IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 20, 2017, accepted September 4, 2017, date of publication October 6, 2017, date of current version November 7, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2760350

Resource Allocation and Interference Management for D2D-Enabled DL/UL Decoupled Het-Nets

ABDULKADIR CELIK[1], (Member, IEEE), REDHA M. RADAYDEH[1], (Senior Member, IEEE), FAWAZ S. AL-QAHTANI[2], (Member, IEEE), AND MOHAMED-SLIM ALOUINI[1], (Fellow, IEEE)

[1]Computer, Electrical, and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

[2]Department of Electrical and Computer Engineering, Texas A&M University, Doha, Qatar

Corresponding author: Abdulkadir Celik (abdulkadir.celik@kaust.edu.sa)

This work was supported by NPRP from the Qatar National Research Fund under Grant 8-1545-2-657.

ABSTRACT In this paper, resource allocation and interference mitigation are investigated for heterogeneous networks where the lowest tier consists of device-to-device (D2D) cells. In order to alleviate *dead-zone* problem, we first consider downlink/uplink (DL/UL) decoupling user association and quantify its capability on interference management and network-wide D2D performance enhancement. Second, we propose an UL fractional frequency reuse scheme where subband (SB) bandwidths are adaptively determined based on: 1) user equipment (UE) density; 2) e-node-B (eNB) density; and 3) on/off switching frequency of small cells. Obtained results show that the adaptive method significantly reduces the number of outage users. Thereafter, a *novel concatenated bi-partite matching* (CBM) method is proposed for joint SB assignment (SA) and resource block allocation (RA) of cellular UEs. Numerical results show that the CBM provides a close performance to exhaustive solution with greatly reduced running time. The CBM is then extended to a *centralized* mode selection, SA, and RA for D2D cells. Alternatively, we develop *offline* and *online* semi-distributed approaches where a D2D-cell can reuse white-list RBs (WRBs), which are not occupied by the adjacent small cells. In the former, D2D-cell members are not aware of intra-cell and inter-cell interference and uniformly distribute their maximum permissible power to WRBs. In the latter, we put D2D sum rate maximization into a convex form by exploiting the proximity gain of D2D UEs. Online distributed solution is then developed by message passing of dual variables and consistency prices. Finally, virtues and drawbacks of the developed approaches are compared and explained.

INDEX TERMS Downlink/uplink decoupling, dead-zone mitigation, inter-tier interference management, truncated channel inversion, concatenated bi-partite graphs, fractional frequency reuse, geometric programming, subband assignment.

I. INTRODUCTION

To fulfill the future demands of the fifth generation (5G) wireless networks, wireless researchers have paid much attention on spectrum efficient solutions. Such data deluge is a natural outcome of the increasing number of mobile devices, Internet of Things, and smart city infrastructures, which inherently introduce machine type communications to cellular networks. To meet these ambitious demands, ultra-dense HetNets have already been considered as a promising solution since densification of the network have the ability to boost network coverage and capacity, while reducing operational and capital expenditures. Therefore, small eNBs (SeNBs) are likely to be standard in 5G networks due to their advantages such as high speed connectivity, enhanced coverage, and improved battery life [1].

Considering the increasing number of machine-type communications and context aware services/applications, another potential spectral efficient approach is D2D communication which provisions three main advantages over conventional HetNets: 1) *Proximity gain* provides high bit rates, low delay, and high energy efficiency as a result of short range between D2D-UEs (DUEs), 2) *Reuse gain* is obtained by sharing the radio resources exploited by higher level tiers, and 3) *Hop gain* is achieved by eliminating the role of eNB by establishing a direct link between DUEs [2].

In the context of hierarchical tiers of HetNets, where tiers 1 and 2 are formed by macro eNBs (MeNBs) and SeNBs, respectively, DUEs can be treated as the lowest tier which utilizes the radio resources in an opportunistic and non-intrusive manner. In orthogonal frequency-division multiple access (OFDMA)-based HetNets, co-tier (tier-$i \leftrightarrow$ tier-i) and cross-tier (tier-$i \leftrightarrow$ tier-j, $i \neq j$) interference take place when multiple UEs operate on the same resource block (RB). Hence, the multi-tiered structure of Het-Nets induces interference management and resource allocation challenges, which are mainly addressed with new solutions in this paper.

A. RELATED WORKS

Some of the recent research efforts on FFR can be exemplified as follows: Xu *et al.* [3] derived the optimal radius to identify inner and outer cell UEs and quantified the achievable throughput of FFR. Relying on realistic irregularly shaped cells, Gonzalez *et al.* [4] developed a soft fractional frequency reuse (SFR) scheme to improve performance. In order to maximize cell throughput and the coverage quality, Zhang *et al.* [5] considered a distributed antenna aided FFR. When MeNBs employed FFR and SFR, authors of [6] studied the coverage probability for picocell users. *Albeit the significant contributions in [3]–[6], authors only considered CUEs without addressing D2D communications.*

In FFR-aided OFDMA cellular systems, achievable spectral efficiency of D2D communication was analytically investigated in [7]. MeNB performance of FFR and SFR is quantified by characterizing the coverage probability and capacity of D2D links [8]. Vlachos *et al.* [9] exploited FFR as the interference-limiting method and developed a multi-objective cell association scheme to orchestrate DUEs. In particular, a D2D-aware sectorized-FFR algorithm is proposed for MeNBs [10]. In return for a reduced spectral efficiency, authors of [11] dedicated a certain subband for the D2D users.

Our work in this paper differs from [3]–[11] in the following regards: While sectorization of the cell area provides a higher frequency reuse factor (FRF), works in [3]–[9] simply divided the cell coverage into inner and outer regions. Furthermore, proposed FFR frameworks in [3]–[5] and [7]–[11] merely considered macrocells and are not suitable for a multi-tiered networks. Moreover, proposed schemes in [3]–[11] considered FFR merely for DL transmission with equipartitioned subband bandwidths. However, UEs are power limited in UL transmission and diverse in required QoS, thus, amount of available bandwidth for cell edge UEs determine the outage performance. Therefore, a dynamic subband bandwith allocation strategy is necessary to use FFR for UL transmission. Finally, single-tier based FFR schemes are not suitable to be used in HetNets as they do not address joint SA and RA for lower tiers.

On the other hand, recent research efforts on interference management and resource allocation problem can be exemplified as follows: In [12], authors constructed an interference based graph to solve an orthogonal resource allocation for DUEs. Reference [13] proposed a pricing-based joint spectrum and power allocation for decentralized interference coordination among DUEs and CUEs. In [14], authors investigated the problem of optimal matching of D2D links and CUEs to form spectrum-sharing partners by constructing a bi-partite graph weighted by outage probabilities. Wu *et al.* [15] targeted energy-efficiency by casting D2D relay selection problem as a minimum weighted bi-partite matching problem. Jiang *et al.* [16] considered an interference-aware communication model and formulated selective caching, and sender-receiver matching as a maximum weighted matching problem. In [17], resource and power allocation is performed via admission control and power allocation for each DUE pair and its CUE partners. Next, a maximum weight bi-partite matching is developed to select a suitable CUE partner to maximize the overall network throughput.

Proposed joint SA and RA method in this paper differs from above works in the following manners: [12]–[17] did not consider a multi-tiered network where co-tier and cross-tier interference management issue is the most challenging scenario. Noting that [12]–[17] also did not address SB assignment problem, the bi-partite matching approach in [14]–[17] are used just for RA, which has the following shortcomings: First, it executes a single matching between RBs and UEs and is not suitable to assign multiple RBs to UEs with high QoS demands. Second, it is not applicable for joint SB and RB allocation, which is one of the most challenging task for sectorized FFR in HetNets.

Existing work on DUDe can be given as follows: Muhammad *et al.* [18] employed a reverse frequency allocation to mitigate interference between macro and small cells. Considering decoupling, authors of [19] developed a model to characterize the UL rate distribution in a HetNet as a function of the association rules. The work in [20] studied a two-tier Het-Net and presented an analytical framework where the MeNBs and SeNBs operate according to a time division duplexing scheme. In [21], an integer problem is discussed to achieve efficient DUE$\leftrightarrow$cell association in order to minimize the interference caused by DUEs on CUEs. For LTE-Unlicensed infrastructure, [22] introduced a machine learning-based algorithm to jointly optimize user association, spectrum allocation, and load balancing for DUDe Het-Nets. *To the best of authors' knowledge, DUDe's dead-zone mitigation ability and its effect on D2D performance are still not investigated and quantified from a large scale network perspective.*

B. MAIN CONTRIBUTIONS

Main contributions of this paper can be summarized as follows:

1) We show that advantages of DUDe for D2D-enabled Het-Nets are twofold: First, CUEs have more tolerance to D2D interference such that more DUEs can utilize the reuse gain on RBs occupied by CUEs. Second, DUEs achieve a significant performance increase as the

interference from cell-edge macrocell UEs (MUEs) are greatly reduced by DUDe. To the best of our knowledge, this is the first study quantifying the benefits of DUDe for D2D communications in multi-tiered Het-Nets.

2) We develop an uplink sectorized-FFR framework with the consideration of full and truncated channel inversion power control. Unlike conventional equipartitioned SB bandwidth approach, we propose an adaptive technique to determine SB bandwidths based on: i) UE density, ii) eNB density, and iii) ON/OFF switching frequency of smallcells. Since transmission power of UEs and available bandwidth are limiting factors to sustain UL QoS requests, obtained results show that proposed method significantly reduces the number of outage users.
3) On the contrary of single-tier oriented schemes, a novel *concatenated bi-partite matching* (CBM) method is developed for joint SA and RA such that the edge weights between SBs and cells are obtained by the optimal solution of the RBs and UEs bi-partite graph. Even though proposed concatenated bi-partite graphs are limited to single matching, it is then generalized to multiple SBs↔eNBs and RBs↔UEs matches. Thereafter, we develop a fast yet high performance algorithm which is shown to have a very close performance to the optimal exhaustive solution.
4) Finally, the CBM is also applied for a centralized mode selection, SA, and RA for DUEs. Alternatively, *offline* and *online* semi-distributed frameworks are developed, where DUEs within a D2D-cell can reuse the set of RBs which are not occupied by the adjacent smallcells, i.e., white-list RBs (WRBs). In the former, D2D-cell members neglect the intra and inter D2D-cell interference and uniformly distribute their maximum permissible power to WRBs. In the latter, D2D sumrate maximization problem is transformed into a convex form and an online distributed solution is developed by message passing of dual variables and consistency prices.

C. NOTATIONS AND PAPER ORGANIZATION

Throughout the paper, sets and their cardinality are denoted with calligraphic and regular uppercase letters (e.g., $|\mathcal{A}| = A$), respectively. Vectors and matrices are represented in lowercase and uppercase boldface (e.g., $\boldsymbol{a}$ and $\boldsymbol{A}$), respectively. The remainder of the paper is organized as follows: Section II introduces the system model and user association schemes. Section III presents dynamic FFR scheme. Section IV formulates the optimal problem, develop the CBM, and then introduce the proposed SA and RA method. Section V addresses the mode selection, SA and RA assignment of DUEs for the centralized and distributed approaches. Numerical results are presented in Section VI, and Section VII concludes the paper.

TABLE 1. Table of notations.

Table of Notations	
Not.	**Description**
Φ_k	PPP for tier-k with density λ_k, $k \in \{1,2\}$.
Φ_u/Φ_d	PPP for CUEs/DUEs with density λ_u/λ_d.
N_U/N_D	Number of CUEs/DUEs for a given PPP realization.
$\mathcal{S}$	Set of SBs
$\mathcal{R}$	Set of RBs $\mathcal{R} = \bigcup_s \mathcal{R}_s$ where $\mathcal{R}_s$ is the set of RBs within SB_s
$\mathcal{C}$	Set of cells $\mathcal{C} = \bigcup_k \mathcal{C}_k$ where $\mathcal{C}_k$ is the set of tier-k cells.
$\mathcal{U}$	Set of UEs, i.e., $\mathcal{U} = \bigcup_c \mathcal{U}_c$
$\mathcal{M}_c^u$	Set of RBs allocated to CUE_u.
M_c	Maximum number of RBs for $\text{CUE}_u \in \mathcal{U}_c$.
P_k	Transmission power of tier-k eNBs.
$P_{c,u}^r$	Transmission power of CUE_u on RB_r, $P_c^u \leq \bar{P}_u$.
P_d^r	Transmission power of DUE_d on RB_r, $P_d^r \leq \bar{P}_d$.
G_i^j	Composite channel gain between node i and j.
$\flat$	Traffic offloading (i.e., bias) factor.
N_0/B	Thermal noise power spectral density/RB bandwidth.
σ_n^2	Thermal noise power.
$x_{c,u}^{s,r}$	Binary allocation variable for $\text{RB}_r \in \mathcal{R}_s$ to $\text{CUE}_u \in \mathcal{U}_c$.
y_c^s	Binary assignment variable for cell_c to SB_s.
$\mathbb{I}_{c,i}^{s,r}$	Interference received by UE_i on RB_r, $i \in \{u,d\}$.
$\gamma_{c,i}^{s,r}$	SINR of UE_i on RB_r, $i \in \{u,d\}$.
$\mathbb{C}_i$	Aggregate rate for UE_i, $i \in \{u,d\}$.
$\bar{\mathbb{C}}_u$	QoS requirement for CUE_u.
α	Prob. of being active (i.e., ON) for SeNBs.
β	Bandwidth proportion between inner and outer SBs.
$\Gamma_{c,u}^{s,r}$	Tolerable interference limit of CUE_u on RB_r
$\mathcal{E}_c^s$	Edge weights of cost matrix $\boldsymbol{\mathcal{E}}$
$\omega_{c,u}^{s,r}$	Edge weights of cost matrix $\boldsymbol{\Omega}_{\boldsymbol{c}}^{\boldsymbol{s}}$
$\pi_{d,r}^c$	Lag. multiplier as a price of reduction in TILs.
$\varpi_{d,r}^c$	Lag. multiplier for power consumption of DUE_d
ϑ_d^r	Lag. multiplier as a consistency price of interference at DUE_d

II. SYSTEM MODEL

We consider UL transmission of a 3-tiered HetNet where each tier represents a particular cell class, i.e., tier-1 consists macrocells, tier-2 comprises of open-access smallcells, and tier-3 is constituted by D2D-cells which are formed by set of DUEs within a smallcell area.[1] Across tiers, eNBs may differ in terms of DL transmit power, P_k, $(k = 1, 2)$ and in their spatial density. We assume that the spatial distribution of eNBs in the tier-k follows a homogeneous Poisson point process (PPP) Φ_k of density λ_k. Likewise, CUEs and DUEs are spatially distributed as PPPs Φ_u and Φ_d of density λ_u and λ_d, respectively. We note that CUE is a generic term used for MUEs, smallcell UEs (SUEs), and DUEs operating in the cellular mode. The number of CUEs and DUEs within Φ_u and Φ_d are denoted as N_U and N_D, respectively.[2] We denote the index set of all cells by $\mathcal{C} = \bigcup_k \mathcal{C}_k = \{c | c \in [1, C]\}$ where $\mathcal{C}_k$ represents the index set of C_k tier-k cells. Similarly, index set of U UEs is given as $\mathcal{U} = \bigcup_c \mathcal{U}_c$ where $\mathcal{U}_c$ denotes the index set of U_c UEs associated with cell c. $\bar{P}_u$ and $\bar{P}_d$ symbolize the maximum transmission power

[1] Herein, we conjecture that closed-access smallcells have intensive to grant access to cell-edge MUEs in return for compensations given by the operator.

[2] Number of UEs/eNBs is assumed to be finite as we consider a finite region where number of points of a PPP realization is finite almost surely [23].

of CUEs and DUEs, respectively. The set of SBs denoted by $\mathcal{S}$ and the index set of the total R UL-RBs denoted by $\mathcal{R} = \bigcup_s \mathcal{R}_s$ where $\mathcal{R}_s$ represents the index set of R_s RBs within SB_s. Each $\text{CUE}_u \in \mathcal{U}_c$ can be assigned at most M_c RBs, which is a design parameter. Set of $M_c^u \leq M_c$ RBs allocated to $\text{CUE}_u \in \mathcal{U}_c$ is denoted as $\mathcal{M}_c^u$. We note that the terms eNB, cell, and their indices will be used interchangeably throughout the paper. Unless it is stated explicitly otherwise, intracell interference is neglected in centralized approaches since eNBs are assumed to allocate RBs of an assigned SB in an orthogonal fashion.

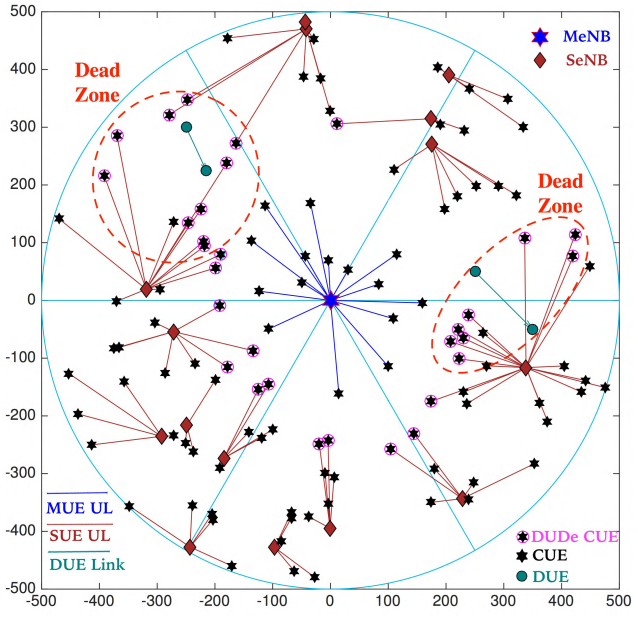


FIGURE 1. Illustration of DUCo/DUDe user association and dead-zone mitigation for $\flat = 0.7$.

In conventional DL/UL coupled (DUCo) scheme, both UL and DL traffic offloading are done by introducing a bias factor, $0 \leq \flat \leq 1$, into the DL received signal strength (RSS) information. Unlike the DUCo, DUDe determines the UL association based on channel gain such that a CUE can be associated with the highest channel gain SeNB in the UL even if it is associated with the MeNB in the DL. For a bias factor of $\flat = 0.7$, a simple DUDe scenario is depicted in Fig. 1, where decoupled CUEs, shown as circled hexstars, are associated with the MeNB and a nearby SeNB for DL and UL, respectively. In DUCo scheme, on the other hand, they are associated with MeNB for both DL and UL. This causes severe interference to nearby SUEs and DUEs, which is a.k.a. *dead-zone* problem.

The composite channel gain between generic nodes i and j is given by

$$G_i^j = K_i^j X_i^j Y_i^j \mathbb{E}\{|H_i^j|^2\} \tag{1}$$

where K_i^j is an antenna parameter constant, $X_i^j = \delta_{i,j}^{-\eta_i^j}$ is the path loss as a function of distance, $\delta_{i,j}$, between the nodes, η_i^j is the path loss exponent, $Y_i^j = 10^{\frac{\xi_i^j}{10}}$ represents the log-normally distributed shadowing, ξ_i^j is a normal random variable representing the variation in received power with a standard deviation of σ_i^j, i.e., $\xi_i^j \sim \mathcal{N}(0, \sigma_i^j)$, and $|H_i^j|^2$ is the exponentially distributed small-scale fading gain from UE_u to eNB_c. In order to avoid ping-pong effects in handover, we especially consider the expected value of $|H_i^j|^2$, $\mathbb{E}\{|H_i^j|^2\}$, throughout the paper. For given user associations, therefore, the signal-to-interference-plus-noise-ratio (SINR) of $\text{CUE}_u \in \mathcal{U}_c$ on $\text{RB}_r \in R_s$ is given by

$$\gamma_{c,u}^{s,r} = \frac{x_{c,u}^{s,r} y_c^s P_c^u G_c^u}{\mathbb{I}_{c,u}^{s,r} + \sigma_n^2}, \quad \forall c \in \mathcal{C}_{1,2},\ \forall u \in \mathcal{U}_c,\ \forall s,\ \forall r \tag{2}$$

where $x_{c,u}^{s,r} \in \{0,1\}$ is a binary variable to indicate the allocation of $\text{RB}_r \in R_s$ for CUE_u, $y_c^s \in \{0,1\}$ is also a binary variable to indicate the assignment of SB_s to cell c, $P_{c,u}^r$ is the transmission power of $\text{CUE}_u \in \mathcal{U}_c$, $\sigma_n^2 = N_0 B$ is the noise variance, B is the bandwidth of a single RB, N_0 is the thermal noise power spectral density $[W/Hz]$ which is assumed to be identical for all nodes, and the aggregated UL interference $\mathbb{I}_{c,u}^{s,r}$ is given as

$$\mathbb{I}_{c,u}^{s,r} = \underbrace{\sum_{\substack{i \in \mathcal{C}_{1,2} \setminus c \\ v \in \mathcal{U}_i}} x_{i,v}^{s,r} y_i^s P_{i,v}^r G_c^v}_{\text{tier}-1/\text{tier}-2\ \text{interference}} + \underbrace{\sum_{\substack{j \in \mathcal{C}_3 \\ d \in \mathcal{U}_j}} x_{j,d}^{s,r} y_j^s P_d^r G_c^d}_{\text{tier}-3\ \text{interference}}, \tag{3}$$

where $\mathcal{C}_{1,2} = \mathcal{C}_1 \cup \mathcal{C}_2$, the first term is the received total interference on eNB_c from CUEs operating on $\text{RB}_r \in \mathcal{R}_s$, and the second term is the total interference from DUEs operating on $\text{RB}_r \in \mathcal{R}_s$. The SINR of DUE_d within D2D-cell c on RB $r \in R_s$ is given by

$$\gamma_{c,d}^{s,r} = \frac{x_{c,d}^{s,r} y_c^s P_d^r G_d^{d'}}{\mathbb{I}_{c,d}^{s,r} + \sigma_n^2}, \quad \forall c \in \mathcal{C}_3,\ \forall d \in \mathcal{U}_c,\ \forall s,\ \forall r, \tag{4}$$

where $G_d^{d'}$ is the composite channel gain between D2D transmitter d and its receiver d', and the aggregated interference is given as

$$\mathbb{I}_{c,d}^{s,r} = \underbrace{\sum_{\substack{i \in \mathcal{C}_{1,2} \\ u \in \mathcal{U}_i}} x_{i,u}^{s,r} y_i^s P_{i,u}^r G_u^{d'}}_{\text{tier}-1/\text{tier}-2\ \text{interference}} + \underbrace{\sum_{\substack{j \in \mathcal{C}_3 \\ v \in \mathcal{U}_j}} x_{j,v}^{s,r} y_j^s P_j^{j'} G_j^{d'}}_{\text{tier}-3\ \text{interference}}, \tag{5}$$

Therefore, aggregated channel capacity for a generic UE_i can be given as

$$\mathbb{C}_i = \sum_{c \in \mathcal{C}} \sum_{s \in \mathcal{S}} \sum_{r \in \mathcal{R}_s} B \log_2 \left(1 + \gamma_{c,i}^{s,r}\right) \ [bps], \quad i \in \{u, d\} \tag{6}$$

For the power control, we previously considered the *full channel inversion power control* where CUEs are enforced to transmit at maximum power even if they are not able to satisfy QoS requirements [24]. As a result, cell-edge CUEs deplete their battery, generate severe interference to other UEs sharing the same RB, and even cause network infeasibility. On the

other hand, *truncated channel inversion power control* [25] eliminates these issues by blocking the unfeasible CUEs as follows

$$P_c^u(\mathcal{M}_c^u, G_c^u) = \begin{cases} \tilde{P}_u/M_c^u, & \text{if } \mathbb{C}_u \geq \overline{\mathbb{C}}_u, M_c^u = M_c \\ 0, & \text{otherwise (truncation outage)} \end{cases}, \tag{7}$$

where the truncation outage event occurs when channel gain of UE cannot satisfy QoS demand even if it is associated with the maximum permissible number of RBs. To avoid outage, CUE_u must guarantee the following condition

$$G_c^u \geq M_c \left(2^{\overline{\mathbb{C}}_u/M_c B} - 1\right) \sigma_n^2 \tag{8}$$

which is derived from the QoS requirement within an interference free environment. Likewise, required minimum number of RBs can also be obtained as

$$M_{c,u}^{\min} = \lceil \min_{z \in \mathbb{R}^+}(z) \text{ s.t. } zB \log_2\left(1 + \frac{\bar{P}_u G_c^u}{z\sigma_n^2}\right) \geq \overline{\mathbb{C}}_u,\ z \geq 1 \rceil \tag{9}$$

which has not a closed form solution, however, it can numerically be evaluated by increasing z from 1 to M_c. That is, if a CUE does not have adequate channel quality to satisfy (8) or $M_{c,u}^{\min} > M_c$, it is in outage and cannot be served.

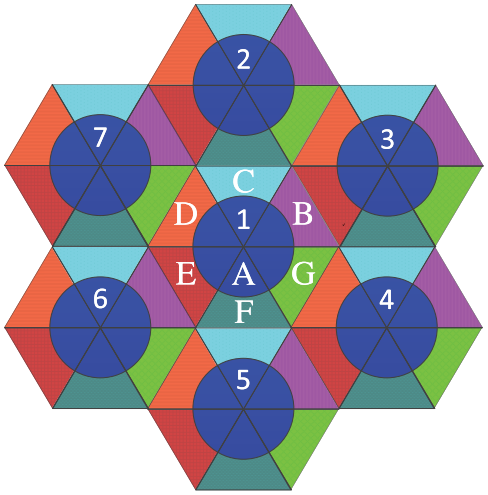


FIGURE 2. Illustration of sectorized-FFR scheme.

III. UL-FFR WITH ADAPTIVE SB BANDWIDTHS

As already advocated by the smallcell forum [26], FFR is a favorable and appropriate method especialy for UL-HetNets as it has low complexity and does not require coordination among MeNBs. For a higher FRF, we prefer a sectorized-FFR scheme to mitigate tier-1↔tier-1 interference by dividing the macrocell area into inner and outer zones, which are further split into N sub-regions as shown in Fig. 2 for $N = 6$, where each color represents a different SB [1]. Even though FFR is generally used as an DL frequency allocation method, it can still provide an extra tier-1 interference mitigation in addition to interference aware power control methods.

Nonetheless, existing works on sectorized-FFR schemes are merely designed for DL transmission and assumes equipartitioned SB bandwidths. When limited transmission power of CUEs considered, however, bandwidth availability, channel conditions, and QoS requirements have an impact on number of users in outage. Therefore, UL-SB bandwidths are required to be determined in an adaptive manner based on UE and eNB densities. In practice, SeNBs operate in an ON/OFF manner where a light loaded SeNB switches off to save energy when traffic load can be transferred to other eNBs. On the other hand, MeNBs operate constantly to provide uninterrupted cellular coverage and connectivity. Defining the inner-zone as the Voronoi tessellation where UEs observe the MeNB as eNB with the highest channel gain, the average number of UEs within the inner-zone can be given as

$$\overline{U}_{in} = \frac{\lambda_u}{\lambda_1 + \alpha\lambda_2} \tag{10}$$

where α is probability of being active for smallcells. Let us consider three non-overlapping user association classes as follows

$$\text{CUE}_u \in \begin{cases} \mathcal{U}_m, & \text{if } P_2\tilde{G}_2^u \leq \flat P_1 \tilde{G}_1^u \\ \mathcal{U}_s, & \text{if } P_1\tilde{G}_1^u < P_2\tilde{G}_2^u \\ \mathcal{U}_b, & \text{if } \flat P_1\tilde{G}_1^u < P_2\tilde{G}_2^u \leq P_1\tilde{G}_1^u \end{cases} \tag{11}$$

where $\mathcal{U}_m$ and $\mathcal{U}_s$ are the set of UEs associated with MeNBs and SeNBs, respectively, $\mathcal{U}_b$ is the set of UEs offloaded/biased to SeNBs, and $\tilde{G}_k^u$ is the highest composite channel gain between CUE_u and the tier-k eNBs, i.e., $\tilde{G}_k^u = \max_{\forall c \in \mathcal{C}_k} \{G_c^u\}$. Following from [27], the probabilities of being in these classes can be derived for ON/OFF switching case as

$$Pr\{u \in \mathcal{U}_m\} = \frac{\lambda_1}{\lambda_1 + \alpha\lambda_2\sqrt{P_2/\flat P_1}} \tag{12}$$

$$Pr\{u \in \mathcal{U}_s\} = \frac{\alpha\lambda_2}{\lambda_1\sqrt{P_1/P_2} + \alpha\lambda_2} \tag{13}$$

$$Pr\{u \in \mathcal{U}_b\} = \left(\frac{\alpha\lambda_2}{\lambda_1\sqrt{\flat P_1/P_2} + \alpha\lambda_2} - \frac{\alpha\lambda_2}{\lambda_1\sqrt{P_1/P_2} + \alpha\lambda_2}\right) \tag{14}$$

Due to the uniformity of users' distribution and non-overlapping class sets, the average number of users in each class is formulated as

$$\overline{U}_m = \frac{\lambda_u}{\lambda_1} Pr\{u \in \mathcal{U}_m\}, \quad \overline{U}_j = \frac{\lambda_u}{\lambda_2} Pr\{u \in \mathcal{U}_j\},\ j \in \{s, b\}, \tag{15}$$

Thus, the average class populations are obtained by substituting (14) into (15) as

$$\overline{U}_m = \frac{\lambda_u}{\lambda_1 + \lambda_2\sqrt{P_2/\flat P_1}}, \quad \overline{U}_s = \frac{\lambda_u}{\lambda_1\sqrt{P_1/P_2} + \lambda_2} \tag{16}$$

$$\overline{U}_b = \left(\frac{\lambda_u}{\lambda_1 \sqrt{\flat P_1/P_2} + \alpha\lambda_2} - \frac{\lambda_u}{\lambda_1 \sqrt{P_1/P_2} + \alpha\lambda_2} \right) \quad (17)$$

Accordingly, the average number of CUEs within outer sub-regions are derived as

$$\overline{U}_{out} = \overline{U}_m + \frac{\alpha\lambda_2}{\lambda_1} \left(\overline{U}_s + \overline{U}_b \right) - \overline{U}_{in} \quad (18)$$

For a bandwidth scale factor, $0 \leq \beta \leq 1$, average bandwidths per UE within inner-zone and outer-zones are given as

$$\overline{W}_{in} = \frac{\beta WR}{\overline{U}_{in}}, \quad \overline{W}_{out} = \frac{(1-\beta)WR}{\overline{U}_{out}} \quad (19)$$

Due to the homogeneous spatial distribution assumption, we require $\overline{W}_{in} = \overline{W}_{out}$ and obtain

$$\beta = \frac{\overline{U}_{in}}{\overline{U}_m + \frac{\alpha\lambda_2}{\lambda_1} \left(\overline{U}_s + \overline{U}_b \right)} = \frac{\lambda_1}{\lambda_1 + \alpha\lambda_2}, \quad (20)$$

which implies that SB bandwidths can be dynamically adapted for different MeNB and SeNB density scenarios. In order to determine SB bandwidths of DUDe, steps in (10)-(20) can be repeated by setting $\flat = P_2/P_1$. So far, we considered tier-1↔tier-1 interference alleviation by adaptively determining the SB bandwidths of sectorized-FFR. SA and RA of SeNBs (SeNBs and MeNBs) in DUCo (DUDe) scheme are addressed in the next section.

IV. INTERFERENCE AND QoS AWARE RB ALLOCATION

Existing FFR schemes mostly deal with single-tier networks and disregard joint SA and RA from a D2D enabling perspective. Assuming that cellular traffic load is satisfied by ensuring certain UL-QoS requirements of CUEs, our goal is to find optimal SA and RA to maximize the D2D sum-rate subject to power constraints and guaranteed QoS levels of CUEs.

A. OPTIMAL RESOURCE ALLOCATION PROBLEM

Considering all possible SB and RA scenarios, $\mathbf{P_0}$ formulates the D2D network capacity maximization problem, where $\boldsymbol{x} = \{x^{s,r}_{c,i} | \forall c, i \in \{u, d\}, s, r\}$ and $\boldsymbol{y} = \{y^s_c | \forall c, s\}$ are the vectorized form of RA and SA binary variables, respectively. Constraints C_1 and C_2 introduce the maximum transmission power constraints for CUEs and DUEs, respectively. C_3 guarantees the necessary QoS requirement for CUEs. C_4 simply states that $\text{RB}_r \in \mathcal{R}_s$ cannot be assigned to $\text{CUE}_u \in \mathcal{U}_c$ unless SB_s is assigned to eNB_c. C_5 ensures that each smallcell is assigned to at least one SB whereas C_6 and C_7 allocate at most M_c RBs for each scheduled CUE in an orthogonal manner. C_8 allocates at least one RB per DUE. In other words, a D2D-cell can be assigned to all S SBs, so that, its members can be allocated up to S RBs each from one of the assigned SBs. C_9 specifies the variable domains.

Due to the integer variables, $\mathbf{P_0}$ is an mixed-integer non-linear programming (MINLP)[3] problem which requires infeasible computational complexity even for moderate sizes of HetNets. Exploiting the nested bi-partite graphs, we solve this MINLP problem in two steps: First, we develop a QoS and interference aware SA and RB assignment for CUEs in the next section. Second, a centralized SA and RB assignment for DUEs are considered in Section V-A, which is followed by semi-distributed offline and online solutions in Section V-B and Section V-C, respectively.

$$\mathbf{P_0}: \underset{\boldsymbol{x},\boldsymbol{y}}{\text{maximize}}\ \mathbb{C} = \sum_d \mathbb{C}_d$$

$$C_1: \text{s.t.} \sum_{r,s} P^u_c x^{s,r}_{c,u} \leq \bar{P}_u, \quad \forall c \in \mathcal{C}_{1,2}, u, s, r$$

$$C_2: \sum_{r,s} P^r_d x^{s,r}_{c,d} \leq \bar{P}_d, \quad \forall c \in \mathcal{C}_3, d, s, r$$

$$C_3: \bar{\mathbb{C}}_u \leq \mathbb{C}_u, \quad \forall u$$

$$C_4: x^{s,r}_{c,u} \leq y^s_c, \quad \forall c \in \mathcal{C}_{1,2}, u, r, s$$

$$C_5: 1 \leq \sum_s y^s_c, \quad \forall c \in \mathcal{C}$$

$$C_6: \sum_{c,s,r} y^s_c x^{s,r}_{c,u} \leq M_c, \quad \forall u, c \in \mathcal{C}_{1,2}, s$$

$$C_7: \sum_u y^s_c x^{s,r}_{c,u} = 1, \quad \forall c \in \mathcal{C}_{1,2}, s, r$$

$$C_8: 1 \leq \sum_{s,r} x^{r,s}_{c,d}, \quad \forall c \in \mathcal{C}_3, d$$

$$C_9: x^{s,r}_{c,u}, x^{s,r}_{c,d}, y^s_c \in \{0, 1\}, \quad P^r_{c,u}, P^r_d \in (0 \cup \mathbb{R}^+)$$

B. CONCATENATED BI-PARTITE MATCHING (CBM)

Even in an optimal SB assignment, interference link and QoS violation among UEs who share the same RB are need to be taken into account. Therefore, considering SA and RA in joint manner is of utmost importance for an improved interference management. Accordingly, we develop a CBM approach to assign SBs and RBs in a joint manner.

Let us first consider a master bi-partite graph $\mathcal{G}(\mathcal{C}, \mathcal{S}, \boldsymbol{\mathcal{E}})$ with vertice set of cells, $\mathcal{C}$, vertice set of SBs, $\mathcal{S}$, and SA cost matrix $\boldsymbol{\mathcal{E}} \in \mathbb{R}^{C \times S}$ with entries $\mathcal{E}^s_c$. For each edge pair between eNB_c and SB_s, also consider a sub-bi-partite graph $\mathcal{G}^s_c(\mathcal{U}_c, \mathcal{R}_s, \boldsymbol{\Omega}^s_c)$ with set of users, $\mathcal{U}_c$, set of RBs $\mathcal{R}_s$, and cost matrix $\boldsymbol{\Omega}^s_c \in \mathbb{R}^{U_c \times R_s}$ with entries $\omega^{s,r}_{c,u}$, $u \in \mathcal{U}_c$, $r \in \mathcal{R}_s$. Sub-partite graphs are nested into $\mathcal{G}(\mathcal{C}, \mathcal{S}, \boldsymbol{\mathcal{E}})$ such that entry $\mathcal{E}^s_c$ is obtained from matching solution of $\mathcal{G}^s_c(\mathcal{U}_c, \mathcal{R}_s, \boldsymbol{\Omega}^s_c)$. This is demonstrated for $M_c = 1$ and $\sum_s y^s_c = 1$ in Fig. 3 where the above bi-partite graph represents the SA and its edge weights are obtained from total edge weights of corresponding sub-bi-partite graph which represents the RA. That is, cost of assigning SB_s for eNB_c, $\mathcal{E}^s_c$, is determined by minimum cost of the optimal bi-partite matching between users of eNB_c and RBs of SB_s. However, the CBM cannot be applied for $M_c > 1$ and $\sum_s y^s_c \geq 1$ since a vertice is not allowed to be matched with multiple vertices. Fortunately, this problem can be handled with insertion of $S-1$ replica eNB_c nodes to $\mathcal{G}(\mathcal{C}, \mathcal{S}, \boldsymbol{\mathcal{E}})$ and $M^u_c - 1$ replica UE_u nodes to the $\mathcal{G}^s_c(\mathcal{U}_c, \mathcal{R}_s, \boldsymbol{\Omega}^s_c)$.

[3]Non-linearity of the problem is caused from multiplication and division of binary variables in SINR expressions.

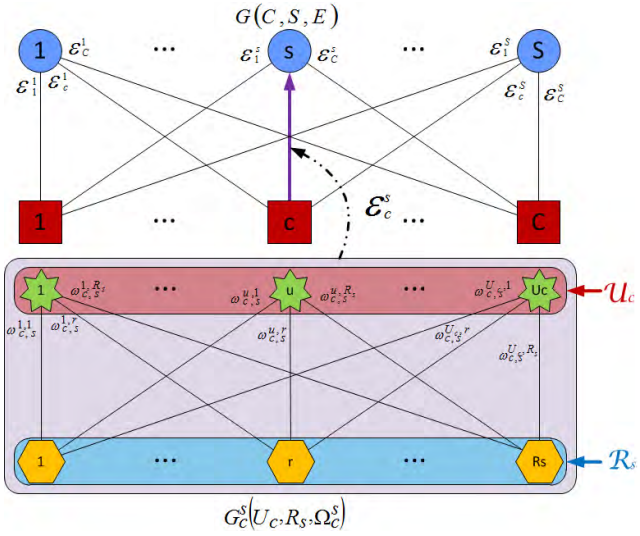


FIGURE 3. **Demonstration of a basic concatenated bi-partite graph for $M_c = 1$ and $\sum_s y_c^s = 1$.**

For a certain M_c^u, CUE_u can tolerate a certain amount of interference without degrading QoS requirement. Accordingly, *tolerable interference limit* (TIL) of $\text{RB}_r \in \mathcal{M}_c^u$ is given in (21), as shown at the bottom of this page, which is derived from aggregated channel capacity expression in (23) where we assume that interference is only tolerated in the RB with the lowest interference.

$$B \sum_{\substack{i \in \mathcal{M}_c^u \\ \mathbb{I}_{c,u}^{s,i} > \mathbb{I}_{c,u}^{s,r}}} \log_2\left(1+\gamma_{c,u}^{s,i}\right) + B\log_2\left(1+\frac{P_c^u G_c^u}{\Gamma_{c,u}^{s,r}+\mathbb{I}_{c,u}^{s,r}+\sigma_n^2}\right) \geq \overline{C}_u \quad (23)$$

If QoS is not satisfied for a certain M_c^u, the inside expression of $[\cdot]^+$ will be negative and no interference will be tolerated. Since an additional user assignment on $\text{RB}_r \in \mathcal{R}_s$ reduces the TIL of CUEs who are already assigned to $\text{RB}_r \in \mathcal{R}_s$, we need to jointly consider the TILs and optimal orthogonal RA in order to evaluate the cost of assigning eNB_c to SB_s, $\mathcal{E}_c^s$.

Therefore, the cost of allocating $\text{RB}_r \in \mathcal{R}_s$ to $\text{CUE}_u \in \mathcal{U}_c$ is formulated as in (22), as shown at the bottom of this page, where the cost is set to total interference of CUE_u on all other CUEs who are allocated to $\text{RB}_r \in \mathcal{R}_s$ if interference from CUE_u does not violate TILs of any of these CUEs. Otherwise, it is set to infinity due to TIL (i.e., QoS) violation. Based on (22), cost of assigning cell c to SB_s can be defined as

$$\mathcal{E}_c^s(\boldsymbol{x}_c^s) = \sum_{u \in \mathcal{U}_c} \sum_{r \in \mathcal{R}_s} x_{c,u}^{s,r} \omega_{c,u}^{s,r} \quad (24)$$

where $\boldsymbol{x}_c^s$ is a solution of sub-bi-partite graph matching and it is subject to intracell orthogonal resource allocation constraints, i.e, $\sum_{r \in \mathcal{R}_c} x_{c,u}^{s,r} = 1, \forall u \in \mathcal{U}_c$, $\sum_{u \in \mathcal{U}_c} x_{c,u}^{s,r} \leq 1$, $\forall r \in \mathcal{R}_s$. Accordingly, we propose a joint SB and RA method in Algorithm 1 which minimizes the total network interference subject to the QoS and maximum transmission power constraints which are already satisfied in (21). It is worth noting that minimizing the total network interference is equivalent to maximizing the available TIL after eNB↔SA so that RBs with higher TIL can host more DUE reuse gain for a given transmission power or higher achievable channel capacity for a given number of DUEs.

In Algorithm 1, Line 1 determines the SB bandwidths using (10)-(20) as explained in Section III. Based on an interference free environment, minimum number of required RBs is evaluated as per (9). In Line 3, CUEs in outage are removed from graphs. Thereafter, RA of MUEs are first performed for DUCo scheme as explained in previous section, and TIL/Cost values initialized. The while loop in Lines 10-22 executes SA and RA for SeNBs (MeNBs and SeNBs) in DUCo (DUDe) as long as there exists some CUEs whose QoS is not guaranteed. At each iteration, Lines 10-22 perform SA and RA and update assignment vectors. Then, edges of decided matchings are eliminated. Lines 16-18 remove the user nodes either if their QoS is guaranteed or not guaranteed even for a maximum number of permissible RBs. In the remaining, M_c^u is increased, subgraphs and TIL/Cost values updated for the next iteration. Note that SA and RA subroutines are in the form of rectangular assignment problem (RAP) which is a generalization of linear assignment problem (LAP). As LAP is generally solved by Munkres Algorithm in the cubic order [28], the main complexity of the Algorithm 1 occurs in line 11 which is in the order of $\mathcal{O}\left(\sum_{s,c \in \mathcal{C}_{1,2}} (\max(M_c U_c, R_s))^3\right)$. This time complexity can significantly be reduced by smart initialization routines and shortest augmenting path algorithms. Therefore, we employ Jonker-Volgenant due to its high performance to handle forbidden (infinite valued) edges even for large rectangular cost matrices [28].

$$\Gamma_{c,u}^{s,r} = \begin{cases} \left[\dfrac{P_c^u G_c^u}{2^{\left(\overline{\mathbb{C}}_u/B - \sum_{i \neq r} \log_2(1+\gamma_{c,u}^{s,i})\right)} - 1} - \mathbb{I}_{c,u}^{s,r} - \sigma_n^2\right]^+, & \text{if } \mathbb{I}_{c,u}^{s,r} < \mathbb{I}_{c,u}^{s,i}, \quad \forall i \in \mathcal{M}_c^u \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

$$\omega_{c,u}^{s,r} = \begin{cases} \sum_{c' \in \mathcal{C} \setminus c} y_{c'}^s \sum_{v \in \mathcal{U}_{c'}} x_{c',v}^{s,r} P_c^u G_{c'}^u, & P_c^u G_{c'}^u \leq \min_{v \in \mathcal{U}_{c'}} (\Gamma_{c',v}^{s,r}), \quad \forall c' \\ \infty, & \text{otherwise (TIL Violation)} \end{cases} \quad (22)$$

Algorithm 1 SB Assignment and Intracell RB Allocation

Input: $\mathcal{C}, \mathcal{S}, \mathcal{R}_s, \mathcal{U}_c$

Output: QoS Guaranteed SA and RA with Minimum Interference.

1: Determine SB bandwidths as per (10)-(20).
2: $M_{c,u}^{min} \leftarrow$ Initialize the minimum number of RBs as in (9).
3: **if** $\{M_{c,u}^{min} > M_c \,||\, G_c^u < M_c(2^{\overline{\mathbb{C}}_u/M_c B} - 1)\sigma_n^2\}$ **then**
4: $\quad \mathcal{U}_c \leftarrow \mathcal{U}_c \setminus \text{UE}_u$ Remove CUE_u from vertice set.
5: **end if**
6: Perform RA for MUEs as in Section III if DUCo is employed.
7: $\boldsymbol{x}_1^\star \leftarrow$ Update RA vector for CUEs.
8: $\Gamma_{c,u}^{s,r}/\boldsymbol{\mathcal{E}}_c^s \leftarrow$ Initialize TIL / Costs
9: $M_c^u \leftarrow 1$
10: **while** $M_c^u < M_c$ & $\mathbb{C}_u < \overline{\mathbb{C}}_u, \exists c, \exists u$ **do**
11: $\quad \left(\mathcal{E}_c^s, \bar{\boldsymbol{x}}_c^s\right) \leftarrow$ Form $\mathcal{E}_c^s$ by substituting $\text{RA}\left(\boldsymbol{\Omega}_c^s\right)$ into (24), $\forall c, s$.
12: $\quad \boldsymbol{y}_1^\star \leftarrow \text{SA}(\boldsymbol{\mathcal{E}})$ Update $\boldsymbol{y}_1^\star$ using SA subroutine
13: $\quad \boldsymbol{x}_1^\star \leftarrow$ Update $\boldsymbol{x}^\star$ based on $\bar{\boldsymbol{x}}_c^s$ corresponding to $\boldsymbol{y}_1^\star$
14: $\quad \mathcal{E}_c^s \leftarrow \infty, \forall c, y_c^s = 1$ Eliminate already matched edges
15: $\quad \omega_{c,u}^{s,r} \leftarrow \infty, \forall c, \forall u, x_{c,u}^{s,r} = 1$ Eliminate already matched edges
16: $\quad$ **if** $\mathbb{C}_u \geq \overline{\mathbb{C}}_u \,||\, \{M_c^u = M_c$ & $\mathbb{C}_u < \overline{\mathbb{C}}_u\}$ **then**
17: $\quad\quad \mathcal{U}_c \leftarrow \mathcal{U}_c \setminus \text{UE}_u$ Remove CUE_u and its replicas, $\forall c, \forall u$
18: $\quad$ **end if**
19: $\quad M_c^u \leftarrow M_c^u + 1$
20: $\quad \mathcal{G}_c^s(\mathcal{R}_s, \mathcal{U}_c, \boldsymbol{\Omega}_c^s) \leftarrow$ Update subgraphs with virtual nodes
21: $\quad \Gamma_{c,u}^{s,r}/\boldsymbol{\mathcal{E}} \leftarrow$ Update TIL / Costs
22: **end while**
return $\boldsymbol{x}^\star, \boldsymbol{y}_1^\star, \Gamma_{c,u}^{s,r}, \boldsymbol{\mathcal{E}}_c^s$

Subband Assignmet (SA) Subroutine

23: **Input:** $\boldsymbol{\mathcal{E}}$

$$\bar{\boldsymbol{y}} \leftarrow \min_{\boldsymbol{y}} \sum_{c,s} y_c^s \mathcal{E}_c^s(\bar{\boldsymbol{x}}_c^s) \text{ (s.t.) } \sum_{s \in \mathcal{S}} y_c^s \leq 1, \sum_{c \in \mathcal{C}_{1,2}} y_c^s = 1$$

24: **Return:** $\bar{\boldsymbol{y}}$

Intracell RB Allocation (RA) Subroutine

25: **Input:** $\boldsymbol{\Omega}_c^s$

$$\bar{\boldsymbol{x}}_c^s \leftarrow \min_{\boldsymbol{x}_c^s} \sum_{\substack{u \in \mathcal{U}_c \\ r \in \mathcal{R}_s}} x_{c,u}^{s,r} \omega_{c,u}^{s,r} \text{ (s.t.) } \sum_{u \in \mathcal{U}_c} x_{c,u}^{s,r} \leq 1, \sum_{r \in \mathcal{R}_s} x_{c,u}^{s,r} = 1$$

26: **Return:** $\bar{\boldsymbol{x}}_c^s$

V. D2D MODE SELECTION AND RESOURCE ALLOCATION

A. CENTRALIZED APPROACH

Contingent upon tier densities and RB availability, the operational mode of DUEs are first selected among cellular mode (CM), dedicated mode (DM), and shared mode (SM). We regard a DUE as CM-capable only if there exists an $\text{eNB}_c^\star$ such that $G_d^{c^\star} > G_d^d$ where $c^\star = \text{argmax}_{c \in \mathcal{C}_{1,2}}(G_d^c)$ and a *cellular mode* CM capable DUE operating on CM associates itself with the $\text{eNB}_c^\star$.[4] Denoting the residual number of unallocated UL-RBs by $\bar{R}_U = R - \sum_{c,u,s,r} x_{c,u}^{s,r}$ and available DL-RBs as $\bar{R}_D$, network density levels can be classified as follows: 1) *Low Density:* $\bar{R}_U \geq N_D$, 2) *Medium Density:* $N_D > \bar{R}_U > 0$, and 3) *High Density:* $\bar{R}_U = 0$.

In low and medium density networks, at most $\bar{R} = \min(\bar{R}_U, \bar{R}_D)$ can operate on CM. In low and medium density networks, the first $\bar{R}$ CM-capable DUEs with the highest $G_d^{c^\star}$ are set to operate on CM. For the remaining UL-RBs, the first $\left[\bar{R}_U - \bar{R}\right]^+$ DUEs with the highest G_d^d are set to operate on DM. Both CM and DM DUEs transmit at maximum transmission power since they have dedicated RBs and our objective is to maximize D2D network capacity. For the remaining unallocated $N_D - \bar{R}_U$ DUEs in the medium density networks and all DUEs in the high density networks, only available option is SM and power allocation must be determined for them. Although CM and DM DUEs enjoys the QoS guarantee by being treated as CUEs, their available RBs can still be reused by other DUEs subject to TIL violation constraints.

The centralized approach is formulated in Algorithm 2 and managed by MeNBs which have the full CSI availability. Algorithm 2 first sorts DUEs with respect to their channel gains. After that, mode selection is performed in lines 3-9 as explained above. After the initialization of cost matrix, while loop between lines 11-18 iteratively performs the SA and RA in order to determine the minimum cost matching as long as there exists feasible cost entries within $\boldsymbol{\mathcal{E}}$. Since RA subroutine operates in RB level granularity, distant DUEs on different cells can be allocated to the same RB which may still provide desirable performance due to the low transmission power and long distance among themselves. The complexity of mode selection session between lines 1-9 is mainly determined by the sorting operation in lines 1-2 with complexity of $\mathcal{O}(N_D \log N_D)$. Since the remaining part is similar to Algorithm 1, the major complexity of Algorithm 2 can be given by $\mathcal{O}\left(N_D \log N_D + \sum_{s,c \in \mathcal{C}_3} (\max(U_c, R_s))^3\right)$.

B. SEMI-DISTRIBUTED OFFLINE SOLUTION

In the centralized approach, assumption of full CSI availability induces high level of communication overhead especially for expected massive amount of D2D users. Another shortcoming of the centralized approach is the orthogonal RA which does not fully exploit the reuse gain of the D2D communications. Therefore, we develop alternative offline and online semi-distributed solutions orchestrated by MeNBs.

An expeditious solution to the signaling deluge of the centralized approach is sending acknowledgments to D2D-cells to inform the set of *black-list* RBs (BRBs) which are occupied by CUEs within the neighboring smallcells,

[4] Since we focus on UL-HetNets, we assume that the eNBs can assure that the DL capacity is no less than the UL capacity by DL transmission power coordination [29] or admission control strategy [17].

Algorithm 2 Centralized Approach

Input: $\boldsymbol{x_1^\star}, \boldsymbol{y_1^\star}, \Gamma_{c,u}^{s,r}$

$\mathcal{U}_{CM} \leftarrow \text{sort}\left(G_d^{c^\star} | d \in \bigcup_{c\in\mathcal{C}_3} \mathcal{U}_c\right)$ Sort DUEs with respect to $G_d^{c^\star}$

$\mathcal{U}_{DM} \leftarrow \text{sort}\left(G_d^{d} | d \in \bigcup_{c\in\mathcal{C}_3} \mathcal{U}_c\right)$ Sort DUEs with respect to G_d^{d}

if $\bar{R}_U > 0$ **then**

$\mathcal{CM} \leftarrow \mathcal{U}_{CM}\left(1 : \min(\bar{R}_U, \bar{R}_D)\right)$ Determine CM DUE set

$\mathcal{DM} \leftarrow \mathcal{U}_{DM}\left(1 : \bar{R}_U - \min(\bar{R}_U, \bar{R}_D)\right)$ Determine DM DUE set

$\mathcal{SM} \leftarrow \bigcup_{c\in\mathcal{C}_3} \mathcal{U}_c \setminus (\mathcal{CM} \cup \mathcal{DM})$ Determine SM DUE set

else

$\mathcal{SM} \leftarrow \bigcup_{c\in\mathcal{C}_3} \mathcal{U}_c$ All DUEs operate on SM

end if

$\boldsymbol{\mathcal{E}}_c^s \leftarrow$ Initialize costs for D2D cell

while $\{\mathcal{E}_c^s | \mathcal{E}_c^s < \infty\} \neq \emptyset$ **do**

$\left(\mathcal{E}_c^s, \boldsymbol{x}_c^s\right) \leftarrow$ Form $\mathcal{E}_c^s$ by substituting $\text{RA}\left(\boldsymbol{\Omega}_c^s\right)$ into (24), $\forall c, s$.

$\boldsymbol{y_2^\star} \leftarrow \text{SA}(\boldsymbol{\mathcal{E}})$ Update $\boldsymbol{y_2^\star}$ using SA subroutine

$\boldsymbol{x_2^\star} \leftarrow$ Update $\boldsymbol{x^\star}$ based on $\boldsymbol{x}_c^s$ corresponding to $\boldsymbol{y_2^\star}$

$\boldsymbol{\mathcal{E}}_c^s \leftarrow \infty, \forall c, y_c^s = 1$ Eliminate assigned cells for next iteration

$\Gamma_{c,u}^{s,r} / \boldsymbol{\mathcal{E}} \leftarrow$ Update TIL / Costs

end while

return $\boldsymbol{x_2^\star}, \boldsymbol{y_2^\star}$

i.e., smallcells who share a common vertex with the D2D cells. These neighboring cells will be referred to as the first smallcell ring which is illustrated in Fig. 2.c. By doing so, D2D-cell members can occupy *white-list* RBs (WRBs) occupied by the second and higher order of rings. DUEs uniformly allocate their maximum permissible power across the WRBs, which yields low transmission powers on each WRB and thus reduced interference thanks to the distance between the D2D-cell and those in the higher order of rings. Since QoS demands of CUEs are not guaranteed in this offline method, MeNBs are required to update BRBs for D2D-cells. Noting that Algorithm 1 already satisfied the QoS requirements of CUEs, if CUEs within a smallcell area experience unexpected service degradation, MeNB release BRB updates for D2D-cells within the first ring of this smallcell. This BRB updates can be extended for the higher of the rings until the DUE interference settles in desired levels. The complexity of the offline approach is dominantly determined by the mode selection complexity, $\mathcal{O}\left(N_D \log N_D\right)$.

C. SEMI-DISTRIBUTED ONLINE SOLUTION

Albeit its simplicity, the offline solution does not tackle the intra-D2D-cell and inter-D2D-cell interference of DUEs, which may result in a significant performance degradation. Denoting the set of WRBs for DUE_d as $\mathcal{W}_d$ and $\mathcal{D}_r = \{\text{DUE}_d | r \in \mathcal{W}_d\}$, we formulate the D2D utility maximization problem subject to TIL and power constraints as follows

$$\mathbf{P_{on}}: \max_{\boldsymbol{p}} B \sum_{r,d\in\mathcal{D}_r} \log_2\left(1+\gamma_d^r\right)$$

$$C_1\text{: s.t. } \frac{1}{\min\limits_{u\in\mathcal{U}_j}\left(\Gamma_{c,u}^{s,r}\right)} \sum_{d\in\mathcal{D}_r} P_d^r G_c^d \leq 1, \quad \forall c \in \mathcal{C}_{1,2},\ \forall r$$

$$C_2\text{: } \frac{1}{\bar{P}_d} \sum_{r\in\mathcal{W}_d} P_d^r \leq 1, \quad \forall d$$

where $\boldsymbol{p}$ is the vector of DUE powers and

$$\gamma_d^r = \frac{P_d^r G_d^{d'}}{\sum_{\substack{c\in\mathcal{C}_{1,2}\\ \forall u\in\mathcal{U}_c}} P_c^u G_u^{d'} + \sum_{\substack{i\neq d\\ i\in\mathcal{D}_r}} P_i^r G_i^{d'} + \sigma_n^2}.$$

Employing geometric programming (GP), we put $\boldsymbol{P_{on}}$ into a convex form by exploiting useful computational and theoretical properties of posynomials and monomials which are defined as follows [30]

Definition 1: A function $f(\boldsymbol{z}|\kappa, \boldsymbol{e}) : \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ *is defined as a monomial such that*

$$f(\boldsymbol{z}|\kappa, \boldsymbol{a}) = \kappa \prod_{j=1}^{n} z_j^{a_j} = \exp\{\boldsymbol{a}^T\tilde{\boldsymbol{z}} + \tilde{\kappa}\} \tag{25}$$

where $\kappa \geq 0$ *is the scalar multiplicative constant,* $\boldsymbol{a} \in \mathbb{R}^n$ *is the exponential constant vector, and* $\tilde{\boldsymbol{z}}$ *is a vector of* $\tilde{z}_j = \log(z_j)$*, and* $\tilde{\kappa} = \log(\kappa)$*. On the other hand, positive sum of monomials is defined as a posynomial*

$$g(\boldsymbol{f}) = \sum_{i=1}^{m} f_i(\boldsymbol{z_i}|\kappa_i, \boldsymbol{a_i}) = \sum_{i=1}^{m} \exp\{\boldsymbol{a_i}^T\tilde{\boldsymbol{z}} + \tilde{\kappa}_i\} \tag{26}$$

Noting that the constraints of $\boldsymbol{P_{on}}$ are already in a posynomial form (also linear), the objective can be put into a posynomial form as follows

$$\max_{\boldsymbol{p}} \left\{ B \sum_{r,d\in\mathcal{D}_r} \log_2\left(1+\gamma_d^r\right) \right\} \simeq \max_{\boldsymbol{p}} \left\{ \sum_{r,d\in\mathcal{D}_r} \log_2\left(\gamma_d^r\right) \right\}$$
$$= \max_{\boldsymbol{p}} \left\{ \log_2\left(\prod_{r,d\in\mathcal{D}_r} \gamma_d^r \right) \right\} = \min_{\boldsymbol{p}} \left\{ \prod_{r,d\in\mathcal{D}_r} \frac{1}{\gamma_d^r} \right\} \tag{27}$$

where we assumed high SINR regime thanks to proximity gain of DUEs.

$\boldsymbol{P_{on}}$ can further be put into a distributed form using dual decomposition methods to separate main dual problem into subproblems for each DUE, which are solved jointly and iteratively using dual variables and consistency prices via message passing. Therefore, DUEs need to know two important metrics: 1) Total interference of DUEs received from all tiers, and 2) Total interference received by eNBs from DUEs along with its TIL amount, which imposes extensive communication overhead. Thus, we mitigate the former problem by

introducing following auxiliary variable

$$\phi_d^r = \sum_{\substack{c\in\mathcal{C}_{1,2}\\ \forall u\in\mathcal{U}_c}} P_c^u G_u^{d'} + \sum_{\substack{i\neq d\\ i\in\mathcal{D}_r}} P_i^r G_i^{d'} \tag{28}$$

which is assumed to be perfectly estimated and broadcasted by DUE_d as a consistency price of the received interference. On the other hand, the latter is resolved by another auxiliary variable

$$\psi_{d,r}^c = \sum_{\substack{i\in\mathcal{D}_r\\ i\neq d}} P_i^r G_j^i - \min_{u\in\mathcal{U}_j}\left(\Gamma_{c,u}^{s,r}\right) \tag{29}$$

which is updated and shared by eNBs as a consistency price of TIL violation. Accordingly, $\boldsymbol{P_{on}}$ can be put into a GP form by change of variables as follows

$$\mathbf{P}_{\mathbf{on}}^{\mathbf{gp}}\colon \min_{\tilde{\boldsymbol{p}}} \sum_{r,d\in\mathcal{D}_r} \log\left(\rho_d^r\right)$$

$$C_1\colon \frac{1}{\tilde{\psi}_{d,r}^c}\exp\{\tilde{P}_d^r\}G_c^d \le 1, \quad \forall r,\ \forall d\in\mathcal{D}_r,\ \forall c\in\mathcal{C}_{1,2}$$

$$C_2\colon \frac{1}{\bar{P}_d}\sum_{r\in\mathcal{W}_d}\exp\{\tilde{P}_d^r\} \le 1, \quad \forall d$$

$$C_3\colon \sum_{c\in\mathcal{C}_{1,2}} P_c^u G_u^{d'} + \sum_{\substack{i\neq d\\ i\in\mathcal{D}_r}}\exp\{\tilde{P}_i^r\}G_i^{d'} - \exp\{\tilde{\phi}_d^r\} = 0, \quad \forall d$$

where $\tilde{z} \triangleq \log(z)$ and $\rho_d^r = \frac{\exp\{-\tilde{P}_d^r\}\left(\exp\{\tilde{\phi}_d^r\}+\sigma_n^2\right)}{G_d^{d'}}$. Thanks to convexity, $\mathbf{P}_{\mathbf{on}}^{\mathbf{gp}}$ has zero duality gap under mild conditions [31], thus, the following Lagrange dual problem solves $\mathbf{P}_{\mathbf{on}}^{\mathbf{gp}}$.

$$\mathbf{P}_{\mathbf{dual}}^{\mathbf{on}}: \max\ \chi\left(\boldsymbol{\pi},\boldsymbol{\varpi},\boldsymbol{\vartheta}\right)$$
$$\text{s.t. } \boldsymbol{\pi}\succeq\mathbf{0}, \quad \boldsymbol{\varpi}\succeq\mathbf{0}$$

where $\boldsymbol{\pi}$ and $\boldsymbol{\varpi}$ are Lagrange multipliers, $\boldsymbol{\vartheta}$ is consistency price, $\chi\left(\boldsymbol{\pi},\boldsymbol{\varpi},\boldsymbol{\vartheta}\right) = \min\left(\mathcal{L}\left(\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\pi},\boldsymbol{\varpi},\boldsymbol{\vartheta}\right)\right)$, and

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{p},\boldsymbol{\phi},\boldsymbol{\pi},\boldsymbol{\varpi},\boldsymbol{\vartheta}\right) = &\sum_{r,d\in\mathcal{D}_r}\log\left(\rho_d^r\right)\\ &+\sum_{\substack{r,c\in\mathcal{C}_{1,2}\\ d\in\mathcal{D}_r}}\pi_{d,r}^c\left(\exp\{\tilde{P}_d^r\}G_j^d-\tilde{\psi}_{d,r}^c\right)\\ &+\sum_d\varpi_d\left(\sum_{r\in\mathcal{W}_d}\exp\{\tilde{P}_d^r\}-\bar{P}_d\right)\\ &+\sum_{r,d\in\mathcal{D}_r}\vartheta_d^r\left(\sum_{\substack{c\in\mathcal{C}_{1,2}\\ \forall u\in\mathcal{U}_c}}P_c^uG_u^{d'}\right.\\ &\left.+\sum_{\substack{i\neq d\\ i\in\mathcal{D}_r}}\exp\{\tilde{P}_i^r\}G_i^{d'}-\exp\{\tilde{\phi}_d^r\}\right)\end{aligned}$$

Using the decomposability property of Lagrangian function [30], $\mathbf{P}_{\mathbf{dual}}^{\mathbf{on}}$ can be separated into sub-problems as follows

$$\begin{aligned}\mathcal{L}_d^r = &\log\left(\rho_d^r\right)+\sum_{c\in\mathcal{C}_{1,2}}\pi_{d,r}^c\left(\exp\{\tilde{P}_d^r\}G_j^d-\tilde{\psi}_{d,r}^c\right)\\ &+\varpi_d\left(\exp\{\tilde{P}_d^r\}-\bar{P}_d\right)\\ &+\vartheta_d^r\left(\sum_{\substack{c\in\mathcal{C}_{1,2}\\ \forall u\in\mathcal{U}_c}}P_c^uG_u^{d'}+\sum_{\substack{i\neq d\\ i\in\mathcal{D}_r}}\exp\{\tilde{P}_i^r\}G_i^{d'}-\exp\{\tilde{\phi}_d^r\}\right)\end{aligned} \tag{30}$$

Based on partial Lagrangian in (30), partial dual problem can be solved in a distributed manner using subgradient method [32] which updates the consistency prices and Lagrange multipliers as per

$$\pi_{d,r}^c(t) = \left[\pi_{d,r}^c(t-1)+a(t)\left(\exp\{\tilde{P}_d^r\}G_j^d-\tilde{\psi}_{d,r}^c\right)\right]^+ \tag{31}$$

$$\varpi_d(t) = \left[\varpi_d(t-1)+b(t)\left(\sum_{r\in\mathcal{W}_d}P_d^r-\bar{P}_d\right)\right]^+ \tag{32}$$

$$\begin{aligned}\vartheta_d^r(t) = &\vartheta_d^r(t-1)+c(t)\\ &\times\left(\sum_{\substack{c\in\mathcal{C}_{1,2}\\ \forall u\in\mathcal{U}_c}}P_c^uG_u^{d'}+\sum_{\substack{i\neq d\\ i\in\mathcal{D}_r}}\exp\{\tilde{P}_i^r\}G_i^{d'}-\exp\{\tilde{\phi}_d^r\}\right)\end{aligned} \tag{33}$$

where $a(t)$, $b(t)$, and $c(t)$ are step sizes and $[\cdot]^+$ ensures dual feasibility of Karush-Kuhn-Tucker (KKT) conditions. Finally, optimal power levels can be obtained from stationarity condition of KKT as

$$\frac{\partial\mathcal{L}\left(\boldsymbol{p}(t),\boldsymbol{\phi}(t),\boldsymbol{\pi}(t),\boldsymbol{\varpi}(t),\boldsymbol{\vartheta}(t)\right)}{\partial\tilde{P}_d^r(t)} = 0, \quad \forall d,\ \forall r \tag{34}$$

which yields the optimal power allocation at time instant t as

$$P_d^{r\star}(t) = \min\left(\bar{P}_d, \exp\left\{\tilde{P}_d^{r\star}(t)\right\}\right) \tag{35}$$

where $\tilde{P}_d^{r\star}(t) = \frac{1}{\sum_{c\in\mathcal{C}_{1,2}}\pi_{d,r}^c(t)G_j^d+\varpi_d(t)+\sum_{\substack{i\neq d\\ i\in\mathcal{D}_r}}\vartheta_i^r(t)G_{i'}^d}$.

Online solution is summarized in Algorithm 3 where first three lines are initialization steps. Thereafter, DUEs and eNBs update their Lagrange multipliers until the maximum change between iterations of optimal power levels is less than ϵ or number of iterations exceed a predetermined threshold, T.

VI. NUMERICAL RESULTS AND ANALYSIS

For the simulations, network area is set to 2.5 km × 2.5 km and proximity between DUE pairs are uniformly distributed between 5m and 50m. Unless it is stated explicitly otherwise, we use the default simulation parameters given in Table 2.

Algorithm 3 Semi-Distributed Online Approach

$t \leftarrow 0$
$0 < P_d^r(t) < \bar{P}_d$, $0 < \pi_{d,r}^c(t)$, $0 < \varpi_d(t)$, $0 < \vartheta_d^r(t)$, $\forall j, d, r$
$\Delta(t) \leftarrow 0$
while $\{\Delta(t) > \epsilon \,||\, t < T\}$ **do**
  DUE$_d$ receives $\pi_{d,r}^c(t)$ from eNB$_c$, $\forall c, \forall d, \ \forall r \in \mathcal{W}_d$
  DUE$_d$ receives $\vartheta_i^r(t)$ from DUE$_i$, $\forall d, \ \forall i \in \mathcal{D}_r$, $i \neq d$
  DUE$_d$ evaluates $P_d^{r\star}(t)$ as per (35) and updates $\varpi_d(t)$ and $\vartheta_d^r(t)$ as per (32) and (33), respectively, $\forall d, \forall r \in \mathcal{W}_d$
  DUE$_d$ shares $P_d^{r\star}(t)$ and $\vartheta_d^r(t)$, $\forall d, \ \forall r \in \mathcal{W}_d$
  $t \leftarrow t + 1$
  $\Delta(t) \leftarrow \max_{\forall d, \forall r \in \mathcal{W}_d} \left(P_d^{r\star}(t+1) - P_d^{r\star}(t) \right)$
  eNB$_c$ updates and shares $\pi_{d,r}^c(t)$ as per (31)
end while
return $p^\star$

TABLE 2. Table of parameters.

Par.	Value	Par.	Value	Par.	Value
η_c^u	4	σ_n^2	$-121.45\ dBm$	R	100
σ_c^u	$8\ dB$	B	$180\ kHz$	$\flat$	0.7
$\mathbb{E}\{\lvert H_c^u\rvert^2\}$	1	$\bar{C}_u$	$1.25\ Mbps$	$\bar{P}_u$	$23\ dBm$
K_c^u	$21.5\ dB$	M_c	3	$\bar{P}_d$	$23\ dBm$
P_1	$46\ dBm$	P_2	$30\ dBm$	λ_u	$7\lambda_2$
λ_1	7	λ_2	$15\lambda_1$	λ_d	$2\lambda_u$

A. IMPACT OF USER ASSOCIATION AND POWER CONTROL SCHEMES ON INTERFERENCE SCENARIOS

We start with an investigation into the impacts of user association and power allocation schemes on different interference scenarios. Fig. 4 illuminates the variation in normalized and aggregated average interference levels with respect to traffic offloading bias factor and different CUE/DUE density levels. The first and second row of Fig. 4 depicts the induced interference under full and truncated channel inversion schemes, respectively. While the left column shows the tier-1↔tier-1 interference, the right column depicts the tier-1↔tier-3 interference cases. It is obvious that the truncation reduces the interference levels by half for both tier-1↔tier-1 and tier-1↔tier-3 interference cases under the DUCo association. This is intuitive as truncated inversion forbids the transmission of the users who cannot satisfy a certain signal reception quality, however, the full inversion allows them to transmit at maximum permissible power, which naturally induces high interference on both MUEs and DUEs. As can be seen from Fig. 4a-4d, interference in DUCo scheme increases with CUE density λ_u and bias factor $\flat$. Higher $\flat$ values associates more cell-edge MUEs to the MeNBs, that is, $\flat = 0$ corresponds to the case that all cellular traffic is offloaded to SeNBs. On the other hand, Fig. 4e-4f clearly shows the performance enhancement comes with the DUDe which provides a stabilized interference level by handling cell-edge MUEs with SeNBs even for high λ_u values.

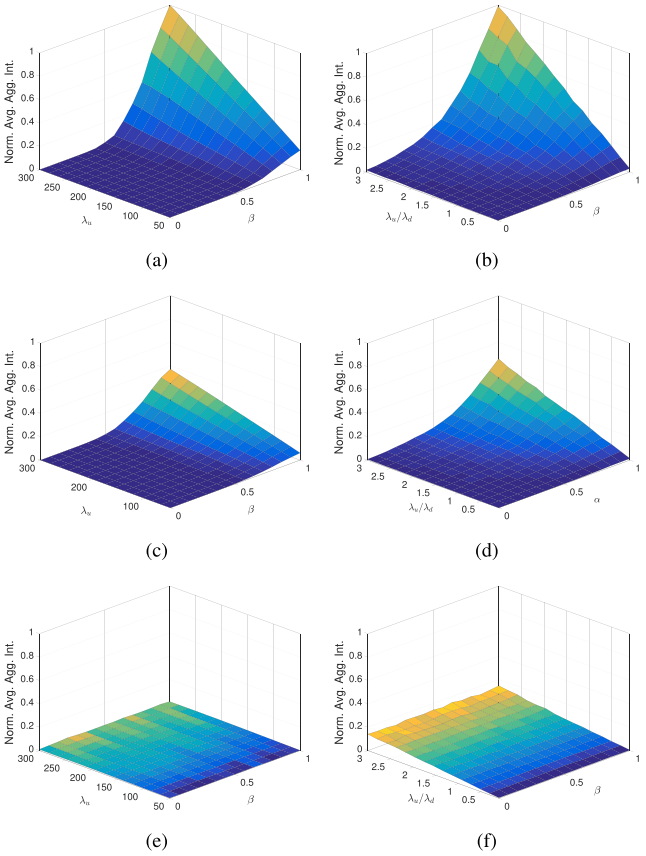


FIGURE 4. Normalized and averaged aggregate interference scenarios under full and truncated channel inversion power control schemes.
(a) DUCo: Tier-1↔Tier-1 interference, with full channel inversion.
(b) DUCo: Tier-1↔Tier-3 interference, with full channel inversion.
(c) DUCo: Tier-1↔Tier-1 interference, truncated channel inversion.
(d) DUCo: Tier-1↔Tier-3 interference, truncated channel inversion.
(e) DUDe: Tier-1↔Tier-1 interference, truncated channel inversion.
(f) DUDe: Tier-1↔Tier-3 interference, truncated channel inversion.

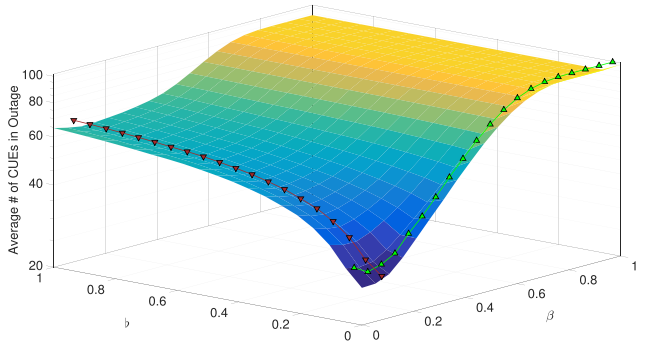


FIGURE 5. Average number of CUEs in outage with respect to bias factor $\flat$ and bandwidth scale factor β.

B. ADAPTIVE FFR AND BANDWIDTH SCALE FACTOR

Fig. 5 demonstrates the effect of bias factor $\flat$ and bandwidth scale factor β for a DUCo scheme with $\alpha = 1$. The increase in outage events is a result of that more and more cell edge users with poor channel conditions are associated with the MeNB for as $\flat$ increases. On the other hand, $\beta = 0$ and $\beta = 1$ are two marginal cases where entire UL bandwidth is merely allocated for inner and outer zone users, respectively.

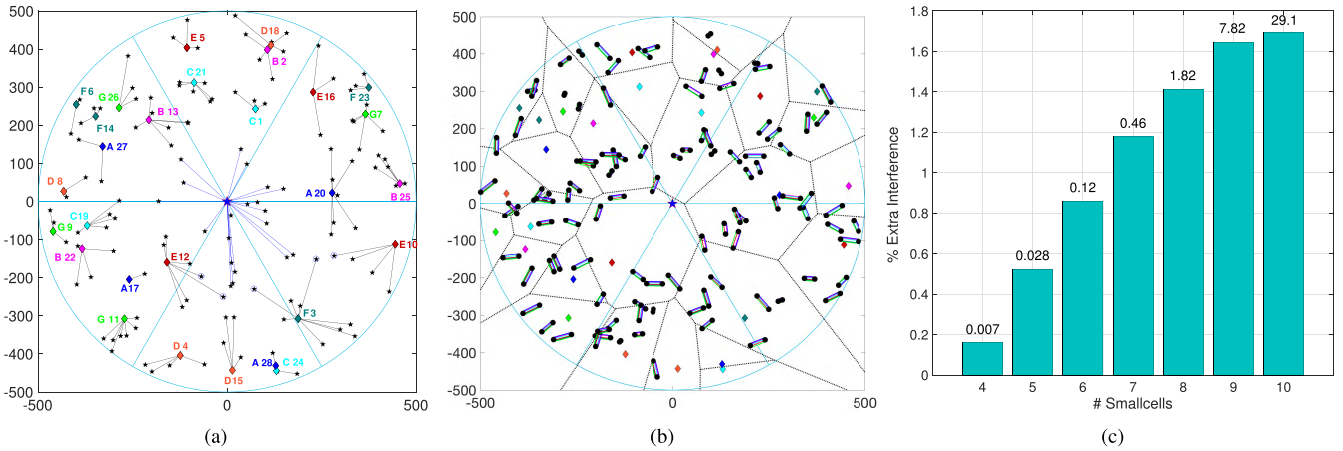


FIGURE 6. (a) Illustration of SA obtained from Algorithm 1. (b) Illustration of RA obtained by Algorithm 2. (c) Comparison of Alg. 1 and exhaustive solution.

Since inner zone users are set to $^{\lambda_u}/_{\lambda_1+\lambda_2}$, the majority of the users fall in outer zone and reduction in their average bandwidth yields high number of outage users. This is mainly connected to the relation between average bandwidth and ρ_u for a certain QoS requirement. That is, a decrease in bandwidth must be compensated by an increase in power to sustain QoS demand. However, this is not possible when the maximum transmission power is reached. Please note that above discussion is an explanation of why truncation and capacity outage is the same in this paper. As pointed out in Section III, the DUDe scheme is a special case of DUCo with $\flat = P_1/P_2$ which is drawn with red colored lower triangle markers. On the other hand, the bandwidth scale factor obtained in (20), $\beta = {}^{\lambda_1}/_{\lambda_1+\lambda_2}$, is drawn with green colored upper triangle markers. From Fig. 5, it is obvious that $\flat = P_1/P_2$ and $\beta = {}^{\lambda_1}/_{\lambda_1+\lambda_2}$ exhibits desirable values. That is, combination of proposed dynamic FFR scheme and DUDe association provides high performance in terms of truncation and capacity outage events.

C. CBM FOR JOINT SA AND RA

In order to provide insights into Algorithm 1, we consider a single macrocell scenario which lies into $500m \times 500m$ area with 28 SeNBs, 130 CUEs, and 100 DUEs as depicted in Fig. 6a where subregions are drawn in light blue colors and SeNB index number appears with assigned SBs in related colors. First of all, the benefit of RB level fine-granularity used both in Algorithm 1 and 2 reveals itself in being able to assign two close proximity SeNBs to the same bandwidth. SeNB pairs, such as (C1-C21, D4-D15, and F6-F14), are assigned to the same band since Algorithm 1 allocates non-overlapping RBs to their members. For example, while SeNB-15 use 5 out of 10 RB within SB D, SeNB 4 exploits 4 RBs from RBs which are not used by SeNB-15. From the SB level coarse-granularity perspective, one can see that Algorithm 1 assign SeNBs (or group of SeNBs) which are far away to the same SB, for example, (C1-C21, C19, C24), (D4-D15, D8, D18), (F3, F6-F14, F23), etc.

D. D2D MODE SELECTION AND RESOURCE ALLOCATION

Contingent upon the SA of SeNBs obtained from Algorithm 1, RA for DUEs is illustrated in Fig. 6b corresponding to Fig. 6a. In Fig. 6b, colored lines between DUEs represents the links on RBs of different SBs. Since Algorithm 2 has a fine-granularity, DUEs are not allocated to RBs which are used nearby CUEs with tight TIL. For example, DUEs within the smallcell areas of SeNBs D4 and D15 are not allocated to SB D RBs since 9 out of 10 RBs are used by CUEs, which can be observed in Fig. 6b where there is no orange color links between DUEs. Similar behaviors can be observed within areas of C1-C21 and F6-F14 for RBs of SBs C and F, respectively. It is also worth mentioning that while most of the cell-edge DUEs are allocated to reuse SB A RBs because they are far away from the inner region and cause low interference due to low transmission power.

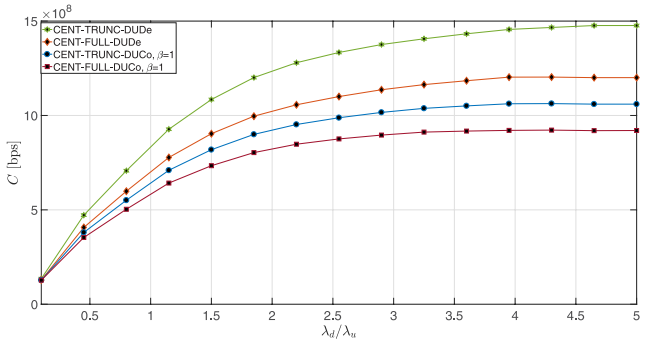
E. COMPARISON OF NESTED BI-PARTITE CONVERSION METHOD WITH EXHAUSTIVE SOLUTION

The comparison between exhaustive solution and proposed nested bi-partite conversion approach is shown in Fig. 6c for Algorithm 1. Sizes of sector, SBs, and RBs are halved (i.e., $N = 3$, $S = 4$, $R = 50$) because considering large number of SBs and RBs requires very large size of combinations. Taking the average of 100 different network realizations, the y-axis shows the extra interference caused from sub-optimality of the proposed approach. For $\lambda_u = R$, the exhaustive solution was able to find SAs with non-overlapping RA, i.e., optimal solution has only thermal noise in the SINR denominator. Thus, the extra interference power percentage of proposed solution is with respect to the thermal noise power, i.e., -121.45 dBm. The numbers at the top of each bar shows the time consumption of the exhaustive solution in units of hours, which are obtained by parallel processing toolbox on a Intel(R) Xeon(R) X5550 processor with 20 cores. Unlike the exhaustive solution, proposed method with the exploitation of Jonker-Volgenant method provides desirable results in at most 10 seconds. As

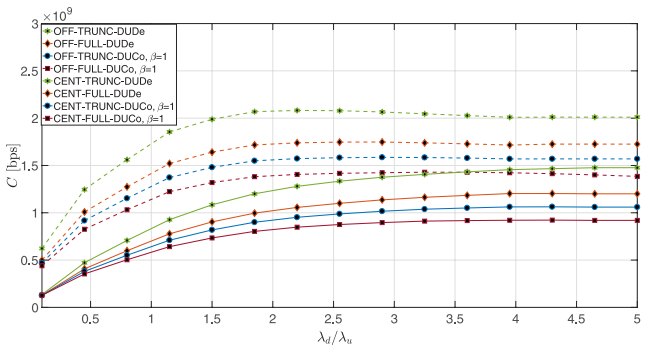
can be seen from Table 3, exhaustive solution of Algorithm 2 takes extremely long time. Noting that both of the algorithms employs nested bi-partite approach in a very similar fashion, we limit ourselves with Fig. 6c.

TABLE 3. Number of combinations for exhaustive solutions.

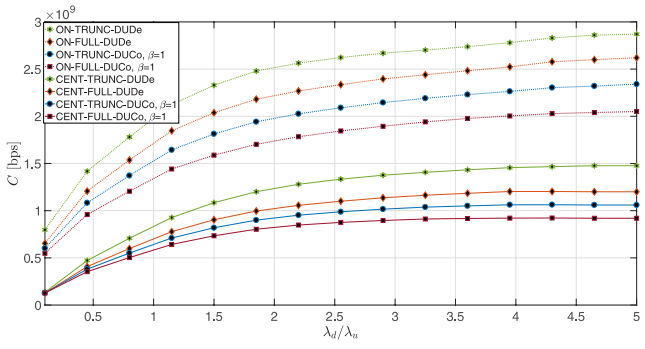
# SBSs	4	5	6	7	8	9	10
Alg. 1	256	1024	4096	1.6×10^4	6.5×10^5	2.6×10^6	1.1×10^6
Alg. 2	6.5×10^5	1.1×10^6	1.7×10^7	2.7×10^8	4.3×10^9	6.9×10^{10}	1.1×10^{12}



(a)



(b)



(c)

FIGURE 7. Performance evaluation of proposed approaches. (a) Centralized Approach. (b) Offline Approach v.s. Centralized Approach. (c) Online Approach v.s. Centralized Approach.

F. PERFORMANCE COMPARISON OF CENTRALIZED AND SEMI-DISTRIBUTED APPROACHES

The D2D aggregate rate performance evaluation is demonstrated for truncated and full channel inversion power control schemes under DUCo and DUDe associations in Fig. 7 where solid, dashed, and dotted lines are used for centralized, offline, and online approaches, respectively. On the other hand, '★', '♦', '•', and '■' markers represent 'DUDe association with truncated inversion', 'DUDe association with full inversion', 'DUCo association with truncated inversion', and 'DUCo association with full inversion', respectively.

The common trend of the aggregate rate curves in Fig. 7 increases with DUE density up to a certain level after which there is no performance increase since the QoS requirement of CUEs behaves as a barrier for reuse gain of D2D communications. Another key observation that is the D2D sumrate is low for under loaded DUE density even if the RB availability is high. The reason is twofold: Firstly, DUEs does not fully exploit the spatial reuse gain, and secondly the transmission power of DUEs are limited despite of relatively high RB availability. For all approaches, aggregate rates of DUDe association outperforms the those of DUCo association. Furthermore, the performance of the truncated inversion is significantly higher than that of full inversion. Indeed, all of these observations are an outcome of the interference comparison made in Section VI-A. Please note that $\flat$ is set to maximum for DUCo scheme and other bias factor values ranges between DUDe and $\flat = 1$ curves.

Fig. 7b (Fig. 7c) evaluates the performance comparison between offline (online) and centralized approaches. It is obvious that both of the distributed methods outperform the centralized approach since the orthogonal allocation does not exploit the reuse gain of D2D communications. That is, centralized approach allocates at most S RBs (each from one of the SBs) while the distributed approaches allocate entire WRBs and distribute the available transmission powers across WRBs. This naturally decrease the transmission power of DUEs in a great amount, cause less interference, and still provide reasonable rates due to close proximity gain of D2D communications. On the other hand, online approach exhibits a much better performance than the offline approach because it takes the intra and inter D2D-cell interference into account. However, it is clear that offline solution still provides desirable performance in between centralized and online approaches for greatly reduced implementation complexity. For the online approach, we initialized the power levels of DUEs by uniformly distributing the available powers to WRBs. Message passing of online method is realized over a hypothetically dedicated control channel in a time scale of long term evolution-Advanced (LTE-A) transmission subframe. During the extensive simulations, we observed that online method converge in 300 iteration on average for $\epsilon = 10^{-4}$ and $T = 1000$.

VII. CONCLUSION

In this study, we investigated the interference management and resource allocation issues of D2D cimmunications underlying HetNets. The interference is mainly mitigated using a D2D-enabled and eNB density adaptive FFR scheme along with DUDe user association. The effect of UL power control

scheme is also considered using full and truncated channel inversion methods. Proposed adaptive SB bandwidth determination method has shown to be reducing the average number of CUEs in outage. For the SB and RB assignment, an MINLP problem is formulated and solved in two steps by developing a nested bi-partite graph conversion. Firstly, SB and RB are assigned for CUEs by ensuring the QoS requirements. Secondly, a similar approach is executed for DUEs by not violating the TIL, i.e., QoS, of CUEs. As an alternative to this orthogonal-centralized method used for DUEs, a semi-distributed approach with offline and online solutions is developed. Extensive simulations are conducted to compare considered schemes and developed approaches, which is also supported by explanation of virtues and drawbacks of each case.

REFERENCES

[1] N. Saquib, E. Hossain, and D. I. Kim, "Fractional frequency reuse for interference management in LTE-advanced hetnets," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 113–122, Apr. 2013.
[2] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, 4th Quart., 2014.
[3] Z. Xu, G. Y. Li, C. Yang, and X. Zhu, "Throughput and optimal threshold for FFR schemes in OFDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2776–2785, Aug. 2012.
[4] D. G. Gonzalez, M. Garcia-Lozano, S. Ruiz Boque, and D. S. Lee, "Optimization of soft frequency reuse for irregular LTE macrocellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2410–2423, May 2013.
[5] J. Zhang, R. Zhang, G. Li, and L. Hanzo, "Distributed antenna systems in fractional-frequency-reuse-aided cellular networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1340–1349, Mar. 2013.
[6] S. Kumar, S. Kalyani, and K. Giridhar, "Spectrum allocation for ICIC-based picocell," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3494–3504, Aug. 2015.
[7] H. Zhu and J. Wang, "Device-to-device communication in cellular networks with fractional frequency reuse," in *Proc. IEEE ICC*, Jun. 2014, pp. 5503–5507.
[8] S. Gupta, S. Kumar, R. Zhang, S. Kalyani, K. Giridhar, and L. Hanzo, "Resource allocation for D2D links in the FFR and SFR aided cellular downlink," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4434–4448, Oct. 2016.
[9] C. Vlachos and V. Friderikos, "MOCA: Multi-objective cell association for device-to-device communications," *IEEE Trans. Veh. Technol.*, to be published.
[10] T. Bansal, K. Sundaresan, S. Rangarajan, and P. Sinha, "R2D2: Embracing device-to-device communication in next generation cellular networks," in *Proc. IEEE INFOCOM*, Apr./May 2014, pp. 1563–1571.
[11] W. Wu and Y. Zhang, "Dedicated resource allocation for D2D communications in cellular systems employing FFR," in *Proc. IEEE WCSP*, Oct. 2014, pp. 1–6.
[12] R. Zhang, X. Cheng, L. Yang, and B. Jiao, "Interference graph-based resource allocation (InGRA) for D2D communications underlaying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3844–3850, Aug. 2015.
[13] R. Yin, G. Yu, H. Zhang, Z. Zhang, and G. Y. Li, "Pricing-based interference coordination for D2D communications in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1519–1532, Mar. 2015.
[14] L. Wang, H. Tang, H. Wu, and G. L. Stüber, "Resource allocation for D2D communications underlay in Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1159–1170, Feb. 2017.
[15] Q. Wu, G. Y. Li, W. Chen, and D. W. K. Ng, "Energy-efficient D2D overlaying communications with spectrum-power trading," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4404–4419, Jul. 2017.
[16] J. Jiang, S. Zhang, B. Li, and B. Li, "Maximized cellular traffic offloading via device-to-device content sharing," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 82–91, Jan. 2016.
[17] D. Feng, L. Lu, Y. Yuan-Wu, G. Y. Li, G. Feng, and S. Li, "Device-to-device communications underlaying cellular networks," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3541–3551, Aug. 2013.
[18] F. Muhammad, Z. H. Abbas, G. Abbas, and L. Jiao, "Decoupled downlink-uplink coverage analysis with interference management for enriched heterogeneous cellular networks," *IEEE Access*, vol. 4, pp. 6250–6260, 2016.
[19] S. Singh, X. Zhang, and J. G. Andrews, "Joint rate and SINR coverage analysis for decoupled uplink-downlink biased cell associations in HetNets," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5360–5373, Oct. 2015.
[20] H. Sun, M. Wildemeersch, M. Sheng, and T. Q. S. Quek, "D2D enhanced heterogeneous cellular networks with dynamic TDD," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4204–4218, Aug. 2015.
[21] H. Elshaer, C. Vlachos, V. Friderikos, and M. Dohler, "Interference-aware decoupled cell association in device-to-device based 5G networks," in *Proc. IEEE 83rd VTC Spring*, May 2016, pp. 1–5.
[22] M. Chen, W. Saad, and C. Yin, "Optimized uplink-downlink decoupling in LTE-U networks: An echo state approach," in *Proc. IEEE ICC*, May 2016, pp. 1–6.
[23] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: General Theory and Structure*, vol. 2. New York, NY, USA: Springer, 2007.
[24] A. Celik, R. M. Radaydeh, F. S. Al-Qahtani, and M.-S. Alouini, "Joint interference management and resource allocation for device-to-device (D2D) communications underlying downlink/uplink decoupled (DUDe) heterogeneous networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
[25] H. ElSawy and E. Hossain, "On stochastic geometry modeling of cellular uplink transmission with truncated channel inversion power control," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4454–4469, Aug. 2014.
[26] "B5G vision: Defining the future HetNet and its evolution to 5G release 7.0," SCF, Bradenton, FL, USA, Tech. Rep. 056.07.01, May 2016. [Online]. Available: http://scf.io/en/documents/056_-_5G_Vision_Defining_the_future_HetNet_and_its_evolution_to_5G.php
[27] H. Ibrahim, H. ElSawy, U. T. Nguyen, and M.-S. Alouini, "Mobility-aware modeling and analysis of dense cellular networks with C-plane/U-plane split architecture," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4879–4894, Nov. 2016.
[28] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, no. 4, pp. 325–340, 1987.
[29] D. Wu, J. Wang, R. Q. Hu, Y. Cai, and L. Zhou, "Energy-efficient resource sharing for mobile device-to-device multimedia communications," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2093–2103, Jun. 2014.
[30] M. Chiang, "Geometric programming for communication systems," *Found. Trends Commun. Inf. Theory*, vol. 2, nos. 1–2, pp. 1–154, 2005.
[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[32] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, Lect. Notes EE392o, 2003.

ABDULKADIR CELIK (S'14–M'16) received the M.S. degree in electrical engineering, the M.S. degree in computer engineering, and the Ph.D. degree with the co-major in electrical engineering and computer engineering from Iowa State University, Ames, IA, USA, in 2013, 2015, and 2016, respectively. He is currently a Post-Doctoral Research Fellow with the Communication Theory Laboratory, King Abdullah University of Science and Technology. His research interests include but not limited to cognitive radio networks, green communications, nonorthogonal multiple access, D2D communications, heterogeneous networks, and optical wireless communications for data center and underwater sensor networks.

REDHA M. RADAYDEH (S'05–M'07–SM'13) was born in Irbid, Jordan, in 1978. He received the B.S. and M.S. degrees in electrical engineering from the Jordan University of Science and Technology (JUST), Irbid, in 2001 and 2003, respectively, and the Ph.D. degree in electrical engineering from The University of Mississippi, Oxford, MS, USA, in 2006. He was with JUST; the King Abdullah University of Science and Technology (KAUST); Texas A&M University at Qatar; and Alfaisal University, Riyadh, Saudi Arabia, as an Associate Professor of electrical engineering. He is currently a Research Consultant with KAUST and also a Researcher with the Department of Electrical Engineering, Texas A&M University, College Station, TX, USA. His research interests include broad topics on wireless communications, and design and performance analysis of wireless networks.

FAWAZ S. AL-QAHTANI (M'10) received the B.Sc. degree in electrical engineering from the King Fahad University of Petroleum and Minerals, Dhahran, Saudi Arabia, in 2000, the M.Sc. degree in digital communication systems from Monash University, Melbourne, VIC, Australia, in 2005, and the Ph.D. degree in electrical and computer engineering from RMIT University, Melbourne, in 2009. He is currently an Assistant Research Scientist with the Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar. He has authored or co-authored over 70 papers in refereed mainstream journals and reputed international conferences. His research interests include channel modeling, applied signal processing, MIMO communication systems, cooperative communications, cognitive radio systems, free-space optical, physical-layer security, visible light communication, device-to-device communication, and power transfer. He received the sponsorship from the Qatar National Research Fund and also from JSERP and NPRP projects, and the Research Excellence Award from Texas A&M University at Qatar in 2013.

MOHAMED-SLIM ALOUINI (S'94–M'98–SM'03–F'09) was born in Tunis, Tunisia. He received the Ph.D. degree in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 1998. He joined the King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, as a Professor of electrical engineering, in 2009. He served as a Faculty Member with the University of Minnesota, Minneapolis, MN, USA, and Texas A&M University at Qatar, Doha, Qatar. His research interests include the modeling, design, and performance analysis of wireless communication systems.