G_{n,j,e}^{RU,T_2} = |w_{n,j,e}^{RU,T_2}|^2 / \Delta\gamma N_0 W$, where w_{n,j,e_2}^{BU,T_2} and $w_{n,j,e}^{RU,T_2}$ are the diagonal elements in $\mathbf{W}_{n,j}^{BU,T_2}$ and $\mathbf{W}_{n,j,\mathcal{M}(e)}^{RU,T_2}$, respectively, and the subscript $\mathcal{M}(e)$ has been defined in Section III-A. To elaborate, for a second-phase BS-to-UE link, w_{n,j,e_2}^{BU,T_2} corresponds to SMC group j and subcarrier block n, while the subscript $e_2 \in \{0, \ldots, \min[N_B, K \cdot \min(N_B, N_U)]\}$ is employed for further distinguishing the multiple selected SMCs associated with UEs from the BS. Similarly, $w_{n,j,e}^{RU,T_2}$, which also corresponds to SMC group j and subcarrier block n, is associated with the second-phase RN-to-UE link between RN $\mathcal{M}(e)$ and the particular UE of SMC-pair e.

For more explicit clarity, a schematic of the transmit and receive beamforming matrices in the first and second transmission phases is presented in Fig. 3.

C. Achievable Spectral Efficiency and Energy Spectral Efficiency

For the sake of convenience, let us first denote the transmit power allocation policy as $\mathcal{P}$, which is a set composed by all transmit power control variables invoked at the BS and/or RNs

[11] In fact, when there are only two matrices to diagonalize, say $\mathbf{A}_0$ and $\mathbf{A}_1$, the diagonalizing matrix may be obtained from the eigenvectors of $\mathbf{A}_0(\mathbf{A}_1)^{-1}$ [35]. This diagonalizing matrix is able to fully diagonalize both $\mathbf{A}_0$ and $\mathbf{A}_1$.

[12] To elaborate a little further, when computing its ZFBF transmit matrix, each transmitter (either the BS or a RN) must take into account an auxiliary SMC, which is also selected from the legitimate SMC candidates and is required for nulling the interference that this particular transmitter imposes on each selected information-bearing SMC of the other transmitters. Furthermore, each auxiliary SMC is employed by its corresponding transmitter to transmit several additional zeros that are padded to the normal data symbols. As a beneficial result, no interference is received at each UE from the transmitter that does not serve this particular UE. For more details of the SMC-based transmission in the second phase, please refer to Algorithm 1 described in Section IV-A.

[13] If at least one SMC of a UE is served by the BS (or a RN), we say that this UE is served by the BS (or the RN).

in both transmission phases. Since receive-BF is employed in conjunction with ZFBF, each SMC transmission may be viewed as a single-input–single-output (SISO) link. Therefore, on the direct links, the receiver's SNR at UE k corresponding to SMCs e_1 and e_2 may be expressed as $\Gamma_{n,j,e_1}^{BU,T_1}(\mathcal{P}) = G_{n,j,e_1}^{BU,T_1} P_{n,j,e_1}^{BU,T_1}$ and $\Gamma_{n,j,e_2}^{BU,T_2}(\mathcal{P}) = G_{n,j,e_2}^{BU,T_2} P_{n,j,e_2}^{BU,T_2}$ for the first and second transmission phases, respectively. The scalar variables P_{n,j,e_1}^{BU,T_1} and P_{n,j,e_2}^{BU,T_2}, which are elements of $\mathcal{P}$, determine the transmit power values for SMCs e_1 and e_2 on the direct links. As a result, the achievable instantaneous SE of the direct links is given by $C_{n,j,e_1}^{BU,T_1}(\mathcal{P}) = (1/2)\log_2(1+\Gamma_{n,j,e_1}^{BU,T_1}(\mathcal{P}))$ and $C_{n,j,e_2}^{BU,T_2}(\mathcal{P}) = (1/2)\log_2(1+\Gamma_{n,j,e_2}^{BU,T_2}(\mathcal{P}))$, which are normalized both by time and by frequency to give units of [bits/sec/Hz]. The factor of 1/2 accounts for the fact that the transmission period is split into two phases.

Similarly, for the SMC-pair e of the DF relaying links, the SNR at RN $\mathcal{M}(e)$ in the first transmission phase is given by $\Gamma_{n,j,e}^{BR,T_1}(\mathcal{P}) = G_{n,j,e}^{BR,T_1} P_{n,j,e}^{BR,T_1}$ and the SNR at UE k in the second transmission phase is formulated as $\Gamma_{n,j,e}^{RU,T_2}(\mathcal{P}) = G_{n,j,e}^{RU,T_2} P_{n,j,e}^{RU,T_2}$. Additionally, $P_{n,j,e}^{BR,T_1}$ and $P_{n,j,e}^{RU,T_2}$ are also elements of $\mathcal{P}$. Since the RNs employ the DF protocol, the achievable SE is limited by the weaker of the two RN-related links [29] and is given by $C_{n,j,e}^{BRU}(\mathcal{P}) = \min[(1/2)\log_2(1+\Gamma_{n,j,e}^{BR,T_1}(\mathcal{P})),(1/2)\log_2(1+\Gamma_{n,j,e}^{RU,T_2}(\mathcal{P}))]$.

Let us now introduce the SMC group selection variable $s_{n,j} \in \{0,1\}$, which indicates that SMC group j, as introduced in Sections III-A and III-B, is selected for subcarrier block n, when $s_{n,j} = 1$, and $s_{n,j} = 0$ otherwise. All SMC group selection variables are scalars and are collected into a set denoted by $\mathcal{S}$. Once again, we emphasize that $\mathcal{G}_n$ denotes the set of possible SMC groups for subcarrier block n. Thus, the total achieved SE is given by (3) (see equation at the bottom of the page), where $\mathcal{E}_{n,j}$ is the set of SMCs in the selected group j on subcarrier block n.

In this work, we adopt the energy dissipation model presented in [36], where the total energy dissipation of the system is assumed to be dependent on several factors, including the number of TAs, the energy dissipation of the RF and baseband circuits, and the efficiencies of the power amplifier, feeder cables, cooling system, mains power supply, and converters. For the sake of simplicity, the total energy dissipation as presented in [36] has been partitioned into a fixed term, and a term that varies with the transmission powers. Thus, the energy dissipation of the system may be characterized by (4) (see equation at the bottom of the page), where P_C^B and P_C^R represent the fixed energy dissipation of each BS and each RN, respectively, while $\xi^B > 1$ and $\xi^R > 1$ are the energy dissipation multipliers of the transmit powers for the BS and the RNs, respectively. The effect of multiple transmit antennas on the total energy dissipation has been included in the terms P_C^B, P_C^R, ξ^B, and ξ^R.

Finally, the ESE of the system is expressed as

$$\eta_E(\mathcal{P},\mathcal{S}) = \frac{C_T(\mathcal{P},\mathcal{S})}{P_T(\mathcal{P},\mathcal{S})}. \tag{5}$$

The objective of this paper is to maximize (5) by appropriately optimizing $\mathcal{P}$ and $\mathcal{S}$.

IV. Semi-Orthogonal Grouping Algorithms

As described in Section II, the BS has to choose $Q_j^{T_1}$ and $Q_j^{T_2}$ SMCs for the first and second transmission phases, respectively. These selected SMCs collectively form the SMC group j. Since the system supports both direct and relaying links, the grouping algorithms described in [7], [8], which were designed for MIMO systems dispensing with relays, may not be directly applied. Instead, we propose a pair of viable grouping algorithms, namely the exhaustive search-based grouping algorithm (ESGA), and the orthogonal component-based grouping algorithm (OCGA).

Note that because there are multiple distributed transmitters in the second transmission phase, each UE designs its receive-BF matrix by jointly considering all the MIMO channel matrices associated with it, as described in Section III-B. However, before applying this method, we have to determine which particular transmitters (out of the BS and RNs) should actively transmit in the second transmission phase based on the results of SMC selection. Note that it is possible that the SMC candidates obtained may lead to higher effective CNRs when a subset of the transmitters are inactive. On the one hand, an additional effect of only activating a subset of transmitters is the reduced number of SMC candidates, which might in turn result in a reduced number of qualified SMCs that satisfy the semi-orthogonality criterion considered. As a result, the achievable spatial multiplexing gain and SE might be degraded. On the

$$C_T(\mathcal{P},\mathcal{S}) = \sum_{n=1}^{N}\sum_{j\in\mathcal{G}_n} s_{n,j}\left[\sum_{e_1\in\mathcal{E}_{n,j}} C_{n,j,e_1}^{BU,T_1}(\mathcal{P}) + \sum_{e_2\in\mathcal{E}_{n,j}} C_{n,j,e_2}^{BU,T_2}(\mathcal{P}) + \sum_{e\in\mathcal{E}_{n,j}} C_{n,j,e}^{BRU}(\mathcal{P})\right] \tag{3}$$

$$P_T(\mathcal{P},\mathcal{S}) = \left(P_C^B + M\cdot P_C^R\right) + \frac{1}{2}\sum_{n=1}^{N}\sum_{j\in\mathcal{G}_n} s_{n,j}\times\left[\xi^B\left(\sum_{e_1\in\mathcal{E}_{n,j}} P_{n,j,e_1}^{BU,T_1} + \sum_{e_2\in\mathcal{E}_{n,j}} P_{n,j,e_2}^{BU,T_2}\right) + \sum_{e\in\mathcal{E}_{n,j}}\left(\xi^B P_{n,j,e}^{BR,T_1} + \xi^R P_{n,j,e}^{RU,T_2}\right)\right] \tag{4}$$

other hand, this SE-reduction effect may be counteracted by the improved CNRs gleaned from the fact that it is easier to generate SMCs that can satisfy a stricter semi-orthogonality criterion, specified by a smaller value of α, when the number of transmitters is lower. For example, in the scenarios where only one or two active transmitters are selected, the UEs can employ receive-BF matrices that create effective DL channel matrices containing completely orthogonal rows by using the SVD or the exact diagonalization method (see Footnote [11]), respectively. To account for this dilemma, for the second transmission phase, the proposed grouping algorithms evaluate a full list of SMCs, which consists of the SMCs obtained from the $2^{M+1}-1$ possible combinations of active transmitters (the BS and M RNs, while ignoring the case when there are no active transmitters). Compared to using a smaller list of SMCs, using a full list of SMCs ensures that achieving a lower-bound SE is always guaranteed, while a higher SE can only be obtained upon increasing the number of transmitters in the system.

A. SMC Checking Algorithm

Both grouping algorithms must evaluate a particular SMC before it may be included into the SMC group to be generated. This evaluation process is depicted in Algorithm 1. More specifically, the algorithm identifies the transmitters associated with each SMC of the current SMC group, denoted by $\mathcal{E}_{n,j}$, in lines 7 to 17. The transmitter associated with the candidate SMC, e_c, is identified in lines 18 to 28. Additionally, as briefly pointed out in Footnote [12], for an active transmitter, if the candidate SMC is associated with a transmission in the second phase, then the auxiliary SMCs, $e^{\varnothing}$ and $e_m^{\varnothing}$, are included for the other active transmitters in lines 23, 26 and 27, to ensure that these potentially interfering transmitters do not impose interference on the candidate SMC.[14] Note that $e^{\varnothing}$ and $e_m^{\varnothing}$ represent auxiliary SMCs invoked by the BS and RNs, respectively. Having determined the transmitters associated with the SMCs, the algorithm checks that the SMCs associated with the same transmitter satisfy the semi-orthogonality criterion of (1) having parameter α in lines 20, 29, and 30. Furthermore, the algorithm ensures that the inclusion of the candidate SMC does not force any of the transmitters to transmit over its maximum number of transmit dimensions, as depicted in lines 29 and 30. Meanwhile, each UE should not receive more than its maximum number of receive dimensions, which is accomplished in lines 12, 32, and 33. Finally, the maximum achievable spatial multiplexing gain should not be exceeded in either the first or second phase, which is ensured by lines 35 and 36. If all of these checks are successful, the algorithm exits with a true condition in line 37.

[14]For distributed transmitters encountered in the second transmission phase, it is not feasible to design a single ZFBF transmit matrix as we did for the BS in the first transmission phase. For the second transmission phase, when $N_B \leq N_U$ and $N_R \leq N_U$, each SMC is associated with a single receive antenna. Consider this case as an example, when the BS is transmitting on a SMC to a particular receive antenna of a UE, an active RN may be transmitting zeros on an auxiliary SMC, which is also selected from the legitimate SMC candidates, to the same receive antenna of that UE. As a beneficial result of this strategy, for each transmitter, the interference imposed by other active transmitters is nulled.

Algorithm 1: SMC checking algorithm

inputs : candidate SMC e_c, current SMC group $\mathcal{E}_{n,j}$, semi-orthogonality parameter α
outputs: **true** or **false**

1 bool `SMCCheck` $(e_c, \mathcal{E}_{n,j}, \alpha)$
2 **begin**
3 $\quad \mathcal{T}^{BS,T_1} \leftarrow \{\}$;
4 $\quad \mathcal{T}^{BS,T_2} \leftarrow \{\}$;
5 $\quad \mathcal{T}_m^{RN,T_2} \leftarrow \{\}, \forall m \in \{1, \cdots, M\}$;
6 $\quad \mathcal{R}_k^{UE,T_2} \leftarrow \{\}, \forall k \in \{1, \cdots, K\}$;
7 $\quad$ **foreach** *SMC* $e_1 \in \mathcal{E}_{n,j}$ **do**
8 $\quad\quad \mathcal{T}^{BS,T_1} \leftarrow \mathcal{T}^{BS,T_1} \cup \{e_1\}$;
9 $\quad$ **end foreach**
10 $\quad$ **foreach** *SMC* $e_2 \in \mathcal{E}_{n,j}$ **do**
11 $\quad\quad \mathcal{T}^{BS,T_2} \leftarrow \mathcal{T}^{BS,T_2} \cup \{e_2\}$;
12 $\quad\quad \mathcal{R}_k^{UE,T_2} \leftarrow \mathcal{R}_k^{UE,T_2} \cup \{e_2\}$;
13 $\quad$ **end foreach**
14 $\quad$ **foreach** *SMC* $e \in \mathcal{E}_{n,j}$ **do**
15 $\quad\quad \mathcal{T}^{BS,T_2} \leftarrow \mathcal{T}^{BS,T_2} \cup \{e\}$;
16 $\quad\quad \mathcal{T}_{\mathcal{M}(e)}^{RN,T_2} \leftarrow \mathcal{T}_{\mathcal{M}(e)}^{RN,T_2} \cup \{e\}$;
17 $\quad$ **end foreach**
18 $\quad$ **if** e_c *is BS transmission in* T_1 **then**
19 $\quad\quad \mathcal{T}^{BS,T_1} \leftarrow \mathcal{T}^{BS,T_1} \cup \{e_c\}$;
20 $\quad\quad$ **if** $\mathcal{T}^{BS,T_1}$ *is not* α*-semi-orthogonal* **or** $|\mathcal{T}^{BS,T_1}| > N_B$ **then return false**;
21 $\quad$ **else if** e_c *is BS transmission in* T_2 **then**
22 $\quad\quad \mathcal{T}^{BS,T_2} \leftarrow \mathcal{T}^{BS,T_2} \cup \{e_c\}$;
23 $\quad\quad \mathcal{T}_m^{RN,T_2} \leftarrow \mathcal{T}_m^{RN,T_2} \cup \{e_m^{\varnothing}\}, \forall m \in \{1, \cdots, M\}$;
24 $\quad$ **else if** e_c *is RN* m *transmission in* T_2 **then**
25 $\quad\quad \mathcal{T}_m^{RN,T_2} \leftarrow \mathcal{T}_m^{RN,T_2} \cup \{e_c\}$;
26 $\quad\quad \mathcal{T}^{BS,T_2} \leftarrow \mathcal{T}^{BS,T_2} \cup \{e^{\varnothing}\}$;
27 $\quad\quad \mathcal{T}_{m'}^{RN,T_2} \leftarrow \mathcal{T}_{m'}^{RN,T_2} \cup \{e_{m'}^{\varnothing}\}, \forall m' \in \{1, \cdots, M\} \setminus m$;
28 $\quad$ **end if**
29 $\quad$ **if** $\mathcal{T}^{BS,T_2}$ *is not* α*-semi-orthogonal* **or** $|\mathcal{T}^{BS,T_2}| > N_B$ **then return false**;
30 $\quad$ **if** $\mathcal{T}_m^{RN,T_2}$ *is not* α*-semi-orthogonal* **or** $|\mathcal{T}_m^{RN,T_2}| > N_R$, $m \in \{1, \cdots, M\}$ **then return false**;
31 $\quad$ **if** e_c *is UE* k *reception in* T_2 **then**
32 $\quad\quad \mathcal{R}_k^{UE,T_2} \leftarrow \mathcal{R}_k^{UE,T_2} \cup \{e_c\}$;
33 $\quad\quad$ **if** $|\mathcal{R}_k^{UE,T_2}| > N_U$, $k \in \{1, \cdots, K\}$ **then return false**;
34 $\quad$ **end if**
35 $\quad$ **if** $|\mathcal{T}^{BS,T_1}| > \min(N_B, KN_U + MN_R)$ **then return false**;
36 $\quad$ **if** $\sum_{k=1}^{K} |\mathcal{R}_k^{UE,T_2}| > \min(N_B, N_R)$ **then return false**;
37 $\quad$ **return true**;
38 **end**

B. ESGA and OCGA

We present our first grouping method in Algorithm 2. Simply put, the ESGA recursively creates new SMC groups by exhaustively searching through all the possible combinations of SMCs and including those that pass the SMC checking algorithm. To elaborate, in the loop ranging from line 3 to line 9, the algorithm searches through all the possible SMCs associated with subcarrier block n, which are collectively denoted by $\mathcal{E}_n$ and satisfy $e_c \in \mathcal{E}_n$. The specific SMCs that satisfy the checks performed in line 4 are appended to the current SMC group in line 5, and the resultant updated SMC group $\mathcal{E}'_{n,j'}$ is appended to the set of SMC groups obtained for subcarrier block n in line

6. Additionally, $\mathcal{E}'_{n,j'}$ is used recursively in line 7 for filling this group and for forming new groups. The computational complexity of ESGA is dependent on the number of SMCs which are semi-orthogonal to each other. The worst-case complexity is obtained when every SMC satisfies the checks performed in line 4, leading to a time-complexity (in terms of the number of SMC groups generated) upper-bounded (not necessarily tight) by $\mathcal{O}(\sum_{n=1}^{N} |\mathcal{E}_n|^{\theta})$, where

$$\theta = \min\left[N_B, K \cdot \min(N_B, N_U) + M \cdot \min(N_B, N_R)\right] + \min\left[\min(N_B, N_R), \sum_{i=1}^{K} L_i^B + L_i^R\right]. \quad (6)$$

In other words, each subcarrier block may be treated independently. For each subcarrier block, $|\mathcal{E}_n|$ SMCs must be checked until the maximum multiplexing gain in both the first and second phases has been attained.

Algorithm 2: Exhaustive search-based grouping algorithm (ESGA)

inputs : set of SMC groups associated with subcarrier block n (initialized as empty set), $\mathcal{G}_n$
current SMC group (initialized as empty set), $\mathcal{E}_{n,j}$
SMCs associated with subcarrier block n, $\mathcal{E}_n$
semi-orthogonality parameter α
outputs: none

```
1  void ESGA (G_n, E_{n,j}, E_n, α)
2  begin
3      foreach e_c ∈ E_n do
4          if SMCCheck (e_c, E_{n,j}, α) then
5              E'_{n,j'} ← E_{n,j} ∪ {e_c};
6              G_n ← G_n ∪ {E'_{n,j'}};
7              ESGA (G_n, E'_{n,j'}, E_n \ e_c, α);
8          end if
9      end foreach
10     return;
11 end
```

The second algorithm, OCGA, is presented in Algorithm 3, which aims to be a lower complexity alternative to ESGA. The OCGA commences by creating a SMC candidate set $\mathcal{E}_c$, whose elements satisfy the checks performed in Algorithm 1, in lines 4 to 6. More specifically, if the current SMC group $\mathcal{E}_{n,j}$ is empty, the algorithm can simply create a new SMC group containing only the candidate SMC that has passed the SMC checks of Algorithm 1 in lines 7 to 10. If the SMC group is not empty, the algorithm adds to it the particular SMC candidate that results in the highest norm of the orthogonal component (NOC), via the Gram-Schmidt procedure [7], [8], in line 20. This process is repeated until the maximum multiplexing gain in both the first and second phases has been attained. When comparing the NOCs obtained for the relaying links, the minimum of the NOCs obtained from the BS-to-RN and RN-to-UE SMCs is used. This is because the information conveyed on the relaying link is limited by the weaker of the two transmissions, which is reflected in the effective channel gains quantified by these norms. If no SMCs satisfy the checks of line 6, the current SMC group is complete, and it is appended to the current set of SMC groups in line 18. Since new groups are only created when the current SMC group is empty, this algorithm results in much fewer groups than ESGA. The algorithmic time-complexity is given by $\mathcal{O}(\sum_{n=1}^{N} |\mathcal{E}_n|)$ as a single group is created for each initially-selected SMC.

Algorithm 3: Orthogonal component-based grouping algorithm (OCGA)

inputs : set of SMC groups associated with subcarrier block n (initialized as empty set), $\mathcal{G}_n$
current SMC group (initialized as empty set), $\mathcal{E}_{n,j}$
SMCs associated with subcarrier block n, $\mathcal{E}_n$
semi-orthogonality parameter α
outputs: none

```
1  void OCGA (G_n, E_{n,j}, E_n, α)
2  begin
3      complete ← true;
4      E_c ← {};
5      foreach e_c ∈ E_n do
6          if SMCCheck (e_c, E_{n,j}, α) then
7              if |E_{n,j}| == 0 then
8                  E'_{n,j'} ← E_{n,j} ∪ {e_c};
9                  OCGA (G_n, E'_{n,j'}, E_n \ e_c, α);
10                 return;
11             else
12                 E_c ← E_c ∪ {e_c};
13                 complete ← false;
14             end if
15         end if
16     end foreach
17     if complete then
18         G_n ← {E_{n,j}};
19     else
20         E'_{n,j'} ← E_{n,j} ∪ arg max_{e_c ∈ E_c} NOC (e_c, E_{n,j});
21         OCGA (G_n, E'_{n,j'}, E_n \ e_c, α);
22     end if
23     return;
24 end
```

Both grouping algorithms may be initialized with an empty SMC group, $\mathcal{E}_{n,j} \leftarrow \{\}$, and an empty set of SMC groups, $\mathcal{G}_n \leftarrow \{\}$, so that they recursively create and fill SMC groups according to their criteria. Additionally, a final step is performed to remove the specific groups, which result in effective channel gains that are less than or equal to that of another group, while having the same transmitters. Therefore, this final step does not reduce the attainable SE or ESE, but reduces the number of possible groups, thus alleviating the computational complexity imposed by the optimization algorithms of Section V-C.

V. SEM/ESEM Problem Formulation and Solution

Having obtained the set of SMC groups $\mathcal{G}_n$ for each subcarrier block n, in this section our aim is to find the optimum power variables contained in $\mathcal{P}$ and optimum SMC-group selection variables contained in $\mathcal{S}$, so that (5) is maximized. We commence by formulating the problem of maximizing the SE of the system as (7)–(13) (see equation at the bottom of the next page).

To elaborate, (7) represents the sum SE of the system, which is formulated in more detail as (3). The constraints (9)–(11) ensure that the maximum instantaneous transmission power constraint is never exceeded in either of the two transmission

phases for the BS and the RNs, while the constraints (8) and (12) ensure that only a single SMC group is selected for each subcarrier block. Finally, (13) restricts the power variables to be non-negative.

A. Relaxed SEM Problem

Although the constraint (13) is affine (hence convex) in the optimization variables contained in $\mathcal{P}$, (8)–(11) are non-convex [32], because (12) imposes a binary constraint on the problem. Furthermore, the objective function given by (7) is not concave, since it is dependent on the binary variables given by $\mathcal{S}$. Thus, (7)–(13) may be classified as a *mixed-integer non-linear programming* (MINLP) problem, which may be solved using branch-and-bound methods [37]. However, these methods typically incur a computational complexity that increases exponentially in the number of discrete variables, which is undesirable for practical implementations. To circumvent this initial setback, we introduce the following auxiliary variables

$$\widetilde{P}^{BU,T_1}_{n,j,e_1} = P^{BU,T_1}_{n,j,e_1}\widetilde{s}_{n,j}, \quad \forall n,j,e_1, \tag{14}$$
$$\widetilde{P}^{BR,T_1}_{n,j,e} = P^{BR,T_1}_{n,j,e}\widetilde{s}_{n,j}, \quad \forall n,j,e, \tag{15}$$
$$\widetilde{P}^{BU,T_2}_{n,j,e_2} = P^{BU,T_2}_{n,j,e_2}\widetilde{s}_{n,j}, \quad \forall n,j,e_2, \tag{16}$$
$$\widetilde{P}^{RU,T_2}_{n,j,e} = P^{RU,T_2}_{n,j,e}\widetilde{s}_{n,j}, \quad \forall n,j,e, \tag{17}$$
$$\widetilde{C}^{BU,T_1}_{n,j,e_1}, \quad \widetilde{C}^{BU,T_2}_{n,j,e_2} \text{ and } \widetilde{C}^{BRU}_{n,j,e}, \quad \forall n,j,e_1,e_2,e, \tag{18}$$

where we have relaxed[15] the binary constraint of (12) to give

$$0 \leq \widetilde{s}_{n,j} \leq 1, \qquad \forall n,j \tag{19}$$

so that we may write (7)–(13) in the hypograph problem [32] form given by (20)–(30),[16] (see equation at the bottom of the page), where $\widetilde{\mathcal{C}}$, $\widetilde{\mathcal{P}}$ and $\widetilde{\mathcal{S}}$ indicate the variable-sets containing their associated auxiliary variables.

It can be seen that the objective function of (7) has been replaced by (20) using the auxiliary rate variables given in (18), and by introducing the hypograph constraints (21)–(24).[17] These additional constraints ensure that the feasible auxiliary rate variables do not exceed their counterparts calculated on each link before using relaxation. As a result, the sum rate given by (20) invoking the feasible auxiliary rate variables does not exceed the sum rate given by (3) either.

As our next step, we prove that the problem described by (20)–(30) is a concave programming problem. Clearly, (20) is affine, hence concave, while (25)–(30) are all affine, and hence convex. Therefore, what remains is to show that constraints (21)–(24) are convex as well. These remaining constraints may be written in the form of

$$C - \frac{s}{2}\log_2\left(1+\frac{GP}{s}\right) \leq 0, \tag{31}$$

where s, P and C are the decision variables. It may be readily verified that $(1+GP)$ is affine and hence concave. Thus, $\log_2(1+GP)$ is concave, since $\log_2(\cdot)$ is concave and non-decreasing as a function of its argument. The function $s\log_2(1+(GP/s))$ is a perspective transformation[18] [32]

[15] In [38], such a relaxation results in a time-sharing solution regarding each subcarrier. In this work, this relaxation may be viewed as time-sharing of each subcarrier block, as multiple SMC groups can then occupy a fraction of each subcarrier block in time. Naturally, the relaxation means that we do not accurately solve the original problem of (7)–(13). However, as shown in [17], [21], [27], the solution to the original problem is still obtained with high probability when using the dual decomposition method on the relaxed problem (as in this work) as the number of subcarriers tends to infinity. It was shown that 8 subcarriers is sufficient for this to be true in the context of [39], while we have shown that 2 subcarriers is sufficient in the context of [17].

[16] Writing the original optimization problem in the hypograph form of (20)–(30) means that minimum per-link or system-wide SE constraints may be readily introduced. However, minimum SE constraints are not considered in this paper as our goal is to find the maximum SE/ESE solutions, which may not be equivalent to the solutions obtained when satisfying minimum SE constraints.

[17] Note that obtaining separate constraints for the first- and second-phase power control variables associated with the relayed transmission is made possible using the DF protocol. This then allowed us to readily derive the optimal power control variables as the decoupled water-filling solutions in Section V-C.1.

[18] Strictly speaking, the perspective transformation also requires that $s > 0$. However, convexity is also preserved for the situation when $s = 0$ as proven in [40].

$$\underset{\mathcal{P},\mathcal{S}}{\text{maximize}} \quad C_T(\mathcal{P},\mathcal{S}) \tag{7}$$

$$\text{subject to} \quad \sum_{j\in\mathcal{G}_n} s_{n,j} \leq 1, \qquad \forall n, \tag{8}$$

$$\sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n} s_{n,j}\left[\sum_{e_1\in\mathcal{E}_{n,j}} P^{BU,T_1}_{n,j,e_1} + \sum_{e\in\mathcal{E}_{n,j}} P^{BR,T_1}_{n,j,e}\right] \leq P^{B}_{max}, \tag{9}$$

$$\sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n} s_{n,j}\sum_{e_2\in\mathcal{E}_{n,j}} P^{BU,T_2}_{n,j,e_2} \leq P^{B}_{max}, \tag{10}$$

$$\sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n} s_{n,j}\sum_{\substack{e\in\mathcal{E}_{n,j}\\ \mathcal{M}(e)=m}} P^{RU,T_2}_{n,j,e} \leq P^{R}_{max}, \qquad \forall m, \tag{11}$$

$$s_{n,j} \in \{0,1\}, \qquad \forall n,j, \tag{12}$$

$$P^{BU,T_1}_{n,j,e_1}, P^{BR,T_1}_{n,j,e}, P^{BU,T_2}_{n,j,e_2}, P^{RU,T_2}_{n,j,e} \geq 0, \qquad \forall n,j,e_1,e_2,e \tag{13}$$

of $\log_2(1+GP)$, which preserves concavity. Finally, $C-(s/2)\log_2(1+(GP/s))$ is convex, since it is the sum of two convex functions. Since (31) is convex, it is clear that constraints (21)–(24) are convex, and so (20)–(30) is a concave programming problem, whose solution algorithm is presented in Section V-C.

B. ESEM Problem

The ESE objective function, given by (32) (see equation at the bottom of the page), is formed by dividing the objective function (20) by $P_T(\widetilde{\mathcal{P}},\widetilde{\mathcal{S}})$, which is obtained by substituting (14)–(17) into (4) and introducing the relaxed variables $\widetilde{s}_{n,j}$.

The objective function (32) is a linear-fractional function, since it is a ratio of two affine functions. Thus the ESEM problem can be solved using the Charnes-Cooper transformation of [24], as given by

$$\widehat{C}^{BU,T_1}_{n,j,e_1} = \widetilde{C}^{BU,T_1}_{n,j,e_1} t, \qquad \forall n,j,e_1, \tag{33}$$

$$\widehat{C}^{BU,T_2}_{n,j,e_2} = \widetilde{C}^{BU,T_2}_{n,j,e_2} t, \qquad \forall n,j,e_2, \tag{34}$$

$$\widehat{C}^{BRU}_{n,j,,e} = \widetilde{C}^{BRU}_{n,j,e} t, \qquad \forall n,j,e, \tag{35}$$

$$\widehat{P}^{BU,T_1}_{n,j,e_1} = \widetilde{P}^{BU,T_1}_{n,j,e_1} t, \qquad \forall n,j,e_1, \tag{36}$$

$$\widehat{P}^{BU,T_2}_{n,j,e_2} = \widetilde{P}^{BU,T_2}_{n,j,e_2} t, \qquad \forall n,j,e_2, \tag{37}$$

$$\widehat{P}^{BR,T_1}_{n,j,e} = \widetilde{P}^{BR,T_1}_{n,j,e} t, \qquad \forall n,j,e, \tag{38}$$

$$\widehat{P}^{RU,T_2}_{n,j,e} = \widetilde{P}^{RU,T_2}_{n,j,e} t, \qquad \forall n,j,e, \tag{39}$$

$$\widehat{s}_{n,j} = \widetilde{s}_{n,j} t, \qquad \forall n,j, \tag{40}$$

$$\underset{\widetilde{\mathcal{C}},\widetilde{\mathcal{P}},\widetilde{\mathcal{S}}}{\text{maximize}} \quad \sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\left[\sum_{e_1\in\mathcal{E}_{n,j}}\widetilde{C}^{BU,T_1}_{n,j,e_1}+\sum_{e_2\in\mathcal{E}_{n,j}}\widetilde{C}^{BU,T_2}_{n,j,e_2}\right]+\left[\sum_{e\in\mathcal{E}_{n,j}}\widetilde{C}^{BRU}_{n,j,e}\right] \tag{20}$$

$$\text{subject to} \quad \frac{\widetilde{s}_{n,j}}{2}\log_2\left(1+\frac{G^{BU,T_1}_{n,j,e_1}\widetilde{P}^{BU,T_1}_{n,j,e_1}}{\widetilde{s}_{n,j}}\right)\geq \widetilde{C}^{BU,T_1}_{n,j,e_1}, \qquad \forall n,j,e_1, \tag{21}$$

$$\frac{\widetilde{s}_{n,j}}{2}\log_2\left(1+\frac{G^{BU,T_2}_{n,j,e_2}\widetilde{P}^{BU,T_2}_{n,j,e_2}}{\widetilde{s}_{n,j}}\right)\geq \widetilde{C}^{BU,T_2}_{n,j,e_2}, \qquad \forall n,j,e_2, \tag{22}$$

$$\frac{\widetilde{s}_{n,j}}{2}\log_2\left(1+\frac{G^{BR,T_1}_{n,j,e}\widetilde{P}^{BR,T_1}_{n,j,e}}{\widetilde{s}_{n,j}}\right)\geq \widetilde{C}^{BRU}_{n,j,e}, \qquad \forall n,j,e, \tag{23}$$

$$\frac{\widetilde{s}_{n,j}}{2}\log_2\left(1+\frac{G^{RU,T_2}_{n,j,e}\widetilde{P}^{RU,T_2}_{n,j,e}}{\widetilde{s}_{n,j}}\right)\geq \widetilde{C}^{BRU}_{n,j,e}, \qquad \forall n,j,e, \tag{24}$$

$$\sum_{j\in\mathcal{G}_n}\widetilde{s}_{n,j}\leq 1, \qquad \forall n, \tag{25}$$

$$\sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\left[\sum_{e_1\in\mathcal{E}_{n,j}}\widetilde{P}^{BU,T_1}_{n,j,e_1}+\sum_{e\in\mathcal{E}_{n,j}}\widetilde{P}^{BR,T_1}_{n,j,e}\right]\leq P^B_{max}, \tag{26}$$

$$\sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\sum_{e_2\in\mathcal{E}_{n,j}}\widetilde{P}^{BU,T_2}_{n,j,e_2}\leq P^B_{max}, \tag{27}$$

$$\sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\sum_{\substack{e\in\mathcal{E}_{n,j}\\ \mathcal{M}(e)=m}}\widetilde{P}^{RU,T_2}_{n,j,e}\leq P^R_{max}, \qquad \forall m, \tag{28}$$

$$0\leq\widetilde{s}_{n,j}\leq 1, \qquad \forall n,j, \tag{29}$$

$$\widetilde{P}^{BU,T_1}_{n,j,e_1},\widetilde{P}^{BR,T_1}_{n,j,e},\widetilde{P}^{BU,T_2}_{n,j,e_2},\widetilde{P}^{RU,T_2}_{n,j,e}\geq 0, \qquad \forall n,j,e_1,e_2,e \tag{30}$$

$$\frac{\displaystyle\sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\left[\sum_{e_1\in\mathcal{E}_{n,j}}\widetilde{C}^{BU,T_1}_{n,j,e_1}+\sum_{e_2\in\mathcal{E}_{n,j}}\widetilde{C}^{BU,T_2}_{n,j,e_2}\right]+\left[\sum_{e\in\mathcal{E}_{n,j}}\widetilde{C}^{BRU}_{n,j,e}\right]}{\left(P^B_C+M\cdot P^R_C\right)+\frac{1}{2}\displaystyle\sum_{n=1}^{N}\sum_{j\in\mathcal{G}_n}\left[\xi^B\left(\sum_{e_1\in\mathcal{E}_{n,j}}\widetilde{P}^{BU,T_1}_{n,j,e_1}+\sum_{e_2\in\mathcal{E}_{n,j}}\widetilde{P}^{BU,T_2}_{n,j,e_2}\right)+\sum_{e\in\mathcal{E}_{n,j}}\left(\xi^B\widetilde{P}^{BR,T_1}_{n,j,e}+\xi^R\widetilde{P}^{RU,T_2}_{n,j,e}\right)\right]} \tag{32}$$

where the auxiliary variable t is given by

$$t = \frac{1}{P_T(\widetilde{\mathcal{P}}, \widetilde{\mathcal{S}})}. \tag{41}$$

Thus, the ESEM problem may be written[19] as (42)–(53) (see equation at the bottom of the page), where $\widehat{\mathcal{C}}$, $\widehat{\mathcal{P}}$ and $\widehat{\mathcal{S}}$ indicate the variable-sets containing their associated transformed variables. It is clear that the objective function (42) is affine, hence concave, while the constraints (47)–(53) are all affine, and hence convex. The constraints (43)–(46) are of the form (31) and are hence convex. Therefore, the problem described by (42)–(53) is a concave programming problem, which can be solved using the algorithm of Section V-C.

[19]Strictly speaking, the constraint $t > 0$ is also needed, but this is guaranteed due to constraint (53).

C. Dual Decomposition Based Solution Algorithm

The dual decomposition method of [17], [41] may be used for conceiving solution algorithms for our SEM and ESEM problems formulated as (20)–(30) and (42)–(53), respectively. We commence by describing the solution algorithm conceived for (42)–(53), which we term the ESEM algorithm. The ESEM algorithm, based on dual decomposition, iterates between calculating the tentative optima of the primal variables, namely $\widehat{C}^{BU,T_1}_{n,j,e_1}$, $\widehat{C}^{BU,T_2}_{n,j,e_2}$, $\widehat{C}^{BRU}_{n,j,e}$, $\widehat{P}^{BU,T_1}_{n,j,e_1}$, $\widehat{P}^{BU,T_2}_{n,j,e_2}$, $\widehat{P}^{BR,T_1}_{n,j,e}$, $\widehat{P}^{RU,T_2}_{n,j,e}$, $\widehat{s}_{n,j}$ as well as t, and updating the dual variables λ^{T_1}, λ^{T_2}, ν_m as well as μ, which will be defined later, until the objective function value converges.

1) Calculating Tentative Optima of Primal Variables: Based on our previous work [17] that employed the dual decomposition and by employing the Karush-Kuhn-Tucker

$$\underset{\widehat{\mathcal{C}},\widehat{\mathcal{P}},\widehat{\mathcal{S}},t}{\text{maximize}} \quad \sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\left[\sum_{e_1\in\mathcal{E}_{n,j}}\widehat{C}^{BU,T_1}_{n,j,e_1}+\sum_{e_2\in\mathcal{E}_{n,j}}\widehat{C}^{BU,T_2}_{n,j,e_2}\right]+\left[\sum_{e\in\mathcal{E}_{n,j}}\widehat{C}^{BRU}_{n,j,e}\right] \tag{42}$$

$$\text{subject to} \quad \frac{\widehat{s}_{n,j}}{2}\log_2\left(1+\frac{G^{BU,T_1}_{n,j,e_1}\widehat{P}^{BU,T_1}_{n,j,e_1}}{\widehat{s}_{n,j}}\right)\geq\widehat{C}^{BU,T_1}_{n,j,e_1}, \qquad \forall n,j,e_1, \tag{43}$$

$$\frac{\widehat{s}_{n,j}}{2}\log_2\left(1+\frac{G^{BU,T_2}_{n,j,e_2}\widehat{P}^{BU,T_2}_{n,j,e_2}}{\widehat{s}_{n,j}}\right)\geq\widehat{C}^{BU,T_2}_{n,j,e_2}, \qquad \forall n,j,e_2, \tag{44}$$

$$\frac{\widehat{s}_{n,j}}{2}\log_2\left(1+\frac{G^{BR,T_1}_{n,j,e}\widehat{P}^{BR,T_1}_{n,j,e}}{\widehat{s}_{n,j}}\right)\geq\widehat{C}^{BRU}_{n,j,e}, \qquad \forall n,j,e, \tag{45}$$

$$\frac{\widehat{s}_{n,j}}{2}\log_2\left(1+\frac{G^{RU,T_2}_{n,j,e}\widehat{P}^{RU,T_2}_{n,j,e}}{\widehat{s}_{n,j}}\right)\geq\widehat{C}^{BRU}_{n,j,e}, \qquad \forall n,j,e, \tag{46}$$

$$\sum_{j\in\mathcal{G}_n}\widehat{s}_{n,j}\leq t, \qquad \forall n, \tag{47}$$

$$\sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\left[\sum_{e_1\in\mathcal{E}_{n,j}}\widehat{P}^{BU,T_1}_{n,j,e_1}+\sum_{e\in\mathcal{E}_{n,j}}\widehat{P}^{BR,T_1}_{n,j,e}\right]\leq t\cdot P^B_{max}, \tag{48}$$

$$\sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\sum_{e_2\in\mathcal{E}_{n,j}}\widehat{P}^{BU,T_2}_{n,j,e_2}\leq t\cdot P^B_{max}, \tag{49}$$

$$\sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\sum_{\substack{e\in\mathcal{E}_{n,j}\\ \mathcal{M}(e)=m}}\widehat{P}^{RU,T_2}_{n,j,e}\leq t\cdot P^R_{max}, \qquad \forall m, \tag{50}$$

$$0\leq\widehat{s}_{n,j}\leq t, \qquad \forall n,j, \tag{51}$$

$$\widehat{P}^{BU,T_1}_{n,j,e_1},\widehat{P}^{BR,T_1}_{n,j,e},\widehat{P}^{BU,T_2}_{n,j,e_2},\widehat{P}^{RU,T_2}_{n,j,e}\geq 0, \qquad \forall n,j,e_1,e_2,e, \tag{52}$$

$$t\cdot\left(P^B_C+M\cdot P^R_C\right)+\frac{1}{2}\sum_{n=1}^{N}\sum_{j\in\mathcal{G}_n}\left[\xi^B\left(\sum_{e_1\in\mathcal{E}_{n,j}}\widehat{P}^{BU,T_1}_{n,j,e_1}+\sum_{e_2\in\mathcal{E}_{n,j}}\widehat{P}^{BU,T_2}_{n,j,e_2}\right)+\sum_{e\in\mathcal{E}_{n,j}}\left(\xi^B\widehat{P}^{BR,T_1}_{n,j,e}+\xi^R\widehat{P}^{RU,T_2}_{n,j,e}\right)\right]=1 \tag{53}$$

optimality conditions [32], we reveal that the tentatively optimal transformed power control variables for the direct SMCs encountered in the problem of (42)–(53) may be formulated as the water-filling solutions of[20]

$$\widehat{P}_{n,j,e_1}^{BU,T_1} = \widehat{s}_{n,j}\left[\frac{1}{(\xi^B\mu + 2\lambda^{T_1})\ln 2} - \frac{1}{G_{n,j,e_1}^{BU,T_1}}\right]^+ = \widehat{s}_{n,j}P_{n,j,e_1}^{BU,T_1} \tag{54}$$

and

$$\widehat{P}_{n,j,e_2}^{BU,T_2} = \widehat{s}_{n,j}\left[\frac{1}{(\xi^B\mu + 2\lambda^{T_2})\ln 2} - \frac{1}{G_{n,j,e_2}^{BU,T_2}}\right]^+ = \widehat{s}_{n,j}P_{n,j,e_2}^{BU,T_2}. \tag{55}$$

Furthermore, the transformed power control variables for the relaying SMCs may be *initially* written as

$$\widehat{P}_{n,j,e}^{BR,T_1} = \widehat{s}_{n,j}\left[\frac{1}{(\xi^B\mu + 2\lambda^{T_1})\ln 2} - \frac{1}{G_{n,j,e}^{BR,T_1}}\right]^+ = \widehat{s}_{n,j}P_{n,j,e}^{BR,T_1} \tag{56}$$

and

$$\widehat{P}_{n,j,e}^{RU,T_2} = \widehat{s}_{n,j}\left[\frac{1}{\left(\xi^R\mu + 2\nu_{\mathcal{M}(e)}\right)\ln 2} - \frac{1}{G_{n,j,e}^{RU,T_2}}\right]^+ = \widehat{s}_{n,j}P_{n,j,e}^{RU,T_2}. \tag{57}$$

Note that the value of $\widehat{s}_{n,j}$ in (54)–(57) is not yet known. Since the SE attainable for a relaying link is limited by the weaker of the BS-to-RN and RN-to-UE links, there is no need to transmit at a high power on the stronger link, if the other link is unable to support the high SE. Thus, the tentatively optimal transformed power control variables provided for the relaying SMC e may be refined by substituting (56), (57) into the right-hand side of

$$\widehat{P}_{n,j,e}^{BR,T_1} = \min\left(\widehat{P}_{n,j,e}^{BR,T_1}, \frac{G_{n,j,e}^{RU,T_2}}{G_{n,j,e}^{BR,T_1}}\cdot\widehat{P}_{n,j,e}^{RU,T_2}\right) \tag{58}$$

and

$$\widehat{P}_{n,j,e}^{RU,T_2} = \min\left(\widehat{P}_{n,j,e}^{RU,T_2}, \frac{G_{n,j,e}^{BR,T_1}}{G_{n,j,e}^{RU,T_2}}\cdot\widehat{P}_{n,j,e}^{BR,T_1}\right). \tag{59}$$

[20]In this paper, $[\cdot]^+$ is equivalent to $\max(0,\cdot)$.

As a result, the tentative estimates of the maximum values that $\widehat{C}_{n,j,e_1}^{BU,T_1}$, $\widehat{C}_{n,j,e_2}^{BU,T_2}$ and $\widehat{C}_{n,j,e}^{BRU}$ can attain are given by

$$\widehat{C}_{n,j,e_1}^{BU,T_1} = \frac{\widehat{s}_{n,j}}{2}\log_2\left(1 + \frac{G_{n,j,e_1}^{BU,T_1}\widehat{P}_{n,j,e_1}^{BU,T_1}}{\widehat{s}_{n,j}}\right), \tag{60}$$

$$\widehat{C}_{n,j,e_2}^{BU,T_2} = \frac{\widehat{s}_{n,j}}{2}\log_2\left(1 + \frac{G_{n,j,e_2}^{BU,T_2}\widehat{P}_{n,j,e_2}^{BU,T_2}}{\widehat{s}_{n,j}}\right), \tag{61}$$

and

$$\widehat{C}_{n,j,e}^{BRU} = \frac{\widehat{s}_{n,j}}{2}\log_2\left(1 + \frac{G_{n,j,e}^{BR,T_1}\widehat{P}_{n,j,e}^{BR,T_1}}{\widehat{s}_{n,j}}\right) = \frac{\widehat{s}_{n,j}}{2}\log_2\left(1 + \frac{G_{n,j,e}^{RU,T_2}\widehat{P}_{n,j,e}^{RU,T_2}}{\widehat{s}_{n,j}}\right), \tag{62}$$

where the value of $\widehat{s}_{n,j}$ remains unknown. However, it is plausible that for the purpose of maximizing the objective function value, $\widehat{s}_{n,j}$, $\forall n,j$ will always be given its maximum value t, if the single SMC group j is selected for subcarrier block n. Thus, the tentatively optimal SMC group j for subcarrier block n is given by the group obtaining the highest value of

$$\sum_{j\in\mathcal{G}_n}\left[\sum_{e_1\in\mathcal{E}_{n,j}}\widehat{C}_{n,j,e_1}^{BU,T_1} + \sum_{e_2\in\mathcal{E}_{n,j}}\widehat{C}_{n,j,e_2}^{BU,T_2}\right] + \sum_{e\in\mathcal{E}_{n,j}}\widehat{C}_{n,j,e}^{BRU}. \tag{63}$$

where $\widehat{s}_{n,j}$ inside the logarithm functions may be canceled out. Additionally, we can ignore the common positive multiplicative factor of t without affecting the maximization of (63). The objective function (42) is maximized when choosing this particular group j for subcarrier block n, while for the remaining groups associated with the same subcarrier block, we set $\widehat{P}_{n,j'\neq j,e_1}^{BU,T_1} = \widehat{P}_{n,j'\neq j,e_2}^{BU,T_2} = \widehat{P}_{n,j'\neq j,e}^{BR,T_1} = \widehat{P}_{n,j'\neq j,e}^{RU,T_2} = \widehat{s}_{n,j'\neq j} = \widehat{C}_{n,j',e_1}^{BU,T_1} = \widehat{C}_{n,j'\neq j,e_2}^{BU,T_2} = \widehat{C}_{n,j'\neq j,e}^{BRU} = P_{n,j'\neq j,e_1}^{BU,T_1} = P_{n,j'\neq j,e_2}^{BU,T_2} = P_{n,j'\neq j,e}^{BR,T_1} = P_{n,j'\neq j,e}^{RU,T_2} = 0$, as these remaining groups are not chosen.

Consequently, the value of t is given by (64) (see equation at the bottom of the page). Note that this is possible without knowing the exact value of $\widehat{s}_{n,j}$, since the factor of $\widehat{s}_{n,j}$ may be canceled out, and thus (64) is only dependent on the dual variables and on the tentatively optimal SMC group selection.

Having identified the tentative optimal SMC group, we set $\widehat{s}_{n,j} = t$ for this selected SMC group corresponding to each subcarrier block n, and we have

$$\widehat{C}_{n,j,e_1}^{BU,T_1} = \frac{t}{2}\log_2\left(1 + G_{n,j,e_1}^{BU,T_1}P_{n,j,e_1}^{BU,T_1}\right), \tag{65}$$

$$\widehat{C}_{n,j,e_2}^{BU,T_2} = \frac{t}{2}\log_2\left(1 + G_{n,j,e_2}^{BU,T_2}P_{n,j,e_2}^{BU,T_2}\right) \tag{66}$$

$$t = \left(P_C^B + M\cdot P_C^R + \frac{1}{2}\sum_{n=1}^{N}\sum_{j\in\mathcal{G}_n}\left[\xi^B\left(\sum_{e_1\in\mathcal{E}_{n,j}}P_{n,j,e_1}^{BU,T_1} + \sum_{e_2\in\mathcal{E}_{n,j}}P_{n,j,e_2}^{BU,T_2}\right) + \sum_{e\in\mathcal{E}_{n,j}}\left(\xi^B P_{n,j,e}^{BR,T_1} + \xi^R P_{n,j,e}^{RU,T_2}\right)\right]\right)^{-1} \tag{64}$$

as well as

$$\widehat{C}^{BRU}_{n,j,e} = \frac{t}{2}\log_2\left(1+G^{BR,T_1}_{n,j,e}P^{BR,T_1}_{n,j,e}\right) = \frac{t}{2}\log_2\left(1+G^{RU,T_2}_{n,j,e}P^{RU,T_2}_{n,j,e}\right), \quad (67)$$

for that selected SMC group. To summarize, given a set of dual variables, the values of power control variables are obtained, resulting in an tentatively optimal SMC group, which obtains the SE values for the corresponding subcarrier block. Therefore, all of the primal variables are obtained for a given set of dual variables. Thus, they are jointly optimized.

2) Updating the Dual Variables: From the derivation of the optimal primal variables described in Section V-C.1, we can see that the constraints (43)–(47) and (51)–(53) are implicitly satisfied. Therefore, we update the dual variables λ^{T_1}, λ^{T_2} and ν_m which are associated with the remaining constraints (48)–(50), respectively. These may be viewed as pricing parameters to ensure that the optimal power control variables satisfy (48)–(50).

Since the Lagrangian of (42)–(53) is differentiable w.r.t. the dual variables, at each iteration i of the solution algorithm, these dual variables may be updated according to (68)–(70) (see equation at the bottom of the page), where $\delta_{\lambda^{T_1}}(i)$, $\delta_{\lambda^{T_2}}(i)$ and $\delta_{\nu_m}(i)$ are appropriately chosen step sizes [41] at iteration i.

The remaining dual variable, μ, which is associated with (53) must also be updated. However, the constraint given by (53) is implicitly satisfied since the value of t is computed from (64). Therefore, we opt for an alternative method based on differentiating the Lagrangian w.r.t t and substituting in the intermediate values of $\widehat{\mathcal{C}}$, $\widehat{\mathcal{P}}$, $\widehat{\mathcal{S}}$ and t. Thus, the updated value of μ is given by (71).

All primal variables are jointly optimized in Section V-C.1 as the optimal power variables are determined by the related dual variables. This leads to the optimal group selection and rate variables, which then allow us to find the optimal t. Given the tentative optima of primal variables, the algorithm proceeds to update the dual variables, which are mostly to ensure that the maximum power constraints are not violated. Using these updated dual variables, the algorithm repeats this process until the objective function value $\widehat{\eta}_E(i)$ at iteration i reaches the predefined convergence threshold, which is given by $|\widehat{\eta}_E(i)-\widehat{\eta}_E(i-1)|<\epsilon$.

The method presented in Section V-C.1 and Section V-C.2 solves the ESEM problem described by (42)–(53). It may also be invoked for solving the SEM problem of (20)–(30), while fixing $\mu=0$ and $t=1$. This is because the ESEM problem considered is simplified to the SEM problem, when we have $\mu=0$ and $t=1$.

$$\lambda^{T_1}(i) = \left[\lambda^{T_1}(i-1) - \delta_{\lambda^{T_1}}(i)\cdot\left(t\cdot P^B_{max} - \sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\left[\sum_{e_1\in\mathcal{E}_{n,j}}\widehat{P}^{BU,T_1}_{n,j,e_1} + \sum_{e\in\mathcal{E}_{n,j}}\widehat{P}^{BR,T_1}_{n,j,e}\right]\right)\right]^+ \quad (68)$$

$$\lambda^{T_2}(i) = \left[\lambda^{T_2}(i-1) - \delta_{\lambda^{T_2}}(i)\cdot\left(t\cdot P^B_{max} - \sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\sum_{e_2\in\mathcal{E}_{n,j}}\widehat{P}^{BU,T_2}_{n,j,e_2}\right)\right]^+ \quad (69)$$

$$\nu_m(i) = \left[\nu_m(i-1) - \delta_{\nu_m}(i)\cdot\left(t\cdot P^R_{max} - \sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\sum_{\substack{e\in\mathcal{E}_{n,j}\\ \mathcal{M}(e)=m}}\widehat{P}^{RU,T_2}_{n,j,e}\right)\right]^+, \quad \forall m \quad (70)$$

$$\begin{aligned}\mu(i) = t\cdot\Bigg(&\sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\left[\sum_{e_1\in\mathcal{E}_{n,j}}\widetilde{C}^{BU,T_1}_{n,j,e_1} + \sum_{e_2\in\mathcal{E}_{n,j}}\widetilde{C}^{BU,T_2}_{n,j,e_2}\right] + \left[\sum_{e\in\mathcal{E}_{n,j}}\widetilde{C}^{BRU}_{n,j,e}\right]\\ &+\lambda^{T_1}(i-1)\cdot\left(P^B_{max} - \sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\left[\sum_{e_1\in\mathcal{E}_{n,j}}\widetilde{P}^{BU,T_1}_{n,j,e_1} + \sum_{e\in\mathcal{E}_{n,j}}\widetilde{P}^{BR,T_1}_{n,j,e}\right]\right)\\ &+\lambda^{T_2}(i-1)\cdot\left(P^B_{max} - \sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\sum_{e_2\in\mathcal{E}_{n,j}}\widetilde{P}^{BU,T_2}_{n,j,e_2}\right)\\ &+\sum_{m=1}^{M}\nu_m(i-1)\cdot\left(P^R_{max} - \sum_{i=1}^{N}\sum_{j\in\mathcal{G}_n}\sum_{\substack{e\in\mathcal{E}_{n,j}\\ \mathcal{M}(e)=m}}\widetilde{P}^{RU,T_2}_{n,j,e}\right)\Bigg)\end{aligned} \quad (71)$$

TABLE I
SIMULATION PARAMETERS USED TO OBTAIN ALL RESULTS IN SECTION VI UNLESS OTHERWISE SPECIFIED

Simulation parameter	Value
Subcarrier block bandwidth, W [Hertz]	180k
Number of RNs, M	$\{0, 1, 2, 4\}$
Number of subcarriers blocks, N	$\{6, 12, 25, 50, 100\}$
Number of UEs, K	$\{2, 10\}$
Antenna configuration, (N_B, N_R, N_U)	$(4, 4, 2)$
Cell radius, [km]	$\{0.75, 1.25, 1.75, 2.25\}$
Ratio of BS-to-RN distance to the cell radius, D_r	0.5
SNR gap of wireless transceivers, $\Delta\gamma$ [dB]	0
Maximum total transmission power of the BS and RNs, P^B_{max} and P^R_{max} [dBm]	$\{0, 10, 20, 30, 40, 50, 60\}$
Fixed power rating of the BS, $P_C^{(B)}$ [Watts] [36], [42]	$32.306N_B$
Fixed power rating of RNs, $P_C^{(R)}$ [Watts] [36], [42]	$21.874N_R$
Reciprocal of the BS power amplifier's drain efficiency, $\xi^{(B)}$ [36], [42]	$3.24N_B$
Reciprocal of the RNs' power amplifier's drain efficiency, $\xi^{(R)}$ [36], [42]	$4.04N_R$
Noise power spectral density, N_0 [dBm/Hz]	−174
Convergence threshold, ϵ	10^{-6}
Number of channel samples	10^4

VI. NUMERICAL RESULTS AND DISCUSSIONS

This section presents the numerical results obtained, when employing the SEM and ESEM algorithms[21] described in Section V to the MIMO-OFDMA multi-relay cellular network considered. The pertinent simulation parameters are given in Table I. Additionally, the path-loss effect is characterized relying on the method and parameters of [30], where the BS-to-UE and RN-to-UE links are assumed to be non-line-of-sight (NLOS) links, since these links are typically blocked by buildings and other large obstructing objects, while the BS-to-RN links are realistically assumed to be line-of-sight (LOS) links, as the RNs may be strategically deployed on tall buildings to create strong wireless backhaul links. Furthermore, independently and randomly generated set of UE locations as well as fading channel realizations were used for each channel sample.

The results of a baseline algorithm is also presented to highlight the improved performance obtained from employing the SEM and ESEM algorithms. This baseline algorithm consists of a random SMC grouping (RG) selection for each subcarrier block and then equal power allocation (EPA) across all the selected SMCs, and will be termed the RG-EPA algorithm.

A. On the Optimality and the Relative Complexity of ESGA and OCGA for Various α Values

Firstly, the behavior of the ESGA and OCGA as a function of α is examined. Note in Fig. 4 that since the ESGA is capable of enumerating all possible SMC groupings, which satisfy (1) for the corresponding α, the optimal SE is attained. The 'normalized optimality gap' is then defined as $(\beta/\beta^*) - 1$, where β^* is the optimal SE obtained from employing the ESGA algorithm, and β is the SE obtained from any other algorithm. We can see from Fig. 4, that the normalized optimality gap of OCGA relative to ESGA is about $-0.005 \sim -0.1$ for the α values considered. However, the number of groups found using ESGA is exponentially increasing with α. By contrast, for OCGA, this number is always significantly lower and gradually becomes less than 200, when α increases to 0.5. In fact, the number of groups found by OCGA is reduced to about 3.5% of that found by ESGA at $\alpha = 0.5$. This demonstrates the viability of using OCGA in the following simulations as a reduced-complexity near-optimum alternative to ESGA. Under the conditions considered in Fig. 4, the optimal ESE solution is the same as the optimal SE solution, as detailed in the next subsection. Therefore, as far as ESEM is concerned, similar conclusions may be drawn regarding the optimality of the two grouping algorithms.

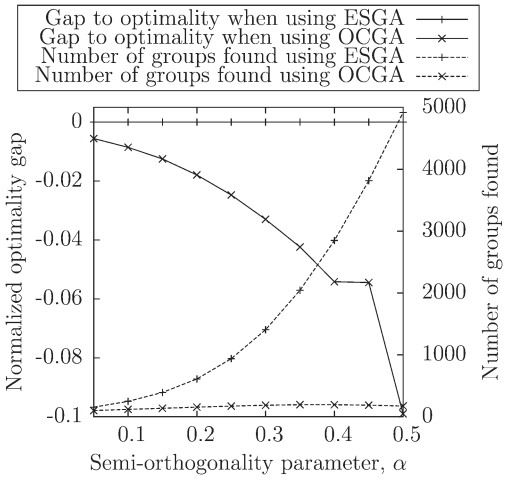

Fig. 4. The optimality gap and total number of SMC groups found when employing the ESGA and OCGA, and using the parameters in Table I with $N = 6$, $K = 2$, $M = 2$, $P^B_{max} = 20$ dBm, $P^R_{max} = 10$ dBm and a cell radius of 0.75 km.

B. The Variation in Achievable SE and ESE for Different Values of P^B_{max} and P^R_{max}

As shown in Fig. 5(a), the achievable SE is monotonically increasing with P^B_{max} and P^R_{max} when using the SEM algorithm. This is not unexpected, since the SEM algorithm optimally allocates all the available power for the sake of achieving the maximum SE. By comparison, we observed from Fig. 5(a) and (b) that both the achievable SE and ESE of the ESEM algorithm saturate at some moderate values of P^B_{max} and/or P^R_{max}. This is because the ESEM algorithm only allocates *just* enough power (that may be lower than the power budget values of P^B_{max} and/or P^R_{max}) for the sake of achieving the maximum ESE. On the other hand, the ESE performance of the SEM algorithm is severely degraded upon further increasing P^B_{max} and/or P^R_{max} after its ESE performance reaches the peak, as shown in Fig. 5(b). This is because the ESE metric is a quasiconcave

[21] In all cases, the step sizes and the initial values of the dual variables described in Section V-C.2 are empirically optimized to give the optimal objective function value in as few iterations as possible, although the exact analytical method for determining the optimal step sizes and initial values still remains an open issue. In our experience, the algorithms converge within just 10 iterations when carefully chosen step sizes are employed, regardless of the size of the problem.

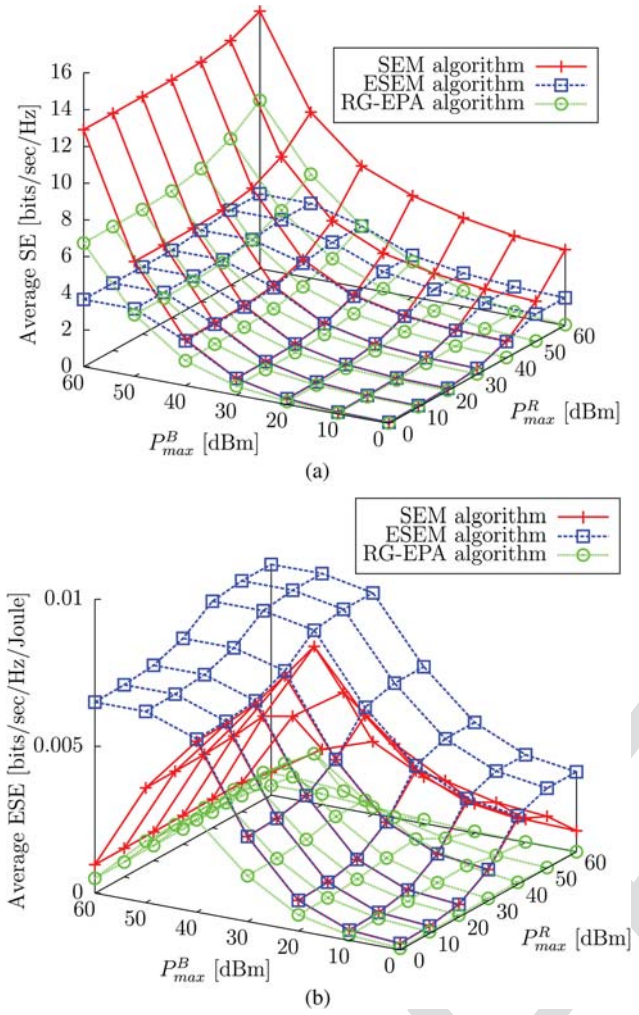

Fig. 5. The average achievable SE and ESE of the SEM, ESEM and RG-EPA algorithms upon varying P^B_{max} and P^R_{max}. The parameters in Table I with $N = 6$, $K = 10$, $M = 2$, $\alpha = 0.1$ and a cell radius of 1.75 km are used. (a) Surface plots of the achievable SE when using the SEM, ESEM and RG-EPA algorithms. (b) Surface plots of the achievable ESE when using the SEM, ESEM and RG-EPA algorithms.

function of the transmit powers—its numerator (i.e. the SE) increases logarithmically with the transmit powers, while its denominator increases linearly with the transmit powers. In fact, the peak ESE of the SEM algorithm is attained at $P^B_{max} = 40$ dBm and $P^R_{max} = 40$ dBm, as seen in Fig. 5(b), and the associated normalized optimality gap is only -0.074. By contrast, the achievable ESE when using the ESEM algorithm also saturates at around $P^B_{max} = 40$ dBm and $P^R_{max} = 40$ dBm.[22] Thus, the operating point of "$P^B_{max} = 40$ dBm and $P^R_{max} = 40$ dBm" may strike an attractive balance between SEM and ESEM. Of course, the required trade-off may be struck on a case-by-case basis in practical systems.

Additionally, the RG-EPA algorithm performs significantly worse in terms of SE when compared to the SEM algorithm, and in terms of ESE when compared to the ESEM algorithm. Furthermore, the RG-EPA algorithm performs even worse than the SEM algorithm in terms of ESE. Although the obtained SE when using the RG-EPA algorithm is, in some cases, higher than the SE obtained when using the ESE algorithm, this performance improvement comes at a great cost to the ESE performance of the RG-EPA algorithm.

[22]Note that when P^B_{max} and P^R_{max} have low/moderate values, the SEM and ESEM algorithms share the same solutions of $\mathcal{P}$ and $\mathcal{S}$.

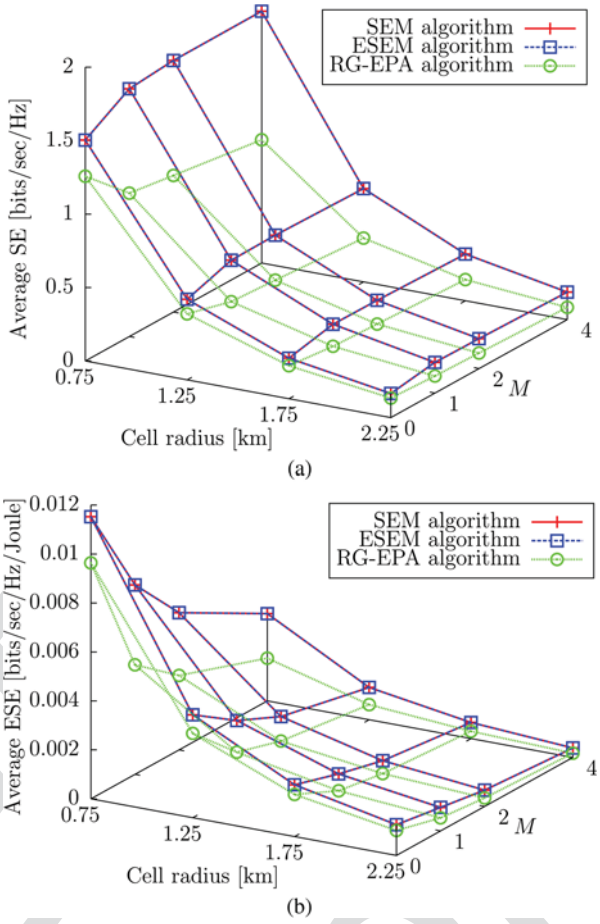

Fig. 6. The average achievable SE and ESE of the SEM, ESEM and RG-EPA algorithms upon varying M and cell radius, and using the parameters in Table I with $N = 6$, $K = 10$, $\alpha = 0.1$, $P^B_{max} = 20$ dBm and $P^R_{max} = 10$ dBm. (a) Surface plots of the achievable SE when using the SEM, ESEM and RG-EPA algorithms. (b) Surface plots of the achievable ESE when using the SEM, ESEM and RG-EPA algorithms.

Finally, note that although both the SE of the SEM algorithm, and the ESE of the ESEM algorithm are non-decreasing as either P^B_{max} or P^R_{max} is increased, the effect of increasing P^B_{max} on the SE or ESE is significantly more pronounced, than that of applying the same increase to P^R_{max}. The intuitive reasoning behind this is that the power available at the BS has a more pronounced effect on the system's performance, since the direct links and, more importantly, the BS-to-RN links rely on the BS. Therefore, increasing P^R_{max} is futile if the BS-to-RN links are not allocated sufficient power to support the RN-to-UE links.

C. The Achievable SE and ESE as a Function of M and the Cell Radius

Fig. 6 illustrates some advantages and disadvantages of employing RNs in the cellular system considered. We observe that the specific low values of the power constraints result in the same solutions for both the SEM and ESEM algorithms. This phenomenon was also shown in Fig. 5.

As evidenced in Fig. 6(a), the attainable SE increases with M, which is a benefit of the additional selection diversity, when forming relaying links. However, the attainable SE does not increase substantially beyond $M = 2$. In fact, only an increase of 0.1% is attained for the SE when M is increased from 2 to 4 at a cell radius of 0.75 km. On the other hand, the cost in terms of ESE is significant (36.4%), as shown in Fig. 6(b). This suggests that employing RNs does not constitute an energy-spectral-efficient technique although it increases the SE of a cellular system, which is partially due to the power amplifier inefficiency and owing to the non-negligible fixed circuit energy dissipation. Note furthermore that both the attainable SE and ESE are decreasing upon increasing the cell radius as a result of the increased path-loss of all the wireless links. However, this reduction is relatively small between a cell radius of 1.75 km and 2.25 km. The reason behind this phenomenon is that both the SEM and ESEM algorithms will selectively serve the UEs nearer to the BS, so that a similar performance may be attained without suffering from a substantial path-loss. This is also the reason why the gain in SE gleaned by employing RNs at a cell radius of 2.25 km seems negligible in Fig. 6(a). Once again, the RG-EPA algorithm performs worse both in terms of SE and ESE performance.

D. The Achievable SE and ESE as a Function of N and N_B

Fig. 7 illustrates the effect of increasing N and N_B on the attainable SE and ESE. Note that in a similar fashion to Fig. 6, the SEM and ESEM algorithms attain the same solutions in the operating region considered.

Observe from both Fig. 7(a) and (b) that the attainable SE and ESE increase upon increasing N_B. This is due to the increased attainable spatial degrees of freedom at the BS in the first transmission phase, which allows for more direct transmissions overall. However, both the SE and ESE are reduced upon increasing N, which suggests that increasing the number of subcarrier blocks does not increase the average efficiency of each block. This is because the power constraints are fixed and thus there is insufficient power for fully exploiting the additional subcarrier blocks. However, note that both total SE and ESE do indeed increase upon increasing N, which may be explicitly seen upon multiplying the results of Fig. 7(a) and (b) by NW. The RG-EPA algorithm performs worse in both cases as expected.

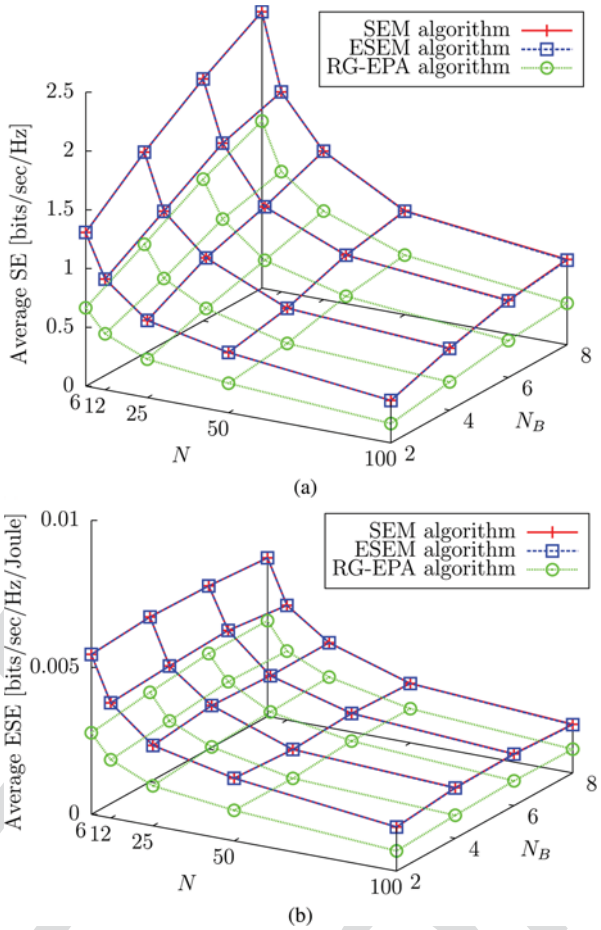

Fig. 7. The average achievable SE and ESE of the SEM, ESEM and RG-EPA algorithms upon varying N and N_B, and using the parameters in Table I with $M = 2$, $K = 10$, $\alpha = 0.1$, $P^B_{max} = 20$ dBm, $P^R_{max} = 10$ dBm and a cell radius of 0.75 km. (a) Surface plots of the achievable SE when using the SEM, ESEM and RG-EPA algorithms. (b) Surface plots of the achievable ESE when using the SEM, ESEM and RG-EPA algorithms.

VII. Conclusion and Future Work

In this paper, firstly a novel transmission protocol based on joint transmit-BF and receive-BF was developed for the multi-relay MIMO-OFDMA cellular network considered. This protocol allows for achieving high-SE performance for the MIMO broadcast network consisting of a BS, multiple RNs and multiple UEs. The associated MIMO channel matrices were mathematically decomposed into multiple MISO channels, which we referred to as SMCs, using receive-BF. By applying ZFBF at the transmitter, the interference between SMC-based concurrent transmissions is completely eliminated, provided that perfect CSI-knowledge is available. For the purposes of obtaining a higher multiplexing gain, the SMCs may be grouped according to the semi-orthogonality criterion. Consequently, a pair of grouping algorithms were proposed, referred to as ESGA and OCGA. The former exhaustively enumerates all of the possible groupings, whereas the latter aims to be a lower-complexity design alternative. Finding the SE-optimal and ESE-optimal SMC groupings as well as their associated optimal power control variables were formulated as optimization problems. With the aid of several variable relaxations and transformations, these optimization problems were transformed into concave optimization problems. Thus, the dual decomposition approach was employed for finding the optimal solutions. We demonstrated that the OCGA constitutes an attractive alternative to ESGA, since it offers a near-optimal performance at a substantially reduced complexity. Furthermore, several numerical results were presented for characterizing the system's attainable SE and ESE performance across a wide range of system parameters, such as the transmit power constraints, cell radius, the number of RNs, the number of BS antennas and the number of subcarrier blocks. Additionally, we demonstrated that our SEM/ESEM algorithms perform significantly better than the benchmark RG-EPA algorithm.

In our future work, we will consider unity frequency reuse multi-relay multi-cell networks. Thus, these networks are interference-limited, rather than noise-limited. Consequently, improved transmission protocols and optimization methods are required for managing both the intra-cell and inter-cell interference to improve the system's SE and ESE performance.

REFERENCES

[1] M. Salem *et al.*, "An overview of radio resource management in relay-enhanced OFDMA-based networks," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 3, pp. 422–438, Apr. 2010.

[2] L. Hanzo, Y. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WIFI and WIMAX: Coherent Versus Non-Coherent and Cooperative Turbo-Transceivers*. Hoboken, NJ, USA: Wiley, 2010, IEEE Press.

[3] C. Han *et al.*, "Green radio: Radio techniques to enable energy-efficient wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 46–54, Jun. 2011.

[4] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[5] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[6] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[7] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[8] N. Ul Hassan and M. Assaad, "Low complexity margin adaptive resource allocation in downlink MIMO-OFDMA system," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3365–3371, Jul. 2009.

[9] G. Raleigh and J. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.

[10] K.-K. Wong, R. Murch, and K. Letaief, "A joint-channel diagonalization for multiuser MIMO antenna systems," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 773–786, Jul. 2003.

[11] W. Ho and Y.-C. Liang, "Optimal resource allocation for multiuser MIMO-OFDM systems with user rate constraints," *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1190–1203, Mar. 2009.

[12] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[13] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.

[14] D. Ng, E. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3618–3631, Oct. 2012.

[15] G. Miao, N. Himayat, and G. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.

[16] G. Miao, N. Himayat, G. Li, and S. Talwar, "Low-complexity energy-efficient scheduling for uplink OFDMA," *IEEE Trans. Commun.*, vol. 60, no. 1, pp. 112–120, Jan. 2012.

[17] K. T. K. Cheung, S. Yang, and L. Hanzo, "Achieving maximum energy-efficiency in multi-relay OFDMA cellular networks: A fractional programming approach," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2746–2757, Jul. 2013.

[18] Z. Shen, R. Chen, J. Andrews, R. Heath, and B. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sep. 2006.

[19] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.

[20] E. Lo *et al.*, "Adaptive resource allocation and capacity comparison of downlink multiuser MIMO-MC-CDMA and MIMO-OFDMA," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 1083–1093, Mar. 2007.

[21] D. Ng, E. Lo, and R. Schober, "Dynamic resource allocation in MIMO-OFDMA systems with full-duplex and hybrid relaying," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1291–1304, May 2012.

[22] G. Brante, I. Stupia, R. D. Souza, and L. Vandendorpe, "Outage probability and energy efficiency of cooperative MIMO with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5896–5907, Nov. 2013.

[23] A. Zappone, P. Cao, and E. Jorswieck, "Energy efficiency optimization in relay-assisted MIMO systems with perfect and statistical CSI," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 443–457, Jan. 2014.

[24] M. Avriel, W. E. Diewert, S. Schaible, and I. Zang, *Generalized Concavity*. New York, NY, USA: Plenum, 1988.

[25] R. Devarajan, S. Jha, U. Phuyal, and V. Bhargava, "Energy-aware resource allocation for cooperative cellular network using multi-objective optimization approach," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1797–1807, May 2012.

[26] W. Dinkelbach, "On nonlinear fractional programming," *Manag. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[27] D. Ng, E. Lo, and R. Schober, "Energy-efficient resource allocation for secure OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2572–2585, Jul. 2012.

[28] C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2946–2957, Aug. 2012.

[29] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[30] "Further advancements for e-utra, physical layer aspects (Release 9)," Sophia-Antipolis Cedex, France, TR 36.814 V9.0.0, Mar. 2010.

[31] S.-J. Kim, A. Magnani, A. Mutapcic, S. Boyd, and Z.-Q. Luo, "Robust beamforming via worst-case SINR maximization," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1539–1547, Apr. 2008.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

[33] L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-Capacity Multi-Functional MIMO Systems: Sphere-Packing, Iterative Detection and Cooperation*. Hoboken, NJ, USA: Wiley, 2009, IEEE Press.

[34] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545–1553, Jul. 2002.

[35] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. New York, NY, USA: Academic, 2010.

[36] G. Auer *et al.*, "D2.3: Energy Efficiency Analysis of the Reference Systems, Areas of Improvements and Target Breakdown," Tech. Rep. INFSO-ICT-247733, Nov. 2010. [Online]. Available: https://bscw.ict-earth.eu/pub/bscw.cgi/d71252/EARTH_WP2_D2.3_v2.pdf **AQ1**

[37] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.

[38] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.

[39] K. Seong, M. Mohseni, and J. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *Proc. IEEE ISIT*, Seattle, Washington, USA, Jul. 2006, pp. 1394–1398.

[40] D. Ng and R. Schober, "Cross-layer scheduling for OFDMA amplify-and-forward relay networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1443–1458, Mar. 2010.

[41] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[42] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *Proc. Future Netw. Mobile Summit*, Florence, Italy, Jun. 2010, pp. 1–8.

Kent Tsz Kan Cheung (S'09) received the B.Eng. degree (first-class honors) in electronic engineering from the Univeristy of Southampton, Southampton, U.K., in 2009. Since then he has been pursuing the Ph.D. degree in wireless communications at the same institution. He was a recipient of the EPSRC Industrial CASE award in 2009, and was involved with the Core 5 Green Radio project of the Virtual Centre of Excellence in Mobile and Personal Communications (Mobile VCE).

His research interests include energy-efficiency, multi-carrier MIMO communications, cooperative communications, resource allocation and optimization.

Shaoshi Yang (S'09–M'13) (https://sites.google.com/site/shaoshiyang/) received the B.Eng. degree in information engineering from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in Jul. 2006, the first Ph.D. degree in electronics and electrical engineering from University of Southampton, U.K., in Dec. 2013, and the second Ph.D. degree in signal and information processing from Beijing University of Posts and Telecommunications (BUPT) in Mar. 2014. He is now working as the IU-ATC Senior Research Fellow in University of Southampton, U.K.

From November 2008 to February 2009, he was an Intern Research Fellow with the Communications Technology Lab (CTL), Intel Labs, Beijing, China, where he focused on Channel Quality Indicator Channel (CQICH) design for mobile WiMAX (802.16 m) standard. His research interests include Multiple-Input–Multiple Output (MIMO) signal processing, multicell joint/distributed signal processing, cooperative communications, green radio and cross-layer interference management. He has published more than 30 research papers on IEEE journals and conferences.

Shaoshi received a number of awards, such as the PMC-Sierra Telecommunications Technology Scholarship, the Electronics and Computer Science Research Studentship and the China Scholarship Council Scholarship. He is a member of IEEE/IET, and a junior member of Isaac Newton Institute for Mathematical Sciences, Cambridge, U.K. He is also a TPC member of several major IEEE journals/conferences, including PIMRC 2012, ICC 2013, ICCVE 2013, ICCVE 2014, ICC 2015 and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS—SPECIAL ISSUE ON RECENT ADVANCES IN HETEROGENEOUS CELLULAR NETWORKS.

Lajos Hanzo (F'XX) (http://www.cspc.ecs.soton.ac.uk) FREng, FIEEE, FIET, Fellow of EURASIP, DSc received the degree in electronics in 1976 and the doctorate in 1983. In 2009 he was awarded the honorary doctorate *Doctor Honoris Causa* by the Technical University of Budapest. **AQ2**

During his 37-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 80+ PhD students, co-authored 20 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10000 pages, published 1400+ research entries at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 100-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award.

He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008–2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. His research is funded by the European Research Council's Senior Research Fellow Grant. For further information on research in progress and associated publications please refer to http://www.cspc.ecs.soton.ac.uk Lajos has 20000+ citations.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1 = Please provide organizational address [36].
AQ2 = Please provide membership history of author Lajos Hanzo.

END OF ALL QUERIES