This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMC.2020.3012451

Distributed Energy Efficient Channel Allocation in Underlay Multicast D2D Communications

Mariem Hmila, Manuel Fernández-Veiga, *Senior Member, IEEE*,
Miguel Rodríguez-Pérez, *Senior Member, IEEE*, and Sergio Herrería-Alonso

Abstract—In this paper, we address the optimization of the energy efficiency of underlay multicast device-to-device (D2MD) communications on cellular networks. In particular, we maximize the energy efficiency of both the global network and the individual users considering various fairness factors such as maximum power and minimum rate constraints. For this, we employ a canonical mixed-integer non-linear formulation of the joint power control and resource allocation problem. To cope with its NP-hard nature, we propose a two-stage semi-distributed solution. In the first stage, we find a stable, yet sub-optimal, channel allocation for D2MD groups using a cooperative coalitional game framework that allows co-channel transmission over a set of shared resource blocks and/or transmission over several different channels per D2MD group. In the second stage, a central entity determines the optimal transmission power for each user in the system via fractional programming. We performed extensive simulations to analyze the resulting energy efficiency and attainable transmission rates. The results show that the performance of our semi-distributed approach is very close to that obtained with a pure optimal centralized one.

Index Terms—Game theory, fractional programming, coalition formation, D2D multicast communication, 5G wireless networks.

1 Introduction

Device-to-device (D2D) communications allow user equipment (UE) devices in close proximity to realize a direct communication without signal relay through an access point. The 3GPP started the standardization on D2D in LTE Release 12 [1] (proximity services, ProSe) for public safety, which was further enhanced with device-based relaying in Release 13, and has subsequently evolved toward the specification of the NR Sidelink in Releases 16 & 17 for 5G and Beyond5G wireless networks [2]. Short range D2D communications attain high data transmission rates with low delays, and improve energy efficiency and radio resource utilization [3], [4]. A broad range of use cases, from proximity-aware services, to public safety communications, vehicle-to-X (V2X) communications and the Internet of Things (IoT) will benefit from the D2D technology.

In applications such as V2X or infotainment, the same piece of data might be requested by a group of users. This multicast scenario may be implemented either as cellular multicast, i.e., the evolved Node-B (eNB) multicasts the content to the group [5], [6], or as mobile data offloading [7], [8], [9], diverting spatially local traffic to other networks such as a D2D side channel. Both approaches reduce the spectrum usage compared to when every user receives through a dedicated channel.

As in cellular multicast, D2D has a natural multicast generalization. In multicast device-to-device (D2MD) communications, a cluster of users with a common interest on a particular content forms a group wherein any UE can act as a transmitter for the rest of the group members. Besides data dissemination, multicast D2D faces its own challenges as to neighbor discovery and cluster membership [10], selection of the head cluster (transmitter UE) [11], and transmission to the weakest receiver, since all the UEs in the cluster must decode the same data stream. At the physical layer, both D2D and D2MD communications can take place either in-band or out-band [12]. In the in-band case, transmitters share the licensed spectrum with other cellular users (underlay communication) while, in the out-band mode, part of the bandwidth is reserved for D2D or D2MD (overlay communication). The gains in spectral efficiency and area throughput due to D2D or D2MD have been extensively analyzed in the literature (e.g., see [9], [13] for recent accounts). For the performance advantages of multicast over unicast D2D, see [14]. In the current 5G standard, NR Sidelink supports transmission in carriers shared with NR or LTE in the licensed spectrum; autonomous UE operation, where the UE senses and selects resources on the sidelink based on network configuration; and groupcast communications too.

Energy efficiency (EE) is another of the recognized key performance indices in 5G networks, and is heavily dependent on the used radio interfaces and the average distance for transmissions [15].EE is defined as the normalized data transmission rate (in bit/s/Hz) divided by the amount of energy used for achieving that reliable transmission rate [16], either from a global network perspective or for an individual user. Since underlay D2MD techniques use less energy in their localized transmissions, they should clearly improve EE too, both locally and globally. However, underlay D2MD also causes co-channel interference, thus hindering the achievable transmission rates and putting a limit on EE. The goal of this paper is to devise efficient interference management

- *M. Hmila, M. Fernández-Veiga, M. Rodríguez-Pérez, and S. Herrería-Alonso are with the atlanTTic Research Center, University of Vigo (Spain)*

This work was supported by the European Regional Development Fund (ERDF) and the Galician Regional Government under agreement for funding the atlanTTic Research Center for Telecommunication Technologies, and by the "Ministerio de Economia, Industria y Competitividad" through grant TEC2017-85587-R, "Programa Estatal de Investigación, Desarrollo e Innovación Orientada a los Retos de la Sociedad" (partly financed with FEDER fundis).

techniques for the successful introduction of underlay D2MD communication in 5G cellular networks.

In our previous work [17], we proposed a centralized solution for optimizing the EE both at the global network level and for individual users in D2MD communications. In this paper, which extends the results previously presented in [18], we build on the same system model to tackle the EE optimization problem with a semi-distributed approach. Here, a two-stage decomposition procedure is adopted. In the first one (the resource allocation sub-problem), a sub-optimal channel allocation is decided, that is, a stable and efficient assignment of resource blocks (RBs) to a set of D2MD groups, based on the outcome of an overlapping coalitional game. We consider new resource sharing cases in which (1) co-channel transmission over a set of shared RBs; and (2) transmission over several different channels per multicast group are allowed simultaneously. Moreover, differently from [17], we also consider that both cellular users and the transmitters in each cluster are rational and independent. Thus, cellular users can choose which clusters to share the RBs with, and vice versa, based on just local information. In the second stage (the power control sub-problem), the optimal transmission power for all the users in the system is obtained by a central entity via fractional programming as in [16], [17]. We compare our solution with the centralized approach previously proposed in [17] and an up-to-date fully distributed matching algorithm [19]. The numerical results show that the performance of the proposed semi-distributed approach is close to the optimal and better than that obtained with fully distributed approaches due to the interplay of the interference-based process of coalition formation and the use of minimal transmission powers.

The rest of the paper is organized as follows. Sect. 2 discusses the related work and Sect. 3 describes the system model. In Sect. 4 we present the coalitional game theory and adapt it to the distributed resource allocation setting. Then, in Sect. 5 we present in detail the coalition formation algorithm required for channel allocation. Simulation results are discussed in Sect. 6. Finally, the main conclusions are laid out in Sect. 7.

2 Related Work

Interference management techniques required in underlay D2MD communications can be broadly classified as centralized or distributed. In the first type, a central scheduler—an access point—manages the pairing between cellular users and D2D clusters to optimize a target metric of interest, like the sum-rate or the energy efficiency. Additionally, it can impose some constraints on the transmission power to limit mutual interference and guarantee a target service level (e.g., a minimum SINR).

A centralized optimization approach for D2MD is applied in [20] for casting the power control (PC) problem as a multi-objective optimization with weighted factors to minimize energy consumption and maximize the number of served links. Joint resource allocation (RA) and PC is considered in [21] to maximize the total EE achieved by all the clusters. In that work, the feasible RBs are first identified based on a threshold on the outage probability, and then a PC algorithm is run to operate the system with the highest EE possible for each feasible channel. Centralized system throughput maximization is canonically formulated in [22] as a Mixed Integer Non-linear Programming (MINLP) problem, subject to maximum power and quality of Service (QoS) constraints for both cellular and D2D users. Due to NP-hardness of the problem, this is decomposed into coupled RA and PC sub-problems. D2MD clusters are arranged based on their individual contribution to the sum throughput when sharing a certain cellular user (CU) resource block (RB). Then, clusters far from each other are allowed to share the same RB as long as the QoS constraints are satisfied. A second RA scheme based on outage probability is also proposed there, but the RA is done assuming the worst case that the D2MD clusters transmit at their maximum allowed power. An incremental approach is adopted in [23], assuming that the QoS levels of CUs are met before the activation of D2MD into the cell. The latter is allowed only if it is still possible to guarantee the QoS requirements among the CUs and the D2MD cluster. As in [22], all users are allowed to transmit using their maximum power, so the system is forced to work in the strong interference-limited regime. D2MD is restricted to zones in the cell area where interference is not too severe, and the best pairing between channels (RBs) and D2MD groups is found by means of a local search procedure (tabu search). An increase in the overall system throughput is demonstrated through numerical experiments.

The effect of spatial and temporal correlation in device mobility is considered in [24]. This work models the problem as a constrained multi-objective optimization (maximize the number of D2D multicast links per resource block and the network energy efficiency) that is solved using a standard evolutionary algorithm (Non-dominated Sorting Genetic Algorithm-II, or NSGA-II). However, the method is sensitive to the tuning of the search algorithm and the trade-off between the performance objectives has to be resolved by a external decision maker. The idea in [25] is to solve the allocation of channels to groups by attempting to optimize the SINR by means of a maximum-weight bipartite matching between channels and D2MD groups, with the Hungarian algorithm, where the weight is the SINR. This SINR relies in turn on a power control algorithm that belongs to the class of fractional programming problems. Further, [25] tests several centralized and distributed clustering algorithms for defining the D2MD groups. Instead of a system-dependent quality signal indicator like the SINR, [26] uses the individual channel gains to allocate the RBs to clusters, in combination with a priority mechanism for assigning the groups having low energy efficiency to the higher-gain channels. Yet not only are the channel coefficients difficult to estimate, even by a central entity, but also the allocation rule is heuristic in nature. A similar centralized, heuristic algorithm is followed by [27], this time focused on maximizing the total EE of the clusters.

Clearly, the assumption that the central controller has complete, accurate information about the users' transmission parameters, channels qualities, etc., and, therefore, can synthesize the optimal solution is unrealistic and complex, for its huge overhead and computational load [28]. Unlike these, in decentralized approaches network devices are responsible for taking decisions concerning the transmission power and resource allocation based only on local observations.

Using non-cooperative games, the authors of [29] propose a resource allocation scheme to maximize the EE of individual users, but only for a setting with a fixed number of users. The same model is used in [30] to study the trade-off between energy and spectral efficiency. There, an iterative algorithm is presented to solve the problem using Nash equilibrium as the finishing condition. Another example of non-cooperative game appears in [31] for downlink RB sharing. Here, a two-level approach was investigated to maximize the number of served D2D users while guaranteeing the quality of service for CUs. Those D2D users who have minimum mutual interference are grouped together, assuming that all transmit at their maximum power. Then, a Stackelberg game is used to allocate the transmission power to them. Cooperative game theory is used to solve the joint power and resource allocation problem in [32]. The model is an overlapping coalition formation game, where devices are grouped into sets (coalitions) composed of a single CU and multiple D2D pairs. Merge and split rules for the coalition members are set up so as to maximize system throughput. Simulation results show that even though the coalition structure attains stability and achieves an optimal solution, it can be unfair for some D2D users. For this reason, a fair D2D resource allocation algorithm is introduced.

While all these works focus on D2D communication, the authors in [19] consider a D2MD model where resource allocation is done using matching theory, whereas power control is solved using fractional programming. The main goal here is to maximize the EE of individual users. Differently, authors in [33], build an underlay D2D multicast real-time video distribution framework. Physical distance is crucial for communication QoS, thus a directed hyper-graph based on distance among users is established to highlight interfering devices and analyse physical location importance. In addition, a complementary metric of distance is introduced to estimate the closeness of UEs in the social domain. In this work, out of range receivers and transmitters share the same RB assigned by the eNB. Later, a multi-leader multi-follower Stakelberg game and a Q-learning-based strategy are used to determine the users' transmission power in order to maximize the network throughput. Unfortunately, with this algorithm it is not clear how many users are sharing the same resource block or what the average distance between a transmitter and a receiver is. Another limitation is that the channel gain between the eNB and a specific user and D2D users link gains are predefined.

The approaches previously discussed can be more neatly categorized according to their degree of shared resources. Let us call *reuse factor* (r) to the number of D2D pairs or D2MD clusters per RB (CU channel), and *split factor* (s) to the number of RBs (CU channels) assigned to a D2D pair or a D2MD cluster. Then we can distinguish four cases:

Dedicated CUs ($r = s = 1$): each D2D pair/D2MD group can only use one CU channel, and each channel supports only a single D2D pair/D2MD group, at most.

Distributed Groups ($s > 1$, $r = 1$): a D2D pair/D2MD group can distribute its messages over distinct CU channels, but a CU cannot share its resources with more than one D2D pair/D2MD group.

Shared CUs ($s = 1$, $r > 1$): a D2D pair/D2MD group can use a single CU channel. However, the channel can be shared among at most r D2D pairs/D2MD groups.

General Case ($s > 1$, $r > 1$): a CU channel can be shared among r D2D pairs/D2MD groups. In addition, a D2D pair/D2MD group can use s different CU channels. In this case, inter-users interference might limit the benefits of D2D pairs/D2MD groups coexistence and reduce spectral and energy efficiency.

Table 1 summarizes the related work discussed in this Section using the proposed taxonomy. Compared with the state-of-the-art, the algorithm presented in this paper introduces the following contributions:

1) We model the up-link resource allocation sub-problem in underlay D2MD communications as an overlapping coalition formation game. The proposed game-theoretic framework uses the ideas of reuse and split factors, and allows to analyze the system behavior in terms of EE for both the system as a whole (global energy efficiency, Section 3.2) and for individual users (max-min energy efficiency, Section 3.1).
2) We develop a distributed coalition formation algorithm that permits D2MD networks to flexibly self-organize themselves into a stable partition and to adapt to environmental changes (Section 5).
3) The distributed model explicitly considers QoS constraints during the coalition formation phase, including maximum power and minimum rate constraints (Section 5).
4) The power allocation sub-problem is optimally solved using Dinkelbach's algorithm. Here, a central entity must be in charge of applying it, thus combining the advantages of distributed and centralized solutions.
5) We evaluate the performance for random networks in varied stochastic spatial configurations (Section 6). The results are compared with the centralized approach in [17] to show near optimality, and with an alternative distributed matching algorithm [19].

3 System Model & Optimization

We focus on a single cell/tier network with one central entity and several users randomly distributed over the cell coverage area, as illustrated in Fig. 1 [17]. On the up-link, M CU users transmit on M orthogonal communication sub-channels or RBs. We assume that D2MD users are grouped into K multicast clusters $\mathcal{D}_k$, $k = 1, \ldots, K$, that can reuse the same communication channels allocated to the CUs for direct communication among their members. Each of these D2MD groups has only one designated transmitter and comprises, therefore, $|\mathcal{D}_k| - 1$ receivers.[1] In this model, the base station suffers the interference caused by the co-channel D2MD transmitters, while the receivers in a D2MD group are affected by the interference caused by the CU user and, possibly, the other transmitters of those D2MD groups sharing the same RB. We briefly reproduce here the two optimization problems to be solved.

1. The case $|\mathcal{D}_k| = 2$ corresponds to a simple unicast communication or D2D pair or dipole.

Table 1
Centralized and Distributed Approaches in D2D and D2MD: State of the Art

Ref	Scenario	Approach	Category	Model	Problem	Objective
[20]	Shared CUs	Optimization	Centralized	D2MD	RA, PC	Spectral efficiency
[21]	General Case	Optimization	Centralized	D2MD	PC	D2MD EE
[27]	Dedicated CUs	Heuristic algorithm	Centralized	D2MD	PC, RA	D2MD EE
[22]	Shared CUs	Graph theory, corner search, STIM	Centralized	D2MD	RA, PC	System throughput
[23]	Dedicated CUs	Tabu search algorithm	Centralized	D2MD	RA, PC	System data rate
[24]	Shared CUs	Multi-objective optimization, NSGA-II	Centralized	D2MD	RA, PC	SE, network EE
[25]	Shared CUs	Optimization, graph theory	Centralized	D2MD	RA, PC	Max SINR
[26]	Dedicated CUs	Optimization	Centralized	D2MD	RA, PC	Min D2MD EE
[29], [30]	General Case	Non-cooperative	Distributed	D2D	RA, PC	Users EE
[31]	General Case	Stackelberg game	Distributed	D2D	PC	System capacity
[32]	Shared CUs	Coalition formation	Distributed	D2D	RA, PC	Coalition sum SINR
[19]	Shared CUs	Matching theory	Distributed	D2MD	RA, PC	Individual EE
[33]	Shared CUs	Multi-leader multi-follower Stackelberg game	Semi-distrib.	D2MD	RA, PC	Network throughput
This paper	General Case	Coalition formation + optimization	Semi-distrib.	D2MD	RA,PC	Global EE, max-min EE

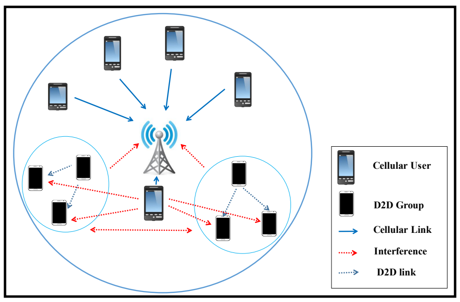


Figure 1. System model.

3.1 MEE Problem — Max-Min Energy Efficiency

The energy efficiency (in bit/Hz/J) for each CU user m is defined as the ratio of its normalized transmission rate to the energy it consumes:

$$\eta_m \triangleq \frac{r_m}{\tau_m + p_m}, \quad m = 1, \dots, M, \tag{1}$$

where r_m is the normalized transmission rate (in bit/s/Hz) of CU m, τ_m is the power consumed by CU m at rest (i.e., when there is nothing to transmit) and p_m is its transmission power. The Shannon's ergodic capacity formula $r_m = \log_2(1+\gamma_m)$ is used for the rate, where γ_m denotes the SINR for CU m.

Similarly, the energy efficiency for each D2MD group $\mathcal{D}_k$ is defined as

$$\zeta_k \triangleq \frac{R_k}{\tau_k + \sum_m P_k^{(m)}}, \quad k = 1, \dots, K, \tag{2}$$

where R_k is the aggregated received rate in group $\mathcal{D}_k$, τ_k is the power used by all the devices in group $\mathcal{D}_k$ at rest and $P_k^{(m)}$ is the transmission power allocated to the designated transmitter of group $\mathcal{D}_k$ over channel m. Note that R_k is constrained by the weakest receiver, i.e., the one with the poorest channel quality, and depends on the number of members of the group:

$$R_k = |\mathcal{D}_k| \sum_{m=1}^{M} x_{k,m} \min_{i \in \mathcal{D}_k} \log_2(1+\gamma_{i,m}), \tag{3}$$

where $x_{k,m}$ is 1 if group $\mathcal{D}_k$ transmits over channel m (or 0, otherwise) and $\gamma_{i,m}$ is the SINR for user i in group $\mathcal{D}_k$ over channel m. We also assume that both CU users and D2MD groups must satisfy individual average power and transmission constraints:

$$p_m \leq \overline{p}_m, \quad m = 1, \dots, M, \tag{4}$$

$$\sum_m P_k^{(m)} \leq \overline{P}_k, \quad k = 1, \dots, K, \tag{5}$$

$$r_m \geq \underline{r}_m, \quad m = 1, \dots, M, \tag{6}$$

$$R_k \geq \underline{R}_k, \quad k = 1, \dots, K. \tag{7}$$

Here, $\overline{p}_m$ (resp., $\overline{P}_K$) denote the maximum transmission power for the CUs (resp., D2D UEs), and $\underline{r}_m$ (resp., $\underline{R}_k$), the minimum acceptable received rate for the CUs (resp., D2D UEs). Now, we can formulate the max-min energy efficiency problem as

$$\text{MEE} = \max_{\mathbf{p}, \mathbf{P}_k, \mathbf{X}} \{\min\{\min_m \eta_m, \min_k \zeta_k\}\}, \tag{8}$$

where $\mathbf{p} = (p_1, \dots, p_M)$ is the power vector allocated to the CUs, $\mathbf{P}_k = (P_k^{(1)}, \dots, P_k^{(M)})$, $k = 1, \dots, K$, is the power vector allocated to the designated transmitter of group $\mathcal{D}_k$ over the M channels and $\mathbf{X} = [x_{k,m}]$ is the $K \times M$ channel allocation matrix containing all the indicator variables.

3.2 GEE Problem — Global Energy Efficiency

The global energy efficiency (GEE) of the cellular network is simply the ratio between the aggregated rate and the total power needed, so we can formulate the optimization problem for maximizing the energy efficiency of the whole system as

$$\text{GEE} = \max_{\mathbf{p}, \mathbf{P}_k, \mathbf{X}} \frac{\sum_m r_m + \sum_k R_k}{\tau + \sum_m p_m + \sum_k \sum_m P_k^{(m)}}, \tag{9}$$

where τ is the total power consumed by all the devices in the network at rest.

4 Overlapping Coalition Game for Channel Allocation

Our approach to simplify the joint resource and power allocation problem is to resort to a classical decomposition: i) a decision-making sub-problem concerning the assignment of RBs to D2MD groups; and ii) an optimization problem for finding the power vector for all the users in the system.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMC.2020.3012451

Here, we tackle the resource allocation sub-problem as a game among the users of the cellular network. Our *players* (the CUs and the D2MD groups) will collaborate with each other in order to minimize the aggregated interference which in turn leads to maximize either the global energy efficiency or the minimum user energy efficiency. Coalitional games are suitable for designing fair, robust, practical and efficient cooperation strategies on communication networks where the players share a common objective. Ultimately, the algorithmic solution to this sub-problem will be *distributed*, while the solution to the whole problem may be considered as *semi-distributed* because, for the power allocation algorithm, we still keep an optimal, centralized approach.[2]

4.1 Overlapping Coalitional Games

Coalitional games allow users to form groups, commonly referred to as *coalitions*. A coalition is useful when several players share a common objective and there exist in the problem structure positive externalities from collaboration among them. In our context, we assume that both CU users and D2MD groups are willing to cooperate to properly share the RBs and achieve higher global and individual EE. A coalition game **G** is formally defined as follows [34].

Definition 1. *A coalition game* $\mathbf{G} = (\mathcal{N}, \mathsf{v}, \mathcal{S})$ *is defined by:*

1) *The set of players* $\mathcal{N}$. *In our setting,* $\mathcal{N} = \mathcal{M} \cup \mathcal{K}$, *with* $\mathcal{M}$ *and* $\mathcal{K}$ *being the sets of CU users and D2MD groups, respectively.*
2) *The partition function* v *that quantifies the worth of a coalition in the game. It associates to each subset* $\mathcal{S}_i$ *of* $\mathcal{N}$ *a real number quantifying the objective value of the coalition.*
3) *The coalition structure* $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_n\}$. *This is the set of formed coalitions, where each coalition* $\mathcal{S}_i$ *is a subset of* $\mathcal{N}$ $(\mathcal{S}_i \subseteq \mathcal{N}, i = 1, \ldots, n)$.

Note in the above definition that neither: (i) $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$ nor (ii) $\cup_{i=1}^{n} \mathcal{S}_i = \mathcal{N}$ are enforced. Therefore, a coalitional game can have overlapped coalitions (condition (i) fails), or have players not included in any coalition (condition (ii) fails). The latter is a simple matter of terminology, since single players can be equally regarded as individual coalitions.

For any partition function v, coalitional games can also have transferable utilities, that is, users can transfer losslessly part of their utilities to another players. In this case, the value of a coalition only depends on the coalition structure (i.e., its members), and not on how the rest of the players have been partitioned. This property makes the game adapt to a more group-rational perspective [34], [35], and for that reason it will be assumed here. Also, it is worth to mention that our coalition game is non-superadditive, so that users will never form a grand coalition since, in such case, the high interference value would force players to act selfishly as in non-cooperative models.

4.2 Coalition Formation

We consider that each coalition is composed of a single CU user (the RB to be shared), and one or more D2MD groups aiming to share the RB with the CU. In this manner, a coalition can be seen as an agreement among various players to act as a single entity for achieving a higher coalition value v. According to the different values that can take the split (s) and the reuse (r) factors, we can distinguish up to three different forms of coalitions:

1) The simplest case occurs when $s = r = 1$. Each coalition is composed of a single CU user and a single D2MD group.
2) The second case appears when $s = 1$ with $r > 1$. In this scenario, each coalition is composed of a single CU and multiple D2MD groups. Note that in this and the former case, **G** remains a traditional coalition formation game.
3) In the third case, we have $s > 1$ and $r \geq 1$, so each D2MD group can be included in different coalitions.[3]

Certainly, a more flexible coalition structure allows players to distribute their resources (time, energy, or money) among the various coalitions they are part of. Here, the transmitter of each D2MD group $\mathcal{D}_k$ has a limited transmission power budget ($\overline{P}_k$) which can be split over up to s channels. As a result, the model turns out to be an **Overlapping Coalition Formation Game** (**OCFG**). The overlapping model leads to better organized coalitions and possibly higher payoffs [36], [37]. Recall that **OCFGs** hold all the properties of a traditional coalition game.

4.3 Partition Function

In our game, CU users and D2MD groups must cooperate to minimize the total co-channel interference among them, thus increasing the system and their own energy efficiency. According to this, the partition function $\mathsf{v}(\mathcal{S}_m)$ for coalition $\mathcal{S}_m$ will represent the total mutual interference among the coalition members. To calculate this value, we must compute the received interference on each side.

First, the interference on each D2MD group $\mathcal{D}_k$, $k = 1, \ldots, K$, in channel $m = 1, \ldots, M$, is determined by its weakest receiver as

$$\alpha_{k,m} = \max_{i \in \mathcal{D}_k} \{p_m \beta_{k,m,i} + \sum_{j \neq k} x_{j,m} P_j^{(m)} h_{j,m,i}\}, \quad (10)$$

where $\beta_{k,m,i}$ is the link gain factor from the CU user in coalition $\mathcal{S}_m$ to receiver i in group $\mathcal{D}_k$ and p_m is the CU transmission power. Each coalition $\mathcal{S}_m$ may include several D2MD groups. Therefore, the inter-groups interference needs to be also considered. Recall that the indicator variable $x_{j,m}$ is equal to 1 if group $\mathcal{D}_j$ is part of coalition $\mathcal{S}_m$ (i.e., $\mathcal{D}_j$ transmits on channel m), otherwise it is set to 0, $h_{j,m,i}$ is the link gain factor from the transmitter in group $\mathcal{D}_j$ in coalition $\mathcal{S}_m$ to receiver i in group $\mathcal{D}_k$ and $P_j^{(m)}$ is its transmission power on channel m. Eventually, the maximum interference value for each receiver $i \in \mathcal{D}_k$ is considered since the rate in multicast communications is always determined by the weakest receiver. We emphasize that this will introduce in the partition function a min-max fairness rule for balancing the interference among the clusters [38]. Second, for each CU user $m = 1, \ldots, M$, the analogous expression for the interference is

$$\Gamma_m = \sum_k x_{k,m} P_k^{(m)} h_{k,m}. \quad (11)$$

2. The problem of finding the optimal transmission power vectors falls into the general class of fractional programming problems, for which there exist efficient mathematical tools. How to solve this particular fractional programming problem is detailed in our previous work [17].

3. When $r = 1$, this form of coalition includes the Distributed Groups case.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMC.2020.3012451

where $h_{k,m}$ is the deterministic link gain from the transmitter of group $\mathcal{D}_k$ to the CU user in coalition $\mathcal{S}_m$ and $P_k^{(m)}$ is its transmission power on channel m.

Finally, when coalition $\mathcal{S}_m$ includes one or more D2MD groups, the interference accumulates on CU users and, therefore, the coalition value is the total interference

$$\mathsf{v}(\mathcal{S}_m) = \Gamma_m + \sum_{\mathcal{D}_k \in \mathcal{S}_m} \alpha_{k,m}. \tag{12}$$

Thus, $\mathsf{v}(\mathcal{S}_m)$ subsumes the total interference received by the CU user and the D2MD groups that are members of the coalition $\mathcal{S}_m$. Recall that, in our game, the number of coalitions is M, the number of CU users, since every coalition represents a potential RB to be shared among D2MD groups.

Besides the choice of the partition function, the game has to specify the users' strategies for forming the final coalitions. This turns out to be rather simple, since any strategy that is profitable [34], i.e., it improves the value of the current partition function, is a candidate solution to the problem. The game would simply be the iterated application of profitable moves, in an asynchronous and distributed manner, until no further improvement can be made.

5 Coalition Formation Algorithm

The main ingredients required to construct a coalition formation algorithm are a preference relation for forming and breaking coalitions, and adequate notations for assessing the stability of partitions.

5.1 Preference Relation

A preference relation defines how two coalition structures are compared. More formally,

Definition 2. *A preference relation or comparison relation denoted by* $\triangleright$ *is an order defined for comparing two coalition structures* $\mathcal{R} = \{\mathcal{R}_1, \ldots, \mathcal{R}_l\}$ *and* $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_p\}$ *that are partitions of the same subset* $\mathcal{A} \subset \mathcal{N}$. $\mathcal{R} \triangleright \mathcal{S}$ *implies that the way* $\mathcal{R}$ *partitions* $\mathcal{A}$ *is preferred to the way* $\mathcal{S}$ *partitions* $\mathcal{A}$.

Our goal is to minimize the total interference over the coalition members. Therefore, if a coalition structure $\mathcal{R}$ is preferred to another coalition structure $\mathcal{S}$, it means that the coalitions in $\mathcal{R}$ result in less total interference than the coalitions in $\mathcal{S}$: $\mathsf{v}(\mathcal{R}) = \sum_{i=1}^{l} \mathsf{v}(\mathcal{R}_i) \leq \mathsf{v}(\mathcal{S}) = \sum_{i=1}^{p} \mathsf{v}(\mathcal{S}_i)$.

5.2 Merge and Split Rules

The partitions in the previous definition are created based on two rules for forming and breaking the coalitions—merge and split—that are defined as follows.

Definition 3. *(**Merge Rule**) Any subset of coalitions* $\{\mathcal{S}_1, \ldots, \mathcal{S}_l\}$ *may be merged whenever the merged form is preferred by the players. In other words, when* $\{\bigcup_{j=1}^{l} \mathcal{S}_j\} \triangleright \{\mathcal{S}_1, \ldots, \mathcal{S}_l\}$, *then* $\{\mathcal{S}_1, \ldots, \mathcal{S}_l\} \rightarrow \{\bigcup_{j=1}^{l} \mathcal{S}_j\}$.

Definition 4. *(**Split Rule**) Any coalition* $\bigcup_{j=1}^{l} \mathcal{S}_j$ *may be split whenever the split form is preferred by the players. In other words, when* $\{\mathcal{S}_1, \ldots, \mathcal{S}_l\} \triangleright \{\bigcup_{j=1}^{l} \mathcal{S}_j\}$, *then* $\{\bigcup_{j=1}^{l} \mathcal{S}_j\} \rightarrow \{\mathcal{S}_1, \ldots, \mathcal{S}_l\}$.

The previous definitions state the general idea behind coalition formation. In our case, the expression "preferred by the player" can be explained as follows. Assume that we have two coalitions $\mathcal{S}_i$ and $\mathcal{S}_j$. The coalition value changes based on the actions of the merge and split rules. A D2MD group $\mathcal{D}_k$ (which might be a single coalition itself) may split from coalition $\mathcal{S}_i$ and merge into coalition $\mathcal{S}_j$. In this case, three conditions need to be satisfied to approve the merge action:

1) The individual interference $\alpha_{k,m}$ on the weakest receiver of group $\mathcal{D}_k$ decreases when merging with $\mathcal{S}_j$.
2) The total received mutual interference of coalition $\mathcal{S}_j \cup \mathcal{D}_k$ does not increase: $\mathsf{v}(\mathcal{S}_j \cup \mathcal{D}_k) \leq \mathsf{v}(\mathcal{S}_j)$.
3) The new coalition structure $\mathcal{S}'$ results in less total interference than the current one: $\mathsf{v}(\mathcal{S}') \leq \mathsf{v}(\mathcal{S}) \Rightarrow \mathcal{S}' \triangleright \mathcal{S}$.

Conversely, the reciprocal conditions will have to hold for considering the splitting of a coalition into two disjoint ones.

Initially, we consider that each CU user and D2MD group forms a single coalition by itself. These isolated coalitions seek to merge, and eventually split, until M coalitions are formed, one for each CU channel. Each of these coalitions can include up to r D2MD groups and each group could be a member of up to s different coalitions. The members of a coalition will be the set of users that causes the minimum mutual interference to its members and to other coalitions. While this does not guarantee that energy efficiency is being maximized, it is a reasonable proxy measure for that goal since: i) for any target SINR, less transmission power will suffice when the interference is lower, thus increasing EE, at least locally; ii) the target transmission rates are easier to achieve when the interference level is tightly controlled; iii) less interference boosts the opportunities for sharing the transmission channels among different groups.

5.3 Coalition Formation Algorithm

We propose Algorithm 1 to implement the overlapping coalition game. Naturally, the algorithm gives explicit rules for both the split and merge actions and, depending on the resource allocation scenario applied, different types of interference are considered. Recall that, when $r = 1$, there is no accumulated interference and a D2MD group will just receive interference from the CU user, and vice versa. Otherwise ($r > 1$), the interference accumulates on the CU users from all the D2MD groups in the coalition and each D2MD group will suffer interference from the CU and the rest of the groups in the coalition.

We focus now on giving a brief account of the algorithm operations. After initializing the counter of the reuse factor r_t and the data structures (line 1), the iterative procedure for determining the coalitions and the transmission powers begins. The outer loop (lines 3–23) enforces the limit on the reuse factor, while the inner loop (lines 8–21) keeps track of the limit on the split factor. Within the inner loop, the optimal constrained transmission powers are calculated for the current coalition structure (line 4), the constraints on the rates are verified (lines 5-6), the aggregate interferences are evaluated (line 7), and the search for a profitable split or merge move is started. For every element in the partition, Algorithm 1 attempts to find myopically a split (lines 10-13) or a merge action (lines 14-17) which decrease the value of the partition function, or equivalently, which decrease the aggregate interference for the coalition. After the move,

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMC.2020.3012451

1: $r_t \leftarrow 1$; $\mathcal{S}_m \leftarrow \{\}$, $m = 1, \ldots, M$
2: **repeat**
3: **while** $r_t \leq r$ **do**
4: Optimize $p_k^{(m)}$, p_m, $k = 1, \ldots, K$, $m = 1, \ldots, M$, using Thm. 2
5: Evaluate R_k using (3), $k = 1, \ldots, K$
6: Evaluate $r_m = \log(1 + \gamma_m)$, $m = 1, \ldots, M$
7: Sort $\mathcal{S}_m$ in ascending order using (10)
8: **repeat**
9: **for** $m = 1, \ldots, M$ **do**
10: Choose $\mathcal{D}_k \in \mathcal{S}_m$
11: **if** $\mathsf{v}(\mathcal{S}_m \setminus \mathcal{D}_k) < \mathsf{v}(\mathcal{S}_m \setminus \mathcal{D}_{k'})$ for some $\mathcal{D}_{k'}$ **then**
12: $\mathcal{S}_m \leftarrow \mathcal{S}_m \setminus \mathcal{D}_k$
13: **end if**
14: Choose $\mathcal{D}_k \notin \mathcal{S}_m$
15: **if** $\mathsf{v}(\mathcal{S}_m \cup \mathcal{D}_k) < \mathsf{v}(\mathcal{S}_m \cup \mathcal{D}_{k'})$ for some $\mathcal{D}_{k'}$ **then**
16: $\mathcal{S}_m \leftarrow \mathcal{S}_m \cup \mathcal{D}_k$
17: **end if**
18: Optimize $p_k^{(m)}$, p_m, $k = 1, \ldots, K$, $m = 1, \ldots, M$, using Thm. 2
19: **end for**
20: Update aggregate interferences using (10), (11) and (12).
21: **until** all $\mathcal{D}_k \in \mathcal{S}_m$ are tried **or** $|\{\mathcal{D}_k : \mathcal{D}_k \in \mathcal{S}_m\}| = s$
22: $r_t \leftarrow r_t + 1$
23: **end while**
24: **until** $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_M\}$ is a stable coalition structure

Algorithm 1: Coalition Formation Algorithm

if any, the transmission powers are again optimized (line 18), and the interference for each channel and cluster is updated (line 20). If no profitable moves are possible, then the limit in the reuse factor is increased to search other non-explored combinations in the partition (line 22). The algorithm finishes when a stable partition structure is found. Stability is discussed in the next Subsection.

The computational complexity of Algorithm 1 is in any specific instance highly dependent on the random locations of the UEs and the channel gain realizations, since these induce different interference strength. Nevertheless, its worst case execution time can be determined as follows. The algorithm terminates after at most r rounds, where r is the reuse factor. In each round, a possible merge and/or split move is considered for up to M coalitions, and a profitable move is tested between its members and non-members, so this step has $O(MK^2)$ complexity, at worst. Besides, the power optimization complexity is solved through the Dinkelbach's algorithm, which has a superlinear convergence rate [39], and the evaluation of our objective function takes $O(1)$ time for a number of variables/constraints which is $O(M + K)$ (rate and power for each user). Therefore, the overall worst-case complexity is $O(rMK^2(M+K))$. We remark that, in practice, the optimization step finishes in just a few iterations (3-4, see [17]) and that our numerical experiments in Section 6 show that a stable coalition can be formed in a low number of iterations too, so the average complexity is typically much smaller than $O(rMK^2(M + K))$.

5.4 Convergence and Stability

Our main result in this Section is a proof that the convergence of the OCFG with Algorithm 1 is guaranteed. To this end, we first define stability as follows, since there exist a number of refinements of the notion of stability in the literature of coalition games [34] useful in different contexts.

Definition 5. *A coalition structure is $\mathbb{D}_{hp}$-stable if it cannot be changed by the merge-and-split process. A coalition structure is Pareto optimal if the following conditions hold:*

1) *The coalition structure is $\mathbb{D}_{hp}$-stable.*
2) *The coalition structure maximizes the sum of utilities of the players.*

Theorem 1. *The outcome of Algorithm 1 is a stable and Pareto optimal coalition structure.*

Proof: Define the state of the process of coalition formation as its current partition structure, i.e., the collection of coalitions formed. A state is called absorbing [34] if the process does not move, that is, there is no incentive for any player to perform a merge or a split action. So, by definition, an absorbing state is $\mathbb{D}_{hp}$-stable. To see that Algorithm 1 reaches in a finite number of steps an absorbing state just note that, according to conditions 2) and 3) in Section 5.2, any merge or split move must be profitable, i.e., it must lead to a lower value for the aggregated interference of the newly formed coalition. Therefore, the sequence of values of the partition function for each D2MD cluster is non-negative and decreasing, and as a consequence the game reaches an absorbing state. By [34, Lemma 5.1], all absorbing states are stable. For the Pareto optimality, the proposed game has transferable utility [28], so the utilities of the player can be distributed so that the social welfare (the sum of utilities) is maximal. □

We remark that, though our game has transferable utilities, it is not symmetric since the partition function depends on the random location of the UEs and not only on the number of users in the clusters.

5.5 Optimal Power Control

Algorithm 1 unfolds an alternating optimization strategy between the sub-problems of channel allocation and optimal power control. This means that, for any given coalition structure, the minimum transmission powers compatible with the constraints are used so as to maximize the EE (global or per-UE). The optimization of transmission powers can be formulated as a fractional programming problem [40] and solved computationally in a very efficient way based on the following result, which is recalled here for the sake of completeness [17, Thm. 1].

Theorem 2. *Let $f(\mathbf{p}) = \sum_m r_m(\mathbf{p}) + \sum_k R_k(\mathbf{p})$, $g(\mathbf{p}) = \tau + \sum_m p_m + \sum_k \sum_m P_k^{(m)}$ and $\mathcal{P}$ the set of feasible vectors for the EE optimization problems. Then $\mathbf{p}^* \in \mathcal{P}$ solves* (9) *if and only if $\mathbf{p}^* = \arg\max_{\mathbf{p}\in\mathcal{P}}\{f(\mathbf{p}) - \lambda^* g(\mathbf{p})\}$, where λ^* is the unique zero of $F(\lambda) = \max_{\mathbf{p}\in\mathcal{P}}\{f(\mathbf{p}) - \lambda g(\mathbf{p})\}$.*

Since, in our case, $f(\mathbf{p})$ is not generally a concave function, we replace it with a surrogate concave minorization function taking the same value as $f(\cdot)$ at $\mathbf{p}$ (see [17] for details). After this change, the maximization problem appearing in Thm. 2

Table 2
System Parameters for the Simulation Experiments

Parameter	Value
Cell radius	500 m
Reuse factor (r)	$\{2, 3, 4, 5\}$
Network density (λ)	250 devices/cell
Split factor (s)	$\{2, 3, 4\}$
Path loss exponent (α)	2.5
Minimum transmission rate	$\{0.1, 0.5\}$ bit/s/Hz
Number of CU users (M)	$\{3, 4, 5, 6, 8, 10, 15\}$
Maximum transmission powers	$[-5, 25]$ dBm
Number of D2MD groups (K)	$\{4, 5, 6, 9, 10, 15, 25\}$
Circuit power	10 dBm

is convex and can be solved with standard methods in linear time [15], [16]. For the case of maximizing the minimum EE, the goal is to maximize $\min_{i\in\mathcal{N}} f_i(\mathbf{p})/g_i(\mathbf{p})$, where now $f_i(\mathbf{p})$ is the rate received by user i, $g_i(\mathbf{p})$ is its power consumption, and $\mathcal{N}$ is the set of UEs. Theorem 2 can be used to prove that this is equivalent to solving $\arg\max_{\mathbf{p}\in\mathcal{P}} \min_i f_i(\mathbf{p}) - \lambda^* g_i(\mathbf{p})$. Since the pointwise minimum of concave functions is a concave function, we immediately see that maximizing the minimum EE has the same structure and can be solved with the same numerical methods as the global EE.

6 Evaluation

In this Section, some simulation results will be presented to evaluate the performance of multicast EE optimization. To obtain these results, we used MATLAB and CVX mathematical package. The default system parameters are summarized in Table 2. We consider different coalition sizes (i.e., reuse factors), overlapping degrees (i.e., split factors), effectiveness of power budget usage (fixed budget per RB or distributed over several RBs), etc. Each simulation experiment has been repeated 200 times and all the graphs display the average value for each performance measure. In these experiments, the number and location of UEs will follow a standard homogeneous Poisson point process (PPP) [41], [42] with density $\lambda = 250$ devices/cell in a cell of radius 500 m. The received signals are assumed to vary due to the path loss according to $P_r = P_t(1+(d/d_0)^\alpha)$ where P_r is the received power, P_t is the transmitted power, d_0 is a reference distance (100 m in our case) and α is the path loss exponent. For D2MD cluster formation, we used two different algorithms: K-Nearest Neighbor (KNN) and Distance Limit (DL) clustering. For both algorithms, the number of head clusters (transmitters of D2MD groups) are predefined and randomly selected. KNN results in clusters of the same size, since the receivers of each cluster are the k closest users to the head cluster, by definition, while DL clusters have different sizes because their receivers are those users within the configured radius [43]. This feature, the fixed or variable size of the clusters, turned out to have little impact on the performance results observed in our numerical experiments, so we only show the results for one of the algorithms, with similar conclusions being valid for the other. Finally, the CU users with the best channel quality are selected to share their RBs with the D2MD clusters [44].

Table 3
EE and Rates with Dedicated CUs

Min. Rate (bit/s/Hz)	GEE (bit/Hz/J)	Avg. Rate (bit/s/Hz)	MEE (bit/Hz/J)	Min. Rate (bit/s/Hz)
0.1	811.96	104.97	178.26	2.13
0.2	799.53	102.86	179.63	2.14
0.3	791.85	101.50	181.30	2.17
0.4	790.65	100.90	183.20	2.19
0.5	775.26	98.13	179.48	2.13

6.1 Dedicated CUs

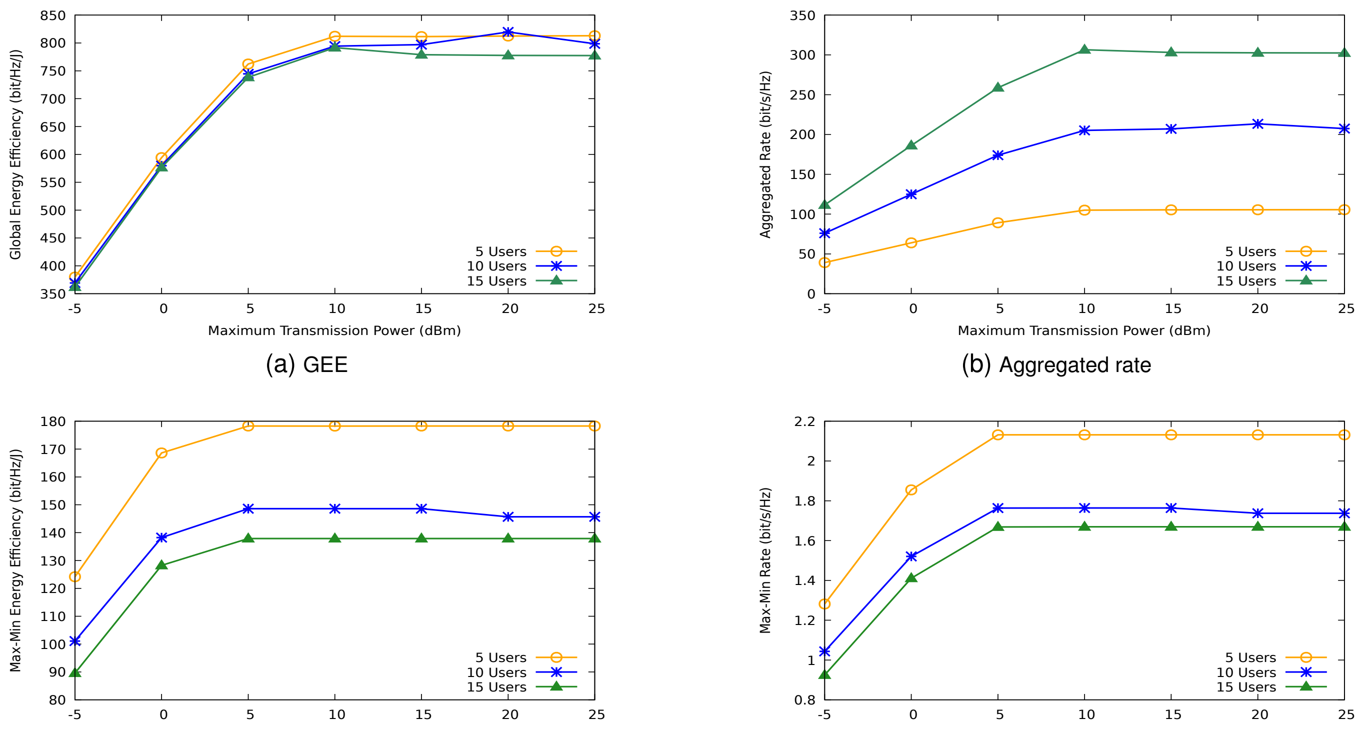
We first investigate the simplest coalition formation game where each D2MD group can only use one CU channel and each channel only supports a single D2MD group ($r = s = 1$). Therefore, there is no accumulated interference on both sides and the value of the coalition is just the sum of the interference received by the CU from its corresponding cluster. The minimum rate per RB is set to 0.1 bit/s/Hz and the maximum transmission power per device is varied in the range of −5 dBm to 25 dBm (the maximum transmission power for UEs in 5G is restricted to 23 dBm [45]). The number of CU users and D2MD clusters are set to $M = K = \{5, 10, 15\}$. We used DL algorithm to create the clusters, such that each cluster has 3 receivers on average with a maximum distance between the head cluster and the receivers of 50 m.

6.1.1 Problem Feasibility

A feasible case means that all the users in the coalitions were able to satisfy the minimum rate threshold with the available transmission power budget. For the GEE problem, feasibility was over 99 %, showing up an average number of non-feasible cases in all the 200 simulation instances equal to 1. Similarly, the feasibility for the MEE problem reached 100 % in all the simulated scenarios. We conclude that, for these settings, problem constraints are not particularly stringent and a solution to both optimization problems exists in almost every case.

6.1.2 Energy Efficiency and Rate Analysis

Both global and minimum EE were evaluated using the parameters defined above. In Fig. 2, we observe the behavior of the EE and the transmission rate. The aggregated rate increases as the number of clusters and CU users grows, but this is not the case for the GEE, where, despite some minor variations related to the simulation error margin, the GEE slightly diminishes with the number of users. However, considering the major difference in the value of the achieved aggregated rate and that the difference in EE is relatively small, we could say that it pays off to sacrifice some of the GEE gain in order to achieve a higher rate. Note also that increasing the transmission power budget beyond a threshold—around 10 dBm for global EE and 5 dBm for minimum EE—does not produce further gains neither in the energy efficiency nor in the rate, due to the strong interference that would result. This saturation point identifies the system capacity (in terms of EE) and is nearly independent of the number of users.



(a) GEE (b) Aggregated rate (c) MEE (d) Max-min rate

Figure 2. Performance with dedicated CUs.

Table 4
EE and Rates with Shared CUs

MIN. RATE (bit/s/Hz)	GEE (bit/Hz/J)	AVG. RATE (bit/s/Hz)	MEE (bit/Hz/J)	MIN. RATE (bit/s/Hz)
0.1	762.19	141.84	120.52	1.42
0.2	739.18	136.57	121.63	1.44
0.3	718.80	130.65	120.60	1.42
0.4	711.76	129.27	120.36	1.42
0.5	702.85	127.03	119.27	1.41

6.1.3 Minimum Rate Constraints

We also test the influence of the minimum rate constraints on both EE and rates. The maximum transmission power is set to 10 dBm, the minimum rate varies from 0.1 bit/s/Hz to 0.5 bit/s/Hz and the number of CUs and D2MD clusters is 5. The results of this testing case are illustrated in Table 3. Now, the GEE and the aggregated rate decrease continuously as we impose tighter minimum rate constraints. This is because users must increase their transmission power to attain the minimum rate, thus raising the amount of interference and preventing any further improvement in the aggregated rate and GEE. For the MEE, it is slowly increasing with the minimum rate constraints augments but drops eventually. The explanation is the same as for GEE.

6.2 Shared CUs

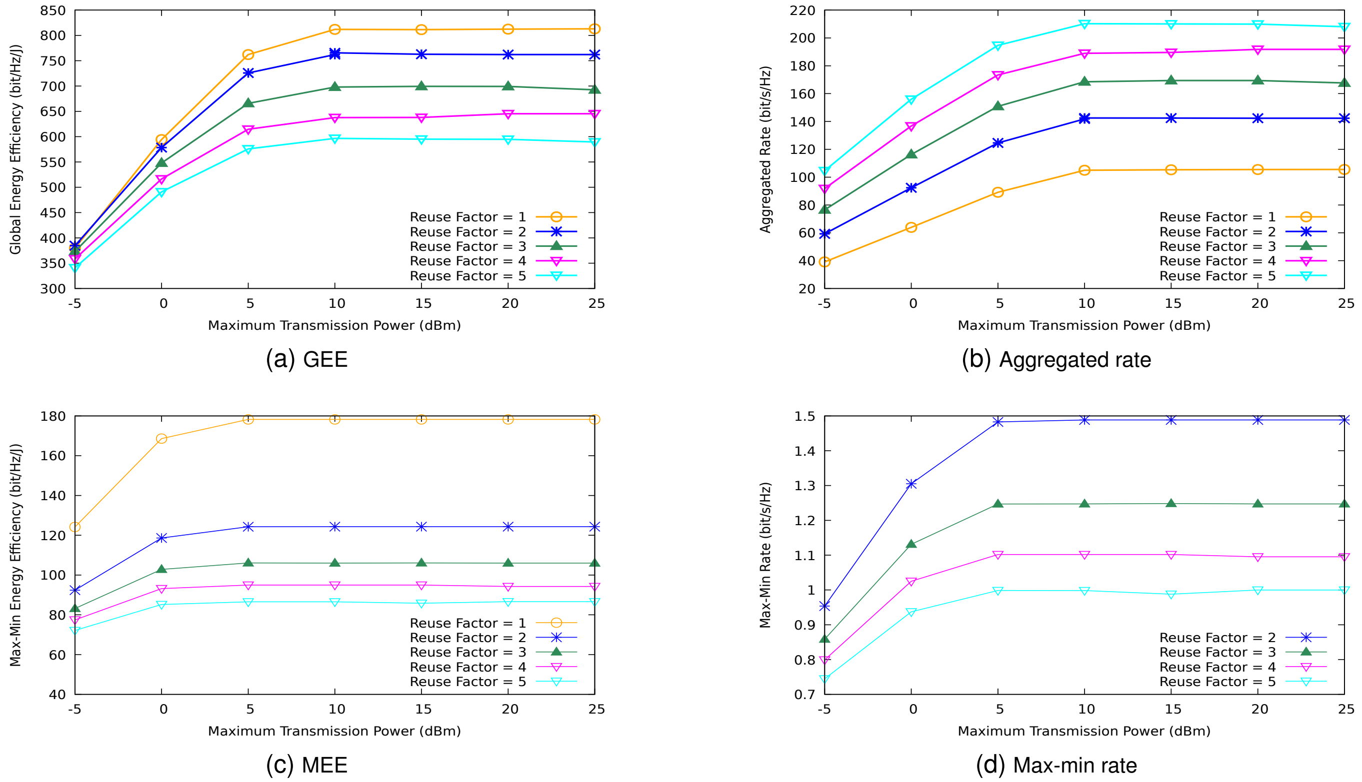
Next, we evaluate the main performance metrics of the system when a coalition is composed of a single CU user and multiple D2MD clusters ($s = 1$ and $r > 1$). The minimum rate per RB is set to 0.1 bit/s/Hz, while the transmission power budget varies from −5 dBm to 25 dBm. The number of CU users is set to $M = 5$, and the number of clusters to $K \in \{10, 15, 20, 25\}$. The reuse factor is $r \in \{2, 3, 4, 5\}$, according to the number of clusters. Again, the clustering technique used here is DL.

6.2.1 Problem Feasibility

For the GEE problem, the existence of non-feasible cases is negligible as long as the reuse factor is kept low. Only with a $r = 5$ the unfeasible cases raise to $55 \approx 25\%$, on average. Though still a low value, it shows that having more degrees of freedom for sharing a channel is not always beneficial for the system design. For the MEE problem, we did not detect unfeasible cases.

6.2.2 Energy Efficiency and Rate Analysis

Fig. 3 shows the EE and rates obtained varying the transmission power budget with different reuse factors. As we noticed in the previous case, the GEE and the aggregated rate increase with the transmission power until hitting a saturation point, and the same happens for the MEE and the minimum rate. Hence, as more power is allocated to users, the EE and rates for both the system and the individual users remain stable. Clearly, when $r = 1$, the network and the individual users achieve the highest EE and rates, which can be explained due to the absence of accumulated interference over both CUs and D2MD groups. However, acceptable values for the EE and rates are achieved with low values of reuse factors like $r = 2$ and $r = 3$, even with low transmission powers. Obviously, for the RBs to support more users or having a



(a) GEE (b) Aggregated rate (c) MEE (d) Max-min rate

Figure 3. Performance with shared CUs.

bigger coalition size, interference must be kept as weak as possible. We recall here that, in many realistic settings, the number of D2MD groups can be larger than the number of RBs, so the reuse of the spectrum would be necessary to allow the transmissions of all users.

6.2.3 Minimum Rate Constraints

We also evaluated the minimum rate constraints when $r = 2$. The obtained results are listed in Table 4. As expected, the higher the minimum rate value, the more transmission power is required to satisfy the requested conditions. As a result, the interference worsens, thus preventing any major improvement in the achieved rate and the achievable rates decrease. In turn, as EE is the ratio of the rate to the consumed power, both MEE and GEE also decrease.

6.3 General Case with Overlapped Coalitions

We test now the benefits of using overlapping coalition games for resource allocation in the general case with $s > 1$ and $r > 1$. The transmission power goes from -5 dBm to 25 dBm, and each D2MD group can distribute it over s coalitions. The minimum total rate is 0.5 bit/s/Hz. The minimum rate per RB for both CU users and D2MD groups is set to 0.1 bit/s/Hz. The number of CUs is set to $M \in \{4, 6, 8\}$ while the number of groups is fixed to $K = 4$ to ensure that the reuse factor $r = 2$ is always satisfied. Finally, the split factor is varied based on the number of available CUs ($s \in \{2, 3, 4\}$). The number of unfeasible problem instances generated under these conditions was again negligible for any combination of parameters.

Table 5
EE and Rates in the General Case

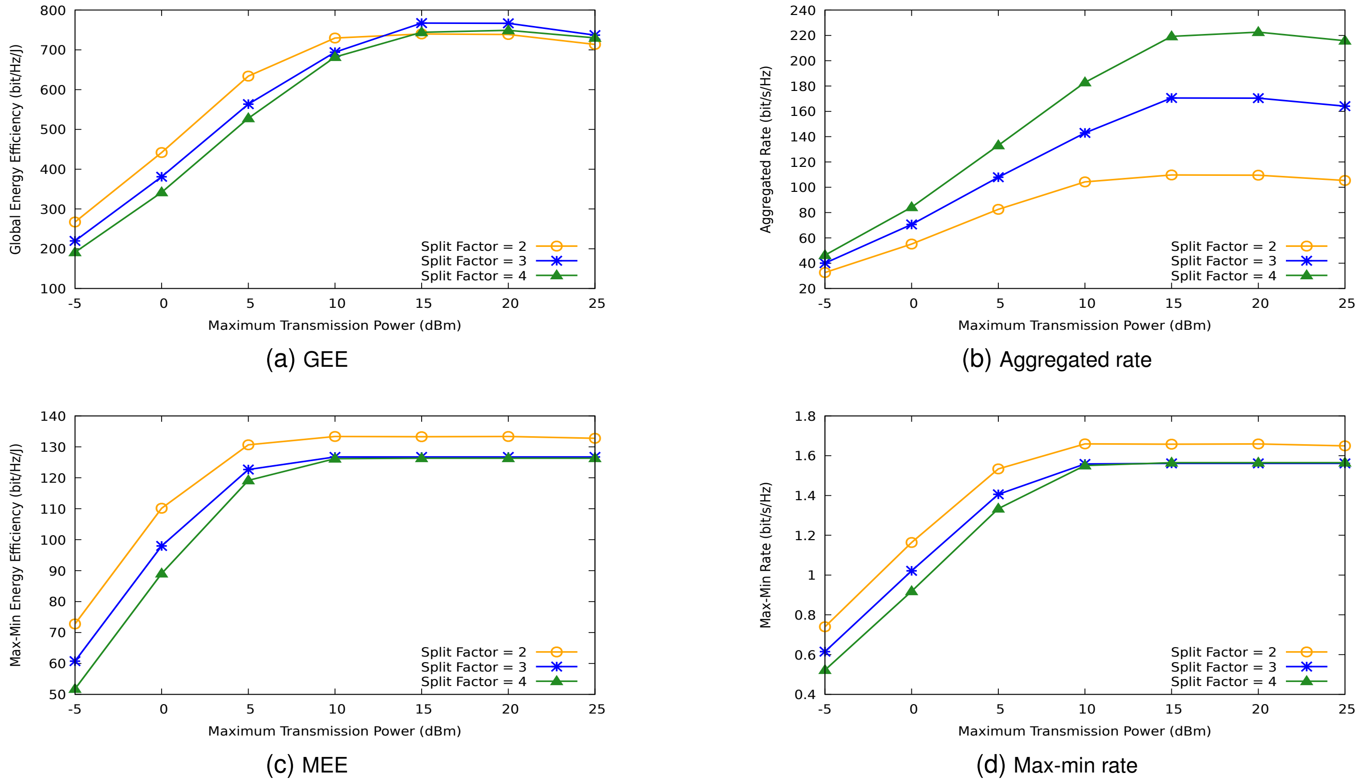
Min. Rate (bit/s/Hz)	GEE (bit/Hz/J)	Avg. Rate (bit/s/Hz)	MEE (bit/Hz/J)	Min. Rate (bit/s/Hz)
0.1	724.48	193.60	123.70	1.51
0.2	696.51	187.16	126.15	1.54
0.3	662.60	179.60	139.81	1.75

6.3.1 Energy Efficiency and Rate Analysis

We evaluated the EE and rates obtained when using overlapped coalitions. As in the previous scenarios, Fig. 4 shows that both the GEE and the aggregated rate increase with higher transmission budgets until they saturate and remain almost constant independently of the possibility of using more power. And the same happens with the MEE and the minimum rate, as shown in Fig. 4. However, note that the split factor affects the EE and the aggregated rate in a different way. It can be seen that, as more resources a D2MD group can have, the less EE it attains. Contrarily, the aggregated rate keeps increasing as more resources are available. Since the difference in the achieved EE over the various power values is not as high as in the previous cases, a higher aggregated rate may compensate for the lower EE.

6.3.2 Minimum Rate Constraints

In this experiment, we set the minimum transmission power for all the users in a coalition to 10 dBm, and fix the minimum rate for D2MD groups to a very low value, 0.01 bit/s/Hz, for each active RB. This is done with the only purpose of



(a) GEE

(b) Aggregated rate

(c) MEE

(d) Max-min rate

Figure 4. Performance in the general case.

avoiding very low quality RBs in the solution, from the point of view of energy efficiency. The minimum rate for CU users and D2MD groups is varied in the range of 0.1 bit/s/Hz to 0.3 bit/s/Hz, and we have picked values for M and K equal to 8 and 4, respectively, with a split factor $s = 4$. The results are shown in Table 5. It is worth to mention that we decreased the range of the minimum rate because, as it increases, the problem feasibility is relatively high. Obviously, the rate and the EE for both the system and individual users decrease as we set tighter constraints. Even for a predefined transmission power, users will always require higher values to satisfy the requested minimum rate. In this case, the interference in the coalition will increase, thus resulting in a lower rate and more consumed transmission power.

6.4 Comparison with Other Approaches

To assess the performance quality of our coalition formation algorithm, we compare it with a greedy algorithm that evaluates all the possible pairings between the available CUs and D2MD groups to select the CU-D2MD pair that achieves the highest GEE. Note that the optimal selection process may be affected by the number of receivers in the D2MD groups. In fact, those groups with more receivers can achieve higher rates with lower power consumptions (i.e., higher GEE). Therefore, to guarantee a fair comparison, we used the KNN algorithm to form 5 groups with the same number of receivers (3 receivers per D2MD group).

We considered the three following cases: i) dedicated CUs, so each coalition is composed of just a CU user and a single D2MD group ($r = s = 1$); ii) shared CUs, so each D2MD group can be part of just a single coalition ($s = 1$), but each coalition can be composed of up to 2 groups ($r = 2$). The number of CU users and D2MD groups is set to $M = 2$ and $K = 4$, respectively; iii) the general case, so we also allow a group to appear in different coalitions. In this setting, the number of CU users and D2MD groups is fixed to $M = K = 4$, and the reuse and split factors are set to $r = s = 2$.

In every test scenario, transmission powers are in the range −5 dBm to 25 dBm and the minimum rate per channel is 0.1 bit/s/Hz. Fig. 5 shows the global EE and aggregated rate obtained in each simulation experiment. Clearly, our coalition formation algorithm performs very close to the optimal, yet with much lower complexity.

In addition, we compare our semi-distributed approach with the iterative matching algorithm proposed in [19], a fully distributed technique. Similar configurations were used, but with $M = 3$, $K \in \{6, 9\}$, $r \in \{2, 3\}$, and $s = 1$. Fig. 6 presents both the system and the individual EE and rates. Clearly, EE increases monotonically with the power budget of a user until reaching a maximum. This happens at, approximately, the same values for both techniques, but note that the distributed approach shows a sharp decline beyond that point. Therefore, our algorithm controls much better the aggregated interference, thus resulting in a significant performance improvement in both EE and rates. Additionally, our semi-distributed approach also permits to maximize the MEE and the minimum rate among network users, thus providing a higher and fairer performance for them.

Finally, we compared this semi-distributed approach with the centralized one proposed in our previous work [17]. Fig. 7 shows that the semi-distributed approach performs slightly better because users make their resource allocation

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMC.2020.3012451

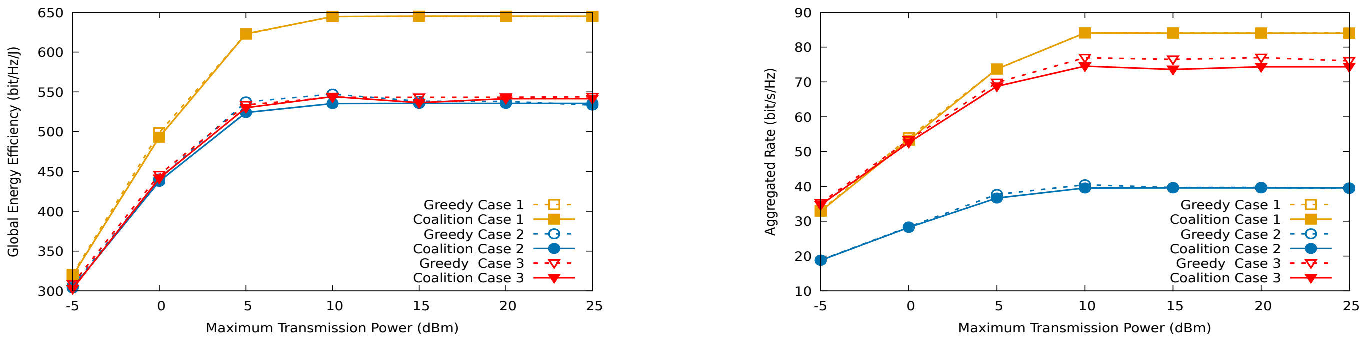
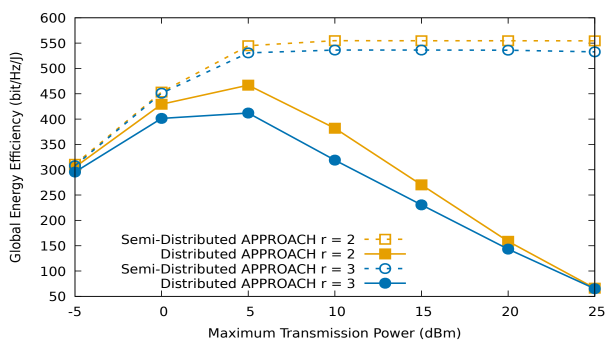
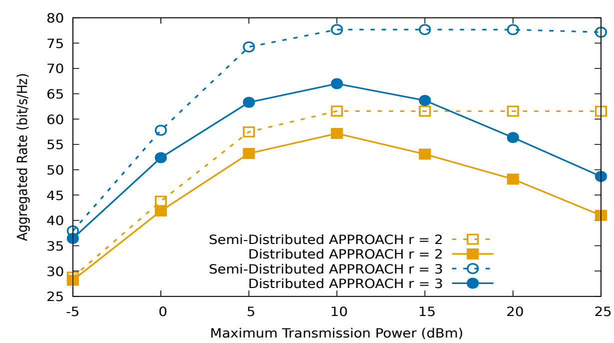


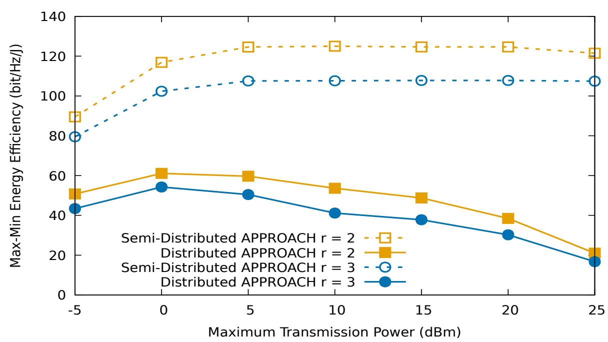
Figure 5. GEE and aggregated rate with the proposed coalition formation algorithm and the greedy one.



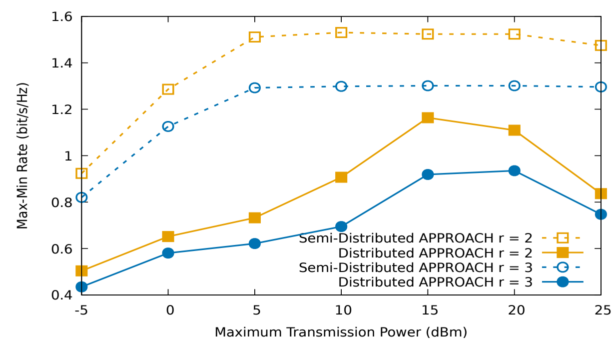
(a) GEE



(b) Aggregated rate



(c) MEE



(d) Max-min rate

Figure 6. Performance comparison: proposed semi-distributed approach vs. the fully distributed one.

decision depend on local information, which is generally more accurate. The advantages of the semi-distributed approach will become more remarkable as the number of users increases since, in this regime, the problem complexity and overhead signaling with a centralized approach are likely to overload the central entity, with a huge burden in both computation and processing time.

6.5 Algorithm Convergence

To evaluate the convergence of the proposed algorithm we devised three test cases: i) Dedicated CUs scenario with each cluster limited to a single RB and the number of CU users and D2MD clusters set to 5; ii) shared CUs scenario with 5 CUs, 10 clusters and $r = 2$, so up to 2 clusters can share a single RB; iii) general sharing with 4 CUs and clusters, and $r = s = 2$. The maximum transmission power for all the devices is 10 dBm and the minimum rate per channel is 0.1 bit/s/Hz. In the general case scenario, clusters can divide the total achieved rate over the $s = 2$ channels, thus the total minimum rate is set to 0.2 bit/s/Hz. For clustering we used the KNN algorithm, so the clusters are homogeneous clusters of size 3.

Simulation results for the convergence of EE in the dedicated CU case are shown in Fig 8. Only a few iterations (2–3) suffice for the semi-distributed algorithm to reach its stationary value, with only marginal improvement in subsequent rounds. Fig 9 illustrates the average aggregated interference per coalition in the shared CUs resource sharing case. The graph shows that most of the reduction in the aggregated interference is attained in the very first merge-split moves of the algorithm, with the remaining ones being useful just for minor changes in the coalitions and for adjusting the transmission power. So, once the stable coalitions are found, no further changes take place into the

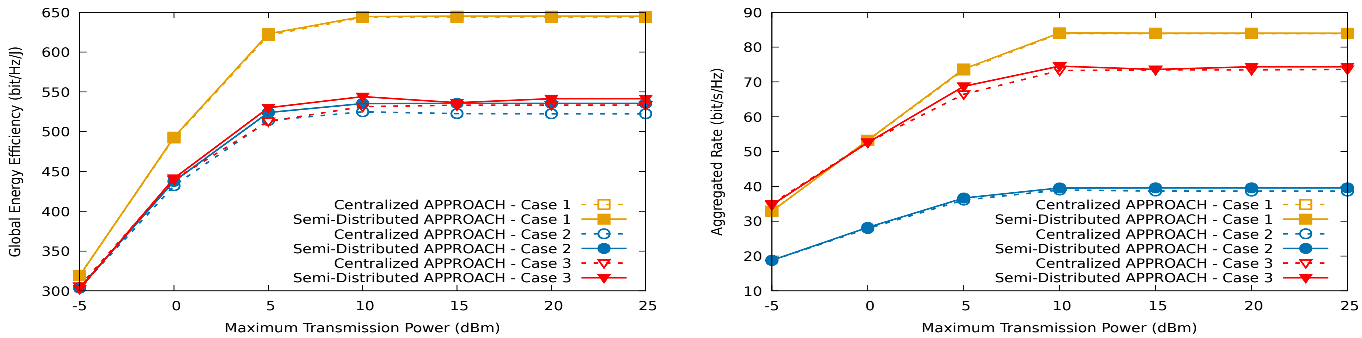


Figure 7. GEE and aggregated rate with the proposed semi-distributed approach and the centralized one.

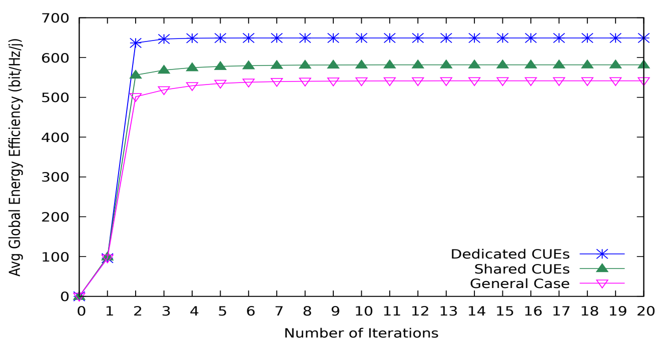


Figure 8. Resource and power allocation algorithm convergence.

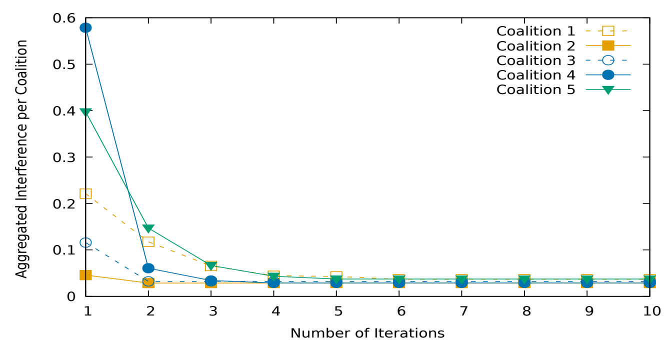


Figure 10. A problem instance: Aggregated interference per coalition vs. Number of iterations.

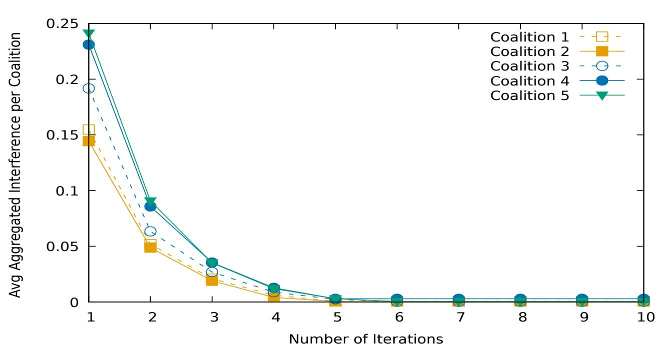


Figure 9. Average aggregated interference per coalition vs. Number of iterations.

structure of the resource allocation, but only in the power control algorithm. As shown, on average, a coalition needs only 3–4 iterations to converge, so the algorithm is fast. The convergence time increases slightly as more clusters are seeking for RBs, e.g., when the number of D2MD clusters is set to 20 in the general sharing case, and the average number of iterations required for a coalition to converge is between 4 and 6 (see Fig. 10). Finally, we remark that in Figs. 9–10 the values of interference are relative to the unit of −20 dBm. Therefore, Figs. 8–10 jointly illustrate that, for typical realizations of the PPP process, convergence to a stable partition of UEs into coalitions happens after just a few execution rounds of Algorithm 1.

6.6 Discussion

Simulation results show that the obtained EE and rates increase with the transmission power budget up to the saturation point, thus remarking the fact that we have uniformly identified the EE vs. capacity trade-off of D2MD wireless networks. Note that, as the interference becomes stronger, users are forced to consume more energy to satisfy the target minimum rate. Therefore, no further improvements can be achieved in rate in exchange of more transmission power. We have also confirmed experimentally that the clustering techniques have little or negligible effect on the obtained EE and rates.

After giving the general idea, we will discuss the different resource sharing scenarios. In the dedicated CUs scenario with a one-to-one allocation between CUs and D2MD groups, the interference is at its minimum level, as it will not accumulate on the side of the CUs neither the D2MD groups. As shown in Fig. 2, in this scenario it is preferred to start with a high number of users because this yields a relatively high aggregated rate with a minor cost in GEE. However, note that this might not be completely fair when considering the users' individual EE, so seeking a moderate number of users will provide an acceptable balance for the loose in the MEE and in the aggregated rate in terms of the gain in the GEE. On the other hand, the shared CUs scenario aims to support more D2MD groups, as multiple groups can share the resources of a single CU. This clearly hardens the challenge since the interference accumulates on both sides from the groups. Even though we kept the same number of resource blocks in the simulation experiments (5 RBs), we managed to support up to 25 clusters with a reuse factor equal to 5. Comparing these scenarios, higher EE values are obviously achieved with dedicated CUs but, for a low transmission power budget, shared CUs are also efficient. The low transmission budget helps to create an environment where the interference is more controlled. As a result, the network capacity is efficiently increased, thus leading to an acceptable trade-off in terms of EE and rate as shown in Fig. 3. Finally, we also allowed that a D2MD group use multiple CUs, i.e., the general case for a resource allocation scenario. To guarantee the efficiency on this latter scenario, D2MD groups require a high transmission power budget to attain high EE and rate values. Compared to the previous cases, this scenario has the most suitable balance in terms of network and individual EE and rates, as shown in Fig. 4. Allocating more resources to D2MD groups, as well as enough transmission power, will guarantee that both the system and individual EE and rates will remain high with the lowest possible energy consumption. As a conclusion, and after examining both the central and the distributed approaches for resource allocation, we conclude that it is possible to take advantage of underlay D2MD communications when properly configured. It is obvious that RBs have limited capacity and that, in some cases (that is, with high reuse factors), more users will result in a negative effect on EE and rate unless the transmission power is tightly controlled. Also, high split factors require more transmission power to attain the desired performance. In summary, the general scenario provides a good balance between system and individual performance, while supporting a moderate number of users.

7 Conclusions

In this paper, we investigated a two-stage semi-distributed solution for the joint power control and resource allocation problem of underlay D2MD communications on cellular networks. A coalitional game model with a partition function based on the aggregate co-channel interference is used to solve in nearly optimal way the allocation of channels under arbitrary constraints on the minimum rate and the maximum power. Overlapping coalitions are allowed for enlarging the solution space. The power control algorithm is instead based on a fractional programming approach. Both phases are coupled, and are run in an alternate way, so the transmission powers are always optimized for every transient coalition structure found so far. The combined framework is unified—allows to maximize system-level EE as well as individual EE— and results in a performance very close to the optimal, obtaining better results than existing approaches and being insensitive to cluster size. We demonstrate as a consequence that the gap between the individual EE of different users is lower than when only the individual EE is maximized. In this respect, our approach obtains more fairness—for a given aggregated transmission rate—than the selfish maximization of EE by a set of independent users.

As a semi-distributed approach, since only the power control phase requires inter-cluster information and the resource allocation is decentralized, the proposed scheme is scalable. Aggregated interference can be easily measured in practice, and, since this is the only signal exchanged during the game, the implementation of the algorithm has low complexity and overhead. Consequently, it realizes the advantages of communications without a central infrastructure like D2D.

Although we focused on least-interferrer channel allocation and power minimization to tightly control the interference regime, other approaches can be investigated. More efficient multicast data delivery with D2D communications can be designed by using reconfigurable intelligent surfaces for mitigating the interference, or separating interference from information-bearing signals at the receivers via non-orthogonal multiple access techniques. The performance analysis of D2MD under those conditions constitutes a promising research direction.

References

[1] 3rd Generation Partnership Project (3GPP), "TS 23.303 V12.0.0.Technical Specification Group Services and System Aspects;Proximity-based services (ProSe); Stage 2(Release 12)," 2014. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.303/23303-c10.zip
[2] ——, "TR 38,836 Study on NR sidelink relay (Rel. 17)."
[3] G. Araniti, A. Raschellà, A. Orsino, L. Militano, and M. Condoluci, "Device-to-device communications over 5G systems: Standardization, challenges and open issues," in *5G mobile communications*. Springer, 2017, pp. 337–360.
[4] M. Höyhtyä, O. Apilo, and M. Lasanen, "Review of latest advances in 3GPP standardization: D2D communication in 5G systems and its energy consumption models," *Future Internet*, vol. 10, no. 1, p. 3, 2018.
[5] S. Pizzi, F. Rinaldi, A. Molinaro, A. Iera, and G. Araniti, "Energy-Efficient Multicast Service Delivery Exploiting Single Frequency Device-To-Device Communications in 5G New Radio Systems," *Sensors*, vol. 18, no. 7, p. 2205, 2018.
[6] R. O. Afolabi, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for OFDMA-based systems: A survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 240–254, 2012.
[7] H. Zhou, H. Wang, X. Li, and V. C. Leung, "A survey on mobile data offloading technologies," *IEEE Access*, vol. 6, pp. 5101–5111, 2018.
[8] L. Militano, A. Orsino, G. Araniti, A. Molinaro, and A. Iera, "A constrained coalition formation game for multihop D2D content uploading," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2012–2024, 2015.
[9] G. Zhao, S. Chen, L. Qi, L. Zhao, and L. Hanzo, "Mobile-Traffic-Aware Offloading for Energy- and Spectral-Efficient Large-Scale D2D-Enabled Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 6, pp. 3251–3264, 2019.
[10] L. Yang, D. Wu, and Y. Cai, "A Distributed Social-Aware Clustering Approach in D2D Multicast Communications," in *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*. IEEE, 2018, pp. 42–47.
[11] Y. Xu, S. Jiang, and J. Wu, "Towards energy efficient device-to-device content dissemination in cellular networks," *IEEE Access*, vol. 6, pp. 25 816–25 828, 2018.
[12] P. Mach, Z. Becvar, and T. Vanek, "In-band device-to-device communication in OFDMA cellular networks: A survey and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1885–1922, 2015.
[13] J. Iqbal, M. A. Iqbal, A. Ahmad, M. Khan, A. Qamar, and K. Han, "Comparison of spectral efficiency techniques in device-to-device communication for 5g," *IEEE Access*, vol. 7, pp. 57 440–57 449, 2019.
[14] Y. Xu, X. Li, and J. Zhang, "Device-to-Device Content Delivery in Cellular Networks: Multicast or Unicast," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4401–4414, 2018.
[15] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, "Energy-efficient power control: A look at 5G wireless technologies," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1668–1683, 2016.
[16] A. Zappone, E. Björnson, L. Sanguinetti, and E. Jorswieck, "Globally optimal energy-efficient power control and receiver design in wireless networks," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2844–2859, Jun. 2017.

[17] M. Hmila, M. Fernández Veiga, M. Rodríguez Pérez, and S. Herrería Alonso, "Energy Efficient Power and Channel Allocation in Underlay Device to Multi Device Communications," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 1–1, 2019.

[18] M. Hmila, M. Fernández Veiga, and M. Rodríguez Pérez, "Distributed resource allocation approach for device-to-device multicast communication," in *14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'18)*. IEEE, 2018.

[19] Z. Zhou, G. Ma, M. Dong, K. Ota, C. Xu, and Y. Jia, "Iterative energy-efficient stable matching approach for context-aware resource allocation in D2D communications," *IEEE Access*, vol. 4, pp. 6181–6196, 2016.

[20] F. Li, Y. Zhang, and M. A. Al-Qaness, "Multi-objective resource allocation scheme for D2D multicast with QoS guarantees in cellular networks," *Applied Sciences*, vol. 6, no. 10, p. 274, Sep. 2016.

[21] A. Bhardwaj and S. Agnihotri, "Energy-and Spectral-Efficiency Trade-off for D2D-Multicasts in Underlay Cellular Networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 546–549, Aug. 2018.

[22] ——, "Multiple D2D Multicasts in Underlay Cellular Networks," *arXiv preprint arXiv:1908.08866*, 2019.

[23] D. D. Ningombam and S. Shin, "Resource-sharing optimization for multicast D2D communications underlaying LTE-A uplink cellular networks," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, 2019, pp. 2001–2007.

[24] B. S. Ghahfarokhi, M. Azadmanesh, and S. K. Khorasani, "Energy and spectrum efficient mobility-aware resource management for d2d multicasting," *Computer Networks*, vol. 146, pp. 47–64, 2018.

[25] S.-Z. Jian and T.-C. Hou, "Enabling efficient device to device multicast communication by clustering and resource sharing," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–6.

[26] J.-K. Zuo and L.-X. Yang, "Energy efficient resource allocation for D2D multicast communications," *Journal of Central South University*, vol. 26, no. 11, pp. 3034–3044, 2019.

[27] F. Jiang, Y. Liu, C. Li, and C. Sun, "Energy-efficient multicast transmission for underlay device-to-device communications: A social-aware perspective," *Mobile Information Systems*, vol. 2017, pp. 1–17, 2017.

[28] Z. Han, *Game theory in wireless and communication networks: theory, models, and applications*. Cambridge University Press, 2012.

[29] Z. Zhou, M. Dong, K. Ota, J. Wu, and T. Sato, "Distributed interference-aware energy-efficient resource allocation for device-to-device communications underlaying cellular networks," in *Global Communications Conference (GLOBECOM), 2014 IEEE*. IEEE, 2014, pp. 4454–4459.

[30] ——, "Energy efficiency and spectral efficiency tradeoff in device-to-device (D2D) communications," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 485–488, Oct. 2014.

[31] X. Chen, R. Q. Hu, and Y. Qian, "Distributed resource and power allocation for device-to-device communications underlaying cellular network," in *Global Communications Conference (GLOBECOM), 2014 IEEE*. IEEE, 2014, pp. 4947–4952.

[32] M.-T. Thi, T. Huynh, A. Radwan, and W.-J. Hwang, "Coalition formation game for joint power control and fair channel allocation in device-to-device communications underlaying cellular networks," *Wireless Personal Communications*, pp. 1–19, 2017.

[33] J. Wang, Y. Sun, B. Wang, B. Wang, A. Wang, S. Li, and Z. Sun, "Resource Allocation for D2D Video Multicast Using Multi-Leader Multi-Follower Stackelberg Game With Personalized Incentives," *IEEE Access*, vol. 7, pp. 117 019–117 028, 2019.

[34] D. Ray and R. Vohra, "Coalition Formation," in *Handbook of Game Theory with Economic Applications*. Elsevier, 2015, pp. 239–326.

[35] L. Song, D. Niyato, Z. Han, and E. Hossain, *Wireless device-to-device communications and networks*. Cambridge University Press, 2015.

[36] T. Wang, L. Song, Z. Han, and W. Saad, "Overlapping coalition formation games for emerging communication networks," *IEEE Netw.*, vol. 30, no. 5, pp. 46–53, 2016.

[37] L. Song, D. Niyato, Z. Han, and E. Hossain, "Game-theoretic resource allocation methods for device-to-device communication," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 136–144, 2014.

[38] H. Boche and M. Schubert, "The structure of general interference functions and applications," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4980–4990, Nov. 2008.

[39] F.-L. Luo, *Signal Processing for 5G*. John Wiley and Sons Ltd, 2016. [Online]. Available: https://www.ebook.de/de/product/26667016/fa_long_luo_signal_processing_for_5g.html

[40] W. Dinkelbach, "On Nonlinear Fractional Programming," *Management Science*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[41] S. Ak, H. Inaltekin, and H. V. Poor, "Gaussian approximation for the downlink interference in heterogeneous cellular networks," in *2016 IEEE International Symposium on Information Theory (ISIT)*, Jul. 2016, pp. 1611–1615.

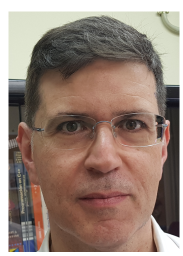
[42] S. Ak, H. Inaltekin, and H. V. Poor, "A Tractable Framework for the Analysis of Dense Heterogeneous Cellular Networks," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 3151–3171, Jul. 2018.

[43] M. Hmila and M. Fernández Veiga, "Energy-efficient power control and clustering in underlay device to multi-device communications," in *Lecture Notes in Computer Science*. Springer International Publishing, 2017, pp. 194–206.

[44] X. Lin, R. Ratasuk, A. Ghosh, and J. G. Andrews, "Modeling, analysis, and optimization of multicast device-to-device transmissions," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4346–4359, Aug. 2014.

[45] 3rd Generation Partnership Prject (3GPP), "5G NR User Equipment (UE) radio transmission and reception - Part2: Range 2 Standalone (3GPP TS 38.101-2 version 15.3.0 Release15)." [Online]. Available: https://www.etsi.org/deliver/etsits/138100138199/13810102/15.03.0060/ts13810102v150300p.pdf

Mariem Hmila received a B.E. degree in computer science and a M.S. degree in intelligent information system from the High Institute of Computer Science and Management (ISIGK), Kairouan, in 2012 and 2014, respectively. In 2019, she received a Ph.D. degree from the University of Vigo, Pontevedra, Spain. Her research interests include device-to-device multicast communication and green communication.



Manuel Fernández-Veiga (M'96–SM'12) received the M.Sc. and Ph.D. degrees in telecommunication engineering from the University of Vigo, Spain, in 1992 and 2001, respectively, where he is currently an Associate Professor. He has authored over 90 papers in international conferences and journals. His research interests include communication theory, performance analysis of computer networks, and wireless networking.

Miguel Rodríguez-Pérez (M'03–SM'18) received the M.Sc. and Ph.D. degrees in telecommunication engineering from the University of Vigo, Spain, in 2001 and 2006, respectively. He is currently an Associate Professor with the Department of Telematics Engineering, University of Vigo, where he is also an Affiliated Member of the Co-located Networking Laboratory. He has published a book and co-authored over 45 conference and journal papers. His research interests are correctly focused on green networking.

Sergio Herrería-Alonso received the M.Sc and Ph.D. degrees in telecommunication engineering from the University of Vigo, Spain, in 2001 and 2006, respectively. Currently, he is an Associate Professor with the Department of Telematics Engineering, University of Vigo, and an affiliated member of the co-located Networking Laboratory. His research interests include quality of service in the Internet and energy-efficient networking. He has authored or co-authored over 40 papers in peer-reviewed international conferences and journals.