1

Distributed Resource Allocation in 5G Cellular Networks

Monowar Hasan and Ekram Hossain

University of Manitoba, Canada

1.1   Introduction

The fifth generation (5G) cellular networks are expected to provide wide variety of high rate (i.e., 300 Mbps and 60 Mbps in downlink and uplink, respectively, in 95 percent of locations and time [1]) multimedia services. The 5G communication platform is seen as a global unified standard with seamless connectivity among existing standards, e.g., High Speed Packet Access (HSPA), Long Term Evolution-Advanced (LTE-A), and Wireless Fidelity (WiFi). Some of the emerging features and trends of 5G networks are: multi-tier dense heterogeneous networks [2], [3], device-to-device (D2D) and machine-to-machine (M2M) communications [3], [4], densification of the heterogeneous base stations (e.g., extensive use of relays and small cells) [5], cloud-based radio access network [3], integrated use of multiple radio access technologies [6], wireless network virtualization [3], massive and 3D MIMO [3], [7], millimeter wave [8], and full duplex [9] communications.

Conventional 3G systems are single-tier and based on code division multiple access (CDMA) technology. In CDMA systems, all network nodes use the same frequency resource and are distinguished from each other by different pseudo-random spreading codes, which are not exactly orthogonal. Therefore the interference among the nodes is closely related to transmit power. Through efficient power control and spreading schemes [10], [11], interference in CDMA systems can be well-managed. 4G systems (such as LTE/LTE-A) employ orthogonal frequency division multiplexing (OFDM) to improve the spectrum efficiency. Due to unplanned deployment of small cells, resource allocation and interference management in 4G networks is quite different to those in single-tier 3G networks. In a heterogeneous 4G network (which is mainly consists with macro and small cells), there is always a major interferer leading to higher dominant-interference-ratio compared to single-tier network [12]. Therefore, advance interference management schemes such as almost blank

subframe (ABS)[1] and coordinated multi-point transmission (CoMP)[2] have been developed for LTE-A based 4G networks.

In future 5G networks, dense deployment of small cells (e.g., pico and femto cells) and heterogeneous nodes (e.g., relays, low power access points, autonomous M2M sensors etc.) has been envisioned to improve the overall network capacity and spectrum efficiency. The 5G cellular wireless systems will have a multi-tier architecture consisting of macrocells, different types of licensed small cells and D2D/M2M networks to serve users with different quality-of-service (QoS) requirements in a spectrum efficient manner. Radio resource management (e.g., interference mitigation and resource allocation) in such ultra dense 5G systems will be extremely complected due to the irregular and pseudo-random network topology, and therefore, existing management schemes may not be sufficient. For example, it has been shown that in single-tier systems, less than 5% network nodes experience interference from more than two major interferers. However, in 5G heterogeneous networks, this number is expected to be 40%, (e.g., almost half of nodes are affected by more than two interferers) [12], [15].

Considering the dense deployment and large number of network nodes, resource allocation and interference management is one of the fundamental research challenges for such multi-tier heterogeneous networks. In this chapter, we consider the radio resource allocation problem in a multi-tier orthogonal frequency division multiple access (OFDMA)-based cellular (e.g., 5G LTE-A) network. In particular, we present three novel approaches for distributed resource allocation in such networks utilizing the concepts of *stable matching*, *factor graph-based message passing*, and *distributed auction*.

Matching theory, a sub-field of economics, is a promising concept for distributed resource management in wireless networks. The matching theory allows low-complexity algorithmic manipulations to provide a decentralized self-organizing solution to the resource allocation problems. In matching-based resource allocation, each of the agents (e.g., radio resources and transmitter nodes) ranks the opposite set using a preference relation. The solution of the matching is able to assign the resources with the transmitters depending on the preferences.

The message passing approach for resource allocation provides low (e.g., polynomial time) complexity solution by distributing the computational load among the nodes in the network. In the radio resource allocation problems, the decision making agents (e.g., radio resources and the transmitters) form a virtual graphical structure. Each node computes and exchanges simple messages with neighboring nodes in order to find the solution of the resource allocation problem.

Similar to matching-based allocation, auction method is also inherited from economics and used in wireless resource allocation problems. Resource allocation algorithms based on auction method provides polynomial complexity solution which are shown to output near-optimal performance. The auction process evolves with a bidding process, in which unassigned agents (e.g., transmitters) raise the cost and bid for resources simultaneously.

---

[1]The ABS approach coordinates the subframe utilization across different cells in the time domain. Without transmitting data signals, only necessary control signals are transmitted in the subframes that are configured as ABSs in an *aggressor* cell (e.g., the dominant interferer cell). Therefore, the user equipments (UEs) in neighboring cells suffering strong interference can be scheduled with higher data transmission priority. For details refer to [13].

[2]The basic idea of CoMP is to avoid the interference among adjacent cells. This can be achieved by a coordinated spatial domain inter-cell scheduling, or transforming the interfering signals to desired signals via joint transmission and reception among multiple transmission points [14].

Once the bids from all the agents are available, the resources are assigned to the highest bidder.

We illustrate each of the modeling schemes with respect to a practical radio resource allocation problem. In particular, we consider a multi-tier network consisting a macro base station (MBS), a set of small cell base stations (SBSs) and corresponding small cell user equipments (SUEs), as well as D2D user equipments (DUEs). There is a common set of radio resources (e.g., resource blocks [RBs]) available to the network tiers (e.g., MBS, SBSs, and DUEs). The SUEs and DUEs use the available resources (e.g., RB and power level) in an underlay manner as long as the interference caused to the macro-tier (e.g., macro user equipments [MUEs]) remains below a given threshold. The goal of resource allocation is twofold: *i*) to allocate the available RBs and transmit power levels to the SUEs and DUEs in order to maximize the spectral efficiency (which is defined by sum data rate of the SUEs and DUEs); and *ii*) to keep the interference caused to macro-tier (e.g., MUEs) by the underlay transmitters (e.g., SBSs and DUEs) within an acceptable limit. We show that due to the nature of the resource allocation problem, the centralize solution is computationally expensive and also incurs huge signaling overhead. Therefore, it may not be feasible to solve the problem by a single centralized controller node (e.g., MBS) especially in a dense network. Hence distributed solutions with low signaling overhead are desirable.

We assume that readers are familiar with the basics of OFDMA-based cellular wireless networks (e.g., LTE-A networks), as well as have preliminary background on theory of computing (e.g., data structures, algorithms, and computational complexity). Followed by a brief theoretical overview of the modeling tools (e.g., stable matching, message passing, and auction algorithm), we present the distributed solution approaches for the resource allocation problem in the aforementioned network setup. We also provide a brief qualitative comparison in terms of various performance metrics such as complexity, convergence, algorithm overhead etc.

The organization of the rest of the chapter is as follows: beginning with the brief overview of 5G multi-tier network architecture in Section 1.2, we present the system model, related assumptions, and the resource allocation problem is presented in Section 1.3. The disturbed solutions for resource allocation problem, e.g., stable matching, message passing, and auction method are discussed in the Sections 1.4, 1.5, 1.6, respectively. The qualitative comparisons among the resource allocation approaches are presented in Section 1.7. We conclude the chapter in Section 1.8 highlighting the directions for future research.

1.2   Multi-tier 5G Cellular : Overview and Challenges

Beyond the previous standards, one of the key requirements of the 5G systems is to provide the better end-user quality of experience. The visions of 5G networks are included but not limited to the following [2], [5], [16]: *i*) Capacity and throughput improvement (e.g., 1000 times of throughput improvement over 4G, cell data rate 10 Gb/s, and signaling loads less than 1-100%); *ii*) Reduced latency (e.g., 2-5 milliseconds end-to-end latency); *iii*) Network densification (approximately 1000 times higher mobile data per unit area, 100-10000 times higher number of connecting devices); *iv*) Improved energy efficiency (e.g., 10 times prolonged battery life).

A promising solution to meet the expectations of 5G performance requirements (such as high throughput and improved energy efficiency) is to enable multiple tiers in the

network architecture [3], [17]. In addition to conventional macrocell-tier (e.g., a MBS with corresponding MUEs), these heterogeneous network tiers may include low power nodes (e.g., pico and femto cells, relays etc.) as well as wireless peer-to-peer (P2P) nodes (e.g., D2D and M2M UEs, sensors etc.). The heterogeneity of different classes of base stations (e.g., macrocells and small cells) not only improve the spectral efficiency but also provide flexible coverage area. With the reduced cell size in small cells (e.g., pico and femto cells), the area spectral efficiency is increased through higher spectrum reuse. Additionally, the coverage can be improved by deploying indoor small cells (in the spots such as home, office buildings, public vehicles etc.). By reusing exiting cellular radio resources, wireless P2P communication (e.g., D2D/M2M communication among UEs and autonomous sensors) underlaying cellular architecture can significantly increase the overall spectrum and energy efficiency. In addition, the network-controlled P2P communications in 5G systems will allow other nodes (such as relay or M2M gateway), rather than the MBS, to control the communications among P2P nodes [18]. It is worth mentioning that the deployments of heterogeneous nodes in 5G systems will significantly have much higher density than present single-tier networks [5]. The evolution towards multi-tier heterogeneous network in future 5G systems is illustrated in Fig. 1.1.
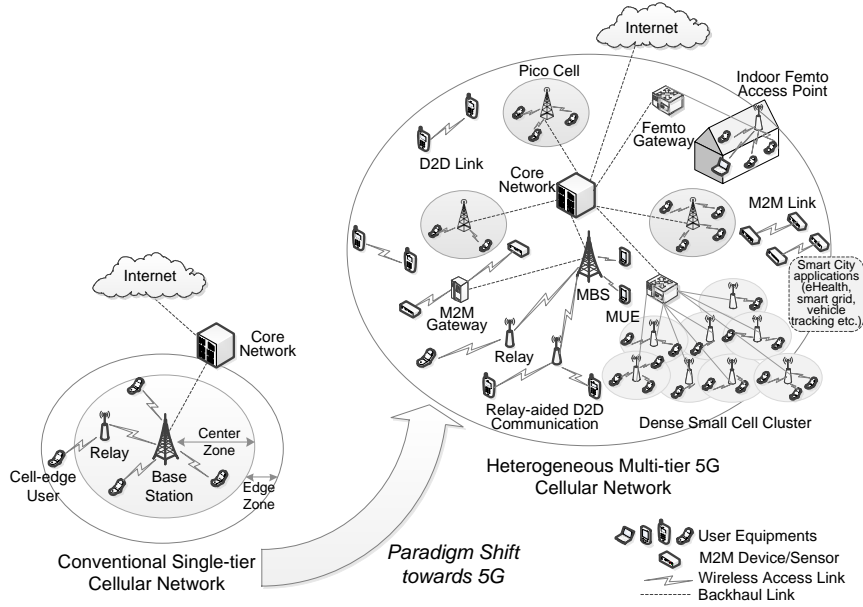
Figure 1.1   From conventional single-tier systems to future heterogeneous multi-tier 5G cellular networks.

1.2.1   *Challenges in Radio Resource Management for Multi-tier Systems*

Regardless of the benefits of multi-tier deployment, as the future heterogeneous network nodes become more dense, the network topology tends to be complicated. Due to the dense deployment of heterogeneous nodes in 5G networks, one approach of improving the resource utilization to use the available resources as a spectrum underlay manner. However, for the underlay communication networks, interference management (e.g., mitigating inter-cell and inter/intra-tier interference) is one of the key challenges. In addition to heterogeneity and dense deployment of wireless devices, coverage and traffic load imbalance due to varying transmit powers of different base stations, different access restrictions (e.g., public, private, hybrid etc.) in different tiers make the interference management and resource allocation problems more challenging than those in conventional single-tier systems. Different channel access priorities and the provision of P2P communication also complicate the dynamics of the interference. Nevertheless, the adoption of multiple tiers in the cellular network architecture will provide better performance in terms of capacity, spectral efficiency, coverage, and power consumption; provided that there exists an efficient inter-tier and intra-tier interference management scheme [17].

Recent studies (e.g., [19], [20], [21]) reveled that the optimum radio resource allocation in future OFDMA-based multi-tier network is generally an NP-hard problem, and hence computationally intractable. The centralized methods (e.g., brute-force, sub-optimal, and heuristic-based approaches) for solving the resource allocation problems in a dense network are not scalable. Besides, the centralize approaches could bring bottleneck to the controller node due to the requirement of global information to manage the whole network tiers (e.g., small cells, D2D/M2M UEs etc.). On the contrary, the distributed or semi-distributed approaches with low signaling overhead are suitable for dense multi-tier networks, where the network nodes (such as SBSs and P2P nodes) perform resource allocation independently or by the minimal assistance of the central controllers (e.g., MBSs). In addition, the distributed methods turns out to be an efficient solution for practical implementations due to reduced computational complexity.

This chapter, therefore, aims to provide an outline of distributed resource management approaches for such multi-tier 5G network architecture. The term *distributed* refers to the fact that the underlay nodes independently determine the allocation with the minimal assistance of MBS. Key mathematical symbols and notations used in the chapter are summarized in Table 1.1.

1.3   **System Model**

1.3.1   *Network Model and Assumptions*

Let us consider a transmission scenario of heterogeneous network as shown in Fig. 1.2. The network consists of one MBS and a set of C cellular MUEs, i.e., $\mathcal{U}^{\mathrm{m}} = \{1, 2, \cdots, C\}$. There are also D D2D pairs and a cluster of S SBSs located within the coverage area of the MBS. The set of SBSs is denoted by $\mathcal{S} = \{1, 2, \cdots S\}$. For simplicity we assume that each SBS serves only one SUE for a single time instance and the set of SUE is given by $\mathcal{U}^{\mathrm{s}}$ where $|\mathcal{U}^{\mathrm{s}}| = S$. The set of D2D pairs is denoted as $\mathcal{U}^{\mathrm{d}} = \{1, 2, \cdots, D\}$. In addition, the d-th element of the sets $\mathcal{U}^{\mathrm{d}_T}$ and $\mathcal{U}^{\mathrm{d}_R}$ denotes the transmitter and receiver UE of the D2D pair $d \in \mathcal{U}^{\mathrm{d}}$, respectively. The set of UEs in the network is given by $\mathcal{U} = \mathcal{U}^{\mathrm{m}} \cup \mathcal{U}^{\mathrm{s}} \cup \mathcal{U}^{\mathrm{d}}$. For

Table 1.1   List of major notations

Notation	Physical Interpretation
● *Network model:*	
$\mathcal{U}^{\mathrm{m}}, \mathcal{U}^{\mathrm{s}}, \mathcal{U}^{\mathrm{d}}$	Set of MUE, SUE, and D2D pairs, respectively
$\mathcal{K}^{\mathrm{T}}, \mathcal{K}^{\mathrm{R}}$	Set of underlay transmitters and receivers, respectively
$\mathcal{N}, \mathcal{L}$	Set of RBs and power levels, respectively
K, N, L	Total number of underlay transmitters, RBs, and power levels, respectively
u_k	The UE associated with underlay transmitter k
$x_k^{(n,l)}, \mathbf{X}$	Allocation indicator, whether transmitter k using resource $\{n,l\}$ and the indicator vector, respectively
$g_{i,j}^{(n)}$	Channel gain between link i, j over RB n
$\gamma_{u_k}^{(n)}$	SINR in RB n for the UE u_k
$\Gamma_{u_k}^{(n,l)}$	Achievable SINR of the UE u_k over RB n using power level l
$p_k^{(n)}$	Transmit power of transmitter k over RB n
R_{u_k}	Achievable data rate for u_k
$I^{(n)}, I_{\max}^{(n)}$	Aggregated interference and threshold limit for the RB n, respectively
$\mathfrak{U}_k^{(n,l)}$	Utility for transmitter k using resource $\{n,l\}$
● *Stable matching:*	
μ	Matching (e.g., allocation) of transmitter to the resources
$i_1 \succeq_j i_2$	Preference relation for agent j (i.e., i_1 is more preferred than i_2)
$\mathfrak{P}_k(\mathcal{N}, \mathcal{L}), \mathfrak{P}_n(\mathcal{K}^{\mathrm{T}}, \mathcal{L})$	Preference profile for the transmitter k and RB n, respectively
● *Message passing:*	
$\delta_{\{n,l\} \to k}\left(x_k^{(n,l)}\right)$	Message delivered by the resource $\{n,l\}$ to the transmitter k
$\delta_{k \to \{n,l\}}\left(x_k^{(n,l)}\right)$	Message from transmitter k to the resource $\{n,l\}$
$\psi_{\{n,l\} \to k}$	Normalized message from the resource $\{n,l\}$ to the transmitter k
$\psi_{k \to \{n,l\}}$	Normalized message from the transmitter k to the resource $\{n,l\}$
$\tau_k^{(n,l)}$	Node marginals for the transmitter k using resource $\{n,l\}$
● *Auction method:*	
$C_k^{(n,l)}$	Cost for transmitter k using resource $\{n,l\}$
$B_k^{(n,l)}$	Data rate (multiplied by a weighting factor) achieved by transmitter k using resource $\{n,l\}$
$\mathfrak{b}_k^{(n,l)}$	Local bidding information available to transmitter k for the resource $\{n,l\}$
ϵ	Minimum bid increment parameter
$\Theta_k = \{n,l\}$	Assignment of resource $\{n,l\}$ to the transmitter k
● *Miscellaneous:*	
$|\mathbf{y}|, |\mathcal{Y}|$	Length (cardinality) of the vector (set) $\mathbf{y}$ ($\mathcal{Y}$)
$y(t)$	Value of variable y at any iteration t
$z := y$	Assignment of the value of variable y to the variable z
/ comment */*	Commented text inside algorithms

notational convenience, we denote by $\mathcal{K}^{\mathrm{T}} = \mathcal{S} \cup \mathcal{U}^{\mathrm{d}_T}$ the set of underlay transmitters (e.g., SBSs and transmitting D2D UEs) and $\mathcal{K}^{\mathrm{R}} = \mathcal{U}^{\mathrm{s}} \cup \mathcal{U}^{\mathrm{d}_R}$ denotes the set of underlay receivers (e.g., SUEs and receiving D2D UEs).

The SBSs and DUEs are underlaid within the *macro-tier* (e.g., MBS and MUEs). Both the macro-tier and the *underlay-tier* (e.g., SBSs, SUEs, and D2D pairs) use the same set
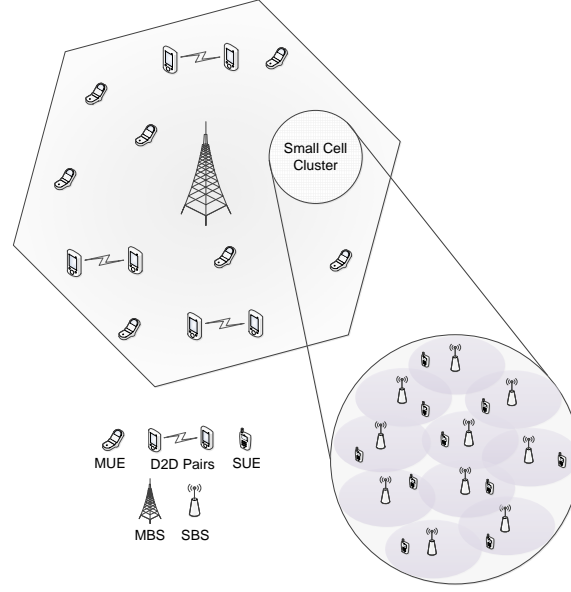
Figure 1.2 Schematic diagram of the heterogeneous network model. The D2D pairs, SBSs, and SUEs are underlaid within the macro-tier by reusing same set of radio resources.

$\mathcal{N} = \{1, 2, \cdots N\}$ of orthogonal RBs[3]. Each transmitter node in the underlay-tier (e.g., SBS and D2D transmitter) selects one RB from the available N RBs. In addition, the underlay transmitters are capable of selecting the transmit power from a finite set of power levels, i.e., $\mathcal{L} = \{1, 2, \cdots L\}$. Each SBS and D2D transmitter should select a suitable RB-power level combination. This RB-power level combination is referred to as *transmission alignment*[4] [23]. For each RB $n \in \mathcal{N}$, there is a predefined threshold $I_{\max}^{(n)}$ for maximum aggregated interference caused by the underlay-tier to the macro-tier. We assume that value of $I_{\max}^{(n)}$ is known to the underlay transmitters by using the feedback control channels. An underlay transmitter (i.e., SBS or transmitter DUE) is allowed to use the particular transmission alignment as long as the cross-tier interference to the MUEs is within the threshold limit.

The system model considered here is a *multi-tier heterogeneous network* since each of the network tiers (e.g., macro-tier and underlay-tier consisting with small cells and D2D UEs) has different transmit power range, coverage region, and specific set of users with different application requirements. It is assumed that the user association to the base stations (either MBS or SBSs) is completed prior to resource allocation. In addition, the potential DUEs are discovered during the D2D session setup by transmitting known synchronization or reference signal (i.e., beacons) [24]. According to our system model, only one MUE is served on each RB to avoid co-tier interference within the macro-tier. However multiple underlay UEs (e.g.,

---

[3]The minimum scheduling unit of LTE-A standard is referred to as an RB. One RB consists of 12 subcarriers (e.g., 180 kHz) in the frequency domain and one sub-feame (e.g., 1 millisecond) in the time domain. For a brief overview of heterogeneous network in the context of LTE-A standard refer to [22, Chapter 1].

[4]Throughout this chapter we use the term *resource* and *transmission alignment* interchangeably.

SUEs and DUEs) can reuse the same RB to improve the spectrum utilization. This reuse causes severe cross-tier interference to the MUEs, and also co-tier interference within the underlay-tier; which leads the requirement of an efficient resource allocation scheme.

1.3.2 Achievable Data Rate

The MBS transmits to the MUEs using a fixed power $p_M^{(n)} > 0$ for $\forall n$. For each underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$, the transmit power over the RBs is determined by the vector $\mathbf{P}_{\mathrm{k}} = \left[p_k^{(1)}, p_k^{(2)}, \cdots, p_k^{(N)} \right]^{\mathsf{T}}$ where $p_k^{(n)} \geq 0$ denotes the the transmit power level of the transmitter k over RB n. The transmit power $p_k^{(n)}$, $\forall n$ must be selected from the finite set of power levels $\mathcal{L}$. Note that if the RB n is not allocated to the transmitter k, the corresponding power variable $p_k^{(n)} = 0$. Since we assume that each underlay transmitter selects only one RB, only one element in the power vector $\mathbf{P}_{\mathrm{k}}$ is non-zero.

All links are assumed to experience independent block fading. We denote by $g_{i,j}^{(n)}$ the channel gain between the links i and j over RB n and defined by $g_{i,j}^{(n)} = \beta_{i,j}^{(n)} d_{i,j}^{-\alpha}$ where $\beta_{i,j}^{(n)}$ denote the channel fading component between link i and j over RB n, $d_{i,j}$ is the distance between node i and j, and α is the path-loss exponent.

For the SUEs, we denote u_k as the SUE associated to SBS $k \in \mathcal{S}$, and for the DUEs, u_k refer to the receiving D2D UE of the D2D transmitter $k \in \mathcal{U}^{\mathrm{d}_T}$. The received signal-to-interference-plus-noise ratio (SINR) for the any arbitrary SUE or D2D receiver, i.e., $u_k \in \mathcal{K}^{\mathrm{R}}, k \in \mathcal{K}^{\mathrm{T}}$ over RB n is given by

$$\gamma_{u_k}^{(n)} = \frac{g_{k,u_k}^{(n)} p_k^{(n)}}{\underbrace{g_{M,u_k}^{(n)} p_M^{(n)}}_{\text{interference from macro-tier}} + \underbrace{\sum_{k' \in \mathcal{K}^{\mathrm{T}}, k' \neq k} g_{k',u_k}^{(n)} p_{k'}^{(n)}}_{\text{interference from underlay-tier}} + \sigma^2} \tag{1.1}$$

where $g_{k,u_k}^{(n)}$ is the link gain between the SBS and SUE (e.g., $u_k \in \mathcal{U}^{\mathrm{s}}, k \in \mathcal{S}$) or the link gain between the D2D UEs (e.g., $u_k \in \mathcal{U}^{\mathrm{d}_R}, k \in \mathcal{U}^{\mathrm{d}_T}$), and $g_{M,u_k}^{(n)}$ is the interference gain between the MBS and the UE u_k. In Equation (1.1), the variable $\sigma^2 = N_0 B_{\mathrm{RB}}$ where B_{RB} is the bandwidth corresponding to an RB and N_0 denotes the thermal noise. Similarly, the SINR for the MUE $m \in \mathcal{U}^{\mathrm{m}}$ over RB n can be written as follows:

$$\gamma_m^{(n)} = \frac{g_{M,m}^{(n)} p_M^{(n)}}{\sum_{k \in \mathcal{K}^{\mathrm{T}}} g_{k,m}^{(n)} p_k^{(n)} + \sigma^2}. \tag{1.2}$$

Given the SINR, the data rate of the UE $u \in \mathcal{U}$ over RB n can be calculated according to the Shannon's formula, i.e., $R_u^{(n)} = B_{\mathrm{RB}} \log_2 \left(1 + \gamma_u^{(n)} \right)$.

1.3.3 Formulation of the Resource Allocation Problem

The objective of resource (i.e., RB and transmit power) allocation problem is to obtain the assignment of RB and power level (e.g., transmission alignment) for the underlay UEs (e.g.,

D2D UEs, and SUEs) that maximizes the achievable sum data rate. The RB and power level allocation indicator for any underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$ is denoted by a binary decision variable $x_k^{(n,l)}$ where

$$x_k^{(n,l)} = \begin{cases} 1, & \text{if the transmitter } k \text{ is trasnmitting over RB } n \text{ with power level } l \\ 0, & \text{otherwise.} \end{cases} \quad (1.3)$$

Note that the decision variable $x_k^{(n,l)} = 1$ implies that $p_k^{(n)} = l$. Let $K = S + D$ denote the total number of underlay transmitters. The achievable data rate of an underlay UE u_k with the corresponding transmitter k is written as

$$R_{u_k} = \sum_{n=1}^{N} \sum_{l=1}^{L} x_k^{(n,l)} B_{\mathrm{RB}} \log_2 \left(1 + \gamma_{u_k}^{(n)} \right). \quad (1.4)$$

The aggregated interference experienced on RB n is given by $I^{(n)} = \sum_{k=1}^{K} \sum_{l=1}^{L} x_k^{(n,l)} g_{k,m_k^*}^{(n)} p_k^{(n)}$, where $m_k^* = \underset{m}{\mathrm{argmax}} \, g_{k,m}^{(n)}$, $\forall m \in \mathcal{U}^{\mathrm{m}}$. In order to calculate the aggregated interference $I^{(n)}$ on RB n we use the concept of reference user [25]. For any RB n, the interference caused by the underlay transmitter k is determined by the highest gains between the transmitter k and MUEs, e.g., the MUE m_k^* who is the mostly affected UE by the transmitter k. Satisfying the interference constraints considering the gain with reference user will also satisfy the interference constraints for other MUEs. As mentioned in Section 1.3.1, an underlay transmitter is allowed to use a particular transmission alignment only when it does not violate the interference threshold to the MUEs, i.e., $I^{(n)} < I_{\max}^{(n)}$, $\forall n$. Mathematically, the resource allocation problem can be expressed by using the following optimization formulation:

$(\mathbf{P1.1})$
$$\max_{x_k^{(n,l)}, \, p_k^{(n)}} \sum_{k=1}^{K} \sum_{n=1}^{N} \sum_{l=1}^{L} x_k^{(n,l)} B_{\mathrm{RB}} \log_2 \left(1 + \gamma_{u_k}^{(n)} \right)$$

subject to:

$$\sum_{k=1}^{K} \sum_{l=1}^{L} x_k^{(n,l)} g_{k,m_k^*}^{(n)} p_k^{(n)} < I_{\max}^{(n)}, \quad \forall n \in \mathcal{N} \quad (1.5\mathrm{a})$$

$$\sum_{n=1}^{N} \sum_{l=1}^{L} x_k^{(n,l)} \leq 1, \qquad \forall k \in \mathcal{K}^{\mathrm{T}} \quad (1.5\mathrm{b})$$

$$x_k^{(n,l)} \in \{0,1\}, \quad \forall k \in \mathcal{K}^{\mathrm{T}}, \, \forall n \in \mathcal{N}, \, \forall l \in \mathcal{L} \quad (1.5\mathrm{c})$$

where

$$\gamma_{u_k}^{(n)} = \frac{g_{k,u_k}^{(n)} p_k^{(n)}}{g_{M,u_k}^{(n)} p_M^{(n)} + \sum_{\substack{k' \in \mathcal{K}^{\mathrm{T}} \\ k' \neq k}} \sum_{l'=1}^{L} x_j^{(n,l')} g_{k',u_k}^{(n)} p_{k'}^{(n)} + \sigma^2}. \quad (1.6)$$

The objective of the resource allocation problem $\mathbf{P1.1}$ is to maximize the data rate of the SUEs and DUEs subject to the set of constraints given by Equations (1.5a)-(1.5c). With the constraint in Equation (1.5a), the aggregated interference caused to the MUEs by the underlay transmitters on each RB is limited by a predefined threshold. The constraint in Equation (1.5b) indicates that the number of RB selected by each underlay transmitter should be at most one and each transmitter can only select one power level at each RB. The binary indicator variable for transmission alignment selection is represented by the constraint in Equation (1.5c).

Corollary 1.3.1 *The resource allocation problem* $\mathbf{P1.1}$ *is a combinatorial non-convex non-linear optimization problem and the centralized solution of the above problem is strongly NP-hard especially for the large set of* $\mathcal{U}$, $\mathcal{N}$, *and* $\mathcal{L}$.

The complexity to solve the above problem using exhaustive search is of $\mathcal{O}\left((NL)^K\right)$. As an example, when $N = 6, L = 3$, and $K = 3 + 2 = 5$, the decision set (e.g., search space) contains 1889568 possible transmission alignments. Considering the computational overhead, it not feasible to solve the resource allocation problem by a single central controller (e.g., MBS) in a practical system; and such centralized solution approach requires all the channel state information (CSI) available to the MBS.

Due to mathematical intractability of solving the above resource allocation problem, in the following we present three distributed heuristic solution approaches, namely, stable matching, factor graph based message passing, and distributed auction-based approaches. The distributed solutions are developed under the assumption that the system is feasible, i.e., given the resources and parameters (e.g., size of the network, interference thresholds etc.), it is possible to obtain an allocation that satisfies all the constraints of the original optimization problem.

1.4   Resource Allocation Using Stable Matching

The resource allocation approach using stable matching involves multiple decision-making agents, i.e., the available radio resources (transmission alignments) and the underlay transmitters; and the solutions (i.e., matching between transmission alignments and transmitters) are produced by individual actions of the agents. The actions, i.e., matching requests and confirmation or rejection are determined by the given *preference profiles*, i.e., the agents hold lists of preferred matches over the opposite set each. The matching outcome yields mutually beneficial assignments between the transmitters and available resources that are individually conducted by such preference lists. In our model, the preference could based on CSI parameters and achievable SINR. *Stability* in matching implies that, with regard to their initial preferences, neither the underlay transmitters nor the MBS (e.g., transmission alignments) have an incentive to alter the allocation.

1.4.1   Concept of Matching

A *matching* (i.e., allocation) is given as an assignment of transmission alignment to the underlay transmitters forming the set $\{k, n, l\} \in \mathcal{K}^{\mathrm{T}} \times \mathcal{N} \times \mathcal{L}$. According to our system model, each underlay transmitter is assigned to only one RB; however, multiple transmitters

can transmit on the same RB to improve spectrum utilization. This scheme corresponds to a *many-to-one* matching in the theory of stable matching. More formally the matching can be defined as follows [26]:

Definition 1.4.1 *A matching μ is defined as a function, i.e., $\mu : \mathcal{K}^{\mathrm{T}} \times \mathcal{N} \times \mathcal{L} \to \mathcal{K}^{\mathrm{T}} \times \mathcal{N} \times \mathcal{L}$ such that*

  i) $\mu(k) \in \mathcal{N} \times \mathcal{L}$ *and* $|\mu_l(n)| \in \{0, 1\}$ *and*
  ii) $\mu(n) \in \{\mathcal{K}^{\mathrm{T}} \times \mathcal{L}\} \cup \{\varnothing\}$ *and* $|\mu(n)| \in \{1, 2, \ldots, K\}$

where $\mu(k) = \{n, l\} \Leftrightarrow \mu(n) = \{k, l\}$ for $\forall k \in \mathcal{K}^{\mathrm{T}}, \forall n \in \mathcal{N}, \forall l \in \mathcal{L}$, and $|\mu(\cdot)|$ denotes the cardinality of matching outcome $\mu(\cdot)$.

The above **Definition 1.4.1** implies that μ is a one-to-one matching if the input to the function is an underlay transmitter. On the other hand, μ is a one-to-many function, i.e., $\mu_l(n)$ is not unique if the input to the function is an RB. The interpretation of $\mu(n) = \varnothing$ implies that for some RB $n \in \mathcal{N}$ the corresponding RB is unused by any underlay transmitter under the matching μ. The outcome of the matching determines the RB allocation vector and corresponding power level, e.g., $\mu \equiv \mathbf{X}$, where

$$\mathbf{X} = \left[x_1^{(1,1)}, \cdots, x_1^{(1,L)}, \cdots, x_1^{(N,L)}, \cdots, x_K^{(N,L)} \right]^{\mathsf{T}}. \tag{1.7}$$

1.4.2   Utility Function and Preference Profile

Let the parameter $\Gamma_{u_k}^{(n,l)} \triangleq \gamma_{u_k | p_k^{(n)} = l}^{(n)}$ denote the achievable SINR of the UE u_k over RB n using power level l (e.g., $p_k^{(n)} = l$) where $\gamma_{u_k}^{(n)}$ is given by Equation (1.6). We express the data rate as a function of SINR. In particular, let $\mathcal{R}\left(\Gamma_{u_k}^{(n,l)} \right) = B_{\mathrm{RB}} \log_2 \left(1 + \Gamma_{u_k}^{(n,l)} \right)$ denote the achievable data rate for the transmitter k over RB n using power level l. The utility of an underlay transmitter for a particular transmission alignment is determined by two factors, i.e., the achievable data rate for a given RB power level combination, and an additional cost function that represents the aggregated interference caused to the MUEs on that RB. In particular, the *utility* of the underlay transmitter k for a given RB n and power level l is given by

$$\mathfrak{U}_k^{(n,l)} = w_1 \mathcal{R}\left(\Gamma_{u_k}^{(n,l)} \right) - w_2 \left(I^{(n)} - I_{\max}^{(n)} \right) \tag{1.8}$$

where w_1 and w_2 are the biasing factors and can be selected based on which network tier (i.e., macro-tier or underlay-tier) should be given priority for resource allocation [23]. As mentioned earlier each underlay transmitter and RB hold a list of preferred matches. The preference profile of an underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$ over the set of available RBs $\mathcal{N}$ and power levels $\mathcal{L}$ is defined as a vector of linear order $\mathcal{P}_k(\mathcal{N}, \mathcal{L}) = \left[\mathfrak{U}_k^{(n,l)} \right]_{n \in \mathcal{N}, l \in \mathcal{L}}$. We denote by $\{n_1, l_1\} \succeq_k \{n_2, l_2\}$ that the transmitter k prefers the transmission alignment $\{n_1, l_1\}$ to $\{n_2, l_2\}$, and consequently, $\mathfrak{U}_k^{(n_1, l_1)} > \mathfrak{U}_k^{(n_2, l_2)}$. Similarly, the each RB holds the preference over the underlay transmitters and power levels given by $\mathcal{P}_n(\mathcal{K}^{\mathrm{T}}, \mathcal{L}) = \left[\mathfrak{U}_k^{(n,l)} \right]_{k \in \mathcal{K}^{\mathrm{T}}, l \in \mathcal{L}}$.

Algorithm 1.1 Assignment of transmission alignments using stable matching

---

Input: The preference profiles $\mathcal{P}_k(\mathcal{N}, \mathcal{L})$, $\forall k \in \mathcal{K}^{\mathrm{T}}$ and $\mathcal{P}_n(\mathcal{K}^{\mathrm{T}}, \mathcal{L})$, $\forall n \in \mathcal{N}$.

Output: The transmission alignment indicator $\mathbf{X} = \left[x_1^{(1,1)}, \cdots, x_1^{(1,L)}, \cdots, x_1^{(N,L)}, \cdots, x_K^{(N,L)} \right]^{\mathsf{T}}$.

1: Initialize $\mathbf{X} := \mathbf{0}$.

2: **while** some transmitter k is unassigned **and** $\mathcal{P}_k(\mathcal{N}, \mathcal{L})$ is non-empty **do**

3:     $\{n_{\mathrm{mp}}, l_{\mathrm{mp}}\} :=$ most preferred RB with power level l_{mp} from the profile $\mathcal{P}_k(\mathcal{N}, \mathcal{L})$.

4:     Set $x_k^{\left(n_{\mathrm{mp}}, l_{\mathrm{mp}}\right)} := 1$. /* *Temporarily assign the RB and power level to the transmitter k* */

5:     $\mathfrak{I}^{(n_{\mathrm{mp}})} := g_{k, m_k^*}^{(n_{\mathrm{mp}})} l_{\mathrm{mp}} + \sum\limits_{\substack{k' \in \mathcal{K}^{\mathrm{T}}, \\ k' \neq k}} \sum\limits_{l'=1}^{L} x_{k'}^{(n_{\mathrm{mp}}, l')} g_{k', m_{k'}^*}^{(n_{\mathrm{mp}})} p_{k'}^{(n_{\mathrm{mp}})}$. /* *Estimate interference of n_{mp}* */

6:     **if** $\mathfrak{I}^{(n_{\mathrm{mp}})} \geq I_{\max}^{(\mathrm{mp})}$ **then**

7:         **repeat**

8:             $\{k_{\mathrm{lp}}, l_{\mathrm{lp}}\} :=$ least preferred transmitter with power level l_{lp} assigned to n_{mp}.

9:             Set $x_{k_{\mathrm{lp}}}^{\left(n_{\mathrm{mp}}, l_{\mathrm{lp}}\right)} := 0$. /* *Revoke assignment due to interference threshold violation* */

10:            $\mathfrak{I}^{(n_{\mathrm{mp}})} := \sum\limits_{k'=1}^{K} \sum\limits_{l'=1}^{L} x_{k'}^{(n_{\mathrm{mp}}, l')} g_{k', m_{k'}^*}^{(n_{\mathrm{mp}})} p_{k'}^{(n_{\mathrm{mp}})}$. /* *Update interference level* */

             /* *Update preference profiles* */

11:            **for each** successor $\{\hat{k}_{\mathrm{lp}}, \hat{l}_{\mathrm{lp}}\}$ of $\{k_{\mathrm{lp}}, l_{\mathrm{lp}}\}$ on profile $\mathcal{P}_{n_{\mathrm{mp}}}(\mathcal{K}^{\mathrm{T}}, \mathcal{L})$ **do**

12:                remove $\{\hat{k}_{\mathrm{lp}}, \hat{l}_{\mathrm{lp}}\}$ from $\mathcal{P}_{n_{\mathrm{mp}}}(\mathcal{K}^{\mathrm{T}}, \mathcal{L})$.

13:                remove $\{n_{\mathrm{mp}}, l_{\mathrm{mp}}\}$ from $\mathcal{P}_{\hat{k}_{\mathrm{lp}}}(\mathcal{N}, \mathcal{L})$.

14:            **end for**

15:        **until** $\mathfrak{I}^{(n_{\mathrm{mp}})} < I_{\max}^{(n_{\mathrm{mp}})}$

16:    **end if**

17: **end while**

1.4.3   Algorithm Development

The matching between transmission alignments to the transmitters is performed in an iterative manner as presented in **Algorithm 1.1**. While a transmitter is unallocated and has a non-empty preference list, the transmitter is temporarily assigned to its first preference over transmission alignments, e.g., the pair of RB and power level, $\{n, l\}$. If the allocation to the RB n does not violate the tolerable interference limit $I_{\max}^{(n)}$, the allocation will persist. Otherwise, until the aggregated interference on the RB n is below threshold, the worst preferred transmitter(s) from the preference list of RB n will be removed even though it was allocated previously. The process terminates when no more transmitters are unallocated. Since the iterative process dynamically updates the preference lists, the procedure above ends up with a local stable matching [27].

The overall stable matching-based resource allocation approach is summarized in **Algorithm 1.2**. Note that **Algorithm 1.1** is executed repeatedly. The convergence of **Algorithm 1.2** occurs when the outcome of two consecutive local matching is similar, e.g., $\mathbf{X}(t) = \mathbf{X}(t-1)$ and as a consequence $R(t) = R(t-1)$, where $R(t) = \sum\limits_{k=1}^{K} R_{u_k}(t)$ denotes the achievable sum rate of the underlay-tier at iteration t.

Algorithm 1.2 Stable matching-based resource allocation

Initialization:
 1: Estimate the CSI parameters from previous time slot.
 2: Each underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$ randomly selects a transmission alignment and the MBS broadcasts the aggregated interference of each RB using pilot signals.
 3: Each underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$ builds the preference profile $\mathcal{P}_k(\mathcal{N}, \mathcal{L})$ from the CSI estimations and the utility function given by Equation (1.8).
 4: For each $n \in \mathcal{N}$, the MBS builds the preference profiles $\mathcal{P}_n(\mathcal{K}^{\mathrm{T}}, \mathcal{L})$.
 5: Initialize number of iterations $t := 1$.

Update:
 6: **while** $\mathbf{X}(t) \neq \mathbf{X}(t-1)$ **and** t is less than some predefined threshold $T_{\max}$ **do**
 7:    MBS obtains a local stable matching $\mathbf{X}(t)$ using **Algorithm 1.1**, calculates the aggregated interference $I^{(n)}(t)$ for $\forall n$ and informs the transmitters.
 8:    Each underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$ updates the preference profile $\mathcal{P}_k(\mathcal{N}, \mathcal{L})$ based on current allocation vector $\mathbf{X}(t)$ and interference level $I^{(n)}(t)$.
 9:    MBS updates the preference profile $\mathcal{P}_n(\mathcal{K}^{\mathrm{T}}, \mathcal{L})$ for $\forall n \in \mathcal{N}$ using $\mathbf{X}(t)$ and $I^{(n)}(t)$.
10:    Update $t := t + 1$.
11: **end while**

Transmission:
12: Use the resources (e.g., the RB and power levels) allocated in the final stage of update phase for data transmission.

1.4.4   Stability, Optimality, and Complexity of the Solution

In this section, we analyze the solution obtained by stable matching approach. The stability, optimality, and the complexity of the algorithm are discussed in the following.

Stability

The notion of stability in the matching μ means that none of the agents (e.g., either underlay transmitters or the resources) prefers to change the allocation obtained by μ. Hence, the matching μ is stable if no transmitter and no resource who are not allocated to each other, as given in μ, prefer each other to their allocation in μ. The transmitters and resources are said to be *acceptable* if the agents (e.g., transmitters and resources) prefer each other to remain unallocated. In addition, a matching μ is called *individually rational* if no agent $\tilde{j}$ prefers unallocation to the matching in $\mu(\tilde{j})$. Before formally defining the stability of matching, we introduce the term *blocking pair* which is defined as

Definition 1.4.2 *A matching μ is **blocked** by a pair of agent (i, j) if they prefer each other to the matching obtain by μ, i.e., $i \succeq_j \mu(j)$ and $j \succeq_i \mu(i)$.*

Using the above definition, the stability of the matching can be defined as follows [28, Chapter 5]:

Definition 1.4.3 *A matching μ is **stable** if it is individually rational and there is no tuple (k, n, l) within the set of acceptable agents such that k prefers $\{n, l\}$ to $\mu(k)$ and n prefers $\{k, l\}$ to $\mu(n)$, i.e., not blocked by any pair of agents.*

The following theorem shows that the solution obtained by the matching algorithm is stable.

Theorem 1.4.4 *The assignment performed in **Algorithm 1.1** leads to a stable allocation.*

Proof. We proof the theorem by contradiction. Let μ be a matching obtained by **Algorithm 1.1**. Let us assume that the resource $\{n, l\}$ is not allocated to the transmitter k, but it belongs to a higher order in the preference list. According to this assumption, the tuple (k, n, l) will block μ. Since the position of the resource $\{n, l\}$ in the preference profile of k is higher compared to any resource $\{\hat{n}, \hat{l}\}$ that is matched by μ, i.e., $\{n, l\} \succeq_k \mu(k)$, transmitter k must select $\{n, l\}$ before the algorithm terminates. Note that, the resource $\{n, l\}$ is not assigned to transmitter k in the matching outcome μ. This implies that k is unassigned with the resource $\{n, l\}$ (e.g., line 9 in **Algorithm 1.1**) and $(k, \hat{n}, \hat{l})$ is a better assignment. As a result, the tuple (k, n, l) will not block μ, which contradicts our assumption. The proof concludes since no blocking pair exists, and therefore, the matching outcome μ leads to a stable matching.

It is worth mentioning that the assignment is stable at each iteration of **Algorithm 1.1**. Since after evaluation of the utility, the preference profiles are updated and the matching subroutine is repeated, a stable allocation is obtained at each iteration.

Optimality

The optimality property of the stable matching approach can be observed using the definition of weak Pareto optimality. Let $\mathcal{R}_\mu$ denote the sum-rate obtained by matching μ. A matching μ is weak Pareto optimal if there is no other matching $\hat{\mu}$ that can achieve a better sum-rate, i.e., $\mathcal{R}_{\hat{\mu}} \geq \mathcal{R}_\mu$ [26].

Theorem 1.4.5 *The stable matching-based resource allocation algorithm is weak Pareto optimal.*

Proof. Let us consider μ to be the stable allocation obtained by **Algorithm 1.1**. For instance, let $\hat{\mu}$ be an arbitrary stable outcome better that μ, i.e., $\hat{\mu}$ can achieve a better sum-rate. Since the allocation $\hat{\mu}$ is better than μ, there exists atleast one resource $\{\hat{n}, \hat{l}\}$ allocated to transmitter k in $\hat{\mu}$, and k is allocated to the resource $\{n, l\}$ in μ. According to our assumption, k prefers $\{\hat{n}, \hat{l}\}$ to $\{n, l\}$, and let $\{\hat{n}, \hat{l}\}$ be allocated to transmitter $\hat{k}$ in μ. It is obvious that resource $\{\hat{n}, \hat{l}\}$ is better than $\{n, l\}$ to transmitter k and $\{k, l\}$ is better than $\{\hat{k}, \hat{l}\}$ to resource $\hat{n}$, i.e., $\{\hat{n}, \hat{l}\} \succeq_k \{n, l\}$ and $\{k, l\} \succeq_{\hat{n}} \{\hat{k}, \hat{l}\}$. By the definition of blocking pair, μ is blocked by $(k, \hat{n}, \hat{l})$ and hence μ is unstable. This contradicts our assumption that μ is a stable allocation. Since there is no stable outcome $\hat{\mu}$ which is better that μ, by definition μ is an optimal allocation.

Complexity

It is possible to show that the stable matching algorithm will iterate for finite number of times.

Theorem 1.4.6 *The RB allocation subroutine terminates after some finite step T'.*

Proof. Let the finite set $\tilde{\mathcal{X}}$ represent the all possible combinations of transmitter-resource matching where each element $\tilde{x}_k^{(n,l)} \in \tilde{\mathcal{X}}$ denotes the resource $\{n, l\}$ is allocated to the transmitter k. Since no transmitter is rejected by the same resource more than once (i.e., line 9 in **Algorithm 1.1**), the finiteness of the set $\tilde{\mathcal{X}}$ ensures the termination of the matching subroutine in finite number of steps.

For each underlay transmitter, the complexity to build the preference profile using any standard sorting algorithm is $\mathcal{O}\left(NL\log(NL)\right)$ (line 8, **Algorithm 1.2**). Similarly, in line 9, the complexity to output the ordered set of preference profile for the RBs is of $\mathcal{O}\left(NKL\log(KL)\right)$. Let $\xi = \sum_{k=1}^{K} |\boldsymbol{\mathcal{P}}_k(\mathcal{N}, \mathcal{L})| + \sum_{n=1}^{N} |\boldsymbol{\mathcal{P}}_n(\mathcal{K}^{\mathrm{T}}, \mathcal{L})| = 2KNL$ be the total length of input preferences in **Algorithm 1.1**, where $|\boldsymbol{\mathcal{P}}_j(\cdot)|$ denotes the length of profile vector $\boldsymbol{\mathcal{P}}_j(\cdot)$. From **Theorem 1.4.6** and [29, Chapter 1] it can be shown that, if implemented with suitable data structures, the time complexity of the RB allocation subroutine is linear in the size of input preference profiles, i.e., $\mathcal{O}(\xi) = \mathcal{O}\left(KNL\right)$. Since the update phase of **Algorithm 1.2** runs at most fixed $T < T_{\max}$ iterations, the complexity of the stable matching-based solution is linear in K, N, L.

1.5   Message Passing Approach for Resource Allocation

In the following, we reformulate the resource allocation problem **P1.1** in such a way that can be solved with a message passing (MP) technique. The MP approach involves computation of the *marginals*, e.g., the messages exchanged between the nodes of a specific graphical model. Among different representations of graphical model, we consider *factor graph* based MP scheme. A factor graph is made up of two different types of nodes, i.e., *function* and *variable* nodes, and an edge connects a function (e.g., factor) node to a variable node if and only if the variable appears in the function. Mathematically, this can be expressed as follows [30]:

Definition 1.5.1 *A factor graph can be represented by a $\mathcal{V}$-$\mathcal{F}$ bipartite graph where $\mathcal{V} = \{v_1, \cdots v_a\}$ is the set of variable nodes and $\mathcal{F} = \{f_1(\cdot), \cdots f_b(\cdot)\}$ is the set of function (e.g., factor) nodes. The connectivity (e.g., edges) of the factor graph can be represented by an $a \times b$ binary matrix $\mathbf{E} = [E_{i,j}]$ where $E_{i,j} = 1$ if the variable node i is connected with the factor node j and $E_{i,j} = 0$, otherwise.*

1.5.1   Overview of the MP Scheme

Before presenting the details resource allocation approach for a heterogeneous scenario, we briefly introduce the generic MP scheme (for the details of factor graph-based MP scheme refer to [30]). Let us consider the maximization of an arbitrary function $f(v_1, \cdots, v_J)$ over all possible values of the argument, i.e., $Z = \max_{\mathbf{v}} f(\mathbf{v})$ where $\mathbf{v} = [v_1, \cdots, v_J]^{\mathrm{T}}$. We denote by $\max_{\mathbf{v}}$ that the maximization is computed over all possible combinations of the elements of the the vector $\mathbf{v}$. The *marginal* of Z with respect to variable v_j is given by $\phi_j(v_j) = \max_{\sim(v_j)} f(\mathbf{v})$ where $\max_{\sim(\cdot)}$ denote the maximization over all variables

except $(\cdot)$. Let us now decompose $f(\mathbf{v})$ into summation of I functions, i.e., $\sum\limits_{i=1}^{I} f_i(\hat{v}_i)$ where $\hat{v}_i$ is a subset of the elements of the vector $\mathbf{v}$ and let $\mathbf{f} = [f_1(\cdot), \cdots, f_I(\cdot)]^{\mathsf{T}}$ is the vector of I functions. In addition, let $\mathfrak{f}_j$ represents subset of functions in $\mathbf{f}$ where the variable v_j appears. Hence the marginal can be rewritten as $\phi_j(v_j) = \max\limits_{\sim(v_j)} \sum\limits_{i=1}^{I} f_i(\hat{v}_i)$. According to the *max-sum* MP strategy the message passed by any variable node v_j to any generic function node $f_i(\cdot)$ is given by $\delta_{v_j \to f_i(\cdot)}(v_j) = \sum\limits_{i' \in \mathfrak{f}_j, i' \neq i} \delta_{f_{i'}(\cdot) \to v_j}(v_j)$. Similarly, the message from function node $f_i(\cdot)$ to variable node v_j is given as $\delta_{f_i(\cdot) \to v_j}(v_j) = \max\limits_{\sim(v_j)} \left(f_i(v_1, \cdots, v_J) + \sum\limits_{j' \in \hat{v}_i, j' \neq j} \delta_{v_{j'} \to f_i(\cdot)}(v_{j'}) \right)$. When the factor graph is cycle free (e.g., there is a unique path connecting any two nodes), all the variables nodes $j = \{1, \cdots, J\}$ can compute the marginals as $\phi_j(v_j) = \sum\limits_{i=1}^{I} \delta_{f_i(\cdot) \to v_j}(v_j)$. Utilizing the general distributive law (e.g., $\max \sum = \sum \max$) [31] the maximization therefore can be computed as $Z = \sum\limits_{j=1}^{J} \max\limits_{v_j} \phi_j(v_j)$.

1.5.2 Reformulation of the Resource Allocation Problem Utilizing MP Approach

In order to solve the resource allocation problem **P1.1** presented in Section 1.3.3 using MP, we reformulate it as a utility maximization problem. Let us define the reward functions $\mathfrak{W}_n(\mathbf{X})$ and $\mathfrak{R}_k(\mathbf{X})$ where the transmission alignment vector $\mathbf{X}$ is given by Equation (1.7). With the constraint in Equation (1.5a), we can define $\mathfrak{W}_n(\mathbf{X})$ as follows:

$$
\mathfrak{W}_n(\mathbf{X}) = \begin{cases} 0, & \text{if } \sum\limits_{k=1}^{K} \sum\limits_{l=1}^{L} x_k^{(n,l)} g_{k,m_k^*}^{(n)} p_k^{(n)} < I_{\max}^{(n)} \\ -\infty, & \text{otherwise.} \end{cases} \tag{1.9}
$$

Similarly to deal with the constraint in Equation (1.5b) we define $\mathfrak{R}_k(\mathbf{X})$ as

$$
\mathfrak{R}_k(\mathbf{X}) = \begin{cases} \sum\limits_{n=1}^{N} \sum\limits_{l=1}^{L} x_k^{(n,l)} B_{\text{RB}} \log_2 \left(1 + \gamma_{u_k}^{(n)} \right) & \text{if } \sum\limits_{n=1}^{N} \sum\limits_{l=1}^{L} x_k^{(n,l)} \leq 1 \\ -\infty & \text{otherwise.} \end{cases} \tag{1.10}
$$

The interpretations of the reward functions in Equations (1.9) and (1.10) are straightforward. Satisfying the interference constraint in Equation (1.5a) does not cost any penalty (e.g., zero reward) in the function $\mathfrak{W}_n(\mathbf{X})$, and in the function $\mathfrak{R}_k(\mathbf{X})$ fulfillment of the RB requirement constraint in Equation (1.5b) gives the desired data rate. However, both in the functions $\mathfrak{W}_n(\mathbf{X})$ and $\mathfrak{R}_k(\mathbf{X})$, the unfulfilled constraints, respectively, given by in Equations (1.5a) and (1.5b), result in infinite cost.

From the Equations (1.9) and (1.10), the resource allocation problem **P1.1** can be rewritten as $\max\limits_{\mathbf{X}} \left(\sum\limits_{n=1}^{N} \mathfrak{W}_n(\mathbf{X}) + \sum\limits_{k=1}^{K} \mathfrak{R}_k(\mathbf{X}) \right)$ and the optimal transmission allocation

vector is therefore given by

$$\mathbf{X}^* = \underset{\mathbf{X}}{\operatorname{argmax}} \left(\sum_{n=1}^{N} \mathfrak{W}_n(\mathbf{X}) + \sum_{k=1}^{K} \mathfrak{R}_k(\mathbf{X}) \right). \tag{1.11}$$

Since our goal is to obtain a distributed solution for the above resource allocation problem, we focus on a single transmission alignment allocation variable, e.g., $x_k^{(n,l)}$. From Equation (1.11) we obtain $x_k^{(n,l)*} = \underset{x_k^{(n,l)}}{\operatorname{argmax}} \phi_k^{(n,l)}\big(x_k^{(n,l)}\big)$ where the marginal $\phi_k^{(n,l)}\big(x_k^{(n,l)}\big)$ is given by

$$\phi_k^{(n,l)}\big(x_k^{(n,l)}\big) = \max_{\sim\big(x_k^{(n,l)}\big)} \left(\sum_{n=1}^{N} \mathfrak{W}_n(\mathbf{X}) + \sum_{k=1}^{K} \mathfrak{R}_k(\mathbf{X}) \right). \tag{1.12}$$

As mentioned in the previous section, $\max\limits_{\sim\big(x_k^{(n,l)}\big)}$ denote the maximization over all variables in $\mathbf{X}$ except $x_k^{(n,l)}$. The marginalization in Equation (1.12) can be computed in a distributed way where each node conveys the solution of a local problem to one another by passing information messages according to the max-sum MP strategy. Note that according to our system model the underlay transmitters and the resources (e.g., transmission alignments) can form a bipartite graph, e.g., each transmission alignment $\{n, l\}$ can be assigned to any of the K transmitters as long as interference to the MUEs on RB n is below threshold. Without loss of generality, let us consider a generic transmission alignment, e.g., RB-power level pair $\{n, l\} \in \mathcal{N} \times \mathcal{L}$ and an underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$. Using the function in Equation (1.9) and utilizing the max-sum MP strategy presented in Section 1.5.1, it is possible to show that the message delivered by the resource $\{n, l\}$ to the transmitter k can be expressed as [32]

$$\delta_{\{n,l\}\to k}\big(x_k^{(n,l)}\big) = \max \sum_{k'\in\mathcal{K}^{\mathrm{T}},\ k'\neq k} \delta_{k'\to\{n,l\}}\big(x_{k'}^{(n,l)}\big)$$

$$\text{subject to:} \quad \sum_{k=1}^{K}\sum_{l=1}^{L} x_k^{(n,l)} g_{k,m_k^*}^{(n)} p_k^{(n)} < I_{\max}^{(n)}. \tag{1.13}$$

Note that the term $\delta_{k\to\{n,l\}}\big(x_k^{(n,l)}\big)$ in the above equation denotes the message from transmitter k to the resource $\{n, l\}$ which can be written as [32]

$$\delta_{k\to\{n,l\}}\big(x_k^{(n,l)}\big) = x_k^{(n,l)} R_{u_k}^{(n,l)} + \max \sum_{\substack{\{n',l'\}\in\mathcal{N}\times\mathcal{L} \\ n'\neq n,\ l'\neq l}} x_k^{(n',l')} R_{u_k}^{(n',l')} + \delta_{\{n',l'\}\to k}\big(x_k^{(n',l')}\big)$$

$$\text{subject to:} \quad \sum_{n=1}^{N}\sum_{l=1}^{L} x_k^{(n,l)} \leq 1 \tag{1.14}$$

where $R_{u_k}^{(n,l)} = B_{\mathrm{RB}} \log_2\left(1 + \Gamma_k^{(n,l)}\right)$ and $\Gamma_k^{(n,l)} \triangleq \gamma_{u_k}^{(n)}\big|_{p_k^{(n)}=l}$.

The interpretation of the Equations (1.13) and (1.14) are as follows: the messages $\delta_{\{n,l\}\to k}(1)$ and $\delta_{k\to\{n,l\}}(1)$ carry the information relative to the use of the resource $\{n, l\}$ by the transmitter k; while the messages $\delta_{\{n,l\}\to k}(0)$ and $\delta_{k\to\{n,l\}}(0)$ carry the information

relative to the lack of transmission over the resource $\{n, l\}$ by the transmitter k. In order to obtain both the messages $\delta_{\{n,l\}\to k}\big(x_k^{(n,l)}\big)$ and $\delta_{k\to\{n,l\}}\big(x_k^{(n,l)}\big)$, it is required to solve the local optimization problem relative to the allocation variable $x_k^{(n,l)}$.

Based on the discussions of Section 1.5.1, the link-wise marginal in Equation (1.12) can be written as [32]

$$\phi_k^{(n,l)}\big(x_k^{(n,l)}\big) = \delta_{\{n,l\}\to k}\big(x_k^{(n,l)}\big) + \delta_{k\to\{n,l\}}\big(x_k^{(n,l)}\big) \tag{1.15}$$

and hence the transmission allocation variable is given by

$$x_k^{(n,l)^*} = \operatorname*{argmax}_{x_k^{(n,l)}} \phi_k^{(n,l)}\big(x_k^{(n,l)}\big). \tag{1.16}$$

At each iteration of the MP-based resource allocation algorithm, at most one message passes through the edge of any given direction (e.g., from transmitters to resources or from resources to transmitters); and each iteration the messages are updated by replacing the previous message sent on the same edge in the same direction [32]. When both the messages given by Equations (1.13) and (1.14) are available, the marginal can be computed using Equation (1.15) and the transmission allocation variable is obtained by Equation (1.16).

1.5.3 Effective Implementation of MP Scheme in a Practical Heterogeneous Network

It is worth noting that, sending messages from resources to transmitters (and vice versa) requires actual transmission on the radio channel. In a practical LTE-A-based 5G system, since the exchange of messages actually involves effective transmissions over the channel, the MP scheme described in the preceding section might be limited by the signaling overhead due to transfer of messages between the transmitters and resources. In the following, we observe that the amount of message signaling can be significantly reduced by some algebraic manipulations. Since the messages carry the information regarding whether any resource is used by any underlay transmitter, each transmitter k actually delivers a real valued vector with two element, i.e., $\boldsymbol{\delta}_{k\to\{n,l\}} = \big[\delta_{k\to\{n,l\}}(1),\ \delta_{k\to\{n,l\}}(0)\big]^{\mathsf{T}}$ and each resource $\{n, l\}$ delivers the vector $\boldsymbol{\delta}_{\{n,l\}\to k} = \big[\delta_{\{n,l\}\to k}(1),\ \delta_{\{n,l\}\to k}(0)\big]^{\mathsf{T}}$. Let us now rewrite the message $\delta_{k\to\{n,l\}}\big(x_k^{(n,l)}\big)$ using the utility function introduced in Equation (1.8) as follows:

$$\delta_{k\to\{n,l\}}\big(x_k^{(n,l)}\big) = x_k^{(n,l)}\mathfrak{U}_k^{(n,l)} + \max \sum_{\substack{\{n',l'\}\in\mathcal{N}\times\mathcal{L} \\ n'\neq n,\ l'\neq l}} x_k^{(n',l')}\mathfrak{U}_k^{(n',l')} + \delta_{\{n',l'\}\to k}\big(x_k^{(n',l')}\big).$$

$$\tag{1.17}$$

By subtracting the constant term $\sum_{\substack{\{n',l'\}\in\mathcal{N}\times\mathcal{L} \\ n'\neq n,\, l'\neq l}} \delta_{\{n',l'\}\to k}(0)$ from the both sides of Equation (1.17) we can obtain the following:

$$
\delta_{k\to\{n,l\}}\big(x_k^{(n,l)}\big) - \sum_{\substack{\{n',l'\}\in\mathcal{N}\times\mathcal{L} \\ n'\neq n,\, l'\neq l}} \delta_{\{n',l'\}\to k}(0) = x_k^{(n,l)}\mathfrak{U}_k^{(n,l)} \;+
$$

$$
\max \sum_{\substack{\{n',l'\}\in\mathcal{N}\times\mathcal{L} \\ n'\neq n,\, l'\neq l}} x_k^{(n',l')}\mathfrak{U}_k^{(n',l')} + \delta_{\{n',l'\}\to k}\big(x_k^{(n',l')}\big) - \delta_{\{n',l'\}\to k}(0). \tag{1.18}
$$

Let us now introduce the parameter $\psi_{\{n,l\}\to k} = \delta_{\{n,l\}\to k}(1) - \delta_{\{n,l\}\to k}(0)$ defined as the normalized message. For instance, consider the vector

$$
\Psi_k = \left[\mathfrak{U}_k^{(1,1)} + \psi_{\{1,1\}\to k}, \cdots, \mathfrak{U}_k^{(1,L)} + \psi_{\{1,L\}\to k}, \cdots, \mathfrak{U}_k^{(N,L)} + \psi_{\{N,L\}\to k}\right]^{\mathsf{T}}
$$

and let us denote by $\big\langle \upsilon_{\{n',l'\}\to k}\big\rangle_{\sim\{n,l\}}$ the maximal entry of the vector Ψ_k without considering the term $\mathfrak{U}_k^{(n,l)} + \psi_{\{n,l\}\to k}$. It can be noted that the terms within the summation in Equation (1.18) are either 0 (e.g., when $x_k^{(n,l)} = 0$) or $\mathfrak{U}_k^{(n',l')} + \psi_{\{n',l'\}\to k}$ (e.g., when $x_k^{(n,l)} = 1$). Since each transmitter requires only a single transmission alignment, when the variable $x_k^{(n,l)} = 0$, only one term in the summation of Equation (1.18) is non-zero. For the case $x_k^{(n,l)} = 1$, no term within the summation of Equation (1.18) is non-zero. Consequently, for $x_k^{(n,l)} = 0$, the maximum rate will be achieved if

$$
\delta_{k\to\{n,l\}}(0) - \sum_{\substack{\{n',l'\}\in\mathcal{N}\times\mathcal{L} \\ n'\neq n,\, l'\neq l}} \delta_{\{n',l'\}\to k}(0) = \big\langle \upsilon_{\{n',l'\}\to k}\big\rangle_{\sim\{n,l\}}. \tag{1.19}
$$

Similarly, when $x_k^{(n,l)} = 1$, the maximum is given by

$$
\delta_{k\to\{n,l\}}(1) - \sum_{\substack{\{n',l'\}\in\mathcal{N}\times\mathcal{L} \\ n'\neq n,\, l'\neq l}} \delta_{\{n',l'\}\to k}(0) = \mathfrak{U}_k^{(n,l)}. \tag{1.20}
$$

Since by definition $\psi_{k\to\{n,l\}} = \delta_{k\to\{n,l\}}(1) - \delta_{k\to\{n,l\}}(0)$, from the Equations (1.19) and (1.20), the normalized messages from the transmitter k to the resource $\{n,l\}$ can be derived as

$$
\psi_{k\to\{n,l\}} = \mathfrak{U}_k^{(n,l)} - \big\langle \upsilon_{\{n',l'\}\to k}\big\rangle_{\sim\{n,l\}}
$$

$$
= \mathfrak{U}_k^{(n,l)} - \big\langle \mathfrak{U}_k^{(n',l')} + \psi_{\{n',l'\}\to k}\big\rangle_{\sim\{n,l\}}. \tag{1.21}
$$

Likewise, from [32], it can be shown that the normalized message sent from the resource $\{n,l\}$ to the transmitter k becomes

$$
\psi_{\{n,l\}\to k} = \delta_{\{n,l\}\to k}(1) - \delta_{\{n,l\}\to k}(0) = -\max_{k'\in\mathcal{K}^{\mathsf{T}}, k'\neq k} \psi_{k'\to\{n,l\}}. \tag{1.22}
$$

For any arbitrary graph, the allocation variables may keep oscillating and might not converge to any fixed point, and the MP scheme may require some heuristic approach to terminate. However, in the context of loopy graphical models, by introducing a suitable weight, the messages given by Equations (1.21) and (1.22) perturb to a fixed point [32, 33]. Accordingly, Equations (1.21) and (1.22) can be rewritten as [32]

$$
\psi_{k \to \{n,l\}} = \mathfrak{U}_k^{(n,l)} - \omega \left\langle \mathfrak{U}_k^{(n',l')} + \psi_{\{n',l'\} \to k} \right\rangle_{\sim\{n,l\}} - (1 - \omega) \left(\mathfrak{U}_k^{(n,l)} + \psi_{\{n,l\} \to k} \right)
$$
(1.23)

$$
\psi_{\{n,l\} \to k} = -\omega \max_{k' \in \mathcal{K}^{\mathrm{T}}, k' \neq k} \psi_{k' \to \{n,l\}} - (1 - \omega) \psi_{k \to \{n,l\}}
$$
(1.24)

where $\omega \in (0, 1]$ denotes the weighting factor for each edge. Notice that when $\omega = 1$, the messages given by Equations (1.23) and (1.24) reduce to the original formulation, e.g., Equations (1.21) and (1.22), respectively. Given the normalized messages $\psi_{k \to \{n,l\}}$ and $\psi_{\{n,l\} \to k}$ for $\forall k, n, l$, the node marginals for the normalized messages can be calculated as $\tau_k^{(n,l)} = \psi_{k \to \{n,l\}} + \psi_{\{n,l\} \to k}$ and hence from Equation (1.16) the transmission alignment allocation can be obtained as

$$
x_k^{(n,l)*} = \begin{cases} 1 & \text{if } \tau_k^{(n,l)} > 0 \text{ and } I^{(n)} < I_{\max}^{(n)} \\ 0 & \text{otherwise.} \end{cases}
$$
(1.25)

1.5.4   Algorithm Development

In line with our discussions and from the expressions derived in Section 1.5.3, the MP-based resource allocation approach is outlined in **Algorithm 1.3**. The underlay transmitters and the resources (e.g., MBS) exchange the messages in an iterative manner. The MBS assigns the resource to the transmitters considering the node marginals, as well as the interference experienced on the RBs. The algorithm terminates when the sum data rate is reached to a steady value, i.e., the allocation vector **X** remains the same in successive iterations.

1.5.5   Convergence, Optimality, and Complexity of the Solution

The convergence, optimality, and complexity of the message passing approach is analyzed in the following subsections.

Convergence and Optimality

As presented in the following theorem, the message passing algorithm converges to fixed messages within fixed number of iterations.

Theorem 1.5.2 *The marginals and the allocation in **Algorithm 1.3** converge to a fixed point.*

Proof. The proof is constructed by utilizing the concept of *contraction mapping* [34, Chapter 3]. Let the vector $\boldsymbol{\psi}(t) = \left[\psi_{1 \to \{1,1\}}(t), \cdots, \psi_{k \to \{n,l\}}(t), \cdots \psi_{K \to \{N,L\}}(t)\right]^{\mathrm{T}}$ represent all the messages exchanged between the transmitters and the resources (e.g., MBS) at iteration t. Let us consider the messages are translated into the mapping $\boldsymbol{\psi}(t + 1) =$

Algorithm 1.3 Resource allocation using message passing

Initialization:

1: Estimate the CSI parameters from previous time slot.
2: Each underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$ selects a transmission alignment randomly and reports to MBS.
3: Initialize $t := 1$, $\psi_{k \to \{n,l\}}(0) := 0$, $\psi_{\{n,l\} \to k}(0) := 0$ for $\forall k, n, l$.

Update:

4: **while** $\mathbf{X}(t) \neq \mathbf{X}(t-1)$ **and** t less than some predefined threshold $T_{\max}$ **do**
5:   Each underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$ sends the message

$$\psi_{k \to \{n,l\}}(t) = \mathfrak{U}_k^{(n,l)}(t-1) - \omega \left\langle \mathfrak{U}_k^{(n',l')}(t-1) + \psi_{\{n',l'\} \to k}(t-1) \right\rangle_{\sim \{n,l\}} - (1-\omega) \left(\mathfrak{U}_k^{(n,l)}(t-1) + \psi_{\{n,l\} \to k}(t-1) \right)$$

  for $\forall \{n,l\} \in \mathcal{N} \times \mathcal{L}$ to the MBS.
6:   For all the resource $\forall \{n,l\} \in \mathcal{N} \times \mathcal{L}$, MBS sends messages

$$\psi_{\{n,l\} \to k}(t) = -\omega \max_{k' \in \mathcal{K}^{\mathrm{T}}, k' \neq k} \psi_{k' \to \{n,l\}}(t-1) - (1-\omega)\, \psi_{k \to \{n,l\}}(t-1)$$

  to each underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$.
7:   Each underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$ computes the marginals as $\tau_k^{(n,l)}(t) = \psi_{k \to \{n,l\}}(t) + \psi_{\{n,l\} \to k}(t)$ for $\forall \{n,l\} \in \mathcal{N} \times \mathcal{L}$ and reports to the MBS.
  /* MBS calculates the allocation vector according to Equation (1.25) */
8:   Set $x_k^{(n,l)} := 0$ for $\forall k, n, l$ /* Initialize the variable to obtain final allocation */
9:   **for each** $k \in \mathcal{K}^{\mathrm{T}}$ and $\{n,l\} \in \mathcal{N} \times \mathcal{L}$ **do**
10:    **if** $\tau_k^{(n,l)}(t) > 0$ **then**
11:      Set $x_k^{(n,l)} := 1$. /* Assign the resource to the transmitter */
12:      $\mathfrak{I}^{(n)} := \sum_{k'=1}^{K} \sum_{l'=1}^{L} x_{k'}^{(n,l')} g_{k',m_{k'}^*}^{(n)} p_{k'}^{(n)}$. /* Calculate interference in RB n */
13:      **if** $\mathfrak{I}^{(n)} \geq I_{\max}^{(n)}$ **then**
14:        **repeat**
15:          $\{\hat{k}, \hat{l}\} := \underset{k' \in \mathcal{K}^{\mathrm{T}}, l' \in \mathcal{L}}{\mathrm{argmax}}\, x_{k'}^{(n,l')} g_{k',m_{k'}^*}^{(n)} p_{k'}^{(n)}$ /* Most interfering transmitter $\hat{k}$ with $p_{\hat{k}}^{(n)} = \hat{l}$ */
16:          Set $x_{\hat{k}}^{(n,\hat{l})} := 0$.   /* Unassigned due to interference threshold violation */
17:          $\mathfrak{I}^{(n)} := \sum_{k'=1}^{K} \sum_{l'=1}^{L} x_{k'}^{(n,l')} g_{k',m_{k'}^*}^{(n)} p_{k'}^{(n)}$.   /* Update interference level */
18:        **until** $\mathfrak{I}^{(n)} < I_{\max}^{(n)}$
19:      **end if**
20:    **end if**
21:   **end for**
22:   MBS calculates the transmission alignment allocation vector $\mathbf{X}(t) = \left[x_k^{(n,l)} \right]_{\forall k,n,l}$ for the iteration t.
23:   Update $t := t + 1$.
24: **end while**

Transmission:

25: Use the allocated transmission alignments (e.g., the RB and power levels) for data transmission.

$\mathbb{T}(\boldsymbol{\psi}(t)) = \left[\mathbb{T}_1^{(1,1)}(\boldsymbol{\psi}(t)), \cdots, \mathbb{T}_K^{(N,L)}(\boldsymbol{\psi}(t)) \right]^{\mathrm{T}}$. From the Equations (1.23) and (1.24) we

can obtain $\psi_{k \to \{n,l\}}(t+1) = \mathbb{T}_k^{(n,l)}(\boldsymbol{\psi}(t))$ as follows:

$$\mathbb{T}_k^{(n,l)}(\boldsymbol{\psi}(t)) = \omega \left(\mathfrak{U}_k^{(n,l)}(t) - \mathfrak{U}_k^{(n',l')}(t) \right) +$$

$$\omega \left(\omega \max_{k' \in \mathcal{K}^{\mathrm{T}}, k' \neq k} \psi_{k' \to \{n',l'\}}(t) + (1-\omega)\psi_{k \to \{n',l'\}}(t) \right) +$$

$$(1-\omega) \left(\omega \max_{k' \in \mathcal{K}^{\mathrm{T}}, k' \neq k} \psi_{k' \to \{n,l\}}(t) + (1-\omega)\psi_{k \to \{n,l\}}(t) \right). \qquad (1.26)$$

For any vector $\mathbf{u}$ and $\mathbf{v}$, any generic mapping $\mathbb{T}$ is a contraction if $\| \mathbb{T}(\mathbf{u}) - \mathbb{T}(\mathbf{v}) \|_\infty \leq \varepsilon \| \mathbf{u} - \mathbf{v} \|_\infty$, where $\varepsilon < 1$ is the modulus of the mapping [34, Chapter 3]. From [33], it can be shown that the mapping $\mathbb{T} : \mathbb{R}^{KNF} \to \mathbb{R}^{KNF}$ is a contraction under the maximum norm, e.g., $\| \mathbb{T}(\boldsymbol{\psi}) \|_\infty = \max_{k \in \mathcal{K}^{\mathrm{T}}, n \in \mathcal{N}, l \in \mathcal{L}} |\mathbb{T}_k^{(n,l)}(\boldsymbol{\psi})|$. Since the contraction mappings have a unique fixed point convergence property for any initial vector, the proof concludes with that fact that message passing algorithm converges to a fixed marginal and hence to a fixed allocation vector $\mathbf{X}$.

The following theorem presents the fixed convergence point of the message passing algorithm is an optimal solution of the original resource allocation problem.

Theorem 1.5.3 *The allocation obtained by message passing algorithm converges to the optimal solution of resource allocation problem* **P1**.**1**.

Proof. The theorem is proved by contradiction. Let us consider that the solution $\widetilde{\mathbf{X}}$ obtained by message passing algorithm is not optimal and let $\mathbf{X}^*$ be the optimal solution obtained by solving **P1**.1. Let us further assume that there are $\chi \leq |\mathbf{X}|$ entries (e.g., allocations) that differ between $\widetilde{\mathbf{X}}$ and $\mathbf{X}^*$. In addition, let $\widetilde{\mathcal{N}} \times \widetilde{\mathcal{L}} \subseteq \mathcal{N} \times \mathcal{L}$ denote the subset of resources for which two allocations differ. For each $\{\tilde{n}, \tilde{l}\} \in \widetilde{\mathcal{N}} \times \widetilde{\mathcal{L}}$ there is a transmitter $\kappa_{\{\tilde{n}, \tilde{l}\}}$ such that $\tilde{x}_{\kappa_{\{\tilde{n},\tilde{l}\}}}^{(\tilde{n},\tilde{l})} = 1$ and $x_{\kappa_{\{\tilde{n},\tilde{l}\}}}^{*(\tilde{n},\tilde{l})} = 0$, and a transmitter $\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}} \neq \kappa_{\{\tilde{n},\tilde{l}\}}$ such that $\tilde{x}_{\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}}^{(\tilde{n},\tilde{l})} = 0$ and $x_{\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}}^{*(\tilde{n},\tilde{l})} = 1$. Hence, the assignment of resource $\{\tilde{n}, \tilde{l}\}$ to transmitter $\kappa_{\{\tilde{n}, \tilde{l}\}}$ implies that the marginal $\tau_{\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}}^{(\tilde{n},\tilde{l})} < 0$ and the following set of inequalities hold for each $\{\tilde{n}, \tilde{l}\} \in \widetilde{\mathcal{N}} \times \widetilde{\mathcal{L}}$:

$$\tau_{\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}}^{(\tilde{n},\tilde{l})} = \omega \left[\left(\mathfrak{U}_{\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}}^{(\tilde{n},\tilde{l})} + \psi_{\{\tilde{n},\tilde{l}\} \to \ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}} \right) - \left(\mathfrak{U}_{\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}}^{(n',l')} + \psi_{\{n',l'\} \to \ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}} \right) \right] < 0 \qquad (1.27)$$

where $\{n', l'\}$ is the resource as represented in Equation (1.21). According to our assumption, the resource $\{n', l'\}$ also belongs to $\widetilde{\mathcal{N}} \times \widetilde{\mathcal{L}}$. Hence, $\sum_{\{\tilde{n},\tilde{l}\} \in \widetilde{\mathcal{N}} \times \widetilde{\mathcal{L}}} \tau_{\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}}^{(\tilde{n},\tilde{l})} = \omega \left(\Delta \mathfrak{U} + \Delta \psi \right)$

where

$$\Delta \mathfrak{U} = \sum_{\{\tilde{n},\tilde{l}\} \in \widetilde{\mathcal{N}} \times \widetilde{\mathcal{L}}} \left(\mathfrak{U}_{\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}}^{(\tilde{n},\tilde{l})} - \mathfrak{U}_{\kappa_{\{\tilde{n},\tilde{l}\}}}^{(\tilde{n},\tilde{l})} \right)$$

$$= \sum_{k=1}^{K} \sum_{N=1}^{N} \sum_{l=1}^{L} x_k^{*(n,l)} \mathfrak{U}_k^{(n,l)} - \sum_{k=1}^{K} \sum_{N=1}^{N} \sum_{l=1}^{L} \tilde{x}_k^{(n,l)} \mathfrak{U}_k^{(n,l)} \qquad (1.28)$$

and $\Delta\psi = \sum\limits_{\{\tilde{n},\tilde{l}\}\in\widetilde{\mathcal{N}}\times\widetilde{\mathcal{L}}} \left(\psi_{\{\tilde{n},\tilde{l}\}\to\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}} - \psi_{\{\tilde{n},\tilde{l}\}\to\kappa_{\{\tilde{n},\tilde{l}\}}} \right)$. After some algebraic manipulations

(for details refer to [33]) we can obtain $\frac{2(1-\omega)}{\omega} \sum\limits_{\{\tilde{n},\tilde{l}\}\in\widetilde{\mathcal{N}}\times\widetilde{\mathcal{L}}} \tau^{(\tilde{n},\tilde{l})}_{\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}} \leq \Delta\mathfrak{U}$. Since $0 < \omega < 1$ and

both the variables $\sum\limits_{\{\tilde{n},\tilde{l}\}\in\widetilde{\mathcal{N}}\times\widetilde{\mathcal{L}}} \tau^{(\tilde{n},\tilde{l})}_{\ddot{\kappa}_{\{\tilde{n},\tilde{l}\}}}$ and $\Delta\mathfrak{U}$ are positive, our assumption that $\widetilde{\mathbf{X}}$ is not

optimal is contradicted and the proof follows.

Complexity

If the message passing algorithm requires $T < T_{\max}$ iterations to converge, it is straightforward to verify that the time complexity at each MBS is of $\mathcal{O}\left(TKNL\right)$. Similarly, considering a standard sorting algorithm that outputs the term $\left\langle \mathfrak{U}_k^{(n',l')} + \psi_{\{n',l'\}\to k} \right\rangle_{\sim\{n,l\}}$ in order to generate the message $\psi_{k\to\{n,l\}}$ with worst-case complexity of $\mathcal{O}\left(NL\log\left(NL\right)\right)$, the overall time complexity at each underlay transmitter is of $\mathcal{O}\left(T(NL)^2\log\left(NL\right)\right)$.

1.6   Auction-Based Resource Allocation

Our final solution approach for the resource allocation is the distributed auction algorithm. The allocation using auction is based on the *bidding* procedure, where the agents (i.e., underlay transmitters) bid for the resources (e.g., RB and power level). The transmitters select the bid for the resources based on the *costs* (e.g., the interference caused to the MUEs) of using the resource. The desired assignment relies on the appropriate selection of the bids. The unassigned transmitters raise the cost of using resource and bid for the resources simultaneously. Once the bids from all the transmitters are available, the resources are assigned to the highest bidder. An overview of auction approach is presented in the following.

1.6.1   Overview of the Auction Approach

In a generic auction-based assignment model, every resource j associated with a cost c_j and each agent i can get the benefit B_{ij} from the resource j. Given the benefit B_{ij}, every agent i who wishes to be assigned with the resource j, needs to pay the cost c_j. The net value (e.g., utility) that an agent i can get from the resource j is given by $B_{ij} - c_j$. The auction procedure involves the assignment of agent i with the resource j' which provides the maximal net value, i.e.,

$$B_{ij'} - c_{j'} = \max_{j} \left\{ B_{ij} - c_j \right\}. \tag{1.29}$$

If the condition given in Equation (1.29) is satisfied for all the agents i, the assignment and the set of costs are referred to as *equilibrium* [35]. However, in many practical problems, obtaining an equilibrium assignment is not straightforward due to the possibility of cycles. In particular, there may be cases where the agents contend for a small number of equally desirable resources without increasing the cost, which creates cycle (e.g., infinite loop) in the auction process. To avoid this difficulty, the notion of *almost equilibrium* is introduced in the

literature. The assignment and the set of costs are said to be almost equilibrium when the net value for assigning each agent i with the resource j' is within a constant $\epsilon > 0$ of being maximal. Hence, in order to be an almost equilibrium assignment, the following condition needs to be satisfied for all the agents [35]:

$$B_{ij'} - c_{j'} \geq \max_{j} \{B_{ij} - c_j\} - \epsilon. \tag{1.30}$$

The condition in Equation (1.30) is known as *ϵ-complementary slackness*. When $\epsilon = 0$, Equation (1.30) reduces to ordinary complementary slackness given by Equation (1.29).

For instance, let the variable $\Theta_i = j$ denote that agent i is assigned with the resource j. In addition, let c_{ij} denote the cost that agent i incurs in order to be assigned with resource j and $\mathfrak{b}_{ij}$ is the bidding information (i.e., highest bidder) available to the agent i about resource j. The auction procedure evolves in an iterative manner. Given the the assignment Θ_i, the set of costs $[c_{ij}]_{\forall ij}$, and the set of largest bidders $[\mathfrak{b}_{ij}]_{\forall ij}$ of previous iteration, the agents locally update the costs and the highest bidders for current iteration. In particular, the costs $c_{ij}(t)$ and bidding information $\mathfrak{b}_{ij}(t)$ available to the agent i about resource j for iteration t are updated from the previous iteration as follows [36]:

$$c_{ij}(t) = \max_{i',i' \neq i} \{c_{ij}(t-1), c_{i'j}(t-1)\} \tag{1.31}$$

$$\mathfrak{b}_{ij}(t) = \max_{i^* \in \underset{i',i' \neq i}{\operatorname{argmax}}\{c_{ij}(t-1), c_{i'j}(t-1)\}} \{\mathfrak{b}_{i^*j}(t-1)\}. \tag{1.32}$$

The above update equations ensure that the agents will have the updated maximum cost of the resource j (i.e., $c_j \triangleq \max_i\{c_{ij}\}$) and the corresponding highest bidder for that resource. Once the update cost and bidding information are available, agent i checks whether the cost of the resource currently assigned to agent i, e.g., $c_{i\Theta_i(t-1)}$ has been increased by any other agents. If so, the current assignment obtained from previous iteration may not be at (almost) equilibrium and the agent needs to select a new assignment, e.g., $\Theta_i(t) = \operatorname{argmax}_j \{B_{ij}(t) - c_{ij}(t)\}$. In order to update the cost for new assignment (e.g., $\Theta_i(t)$) for any iteration t, the agent will use the following cost update rule [36]:

$$c_{ij}(t) = c_{ij}(t-1) + \Delta_i(t-1) \tag{1.33}$$

where Δ_i is given by

$$\Delta_i(t-1) = \max_j \{B_{ij}(t-1) - c_{ij}(t-1)\} - \max_{j' \neq \Theta_i(t)} \{B_{ij'}(t-1) - c_{ij'}(t-1)\} + \epsilon. \tag{1.34}$$

The variable $\max_j \{B_{ij}(t-1) - c_{ij}(t-1)\}$ and $\max_{j' \neq \Theta_i(t)} \{B_{ij'}(t-1) - c_{ij'}(t-1)\}$ denote the maximum and second maximum net utility, respectively. Note that Δ_i is always greater than zero as $\epsilon > 0$ and by definition $\max_j \{B_{ij}(t-1) - c_{ij}(t-1)\} > \max_{j' \neq \Theta_i(t)} \{B_{ij'}(t-1) - c_{ij'}(t-1)\}$. Since $c_{i\Theta_i(t)}(t)$ is the highest cost for iteration t, agent i can also update the bidding information, e.g., $\mathfrak{b}_{i\Theta_i(t)}(t) = i$. Accordingly, the cost update rule using Δ_i as given in Equation (1.33) ensures that the assignment and the set of costs are almost at equilibrium [36].

1.6.2 Auction for Radio Resource Allocation

Based on the discussion provided in the preceding section, in the following, we present the auction-based resource allocation scheme. We present the cost model and use the concept of auction to develop the resource allocation algorithm in our considered heterogeneous network setup.

Cost Function

Let us consider the utility function given by Equation (1.8). Recall that the term $w_2\left(I^{(n)} - I_{\max}^{(n)}\right)$ in Equation (1.8) represents the cost (e.g., interference caused by underlay transmitters to the MUE) of using the RB n. In particular, when the transmitter k is transmitting with power level l, the cost of using RB n can be represented by

$$c_k^{(n,l)} = w_2\left(I^{(n)} - I_{\max}^{(n)}\right) = w_2\left(\sum_{k'=1}^{K}\sum_{l'=1}^{L} x_{k'}^{(n,l')} g_{k',m_{k'}^*}^{(n)} p_{k'}^{(n)} - I_{\max}^{(n)}\right)$$

$$= w_2\left(g_{k,m_k^*}^{(n)} l + \sum_{k'\in\mathcal{K}^{\mathrm{T}}, k'\neq k}\sum_{l'=1}^{L} x_{k'}^{(n,l')} g_{k',m_{k'}^*}^{(n)} p_{k'}^{(n)} - I_{\max}^{(n)}\right). \qquad (1.35)$$

Let the parameter $C_k^{(n,l)} = \max\{0, c_k^{(n,l)}\}$ and accordingly the cost $C_k^{(n,l)} = 0$ only if $I^{(n)} \leq I_{\max}^{(n)}$. Notice that using the cost term we can represent Equation (1.8) as

$$\mathfrak{U}_k^{(n,l)} = w_1\mathcal{R}\left(\Gamma_{u_k}^{(n,l)}\right) - w_2\left(I^{(n)} - I_{\max}^{(n)}\right) = B_k^{(n,l)} - c_k^{(n,l)} = B_k^{(n,l)} - C_k^{(n,l)}$$

where $B_k^{(n,l)} = w_1\mathcal{R}\left(\Gamma_{u_k}^{(n,l)}\right)$, and $c_k^{(n,l)}$ is given by Equation (1.35). The variable $B_k^{(n,l)}$ is proportional to the data rate achieved by transmitter k using resource $\{n,l\}$. Analogous to the discussion of previous section, $\mathfrak{U}_k^{(n,l)}$ represents the net benefit that transmitter k obtains from the resource $\{n,l\}$.

Let $\mathfrak{b}_k^{(n,l)}$ denote the local bidding information available to transmitter k for the resource $\{n,l\}$. For notational convenience, let us assume that $\Theta : [k]_{k=1,\cdots,K} \rightarrow [\{n,l\}]_{\substack{n=1,\cdots,N \\ l=1,\cdots,L}}$ denotes the mapping between the transmitters and the resources, i.e., $\Theta_k = \{n,l\}$ represents the assignment of resource $\{n,l\}$ to transmitter k. Hence we represent by $C_k^{\Theta_k}$ the cost of using the resource $\{n,l\}$ obtained by the assignment $\Theta_k = \{n,l\}$. Similarly, given $\Theta_k = \{n,l\}$ the variable $\mathfrak{b}_k^{\Theta_k} \equiv \mathfrak{b}_k^{(n,l)}$ denotes the local bidding information about the resource $\{n,l\}$ available to the transmitter k. Note that $\Theta_k = \{n,l\}$ also implies $x_k^{(n,l)} = 1$. In other words, $\Theta_k = \{n,l\}$ denote the non-zero entry of the vector $\mathbf{x}_k = \left[x_k^{(n,l)}\right]_{\forall n,l}$. Since each underlay transmitter k selects only one resource $\{n,l\}$, only a single entry in the vector $\mathbf{x}_k$ is non-zero.

Update of Cost and Bidder Information

In order to obtain the updated cost and bidding information, we utilize similar concept given by Equations (1.31)-(1.34). At the beginning of the auction procedure, each underlay

transmitter updates the cost as $C_k^{(n,l)}(t) = \max\limits_{k' \in \mathcal{K}^{\mathrm{T}}, k' \neq k} \left\{ C_k^{(n,l)}(t-1), C_{k'}^{(n,l)}(t-1) \right\}$. In addition, as described by Equation (1.32), the information of maximum bidder is obtained by $\mathfrak{b}_k^{(n,l)}(t) = \mathfrak{b}_{k^*}^{(n,l)}(t-1)$ where $k^* = \operatorname*{argmax}\limits_{k' \in \mathcal{K}^{\mathrm{T}}, k' \neq k} \left\{ C_k^{(n,l)}(t-1), C_{k'}^{(n,l)}(t-1) \right\}$. When the transmitter k needs to select a new assignment, i.e., $\Theta_k = \{\hat{n}, \hat{l}\}$, the transmitter increases the cost of using the resource, e.g., $C_k^{(\hat{n},\hat{l})}(t) = C_k^{(\hat{n},\hat{l})}(t-1) + \Delta_k(t-1)$, and $\Delta_k(t-1)$ is given by

$$\Delta_k(t-1) = \max_{\{n',l'\} \in \mathcal{N} \times \mathcal{L}} \mathfrak{U}_k^{(n',l')}(t-1) - \max_{\substack{\{n',l'\} \in \mathcal{N} \times \mathcal{L} \\ n' \neq \hat{n}, l' \neq \hat{l}}} \mathfrak{U}_k^{(n',l')}(t-1) + \epsilon \qquad (1.36)$$

where $\epsilon > 0$ indicates the minimum bid requirement parameter. Similar to Equation (1.34), the term $\max\limits_{\{n',l'\} \in \mathcal{N} \times \mathcal{L}} \mathfrak{U}_k^{(n',l')}(t-1) - \max\limits_{\substack{\{n',l'\} \in \mathcal{N} \times \mathcal{L} \\ n' \neq \hat{n}, l' \neq \hat{l}}} \mathfrak{U}_k^{(n',l')}(t-1)$ denotes the difference between the maximum and the second to the maximum utility value. In the case when the transmitter k does not prefer to be assigned with a new resource, the allocation from the previous iteration will remain unchanged, i.e., $\Theta_k(t) = \Theta_k(t-1)$, and consequently, $\mathbf{x}_k(t) = \mathbf{x}_k(t-1)$.

1.6.3   Algorithm Development

Algorithm 1.5 outlines the auction-based resource allocation approach. Each transmitter locally executes **Algorithm 1.4** and obtains a temporary allocation. When the execution of **Algorithm 1.4** is finished, each underlay transmitter k reports to the MBS the local information, e.g., choices for the resources, $\mathbf{x}_k = \left[x_k^{(n,l)} \right]_{\forall n,l}$. Once the information (e.g., output parameters from **Algorithm 1.4**) from all the transmitters are available to the MBS, the necessary parameters (e.g., input arguments required by **Algorithm 1.4**) are calculated and broadcast by the MBS. **Algorithm 1.4** repeated iteratively until the allocation variable $\mathbf{X} = [\mathbf{x}_k]_{\forall k} = \left[x_1^{(1,1)}, \cdots, x_1^{(1,L)}, \cdots, x_1^{(N,L)}, \cdots, x_K^{(N,L)} \right]^{\mathsf{T}}$ for two successive iterations becomes similar.

1.6.4   Convergence, Complexity, and Optimality of the Auction Approach

In the following subsections we analyze the convergence, complexity, and optimality of the solution obtained by auction algorithm.

Convergence and Complexity

For any arbitrary fixed $\epsilon > 0$, the auction approach is guaranteed to converge to a fixed assignment. The following theorem shows that the auction process terminates within a fixed number of iterations.

Theorem 1.6.1 *The auction process terminates in a finite number of iterations.*

Algorithm 1.4 Auction method for any underlay transmitter k

Input: Parameters from previous iteration: an assignment $\mathbf{X}(t-1) = [\mathbf{x}_1(t-1), \cdots \mathbf{x}_K(t-1)]^\mathsf{T}$, aggregated interference $I^{(n)}(t-1)$ for $\forall n$, cost of using resources $\mathbf{C}(t-1) = \left[C_k^{(n,l)}(t-1)\right]_{\forall k,n,l}$ and the highest bidders of the resources $\mathfrak{B}(t-1) = [\mathfrak{B}_k(t-1)]_{\forall k}$ where $\mathfrak{B}_k(\cdot) = \left[\mathfrak{b}_k^{(n,l)}(\cdot)\right]_{\forall n,l}$.

Output: The allocation variable $\mathbf{x}_k(t) = \left[x_k^{(n,l)}\right]_{\forall n,l}$, updated costs $\mathbf{C}_k(t) = \left[C_k^{(n,l)}(t)\right]_{\forall n,l}$, and bidding information $\mathfrak{B}_k(t) = \left[\mathfrak{b}_k^{(n,l)}(t)\right]_{\forall n,l}$ at current iteration t for the transmitter k.

1: Initialize $\mathbf{x}_k(t) := \mathbf{0}$.
2: For all the resources $\{n,l\} \in \mathcal{N} \times \mathcal{L}$,

- Obtain the transmitter $k^* := \underset{k' \in \mathcal{K}^\mathsf{T}, k' \neq k}{\operatorname{argmax}} \left\{ C_k^{(n,l)}(t-1), C_{k'}^{(n,l)}(t-1) \right\}$ and update the highest bidder as $\mathfrak{b}_k^{(n,l)}(t) := \mathfrak{b}_{k^*}^{(n,l)}(t-1)$.

- Update the cost as $C_k^{(n,l)}(t) := \underset{k' \in \mathcal{K}^\mathsf{T}, k' \neq k}{\max} \left\{ C_k^{(n,l)}(t-1), C_{k'}^{(n,l)}(t-1) \right\}$.

/ Let $\Theta_k(t-1)$ denote the assignment of transmitter k at previous iteration $t-1$, i.e., $\Theta_k(t-1)$ represents the non-zero entry in the vector $\mathbf{x}_k(t-1)$. Since each transmitter uses only one transmission alignment, only a single entry in the vector $\mathbf{x}_k(t-1)$ is non-zero. When cost is greater than previous iteration and the transmitter k is not the highest bidder, update the assignment */*

3: **if** $C_k^{\Theta_k(t-1)}(t) \geq C_k^{\Theta_k(t-1)}(t-1)$ **and** $\mathfrak{b}_k^{\Theta_k(t-1)}(t) \neq k$ **then**
4:     $\{\hat{n}, \hat{l}\} := \underset{\{n',l'\} \in \mathcal{N} \times \mathcal{L}}{\operatorname{argmax}} \mathfrak{U}_k^{(n',l')}(t)$. */* Obtain the best resource for transmitter k */*
5:     $\mathfrak{I}^{(\hat{n})} := g_{k,m_k^*}^{(\hat{n})} \hat{l} + I^{(\hat{n})}$. */* Calculate additional interference caused by transmitter k for using RB $\hat{n}$ */*
6:     **if** $\mathfrak{I}^{(\hat{n})} < I_{\max}^{(\hat{n})}$ **then**
7:         Set $x_k^{(\hat{n},\hat{l})} := 1$.   */* e.g., $\Theta_k(t) = \{\hat{n}, \hat{l}\}$ */*
8:         Update the highest bidder for the resource $\{\hat{n}, \hat{l}\}$ as $\mathfrak{b}_k^{(\hat{n},\hat{l})}(t) := k$.
9:         Increase the cost for the resource $\{\hat{n}, \hat{l}\}$ as $C_k^{(\hat{n},\hat{l})}(t) = C_k^{(\hat{n},\hat{l})}(t-1) + \Delta_k(t-1)$ where $\Delta_k(t-1)$ is given by Equation (1.36).
10:     **else**
11:         Keep the assignment unchanged from previous iteration, i.e., $\mathbf{x}_k(t) := \mathbf{x}_k(t-1)$.
12:     **end if**
13: **else**
14:     Keep the assignment unchanged from previous iteration, i.e., $\mathbf{x}_k(t) := \mathbf{x}_k(t-1)$.
15: **end if**

Proof. According to our system model, each underlay transmitter selects only one transmission alignment. Hence, once each resource receives at least one bid (which implies that each transmitter is assigned to a resource), the auction process must terminate. Now if any resource $\{n,l\}$ receives a bid in $\hat{t}$ iterations, the cost must be greater than the initial price by $\hat{t}\epsilon$. As a result, the resource $\{n,l\}$ becomes costly to be assigned when compared to any resource $\{n',l'\}$ that has not received any bid yet. The argument follows that there are two possibilities, e.g., *i)* the auction process terminates in a finite iterations with each transmitter assigned to a resource, regardless of every resource receives a bid; or *ii)* the auction process

---

Algorithm 1.5 Auction-based resource allocation

Initialization:

 1: Estimate the CSI parameters from the previous time slot.
 2: Each underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$ randomly selects a transmission alignment and reports to the MBS.
 3: MBS broadcasts the assignment of all transmitters, aggregated interference of each RB, the costs and the highest bidders using pilot signals.
 4: Initialize number of iterations $t := 1$.

Update:

 5: **while** $\mathbf{X}(t) \neq \mathbf{X}(t-1)$ **and** t is less than some predefined threshold $T_{\max}$ **do**
 6:     Each underlay transmitter $k \in \mathcal{K}^{\mathrm{T}}$ locally runs the **Algorithm 1.4** and reports the assignment $\mathbf{x}_k(t)$, the cost $\mathbf{C}_k(t)$ and the bidding information $\mathfrak{B}_k(t)$ to the MBS.
 7:     MBS calculates the aggregated interference $I^{(n)}(t)$ for $\forall n$, the allocation variable $\mathbf{X}(t)$, information about highest bidders $\mathfrak{B}(t)$, the cost $\mathbf{C}(t)$, and broadcast to the underlay transmitters.
 8:     Update $t := t+1$.
 9: **end while**

Transmission:

10: Use the resources (e.g., the RB and power levels) allocated in the final stage of update phase for data transmission.

---

continues for a finite number of iterations and each resource will receive at least one bid, therefore, the algorithm terminates.

At termination, the solution (e.g., allocation) obtained is almost at equilibrium, e.g., the condition in Equation (1.30) is satisfied for all the underlay transmitters. Since the algorithm terminates after a finite number of iterations, we can show that the algorithm converges to a fixed allocation and the complexity at each transmitter is linear to the number of resources.

Theorem 1.6.2 *The auction algorithm converges to a fixed allocation with the number of iterations of* $\mathcal{O}\left(TKNL \left\lceil \dfrac{\max\limits_{k,n,l} B_k^{(n,l)} - \min\limits_{k,n,l} B_k^{(n,l)}}{\epsilon} \right\rceil \right)$.

Proof. The proof follows from the similar argument presented in **Theorem 1.6.1**. In the worst case, the total number of iterations in which a resource can receive a bid is no more than $\Upsilon = \left\lceil \dfrac{\max\limits_{k,n,l} B_k^{(n,l)} - \min\limits_{k,n,l} B_k^{(n,l)}}{\epsilon} \right\rceil$ [36]. Since each bid requires $\mathcal{O}(NL)$ iterations, and each iteration involves a bid by a single transmitter, the total number of iterations in **Algorithm 1.5** is of $\mathcal{O}(KNL\Upsilon)$. For the convergence, the allocation variable $\mathbf{X}$ needs to be unchanged for at least $T \geq 2$ consecutive iterations. Hence, the overall running time of the algorithm is $\mathcal{O}(TKNL\Upsilon)$.

Note that for any transmitter node $k \in \mathcal{K}^{\mathrm{T}}$, the complexity of the auction process given by **Algorithm 1.4** is linear with number of resources for each of the iterations.

Optimality

In the following we show that the data rate obtained by the auction algorithm is within $K\epsilon$ of the maximum data rate obtained by solving the original optimization problem **P1.1**.

Theorem 1.6.3 *The data rate obtained by the distributed auction algorithm is within $K\epsilon$ of the optimal solution.*

Proof. We construct the proof by using an approach similar to that presented in [36]. The data rate obtained by any assignment will satisfy the following condition:

$$\sum_{k=1}^{K} R_{u_k} \leq \sum_{\{n,l\}\in\mathcal{N}\times\mathcal{L}} \widehat{C}^{(n,l)} + \sum_{k=1}^{K} \max_{\{n,l\}\in\mathcal{N}\times\mathcal{L}} \left\{ B_k^{(n,l)} - \widehat{C}^{(n,l)} \right\} \qquad (1.37)$$

where $\widehat{C}^{(n,l)} = \max_{k'\in\mathcal{K}^{\mathrm{T}}} C_{k'}^{(n,l)}$, $B_k^{(n,l)} = w_1 \mathcal{R}\left(\Gamma_{u_k}^{(n,l)}\right)$ and R_{u_k} is given by Equation (1.4). The inequality given by Equation (1.37) is satisfied since the first term in the right side of the inequality, e.g., $\sum_{\{n,l\}\in\mathcal{N}\times\mathcal{L}} \widehat{C}_{(n,l)}$ is equal to $\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{l=1}^{L} x_k^{(n,l)} C_k^{(n,l)}$ and the second term is not less than $\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{l=1}^{L} x_k^{(n,l)} \left(B_k^{(n,l)} - \widehat{C}^{(n,l)} \right)$. Let the variable $A^* \triangleq \max_{\mathbf{X}} \sum_{k=1}^{K} R_{u_k} = \sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{l=1}^{L} x_k^{(n,l)} B_{\mathrm{RB}} \log_2\left(1 + \gamma_{u_k}^{(n)}\right)$ denote the optimal achievable data rate. In addition, let the variable D^* be defined as

$$D^* \triangleq \min_{\substack{\widehat{C}^{(n,l)} \\ \{n,l\}\in\mathcal{N}\times\mathcal{L}}} \left\{ \sum_{\{n,l\}\in\mathcal{N}\times\mathcal{L}} \widehat{C}^{(n,l)} + \sum_{k=1}^{K} \max_{\{n,l\}\in\mathcal{N}\times\mathcal{L}} \left\{ B_k^{(n,l)} - \widehat{C}^{(n,l)} \right\} \right\}. \qquad (1.38)$$

Hence from Equation (1.37), we can write $A^* \leq D^*$. Since the final assignment and the set of costs are almost at equilibrium, for any underlay transmitter k, the condition $\sum_{n=1}^{N}\sum_{l=1}^{L} x_k^{(n,l)} \left(B_k^{(n,l)} - \widehat{C}^{(n,l)} \right) \geq \max_{\{n,l\}\in\mathcal{N}\times\mathcal{L}} \left\{ B_k^{(n,l)} - \widehat{C}^{(n,l)} \right\} - \epsilon$ will hold. Consequently, we can obtain the following inequality:

$$D^* \leq \sum_{k=1}^{K} \left(\sum_{n=1}^{N}\sum_{l=1}^{L} x_k^{(n,l)} \widehat{C}^{(n,l)} + \max_{\{n,l\}\in\mathcal{N}\times\mathcal{L}} \left\{ B_k^{(n,l)} - \widehat{C}^{(n,l)} \right\} \right)$$

$$\leq \sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{l=1}^{L} x_k^{(n,l)} B_k^{(n,l)} + K\epsilon \leq \sum_{k=1}^{K} R_{u_k} + K\epsilon \leq A^* + K\epsilon. \qquad (1.39)$$

Since $A^* \leq D^*$, the data rate achieved by the auction algorithm is within $K\epsilon$ of the optimal data rate A^* and the proof follows.

1.7   Qualitative Comparison Among the Resource Allocation Schemes

In this section, we compare the different resource allocation schemes discussed above based on several criteria (e.g., flow of algorithm execution, information requirement and algorithm

Table 1.2 Comparison among different resource allocation approaches

Criterion	Schemes			
	COS	Stable matching	Message passing	Auction method
Type of the solution	Centralized	Distributed	Distributed	Distributed
Algorithm execution	MBS solves the resource optimization problem (e.g., **P1.1**)	MBS and underlay transmitters locally update the preference profiles, MBS runs the matching subroutine	MBS and underlay transmitters alliteratively exchange the messages, MBS computes the marginals and selects allocation	Each underlay transmitters locally runs the auction subroutine, MBS collects the parameters from all the transmitters and broadcast required parameters needed for the auction subroutine
Optimality	Optimal	Weak Pareto optimal	Optimal subject to the weight ω	Within $K\epsilon$ to the optimal
Complexity	$\mathcal{O}\left((NL)^K\right)$ at the MBS	$\mathcal{O}\left(TNL\log(NL)\right)$ at the transmitters, $\mathcal{O}(TKNL)$ at the MBS	$\mathcal{O}\left(T(NL)^2\log(NL)\right)$ at the transmitters, $\mathcal{O}(TKNL)$ at the MBS	For each iteration linear with N, L at the transmitters, overall running time $\mathcal{O}(TKNL\Upsilon)$
Convergence behavior	N/A	Converges to a stable matching and hence to a fixed allocation	Converges to a fixed marginal and to a fixed allocation	Converges to a fixed allocation within $K\epsilon$ of the optimal
Information required by the MBS	Channel gains (e.g., CSI parameters) between all the links of the network	The preference profiles and the channel gains $G_k^{(n)} = \left[g_{k,m_k^*}^{(n)}\right]_{\forall k,n}$	The messages $\left[\psi_{k \to \{n,l\}}\right]_{\forall k,n,l}$ and the channel gains $G_k^{(n)} = \left[g_{k,m_k^*}^{(n)}\right]_{\forall k,n}$	The channel gains $G_k^{(n)} = \left[g_{k,m_k^*}^{(n)}\right]_{\forall k,n}$, local assignments $\mathbf{x}_k$, the cost $\mathbf{C}_k$, and the bidding information $\mathfrak{B}_k$ for $\forall k$
Algorithm overhead	High (exponential) computational complexity, requirement of all CSI parameters of the network	Build the preference profiles, exchange information to update preference profiles, execution of matching subroutine	Calculation and exchange of messages, computation of the marginals	Computation and exchange of the parameters, e.g., $I^{(n)}$ for $\forall n$, the allocation vector $\mathbf{X}$, information about highest bidders $\mathfrak{B}$, the cost vector $\mathbf{C}$

overhead, complexity and optimality of the solution, convergence behavior etc.). We term the centralize solution (which can be obtained by solving the optimization problem **P1.1**) as COS (centralized optimal scheme) and compare it with the distributed solutions. A comparison among the resource allocation schemes is presented in Table 1.2.

1.8   Chapter Summary and Conclusion

We have presented three comprehensive distributed solution approaches for the future 5G cellular mobile communication systems. Considering a heterogeneous multi-tier 5G network, we have developed distributed radio resource allocation algorithms using three different mathematical models (e.g., stable matching, message passing, and auction method). The properties (e.g., convergence, complexity, optimality) of these distributed solutions are also briefly analyzed. To this end, a qualitative comparison of these schemes is illustrated.

The solution tools presented in this chapter can also be applicable to address the resource allocation problems in other enabling technologies for 5G systems. In particular, the mathematical tools presented in this chapter open up new opportunities to investigate other network models, such as resource allocation problems for wireless virtualization [37] and cloud-based radio access networks [38]. In such systems, these modeling tools need to be customized accordingly based on the objective and constraints required for the resource allocation problem.

In addition to the presented solutions, there are few game theoretical models which have not been covered in this chapter. However, these game models can also be considered as potential distributed solution tools. Different from traditional cooperative and non-cooperative games, the game models (such as mean field games [39], [40], evolutionary games [41] etc.) are scalable by nature, and hence applicable to model such large heterogeneous 5G networks. Utilizing those advanced game models for the resource allocation problems and analyzing the performance (e.g., data rate, spectrum and energy efficiency etc.) of 5G systems could be an interesting area of research.

References

[1] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, May 2014.

[2] P. Demestichas, A. Georgakopoulos, D. Karvounas, K. Tsagkaris, V. Stavroulaki, J. Lu, C. Xiong, and J. Yao, "5G on the horizon: Key challenges for the radio-access network," *IEEE Vehicular Technology Magazine*, vol. 8, no. 3, pp. 47–53, September 2013.

[3] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 106–112, April 2014.

[4] M. Tehrani, M. Uysal, and H. Yanikomeroglu, "Device-to-device communication in 5G cellular networks: Challenges, solutions, and future directions," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 86–92, May 2014.

[5] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 82–89, February 2014.

[6] N. Himayat, S. ping Yeh, A. Panah, S. Talwar, M. Gerasimenko, S. Andreev, and Y. Koucheryavy, "Multi-radio heterogeneous networks: Architectures and performance," in *International Conference on Computing, Networking and Communications (ICNC)*, February 2014, pp. 252–258.

[7] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, February 2014.

[8] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[9] S. Talwar, D. Choudhury, K. Dimou, E. Aryafar, B. Bangerter, and K. Stewart, "Enabling technologies and architectures for 5G wireless," in *IEEE MTT-S International Microwave Symposium (IMS)*, June 2014, pp. 1–4.

[10] V. Subramanian, R. Berry, and R. Agrawal, "Joint scheduling and resource allocation in CDMA systems," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2416–2432, May 2010.

[11] S.-J. Oh, D. Zhang, and K. Wasserman, "Optimal resource allocation in multiservice CDMA networks," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 811–821, July 2003.

[12] K. Pedersen, Y. Wang, S. Strzyz, and F. Frederiksen, "Enhanced inter-cell interference coordination in co-channel multi-layer LTE-advanced networks," *IEEE Wireless Communications*, vol. 20, no. 3, pp. 120–127, June 2013.

[13] "LTE Advanced: Heterogeneous networks," January 2011, White Paper. [Online]. Available: https://www.qualcomm.com/documents/lte-heterogeneous-networks

[14] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced: Deployment scenarios and operational challenges," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 148–155, February 2012.

[15] Y. Zhou, L. Liu, H. Du, L. Tian, X. Wang, and J. Shi, "An overview on intercell interference management in mobile cellular networks: From 2G to 5G," in *IEEE International Conference on Communication Systems (ICCS)*, Nov 2014, pp. 217–221.

[16] "5G: A technology vision," February 2014, White Paper. [Online]. Available: http://www.huawei.com/5gwhitepaper/

[17] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 118–127, June 2014.

[18] M. Hasan and E. Hossain, "Resource allocation for network-integrated device-to-device communications using smart relays," in *IEEE Global Communications Conference Workshops (GC Wkshps)*, 2013, pp. 597–602.

[19] M. Arslan, J. Yoon, K. Sundaresan, S. Krishnamurthy, and S. Banerjee, "A resource management system for interference mitigation in enterprise OFDMA femtocells," *IEEE/ACM Transactions on Networking*, vol. 21, no. 5, pp. 1447–1460, October 2013.

[20] H. Zhang, X. Chu, W. Ma, W. Zheng, and X. Wen, "Resource allocation with interference mitigation in OFDMA femtocells for co-channel deploymenta," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, 2012. [Online]. Available: http://dx.doi.org/10.1186/1687-1499-2012-289

[21] W.-H. Kuo and W. Liao, "Utility-based radio resource allocation for QoS traffic in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2714–2722, July 2008.

[22] E. Hossain, L. B. Le, and D. Niyato, *Radio resource management in multi-tier cellular wireless networks*. John Wiley & Sons, 2013.

[23] P. Semasinghe, E. Hossain, and K. Zhu, "An evolutionary game for distributed resource allocation in self-organizing small cells," *IEEE Transactions on Mobile Computing*, vol. 14, no. 2, pp. 274–287, Feb 2015.

[24] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, and Z. Turányi, "Design aspects of network assisted device-to-device communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, March 2012.

[25] K. Son, S. Lee, Y. Yi, and S. Chong, "REFIM: A practical interference management in heterogeneous wireless access networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 6, pp. 1260–1272, June 2011.

[26] E. Jorswieck, "Stable matchings for resource allocation in wireless networks," in *17th International Conference on Digital Signal Processing (DSP)*, July 2011, pp. 1–8.

[27] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *American Mathematical Monthly*, pp. 9–15, 1962.

[28] A. E. Roth and M. A. O. Sotomayor, *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge University Press, 1992, no. 18.

[29] G. O'Malley, "Algorithmic aspects of stable matching problems," Ph.D. dissertation, University of Glasgow, 2007.

[30] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, February 2001.

[31] S. Aji and R. McEliece, "The generalized distributive law," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 325–343, March 2000.

[32] A. Abrardo, M. Belleschi, P. Detti, and M. Moretti, "A min-sum approach for resource allocation in communication systems," in *IEEE International Conference on Communications (ICC)*, June 2011, pp. 1–6.

[33] M. Moretti, A. Abrardo, and M. Belleschi, "On the convergence and optimality of reweighted message passing for channel assignment problems," *IEEE Signal Processing Letters*, vol. 21, no. 11, pp. 1428–1432, November 2014.

[34] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: Numerical methods.* Prentice-Hall, Inc., 1989.

[35] D. Bertsekas, "Auction algorithms," in *Encyclopedia of Optimization*, C. Floudas and P. Pardalos, Eds. Springer US, 2001, pp. 73–77.

[36] M. Zavlanos, L. Spesivtsev, and G. Pappas, "A distributed auction algorithm for the assignment problem," in *47th IEEE Conference on Decision and Control*, December 2008, pp. 1212–1217.

[37] C. Liang and F. Yu, "Wireless network wirtualization: A survey, some Research issues and challenges," *IEEE Communications Surveys and Tutorials*, 2014.

[38] M. Hadzialic, B. Dosenovic, M. Dzaferagic, and J. Musovic, "Cloud-RAN: Innovative radio access network architecture," in *55th IEEE International Symposium (ELMAR)*, September 2013, pp. 115–120.

[39] M. Manjrekar, V. Ramaswamy, and S. Shakkottai, "A mean field game approach to scheduling in cellular systems," in *IEEE International Conference on Computer Communications*, April 2014, pp. 1554–1562.

[40] H. Tembine, R. Tempone, and P. Vilanova, "Mean field games for cognitive radio networks," in *American Control Conference (ACC)*, June 2012, pp. 6388–6393.

[41] H. Tembine, E. Altman, R. El-Azouzi, and Y. Hayel, "Evolutionary games in wireless networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 3, pp. 634–646, June 2010.

Additional Reading

- *5G and Heterogeneous Networks:*

  [i] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

  [ii] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5G era," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 90–96, February 2014.

  [iii] P. Pirinen, "A brief overview of 5G research activities," in *1st International Conference on 5G for Ubiquitous Connectivity (5GU)*, November 2014, pp. 17–22.

  [iv] Y. Lee, T. Chuah, J. Loo, and A. Vinel, "Recent advances in radio resource management for heterogeneous LTE/LTE-A networks," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 4, pp. 2142–2180, 2014.

- *Stable Matching:*

  [v] K. Iwama and S. Miyazaki, "A survey of the stable marriage problem and its variants," in *International Conference on Informatics Education and Research for Knowledge-Circulating Society (ICKS)*, January 2008, pp. 131–136.

  [vi] X. Feng, G. Sun, X. Gan, F. Yang, X. Tian, X. Wang, and M. Guizani, "Cooperative spectrum sharing in cognitive radio networks: a distributed matching approach," *IEEE Transactions on Communications*, vol. 62, no. 8, pp. 2651–2664, August 2014.

  [vii] A. Leshem, E. Zehavi, and Y. Yaffe, "Multichannel opportunistic carrier sensing for stable channel access control in cognitive radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 1, pp. 82–95, January 2012.

- *Message Passing:*

  [viii] M. Hasan and E. Hossain, "Distributed resource allocation for relay-aided device-to-device communication: A message passing approach," *IEEE Transactions on Wireless Communications*, 2014.

  [ix] A. Abrardo, M. Belleschi, P. Detti, and M. Moretti, "Message passing resource allocation for the uplink of multi-carrier multi-format systems," *IEEE Transactions on Wireless Communications*, vol. 11, no. 1, pp. 130–141, 2012.

[x] K. Yang, N. Prasad, and X. Wang, "A message-passing approach to distributed resource allocation in uplink DFT-Spread-OFDMA systems," *IEEE Transactions on Communications*, vol. 59, no. 4, pp. 1099–1113, 2011.

- *Auction Algorithm:*

[xi] Y. Zhang, C. Lee, D. Niyato, and P. Wang, "Auction approaches for resource allocation in wireless systems: A survey," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1020–1041, 2013.

[xii] I. Koutsopoulos and G. Iosifidis, "Auction mechanisms for network resource allocation," in *8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, May 2010, pp. 554–563.

[xiii] K. Yang, N. Prasad, and X. Wang, "An auction approach to resource allocation in uplink OFDMA systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4482–4496, November 2009.

[xiv] M. Bayati, B. Prabhakar, D. Shah, and M. Sharma, "Iterative scheduling algorithms," in *26th IEEE International Conference on Computer Communications. (INFOCOM)*, May 2007, pp. 445–453.