Energy Savings in 5G Automotive Nomadic Relaying Networks

vorgelegt von

Dipl. -Ing.

Zhe Ren

geb. in LiaoNing, China

von der Fakultät IV - Elektrotechnik und Informatik

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Ingenieurwisschenschaften

- Dr. -Ing. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. Hans-Joachim Grallert

Gutachter: Prof. Dr.-Ing. Sławomir Stańczak

Gutachter: Prof. Petar Popovski

Gutachter: Dr.-Ing. habil. Gerhard Wunder

Tag der wissenschaftlichen Aussprache: 01. April 2016

Berlin 2017

Zusammenfassung

Die Anforderung von künftigen Hochgeschwindigkeitsdatenzugängen erzeugt fundamentale Herausforderungen an die Gestaltung von modernen Fahrzeugtelematiksystemen und drahtlosen Kommunikationssystemen. Daraus resultierend ruft das Konzept des *Nomadic Relaying Networks* enormes Interesse sowohl in der Wissenschaft als auch aus der Industrie hervor. Ein Nomadic Relaying Network besteht aus zufällig verteilten *Nomadic Nodes* (z.B. parkenden Fahrzeugen mit On-Board Relay Komponenten), die nicht durch einen Mobilfunknetzbetreiber bereitgestellt werden. Diese ermöglichen die Möglichkeit der Multi-hop Kommunikation zwischen Benutzern und Basisstationen. Um die künftigen Konnektivitätsanforderungen effizient erfüllen zu können, arbeiten die Nomadic Nodes in einer selbstorganisierten Weise und werden auf einer nachfrageorientierten Basis aktiviert beziehungsweise deaktiviert.

Der Schwerpunkt dieser Arbeit liegt auf der Entwicklung einer Optimierungsarchitektur und von Optimierungsalgorithmen für Nomadic Relaying Networks. Unter Berücksichtigung von Nutzeranforderungen und der zur Verfügung stehenden Netzwerkkapazität wird ein mathematisches Optimierungsframework entwickelt. Die fundamentale Randbedingung der Optimierung liegt darin, die Nutzeranforderungen durch die verfügbare Bandbreite in dem System zu erfüllen. Darüber hinaus wird ein Lastkopplungsmodell etabliert, um die Nutzeranforderungen, Verbindungszuordnungen und Sendeleistungen mit den Auslastungen der Funkzellen zu verknüpfen. Die grundlegenden Eigenschaften des Optimierungsframeworks werden analysiert, um effiziente Optimierungsverfahren durchführen zu können.

Auf Basis des Optimierungsframeworks werden zentrale Aktivierungsalgorithmen, verteilte Zellauswahlalgorithmen, sowie dezentrale Algorithmen zur Optimierung des Energieverbrauchs in Nomadic Relaying Networks vorgeschlagen. In den zentralen Aktivierungsalgorithmen werden eine Reihe von linearen Programmen in einer zentralisierten Einheit ausgeführt, um die Verbindungszuordnungen zu ermitteln und die Zelle entsprechend zu aktivieren beziehungsweise zu deaktivieren. Überdies werden verteilte Algorithmen zur Zellauswahl, Zugriffskontrolle, sowie Deaktivierung der Zellen vorgeschlagen, um eine praktische Optimierungsimplementierung

aufzuzeigen. Weiterhin wird ein verteilter Energieregelungsalgorithmus entwickelt, um die Nutzer-anforderungen während der Übergangsphase der Aktivierung zu garantieren sowie weitere En-ergieeinsparungen zu erzielen. Es werden Simulationen durchgeführt, um die theoretischen Ergebnisse zu verifizieren. Durch intensive Untersuchungen der vorgeschlagenen Algorithmen unter realistischen Simulationsannahmen, kann eine signifikante Verbesserung der Energieef-fizienz festgestellt werden.

Abstract

The requirements of ubiquitous connectivity of high speed data access create fundamental challenges for the design of future wireless communication systems and modern vehicle telematic systems. Against this background, the concept of a *nomadic relaying network* is of enormous interest to both academia and industry. A nomadic relaying network consists of randomly distributed *nomadic nodes* (e.g. parked vehicles with on-board relay infrastructure) that are not deployed by an operator and offer the possibility of multi-hop relaying between users and base stations. The nomadic nodes operate in a self-organized fashion and are activated and deactivated on a demand-driven basis, efficiently addressing future connectivity requirements.

The focus of this thesis is the optimization architecture and optimization algorithms of the nomadic relaying network. A mathematical optimization framework is developed through consideration of the user requirements and the network capacity. The optimization constraints are formulated to ensure that the available bandwidth supports the data access requirements of all the users in the system. Furthermore, a load-coupling model is established to connect rate requirements, network assignments and transmission powers with the cell loads. Moreover, the fundamental properties of the optimization framework are analyzed for performing efficient optimization algorithms.

Based on this optimization framework, centralized node activation algorithms, distributed cell selection algorithms, and distributed power control algorithms are proposed for energy savings in the nomadic relaying network. In the centralized node activation algorithms, a series of linear programs is performed in a centralized manner, in order to optimize the network assignments and to activate or deactivate the nomadic nodes. In addition, a distributed cell selection, admission control and cell switch algorithm is introduced, in order to enable more practical implementations of the energy-saving algorithms. Furthermore, a distributed power control algorithm is proposed, in order to optimize the transient performance during the activation of nomadic nodes, as well as to achieve further energy savings. Simulation results are provided to confirm the theoretical analysis and the convergence of the proposed algorithms. Furthermore, a significant network-

wide benefit in terms of energy efficiency is identified by intensively evaluating the performance of the proposed algorithms under realistic simulation assumptions.

Contents

Chapter 1.

Background and Introduction

The telecommunication industry is entering a revolutionary era that is characterized by high demands on mobile data volume, large numbers of connected devices and the eco-social significance of sustainability. A large variety of innovative services and access technologies are under development, whereby the automotive sector attracts special interests. In this introductory chapter, we briefly review the history of and innovations in both automotive communication and telecommunication systems. Connecting both technology fields, we provide an insight into the nomadic relaying network, which is a novel deployment option for the next-generation mobile communication system. Subsequently, we outline the main contributions and contents of this thesis, before explaining the notation conventions at the end of this chapter.

1.1. The History of Automotive Communication

The origin of automotive communication dates back to wired in-vehicle networking in the 1990s, when the exponential increase of electronic systems gradually turned the modern vehicle from a mechanic and hydraulic machine into a complex distributed computer system [1]. The concept of *from P2P to Internetworking* then became infamous for vastly reducing the cabling costs to interconnect the large number of Electronical Control Units (ECUs). Bus systems such as Local Interconnect Network (LIN) and Control Area Network (CAN) were created in response to the initial demand for low-cost, low-rate and reliable communications, enabling applications such as the Electronic Stability Program (ESP), Antilock Braking System (ABS) and many other comfort functions. With advancements in multimedia technology and the increased functional safety requirements, dependable high-data-rate communications emerged in the automotive do-

main, accompanying the standardization of many high-performance communication protocols such as Media Oriented Systems Transport (MOST) and FlexRay for addressing entertaining systems as well as many x-by-wired functions with high communication requirements.

The irreplaceable advantages of the flexibility of wireless communication later attracted later enormous interest, leading to the further extension of automotive communication from *Wired to Wireless* systems. As the 21st century began, the prosperity of mobile device productions and the rapid development of mobile standards pushed the automotive Original Equipment Manufacturers (OEMs) to indispensably equip the modern vehicle with wireless modules. While dedicated Radio Frequency (RF) techniques, such as the remote keyless system, were developed separately by the OEMs, the in-vehicle networking was also well complemented by specification amendments of wireless communication systems such as the Bluetooth Hands Free Profile [2], in order to address the special requirements of automotive applications.

The automotive communication industry evolved *from in-vehicle networking to Telematics* with the maturity of Intelligent Transportation System (ITS) and Third Generation Partnership Program (3GPP) systems. By extending Wireless Local Area Network (WLAN) (IEEE 802.11) technology with the Wireless Communication in Vehicular Environments (WAVE), IEEE 802.11p system became the de facto Dedicated Short-Range Communications (DSRC) standard for improving traffic safety and efficiency in the framework of Vehicular Ad Hoc Network (VANET) [3]. A large number of application scenarios for avoiding traffic accidents and reducing unnecessary emissions are defined based on the Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications that are enabled by the IEEE 802.11p systems [4]. Due to the lack of eco-social deliberation, however, the roll out of the ITS standards stagnated, preventing the vehicles from benefiting from the innovations in wireless connectivity. On the other hand, the successful commercialization of a series of 3GPP cellular systems has facilitated a broad spectrum of telematic services based on mobile internet, including traffic efficiency services, fleet management applications and entertainment programs. Nowadays, almost all the major car manufactures have introduced branding telematic services depending on cellular connectivity, such as the BMW ConnectedDrive, Audi Connect or Mercedes Command Online. Moreover, the trend of connecting vehicles to the internet is being and will continue to be further promoted.

As the modern information society is vigorously changing the quality of our lives, novel applications are arriving on the horizon of the mainstream automotive industry. Although they are still in their infancy, the promising future of Machine to Machine (M2M) communications and Internet of Things (IoT) [5] engenders both challenges and opportunities regarding the con-

nectivity of a large number and variety of wireless sensors and Consumer Electronics (CE), as well as the exploitation of new services. Besides this, the continuous development of ubiquitous/pervasive computing and cloud services creates new dimensions for innovations related to vehicular connectivity. Furthermore, the mobile nature of the vehicles offers the possibilities of delivering dynamic services, where wireless communication capability plays a fundamental role in keeping the mobility under control. In general, the evolution of the automotive communication system reflects exactly the evolution of customer interests, the eco-social condition and the maturity level of the electronics industry. The integration of wireless communication technologies into the vehicle is *Still Drastically Expanding*, especially in combination with modern cellular technologies, which have been steadily evolving over the past decades.

1.2. The Evolution of Mobile Cellular Systems

Mobile cellular systems are often divided into "Generations", with each generation making revolutionary improvements in both technological fundamentals and service enhancements. Before the formation of 3GPP in 1998, voice services were of major concern in the First Generation (1G) analog and the Second Generation (2G) digital networks. Whereas the 1G cellular systems were launched only by regional operators, such as Nippon Telegraph and Telephone (NTT) in Japan and Nordic Mobile Telephone (NMT) in the Nordic countries, the Time Division Multiple Access (TDMA) based Global System for Mobile Communications (GSM) began to dominate the global market in the 1990s. Apart from the fact that the digitally encrypted voice signals significantly increased the efficiency and flexibility of 2G systems, data services such as Short Message Services (SMS) and Multi Media Messages (MMS) gradually became an indispensable part of everyday life. In order to address the increased demand for high data rate applications, General Packet Radio Service (GPRS) and Enhanced Data Rate for GSM Evolution (EDGE), sometimes branded as 2.5G, were added to GSM, introducing packet switch and achieving a peak rate of up to 1Mbps.

The pre-Third Generation (3G) systems attained huge commercial successes, however, the fundamental limitations of TDMA systems required later revolutionary changes to enable large scale mobile data services. In this context, the foundation of 3GPP arose to take the leading role in the world-wide standardization of 3G systems. The technological revolution for the 3G systems lies in the application of Code Division Multiple Access (CDMA) technique, with which a higher spectrum efficiency and system capacity can be achieved. 3G systems, such as the Universal Mobile Telecommunications System (UMTS), were commercialized in the early

2000s. They further evolved into post-3G systems, such as High Speed packet Access (HSPA), providing 14Mbps and 5.8Mbps peak data rates in downlink and uplink, respectively.

An advanced milestone was set up in 2008, when Long Term Evolution (LTE) was introduced to satisfy the Fourth Generation (4G) requirements set by International Mobile Telecommunications (IMT)-advanced in a step-by-step manner. Based on Orthogonal Frequency Division Multiplexing (OFDM), an LTE system reaches a peak rate of 100 Mbps and supports high mobility of up to 500km/h. Moreover, the all-IP System Architecture Evolution (SAE) enables a more efficient and flexible network deployment. The LTE-Advanced (Release 10) is selected to be a 4G conform system in 2010 [6] and is now under rollout by several operators. In the upcoming Releases 11/12, Self-Organizing Network (SON), interworking with WiFi, small cell deployment and location-based services have been treated as the main working focuses [7].

As the LTE marshes towards a powerful future of the cellular system, world-wide experts have recently collaborated in aiming for the foundation of the Fifth Generation (5G) systems, where novel services, network scalability and system efficiency are among the most important objectives. High data volume, short latency and low energy cost are the key performance targets in the 5G systems, while novel system concepts for ultra-dense networks, massive Multiple Input Multiple Output (MIMO) and moving networks are designed for the forthcoming generation [8]. BMW, which represents the automotive industry, is actively participating in developing the 5G standards. The main focuses are, on the one hand, delivering traffic safety and efficiency functions through novel cellular reliable communication and, on the other hand, enabling a dynamic network access option that makes use of the automotive telematic systems: the nomadic relaying network.

1.3. Nomadic Relaying Networks

The emerging requirements for the universal connectivity of everything create new challenges for the design of modern vehicle telematic systems. In this context, an innovative concept of a nomadic relaying network that combines both automotive and cellular communication systems is established as an important 5G component [9]. In response to the boosting traffic dynamics, a nomadic relaying network (or a nomadic network for short) is designed to efficiently extend the cellular system with the help of the vehicle telematic systems and multi-hop communication technologies. A nomadic network consists of randomly distributed non-operator-deployed nodes (e.g. parked vehicles with on-board relay infrastructure) offering the possibility of multi-hop

relaying between User Equipments (UEs) and Base Stations (BSs). Note that relaying here is a logical concept that defines the multi-hop transmission and the relaying of user data, and it can be implemented by different multi-hop technologies. While the location of operator-deployed Relay Nodes (RNs) is optimized by means of network planning tools, the location of the RNs in a nomadic network, referred to as *nomadic RNs* or *nomadic nodes*, is out of the control of the network operators, and is considered to be random. Moreover, their availability and position may change over time (hence, the term "nomadic") due to battery state or node movement. The nomadic RNs operate in a self-organized fashion and are generally activated or deactivated based on capacity, coverage, load balancing or energy efficiency demands. Therefore, the concept of a nomadic network describes an effective extension of the cellular infrastructure that allows for a dynamic network deployment.

1.3.1. Enabling Technologies

The recent advancements and trends in both cellular and automotive areas support the deployment of the nomadic relaying network concept. In general, the standardization enhancements in multi-hop communications and the framework of SON plug-and-play capability give options for protocols and architectures, while the improvements of the connected car platforms and vehicle telematic systems serve as enablers in the automotive perspective for the nomadic network.

Firstly, the standardization of LTE Relaying has been introduced in Release 10, aiming at a cost-efficient coverage and capacity extension in the heterogeneous network [10]. An LTE L3-RN [11] is defined to be a low-power evolved Node B (eNodeB) that supports all the eNodeB functionalities and is seen by the UEs as a regular eNodeB, whereas it also supports a subset of UE functionalities in order to wirelessly connect to the donor eNodeB. Note that we use BS to denote all types of base stations in a cellular system, including, e.g., the NodeB and the eNodeB. With respect to the spectrum usage, relay operations have been further specified into inband (Type 1 and Type 1b) and outband (Type 1a) types. Whereas Type 1 RNs utilize the same bandwidth for both BS-RN and RN-UE links, Type 1b RNs assume a physical isolation in order to ensure a limited interference level between them. Furthermore, a Type 1a is an outband RN that is characterized by the same set of features as the Type 1 relay node. An important advantage of RN deployment is that the architectural functionalities, such as Mobility Management Entity (MME) and Service Gateway (S-GW), are all proxied by the donor eNodeBs. This means that only the Radio Access Network (RAN) protocol needs to be adapted for admitting RNs into the network. Besides the standardization of relaying, the enhancement of WiFi integration

and the investigation of Device to Device (D2D) communication in Release 11/12 provides further relaying and backhauling possibility for multi-hop communications, extending the device compatibility to redirect traffic through further access technologies and spectral bands.

Another important enabling technology for the nomadic network is the framework of SON, which could enable the network adaptation within a few minutes [12]. The nomadic nodes are due to vehicle being mobility non-stationary, i.e., an exchange of management information is more frequently required than in traditional network planning scenarios. Whereas the manually reconfiguration of such dynamic networks is obviously not practical, the concept of SON, which has been consolidated and is further evolving, makes the management of nomadic nodes feasible. While the framework on SON-relays [13] further narrows the gap in implementing the nomadic network, the SAE working groups are also actively working towards a more generic and powerful tool for the provisioning of higher layer services for plug-and-play terminals [14].

Meanwhile, in the automotive industry, the commercialization of telematic systems derived from the cutting-edge advancements of wireless communication system opens another door for the realization of nomadic networks. For instance, many car manufactures, public transportation carriers and plug-in gadgets are providing WiFi hotspot functions that facilitate internet for WiFi compatible CE through cellular internet accesses. Moreover, the potential concept of deploying moving femtocells in the vehicles has also been discussed in research contributions [15]. On the other hand, the maturity of the connected car service platforms that are maintained by the automotive OEMs further enables the centralized management and the potential remote on-demand activation and deactivations of the nomadic nodes.

1.3.2. Functional Architecture

Considering the future 5G architecture of the cellular communication systems, we elaborate on a potential architecture for the nomadic relaying network that integrates future cellular systems with the automotive service platform and nomadic nodes. The architecture enables both centralized and distributed management mechanisms for enhancing cellular capabilities through nomadic nodes.

Fig 1.1 illustrates a simplified architectural design that contains only necessary functional elements for performing optimizations to the nomadic relaying network. We aim to provide both centralized cloud-based and distributed RAN-based optimization mechanisms. Therefore, the
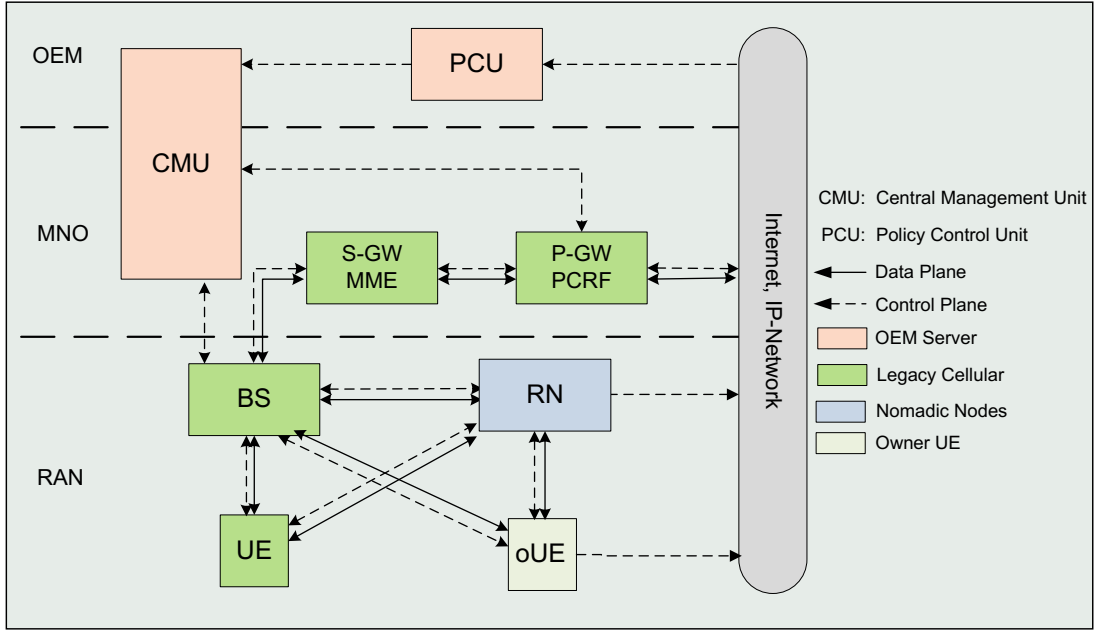
Figure 1.1.: Functional Architecture of the Nomadic Relaying Network.

RAN components are also displayed, where the nomadic RN and the owner UE are defined. Firstly, a nomadic node implements the LTE relay specifications, through which it inherits all BS functionalities and reaches the key Evolved Packet Core (EPC) components, such as MME, S-GW and Packet Gateway (P-GW), for performing BS functionalities upon activation. Without specifying RAN protocol, we assume the access link of the nomadic nodes can be realized by different radio access technologies. Apart from the RAN extension, a Centralized Management Unit (CMU) is also located between Mobile Network Operator (MNO) and OEM for centralized optimization. Whereas the centralized management unit has been discussed in the framework 5G architecture [16], we assume an interface or joint operation of CMU between MNO and OEM, and thus, the block of CMU crosses two network levels. Furthermore, a Policy Control Unit (PCU) that is maintained by the OEMs can be updated dynamically by the nomadic RNs and owner UEs through secure internet access. The OEM side policy can be further fed to the Policy and Charging Rules Function (PCRF) of the MNOs for potential business cooperation.

In Fig 1.1, the dashed arrows are the flows of control signals for configurations, channel feedbacks and context information. The channel and context measurements are then available for centralized optimizations at CMU for performing centralized optimizations. Based on the control plane signals, there are in general two main mechanisms for optimizing the nomadic network. Firstly, a centralized optimization can be carried out at CMU based on the channel feedback and the user policy, whereby the optimization decision can be sent to the BSs and then further to the entire RAN. Secondly, a distributed mechanism can be realized within the RAN by simply forward-

ing the control policy to the BSs, which then update the Radio Resource Management (RRM) algorithms for the RNs and UEs. Besides the main mechanisms, a deterministic control of the nomadic RN should also be possible such that the owner UE may directly activate the RN, such as by sending activation signals to the RN. Note that the admission is finally made by the RAN and the core network, even if the centralized optimization provides deterministic results. This is due to the fact that the dynamic change in the network may lead to cell over load or coverage loss, which generally cannot be accepted by the MNOs.

The data plane flows are shown in solid lines, and it can be observed that they traverse only through the legacy network extended by the RNs. The case when the RNs have a backhaul link through WiFi or other third party connections is omitted here, since it requires further authentication and management mechanisms that are beyond the scope of this thesis. However, we do not exclude the possibility of having a variety of backhaul solutions as extensions for further performance enhancements.

1.3.3. Further Challenges and Opportunities

The nomadic relaying network, due to its randomness and the relation to vehicles, raises significant challenges, especially in the aspects of management and business. The main challenges, both technical and non-technical, are as follows.

- Due to the large diversity of nomadic relays, the management of such a dynamic network becomes critical. In particular, the generated dynamic interference requires enhanced RRM solutions.

- The business cooperation between network operators, car manufactures and other private stakeholders should result in a reasonable compromise to enlarge the benefits of every participant in the business.

On the other hand, new opportunities come along with the challenges thanks to the recent development in both the technical and business sectors.

- A large number of potential nomadic RNs, both privately owned vehicles and car fleets (car rental, car sharing, taxi), are available for performing network optimization.

- More space for antenna design allows for the possibility of further backhaul link enhancements and implementation options for relaying and multi-hop communications.

- The low-power nature of small cell and the rising trend for electro-vehicles ease the critical energy consumption problem of standing vehicles.

1.4. Contribution and Organization of the Thesis

In this thesis, we focus on the technical aspects of the challenges of realizing the innovative concept of nomadic networks. Based on the current enabling technologies and the functional architecture, we propose both centralized and distributed algorithms for operating the nomadic relaying network. We study the network-wide benefits in terms of energy savings by evaluating the performance of the proposed algorithms considering realistic network configurations and the opportunities mentioned in the previous subsection. The major contributions of this thesis can be summarized as follows.

- A functional architecture is designed to enable both centralized and distributed optimizations of the nomadic network.

- A novel and abstracted model of the nomadic network is presented, where we assume that the decisions on assignment, routing and power control are the main control parameters for network optimization.

- We establish a load function under a generic interference model and elucidate the fundamental properties of the load function.

- Based on the load function, a generic optimization framework is formulated where the UE Quality of Service (QoS) satisfaction, load balancing and power limitation express the essential constraints of the nomadic network.

- We propose centralized iterative algorithms for assignment/routing optimization for energy savings, with the help of the nomadic nodes.

- We propose a distributed energy-aware cell selection and admission control algorithm which can be practically implemented in future cellular systems.

- We propose a distributed power control algorithm that ensures the transitional performance of the network during the cell activation procedure in the nomadic network.

As logically depicted in Figure 1.2, the rest of the thesis is organized as follows: By abstracting the concept of nomadic relaying network, Chapter 2 introduces a system model that includes
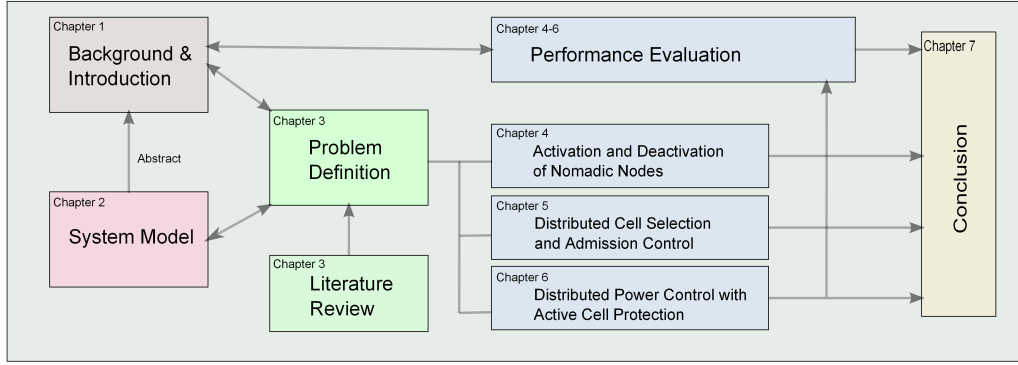
Figure 1.2.: Organization of the Thesis.

network deployment, resource utilization and a link rate model. Then, the energy-saving optimization problem is formulated in Chapter 3, with an extensive literature review of the current approaches. Following the problem definition, Chapter 4, Chapter 5 and Chapter 6 discuss in detail the optimization algorithm for centralized activation and deactivation, distributed cell selection and admission control and distributed power control algorithms, respectively. The corresponding performance evaluations are given in each chapter, confirming the theoretical analyses and the performance benefit. Finally, conclusions are drawn and suggestions are made in Chapter 7.

1.5. Notation Conventions

Scalars, vectors and matrices are written in regular, bold lower case and bold upper case letters, respectively. For instance, we denote $x \in \{0, 1\}$ and $\mathbf{X} \in \{0, 1\}^{(M+K) \times (N+K)}$ to be the assignment variable and the assignment matrix, while we use $\rho \in [0, 1]$ and $\boldsymbol{\rho} \in [0, 1]^{M+K}$ for the load and the load vector, respectively. Furthermore, we use the corresponding lower case letter to denote the vectorization of a matrix, which is defined by column-wise stacking the entries of the matrix in a vector, e.g., $\mathbf{x} = \text{vec}(\mathbf{X})$. The identity matrix of size $K \times K$ is written as $\mathbf{I}^K$, while $\mathbf{1}^L(\mathbf{0}^L)$ and $\mathbf{1}^{M \times N}(\mathbf{0}^{M \times N})$ refer to a column vector of length L and an $M \times N$ matrix of ones (zeros), respectively. If not specified, $\mathbf{1}(\mathbf{0})$ and $\mathbf{I}$ are matrices or vectors with the proper size for matrix operator. For any two matrices (or vectors) $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \cdot \mathbf{B}$ (or simply $\mathbf{AB}$) denotes the normal matrix product and $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product. If $\mathbf{A}$ and $\mathbf{B}$ are of the same size, $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard matrix product, while both $\mathbf{A} \leq (\text{or} <)\mathbf{B}$ and $\mathbf{A} \geq (\text{or} >)\mathbf{B}$ should be understood element-wise. Assume $\mathbf{F} : \mathcal{X} \to \mathcal{Y}$ to be a function, where $\mathcal{X}$ and $\mathcal{Y}$ are the domain and the image of the function, respectively. Throughout this thesis, we use $\mathbf{J}_{\mathbf{F}}^{\mathbf{x}}(\mathbf{x})$, $\mathbf{x} \in \mathcal{X}$, to denote the Jacobien of function $\mathbf{F}(\cdot)$ with respect to $\mathbf{x}$. Furthermore, given $\mathbf{x}$, the i-th

row and the an entry i, j in $\mathbf{J_F^x}(\mathbf{x})$ (gradient) are denoted as $\mathbf{J_{F_i}^x}(\mathbf{x})$ and $\mathbf{J_{F_i}^{x_j}}(\mathbf{x})$, respectively.

1.6. Publications and Copyright Information

Parts of this thesis have already been published or submitted as conference proceedings and journal articles in [17–23]. These parts, which are, up to minor modifications, identical with the corresponding scientific publication, are ©2013-2015 IEEE.

During the time of my Ph.D., we were also able to produce scientific findings in the area of handover optimization, interference identification, etc. The results in [24–26] are achieved but not included in this thesis.

The publications achieved during the thesis are listed below:

- Z. Ren, S. Stańczak, P. Fertl, and F. Penna, "Energy-Aware Activation of Nomadic Relays for Performance Enhancement in Cellular Networks," in *Proceedings of IEEE International Conference on Communications (ICC), Sydney, Australia*, June 2014, pp. 1–6

- Z. Ren, S. Stańczak, and P. Fertl, "Activation of Nomadic Relays in Dynamic Interference Environment for Energy Savings," in *Proceedings of IEEE Global Conference on Communications (GLOBECOM), Austin, Texas*, December 2014, pp. 1–6

- Z. Ren, S. Stańczak, M. Shabeeb, P. Fertl, and L. Thiele, "A Distributed Algorithm for Energy Saving in Nomadic Relaying Networks," in *Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA*, November 2014, pp. 1–5

- Z. Ren, M. Jäger, S. Stańczak, and P. Fertl, "Distributed Power Control with Active Cell Protection in Future Cellular Systems," in *IEEE International Conference on Communications (ICC), London, UK*, June 2015, pp. 1–6

- Z. Ren, S. Stańczak, and P. Fertl, "An Optimization Framework for Energy Saving in 5G Nomadic Relaying Networks," *Preprint*, 2015

- Ömer Bulakci, Z. Ren, C. Zhou, J. Eichinger, P. Fertl, and S. Stańczak, "Dynamic Nomadic Node Selection for Performance Enhancement in Composite Fading/Shadowing Environments," in *Processing of IEEE Vehicular Technology Conference Spring (VTC-Spring), Soul, Korea*, May 2014, pp. 1–6

- Ömer Bulakci, Z. Ren, C. Zhou, J. Eichinger, P. Fertl, D. Gozalvez-Serreno, and S. Stańczak, "Towards Flexible Network Deployment in 5G: Nomadic Node Enhancement to Heterogeneous Networks," in *Workshop of IEEE International Conference on Communications (ICC Workshop), London, UK*, June 2015, pp. 1–6

- Z. Ren, P. Fertl, Q. Liao, F. Penna, and S. Stańczak, "Street Specific Handover Optimization in Future Cellular Networks," in *Processing of IEEE Vehicular Technology Conference Spring (VTC-Spring), Dresden, Germany*, June 2013, pp. 1–5

- Q. Liao, F. Penna, S. Stańczak, Z. Ren, and P. Fertl, "Context-Aware Handover Optimization for Relay-Aided Vehicular Terminals," in *Proceedings of IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Darmstadt, Germany*, June 2013, pp. 1–5

- F. Penna, S. Stańczak, Z. Ren, and P. Fertl, "MMSE Interference Identification in LTE Networks," in *Proceedings of IEEE International Conference on Communications (ICC), Sydney, Australia*, June 2014, pp. 1–6

Chapter 2.

System Model

In this chapter, a downlink model of a nomadic network is introduced as the framework for performing network optimizations. By adapting a generic interference model, the data link rate is modeled as a function of transmission powers, network loads and channel measurements. Combining the data link rate and the network assignment, a network load coupling model is established where the transmission power vector and the network assignment matrix are included as control parameters. We elaborate some fundamental constraints for network optimizations with respect to these parameters. In particular, resource sharing schemes for interference coordination are discussed. Note that parts of the work in this chapter are based on the publications in [17–20].

2.1. Network Deployment and Connection Assignment

Consider a downlink model of a nomadic network with M BSs, K RNs and N UEs. The sets of BSs, RNs and UEs are denoted by, respectively, $\mathcal{B}$, $\mathcal{R}$ and $\mathcal{U}$. As illustrated in Fig 2.1, we use direct links, access links and relay links to refer to the BS-UE, RN-UE and BS-RN links, respectively. Throughout the thesis, notations with superscripts (m), (n) and (k) are, respectively, variables associated with BSs, UEs and RNs, while the pairs (m,n), (k,n) and (m,k) are used to denote to the direct links, access links and relay links, correspondingly. Regarding relaying operation, the following assumptions hold throughout the manuscript:

Assumption 2.1. L3 relaying [11]: each RN has all the RRM functionalities of a BS and is seen by the UEs as a conventional BS.

Assumption 2.2. Only one-hop relaying: there exists no connection between RNs.
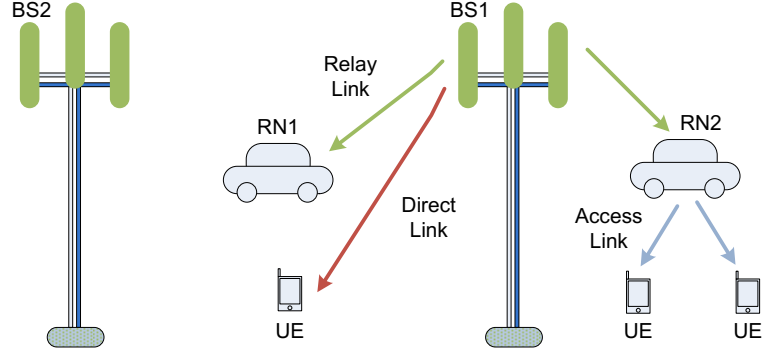
Figure 2.1.: Cells, nodes and links in a noamdic relay network.

The term *cell* refers to a BS or an RN to which a UE can be connected, whereas *node* is a UE or an RN that is searching access to a cell. Then, we have the multi-user downlink scenario in which $M + K$ cells are allocated some frequency spectrum to serve $N + K$ nodes. The frequency bandwidth (in Hz) at each cell is fixed and the bandwidths of all cells are grouped in a vector $\mathbf{b} = \begin{pmatrix} \mathbf{b}^{(m)} \\ \mathbf{b}^{(k)} \end{pmatrix} = (b_1, \ldots, b_M, b_{M+1}, \ldots b_{M+K})^T$ where $\mathbf{b}^{(m)} = (b_1^{(m)}, \ldots, b_M^{(m)})^T$ and $\mathbf{b}^{(k)} = (b_1^{(k)}, \ldots, b_K^{(k)})^T$ refer to the bandwidths allocated to the BSs and RNs, respectively.

Furthermore, both centralized and distributed approaches are considered in this work for optimizing the network assignment.

- The network connection assignments between the cells and nodes can be controlled by the CMU, which performs a centralized optimization (Chapter 4).

- If a distributed operation is targeted, the assignments can be also decided directly by the nodes based on some network measurements (Chapter 5).

Let $x_{i,j}$ denote the assignment variable: $x_{i,j} = 1$ if there is an active connection between cell i and node j, and $x_{i,j} = 0$ otherwise. These variables are clustered in the assignment matrix defined to be:

$$\mathbf{X} \triangleq \begin{pmatrix} \mathbf{X}^{(m,n)} & \mathbf{X}^{(m,k)} \\ \mathbf{X}^{(k,n)} & \mathbf{X}^{(k,k)} \end{pmatrix} \in \{0,1\}^{(M+K)\times(N+K)}. \tag{2.1}$$

Here and hereafter, $\mathbf{X}^{(m,n)} \in \{0,1\}^{M\times N}$, $\mathbf{X}^{(k,n)} \in \{0,1\}^{K\times N}$ and $\mathbf{X}^{(m,k)} \in \{0,1\}^{M\times K}$ are assignment matrices for the direct, access and relay links, respectively. Note that $\mathbf{X}^{(k,k)} \in \{0,1\}^{K\times K}$ is the connection between RNs and is set to be an all zero matrix according to

Assumption 2.2. Therefore, throughout the manuscript, we have $\mathbf{X}^{(k,k)} \equiv \mathbf{0}$ and

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(m,n)} & \mathbf{X}^{(m,k)} \\ \mathbf{X}^{(k,n)} & \mathbf{0} \end{pmatrix}. \tag{2.2}$$

Assumption 2.3. We consider Full connectivity scenario where each node is connected to the network either via a BS or an RN.

Assumption 2.3 is ensured by imposing the following condition:

$$\mathbf{X}^T \cdot \mathbf{1}^{M+K} = \mathbf{1}^{N+K}. \tag{2.3}$$

The left hand side of equation (2.3) computes the column-wise sum of $\mathbf{X}$, and therefore equation (2.3) guarantees that every node is connected to exact one cell. Note that this equation also implies that an RN is always connected to the network, even though it is sometimes not actively transmitting and receiving data.

2.2. QoS Model and Rate Assignment

In the previous work on network planning [27], a concept of rate density per area is used for optimizing the long term network performance. Furthermore, a queuing model of the user rate distribution is given in [28] for the SON relaying framework. A location based coverage and assignment strategy is assumed for both works, aiming at maximizing the network capacity with the help of network and relay planning tools. In this work, according to the on-the-fly concept of the nomadic network, the optimization is based on a short-term rate requirement with a potential irregular rate distribution.

Assumption 2.4. We adopt a simple constant rate QoS model that represents the short-term average rate of the UEs in the network. Further, we assume that the temporal and spatial change of the rate QoS is insignificant and negligible within the time scale of the network optimization.

The QoS requirements of a finite number of UEs are given by a vector of minimum rates (in bit/s), which is denoted by $\mathbf{r}^{(n)} = (r_1^{(n)}, \ldots, r_N^{(n)})^T$. For each RN, the rate requirements of all the connected nodes should be satisfied by the backhaul link (relay link). This means, the backhaul

link of RN $k \in \mathcal{R}$ must support the sum of the rate requirements of the UEs connected to it:

$$r_k^{(\mathrm{k})} = \sum_{j \in \mathcal{U}} r_j^{(\mathrm{n})} x_{k,j}, \quad \text{for } k \in \mathcal{R} . \tag{2.4}$$

By using the RN rate vector defined to be $\mathbf{r}^{(\mathrm{k})} = (r_1^{(\mathrm{k})}, \dots, r_K^{(\mathrm{k})})^T$, we can write (2.4) in a matrix form as

$$\mathbf{r}^{(\mathrm{k})} = \mathbf{X}^{(\mathrm{k,n})} \mathbf{r}^{(\mathrm{n})}. \tag{2.5}$$

For a better readability, we also use the vector $\mathbf{r} := \left(\begin{smallmatrix} \mathbf{r}^{(\mathrm{n})} \\ \mathbf{r}^{(\mathrm{k})} \end{smallmatrix} \right) = (r, \dots, r_N, r_{N+1}, \dots r_{N+K})^T$ to refer to the rate requirements of all nodes.

Assumption 2.5. For centralized optimization, we assume that $\mathbf{r}^{(\mathrm{n})}$ is known at the CMU or can be estimated reliably.

Assumption 2.6. A node can be connected to multiple nodes [29] and a value of $x_{i,j}$ between $[0,1]$ indicates the portion of the data traffic that should be delivered through link (i,j).

The optimization problem with $x_{i,j} \in [0,1]$ is a relaxation problem of the original one. Note that heuristic solutions for the original problem can be found by mapping (e.g., rounding) the solution for the relaxed problem back to the original domain $\{0,1\}^{(M+K) \times (N+K)}$. Note that further adjustments may be needed to avoid violations of the original QoS constraints. The assignment matrix $\mathbf{X} \in [0,1]^{(M+K) \times (N+K)}$ can be also interpreted as a routing matrix. Throughout this manuscript, we use the phrase (rate) assignment matrix and (rate) routing matrix interchangeably for $\mathbf{X}$. Furthermore, a value $x_{i,j} > 1$ is also feasible and it indicates that rate throughput of the node is higher than its minimum demand. Thus, for satisfying the QoS requirements of all the UEs, the following condition should be satisfied:

$$(\mathbf{X}^T \cdot \mathbf{1}^{M+K}) \circ \mathbf{r} \geq \mathbf{r}. \tag{2.6}$$

It means, each node needs to be connected to the network and sum rate of every node must be larger than or equal to the minimum requirement $\mathbf{r}.$. If $\mathbf{r} > \mathbf{0}$, we can equivalently write the condition as:

$$\mathbf{X}^T \cdot \mathbf{1}^{M+K} \geq \mathbf{1}^{N+K}. \tag{2.7}$$

Note that this is a sufficient condition for (2.6), since by Assumption 2.6 all the RNs should be connected to the network which is implied by (2.7), whereas by (2.6), no constraints on an RN need to be satisfied if no UE is connected to it and the rate requirement of the RN is 0.

Assumption 2.7. Throughout this manuscript, all RNs need to be connected to the network

for communicating the necessary control plane data which is assumed to consume a negligible amount of resources of the system.

Therefore, we take (2.7) as the general constraint for system design.

2.3. Link Rate Model

The achieved rate per channel or Spectral Efficiency (SE) of a link (i, j) (in bits/s/Hz or bit-s/channel use) is approximated by the Shannon's capacity formulation [11, 30]:

$$\omega_{i,j} = \zeta_b \cdot \log(1 + \zeta_s \cdot \tau_{i,j}), \tag{2.8}$$

where $0 \leq \zeta_b \leq 1$ is the bandwidth efficiency and $0 \leq \zeta_s \leq 1$ refers to the Signal-to-Interference-plus-Noise-Ratio (SINR) efficiency. Depending on link transmission techniques such as MIMO or Adaptive Modulation and Coding (AMC), ζ_b and ζ_s may have different values.

Assumption 2.8. Without loss of generality, we assume $\zeta_b = 1$ and $\zeta_s = 1$ to simplify the notation throughout the manuscript.

We follow the previous studies on the probabilistic interference modeling [31–33] to formulate a generic interference model by defining an interference relation matrix of the same size and the same structure as the assignment matrix:

$$\mathbf{S} \triangleq \begin{pmatrix} \mathbf{S}^{(m,n)} & \mathbf{S}^{(m,k)} \\ \mathbf{S}^{(k,n)} & \mathbf{S}^{(k,k)} \end{pmatrix} \in \{0, 1\}^{(M+K) \times (N+K)}. \tag{2.9}$$

Here and hereafter, an entry $s_{i,j}$ indicates whether the cell i can cause interferences to node j or not: If $s_{i,j} = 0$, it means cell i never transmits at the frequency resources that is allocated to node j.

Further, we define the vector of transmission powers of the cells to be

$$\mathbf{p} \triangleq [\begin{smallmatrix} \mathbf{p}^{(m)} \\ \mathbf{p}^{(k)} \end{smallmatrix}] = [p_1, \ldots p_M, p_{M+1}, \ldots, p_{M+K}]^T. \tag{2.10}$$

In a real system, there are limitations on the transmission powers so that there holds:

$$\mathbf{0} \leq \mathbf{p} \leq \hat{\mathbf{p}}, \tag{2.11}$$

where $\hat{\mathbf{p}}$ is the maximal transmission power vector that depends on the hardware of the cell and some system parameters. Furthermore, define $\rho_i \in [0,1]$ to be the load or the activity level, which is also the interference scaling factor of the cell i and is explained in detail in Section 2.5.

Then, the SINR of link (i,j) can be computed as

$$\tau_{i,j} = \frac{p_i g_{i,j}}{\sum_{d\in\mathcal{B}\bigcup\mathcal{R}} p_d g_{d,j} s_{d,j} \rho_d + \sigma_j}, \qquad (2.12)$$

where σ_j and $g_{i,j}$ refer to the receiver-side noise power and channel gain for link (i,j), respectively.

Assumption 2.9. We assume a constant channel gain that captures the average propagation fluctuation including pathloss and shadow fading. Within the time scale of the optimization, the impact of small scale fading is averaged out.

Hence, the channel gain model can be written as

$$g_{i,j} = \frac{g_f}{d_{i,j}^\gamma}, \qquad (2.13)$$

where $d_{i,j}$ is the path length of link (i,j) and γ is the pathloss exponent that depends on factors such as carrier frequency, antenna heights and antenna gain. Furthermore, g_f is the shadow fading which is spatially log-normal distributed with variance σ_f and correlation distance d_f. Note that $\tau_{i,j} > 0$ always holds, since both received power and interference plus noise power are positive values. A positive $\tau_{i,j}$ exists also for the case when both $i,j \in \mathcal{R}$, however, we do not allow a connection between RNs in this work.

2.4. Interference Management

Considering the network in Fig 2.1, where both RNs are assigned to BS1, the *full duplex full reuse* resource utilization is depicted in Fig 2.2. If no interference management scheme is applied, i.e., $\mathbf{S} = \mathbf{1}^{(M+K)\times(N+K)}$, a node is interfered by all other cells except for the serving cell. In this case, an RN reuses the whole bandwidth for the access link transmission and causes interference to all the nodes served by the other RNs and BSs including its own relay link. However, self-interference should be avoided for practical operation through interference cancellation or duplexing techniques, since the transmitter and the receiver of an RN are located at the same place and may interfere severely with each other. A more general statement regarding
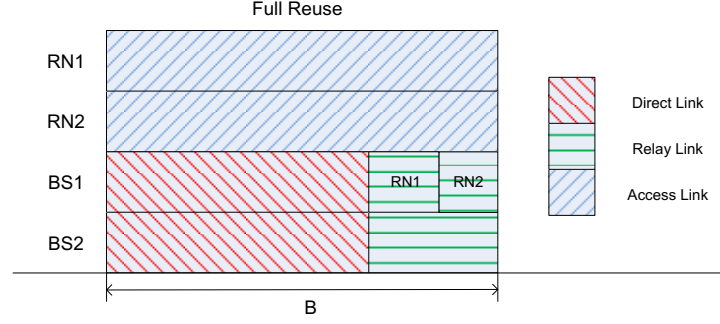
Figure 2.2.: Full reuse relay resource utilitzation model.

interference is that if the channel gain $g_{i,j}$ is very high, interference should be excluded from cell i to node j. Furthermore, if the access network of the nomadic nodes operates in another band, interferences are automatically avoided between backhaul links and access links. In the following, we discuss the approaches for interference coordination in a nomadic network and its impact on the interference relation matrix **S**.

2.4.1. RN-RN Interference

The term RN-RN interference refers to the interference generated by the access link transmission of the RNs to the relay link transmissions [34]. In particular, *self-interference*, which is the interference from access link to the relay link of the same RN, severely restricts the performance of the relay link. Practically, if full reuse is applied, self-interference cancellation or access/relay link isolation is necessary for an efficient RN operation. In case of the nomadic relay network, the physical isolation (LTE Type I.(b) relay [11]), which can be understood as $g_{i,j}s_{i,j} = 0$, between transmit and receive antennas is not feasible. Hence, dealing with self-interference and RN-RN interference is an essential challenge for performing in-band relaying in nomadic relaying networks. In LTE, the RNs can be configured to blank some resources on the access link transmission for the relay links. In Fig 2.3, we illustrate two configuration schemes that address the RN-RN interference and self-interference, respectively. The scheme (a) in Fig 2.3 is called *synchronized in-band*, since all the RNs are configured to exclude the same access link resources which can be then used for relay link transmissions. The disadvantage of these schemes is obviously the possible waste due to the unused resources that are reserved for the relay links of some RNs.
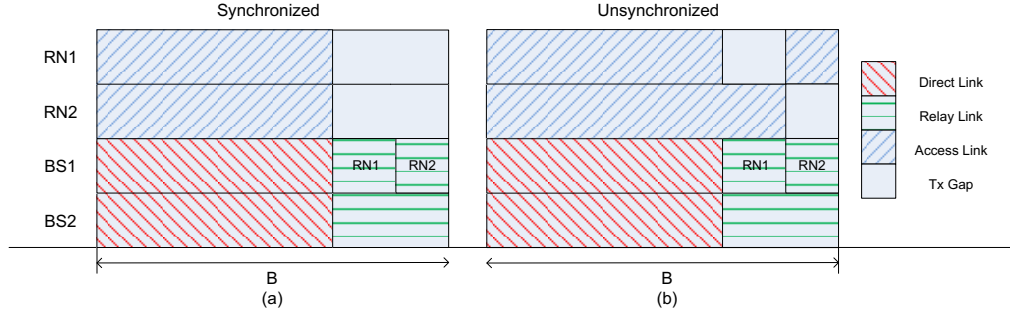
Figure 2.3.: In-band relays resource utilization models.

For scheme (a), the interference relation matrix can be written as:

$$s_{i,j} = \begin{cases} 0 & \text{for } i,j \in \mathcal{R}, \\ 1 & \text{otherwise.} \end{cases} \tag{2.14}$$

The scheme (b) in contrast is called *unsynchronized in-band*, where RNs split a random portion of frequency resources for their relay links only to avoid self-interference and a certain probability of RN-RN interference exists. In this scheme, $\mathbf{S}^{(k,k)}$ has zeros only on the diagonal element while the whole matrix $\mathbf{S}$ is 1 elsewhere, such that:

$$s_{i,j} = \begin{cases} 0 & \text{for } i = j \in \mathcal{R}, \\ 1 & \text{otherwise.} \end{cases} \tag{2.15}$$

2.4.2. General Interference Coordination Constraints

A more general approach for interference coordination is to exclude interferences between closely located transmitters and receivers. Interference should be avoided if the channel between interfering source and the receiving node is very high. Hence, a more general interference coordination scheme can be formulated as:

$$s_{i,j} = \begin{cases} 0 & \text{for } g_{i,j} \geq g_t, \\ 1 & \text{otherwise,} \end{cases} \tag{2.16}$$

where g_t is a threshold for coordination. If cell i and node j are close to each other, the channel $g_{i,j}$ will be in general very large with high possibility and hence coordination might be required.
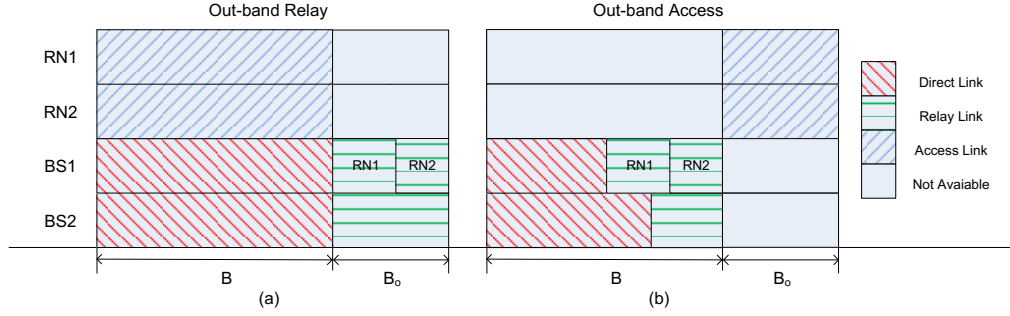
Figure 2.4.: Out-band relays resource utilization models.

This is important for the nomadic relaying network since several vehicles might be parking in the vicinity of each other. In particular, if i and j are the same RN and no isolation between the transmitter and receiver antennas is possible, we have $g_{i,j} \approx 1$ implying a high self-interference channel gain. In this case, we need to perform resource split (typically half-duplex) to put the RN in an operation mode without self-interference.

2.4.3. Outband Relaying and Outband Access Network

Due to the existence of multiple standards (GSM, UMTS, WiFi, etc) and the discussion on LTE unlicensed, secondary bandwidth for relay link or access link can be possible. Definitely, the cost due to the extra band should be taken into account when optimizing the nomadic relay network. In Fig 2.4, we illustrate two possibilities to incorporate the extra bandwidth into the nomadic relaying network.

Scheme (a), namely, the *Out-band Relay* mode, is depicted where an extra band b_o is dedicated for the relay link transmissions. In this case, the equivalent interference scenario is established as for the synchronized in-band case, except that more spectrum is available. On the other hand, the scheme (b) *Out-band Access* describes a network where the nomadic RNs have dedicated resources, e.g., WiFi or the future LTE unlicensed, for its access link transmissions. In this case, $\mathbf{S}$ is not enough to describe the interference relation in the network. However, we can decompose the system into two orthogonal networks, such that two independent interference relation matrices can be formulated as $\mathbf{S}^i = (\mathbf{S}^{(m,n)} \ \mathbf{S}^{(m,k)})$ and $\mathbf{S}^o = \mathbf{S}^{(k,n)}$. For both cases, we can use the generalized interference coordination scheme in (2.16) to avoid closely located nodes from interfering each other.

Without specifying the techniques for combating interference and ignoring the resource reservation for interference management, we make the following assumptions for interference coordination throughout the manuscript:

Assumption 2.10. $s_{i,j} = 0$ if cell i and node j are close and we choose a threshold $g_t = 100\ dB$. This includes the removal of the self-interference, where both i and j are the same node. This can be done by ignoring the additional resource partitioning or scheduling constraints to achieve the interference coordination scheme.

In particular, this assumption indicates that self-interference can be avoided without resource constraints. This can be realized by a full-duplex mode with interference cancellation or antenna isolation, out-band relaying operation, or a half-duplex relay mode by ignoring the backhaul link resource consumption.

2.5. Load Coupling Model

It follows from (2.8) that the bandwidth $b_{i,j} > 0$ which is needed at cell i to satisfy the rate requirement r_j of node j is equal to

$$b_{i,j} = \frac{r_i}{\omega_{i,j}}, \tag{2.17}$$

where the spectral efficiency is defined in (2.8). Then, we define the load induced by node j at cell i to be

$$\rho_{i,j} = \frac{b_{i,j}}{b_i} = \frac{r_j}{b_i \omega_{i,j}}, \tag{2.18}$$

where $b_i > 0$ is the total bandwidth at cell i. Now, we define $\boldsymbol{\rho} \in \mathbb{R}_+^{M+K}$, to be the vector of loads at the cells:

$$\boldsymbol{\rho} \triangleq \begin{bmatrix} \boldsymbol{\rho}^{(\mathrm{m})} \\ \boldsymbol{\rho}^{(\mathrm{k})} \end{bmatrix} = [\rho_1, \ldots \rho_M, \rho_{M+1}, \ldots, \rho_{M+K}]^T, \tag{2.19}$$

where the i-th entry of the load vector yields

$$
\begin{aligned}
\rho_i &= \rho_i^{(1)} + \rho_i^{(2)}, \quad i \in \mathcal{B} \bigcup \mathcal{R} \\
&= \sum_{j \in \mathcal{U}} \rho_{i,j}^{(1)} x_{i,j} + \sum_{k \in \mathcal{R}} \rho_{i,k}^{(2)} x_{i,k} \\
&= \sum_{j \in \mathcal{U}} \frac{r_j^{(\mathrm{n})}}{b_i \omega_{i,j}} x_{i,j} + \sum_{k \in \mathcal{R}} \frac{r_k^{(\mathrm{k})}}{b_i \omega_{i,k}} x_{i,k} \\
&= \underbrace{\sum_{j \in \mathcal{U}} \frac{r_j^{(\mathrm{n})}}{b_i \omega_{i,j}} x_{i,j}}_{\text{direct/access links}} + \underbrace{\sum_{k \in \mathcal{R}} \sum_{j \in \mathcal{U}} \frac{r_j^{(\mathrm{n})}}{b_i \omega_{i,k}} x_{i,k} x_{k,j}}_{\text{relay links}}.
\end{aligned}
\tag{2.20}
$$

Herein, $\rho_{i,j}^{(1)}$, $\rho_i^{(1)}$ and $\rho_{i,j}^{(2)}$, $\rho_i^{(2)}$ refer to the load corresponding to the UEs (direct/access links) and RNs (relay links), respectively. Note that the definition "load" can be also interpreted as activity level in [27], which is the fraction of necessary resources to satisfy the QoS of the connected nodes. It also expresses the possibility that the cell utilizes a certain block of time-frequency resource if no scheduling preference is specified. Using $\mathbf{B} = \left(\rho_{i,j} \right) \in \mathbb{R}_{++}^{(M+K) \times (N+K)}$ and $\mathbf{W} = \left(\frac{1}{b_i \omega_{i,j}} \right) \in \mathbb{R}_{++}^{(M+K) \times (N+K)}$, the load vector can be also expressed in a matrix form as follows

$$
\boldsymbol{\rho} = (\mathbf{B} \circ \mathbf{X}) \cdot \mathbf{1} = (\mathbf{W} \circ \mathbf{X}) \cdot \mathbf{r},
\tag{2.21}
$$

where $\mathbf{B} \triangleq \mathbf{B}(\boldsymbol{\rho}, \mathbf{p}, \mathbf{x})$ and $\mathbf{W} \triangleq \mathbf{W}(\boldsymbol{\rho}, \mathbf{p}, \mathbf{x})$ depend in general on $\boldsymbol{\rho}$, $\mathbf{p}$ and $\mathbf{x}$. Thus, the total load is determined by the function $\mathbf{F} = [F_1, \ldots, F_{M+K}] : \mathbb{R}^{M+K+\mathrm{L}} \to \mathbb{R}^{M+K}$ given by

$$
\boldsymbol{\rho} = \mathbf{F}(\boldsymbol{\rho}, \mathbf{p}, \mathbf{X}),
\tag{2.22}
$$

where $\mathbf{F}(\boldsymbol{\rho}, \mathbf{p}, \mathbf{X}) = \mathbf{F}^{(1)}(\boldsymbol{\rho}, \mathbf{p}, \mathbf{x}) + \mathbf{F}^{(2)}(\boldsymbol{\rho}, \mathbf{p}, \mathbf{x})$ with $\rho_i^{(1)} = F_i^{(1)}(\boldsymbol{\rho}, \mathbf{p}, \mathbf{x})$ and $\rho_i^{(2)} = F_i^{(2)}(\boldsymbol{\rho}, \mathbf{p}, \mathbf{x})$. Here and hereafter, we define

$$
\mathbf{x} \triangleq \mathrm{vec}(\mathbf{X}) = (\bar{x}_1, \ldots \bar{x}_{\mathrm{L}})^T,
\tag{2.23}
$$

where

$$
\mathrm{L} = (M + K) \cdot (N + K)
\tag{2.24}
$$

and

$$
x_{i,j} = \bar{x}_{(M+K) \cdot (j-1) + i}.
\tag{2.25}
$$

Similarly, $\mathbf{x}^{(\mathrm{m,n})}$, $\mathbf{x}^{(\mathrm{m,k})}$, $\mathbf{x}^{(\mathrm{k,n})}$ and $\mathbf{x}^{(\mathrm{k,k})}$ refer to vectorization of the corresponding block matrix. Furthermore, we use interchangeably $\mathbf{x}$ and $\mathbf{X}$ as the input argument of functions, such that, e.g., $\mathbf{F}(\mathbf{x})$ is equivalent to $\mathbf{F}(\mathbf{X})$. In the following, we distinguish two interference models, namely, the *static interference model* and the *dynamic interference model*.

2.5.1. Static Interference Model

If there exists no direct dependency between the current load of the cells $\boldsymbol{\rho}$ and the current spectral efficiency $\omega_{i,j}$ of link (i,j), the load is then determined only by the assignment matrix and the power vector. In such cases, we use a decoupled load model and write

$$\boldsymbol{\rho} = \mathbf{F}(\boldsymbol{\rho}', \mathbf{p}, \mathbf{x}) = \mathbf{F}(\mathbf{p}, \mathbf{x}), \tag{2.26}$$

where $\boldsymbol{\rho}'$ is a constant load vector that is not related to $\boldsymbol{\rho}$. Particularly, under the *worst-case interference* model we have $\boldsymbol{\rho}' = \mathbf{1}$. As previously mentioned, the worst-case interference model is adopted in several studies [17, 29, 35–37], resulting in a conservative estimation of the interference situation but fulfilling the minimum user QoS. The assumption is important if a small outage is required and low latency is targeted.

2.5.2. Dynamic Interference Model

Adopted in [18, 27, 31–33, 38–40], the dynamic interference model captures the statistical dependency between the interference and the current load of the network. The model is proven to be very accurate for LTE downlink in [41], since in a real system, a cell is not fully utilizing the resources in the power, time and frequency domain, therefore less interference is experienced by the users. In this case, we can have the load coupling model as in (2.22).

2.5.3. Load Constraints

In a real system, a load larger than 1 (overload) means that more resources are needed than the available amount at a cell to satisfy the minimum QoS of the attached nodes. Therefore, the real load in a practical system is given by

$$\bar{\boldsymbol{\rho}} = \min(\boldsymbol{\rho}, \mathbf{1}) = \min(\mathbf{F}(\boldsymbol{\rho}, \mathbf{p}, \mathbf{x}), \mathbf{1}) \tag{2.27}$$

where the "min" operation is taken component-wise. If overload happens, user satisfaction cannot be fully guaranteed. Therefore, in order to avoid such situations, the following condition should be satisfied:

$$\boldsymbol{\rho} = \mathbf{F}(\boldsymbol{\rho}, \mathbf{p}, \mathbf{x}) < \mathbf{1}. \tag{2.28}$$

Along with (2.7), (2.28) expresses the fundamental QoS constraints for performing network optimization.

Chapter 3.

Problem Definition and Literature Review

Based on the model and the constraints discussed in Chapter 2, a generic optimization framework for nomadic relaying networks is proposed at the beginning of this chapter. Subsequently, the objective function of the optimization framework is formulated as the sum of the total network energy consumption. To this end, the energy efficiency fundamentals are examined and the energy consumption models from the related works are adopted. Finally, the energy-saving optimization problem is devided into the two sub-problems, and the corresponding state-of-the-art solutions are discussed.

3.1. A Generic Optimization Framework

In Chapter 2, the nomadic relaying network is modeled as a network controlled by the cell transmission power vector $\mathbf{p}$ and the assignment matrix $\mathbf{x}$. Proper values of $\mathbf{p}$ and $\mathbf{x}$ should be chosen, so that no overload occurs in any cell. This is to say that (2.28) must be always satisfied. On the other hand, the minimum rate requirements of all UEs should be supported by the network as in (2.6) or (2.7), whereas the transmission power is upper-bounded by (2.11). From these points, we can formulate a generic optimization problem in a nomadic relaying network as follows:

$$\min_{\mathbf{x},\mathbf{p}} \quad \mathrm{U}(\boldsymbol{\rho}, \mathbf{x}.\mathbf{p}) \tag{3.1a}$$

$$\text{subject to} \quad \mathbf{X}^T \cdot \mathbf{1} \geq \mathbf{1}, \quad \mathbf{x} \geq \mathbf{0} \tag{3.1b}$$

$$\mathbf{0} \leq \mathbf{p} \leq \hat{\mathbf{p}} \tag{3.1c}$$

$$\boldsymbol{\rho} = \boldsymbol{\rho}(\mathbf{x}, \mathbf{p}) \leq \mathbf{1}. \tag{3.1d}$$

Herein, $\boldsymbol{\rho}(\mathbf{x}, \mathbf{p})$ is the load induced by the assignments $\mathbf{x}$ and power vector $\mathbf{p}$, whereas $\mathrm{U}(\boldsymbol{\rho}, \mathbf{x}.\mathbf{p})$ is a cost function that captures the coverage, capacity, load balancing, energy saving or other network design objectives. For instance:

- if the total network rate throughput is considered, the objective is to maximize the sum of the rates, which is equivalent to minimize $\mathrm{U}(\boldsymbol{\rho}, \mathbf{x}.\mathbf{p}) = -\mathbf{r}^T \cdot (\mathbf{X}^T \cdot \mathbf{1})$;

- if load balancing is the optimization objective, we can minimize the maximal load in the network, in which case we have $\mathrm{U}(\boldsymbol{\rho}, \mathbf{x}.\mathbf{p}) = |\boldsymbol{\rho}|_\infty$.

In this work, we focus on the energy saving problem. In the following, we review the energy consumption models in order to formulate the objective function of minimizing total network energy consumption.

3.2. Energy Saving Problem

Power consumption and carbon emissions are becoming an eminent problem for Information and Communication Technology (ICT) systems, especially due to the radio access network of the cellular systems [42–44]. In particular, the increasing electricity costs and the large amount of energy consumption at the BSs significantly increase the Operational Expenditure (OPEX) of the operators ($3000/$30000 per BS per year for on/off-grid BSs [42]). Adding the fact that a large number of BS sites are required for covering the expanding metropolitan regions, a substantial amount of energy is needed for delivering data and services through cellular systems. On the other hand, parking vehicles do not have a power supply, meaning that high energy efficiency becomes one of the main requirements in the radio system design at the RN side. Therefore, the energy-saving problem is a key issue when designing optimization algorithms for nomadic relaying networks.

3.2.1. Energy Efficiency: From Theory to Practice

In the field of ICT, the fundamental theory that connects energy and information is Shannon's theory of information [30]. From the link-level perspective, such as in (2.8), the amount of information that can be transmitted reliably per channel use increases *logarithmically* with respect to the transmission power, whereas it increases *linearly* with respect to the transmission

bandwidth. Therefore, the theoretical optimal power allocation yields when the whole bandwidth is fully utilized. This means that a higher transmission power implies a higher spectrum efficiency but also a lower energy efficiency. From the network perspective, network deployments with smaller cell radii significantly decrease the pathloss and hence reduce the demand on the total cell transmission power. A large number of research papers have proven the energy efficiency in terms of radio transmissions in small cell deployments [45–47]. The presumption of both link- and network-level theories is, however, that no extra expenditures for hardware energy consumption and deployment cost exist. Analyses of energy efficiency that take into account such practical factors are given in [46, 48], where the authors highlight the importance of including the segment of energy consumption that is not directly related to the transmission power that radiated from the transmission antenna to the receiver antenna. The fundamental trade-offs between energy efficiency and spectral efficiency have been well explained in [49] with hardware and deployment considerations and in [50] with economic analyses, where the optimal energy efficiency is achieved at a watershed by jointly considering hardware/site cost and transmission energy. Therefore, the energy modeling must take into account the practical constraints on hardware and deployment costs in the cellular systems.

3.2.2. Energy Metric: Static and Dynamic Energy

The common metric for energy efficiency is Energy Consumption Ratio (ECR) which is defined as the ratio of the total energy consumption to the delivered bits of information. Since we assume a fixed QoS model and consider the downlink scenario, the only metric in this work is the total energy consumption of BSs and RNs in the network. The energy consumption models of different types of BSs have been intensively studied in literature [51–53]. Breakdown analyses have been carried out for different types of cells, establishing a model that contains both static energy consumption (constant) and dynamic energy consumption (varying based on load and power). The static energy consumption is due to the power supply, server operations, cooling system, and so on, and it is almost constant at a cell site. The dynamic part, on the other hand, comes from hardware components such as Digital Signal Processing (DSP) and Power Amplifier (PA) which depend on the transmission power and load of the system and can be significant in a heavily loaded cell with a large transmission power. Due to the low efficiency of PA at low load situations, conventional BSs consume almost the fixed amount of dynamic energy independent of load. Innovations and optimizations on hardware and software enable a scalable energy consumption that fits to the traffic load of the network to achieve dynamic energy savings [54]. Summarizing all the models given in those works, we establish a generic

formulation of the energy consumption of a given cell as

$$
\mathrm{U}_{cell}(p, \rho) = \underbrace{c \cdot ||p \cdot \rho||_0}_{\text{static energy}} + \underbrace{d \cdot ||p \cdot \rho||_1}_{\text{dynamic energy}} .
\tag{3.2}
$$

Herein, l_0-norm and l_1-norm represent, respectively, the static and the dynamic energy consumption, whereby ρ and p are the current load and transmission power of the cell, respectively. Note that c depends on the type of the cell, and in particular, on the transmission power of the cell. Furthermore, $||p \cdot \rho||_1$ is the total output power of the cell, while $d||p \cdot \rho||_1$ is the total input power of the cell. Therefore, $d > 1$ can be understood as the inverse of the dynamic power efficiency of the cell. This model fits the different models in the literature (e.g., [54, Fig. 9] and [43, Fig. 1]), and suitable parameters of c and d need to be selected to reflect different types of cells. Based on this general energy model, we can formulate the objective function of our optimization framework as

$$
\mathrm{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}, \mathbf{x}, \mathbf{p}) := \sum_{i \in \mathcal{B} \bigcup \mathcal{R}} c_i ||p_i \rho_i||_0 + d_i ||p_i \rho_i||_1,
\tag{3.3}
$$

where the vector $\mathbf{c} = (c_1, \ldots, c_{M+K}) \in \mathbf{R}_+^{(M+K) \times 1}$ and $\mathbf{d} = (d_1, \ldots, d_{M+K}) \in \mathbf{R}_+^{(M+K) \times 1}$ are coefficients for the static and the dynamic energy consumption of active cells, respectively.

3.2.3. Energy Saving: Mechanisms and Algorithms

The energy-saving opportunities at different layers of the cellular networks are extensively investigated and overviewed in several survey papers [42–44, 54–57]. The papers review the energy saving techniques from different perspectives, where the main opportunities for energy savings lie in hardware enhancements, deployment strategies, transmission and RRM schemes, etc. In [42], advanced radio energy-saving transmission and RRM techniques are reviewed at both the cell level and the network level. Authors in [43] emphasize the importance of energy-efficient hardware components with extensive reviews on energy-saving hardware techniques. In [44] and [55], load-dependent RRM, small cell deployment and MIMO-OFDM optimizations are seen as the key directions for realizing an energy-efficient network. With the focus on the small cell deployment, [54] discusses the potential techniques to adapt the energy consumption to the daily varying load pattern, whereas the sleep mode mechanisms of small cell are investigated in [56], to further increase the energy efficiency of small cell deployment. In [57], RRM schemes to achieve high energy efficiency are reviewed, including power allocation, interference management and routing strategies, showing another direction for future green radio systems.

We summarize the energy-saving techniques from mathematical perspective, according to the model in (3.2) and (3.3). Firstly, in order to reduce static energy consumption, either energy-efficient hardware components need to be installed or operational switching-off mechanisms should be performed. Hardware efficiency, especially the power amplifier, impacts significantly on the total energy consumption in a cell. Therefore, suggestions for improvement that should be paid great attention, are, e.g., upgrading PA material to Aluminum Gallium nitride (in [44] and the references therein), locating PA near the antenna ([57]) and reducing the crest factor for higher PA linearity ([43]). Addressing the varying QoS profiles with respect to the time or space domains [51–53], operational switching-off of components, including site shutdown, sleeping mode and discontinuous transmission (DTX) mechanisms, directly reduces the static energy consumption. While site shut-down completely closes the cell and implies a zero energy consumption, sleep mode is the shutdown of some energy-consuming components in the cell, providing the ability of being active again quickly to cater to the fast load fluctuation. DTX can be understood as the short-term and partial sleep of a cell, and it enables a higher granularity for saving the static energy consumption. The key motivation behind the switching-off mechanisms is the redundancy of cells and radio resources, especially in a dense heterogeneous deployment where small cell coverage is overlapping with the macro coverage. The framework for switching-off components is the Dynamic Power Management (DPM) [58], in which different components are observed and can be turned-off on demand. While a DPM framework is required to passively monitor the change in the network for opportunistic or statistical switching-off, user assignments and load balancing algorithms, which are presented in Subsections 3.3 and 3.4, are the key enablers for proactively switching-off nodes in order to reduce the consumption of static energy.

It is worth noting that hardware upgrades also significantly increase the dynamic energy efficiency (impact on d). On the other hand, the total output power, which equals $p \cdot \rho$, can be reduced through efficient RRM schemes. As mentioned in the trade-off study [49], an efficient power and bandwidth allocation optimizes the power consumption for radio transmission. Furthermore, Shannon's equation also tells us that noise, interference and path loss are the only facts due to which the radiated energy is not fully or efficiently delivered to the receiver as information. Therefore, it is intuitive that the small cell deployment requires short path transmission and hence lowers the pathloss in the air so as to improve the dynamic energy efficiency. However, for a dense small cell deployment, although sleep mode could compensate the extra static energy consumption, the associated site cost cannot be eliminated. Besides small cell deployment, transmission schemes that exploit time and space diversity also reduce the energy loss in the air and hence increase the energy efficiency. Cooperative communications (MIMO, Coordinated Multipoint Transmission and Reception (CoMP), network coding, etc.), which compensate the
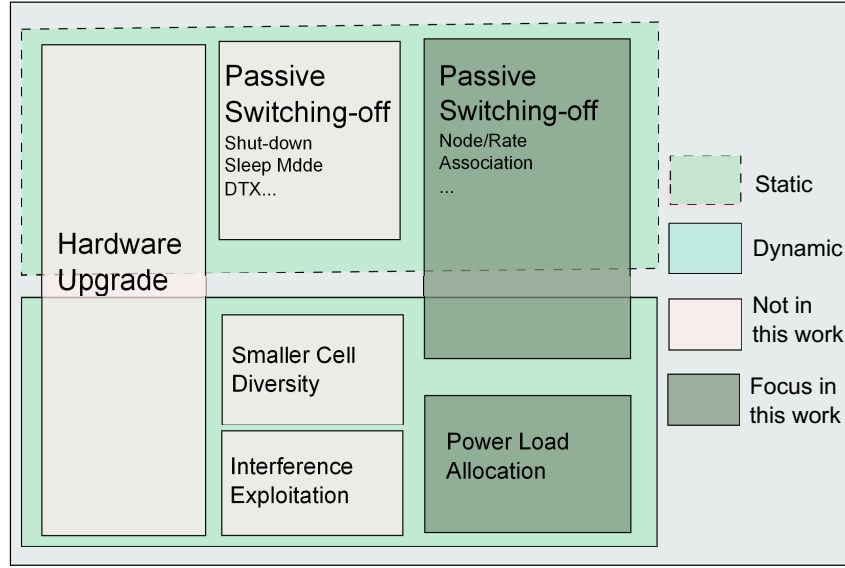
Figure 3.1.: Energy saving mechanisms.

energy loss, can significantly increase the spectrum efficiency without increasing the transmission power, also resulting in a lower dynamic energy consumption.

A graphical depiction of the energy-saving mechanisms is given in Fig 3.1, where the focus of this work is marked in a dark color. Considering our optimization framework, we divide our problem into two decoupled sub-problems and classify the algorithms into the two categories:

- optimization over the assignment variables $\mathbf{x}$, which addresses the problem of the activation and deactivation of nomadic nodes based on energy-saving demands;

- optimization of transmission powers $\mathbf{p}$, which aims at the optimal dynamic energy and the transient load balancing during the activation procedure of nomadic nodes.

3.3. Nomadic Nodes Activation Problem

Due to the large number of RNs in the network, one of the major tasks in the nomadic network is to select a subset of the RNs for a given network optimization objective. The admission of additional network nodes brings data-access possibilities, as well as sources of interference. Hence, activation and deactivation of network elements should be carefully considered in a demand-driven fashion. The node activation problem can be reformulated as a two-step assignment problem, i.e., Relay Selection (RS) and User Association (UA). Whereas RS refers to the selection and subsequent assignment of RNs to BSs, UA is the assignment of UEs either to RNs

(two-hop relaying) or to BSs (direct communications). Further, a rate assignment interpretation yields if, instead of assigning a node to a cell, the rate of a node is split and a portion of the rate is assigned to a cell. We adapt in this work the latter case according to Assumption 2.6.

3.3.1. Problem Formulation

In a cell activation problem, we assume a fixed transmission power of all nodes, i.e., $\mathbf{p} = \tilde{\mathbf{p}}$. This yields a special instance of the objective function in (3.3):

$$\min_{\mathbf{X}} \quad U_{\boldsymbol{\rho}}(\boldsymbol{\rho})|_{\mathbf{p}=\tilde{\mathbf{p}}} = \sum_{i=1}^{M+K} c_i ||\tilde{p}_i \rho_i||_0 + d_i ||\tilde{p}_i \rho_i||_1 \tag{3.4a}$$

$$\text{subject to} \quad \mathbf{X}^T \cdot \mathbf{1} \geq \mathbf{1}, \tag{3.4b}$$

$$\boldsymbol{\rho} \leq \mathbf{1}, \tag{3.4c}$$

$$\mathbf{x} \geq \mathbf{0}. \tag{3.4d}$$

The problem is in general discontinuous, due to the l_0-norm. Note that the problem can be seen as a general load-balancing problem over the assignments. Furthermore, the load function in (2.26) is proven to be non-convex with respect to the assignment $\mathbf{x}$ in [17], which requires further relaxations. In Chapter 4 and Chapter 5, we elaborate on the fundamental properties of the optimization problem and present solutions by relaxation and reformulation techniques. In the following, we give an overview on the related works on assignment optimizations for load-balancing and energy-saving purposes.

3.3.2. State of the Art

IThe optimization of UA in a single-hop network has been intensively studied in the context of network planning [27, 29, 35–40, 59–65]. Based on the static interference model, the UA problem is investigated for load balancing or energy-saving purposes in [17, 29, 35–37, 59, 60]. In [59], off-line optimization and on-line handover algorithms are proposed for load balancing, where a proportional fair utility is targeted. The paper jointly optimizes fractional reuse and user association such that both load balancing and interference avoidance gains can be achieved. Later in [60], the authors study the user association policies in heterogeneous networks, proposing practical schemes such as range expansion, resource negotiation and dynamic interference management. The optimal user assignment scheme in [36] is proposed in order to optimize a

generic α-utility function in a distributed manner, through which the performance on delay, throughput and max-load can be optimized.

The above-mentioned optimization objectives are all convex functions of load. However, the energy consumption is even not continuous, meaning that convex optimization methods cannot be directly applied. The energy-saving problem is addressed in [37] by a two-step algorithm, where a greedy algorithm is applied to find a BS operation mode and a Linear Program (LP) based UA optimization is used to further reduce the dynamic energy. More proactive algorithms for energy saving that switch off BS are given in [29] for single-Radio Access Technology (RAT) and in [35] for multi-RAT networks, where the l_0-norm is approximated by a concave function that can be minimized by a Majorization Minimization (MM)-algorithm. The UA optimization is further discussed in [27, 38–40, 61–65] under a dynamic interference model that has been explained in Chapter 2. In [61], a coupled queue model is applied in order to find the optimal load allocation between two BSs. In [27, 62], a flow-level traffic model is introduced and a UA algorithm is given jointly with antenna tilts optimization for load balancing in the framework of SON, where the load coupling is iteratively decoupled as the static interference model. In [39, 40, 63], the authors fundamentally analyze the load coupling function and propose practical algorithms for network planning in both legacy and heterogeneous networks. The dynamic model is also adopted in [64] for utility maximization in the presence of complementary networks for off-loading. The load estimation problem under the dynamic interference model is discussed in [38], where an energy-saving algorithm is given based on [29, 35] (considering both static and dynamic energy consumptions). Moreover, the energy-saving optimization is modeled as a load-minimization problem under the dynamic interference model in [65], where only the dynamic energy consumption is considered.

In the context of multi-hop networks, the topic of RS has been theoretically elaborated in [66], where the approach is based on link-level metrics rather than a network-wide performance enhancement. Therein, two selection criteria (max-min and harmonic mean of the two hops) are proposed. Relay-based network optimization algorithms are proposed for load balancing and throughput optimization, as well as for energy savings, in [67–74]. In [67], a joint routing and link-scheduling algorithm is proposed in order to maximize the overall throughput, where no reuse among the links is considered. Algorithms for further exploiting the spatial reuse of the radio resources among the access links and between relay and access links are proposed in [68, 69], whereas the resource split between relay links and access links are further optimized jointly with link-schedule algorithm in [72], achieving significant throughput improvements. In [73], a distributed relay reassociation scheme is proposed, and simulation shows a reducing call

block rate. The relay-aided energy-saving problem is addressed in [70, 71], where the RNs are used to redirect data in order to enable more BSs to enter sleep mode. In [74], a generalized assignment problem is formulated and a tabu search approach is given in order to find the optimal BSs modes for static energy savings. In all these works, fixed relay deployment and fixed user assignment policy are considered. Further, only static interference model is used to model the spectral efficiency. In our works [17–19] and in this thesis, we focus on the joint RS-UA problem for energy savings in a multi-cell nomadic relaying network with both on-line and off-line algorithms.

3.4. Power Control Problem

After finding the optimal assignments, some nomadic nodes need to be activated by performing power ramping in order to serve the UEs according to the optimal assignments. In this situation, further optimization based on the transmission power can be done to reduce dynamic energy consumption. Another problem occurs when a nomadic node is entering the network, since the additional interference may significantly deteriorate the SE of the UEs in the vicinity. More resources are then needed to satisfy the affected UEs, and this may lead to an overloaded serving cell. Therefore, suitable power-control mechanisms need to be employed to keep the cells from being overloaded during the transient phases.

3.4.1. Problem Formulation

Fixing the assignments and assuming the same d_i for every cell, the energy-saving problem reduces to an l_1-norm minimization problem given by

$$\min_{\mathbf{p}} \quad \mathbf{p}^T \cdot \boldsymbol{\rho} \tag{3.5a}$$

$$\text{subject to} \quad \mathbf{0} \leq \mathbf{p} \leq \hat{\mathbf{p}} \tag{3.5b}$$

$$\boldsymbol{\rho} \leq \mathbf{1} \tag{3.5c}$$

The difficulty in solving the problem above lies in the complexity of the coupling between power vector $\mathbf{p}$ and the load vector $\boldsymbol{\rho}$. Another challenge is to develop a distributed iterative power control algorithm, under which the transient performance can be guaranteed.

Denoting (t) to be the index of an iteration step, the iterative algorithm can be described as

$$\mathbf{p}(t+1) = \mathbf{T}(\mathbf{p}(t), \boldsymbol{\rho}(t)) \tag{3.6a}$$

$$\mathbf{0} \leq \mathbf{p}(t+1) \leq \hat{\mathbf{p}}, \tag{3.6b}$$

$$\boldsymbol{\rho}(t+1) \leq \mathbf{1}, \ \text{if} \ \boldsymbol{\rho}(t) \leq \mathbf{1}, \tag{3.6c}$$

where $\mathbf{T} : \mathbb{R}_+^{M+K} \rightarrow \mathbb{R}_+^{M+K}$ is a map that determines the update algorithm for iteratively adjusting the transmission powers. The algorithm should converge to the optimal power vector $\mathbf{p}^*$, i.e., $\mathbf{p}(t+1)|_{t\rightarrow\infty} = \mathbf{p}^*$, such that $\mathbf{p}^*$ solves Problem 3.5. In the next subsection, we review the previous works in this area.

3.4.2. State of the Art

Power control is one of the fundamental mechanisms for resource allocation in wireless communication systems. Early works on power control have focused on centralized power allocation for balancing Signal to Interference Ratio (SIR), such that the minimum SIR of all links is maximized [75]. For practical implementation issues, Distributed Power Control (DPC) has been extensively studied in the context of single-carrier networks including the uplink channel and the distributed wireless mesh networks [76–82]. Based on a noiseless power control scheme in [76], the authors propose in [77] an iterative DPC algorithm that converges to an optimal power vector where a linear interference plus noise model is considered. In both papers, the total transmission power is minimized by performing the distributed power control algorithms. The idea is further extended in [78] by adding Active Link Protection (ALP) such that the QoS for the users do not drop below the requirements during the transient phase. The energy-robustness trade-off of ALP/DPC is discussed in [79], where the authors propose an algorithm, denoted as the Robust Distributed Power Control (RDPC), to dynamically adjust the control parameter. It also shows, from the optimization perspective that the algorithm optimizes a compromise between the total power and some indicators for robustness. Furthermore, the ALP/DPC framework has been extended in [80, 81] in the context of Standard Interference Functions (SIFs) and in [82] in the framework of General Interference Functions (GIFs), which reflects the case of zero noise interference. Incorporating the load coupling model, the authors in [83] prove that full load is optimal considering the dynamic energy consumption and develop an algorithm that minimizes the total transmission energy. In this work, we develop an Active Cell Protection (ACP) power control algorithm to optimize the total dynamic energy for the nomadic network.

Chapter 4.

Activation of Nomadic Nodes

Based on [17, 18], this chapter focuses on the optimization problem in (3.4) which is a user assignment and relay selection problem. We first analyze the fundamental properties of the load coupling function in (2.22) under both the static and the dynamic interference models. Then, we propose different approximation, relaxation and reformulation techniques to enable efficient centralized off-line algorithms. The proposed algorithms are intensively evaluated under different network and relaying configurations.

4.1. Properties of the Energy Saving Optimization

The complexity of Problem (3.4) lies mainly in the dependency between the load vector $\boldsymbol{\rho}$ and the assignment matrix $\mathbf{x}$, in other word, in the convexity of the feasible region for the optimization problem. Furthermore, the l_0-norm implies a discontinuous relation between the objective value $\mathrm{U}_{\boldsymbol{\rho}}$ and the load vector $\boldsymbol{\rho}$, which is in turn induced by the optimization variable $\mathbf{x}$ according to the load function. Therefore, we investigate the fundamental properties of the load function and the objective function by verifying the convexity, continuity and further properties with respect to the optimization variable $\mathbf{x}$. Before starting discussing the properties, we list some definitions to support our analysis.

Definition 4.1 (Monotonicity)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is called monotonically increasing if $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ such that $\mathbf{x}_1 \geq \mathbf{x}_2$, we have $f(\mathbf{x}_1) \geq f(\mathbf{x}_2)$.

Lemma 4.1. Let $\mathbf{x} \in \mathbb{R}_+^n$ for arbitrary $n \geq 1$. The monotonicity property is fulfilled, if $f : \mathbb{R}_+^n \to \mathbb{R}_{++}$ is continuously differentiable over $\mathbf{x} \in \mathbb{R}_{++}^n$ with only non-negative gradients: $\mathbf{J}_f^{\mathbf{x}}(\mathbf{x}) \geq \mathbf{0}$ for all $\mathbf{x} \in \mathbb{R}_{++}^n$.

Proof. Let $\mathbf{z}, \mathbf{z}' \in \mathbb{R}_{++}^n$ be arbitrary and, without loss of generality, assume that $\mathbf{z} \le \mathbf{z}'$. Now let $\mathbf{z}^{(i)} = (0, \ldots, 0, z_i' - z_i, 0, \ldots, 0)$, $1 \le i \le n$, be a vector with zeros everywhere except for the i-th position which is equal to $z_i' - z_i \ge 0$. Since $\mathbf{J}_f^{\mathbf{x}}(\mathbf{x}) \ge \mathbf{0}$ for all $\mathbf{x} \in \mathbb{R}_{++}^n$, we have

$$f(\mathbf{z}) \le f(\mathbf{z} + \mathbf{z}^{(1)}) \le f(\mathbf{z} + \mathbf{z}^{(1)} + \mathbf{z}^{(2)}) \le \ldots \le f(\mathbf{z} + \textstyle\sum_{i=1,\ldots,n} \mathbf{z}^{(i)}) = f(\mathbf{z}'). \qquad \square$$

Definition 4.2 (Monotonicity)**.** A vector-valued function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ is called element-wise monotonically increasing if $\forall i \in 1, \ldots, m$, F_i is monotonically increasing over $\mathbf{x} \in \mathbb{R}_{++}^n$.

Definition 4.3 (Scalability)**.** A function $f : \mathbb{R}_+^n \to \mathbb{R}$ is called scalable (or inverse scalable) if $f(\alpha \mathbf{x}) < \alpha f(\mathbf{x})$ (or $f(\alpha \mathbf{x}) > \alpha f(\mathbf{x})$) for all $\mathbf{x} \in \mathbb{R}_+^n$ and all $\alpha > 1$.

Definition 4.4 (Scalability)**.** A vector-valued function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ is called element-wise scalable (or inverse scalable) if $\forall i \in 1, \ldots, m$, F_i is scalable (or inverse scalable) over $\mathbf{x} \in \mathbb{R}_{++}^n$.

Definition 4.5 (SIF [38, 81])**.** A function $f : \mathbb{R}_+^n \to \mathbb{R}_{++}$ is called an SIF, if it is scalable and monotonically increasing.

Based on these definitions, we provide in the following some basic properties of the load function defined in (2.22) or (2.26). For brevity, we omit the power vector $\mathbf{p}$ in the notation of the load function since constant transmission powers are assumed at every cell in this chapter.

4.1.1. Load Function under the Static Interference Model

It is shown in [17] that the load constraint is not convex under the worst-case interference model, and therefore certain relaxations are necessary for heuristically solving the problem. Here we give more general conclusions for the load function and the corresponding constraints. First, we reformulate the constraints in (3.4b) and (3.4c) into the standard form:

$$\mathbf{p}_i^T \mathbf{x} \ge 1, \quad \text{for } i \in \{1, \ldots, N + K\} \tag{4.1a}$$

$$F_i(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{Q}_i \mathbf{x} + \mathbf{q}_i^T \mathbf{x} \le 1, \quad \text{for } i \in \{1, \ldots, M + K\}, \tag{4.1b}$$

where the details of $\mathbf{p}_i$, $\mathbf{Q}_i$ and $\mathbf{q}_i$, which are constant matrices in case of the static interference model, are derived and given in (4.2), (4.9) and (4.10), respectively.

First, it is obvious from (A.4) in the Appendix that the vectorization of the left hand side of (3.4b) can be written as $\mathbf{X}^T \mathbf{1} = (\mathbf{1}^T \mathbf{X})^T = (\mathbf{I} \otimes \mathbf{1}^T)\mathbf{x}$. Let $\mathbf{e}_i^T$ denote the i-th row of the identity matrix of a proper size for matrix operations, then the i-th row of $(\mathbf{I} \otimes \mathbf{1}^T)\mathbf{x}$ is $\mathbf{e}_i^T(\mathbf{I} \otimes \mathbf{1}^T)\mathbf{x}$.

Therefore, it can be concluded that

$$\mathbf{p}_i = \mathbf{e}_i^T (\mathbf{I} \otimes \mathbf{1}^T). \tag{4.2}$$

According to (2.21), the load function can be formulated as:

$$
\begin{aligned}
\mathbf{F}(\mathbf{x}) &= (\mathbf{B} \circ \mathbf{X}) \cdot \mathbf{1} = (\mathbf{W} \circ \mathbf{X}) \cdot \mathbf{r} = (\mathbf{W} \circ \mathbf{X}) \begin{pmatrix} \mathbf{r}^{(n)} \\ \mathbf{r}^{(k)} \end{pmatrix} \\
&= \begin{bmatrix} \underbrace{(\mathbf{W}^{(m,n)} \circ \mathbf{X}^{(m,n)}) \mathbf{r}^{(n)}}_{\lambda^{(m,n)}} + \underbrace{(\mathbf{W}^{(m,k)} \circ \mathbf{X}^{(m,k)}) \mathbf{r}^{(k)}}_{\lambda^{(m,k)}} \\ \underbrace{(\mathbf{W}^{(k,n)} \circ \mathbf{X}^{(k,n)}) \mathbf{r}^{(n)}}_{\lambda^{(k,n)}} \end{bmatrix},
\end{aligned} \tag{4.3}
$$

where the three blocks $\lambda^{(m,n)}$, $\lambda^{(k,n)}$ and $\lambda^{(m,k)}$ can be further written with the help of (A.3) and (A.5) - (A.8) in the Appendix as:

$$\lambda^{(m,n)} \overset{(A.6)}{=\joinrel=} ([\mathbf{r}^{(n)}]^T \otimes \mathbf{I}^M) \mathrm{diag}\left(\mathbf{W}^{(m,n)}\right) \mathbf{x}^{(m,n)}, \tag{4.4a}$$

$$\lambda^{(k,n)} \overset{(A.6)}{=\joinrel=} ([\mathbf{r}^{(n)}]^T \otimes \mathbf{I}^K) \mathrm{diag}\left(\mathbf{W}^{(k,n)}\right) \mathbf{x}^{(k,n)}, \tag{4.4b}$$

$$
\begin{aligned}
\lambda^{(m,k)} &\overset{(A.8)}{=\joinrel=} \left(\mathbf{W}^{(m,k)} \circ \mathbf{X}^{(m,k)} \circ (\mathbf{1}^M \otimes [\mathbf{r}^{(k)}]^T)\right) \mathbf{1}^K \\
&\overset{(A.5)}{=\joinrel=} \underbrace{([\mathbf{1}^K]^T \otimes \mathbf{I}^M) \cdot \mathrm{diag}\left(\mathbf{W}^{(m,k)}\right)}_{\mathbf{D}_1} \cdot \left(\mathrm{vec}(\mathbf{1}^M \otimes [\mathbf{r}^{(k)}]^T) \circ \mathbf{x}^{(m,k)}\right) \\
&= \mathbf{D}_1\left(\mathrm{vec}(\mathbf{1}^M [\mathbf{r}^{(n)}]^T [\mathbf{X}^{(k,n)}]^T) \circ \mathbf{x}^{(m,k)}\right) \\
&\overset{(A.3)}{=\joinrel=} \mathbf{D}_1\left(\left((\mathbf{I}^K \otimes (\mathbf{1}^M [\mathbf{r}^{(n)}]^T)) \cdot \mathrm{vec}([\mathbf{X}^{(k,n)}]^T) \right) \circ \mathbf{x}^{(m,k)}\right) \\
&= \mathbf{D}_1\left(\underbrace{(\mathbf{I}^K \otimes (\mathbf{1}^M [\mathbf{r}^{(n)}]^T) \mathbf{\Pi}_0^T}_{\mathbf{D}_2} \cdot \mathbf{x}^{(k,n)}) \circ \mathbf{x}^{(m,k)}\right).
\end{aligned} \tag{4.4c}
$$

In (4.4c), $\mathbf{\Pi}_0$ is a permutation matrix of length of $K \times N$ such that $[\mathbf{x}^{(k,n)}]^T \mathbf{\Pi}_0 = \mathrm{vec}([\mathbf{X}^{(k,n)}]^T)^T$. Given this, the i-th row of $\lambda^{(m,k)}$ can be further written according to (A.9) as:

$$
\begin{aligned}
\mathbf{e}_i^T \cdot \lambda^{(m,k)} &= \mathbf{e}_i^T \mathbf{D}_1\left((\mathbf{D}_2 \cdot \mathbf{x}^{(k,n)}) \circ \mathbf{x}^{(m,k)}\right) \\
&= \left((\mathbf{e}_i^T \mathbf{D}_1) \circ (\mathbf{D}_2 \cdot \mathbf{x}^{(k,n)}) \circ \mathbf{x}^{(m,k)}\right) \cdot \mathbf{1} \\
&= [\mathbf{x}^{(k,n)}]^T \mathbf{D}_2^T \mathrm{diag}\left(\mathbf{e}_i^T \mathbf{D}_1\right) \mathbf{x}^{(m,k)}.
\end{aligned} \tag{4.5}
$$

Now, let $\mathbf{\Pi}_a \in \mathbb{R}^{N \times (M+K)}$ and $\mathbf{\Pi}_b \in \mathbb{R}^{K \times (M+K)}$ be permutation matrices such that

$$\mathbf{x} = \begin{pmatrix} \mathbf{\Pi}_a \cdot \begin{pmatrix} \mathbf{x}^{(m,n)} \\ \mathbf{x}^{(k,n)} \end{pmatrix} \\ \mathbf{\Pi}_b \cdot \begin{pmatrix} \mathbf{x}^{(m,k)} \\ \mathbf{0} \end{pmatrix} \end{pmatrix}. \tag{4.6}$$

Considering this, we can rewrite the i-th row of $\lambda^{(m,k)}$ as

$$\mathbf{e}_i^T \cdot \lambda^{(m,k)} = \frac{1}{2} \cdot \mathbf{x}^T \cdot \begin{pmatrix} \mathbf{0} & \mathbf{T}_i \\ \mathbf{T}_i^T & \mathbf{0}, \end{pmatrix} \cdot \mathbf{x} \tag{4.7}$$

with

$$\mathbf{T}_i = \mathbf{\Pi}_a \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{D}_2^T \mathrm{diag}\left(\mathbf{e}_i^T \mathbf{D}_1\right) & \mathbf{0} \end{pmatrix} \mathbf{\Pi}_b^T. \tag{4.8}$$

Thus, by comparing (4.4a), (4.4b) and (4.4c) with (4.1b), we finally obtain

$$\mathbf{Q}_i = \begin{cases} \begin{pmatrix} \mathbf{0} & \mathbf{T}_i \\ \mathbf{T}_i^T & \mathbf{0} \end{pmatrix}, & \text{for } i \in \mathcal{B}, \\ \mathbf{0}, & \text{for } i \in \mathcal{R}. \end{cases} \tag{4.9}$$

and

$$\mathbf{q}_i = \begin{cases} \mathbf{e}_i^T \cdot ([\mathbf{r}^{(n)}]^T \otimes \mathbf{I}^M) \mathrm{diag}\left(\mathbf{W}^{(m,n)}\right), & \text{for } i \in \mathcal{B}, \\ \mathbf{e}_i^T \cdot ([\mathbf{r}^{(n)}]^T \otimes \mathbf{I}^K) \mathrm{diag}\left(\mathbf{W}^{(k,n)}\right), & \text{for } i \in \mathcal{R}. \end{cases} \tag{4.10}$$

Lemma 4.2. For the static interference model, the load function $\rho = \mathbf{F}(\mathbf{x})$ defined in (2.26) is element-wise *monotonically increasing* in $\mathbf{x}$.

Proof. Considering (4.1b), it is obvious that the matrices $\mathbf{Q}_i$ and $\mathbf{q}_i$ only have non-negative elements, since all the component blocks are non-negative. Then, $\mathbf{J}_{\mathbf{F}_i}^{\mathbf{x}}(\mathbf{x}) = \mathbf{Q}_i \mathbf{x} + \mathbf{q}_i \geq \mathbf{0}$, $\forall \mathbf{x} \geq \mathbf{0}$. Therefore, Lemma 4.1 is fulfilled which completes the proof. □

Lemma 4.3. For the static interference model, the load function $\rho = \mathbf{F}(\mathbf{x})$ as in (2.26) is element-wise *inverse scalable* for $\mathbf{x} > \mathbf{0}$.

Proof. It suffices to prove that i-th component of $\boldsymbol{\rho} = \mathbf{F}(\mathbf{x})$ is inverse scalable. Let $\alpha > 1$ be arbitrary. By inserting $\alpha \mathbf{x}$ into the i-th component of the standard form, we have

$$\begin{aligned} \mathbf{F}_i(\alpha \mathbf{x}) &= \alpha^2 \mathbf{x}^T \mathbf{Q}_i \mathbf{x} + \alpha \mathbf{q}_i \mathbf{x} \\ &= \alpha(\alpha - 1) \mathbf{x}^T \mathbf{Q}_i \mathbf{x} + \alpha(\mathbf{x}^T \mathbf{Q}_i \mathbf{x} + \mathbf{q}_i \mathbf{x}) \\ &= \alpha(\alpha - 1) \mathbf{x}^T \mathbf{Q}_i \mathbf{x} + \alpha \mathbf{F}_i(\mathbf{x}). \end{aligned}$$

Due to the existence of RNs and backhaul links, the trivial case when $\mathbf{Q}_i$ is an all-zero matrix can be excluded. Therefore, $\alpha(\alpha-1) \mathbf{x}^T \mathbf{Q}_i \mathbf{x} > 0$ for $\mathbf{x} > \mathbf{0}$, from which we have $\mathbf{F}_i(\alpha \mathbf{x}) > \alpha \mathbf{F}_i(\mathbf{x})$. □

Lemma 4.3 states that the load increases faster than the assignment $\mathbf{x}$. This is because that the cross product increases quadratically with respect to the assignment $\mathbf{x}$. Furthermore, the inverse scalability condition in Definition 4.3 cannot be satisfied, and therefore, $\mathbf{F}_i$ is in general not an SIF in $\mathbf{x}$.

Proposition 4.1. For the static interference model, the load function $\rho_i = \mathbf{F}_i(\mathbf{x})$ defined in (2.26) is in general *neither convex nor concave.*

Proof. Let $\mathbf{v}_1 = \mathbf{1}^L$ and $\mathbf{v}_2 = \left[\begin{smallmatrix} \mathbf{1}^{N \times (M+K)} \\ -\mathbf{1}^{K \times (M+K)} \end{smallmatrix}\right]$. Assume that the trivial case when $\mathbf{Q}_i$ is an all zero matrix is, due to the RNs and backhaul links, not possible. Then, we have $\mathbf{v}_1^T \mathbf{Q}_i \mathbf{v}_1 = \sum_i \sum_j \mathbf{T}_i > 0$ and $\mathbf{v}_2^T \mathbf{Q}_i \mathbf{v}_2 = -\sum_i \sum_j \mathbf{T}_i < 0$. Hence, $\mathbf{Q}_i$ is in general neither positive nor negative definite which completes the proof by following [84]. $\qquad\square$

Proposition 4.1 is derived based on the relation between convexity and definiteness of quadratic forms, details of which can be found in Appendix B.1. In Proposition 4.1, the load constraints render a non-convex set due to the cross term, for which relaxation techniques are required to efficiently find heuristic solutions. Note that there are two special cases that lead to linear load functions and linear constraints.

Assumption 4.1. If we assume that the relay backhaul link is not able to forward data ($\mathbf{T}_i \rightarrow \infty$). Hence, $\mathbf{x}^{(m,k)}$ and $\mathbf{x}^{(k,n)}$ are forced to be zero and no RN operation is possible. In this case, the network is equivalent to a *macro-only* network;

Assumption 4.2. If we assume perfect backhaul links ($\mathbf{T}_i \rightarrow 0$), the quadratic part of (4.1b) disappears and it implies a *femto-cell* network with, e.g., fixed backhaul links.

4.1.2. Load Function under the Dynamic Interference Model

The dependency between $\boldsymbol{\rho}$ and $\mathbf{x}$ becomes more complicated under the dynamic interference model. In this case, no explicit relation between the load vector and the assignment matrix can be formulated.

Proposition 4.2. Let $\mathbf{x} \in \mathbb{R}_+^{(M+K) \times (N+K)}$. The load function $\rho_i = \mathbf{F}_i(\boldsymbol{\rho}, \mathbf{x})$, $i \in \mathcal{B} \bigcup \mathcal{R}$ defined by (2.26) is an SIF with respect to $\boldsymbol{\rho} \in \mathbb{R}^{M+K}$.

Proof. This proposition is proved in [38] by claiming that the load function is positive and concave in $\boldsymbol{\rho} \in \mathbb{R}_+^{M+K}$ for given assignments. We present here a proof that utilizes the definition

of SIF. First, consider the spectral efficiency. For all $\alpha > 1$, we have

$$
\begin{aligned}
\omega_{i,j}(\alpha\boldsymbol{\rho}) &= \log(1 + \frac{p_i g_{i,j}}{\alpha \sum\limits_{d \in \mathcal{B} \bigcup \mathcal{R}, d \neq i} p_d g_{d,j} s_{d,j} \rho_d + \sigma_j}) \\
&> \log(1 + \frac{p_i g_{i,j}}{\alpha(\sum\limits_{d \in \mathcal{B} \bigcup \mathcal{R}, d \neq i} p_d g_{d,j} s_{d,j} \rho_d + \sigma_j)}) \\
&> \frac{1}{\alpha}\log(1 + \frac{p_i g_{i,j}}{\sum\limits_{d \in \mathcal{B} \bigcup \mathcal{R}, d \neq i} p_d g_{d,j} s_{d,j} \rho_d + \sigma_j}) = \frac{1}{\alpha}\omega_{i,j}(\boldsymbol{\rho}).
\end{aligned}
\tag{4.11}
$$

Herein, the first inequality is directly obtained by amplifying the denominator in the expression of SINR. For the second inequality, let $f(x) = (1+x)^\beta - (1 + \beta x)$ for $\beta < 1$ and $x > 0$ and note that $f'(x) = \beta(1+x)^{\beta-1} - \beta < 0$. Therefore, $f(x) = (1+x)^\beta - (1+\beta x) < f(0) = 0$ for $\beta < 1$ and $x > 0$, i.e., $\log(1 + \beta x) > \beta\log(1+x)$. Let $\alpha = \frac{1}{\beta} > 1$, we have $\log(1 + \frac{1}{\alpha}x) > \frac{1}{\alpha}\log(1+x)$, which directly leads to the second inequality.

In order to show the scalability of the load function we need to show that $\mathbf{F}_i(\alpha\boldsymbol{\rho}) < \alpha\mathbf{F}_i(\boldsymbol{\rho})$ holds for all $\alpha > 1$ and $i \in \mathcal{B} \bigcup \mathcal{R}$. Without loss of generality, we can prove the inequality holds for an arbitrary i-th component of the load function. To this end, we use (4.11) to conclude that:

$$
\mathbf{F}_i(\alpha\boldsymbol{\rho}) = \sum_{j \in \mathcal{U} \bigcup \mathcal{R}} \frac{r_j x_{i,j}}{b_i \omega_{i,j}(\alpha\boldsymbol{\rho})} < \sum_{j \in \mathcal{U} \bigcup \mathcal{R}} \frac{r_j x_{i,j}}{\frac{1}{\alpha}b_i \omega_{i,j}(\boldsymbol{\rho})} = \alpha\mathbf{F}_i(\boldsymbol{\rho}).
$$

To show monotonicity, we can calculate the Jacobian to obtain:

$$
\mathbf{J}_{\mathbf{F}_i}^{\rho_{\tilde{i}}}(\boldsymbol{\rho}, \mathbf{x}) = \begin{cases} \sum\limits_{j \in \mathcal{U} \bigcup \mathcal{R}} \frac{r_j x_{i,j}}{b_i \omega_{i,j}(\boldsymbol{\rho})} \frac{p_{\tilde{i}} g_{\tilde{i},j}/p_i g_{i,j}}{\ln(1+\tau_{i,j})(\tau_{i,j}^{-2} + \tau_{i,j}^{-1})} & i \neq \tilde{i}, \\ 0 & i = \tilde{i}. \end{cases}
\tag{4.12}
$$

It is obvious that $\mathbf{J}_{\mathbf{F}_i}^{\rho_{\tilde{i}}}(\boldsymbol{\rho}, \mathbf{x}) \geq 0$ for all $i, \tilde{i} \in \mathcal{B} \bigcup \mathcal{R}$ so that $\mathbf{J}_{\mathbf{F}}^{\boldsymbol{\rho}}$ has only non-negative elements. Hence, we can conclude monotonicity from Lemma 4.1 which completes the proof. $\square$

Proposition 4.3. Let $\mathcal{X} := \{\mathbf{x} \in [0,1]^L | \exists_{\boldsymbol{\rho}>\mathbf{0}} \boldsymbol{\rho} \geq \mathbf{F}(\boldsymbol{\rho}, \mathbf{x})\}$ and assume that $\mathcal{X} \neq \emptyset$. Then, there exists a *continuous function* $\mathbf{G} : \mathcal{X} \mapsto \mathbb{R}^{M+K}$ relating $\boldsymbol{\rho}$ to $\mathbf{x}$:

$$
\boldsymbol{\rho} = \mathbf{G}(\mathbf{x}), \text{ for } \mathbf{x} \in \mathcal{X}.
\tag{4.13}
$$

Proof. Let $\mathbf{x} \in \mathcal{X} \neq \emptyset$ be arbitrary. Then, by [81], we know that there exists $\boldsymbol{\rho}(\mathbf{x}) > 0$ such that

$$
\boldsymbol{\rho}(\mathbf{x}) = \mathbf{F}(\boldsymbol{\rho}(\mathbf{x}), \mathbf{x}), \text{ for } \mathbf{x} \in \mathcal{X}.
\tag{4.14}
$$

Moreover, $\boldsymbol{\rho}(\mathbf{x}) > 0$ is the unique fixed-point of $\mathbf{F}(\cdot, \mathbf{x})$. Now let

$$\mathbf{G}(\mathbf{x}) = \boldsymbol{\rho}(\mathbf{x}) = \mathbf{F}(\boldsymbol{\rho}(\mathbf{x}), \mathbf{x}) \tag{4.15}$$

and note that $\mathbf{G}$ maps elements of $\mathcal{X}$ into $\mathbb{R}_+^{M+K}$. Moreover, due to the uniqueness of the fixed point, we can conclude that, for any $\boldsymbol{\rho}(\mathbf{x}^{(1)}) > 0$ and $\boldsymbol{\rho}(\mathbf{x}^{(2)}) > 0$, $\boldsymbol{\rho}(\mathbf{x}^{(1)}) \neq \boldsymbol{\rho}(\mathbf{x}^{(2)})$ implies $\mathbf{x}^{(1)} \neq \mathbf{x}^{(2)}$. Therefore, $\mathbf{G} : \mathcal{X} \mapsto \mathbb{R}^{M+K}$ is a function. It is continuous because $\boldsymbol{\rho}(\mathbf{x})$ and $\mathbf{F}(\cdot, \mathbf{x})$ are both continuous, and the concatenation of continuous functions is continuous. $\qquad\square$

Moreover, $\mathbf{G}(\mathbf{x}) > 0$ is the unique fixed-point of $\mathbf{F}(\cdot, \mathbf{x})$ and can be found (if exists) iteratively by the following fixed-point algorithm:

$$\boldsymbol{\rho}(n+1) = \mathbf{F}(\boldsymbol{\rho}(n), \mathbf{x}), \text{ for } \mathbf{x} \in \mathcal{X}. \tag{4.16}$$

In other words, if $\mathcal{X} \neq \emptyset$, the algorithm converges to the unique fixed-point $\mathbf{G}(\mathbf{x})$ defined in (4.14).

Proposition 4.4. $\mathbf{G}(\mathbf{x})$ is *continuously differentiable* on $\mathcal{X} := \{\mathbf{x} \in [0,1]^{\mathrm{L}} | \exists_{\boldsymbol{\rho} > \mathbf{0}} \boldsymbol{\rho} \geq \mathbf{F}(\boldsymbol{\rho}, \mathbf{x})\}$.

Proof. By (4.15), it is sufficient to show that the function $\boldsymbol{\rho}(\mathbf{x}), \mathbf{x} \in \mathcal{X}$, is continuously differentiable. To this end, define $\tilde{\mathbf{F}} : \mathbb{R}^{M+K} \times \mathcal{X} \to \mathbb{R}^{M+K}$ to be $\tilde{\mathbf{F}}(\boldsymbol{\rho}, \mathbf{x}) := \boldsymbol{\rho} - \mathbf{F}(\boldsymbol{\rho}, \mathbf{x})$ and consider $\tilde{\mathbf{F}}(\boldsymbol{\rho}, \mathbf{x}) = 0$, which is an implicit function between $\boldsymbol{\rho}$ and $\mathbf{x}$ since $\boldsymbol{\rho} = \boldsymbol{\rho}(\mathbf{x})$ depends on $\mathbf{x} \in [0,1]^{\mathrm{L}}$. Therefore, by the implicit function theorem in Appendix C.1, it suffices to show that for all $(\boldsymbol{\rho}, \mathbf{x}) \in \mathbb{R}^{M+K} \times \mathcal{X}$: (i) $\tilde{\mathbf{F}}$ is differentiable, and (ii) the Jacobian $\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\rho}}(\boldsymbol{\rho}, \mathbf{x})$ with respect to $\boldsymbol{\rho}$ is invertible.

By examining Jacobian of $\mathbf{F}$ with respect to $\boldsymbol{\rho}$ given by (4.17) and with respect to $\mathbf{x}$ given by (4.21), we can conclude that partial derivatives of $\tilde{\mathbf{F}}$ *exist and are continuous*. Therefore, $\tilde{\mathbf{F}}$ is differentiable according to [85] which shows (i) is satisfied.

In order to prove (ii), we first show in the following that the Jacobian $\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\rho}}(\boldsymbol{\rho}, \mathbf{x})$ is a Generalized Diagonally Dominant Matrix (GDM) on $\mathbb{R}^{M+K} \times \mathcal{X}$. By the definition of $\tilde{\mathbf{F}}$, we have the Jacobian matrix $\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\rho}} = \mathbf{I} - \mathbf{J}_{\mathbf{F}}^{\boldsymbol{\rho}}$ with $\mathbf{J}_{\tilde{\mathbf{F}}_i}^{\rho_{\tilde{i}}}$ (the entry at row i and column $\tilde{i}$) given by

$$\mathbf{J}_{\tilde{\mathbf{F}}_i}^{\rho_{\tilde{i}}} = \begin{cases} -\displaystyle\sum_{j \in \mathcal{U} \bigcup \mathcal{R}} \frac{r_j^{(\mathrm{n})} x_{i,j}}{b_i \omega_{i,j}(\boldsymbol{\rho})} \frac{p_{\tilde{i}} g_{\tilde{i},j} / p_i g_{i,j}}{\ln(1+\tau_{i,j})(\tau_{i,j}^{-2}+\tau_{i,j}^{-1})} & i \neq \tilde{i}, \\ 1 & i = \tilde{i}. \end{cases} \tag{4.17}$$

Note that both i and $\tilde{i}$ are indices for transmitters (BS or RN) and j is the index for a receiver

(UE or RN). For any $\boldsymbol{\rho} > \mathbf{0}$,

$$\mathbf{J}_{\tilde{\mathbf{F}}_i}^{\rho} \cdot \boldsymbol{\rho} = \sum_{\tilde{i} \in \mathcal{B} \bigcup \mathcal{R}} \rho_{\tilde{i}} \mathbf{J}_{\mathbf{F}_i}^{\rho_{\tilde{i}}} = \rho_i - \sum_{\tilde{i} \neq i} \rho_{\tilde{i}} \mathbf{J}_{\mathbf{F}_i}^{\rho_{\tilde{i}}}, \tag{4.18}$$

where $\mathbf{J}_{\tilde{\mathbf{F}}_i}^{\rho}$ is the i-th row of $\mathbf{J}_{\tilde{\mathbf{F}}}^{\rho}$ and

$$\sum_{\tilde{i} \neq i} \rho_{\tilde{i}} \mathbf{J}_{\mathbf{F}_i}^{\rho_{\tilde{i}}} = \sum_{j \in \mathcal{U} \bigcup \mathcal{R}} \frac{r_j x_{i,j}}{b_i \omega_{i,j}(\boldsymbol{\rho})} \frac{(\sum_{\tilde{i} \neq i} \rho_{\tilde{i}} p_{\tilde{i}} g_{\tilde{i},j})/p_i g_{i,j}}{\ln(1 + \tau_{i,j})(\tau_{i,j}^{-2} + \tau_{i,j}^{-1})}.$$

Now let $f(x) = (1 + x)\ln(1 + x) - x$ for $x > 0$, and note that $f'(x) = \ln(1 + x) > 0$. Hence, $f(x) > f(0) = 0$, implying that $(1 + x)\ln(1 + x)/x > 1$ for all $x > 0$. Thus,

$$\ln(1 + \tau_{i,j})(\tau_{i,j}^{-2} + \tau_{i,j}^{-1})$$
$$= \tau_{i,j}^{-1}\ln(1 + \tau_{i,j})(1 + \tau_{i,j})/\tau_{i,j} > \tau_{i,j}^{-1} \, for \, all \, \tau_{i,j}.$$

Furthermore, it can be easily verified that

$$(\textstyle\sum_{\tilde{i} \neq i} \rho_{\tilde{i}} p_i g_{\tilde{i},j})/p_i g_{i,j} < (\textstyle\sum_{\tilde{i} \neq i} \rho_{\tilde{i}} p_{\tilde{i}} g_{\tilde{i},j} + \sigma_j)/p_i g_{i,j} = \tau_{i,j}^{-1}.$$

All these bounds together with (4.18) yield

$$\mathbf{J}_{\tilde{\mathbf{F}}_i}^{\rho} \cdot \boldsymbol{\rho} > \rho_i - \sum_{j \in \mathcal{U} \bigcup \mathcal{R}} \frac{r_j x_{i,j}}{b_i \omega_{i,j}(\boldsymbol{\rho}))} = 0. \tag{4.19}$$

According to (4.17), all the diagonal elements of $\mathbf{J}_{\tilde{\mathbf{F}}}^{\rho}$ are equal to one, while the off-diagonals are all negative. Therefore, we can conclude from Definition C.1 in Appendix C.2 that $\mathbf{J}_{\tilde{\mathbf{F}}}^{\rho}$ is an invertible GDM. This proves (ii) and completes the proof. $\square$

Moreover, the Jacobian of $\mathbf{G}$ yields according to implicit function theorem

$$\mathbf{J}_{\mathbf{G}}^{\mathbf{x}}(\mathbf{x}) = -\mathbf{J}_{\tilde{\mathbf{F}}}^{\rho}(\boldsymbol{\rho}, \mathbf{x})^{-1}\mathbf{J}_{\tilde{\mathbf{F}}}^{\mathbf{x}}(\boldsymbol{\rho}, \mathbf{x}), \tag{4.20}$$

where $\mathbf{J}_{\tilde{\mathbf{F}}}^{\rho}(\boldsymbol{\rho}, \mathbf{x}) \in \mathbb{R}^{(M+K) \times (M+K)}$ can be found in (4.12) and $\mathbf{J}_{\tilde{\mathbf{F}}}^{\mathbf{x}}(\boldsymbol{\rho}, \mathbf{x}) \in \mathbb{R}^{(M+K) \times (L}$ is the

Jacobian with respect to $\mathbf{x}$ given by

$$
\mathbf{J}_{\tilde{\mathbf{F}}_i}^{\mathbf{x}_h} = \begin{cases} -\dfrac{r_j^{(\mathrm{n})}}{b_i \omega_{i,j}(\boldsymbol{\rho})} & \text{for } i = \tilde{i}, \ j \in \mathcal{U}, \\ -\dfrac{r_k^{(\mathrm{k})}}{b_i \omega_{i,j}(\boldsymbol{\rho})} & \text{for } i = \tilde{i}, \ j \in \mathcal{R}, \\ -\dfrac{r_j^{(\mathrm{n})}}{b_i \omega_{i,\tilde{i}}(\boldsymbol{\rho})} x_{i,\tilde{i}} & \text{for } i \in \mathcal{B}, \ \tilde{i} \in \mathcal{R}, \ j \in \mathcal{U}, \\ 0 & \text{otherwise} \end{cases} .
\tag{4.21}
$$

Herein, $h = \tilde{i} + (j-1)(M+K)$ indicates the assignment between node $\tilde{i}$ and j. Note that we also call $\boldsymbol{\rho} = \mathbf{G}(\mathbf{x})$ the explicit load function.

Proposition 4.5. The explicit load function $\boldsymbol{\rho} = \mathbf{G}(\mathbf{x})$ is element-wise *monotonically increasing.*

Proof. According to Lemma 4.1, it suffices to prove that the Jacobian of $\mathbf{G}(\mathbf{x})$ is non-negative for all $\mathbf{x} \geq 0$, i.e., $\mathbf{J}_{\mathbf{G}}^{\mathbf{x}}(\mathbf{x}) \geq \mathbf{0}$ element-wise. First, we know that $\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\rho}}(\boldsymbol{\rho}, \mathbf{x})$ is a GDM and the inverse of a GDM has only non-negative elements [86]. Furthermore, by examining (4.21), it can be easily concluded that $\mathbf{J}_{\tilde{\mathbf{F}}}^{\mathbf{x}}(\boldsymbol{\rho}, \mathbf{x})$ is a non-positive matrix. Therefore, $\mathbf{J}_{\mathbf{G}}^{\mathbf{x}}(\mathbf{x}) \geq \mathbf{0}$ element-wise and $\mathbf{G}(\mathbf{x})$ is element-wise monotonically increasing under the dynamic interference model. $\qquad\square$

Proposition 4.6. The explicit load function $\boldsymbol{\rho} = \mathbf{G}(\mathbf{x})$ is element-wise *inverse scalable.*

Proof. Let $\alpha > 1$ be arbitrary. For all $\mathbf{x} \in \mathbb{R}_+^{\mathrm{L}}$, we have

$$
\mathbf{G}(\alpha \mathbf{x}) = \mathbf{F}(\mathbf{G}(\alpha \mathbf{x}), \alpha \mathbf{x}) > \alpha \mathbf{F}(\mathbf{G}(\alpha \mathbf{x}), \mathbf{x}) > \alpha \mathbf{F}(\mathbf{G}(\mathbf{x}), \mathbf{x}) = \alpha \mathbf{G}(\mathbf{x}).
$$

Herein, the first inequality holds due to the scalability of the load function in the static interference model as in (4.3), whereas the second inequality holds due to the monotonicity of the static load function as in (4.2) and Proposition 4.5. $\qquad\square$

This Proposition 4.6 shows that load scales faster than the increase rate of $\mathbf{x}$. On one hand, this is due to the fact that the cross term results in quadratic dependency between $\mathbf{x}$ and $\boldsymbol{\rho}$. On the other hand, due to the load coupling, the entries in the load vector increase mutually each other as one entry increases. Positive convex functions may have the property of inverse scalability, however, same as for $\mathbf{F}(\mathbf{x})$, the convexity cannot be concluded for $\mathbf{G}(\mathbf{x})$.

4.1.3. Objective Function and Optimization Simplification

From Proposition 4.2 and Proposition 4.5, we know that the load function is monotonically increasing under both the static and the dynamic interference models. We denote $\rho(\mathbf{x})$ to be the load vector induced by the assignment $\mathbf{x}$, such that $\rho(\mathbf{x}) = \mathbf{F}(\mathbf{x})$ for the static and $\rho(\mathbf{x}) = \mathbf{G}(\mathbf{x})$ for the dynamic interference model. Based on this, we can derive further properties of the objective function and simplify the optimization problem.

Proposition 4.7. The objective function $\mathrm{U}_\rho(\rho)$ (or $\mathrm{U}_\rho(\rho(\mathbf{x}))$) in (3.4a) is element-wise *monotonically increasing* with respect to $\rho \in \mathbb{R}_+^{(M+K)}$ (or $\mathbf{x} \in \mathbb{R}_+^{\mathrm{L}}$).

Proof. It is obvious that both l_0-norm and l_1-norm are *monotonically increasing* (non-decreasing). Therefore, $\mathrm{U}_\rho(\rho)$, which is the linear combination of l_0-norms and l_1-norms, is also *monotonically increasing*. Due to the monotonicity of load function, we have $\rho(\mathbf{x}_1) \geq \rho(\mathbf{x}_2)$ for $\mathbf{x}_1 \geq \mathbf{x}_2$. Hence, $\mathrm{U}_\rho(\rho(\mathbf{x}_1)) - \mathrm{U}_\rho(\rho(\mathbf{x}_2)) \geq 0$ for $\mathbf{x}_1 \geq \mathbf{x}_2$, which completes the proof.

$\square$

Proposition 4.8. If the energy saving problem in (3.4) is feasible, it has at least one global minimum on the boundary of the inequality constraints $\mathbf{X}^T \cdot \mathbf{1} \geq \mathbf{1}$, i.e. $\exists \mathbf{Y}$ as minimizer, such that $\mathbf{Y}^T \cdot \mathbf{1} = \mathbf{1}$

Proof. Without loss of generality, let $\mathbf{X}$ minimize the objective function with the inequality $\mathbf{X}^T \cdot \mathbf{1} = \mathbf{a} \geq \mathbf{1}$. Let $\mathbf{Y}$ denote the column-wise normalization of $\mathbf{X}$ over $\mathbf{a} \geq \mathbf{1}$ such that $y_{i,j} = x_{i,j}/a_j$. Then, $\mathbf{Y}^T \cdot \mathbf{1} = \mathbf{1}$ and $\mathbf{Z} = \mathbf{X} - \mathbf{Y} \geq \mathbf{0}$. Due to the monotonicity of the load function, i.e., due to Proposition 4.2 and Proposition 4.5, we can show that the load vector $\rho(\mathbf{Y}) \leq \rho(\mathbf{Y} + \mathbf{Z}) = \rho(\mathbf{X}) \leq \mathbf{1}$ is feasible. Furthermore, we know $\mathrm{U}_\rho(\rho(\mathbf{Y})) \leq \mathrm{U}_\rho(\rho(\mathbf{X}))$ due to Proposition 4.7. Since $\mathbf{X}$ is a minimizer of the problem, $\mathrm{U}_\rho(\rho(\mathbf{Y})) \geq \mathrm{U}_\rho(\rho(\mathbf{X}))$. Therefore, $\mathrm{U}_\rho(\rho(\mathbf{Y})) = \mathrm{U}_\rho((\rho\mathbf{X}))$ and $\mathbf{Y}$ is also a minimizer. $\square$

By replacing the inequality constraint with an equality constraint as in Problem (4.22), we can reformulate an optimization problem that achieves the same optimal value as the original one.

This will reduce the complexity for solving Problem (3.4).

$$\min_{\mathbf{x}} \quad \mathrm{U}_\rho(\boldsymbol{\rho}) = \mathrm{U}_\rho(\boldsymbol{\rho}(\mathbf{x})) \tag{4.22a}$$

$$\text{s.t.} \quad \mathbf{X}^T \cdot \mathbf{1} = \mathbf{1} \tag{4.22b}$$

$$\boldsymbol{\rho}(\mathbf{x}) \leq \mathbf{1} \tag{4.22c}$$

$$\mathbf{x} \geq \mathbf{0} \qquad\qquad . \tag{4.22d}$$

In general, this problem is still difficult to solve due to the non-continuous objective function and the non-convex constraints. In Section 4.2, we present relaxation techniques and centralized optimization algorithms to solve the problem in an efficient way.

4.2. Relaxation, Reformulation and Algorithms

4.2.1. Load Function under the Static Interference Model

The first relaxation approach is to define a convex subset of the non-convex constraint by reducing the dimension of the optimization variables.

Lemma 4.4. Assuming the static interference model, the load function $\mathbf{F}(\mathbf{x})$ defined in (2.26) becomes linear by fixing either the relay link assignment $\mathbf{x}^{(m,k)}$ or the access link assignment $\mathbf{x}^{(k,n)}$.

Proof. The proof can be done by investigating (4.5) and reformulating the load function as:

$$\mathbf{F}_i(\mathbf{x}) = \begin{cases} (\mathbf{x}^{(k,n)})^T \mathbf{D}_2^T \mathrm{diag}\left(\mathbf{e}_i^T \mathbf{D}_1\right) \mathbf{x}^{(m,k)} + \mathbf{q}_i^T \mathbf{x}, & i \in \mathcal{B}, \\ \mathbf{q}_i^T \mathbf{x}, & i \in \mathcal{R}, \end{cases}$$

where both $\mathbf{D}_1$ and $\mathbf{D}_2$ can be found in (4.4). Then, a linear function of $\mathbf{F}_i$ yields by fixing either $\mathbf{x}^{(m,k)}$ or $\mathbf{x}^{(k,n)}$. $\qquad\square$

This results in the algorithm which is called Iterative Backhaul Updating (IBU) in [17]. Basically, IBU transfers the load function into a linear function by alternatingly fixing $\mathbf{x}^{(m,k)}$ and $\mathbf{x}^{(k,n)}$. Fig. 4.1 (a) shows how IBU works with the help of a simple example: Consider the constraint $f(x_1, x_2) \leq 0.4$ which is illustrated as the non-convex feasibility region in Fig. 4.1. Beginning at point $(x_1^{(l)}, x_2^{(l)})$ and fixing $x_1 = x_1^{(l)}$, the first step of an optimization iteration has the vertical
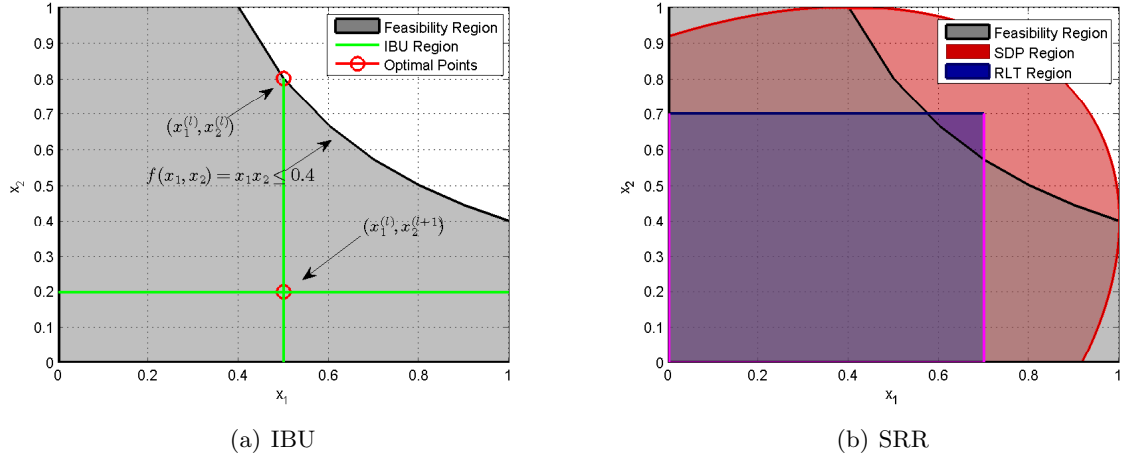
Figure 4.1.: Illustration of the relaxation techniques.

line as its feasible region and ends up with, e.g., $(x_1^{(l)}, x_2^{(l+1)})$ as the optimum. Then, the second step optimizes the problem over the horizontal line determined by point $(x_1^{(l)}, x_2^{(l+1)})$. These iterations continue with the next vertical and afterwards the next horizontal line as the set for optimization. The optimal point from the last iteration is always contained in the set for optimization of the next iteration. This ensures convergence in the objective since the sequence of optimal values is non-increasing by including the optimal point from the last iteration.

Another relaxation techniques in [17] that turns the non-convex constraint set into a convex set is based on Semi-Definite Programming (SDP) relaxation and Reformulation-Linearization Technique (RLT). The basic concept for SDP and RLT, explained in detail in Appendix B.2 and Appendix B.3, is to introduce a non-convex constraint $\bar{\mathbf{X}} - \mathbf{x}\mathbf{x}^T = 0$ and to replace the quadratic form of $\mathbf{x}$ with a linear term of $\bar{\mathbf{X}}$. In the case of the SDP relaxation, $\bar{\mathbf{X}} = \mathbf{x}\mathbf{x}^T$ is relaxed to $\bar{\mathbf{X}} - \mathbf{x}\mathbf{x}^T \succeq 0$, or equivalently as in (4.23d), while linear relations are derived with RLT by multiplying the boundaries of the original variables with the non-convex constraints. Note that "$\succeq$" means that the matrix is positive semi-definite. The relaxation combining both techniques is called in this thesis the SDP and RLT Relaxation (SRR). After performing SRR,

the original constraints are transformed into:

$$\operatorname{tr}(\tilde{\mathbf{P}}_i\tilde{\mathbf{X}}) = \operatorname{tr}(\begin{pmatrix} -2 & \mathbf{p}_i^T \\ \mathbf{p}_i & \mathbf{0}^{\mathrm{L}\times\mathrm{L}} \end{pmatrix}\tilde{\mathbf{X}}) = \mathbf{0}, \quad \text{for } i \in \{1, \dots, N+K\}, \tag{4.23a}$$

$$\operatorname{tr}(\tilde{\mathbf{Q}}_i\tilde{\mathbf{X}}) = \operatorname{tr}(\begin{pmatrix} -2 & \mathbf{q}_i^T \\ \mathbf{q}_i & \mathbf{Q}_i \end{pmatrix}\tilde{\mathbf{X}}) \le \mathbf{0}, \quad \text{for } i \in \{1, \dots, M+K\}, \tag{4.23b}$$

$$\mathbf{0} \le \mathbf{x} \le \mathbf{1}, \tag{4.23c}$$

$$\tilde{\mathbf{X}} := \begin{pmatrix} 1 & \mathbf{x}^T \\ \mathbf{x} & \bar{\mathbf{X}} \end{pmatrix} \succeq 0, \tag{4.23d}$$

$$\bar{\mathbf{X}} \ge \mathbf{0}^{\mathrm{L}\times\mathrm{L}}, \tag{4.23e}$$

$$\bar{\mathbf{X}} - \mathbf{1}\mathbf{x}^T - \mathbf{x}\mathbf{1}^T \ge -\mathbf{1}^{\mathrm{L}\times\mathrm{L}}, \tag{4.23f}$$

$$\bar{\mathbf{X}} - \mathbf{x}\mathbf{1}^T \le \mathbf{0}^{\mathrm{L}\times\mathrm{L}}. \tag{4.23g}$$

For the same example as for IBU, the SDP relaxation spans a convex ellipse, while a convex polytope is generated by the RLT relaxation (4.23e)-(4.23g) as shown in Fig. 4.1 (b). The intersection of the ellipse and the square then renders a convex set that approximates the original non-convex set determined by the load function. Compared with IBU, this enables a one-step optimization with apparently more area of the original feasible set being included for optimization, however imposes strong computation complexity and hazards the violation of the load constraints.

In fact, the load function involves only cross products as quadratic forms and is therefore a non-convex *bilinear form* (See Appendix B.1). Optimization problems containing bilinear constraints are in general Non-deterministic Polynomial-time (NP)-hard [87]. However, we show in the following that an exact equivalent reformulation is possible which allows for a linear optimization in another domain.

Lemma 4.5. Consider the optimization problem

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) \tag{4.24a}$$

$$\text{s.t.} \quad \mathbf{x} \in \mathcal{X} := \{\mathbf{x} \mid g(\mathbf{x}) \le 0\}, \tag{4.24b}$$

and the substitution problem

$$\min_{\mathbf{z}} \quad f(\mathbf{T}(\mathbf{z})) \tag{4.25a}$$

$$\text{s.t.} \quad \mathbf{z} \in \mathcal{N} := \{\mathbf{x} \mid g(\mathbf{T}(\mathbf{z})) \le 0\}. \tag{4.25b}$$

with a surjective mapping $\mathbf{T} : \mathcal{N} \to \mathcal{X}$, i.e., $\forall \mathbf{x} \in \mathcal{X}, \exists \mathbf{z} \in \mathcal{N}$ with $\mathbf{x} = \mathbf{T}(\mathbf{z})$. If $\mathbf{z}^*$ minimizes Problem (4.25) then, $\mathbf{x}^* = \mathbf{T}(\mathbf{z}^*)$ minimizes Problem (4.24).

Proof. If $\mathbf{z}^*$ minimizes Problem (4.25), then, $f(\mathbf{x}^*) = f(\mathbf{T}(\mathbf{z}^*)) \leq f(\mathbf{T}(\mathbf{z})), \forall \mathbf{z} \in \mathcal{N}$. Since $\mathbf{T}$ is surjective, $\forall \mathbf{x} \in \mathcal{X}, \exists \mathbf{z} \in \mathcal{N}$ such that $\mathbf{x} = \mathbf{T}(\mathbf{z})$ and $f(\mathbf{x}^*) \leq f(\mathbf{T}(\mathbf{z})) = f(\mathbf{x})$. Therefore, $\mathbf{x}^*$ is the minimizer of Problem (4.24). $\qquad\square$

Proposition 4.9. Assume the static interference model and $\mathbf{X}^{(k,k)} = \mathbf{0}^{(k,k)}$. Define

$$\mathcal{X} := \{\mathbf{x} \mid \text{subject to (4.22b), (4.22c), (4.22d)}\} \in \mathbb{R}_+^{(M \times N + M \times K + K \times N) \times 1} \tag{4.26}$$

Let $\mathbf{z} = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$ with $\mathbf{u} \in \mathbb{R}_+^{(M \times N) \times 1}$ and $\mathbf{v} \in \mathbb{R}_+^{(M \times K \times N) \times 1}$. Then, consider the function $\mathbf{T} : \mathbb{R}_+^{M \times N + M \times K \times N} \to \mathbb{R}_+^{M \times N + M \times K + K \times N}$:

$$\mathbf{x}^{(m,n)} = \mathbf{u} \tag{4.27a}$$

$$\mathbf{x}^{(k,n)} = \mathbf{V}^T \cdot \mathbf{1}^M \tag{4.27b}$$

$$\mathbf{X}^{(m,k)} = \mathbf{V}(\mathbf{r}^{(n)} \otimes \mathbf{I}^K)\text{diag}\left(\mathbf{r}^{(k)} + \epsilon\right)^{-1} \tag{4.27c}$$

where $\mathbf{V} \in \mathbb{R}^{M \times (K \times N)}$ is given by $\mathbf{v} = \text{vec}(\mathbf{V})$ and $\epsilon > 0$ is a sufficiently small value that ensures the invertability of $\text{diag}\left(\mathbf{r}^{(k)} + \epsilon\right)^{-1}$.

By applying $\mathbf{T}$ to (4.26), we have $\mathcal{N} : \{\mathbf{z} \mid \mathbf{x} = \mathbf{T}(\mathbf{z}), \mathbf{x} \in \mathcal{X}\}$ whereas $\mathcal{N}$ is a *convex set*:

$$\mathcal{N} : \{\mathbf{z} \mid \mathbf{H}\mathbf{z} \leq \mathbf{1}, \mathbf{D}\mathbf{z} = \mathbf{1}\} \in \mathbb{R}_+^{(M \times N \times K + M \times N) \times 1} \tag{4.28}$$

for some given matrices $\mathbf{H} \in \mathbb{R}_+^{(M+K) \times (M \times N + M \times K \times N)}$ and $\mathbf{D} \in \mathbb{R}_+^{M \times (M \times N + M \times K \times N)}$. Furthermore, the mapping $\mathbf{T} : \mathcal{N} \to \mathcal{X}$ is *surjective*.

Proof. First, we prove that the constraints (4.22b) − (4.22d) for $\mathbf{x}$ hold under the mapping $\mathbf{x} = \mathbf{T}(\mathbf{z})$, for $\mathbf{z} \in \mathcal{N}$.

i) (4.22d) obviously holds, since $\mathbf{T}$ maps $\mathcal{N}$ only into non-negative values according to (4.27).

ii) Regarding (4.22c) or (4.1b), we insert $\mathbf{x} = \mathbf{T}(\mathbf{z})$ into the load function $\mathbf{F}(\mathbf{x})$ defined in (2.26) and show in the following that $\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{T}(\mathbf{z})) = \mathbf{H}\mathbf{z}$. Using the matrix operation rules in

Appendix A, we obtain:

$$\mathbf{x}^{(k,n)} \overset{(4.27b)}{=} \mathbf{V}^T \mathbf{1}^M \overset{(A.4)}{=} (\mathbf{I}^{K\times N} \otimes [\mathbf{1}^M]^T)\mathbf{v} \tag{4.29a}$$

$$\mathbf{x}^{(m,k)} \overset{(4.27c)}{=} \mathbf{V}(\mathbf{r}^{(n)} \otimes \mathbf{I}^K)\mathrm{diag}\left(\mathbf{r}^{(k)} + \epsilon\right)^{-1}$$

$$\overset{(A.5)}{=} \mathrm{vec}(\mathbf{V}(\mathbf{r}^{(n)} \otimes \mathbf{I}^K)) \circ \mathrm{vec}(\mathbf{1}^M \otimes [\mathbf{r}^{(k)}{}_*]^T) \tag{4.29b}$$

$$\overset{(A.2)}{=} ([\mathbf{r}^{(n)}]^T \otimes \mathbf{I}^{M\times K})\mathbf{v} \circ \mathrm{vec}(\mathbf{1}^M \otimes [\mathbf{r}^{(k)}{}_*]^T).$$

where a matrix with subscript "$*$", e.g., $\mathbf{A}_*$, denotes to the element-wise inverse of $\mathbf{A} + \epsilon$, i.e., $\mathbf{A}_{*i,j} = \frac{1}{\mathbf{A}_{i,j}+\epsilon}$. Then we can further extend (4.4a)-(4.4c) as:

$$\lambda^{(m,n)} \overset{(4.4a)}{=} ([\mathbf{r}^{(n)}]^T \otimes \mathbf{I}^M)\mathrm{diag}\left(\mathbf{W}^{(m,n)}\right)\mathbf{x}^{(m,n)}$$

$$\overset{(4.27a)}{=} \underbrace{([\mathbf{r}^{(n)}]^T \otimes \mathbf{I}^M)\mathrm{diag}\left(\mathbf{W}^{(m,n)}\right)}_{\mathbf{A}_1} \cdot \mathbf{u} \tag{4.30a}$$

$$\lambda^{(k,n)} \overset{(4.4b)}{=} ([\mathbf{r}^{(n)}]^T \otimes \mathbf{I}^K)\mathrm{diag}\left(\mathbf{W}^{(k,n)}\right)\mathbf{x}^{(k,n)}$$

$$\overset{(4.29a)}{=} \underbrace{([\mathbf{r}^{(n)}]^T \otimes \mathbf{I}^K)\mathrm{diag}\left(\mathbf{W}^{(k,n)}\right)(\mathbf{I}^{K\times N} \otimes [\mathbf{1}^M]^T)}_{\mathbf{B}_2} \cdot \mathbf{v} \tag{4.30b}$$

$$\lambda^{(m,k)} \overset{(4.4c)}{=} ([\mathbf{1}^K]^T \otimes \mathbf{I}^M)\mathrm{diag}\left(\mathbf{W}^{(m,k)}\right)(\mathrm{vec}(\mathbf{1}^M \otimes [\mathbf{r}^{(k)}]^T) \circ \mathbf{x}^{(m,k)}))$$

$$\overset{(4.29b)}{=} ([\mathbf{1}^K]^T \otimes \mathbf{I}^M)\mathrm{diag}\left(\mathbf{W}^{(m,k)}\right) \cdot (([\mathbf{r}^{(n)}]^T \otimes \mathbf{I}^{M\times K})\mathbf{v} \circ \dots$$

$$\dots \mathrm{vec}(\mathbf{1}^M \otimes [\mathbf{r}^{(k)}{}_*]^T) \circ \mathrm{vec}(\mathbf{1}^M \otimes [\mathbf{r}^{(k)}]^T))$$

$$= \underbrace{([\mathbf{1}^K]^T \otimes \mathbf{I}^M)\mathrm{diag}\left(\mathbf{W}^{(m,k)}\right)([\mathbf{r}^{(n)}]^T \otimes \mathbf{I}^{M\times K})}_{\mathbf{B}_1} \cdot \mathbf{v}. \tag{4.30c}$$

Then, we have

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{T}(\mathbf{z})) = \begin{pmatrix} \lambda^{(m,n)} + \lambda^{(m,k)} \\ \lambda^{(k,n)} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{0} \end{pmatrix}\mathbf{u} + \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix}\mathbf{v} = \mathbf{H} \cdot \mathbf{z}, \tag{4.31}$$

where

$$\mathbf{H} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{0} & \mathbf{B}_2 \end{pmatrix}. \tag{4.32}$$

Thus, $\mathbf{F}(\mathbf{x}) \leq \mathbf{1}$ if and only if $\mathbf{Hz} \leq \mathbf{1}$. This implies that $\forall \mathbf{z} \in \mathcal{N}$, $\mathbf{x} = \mathbf{T}(\mathbf{z})$ satisfies (4.22c).

iii) At last, we show that the equality constraint in (4.22b) is partly implied by the definition of (4.27c) and partly equivalent to the condition $\mathbf{Dz} = \mathbf{1}$. The equality in (4.22b) can be

decomposed into two equalities:

$$(\mathbf{X}^{(\mathrm{m,k})})^T\mathbf{1}^M = \mathbf{1}^K, \tag{4.33a}$$

$$(\mathbf{X}^{(\mathrm{m,n})})^T\mathbf{1}^M + (\mathbf{X}^{(\mathrm{k,n})})^T\mathbf{1}^K = \mathbf{1}^N. \tag{4.33b}$$

From the definition of (4.27c), we have:

$$
\begin{aligned}
(\mathbf{X}^{(\mathrm{m,k})})^T\mathbf{1}^M &= \mathrm{diag}\left(\mathbf{r}^{(\mathrm{k})} + \epsilon\right)^{-1}([\mathbf{r}^{(\mathrm{n})}]^T \otimes \mathbf{I}^K)\mathbf{V}^T\mathbf{1}^M \\
&\overset{(4.27\mathrm{b})}{=} \mathrm{diag}\left(\mathbf{r}^{(\mathrm{k})} + \epsilon\right)^{-1}([\mathbf{r}^{(\mathrm{n})}]^T \otimes \mathbf{I}^K)\mathbf{x}^{(\mathrm{k,n})} \\
&= \mathrm{diag}\left(\mathbf{r}^{(\mathrm{k})} + \epsilon\right)^{-1}\cdot(\mathbf{X}^{(\mathrm{k,n})}\mathbf{r}^{(\mathrm{n})}) \\
&= \mathrm{diag}\left(\mathbf{r}^{(\mathrm{k})} + \epsilon\right)^{-1}\cdot\mathbf{r}^{(\mathrm{k})} = \mathbf{1}.
\end{aligned}
$$

.

As for (4.33b), we have

$$
\begin{aligned}
&(\mathbf{X}^{(\mathrm{m,n})})^T\mathbf{1}^M + (\mathbf{X}^{(\mathrm{k,n})})^T\mathbf{1}^K \\
=\ & (\mathbf{I}^N \otimes [\mathbf{1}^M]^T)\mathrm{vec}(\mathbf{X}^{(\mathrm{m,k})}) + (\mathbf{I}^N \otimes [\mathbf{1}^K]^T)\mathrm{vec}(\mathbf{X}^{(\mathrm{k,n})}) \\
\overset{(4.27\mathrm{a}),(4.29\mathrm{a})}{=}\ & (\mathbf{I}^N \otimes [\mathbf{1}^M]^T)\mathbf{u} + (\mathbf{I}^N \otimes [\mathbf{1}^K]^T)(\mathbf{I}^{K\times N} \otimes [\mathbf{1}^M]^T)\mathbf{v} \\
=\ & \mathbf{Dz} = \mathbf{1},
\end{aligned}
$$

where

$$\mathbf{D} = [(\mathbf{I}^N \otimes (\mathbf{1}^M)^T),\quad (\mathbf{I}^N \otimes (\mathbf{1}^K)^T)(\mathbf{I}^{K\times N} \otimes (\mathbf{1}^M)^T]. \tag{4.34}$$

Now, (4.22c)-(4.22d) are all true and it means that if $\mathbf{z} \in \mathcal{N}$ with $\mathbf{H}$ as in (4.32) and $\mathbf{D}$ as in (4.34), then $\mathbf{x} = \mathbf{T}(\mathbf{z}) \in \mathcal{X}$.

For surjection, let us consider the function $\tilde{\mathbf{T}} : \mathcal{X} \to \mathbb{R}_+^{M\times N + M\times K\times N}$ defined in (4.35):

$$\mathbf{u} = \mathbf{x}^{(\mathrm{m,n})}, \tag{4.35a}$$

$$\mathbf{V} = ([\mathbf{1}^N]^T \otimes (\mathbf{X}^{(\mathrm{m,k})})) \circ (\mathbf{1}^M \otimes [\mathbf{x}^{(\mathrm{k,n})}]^T). \tag{4.35b}$$

We prove in the following that $\tilde{\mathbf{T}}$ is the inverse operation of $\mathbf{T}$, i.e, $\forall \mathbf{x} \in \mathcal{X}$, $\mathbf{T}(\mathbf{z}) = \mathbf{T}(\tilde{\mathbf{T}}(\mathbf{x})) = \mathbf{x}$. This can be done by showing that the conditions in (4.29) hold by applying the composition $\mathbf{T}(\tilde{\mathbf{T}}(\mathbf{x}))$ to $\mathbf{x} \in \mathcal{X}$ according to (4.29) and (4.35). First of all, (4.27a) is obvious. For (4.27b),

we have:

$$
\mathbf{V}^T \mathbf{1}^M \stackrel{(4.35a)}{=} \left(\mathbf{1}^N \otimes [\mathbf{X}^{(m,k)}]^T \right) \circ \left([\mathbf{1}^M]^T \otimes \mathrm{vec}(\mathbf{X}^{(k,n)}) \right) \mathbf{1}^M
$$

$$
\stackrel{(A.10)}{=} \mathrm{vec}(\mathbf{X}^{(k,n)}) \circ (\mathbf{1}^N \otimes [\mathbf{X}^{(m,k)}]^T) \mathbf{1}^M
$$

$$
= \mathrm{vec}(\mathbf{X}^{(k,n)}) \circ (\mathbf{1}^N \otimes [\mathbf{1}^K]^T) = \mathrm{vec}(\mathbf{X}^{(k,n)}).
$$

Furthermore, we can apply the matrix operation rule in Appendix A.11 for proving (4.27c):

$$
\mathbf{V}(\mathbf{r}^{(n)} \otimes \mathbf{I}^K) \mathrm{diag}\left(\mathbf{r}^{(k)} + \epsilon \right)^{-1}
$$

$$
\stackrel{(4.35b)}{=} \left([\mathbf{1}^N]^T \otimes \mathbf{X}^{(m,k)} \right) \circ \left(\mathbf{1}^M \otimes [\mathrm{vec}(\mathbf{X}^{(k,n)})]^T \right) (\mathbf{r}^{(n)} \otimes \mathbf{I}^K) \mathrm{diag}\left(\mathbf{r}^{(k)} + \epsilon \right)^{-1}
$$

$$
\stackrel{(A.11)}{=} \mathbf{X}^{(m,k)} \circ \left((\mathbf{1}^M \otimes \mathrm{vec}(\mathbf{X}^{(k,n)})) \mathbf{r}^{(n)} \otimes \mathbf{I}^K) \mathrm{diag}\left(\mathbf{r}^{(k)} + \epsilon \right)^{-1} \right)
$$

$$
= \mathbf{X}^{(m,k)} \circ \left((\mathbf{1}^M \otimes [\mathbf{r}^{(k)}]^T) \circ (\mathbf{1}^M \otimes \mathbf{r}^{(k)}_*) \right)
$$

$$
= \mathbf{X}^{(m,k)} \circ (\mathbf{1}^M \otimes [\mathbf{1}^K]^T) = \mathbf{X}^{(m,k)}.
$$

Hence, $\forall \mathbf{x} \in \mathcal{X}$, $\exists \mathbf{z} = \tilde{\mathbf{T}}(\mathbf{x}) \in \mathcal{N}$ such that $\mathbf{T}(\mathbf{z}) = \mathbf{T}(\tilde{\mathbf{T}}(\mathbf{x})) = \mathbf{x}$. Then, $\mathbf{T} : \mathcal{N} \to \mathcal{X}$ is surjective which completes and proof. $\square$

From the mathematic point of view, the reformulation technique applies the concept of Reduced Reformulation-Linearization Technique (RRLT) to a bilinear form [87]. We will call the corresponding algorithms also RRLT. By multiplying the non-linear constraints (here, $\mathbf{F}(\mathbf{x}) \leq \mathbf{1}$) with the linear equality constraints (here, $\mathbf{X}^T \cdot \mathbf{1} = \mathbf{1}$), the reduction of non-linearity can be expected. For instance, the set $\mathcal{A} : \{(x,y,z)|x \leq yz, y=1\}$ is non-linear, however, by multiplying $x \leq yz$ with $y = 1$ we have an identical set $\tilde{\mathcal{A}} : \{(x,y,z)|x = z, y=1\}$.

Moreover, another interpretation from the modeling point of view can also explain the linear reformulation. Assume $v_{i,(k \cdot j)}$, an element in $\mathbf{V}$, to be the end-to-end routing variable indicating that UE j is connected to BS i through RN k. Then, the operation in (4.27b) states the fact that the sum of all the flows that originate from UE j through RN k to all BSs is equal to the flow between UE j and RN k, i.e., $\sum_i v_{i,(k \cdot j)} = x_{k,j}$. The operation in (4.27c) expresses that the rate routed from RN k to BS i is the sum of the flows that originate from all the UEs through RN k to BS i. This can be also written as $x_{i,k}(r_k^{(k)} + \epsilon) = x_{i,k} \sum_j (\sum_i v_{i,(k \cdot j)} r_j^{(n)} + \epsilon) = \sum_j (v_{i,(k \cdot j)} r_j^{(n)} + \epsilon)$ or $x_{i,k} = (\sum_j v_{i,(k \cdot j)} r_j^{(n)} + \epsilon) / (\sum_j (x_{k,j} r_j^{(n)} + \epsilon))$. On the other hand, the backwards transformation in (4.35b) indicates a multiplication of the rate routing: $v_{i,(k \cdot j)} = x_{i,k} x_{k,j}$. Both models are in fact equivalent and the differences are: in the transformed domain $\mathbf{z}$, the backhaul link assignments $x_{i,k} = (\sum_j v_{i,(k \cdot j)} r_j^{(n)} + \epsilon) / (\sum_j (\sum_i v_{i,(k \cdot j)} r_j^{(n)} + \epsilon))$ is a non-linear expression, whereas the original
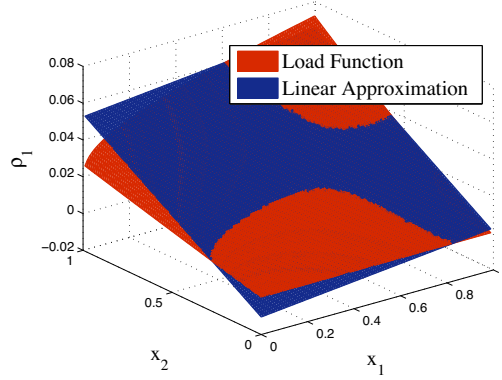
Figure 4.2.: Load function and approaximation, where $x_1 = 0.7$ and $x_2 = 0.7$

bilinear cross term $x_{i,k} \sum_j x_{k,j} r_j^{(\mathrm{n})}$ can be transformed into a linear combination in the new domain as $\sum_j v_{i,(k \cdot j)} r_j^{(\mathrm{n})}$. If a linear optimization problem does not require to directly take into account the backhaul link assignment, a simple LP in the transformed the domain can be formulated under the static interference model.

4.2.2. Load Function under Dynamic Interference Model

In the static interference model, $\mathbf{H}$ and $\mathbf{D}$ are constant matrices. However, in the dynamic interference model, the matrices depend on the load and in turn on the assignments ($\mathbf{x}$ or $\mathbf{z}$). Therefore, non-linearity occurs again and further approximations and relaxations are necessary for solving the problem under the dynamic interference model.

Since the load function is continuous differentiable, the first linearization approach is to use the first-order Taylor series to approximate the load $\boldsymbol{\rho}$ as

$$\boldsymbol{\rho} = \mathbf{G}(\mathbf{x}) \approx \tilde{\mathbf{G}}(\mathbf{x}) = \mathbf{G}(\mathbf{x}^*) + \mathbf{J}_G^{\mathbf{x}}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*), \tag{4.36}$$

where $\mathbf{x}^*$ is the point of expansion. Knowing $\mathbf{x}^*$, a constant value of the load vector $\mathbf{G}(\mathbf{x}^*)$ can be computed according Proposition 4.3, while the Jacobian $\mathbf{J}_G^{\mathbf{x}}(\mathbf{x}^*)$ can be also calculated by (4.20). Therefore, a linear function yields which approximates the original non-linear load coupling function. It can be seen in Fig. 4.2 that the linear approximation results in a reasonable estimation of the load function. Especially at the region near the point of expansion, the approximation is quite close the original load function. With this approximation, a fast one-step optimization can be performed to find out the optimal solution, however, without guarantee for satisfying the original load constraints. Therefore, further steps may be required to reassign the

extra users or redirect part of the QoS to avoid overloading some cells. Since the approximation is more precise near the point of expansion, an intuitive solution is to perform iterative optimizations, whereby the assignment should not be changing too much between iterations.

Now, we give another important contribution of this thesis - an important property that relates both the static and the dynamic load functions.

Proposition 4.10. Assume $\mathcal{X} := \{\mathbf{x} \geq \mathbf{0} | \exists_{\boldsymbol{\rho}>0} \mathbf{F}(\boldsymbol{\rho}, \mathbf{x}) \leq \boldsymbol{\rho}\} \neq \emptyset$. $\mathbf{F}(\boldsymbol{\rho}, \mathbf{x}) \leq \boldsymbol{\rho}$ implies $\mathbf{G}(\mathbf{x}) \leq \boldsymbol{\rho}$.

Proof. Assume $\mathbf{G}(\mathbf{x}) = \boldsymbol{\rho}'$. Since $\mathbf{F}(\boldsymbol{\rho}, \mathbf{x})$ is an SIF in $\boldsymbol{\rho}$, we know from [81, Lemma 2] that the fixed-point iteration $\boldsymbol{\rho} = \mathbf{F}(\boldsymbol{\rho}, \mathbf{x})$, if exists, generates a monotone decreasing sequence that starts from $\boldsymbol{\rho}$ and converges to $\boldsymbol{\rho}'$. Hence, $\mathbf{G}(\mathbf{x}) = \boldsymbol{\rho}' \leq \mathbf{F}(\boldsymbol{\rho}, \mathbf{x}) \leq \boldsymbol{\rho}$. $\square$

This proposition enables to iteratively relax the dynamic load coupling constraint $\mathbf{G}(\mathbf{x}) \leq \mathbf{1}$ with the static interference model:

$$\mathbf{F}(\boldsymbol{\rho}^{(\mathrm{n})}, \mathbf{x}) \leq \boldsymbol{\rho}^{(\mathrm{n})}. \tag{4.37}$$

If the sequence of load vectors is non-increasing, i.e., $\boldsymbol{\rho}^{(\mathrm{n})} \leq \boldsymbol{\rho}^{(\mathrm{n}-1)} \leq \ldots \leq \mathbf{1}$, then original load constraint is fulfilled, since $\mathbf{G}(\mathbf{x}) \leq \boldsymbol{\rho}^{(\mathrm{n})} \leq \boldsymbol{\rho}^{(\mathrm{n}-1)} \leq \ldots \leq \mathbf{1}$ as the iteration continues. There are different ways for generating the non-increasing sequence of the load vectors. For instance, we can ensure the inequality by taking $\boldsymbol{\rho}^{(\mathrm{n})} = \mathbf{F}(\boldsymbol{\rho}^{(\mathrm{n}-1)}, \mathbf{x}^{(\mathrm{n})}) \leq \boldsymbol{\rho}^{(\mathrm{n}-1)}$. Furthermore, a heuristic example that can be explained by Proposition 4.10 is the optimization by iteratively shutting down and excluding cells. In this case, the load of an excluded cell is kept at zero and we can still apply the worst-case interference model to the rest of the network.

Finally, the set for the optimization can be either directly written as a convex set under the static interference model (Proposition 4.9) or iteratively reformulated as a series of convex set under the dynamic interference model (Proposition 4.9 and Proposition 4.10) by choosing a proper non-increasing load sequence. Therefore, convex optimizations can be carried out for solving the energy saving problem in nomadic relaying networks. In the following, we discuss how to deal with the l_0-norm and show different algorithms based on the properties of load function.

4.2.3. Iterative Algorithms for Energy Savings

In order to tackle the discreteness of the objective function, we adopt the same approach as in [29] by approximating the l_0-norm using a *strictly concave function* [88]. Note that $|\rho_i|_0 = |\mathbf{e}_i^T \mathbf{X} \mathbf{1}|_0$,

and hence l_0-norm can be either expressed with respect to the load vector $\boldsymbol{\rho}$ (4.38) or with respect to the assignment matrix $\mathbf{X}$ (4.39).

$$\tilde{\mathrm{U}}_{\boldsymbol{\rho}}(\boldsymbol{\rho}) = \sum_{i=1}^{M+K} \frac{c_i(\log(\epsilon + \rho_i) - \log(\epsilon))}{\log(1 + \epsilon^{-1})} + d_i \rho_i \approx \mathrm{U}_{\rho}(\boldsymbol{\rho}) \tag{4.38}$$

$$\tilde{\mathrm{U}}_{\boldsymbol{\rho},\mathbf{x}}(\boldsymbol{\rho}, \mathbf{x}) = \sum_{i=1}^{M+K} \frac{c_i(\log(\epsilon + \mathbf{e}_i^T \mathbf{X} \mathbf{1}) - \log(\epsilon))}{\log(1 + \epsilon^{-1})} + d_i \rho_i \approx \mathrm{U}_{\rho}(\boldsymbol{\rho}) \tag{4.39}$$

Herein, ϵ is a small constant and $\lim_{\epsilon \to 0} \tilde{\mathrm{U}}_{\rho}(\boldsymbol{\rho}) = \lim_{\epsilon \to 0} \tilde{\mathrm{U}}_{\boldsymbol{\rho},\mathbf{x}}(\boldsymbol{\rho}, \mathbf{x}) = \mathrm{U}_{\rho}(\boldsymbol{\rho}) = \mathrm{U}_{\rho}(\boldsymbol{\rho}(\mathbf{x}))$. Yet, the approximated objective function is not easy to minimize due to the concave form of the approximation. Although a global optimum is out of reach, we can proceed essentially as in [29] to find almost optimal solutions by using the MM techniques.

MM-Algorithm

First, we briefly explain the MM-algorithm [89] under varying feasible region by looking at the following minimization problem:

$$\min_{\mathbf{x} \in \mathcal{Z}} f(\mathbf{x})$$

where $f : \mathcal{Z} \to \mathbb{R}$. Then, a function $g : \mathcal{Z}^{(n)} \to \mathbb{R}$ is called a majorization function of f over $\mathcal{Z}^{(n)} \in \mathcal{Z}$ with $\mathbf{x}^{(n)} \in \mathcal{Z}^{(n)}$, if $g(\mathbf{x}|\mathbf{x}^{(n)}) \geq f(\mathbf{x})$ with $g(\mathbf{x}^{(n)}|\mathbf{x}^{(n)}) = f(\mathbf{x}^{(n)})$. Then, the iterative process of the minimization problems

$$\mathbf{x}^{(n+1)} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{Z}^{(n)}} g(\mathbf{x}|\mathbf{x}^{(n)})$$

converge to a local minimum or a saddle point of the original problem as n goes to infinity, since $f(\mathbf{x}^{(n+1)}) \leq g(\mathbf{x}^{(n+1)}|\mathbf{x}^{(n)}) \leq g(\mathbf{x}^{(n)}|\mathbf{x}^{(n)}) = f(\mathbf{x}^{(n)})$. Note that the condition $\mathbf{x}^{(n)} \in \mathcal{Z}^{(n)}$ is necessary, because it ensures that the inequality $g(\mathbf{x}^{(n+1)}|\mathbf{x}^{(n)}) \leq g(\mathbf{x}^{(n)}|\mathbf{x}^{(n)})$ holds. Then, the key issue for applying MM algorithm to our optimization problem is the choice of the majorization functions.

Static Energy Consumption

The static energy consumption may dominate the total energy consumption of a cell, i.e., $c_i \gg d_i$ for cell i and the part of the dynamic energy consumption can be ignored. This assumption

is widely used for formulating optimization algorithms [17, 18, 29]. The advantage is that the energy consumption approximation becomes a strict concave function with respect to the assignment $\mathbf{x}$.

For an arbitrary concave function $f(x)$, we have in general a set of majorization functions:

$$g(x, x^*) = \mathbf{J}_f^x(x^*)(x - x^*) + f(x^*),$$

since $\mathbf{J}_f^x(x^*)(x - x^*) > f(x) - f(x^*)$ holds for any strict concave functions.

Denote

$$\tilde{\mathbf{U}}_{\mathbf{x}}(\mathbf{x}) = \sum_{i=1}^{M+K} \frac{c_i(\log(\epsilon + \mathbf{e}_i^T \mathbf{X1}) - \log(\epsilon))}{\log(1 + \epsilon^{-1})}$$

to be the objective approximation by ignoring the dynamic energy consumption. Then, then majorization function can be calculated as

$$\tilde{\mathbf{U}}_{\mathbf{x}}(\mathbf{x}|\mathbf{x}^*) = \mathbf{J}_{\tilde{\mathbf{U}}}^{\mathbf{x}}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + \tilde{\mathbf{U}}_{\mathbf{x}}(\mathbf{x}^*)\tilde{\mathbf{U}}_{\mathbf{x}}(\mathbf{x})$$

for any given assignment $\mathbf{x}^*$. Thus, we can formulated the iterative process with each iteration minimizing an LP as

$$\mathbf{x}^{(n+1)} = \text{argmin}_{\mathbf{x}} \tilde{\mathbf{U}}_{\mathbf{x}}(\mathbf{x}|\mathbf{x}^{(n)}).$$

Connecting the different relaxation techniques in Section 4.2, we formulate in the following Algorithm 1 (IBU), Algorithm 2 (SRR) and Algorithm 3 (Sequential Linear Reformulation (SLR)), respectively. Herein, we choose $\epsilon_t > 0$ with a typical value of 10^{-2} as the termination scalar. Note that the only difference between Algorithm 2 and Algorithm 3 is that they have different feasibility regions. Whereas $\mathcal{X}^*$ denotes the set spanned by the SRR as in (4.23a)-(4.23g), $\mathcal{X}^{(n)}$ is the set determined by the approximation (4.36) at point $\mathbf{x}^{(n)}$ according to (3). Note that heuristic rounding operations may be required to avoid violation of the load function after finding an optimal assignment by using Algorithms, however, we do not explicitly state the operation here.

Algorithm 1 Iterative Backhaul Updating (IBU)

Let n = 0, initialize $\mathbf{x}^{(n)}$
loop
  Let q = 0, initialize $\mathbf{x}^{(n,q)}$
  loop
    Solve LP: $\mathbf{x}^{(n,q)'} = \text{argmin}_{\mathbf{x}\in\mathcal{X},\mathbf{x}^{(m,k)}=(\mathbf{x}^{(m,k)})^{(n,q)}}\tilde{U}_{\mathbf{x}}(\mathbf{x}|\mathbf{x}^{(n,q)})$
    Solve LP: $\mathbf{x}^{(n,q+1)} = \text{argmin}_{\mathbf{x}\in\mathcal{X},\mathbf{x}^{(k,n)}=(\mathbf{x}^{(k,n)})^{(n,q)'}}\tilde{U}_{\mathbf{x}}(\mathbf{x}|\mathbf{x}^{(n,q)'})$
    if $\tilde{U}_{\mathbf{x}}(\mathbf{x}^{(n,q)}|\mathbf{x}^{(n)}) - \tilde{U}_{\mathbf{x}}(\mathbf{x}^{(n,q+1)}|\mathbf{x}^{(n)}) < \epsilon_t$ **then**
      $\tilde{U}_{\mathbf{x}}^{(n)} = \tilde{U}_{\mathbf{x}}(\mathbf{x}^{(n,q)}|\mathbf{x}^{(n)})$, break
    else
      q ← q + 1
    end if
  end loop
  if $\tilde{U}_{\mathbf{x}}(\mathbf{x}^{(n,q)}|\mathbf{x}^{(n)}) - \tilde{U}_{\mathbf{x}}^{(n)} < \epsilon_t$ **then**
    break
  else
    n ← n + 1, $\mathbf{x}^{(n)} \leftarrow \mathbf{x}^{(n,q)}$
  end if
end loop

Algorithm 2 SDP-RLT Relaxation (SRR)

Let n = 0, initialize $\mathbf{x}^{(n)}$
loop
  Solve LP: $\mathbf{x}^{(n+1)} = \text{argmin}_{\mathbf{x}\in\mathcal{X}^*}\tilde{U}_{\mathbf{x}}(\mathbf{x}|\mathbf{x}^{(n)})$
  if $\tilde{U}_{\mathbf{x}}(\mathbf{x}^{(n+1)}|\mathbf{x}^{(n)}) - \tilde{U}_{\mathbf{x}}(\mathbf{x}^{(n)}|\mathbf{x}^{(n)}) < \epsilon_t$ **then**
    break
  else
    n ← n + 1
  end if
end loop

Algorithm 3 Sequential Linearization Relaxation (SLR)

Let n = 0, initialize $\mathbf{x}$ with $\mathbf{x}^{(n)}$
loop
  Solve LP: $\mathbf{x}^{(n+1)} = \text{argmin}_{\mathbf{x}\in\mathcal{X}^{(n)}}\tilde{U}_{\mathbf{x}}(\mathbf{x}|\mathbf{x}^{(n)})$
  if $\tilde{U}_{\mathbf{x}}(\mathbf{x}^{(n+1)}|\mathbf{x}^{(n)}) - \tilde{U}_{\mathbf{x}}(\mathbf{x}^{(n)}|\mathbf{x}^{(n)}) < \epsilon_t$ **then**
    break
  else
    n ← n + 1
  end if
end loop

Dynamic Energy Consumption

The dynamic energy consumption, which is the l_1-norm of the load function, is neither concave nor convex in the original domain of $\mathbf{x}$ according to Proposition 4.1. In order to enable efficient algorithmic solutions, optimization in transformed domain according to Proposition 4.9 should be considered. Assume the static interference model with a static load of $\boldsymbol{\rho}^*$. In the domain $\mathbf{z}$, the load function $\boldsymbol{\rho} = \mathbf{F}(\boldsymbol{\rho}^*, \mathbf{T}(\mathbf{z}))$ is a linear function and the approximation $\tilde{\mathrm{U}}_{\boldsymbol{\rho}}(\boldsymbol{\rho}) = \tilde{\mathrm{U}}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^*, \mathbf{F}(\mathbf{T}(\mathbf{z})))$ is a strictly concave function. Hence, it is possible to formulate majorization functions by using the concave property of the function $\tilde{\mathrm{U}}_{\boldsymbol{\rho}}(\mathbf{F}(\mathbf{T}(\mathbf{z})))$ for performing iteratively optimizations. In Proposition 4.11, we give a general guideline on how to find majorization functions. Then, we formulate optimization algorithms for both the static and the dynamic interference models.

Proposition 4.11. Denote $\mathcal{N}^{(n)} : \{\mathbf{z} \in \mathbb{R}_+^{(M \times K \times N) \times 1} | \exists_{\boldsymbol{\rho}^{(n)} > 0} \mathbf{F}(\boldsymbol{\rho}^{(n)}, \mathbf{T}(\mathbf{z})) \leq \boldsymbol{\rho}^{(n)}\} \neq \emptyset$ and let $\tilde{\mathrm{U}}_{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z}) = \tilde{\mathrm{U}}_{\boldsymbol{\rho}}(\mathbf{F}(\boldsymbol{\rho}^{(n)}, \mathbf{T}(\mathbf{z})))$. Then,

$$\tilde{\mathrm{U}}_{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z}|\mathbf{z}^{(q)}) = \tilde{\mathrm{U}}_{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z}^{(q)}) + \mathbf{J}_{\tilde{\mathrm{U}}_{\mathbf{z}}}^{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z}^{(q)})(\mathbf{z} - \mathbf{z}^{(q)}) \tag{4.40}$$

is the majorization function of the approximation $\tilde{\mathrm{U}}_{\boldsymbol{\rho}}(\boldsymbol{\rho}(\mathbf{z}))$ over $\mathcal{N}^{(n)}$, where $\boldsymbol{\rho}(\mathbf{z})$ is the load induced by $\mathbf{x} = \mathbf{T}(\mathbf{z})$, i.e., $\boldsymbol{\rho}(\mathbf{z}) = \mathbf{F}(\boldsymbol{\rho}^{(n)}, \mathbf{x}) = \mathbf{F}(\boldsymbol{\rho}^{(n)}, \mathbf{T}(\mathbf{z}))$ for the static interference model or $\boldsymbol{\rho}(\mathbf{z}) = \mathbf{G}(\mathbf{x}) = \mathbf{G}(\mathbf{T}(\mathbf{z}))$ for the dynamic interference model.

Proof. For the static interference assumption, we have $\boldsymbol{\rho}(\mathbf{z}) = \mathbf{F}(\boldsymbol{\rho}^{(n)}, \mathbf{T}(\mathbf{z}))$, while for the dynamic interference assumption $\boldsymbol{\rho}(\mathbf{z}) = \mathbf{G}(\mathbf{T}(\mathbf{z})) \leq \mathbf{F}(\boldsymbol{\rho}^{(n)}, \mathbf{T}(\mathbf{z}))$ holds over $\mathbf{z} \in \mathcal{N}^{(n)}$ according to Proposition 4.10. Then, we have for both cases

$$\boldsymbol{\rho}(\mathbf{z}) \leq \mathbf{F}(\boldsymbol{\rho}^{(n)}, \mathbf{T}(\mathbf{z})), \quad \forall \mathbf{z} \in \mathcal{N}^{(n)}. \tag{4.41}$$

Furthermore, it can be concluded due to the monotonicity of the objective function that

$$\tilde{\mathrm{U}}_{\boldsymbol{\rho}}(\boldsymbol{\rho}(\mathbf{z})) \leq \tilde{\mathrm{U}}_{\boldsymbol{\rho}}(\mathbf{F}(\boldsymbol{\rho}^{(n)}, \mathbf{T}(\mathbf{z}))) = \tilde{\mathrm{U}}_{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z}), \quad \forall \mathbf{z} \in \mathcal{N}^{(n)}. \tag{4.42}$$

As the composition $\tilde{\mathrm{U}}_{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z})$ is a concave function in $\mathbf{z}$, we have the inequality

$$\tilde{\mathrm{U}}_{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z}) \leq \tilde{\mathrm{U}}_{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z}^{(q)}) + \mathbf{J}_{\tilde{\mathrm{U}}_{\mathbf{z}}}^{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z}^{(q)})(\mathbf{z} - \mathbf{z}^{(q)}) = \tilde{\mathrm{U}}_{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z}|\mathbf{z}^{(q)}), \quad \forall \mathbf{z} \in \mathcal{N}^{(n)}. \tag{4.43}$$

By combining both (4.42) and (4.43), the proof is complete. $\qquad\square$

Algorithm 4 Reduced Reformulation Linearization Technique (RRLT)

Let n = 0, initialize $\mathbf{z}^{(n)}$
loop
   Solve LP: $\mathbf{z}^{(n+1)} = \mathrm{argmin}_{\mathbf{z} \in \mathcal{N}} \tilde{U}_{\mathbf{z}}(\mathbf{1}, \mathbf{z} | \mathbf{z}^{(n)})$
   if $\tilde{U}_{\mathbf{x}}(\mathbf{x}^{(n+1)} | \mathbf{x}^{(n)}) - \tilde{U}_{\mathbf{x}}(\mathbf{x}^{(n)} | \mathbf{x}^{(n)}) < \epsilon_t$ **then**
     break
   else
     $n \leftarrow n + 1$
   end if
   $\mathbf{x} = \mathbf{T}(\mathbf{z})$
end loop

Algorithm 5 Dynamic Reduced Reformulation Linearization Technique (DRRLT)

Let n = 0, initialize $\mathbf{z}^{(n)}$
loop
   Let n = 0, initialize $\mathbf{z}^{(n,q)}$
   loop
     Solve LP: $\mathbf{z}^{(n,q+1)} = \mathrm{argmin}_{\mathbf{z} \in \mathcal{N}^{(n)}} \tilde{U}_{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z} | \mathbf{z}^{(n,q)})$
     if $\tilde{U}_{\mathbf{z}}(\mathbf{z}^{(n,q)} | \mathbf{z}^{(n)}) - \tilde{U}_{\mathbf{z}}(\mathbf{z}^{(n,q+1)} | \mathbf{z}^{(n)}) < \epsilon_t$ **then**
       $\tilde{U}_{\mathbf{z}}^{(n)} = \tilde{U}_{\mathbf{z}}(\mathbf{z}^{(n,q)} | \mathbf{z}^{(n)})$, break
     else
       $q \leftarrow q + 1$
     end if
   end loop
   if $\tilde{U}_{\mathbf{z}}(\mathbf{z}^{(n,q)} | \mathbf{z}^{(n)}) - \tilde{U}_{\mathbf{z}}^{(n)} < \epsilon_t$ **then**
     break
   else
     $n \leftarrow n + 1$, $\mathbf{z}^{(n)} \leftarrow \mathbf{z}^{(n,q)}$, update $\boldsymbol{\rho}^{(n)}$ according to $\mathbf{z}^{(n)}$
   end if
   $\mathbf{x}^{(n)} = \mathbf{T}(\mathbf{z}^{(n,q)})$
end loop

With the help of Proposition 4.11, we can construct MM-functions for both the worst-case and the dynamic interference models. For the worst-case interference model, $\boldsymbol{\rho}^{(n)}$ is chosen to be $\mathbf{1}$. Then, in each iteration of the MM-algorithm, we need to minimize the following LP:

$$\mathbf{z}^{(q+1)} = \mathrm{argmin}_{\mathbf{z}} \tilde{U}_{\mathbf{z}}(\mathbf{1}, \mathbf{z} | \mathbf{z}^{(q)}). \tag{4.44}$$

In Algorithm 4, the RRLT is performed with a fixed convex set $\mathcal{N}$ for optimization. In Algorithm 5, we introduce the Dynamic Reduced Reformulation-Linearization Technique (DRRLT), which comprises two tiers of iterations. In each of the outer iterations, a non-increasing sequence $\{\boldsymbol{\rho}^{(n)}\}$ is computed and used as the static load for the LP of each inner iteration:

$$\mathbf{z}^{(n,q)} = \mathrm{argmin}_{\mathbf{z} \in \mathcal{N}} \tilde{U}_{\mathbf{z}}(\boldsymbol{\rho}^{(n)}, \mathbf{z} | \mathbf{z}^{(q)}). \tag{4.45}$$
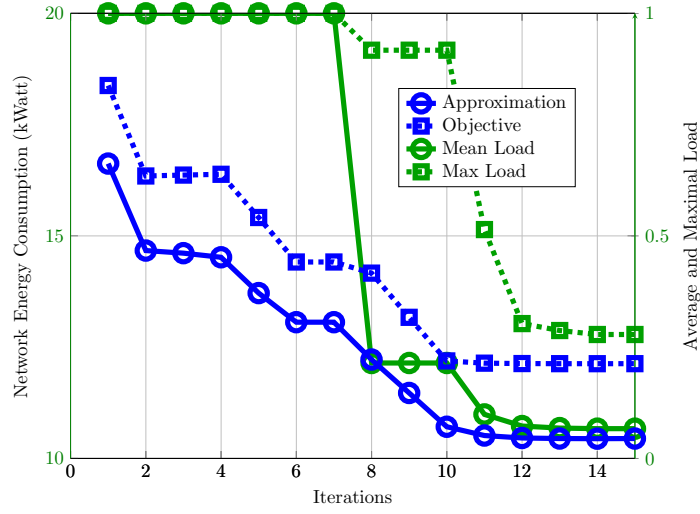
Figure 4.3.: Convergence of the DRRLT algorithm.

Furthermore, convergence is ensured for DRRLT if the load sequence $\{\boldsymbol{\rho}^{(n)}\}$ is monotonically decreasing. Note that if we choose a constant sequence of $\{\boldsymbol{\rho}^{(n)}\}$, we have an optimization scenario under the static interference model. If we choose element-wise l_0-norm of $\{\boldsymbol{\rho}^{(n)}\}$ as the load vector of the static interference model for the inner iterations, an algorithm yields such that the BSs and RNs are iteratively shut-down, whereas a static (typically worst-case) interference model is assumed for the active cells. We can also choose to update the load by $\mathbf{F}(\boldsymbol{\rho}^{(n)}, \mathbf{x}^{(n+1)})$ or by finding the fixed point, i.e. $\boldsymbol{\rho}^{(n+1)} = \mathbf{G}(\mathbf{x}^{(n+1)})$ for the next iteration. This means, after having a solution for the problem under a static load $\boldsymbol{\rho}^{(n)}$, we can update the load situation, where less interferences can be expected. Then, we have $\boldsymbol{\rho}^{(n+1)} \leq \boldsymbol{\rho}^{(n)} \ldots \boldsymbol{\rho}^{(0)} \leq \mathbf{1}$ guarantees both convergence and feasibility. Therefore, the algorithm for the dynamic interference model outperforms theoretically the worst-case algorithm.

In Fig 4.3, the convergence of the DRRLT algorithm is shown. The simulation scenario is taken the same as in Section 4.3. It can be clearly seen that the energy consumption approximation (4.38), which is the objective of our optimization problem, is monotonically decreasing by iterations. We can further confirm that the real energy consumption behaves similarly to the approximation. Further, the maximum load as well as the mean load decrease monotonically in accordance with the theory and the performance of the objective function. Note that in Fig. 4.3, the first inner iteration ends at the 7-th overall iteration and the second ends at the 10-th overall iteration, both resulting in substantial load reductions and energy reductions.

4.3. Performance Evaluation

4.3.1. Simulation Scenarios and Methodology

We simulate two deployment scenarios (urban and suburban) with 19 BSs in a two-tier hexagonal layout. While an Inter Site Distance (ISD) of 500 m and a number of 300 UEs are assumed for the urban scenario, the suburban scenario has an ISD of 1500 m and 100 UEs in the area. The UEs are uniformly distributed within the coverage of the BSs with a uniform data rate distribution from 0 to 200 kbps (except Fig. 4.4 with varying rate), while 50 and 150 RNs are assumed to be randomly distributed in urban and suburban scenarios, respectively (except Fig. 4.6(a) with varying density). If not specified, the static (c_i) and dynamic (d_i) energy consumption of a BS are 1000 Watt and 100 Watt, while the RNs has 10 Watt for the static and 1 Watt for the dynamic energy consumption if not explicitly otherwise specified. Assumption 2.10 is taken for interference coordination, whereas other assumptions in the network can be found in Chapter 2. Furthermore, we perform 200 iterations to average the statistical uncertainty. As a summary, the system setup is listed in Tab. 6.1.

Table 4.1.: Simulation Setup.

Baseline Dployment Scenario	
layout:	19 BSs in hexagon shape
ISD:	Urban 500 m, Suburban 1500 m
number of RNs:	Urban 100, Suburban 50
number of UEs:	Urban 300, Suburban 200
Baseline Transmission Parameters	
transmission power	46 dBm for BS & 23 dBm for RN
BS static energy consumption	1000 Watt for BS & 50 Watt for RN
available bandwidth (in-band)	10 MHz @2 GHz
antenna configuration	2 antennas for BSs, RNs and UEs
path loss model for all links	as in Table A.2.1.1.2-3 in [11]
noise figure	5 dB at UE & RN
Varying Simulation Parameters	
user data rate	[1 10 100 1000] kbps
RN/BS energy consumption ratio	[0 0.001 0.01 0.1 1 10]
dynamic/static energy consumption ratio	[0 0.001 0.01 0.1 1 10]
number of nomadic relays	[0 50 100 150 200]
nomadic relay antenna gain	[0 3 6 9 12] dB

In this work, we have proposed optimization algorithms for both static and dynamic interference models. Furthermore, Assumption 4.1 and Assumption 4.2 are also considered. The two special cases can be also seen as the upper and lower bound of the algorithms, respectively. With all the above mentioned scenarios, we compare the following algorithms:

- IBU: the IBU algorithm as proposed in Algorithm 1;

- SRR: the SRR algorithm as proposed in Algorithm 2;

- SLR: the algorithm in Algorithm 3;

- RRLT: the algorithm in Algorithm 4;

- DRRLT: the RRLT algorithm in Algorithm 5;

- M-DRRLT: macro-only network, which is the upper-bound of DRRLT algorithm;

- F-DRRLT: femto network, which is the lower-bound of DRRLT algorithm.

4.3.2. Impact of UE QoS Requirements

In Fig. 4.4, the algorithms are evaluated in both (a) urban and (b) suburban scenarios with varying user rate profiles that reflect the different data traffic activities during the different day times. It is clear from the figure that in both cases the proposed RRLT algorithms (RRLT, DRRLT, M-RRLT and F-DRRLT) outperform significantly the other algorithms that only consider static energy (IBU, SRR and SLR), since (i) the reformulation allows for a more complete search towards the optima and (ii) a more accurate optimization can be achieved by taking into account the dynamic energy consumption.

Furthermore, by comparing RRLT and DRRLT, it can be concluded that significantly more energy savings are achieved by using the dynamic interference model. In particular, a huge energy saving potential is identified if the average user rate requirement is very low. In the urban scenario (which is an interference limited scenario), about 75% of the network energy consumption can be reduced by the DRRLT, whereas in the suburban scenario, we can expect 50% energy savings. The gain vanishes, however, as the rate increases due to the fact that higher rate indicates higher load of the system. With high load, it is difficult to handover totally the load of a certain BS to switch it off.

In the urban scenario, the F-DRRLT achieves only significant gains compared with DRRLT in
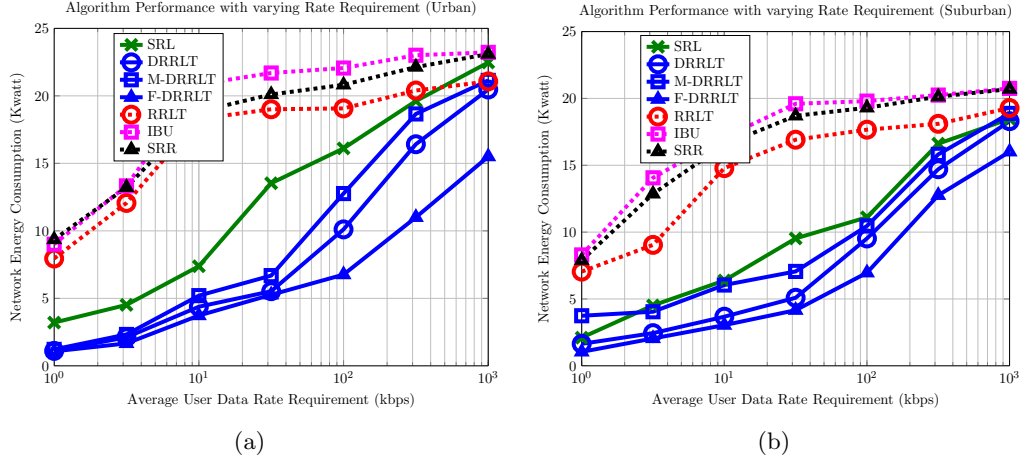
Figure 4.4.: Energy saving performance in (a) the urban scenario and (b) the suburban scenario.

the high rate region. Furthermore, the M-DRRLT performance is close to DRRLT for all rate profiles, since the coverage areas of the BSs overlap with each other, and therefore a UE can be connected to multiple BSs without the need for help of the RNs. This is different for the suburban scenario, where in general less energy consumptions than the urban scenario are seen due to the lower user density. It is worth noting that the lower bound (F-DRRLT) is quite close to DRRLT in this case, since only limited number of UEs can be connected to the RNs due to the large area and small RN coverage. In the low rate region, the macro-only scheme requires significantly more energy than DRRLT, since the BS coverage are limited without RNs, and hence the UEs are not able to connect to many BSs.

4.3.3. Impact of the Energy Consumption Model

Modern BSs may have various energy consumption models. Furthermore, different RN energy consumption profiles may also occur due to different hardware and software implementations. We apply in this subsection the DRRLT algorithm to both scenarios and analyze the impact of different energy consumption models on the total energy saving performance.

Fig. 4.5.(a) shows the impact of the dynamic energy consumption on the average load and on the total number of active BSs. We choose a varying ratio ([0 0.001 0.01 0.1 1 10]) of the dynamic energy consumption over the static energy consumption. As the dynamic energy consumption increases, a higher total energy consumption of the network is required and more BSs tend to be active. However, the average load is decreasing, since a higher spectral efficiency is targeted in order to limit the dynamic energy consumption. If the dynamic energy consumption dominates, a spectral efficiency based cell selection methods achieve the optimum. In the objective function,
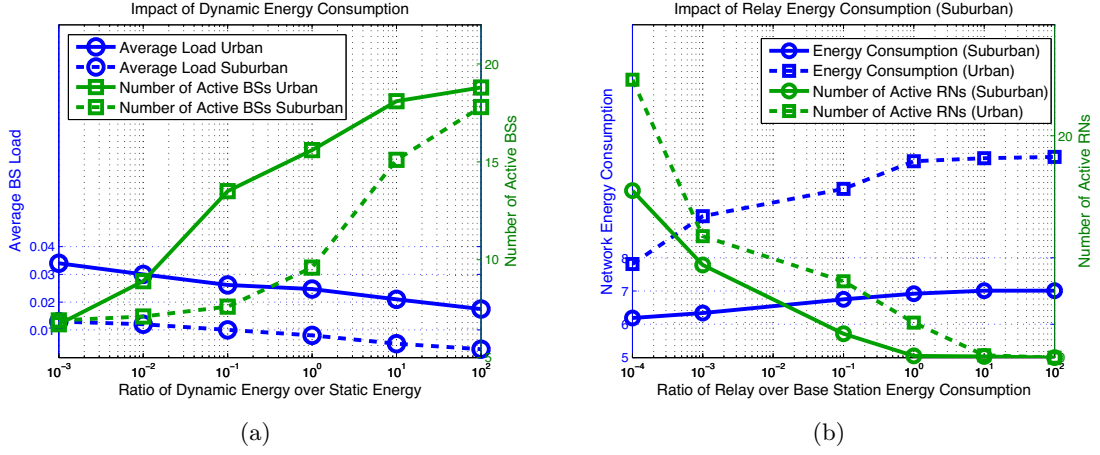
Figure 4.5.: Impact of (a) the dynamic energy consumption and (b) the relay energy consumption model.

the l_1-norm has a load balancing effect, whereas the l_0-norm indeed leads to load unbalancing. In general, shutting down cells trades-off the static energy consumption with resource efficiency. For instance, Fig. 4.5.(a) shows that for a ratio of 10, almost all BSs need to be active in order to achieve a higher spectral efficiency.

Fig. 4.5.(b) shows the total network energy consumption as well as the number of active RNs versus the energy consumption ratio of an RN to a BS. We can conclude that higher RN energy consumption reduces drastically the number of active RNs. If the RN energy cost is at the similar level of a BS, almost no RN will be activated. On the other hand, the energy increase due to higher dynamic energy consumptions is not significant, especially in the suburban scenario where fewer UEs are connected to the network through an RN.

4.3.4. Impact of Relay Density and Configurations

As a central role in the nomadic relaying network, the nomadic RNs may significantly affect the energy saving performance of the proposed algorithms. Therefore, we focus on analyzing the impact of relay specifications on the total energy saving performance. An essentially important factor here is the density of the nomadic relays. Intuitively a higher density means a higher diversity of the antennas of the nomadic RNs and leads to higher probability of suitable candidates for redirecting data traffic for energy savings. This can be justified by Fig. 4.6(a), in which a higher number of RNs in the network significantly reduce the total energy consumption. Although the number of active RNs is similar for the both scenarios, the algorithm achieves more energy savings in the suburban scenario (up to 50% here).

(a) Impact of relay density   (b) Impact of relay antenna gain
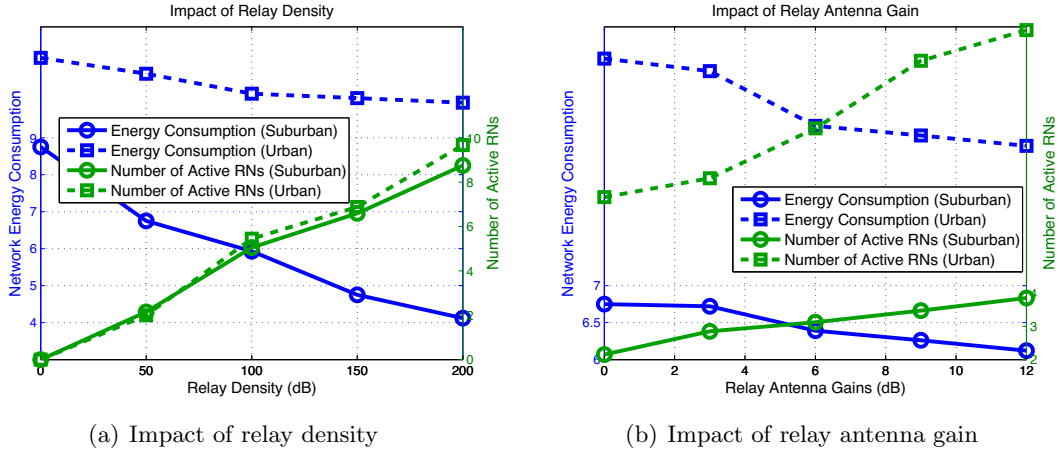
Figure 4.6.: Impact of nomadic relay density, antenna gains and energy consumption.

Another performance improving factor for the nomadic relaying networks is the antenna performance of the vehicular on-board relays. Advanced and smart antenna designs may significantly enhance the probability of correctly decoding the received signals at the nomadic relays. This means as well a higher end to end performance for the UEs that are connected through RNs into the network. For this reason, we assume a varying backhaul SINR gain from 0 to 12 dB at the nomadic RNs and evaluate the DRRLT algorithm in both the urban and the suburban scenarios. It can be seen in Fig. 4.6(b) that limited energy savings are achieved in the suburban scenario, while higher gains can be expected in the urban scenario. This is due to the fact that there are a large number of RNs and UEs in the urban scenario, so that more UEs can access the network through RNs which are equipped with boosted backhaul antenna.

Chapter 5.

Distributed Algorithms for Cell Selection and Admission Control

The algorithms in Chapter 4 require Channel State Information (CSI) at the central management unit. In practice, CSI reporting delay, signaling overhead as well as computational complexity of the centralized optimization lead to performance degradation, especially in high mobility scenarios. Therefore, distributed RAN-based algorithms for energy saving in nomadic relaying networks is proposed based on [19]. In the distributed algorithms, RNs and UEs select their access points based on the broadcasted radio link and network measurements, whereas the cells enter sleep mode when no access request arrives for a certain period of time. This chapter begins with a short introduction to the radio measurements in LTE. Then, the proposed cell selection and admission control algorithms is given with proof for convergence. At the end, numerical results are presented and compared with the results achieved by the centralized algorithms.

5.1. Radio Measurements for Cell Selection

In the LTE network, there are different cell selection procedures for determining the camping cell of a node. According to the model in Chapter 2, these cell selection procedures modify the assignment matrix $\mathbf{X}$ in a distributed manner. The UEs can perform *initial cell selection* by cell searching to read out system information and then choose a cell for random access. The initial cell selection defines the procedure from the state of "power-up" to "connected". Furthermore, the cell re-election procedure is defined for the transfer from the state of being disconnected from a cell to be re-connected to a cell (not necessarily the same cell), whereas handover procedures are defined for changing a camping cell. These happen mostly due to the change of the radio

link or network conditions that occur during UE movements or network load changes.

In order to enable the aforementioned cell selection procedures, various radio measurements are specified in LTE. Comparably to the scrambling code for separating Wideband Code Division Multiple Access (WCDMA) BSs, LTE BSs are identified by a hierarchical cell-searching procedure [90]. Primary Synchronization Signal (PSS) and Secondary Synchronization Signal (SSS) signals, which carry the cell identity, are transmitted periodically at predefined frames of the Physical Resource Block (PRB)s, providing the ability to distinguish between 504 physical cell identities. Then, the UE is able to decode the Physical Broadcast Channel (PBCH) of different cells and extract the corresponding Master Information Blocks (MIBs) and System Information Blocks (SIBs), which contain parameters for cell selection, re-selection and handover decisions. The radio measurements, including Reference Signal Received Power (RSRP), Received Signal Strength Indicator (RSSI) and Reference Signal Received Quality (RSRQ), which are specified in [91] to reflect the quality of the radio link and network, are measurable at a receiver side:

- RSRP is defined as the linear average over the power contributions (in Watt) of the resource elements that carry cell-specific reference signals within the considered measurement frequency bandwidth. Considering the system model in Chapter 2, RSRP of cell i at node j can be formulated as: $p_i g_{i,j}/n_{rb}$, where n_{rb} is the number of PRBs of the entire Evolved-UMTS Terrestrial Radio Access (E-UTRA) carrier bandwidth.

- RSSI is the received wide-band power, including thermal noise and noise generated in the receiver, within the bandwidth defined by the receiver pulse shaping filter. RSSI can be understood as the sum of all the average powers plus noise, which is according to our model: $\sum_{i \in \mathcal{B} \bigcup \mathcal{R}} p_i g_{i,j} \rho_i + \sigma_j$.

- RSRQ is defined as the ratio $n_{rb} \times \text{RSRP}/\text{RSSI}$, where the measurements RSRP and RSSI shall be made over the same set of resource blocks. Connecting the computation of RSRP and RSSI, RSRQ can be formulated as $\frac{p_i g_{i,j}}{\sum_{i \in \mathcal{B} \bigcup \mathcal{R}} p_i g_{i,j} \rho_i + \sigma_j}$. It reflects the quality of the link and approximates the UE experienced SINR.

Conventionally, simple cell selection criteria are applied in the network such that the cell with the strongest RSRP or RSRQ value is selected, aiming at the best spectral efficiency and network coverage. Since the fluctuation of channel results in fast varying physical layer measurements, certain mechanisms are employed to improve the stability of the cell selection procedures. For instance, measurements averaging, Time to Trigger (TTT) and different handover thresholds are defined for intra-RAT, inter-RAT and inter-operator handovers, in order to avoid unnecessary

ping-pongs as well as to achieve load balancing purposes [24, 25]. Moreover, radio measurements are carried out with different intervals for different types of handovers, in accordance with the handover frequencies to reduce unnecessary measurements.

In general, the cell selection, re-selection and handover criterion can be formulated as

$$i = \text{argmax}_i(m_i + h_i), \tag{5.1}$$

where i is the cell to be selected and m_i is the description of the selection criterion, which can be radio measurement values, e.g., RSRP or RSRQ. Furthermore, h_i is a cell specific threshold which carries cell selection preferences and user context information. In this work, we ignore the specific offset, i.e., $h_i = 0$ and focus on choosing a proper criterion m_i for potential energy saving in the nomadic relaying networks. On the other hand, *admission control* and *switch on/off* mechanisms are applied at the cell such that overload avoidance is granted and site shut-down for energy saving can be achieved.

5.2. Cell Selection and Admission Control for Energy Savings

5.2.1. Cell Selection Criterion

Consider the optimization problem in (4.22) and aim at performing practical distributed algorithms. The constraints in (4.22b) and (4.22d) can be fulfilled by running distributed cell selection algorithms, i.e., each UE or RN chooses one access point based on the available measurements. By doing this, the assignment $\mathbf{X}$ is chosen in a decentralized manner such that the equality constraint in (4.22b) and (4.22d) is satisfied with discrete assignments on $\{0, 1\}^{(M+K)\times(N+K)}$.

The constraint in (4.22c) is addressed in the next subsection. In this subsection, we aim at finding the criteria m_i for improving the energy efficiency. As the first step, consider only the static energy consumption. Denote t to be the index for an iteration step or the time slot of the iteration step. Let $\boldsymbol{\rho}^{(\text{t})} \in \mathbb{R}^{M+K}$ be the load vector at iteration step t. Then, the distributed cell selection criterion for energy saving in time slot (t) can be formulated as in (i)-(iii):

(i) Each cell (BS or RN) i broadcasts its current load $\rho_i^{(\text{t})}$ and other standard LTE reference signals.

(ii) Each RN j selects a BS from the set of non-overloading and accessible BSs (denoted as $\bar{\mathcal{B}}_j$) by solving the following problem:

$$i = \operatorname{argmin}_{i \in \bar{\mathcal{B}}_j} \theta_{i,j}, \tag{5.2}$$

where

$$\theta_{i,j} = \frac{1}{(\rho_i^{(\mathrm{t})} + \epsilon) \omega_{i,j} b_i}. \tag{5.3}$$

(iii) Each UE j selects a cell i from the set of non-overloading and accessible BSs and RNs ($\bar{\mathcal{B}}_j$ and $\bar{\mathcal{R}}_j$) according to:

$$i = \operatorname{argmin}_{i \in \bar{\mathcal{B}}_j \bigcup \bar{\mathcal{R}}} \phi_{i,j} \tag{5.4}$$

where

$$\phi_{i,j} = \begin{cases} \frac{1}{(\rho_i^{(\mathrm{t})} + \epsilon) \omega_{i,j} b_i} & i \in \bar{\mathcal{B}}_j \\ \frac{1}{(\rho_i^{(\mathrm{t})} + \epsilon) \omega_{i,j} b_i} + \frac{\sum\limits_{h \in \bar{\mathcal{B}}} x_{hi}}{(\rho_h^{(\mathrm{t})} + \epsilon) \omega_{h,j} b_h} & i \in \bar{\mathcal{R}}_j. \end{cases} \tag{5.5}$$

Herein, the set of non-overloading and accessible cells is identified by taking into account both the broadcasted load $\rho_i^{(\mathrm{t})}$ and the resource consumption (denoted by $\rho_{i,j}^{(\mathrm{t})}$) for cell i to satisfy node j through link (i, j). If the remaining amount of resources of cell i supports to serve node j through link (i, j), i.e., if

$$1 - \rho_i^{(\mathrm{t})} \geq \rho_{i,j}^{(\mathrm{t})}, \tag{5.6}$$

then cell i belongs to the set of candidates.

Note that the criterion also applies to the legacy network or femtocell deployment, where no RN exists and the step (ii) can be jumped over. In this case, the cell selection criterion can be understood as choosing a cell with the largest product of link SINR and cell load, i.e., $i = \operatorname{argmax}_{i \in \bar{\mathcal{B}}} \rho_i \omega_{i,j}$. This modifies the conventional cell selection methods which take mostly only the link quality into account. The criterion adds bias to camp on a cell with higher load, so that the cell with low load are able to further handover nodes to other cells to reduce its own load. If a cell does not serve any node, it can be turn-off for power saving purposes.

5.2.2. Admission Control Scheme

The cell selection algorithm may violate the constraint in (3.4c), since the broadcasted load information of a cell may become outdated after other nodes have connected to the cell due to uncoordinated accesses. Therefore, we introduce an admission control mechanism at the cells: the access is only allowed when the constraint $\rho_i < 1$ holds after attaching the node, otherwise, a rejection will be sent and the RN or UE should keep the previous cell connection such that:

$$
x_{i,j}^{(t+1)} = \begin{cases} 1, & \text{if cell } i \text{ is selected and } \rho_i^{(t,c)} + \rho_{i,j}^{(t)} \leq 1 \\ x_{i,j}^{(t)}, & \text{otherwise,} \end{cases}
\tag{5.7}
$$

where $\rho_i^{(t,c)}$ is the locally measured real load of cell i at step (t) after a certain number of random accesses have been performed. Note that $\rho_i^{(t,c)}$ is different than the broadcasted load information $\rho_i^{(t)}$, since other nodes may access or leave the cell before the load information can be updated.

5.2.3. Switch-on/off Mechanism

Furthermore, in order to save static energy, sleeping mode or cell switch-off mechanisms can be introduced. In order to enter the sleeping mode or to switch-off a cell i, we propose that the following threshold based condition must be fulfilled:

$$
\rho_i^{(t)} \leq h_s, \text{for } t \in [t_0 - t_s, t_0],
\tag{5.8}
$$

where h_s should be a small positive value of the load threshold. Furthermore, t_s and t_0 are TTT and the current time stamp, respectively. This simple mechanism enables a cell to enter sleep mode at t_0 if it has low load (below h_s) for a certain period of time (t_s). If a cell has entered sleep mode, parts of its components can still be actively sending reference signals to inform its presence as an access point to the UEs and RNs. The cell will become active again if a certain number of access attempts are observed. However, from the formulation of (5.3) and (5.5), a load value of zero means a very low ranking for cell selection. Therefore, an access attempt to that cell only happens if there are no other active candidate cells. If a cell is completely shut-down, it will not send any reference signals and UEs are not able to detect the cell. In this case, some location based database can be used to map the corresponding channel and signal quality [92]. Nevertheless, there must be at least one component that actively waits for a control command to activate the cell.

5.3. Algorithm and Convergence Analysis

In this section, we propose distributed algorithms and analyze the convergence performance for both the static and the dynamic interference models. We show that the approximation of the original objective function monotonically decreases. Furthermore, we show numerical results to confirm that the original objective has the identical behavior as the approximation.

5.3.1. Static Interference Model

Assuming the static interference model, we can formulate the distributed algorithm for energy saving by summarizing all the three functional blocks (5.2.1-5.2.3) as in Algorithm 6 and Algorithm 7. Whereas we only consider the static energy consumption in Algorithm 6, Algorithm 7 takes into account as well the dynamic energy consumption. Note that Algorithm 7 is a generalization of Algorithm 6 and the traditional RSRP/RSRQ based cell selection algorithm. If $\mathbf{d} = \mathbf{0}$, the algorithm is equivalent to Algorithm 6, whereas it becomes the conventional algorithm if $\mathbf{c} = \mathbf{0}$.

Proposition 5.1. The proposed algorithms (Algorithm 6 and Algorithm 7) converge and iteratively minimizes the approximation of the objective function under the static interference model as:

$$\tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t+1)}) \leq \tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t)}), \tag{5.9}$$

where $\boldsymbol{\rho}^{(t)} = \mathbf{F}(\mathbf{x}^{(t)})$ with $\mathbf{x}^{(t)}$ denoting the network assignment at step (t).

Proof. Since $\tilde{U}(\boldsymbol{\rho})_{\boldsymbol{\rho}} \geq 0$ holds for all $\boldsymbol{\rho} \geq \mathbf{0}$, it suffices for convergence in objective to show that the sequence $\{\tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t)})\}_{t=0}^{\infty}$ is non-increasing. In order to prove this, we denote $\boldsymbol{\rho}^{(t')}$ and $x_{i,j}^{(t')}$ to be, respectively, the intermediate states of the load state vector and the assignment of link (i, j) after step (ii) and before step (iii) in time slot (t). Then, by the strict concavity of $\tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho})$ given by (4.38), we have:

$$\tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t')}) - \tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t)}) \leq \mathbf{J}_{\tilde{U}}^{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t)})(\boldsymbol{\rho}^{(t')} - \boldsymbol{\rho}^{(t)}). \tag{5.10}$$

Algorithm 6 Distributed Cell Selection and Admission Control for Static Energy Saving (DCAS)

Let t = 0, initialize $\mathbf{x}$ with $\mathbf{x}^{(t)}$, calculate $\rho^{(t)}$
loop
  @RNs/BSs: Broadcasting $\rho^{(t)}$
  @RNs: selecting a BS according to (5.3)
  @UEs: selecting a BS or an RN according to (5.5)
  @RNs/BSs: performing admission control according to (5.7)
  @RNs/BSs: performing switching-on/off according to (5.8)
  $t \leftarrow t + 1$
end loop

Algorithm 7 Distributed Cell Selection and Admission Control for Generic Energy Saving (DCAG)

Let t = 0, initialize $\mathbf{x}$ with $\mathbf{x}^{(t)}$, calculate $\rho^{(t)}$
loop
  @RNs/BSs: Broadcasting $\rho^{(t)}$
  @RNs: selecting a BS according to
    $i = \operatorname{argmin}_{i \in \bar{\mathcal{B}}_j} \theta_{i,j}, \quad \theta_{i,j} = \frac{k(\rho_i)}{\omega_{i,j}b_i}$ with $k(\rho_i) = c_i/(\rho_i + \epsilon)(\log(1 + \epsilon^{-1})) - d_i$
  @UEs: selecting a BS or an RN according to
    $$i = \operatorname{argmin}_{i \in \bar{\mathcal{B}}_j \bigcup \bar{\mathcal{R}}} \phi_{i,j}, \quad \phi_{i,j} = \begin{cases} \theta_{i,j} & i \in \bar{\mathcal{B}}_j \\ \frac{k(\rho_i)}{\omega_{i,j}b_i} + \sum\limits_{h \in \bar{\mathcal{B}}} x_{hi}\frac{k(\rho_i)}{\omega_{h,i}b_h} & i \in \bar{\mathcal{R}}_j \end{cases}$$
  @RNs/BSs: performing admission control according to (5.7)
  @RNs/BSs: performing switching-on/off according to (5.8)
  $t \leftarrow t + 1$
end loop

The right hand side of (5.10) can be further written as:

$$\mathbf{J}_{\tilde{\mathbf{U}}}^{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t)})(\boldsymbol{\rho}^{(t')} - \boldsymbol{\rho}^{(t)}) = \sum_{i \in \mathcal{B}}\left(\frac{c_i}{(\rho_i^{(t)} + \epsilon)\log(1 + \epsilon^{-1})} + d_i\right)\sum_{j \in \mathcal{R}}\frac{r_j^{(k)}(x_{i,j}^{(t')} - x_{i,j}^{(t)})}{b_i\omega_{i,j}} \quad (5.11)$$

$$= \sum_{i \in \mathcal{B}}\sum_{j \in \mathcal{R}}\frac{r_j^{(k)}k(\rho_j)(x_{i,j}^{(t')} - x_{i,j}^{(t)})}{b_i\omega_{i,j}} \quad (5.12)$$

$$= \sum_{i \in \bar{\mathcal{B}}_j}\sum_{j \in \tilde{\mathcal{R}}}\frac{r_j^{(k)}k(\rho_j)(x_{i,j}^{(t')} - x_{i,j}^{(t)})}{b_i\omega_{i,j}}, \quad (5.13)$$

where

$$k(\rho_i) = c_i/(\rho_i + \epsilon)(\log(1 + \epsilon^{-1})) - d_i.$$

Note that $\tilde{\mathcal{R}}$ is the set of admitted RNs, and the last equality holds since the admission control rule at the BSs ensures that $x_{i,j}^{(t')} - x_{i,j}^{(t)} = 0$ for $j \notin \tilde{\mathcal{R}}$ and the cell selection criterion makes both $x_{i,j}^{(t')} = x_{i,j}^{(t)} = 0$ if $i \notin \bar{\mathcal{B}}_j$.

Furthermore, we have for any RN j in $\tilde{\mathcal{R}}$:

$$\sum_{i \in \bar{\mathcal{B}}_j} \frac{k(\rho_j)(x_{i,j}^{(t')} - x_{i,j}^{(t)})}{b_i \omega_{i,j}} \leq 0, \tag{5.14}$$

since only one BS is selected and the index i corresponding to $x_{i,j}^{(t')} = 1$ is the minimizer of $\frac{k(\rho_j)}{b_i \omega_{i,j}}$. This shows that $\tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t')}) - \tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t)}) \leq 0$ holds. Without details of proof, $\tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t+1)}) - \tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t')}) \leq 0$ yields analogously after step (iii). Therefore, $\tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t+1)}) \leq \tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t')}) \leq \tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t)})$ and Algorithm 7 iteratively reduces the energy consumption of the network. Note that since Algorithm 6 can be understood as a special case of Algorithm 7, we can also conclude the convergence of Algorithm 6. □

5.3.2. Dynamic Interference Model

With the dynamic interference model, the spectral efficiency changes as the iterations continue, i.e., $\omega_{i,j}^{(t)} = \omega_{i,j}(\boldsymbol{\rho}^{(t)}) \neq \omega_{i,j}^{(t+1)}$. Therefore, the derivation in (5.11) is in general not true under the dynamic interference model. However, we can modify the rules for admission control by making use of Proposition 4.10 to ensure the monotonic decrease in the objective approximation.

The algorithm for the dynamic interference model is given in Algorithm 8. Note that it can be seen as the generalization of Algorithm 7, since it becomes Algorithm 7 if we take $\mathbf{1}$ instead of $\boldsymbol{\rho}^{(t)}$ for calculating the spectral efficiency and performing admission control.

Algorithm 8 Generic Distributed Cell Selection and Admission Control GDCA

Let $t = 1$, initialize $\mathbf{x}$ with $\mathbf{x}^t$, $\mathbf{x}$ with $\boldsymbol{\rho}^t = \mathbf{1}$
loop
　@RNs/BSs: Broadcasting $\boldsymbol{\rho}^{(t)}$
　@RNs: selecting a BS according to
　　　$i = \text{argmin}_{i \in \bar{\mathcal{B}}_j} \theta_{i,j}, \quad \theta_{i,j} = \frac{k(\rho_i^{(t)})}{\omega_{i,j}(\boldsymbol{\rho}^{(t)})b_i}$ with $k(\rho_i^{(t)}) = c_i / (\rho_i^{(t)} + \epsilon)(\log(1 + \epsilon^{-1})) - d_i$
　@UEs: selecting a BS or an RN according to
　　　$i = \text{argmin}_{i \in \bar{\mathcal{B}}_j \bigcup \bar{\mathcal{R}}} \phi_{i,j}, \quad \phi_{i,j} = \begin{cases} \theta_{i,j} & i \in \bar{\mathcal{B}}_j \\ \frac{k(\rho_i^{(t)})}{\omega_{i,j}(\boldsymbol{\rho}^{(t)})b_i} + \sum_{h \in \bar{\mathcal{B}}} x_{hi} \frac{k(\rho_i^{(t)})}{\omega_{h,i}(\boldsymbol{\rho}^{(t)})b_h} & i \in \bar{\mathcal{R}}_j \end{cases}$
　@RNs/BSs: performing admission control such that $\boldsymbol{\rho}^{(t+1)} \leq \boldsymbol{\rho}^{(t)}$
　@RNs/BSs: performing switching-on/off according to (5.8)
　$t \leftarrow t + 1$
end loop

Proposition 5.2. Algorithm 8 is a distributed energy saving algorithm that has the following

features: (i) Feasibility: the constraint $\boldsymbol{\rho} \leq \mathbf{1}$ is always fulfilled; (ii) Convergence: approximation of the objective function decreases iteratively.

Proof. Since the admission control ensures that $\boldsymbol{\rho}^{(t+1)} \leq \boldsymbol{\rho}^{(t)} \leq \ldots \leq \mathbf{1}$, the feasibility condition in (i) is obviously true. For (ii), we need to prove the convergence in the objective, i.e., $\tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t+1)}) - \tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t)}) \leq 0$. This can be done analogously as the proof of Proposition 5.1. Consider the first sub-step of iteration (t) when establishing the access link connections between RNs and UEs. According to Proposition 4.11, we can formulate the majorization functions of $\tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho})$ over $\mathbf{0} \leq \boldsymbol{\rho} \leq \boldsymbol{\rho}^{(t)}$ as

$$\tilde{U}_{\boldsymbol{\rho}}(\boldsymbol{\rho}^{(t)}) + \mathbf{J}_{\tilde{U}}^{\mathbf{x}}(\mathbf{x}^{(t)})(\mathbf{x}^{(t')} - \mathbf{x}^{(t)}). \tag{5.15}$$

Therefore, if the admission control ensures that $\mathbf{0} \leq \boldsymbol{\rho}^{(t+1)} \leq \boldsymbol{\rho}^{(t)}$, we have:

$$
\begin{aligned}
\tilde{U}(\boldsymbol{\rho}^{(t')}) - \tilde{U}(\boldsymbol{\rho}^{(t)}) &\leq \mathbf{J}_{\tilde{U}}^{\mathbf{x}}(\mathbf{x}^{(t)})(\mathbf{x}^{(t')} - \mathbf{x}^{(t)}) \\
&= \sum_{i \in \mathcal{B}} \Big(\frac{c_i}{(\rho_i^{(t)} + \epsilon) \log(1 + \epsilon^{-1})} + d_i \Big) \sum_{j \in \mathcal{R}} \frac{r_j^{(k)}(x_{i,j}^{(t')} - x_{i,j}^{(t)})}{b_i \omega_{i,j}(\boldsymbol{\rho}^{(t)})} \\
&= \sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{R}} \frac{r_j^{(k)} k(\rho_i^{(t)})(x_{i,j}^{(t')} - x_{i,j}^{(t)})}{b_i \omega_{i,j}(\boldsymbol{\rho}^{(t)})} \\
&= \sum_{i \in \bar{\mathcal{B}}_j} \sum_{j \in \tilde{\mathcal{R}}} \frac{r_j^{(k)} k(\rho_i^{(t)})(x_{i,j}^{(t')} - x_{i,j}^{(t)})}{b_i \omega_{i,j}(\boldsymbol{\rho}^{(t)})} \\
&\leq 0.
\end{aligned}
$$

Without details of proof, we can conclude that $\tilde{U}(\boldsymbol{\rho}^{(t+1)}) \leq \tilde{U}(\boldsymbol{\rho}^{(t')}) \leq \tilde{U}(\boldsymbol{\rho}^{(t)})$, and therefore, the objective monotonically decreases. $\square$

Note that a cell does not allow to increase its load. This means a node is only accepted by the cell if the total load decreases compared with the last iteration step. In practical networks, bursting user data accesses may happen, whereas the algorithm denies new accesses such that new users are not allowed to join the network. Therefore, a relaxation of the admission control mechanism may be required to perform practical optimizations. If we assume that the load and interference situation remains at the same level between two consecutive iteration steps, i.e., $\boldsymbol{\rho}^{(t+1)} \approx \boldsymbol{\rho}^{(t)}$, we can directly apply the admission control scheme in Algorithm 7, i.e., each cell only controls if a newly attached node leads to overloading. In this case, we indeed assume a static interference scenario between the two iteration steps.

In Fig. 5.1, the numerical results of the convergence performance of the Algorithm 8 is illustrated (Simulation Scenario is described in Section 5.4). Both the objective and the approx-
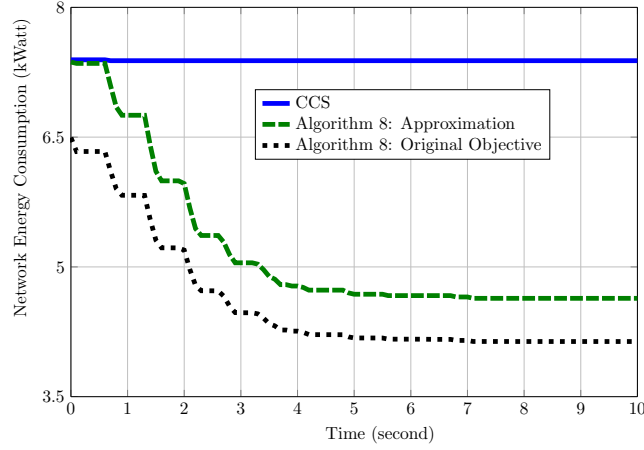
Figure 5.1.: Convergence of the algorithm.

imation monotonically decrease by iteration steps. Furthermore, it can be observed that the approximation and the real total energy consumption are close to each other and have similar convergence behavior using the proposed algorithm. Note that we choose as for comparison the Closest Cell Selection (CCS) algorithm in which the UEs and RNs select the closest cell for data transmission. The CCS algorithm tends to activate all the BSs, since the UEs and RNs are uniformly distributed and are trying to connect to the cells close to them.

5.4. Performance Evaluation

We evaluate the proposed algorithms in a nomadic relaying network with 7 BSs in a hexagon layout and with an ISD of 1000 m. In the coverage of the BSs, 150 UEs and a certain number of nomadic RNs are randomly dropped according to a uniform distribution. The other network and transmission parameters for the simulation are taken as in Table 6.1 in Chapter 4. The static and dynamic energy consumption for an active BS is assumed to be 1 kWatt and 0.1 kWatt, respectively, while an active RN consumes 10 Watt static and 1 Watt dynamic energy. Both BSs and RNs are allocated 10 MHz bandwidth at 2 GHz, whereas the transmission powers of the BSs and the RNs are assumed to be 46 dBm and 23 dBm, respectively. Directional antennas are equipped at the BSs and omni-directional antennas are assumed for the RNs. Furthermore, the radio propagation model is chosen according to the 3GPP recommendations [11] and the noise figure is set to 5 dB at all nodes in the network. The time interval for broadcasting system information and for running the distributed algorithm is chosen to be 100 ms. We ignore the time for further signaling procedures and assume cell selection, handover and admission control can be successfully done within this time interval. Furthermore, we perform 200 iterations to
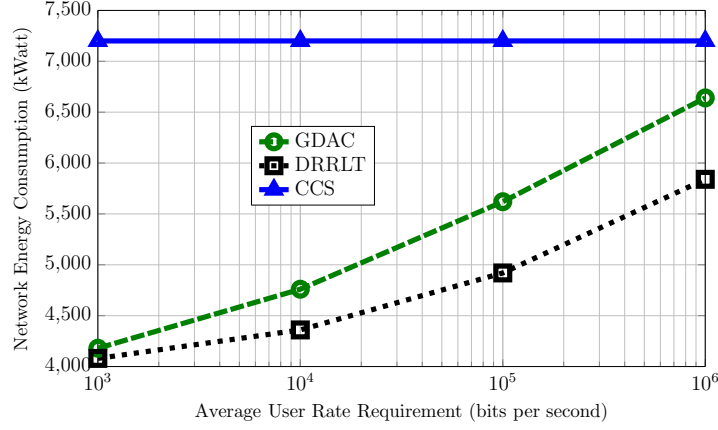
Figure 5.2.: Network energy consumption against user rate requirements.

average the statistical uncertainty.

The following algorithms, including the conventional cell selection algorithm (CCS), centralized algorithm (DRRLT) and distributed algorithm (GDCA), are compared.

- CCS: the nodes select the closest cell for data transmission;

- DRRLT: as proposed in Algorithm 5 in Chapter 4;

- GDCA: as proposed in Algorithm 8.

5.4.1. Impact of UE QoS Requirements

First, we choose to evaluate the network performance against different UE data requirements that vary from 1kbps to 1 Mbps. In Fig. 5.2, it can be seen that the proposed algorithm significantly reduce the energy consumption compared with the conventional cell selection algorithm. It can be seen that both the centralized DRRLT and the distributed GDCA significantly reduce the total energy consumption. Particularly in the low rate region, around 40% energy savings can be expected. As the average rate requirement of UE increases, the energy saving gain decreases since the BSs are becoming more and more overloaded. However, the centralized algorithm may save more energy compared with the distributed algorithm if the user average rate requirement is high. This can be attributed to fact that if the cell is highly loaded and multiple access requests from different nodes reach it, the cell cannot choose the correct UE according to GDCA to globally optimize the total network energy consumption.
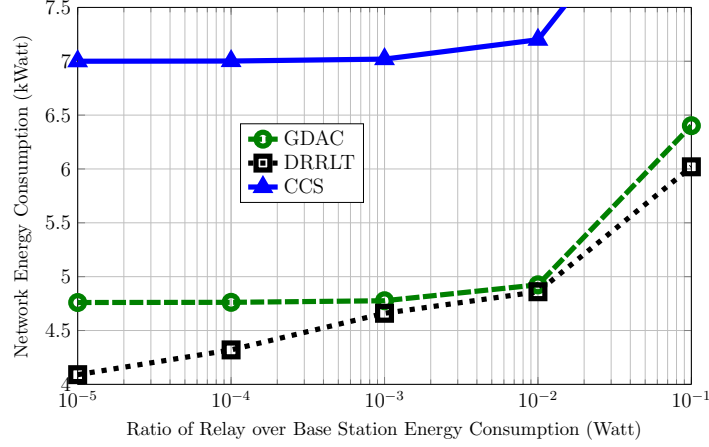
Figure 5.3.: Network energy consumption against per relay energy consumption.

Further evaluations are based on the configurations of the nomadic RNs, including the density of nomadic RNs, the per RN energy consumption and the RN backhaul link SINR gain. These evaluations set up the basic requirements for designing the nomadic network and relay infrastructure in order to achieve performance gain in terms of energy savings.

5.4.2. Impact of the Energy Consumption Model

Fig. 5.3 depicts the energy saving performance with respect to the varying RN static energy consumptions ([0.01 0.1 1 10 100] Watt), where an average rate requirement of 10 kbps is assumed. Logically for both algorithms, the total energy consumption increases as the RN energy consumption increases. If each active RN consumes lower than 10 Watt, corresponding to 1% of the energy consumption of a BS, the energy saving performance is close to the case when RN causes no energy consumption. The total energy consumption only significantly increases when an RN reaches a level of 10% energy consumption of a BS. Note that the energy consumption of a low power node is in practice at the level of 10 Watt, and therefore, the concept of nomadic network has the potential to save energy in a realistic network.

5.4.3. Impact of Relay Density and Antenna

In Fig. 5.4 (a), the energy saving performance is depicted for different densities of RNs, where the UE average rate requirements is set to 100 kbps. It can be easily concluded from the figure that higher RN density results in more energy savings by using the proposed algorithm. In

(a) Impact of relay density
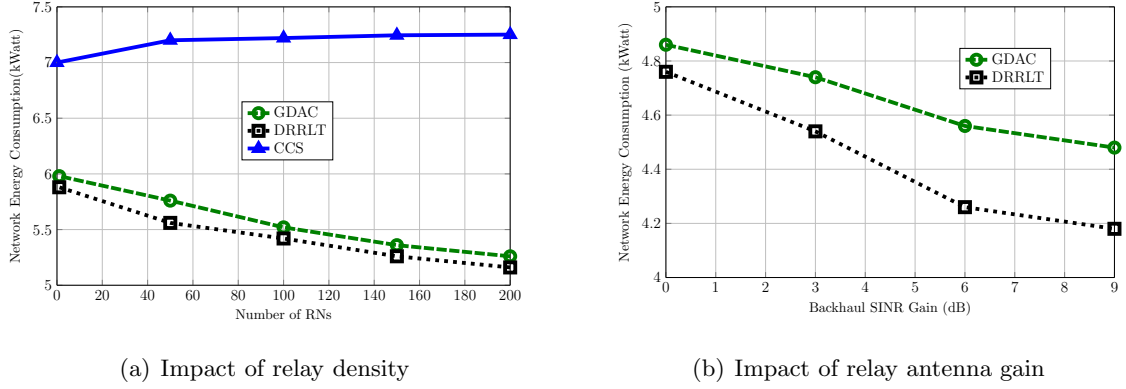
(b) Impact of relay antenna gain

Figure 5.4.: Impact of nomadic relay density and antenna gains.

contrast, higher RN density leads to slightly more energy consumption by using the SINR based algorithm, since more RNs are activated without switching off any BSs. As the number of RNs increases from 50 to 200, around 10% less energy consumption is required to the support the QoS of all the UEs. The performance improvement is due to the fact that the extra RNs imply a higher possibility of having suitable RNs to redirect data traffics for energy saving purposes. It is worth noting that the proposed the distributed algorithm performs comparably to the centralized algorithm which relies strongly on channel feedbacks and computational complexity.

At last, we evaluate the network performance considering another important aspect - the RN backhaul performance gain. Since more space is available for designing the vehicle antennas, the backhaul link performance gain can be achieved through smart antenna designs or advanced signal processing techniques such as MIMO and interference cancellation. In Fig. 5.4 (b), we assume that the backhaul SINR gain varies from 0 dB to 9 dB, where the average rate requirement is 10 kbps. The backhaul link SINR gain increases the BS coverage on one hand, and reduces the backhaul link resource consumption on the other hand. Hence, it can be seen that 10% more energy savings in a low rate scenario can be achieved if 9 dB the backhaul link gain can be achieved. With the centralized DRRLT algorithm, more energy saving potentials, however, limited up to 10%, can be achieved with higher antenna gains compared with the proposed distributed algorithm in this Chapter. Therefore, the distributed algorithm is a more practical and suitable implementation for energy saving in the nomadic relaying networks.

Chapter 6.

Distributed Power Control with Active Cell Protection

This chapter is written based on [20] and we focus on the power control problem for reducing the dynamic energy consumption. First, we introduce a new SIF - the power interference function, based on which, we derive the feasibility and optimality condition for the energy saving problem. Then, we propose a distributed power control algorithm to keep the active cells remaining active during the transient phase of the activation procedure, i.e., they have enough resources to support the QoS of the connected nodes. The convergence of the algorithm is analyzed under both the static and the dynamic interference model. Furthermore, the algorithm is enhanced by adaptively updating a control parameter to avoid violating the power constraints in practical system.

6.1. Power, Interference and Load Coupling

6.1.1. Explicit Power Load Function

In chapter 4, we have proven the existence of an explicit load function that takes only the assignment vector $\mathbf{x}$ as its input argument. Fixing the assignment, we can have a similar conclusion on the existence of an *explicit power load function*.

Proposition 6.1. Let $\mathcal{P} : \{\mathbf{p} \in \mathbb{R}_+^{M+K} | \exists_{\boldsymbol{\rho} \geq \mathbf{0}}, \boldsymbol{\rho} > \mathbf{F}(\boldsymbol{\rho}, \mathbf{p})\} \neq \emptyset$. Then, exists a *continuous differentiable function* $\mathbf{H} : \mathcal{P} \to \mathbb{R}_+^{M+K}$ such that $\boldsymbol{\rho} = \mathbf{H}(\mathbf{p})$ takes power vector $\mathbf{p}$ as argument.

Proof. The proof is omitted since it can be done the same way as the proof of Proposition 4.3 and Proposition 4.4. ☐

Lemma 6.1. If $\alpha > 1$ and $\mathbf{p} > \mathbf{0}$, $\mathbf{H}(\alpha\mathbf{p}) < \mathbf{H}(\mathbf{p})$ element-wise.

Proof. Let $\boldsymbol{\sigma} \in \mathbb{R}_{++}^{N+K}$ denote the vector that contains the receiver-side noise. Then, we extend the arguments of $\mathbf{H}$ to $\boldsymbol{\rho}$, $\mathbf{p}$ and the noise vector $\boldsymbol{\sigma}$, i.e., $\mathbf{F}(\boldsymbol{\rho}, \mathbf{p}) = \mathbf{F}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma})$. Furthermore, let $\mathbf{H}(\mathbf{p}, \boldsymbol{\sigma})$ denote the explicit power load function implied by

$$\tilde{\mathbf{F}}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma}) \triangleq \boldsymbol{\rho} - \mathbf{F}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma}) = \mathbf{0}. \tag{6.1}$$

According to the implicit function theorem,

$$\mathbf{J}_{\mathbf{H}}^{\boldsymbol{\sigma}}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma}) = -\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\rho}}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma})^{-1}\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\sigma}}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma}), \tag{6.2}$$

where

$$\mathbf{J}_{\mathbf{F}_i}^{\sigma_j}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma}) = -\frac{r_j x_{i,j}}{b_i \omega_{i,j}(1 + \tau_{i,j})(\sum_{d \neq i} p_d g_{d,j} \rho_d s_{d,j} + \sigma_j)} \tag{6.3}$$

According to [18, Prop.2], the inverse of $\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\rho}}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma})$ exists with only non-negative elements. Moreover, it can be easily concluded from (6.3) that $\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\sigma}}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma})$ is a matrix with only negative entries. Thus, $\mathbf{J}_{\mathbf{H}}^{\boldsymbol{\sigma}}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma}) > \mathbf{0}$ for $\boldsymbol{\sigma} > 0$, indicating a monotonically increasing property on $\boldsymbol{\sigma}$ according to Lemma 4.1. Therefore, for $\alpha > 1$,

$$\mathbf{H}(\alpha\mathbf{p}, \boldsymbol{\sigma}) = \mathbf{H}(\mathbf{p}, \boldsymbol{\sigma}/\alpha) < \mathbf{H}(\mathbf{p}, \boldsymbol{\sigma}). \tag{6.4}$$

The equality yields since $\mathbf{F}(\boldsymbol{\rho}, \alpha\mathbf{p}, \boldsymbol{\sigma}) = \mathbf{F}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma}/\alpha)$, and therefore, $\boldsymbol{\rho} = \mathbf{F}(\boldsymbol{\rho}, \alpha\mathbf{p}, \boldsymbol{\sigma})$ and $\boldsymbol{\rho} = \mathbf{F}(\boldsymbol{\rho}, \mathbf{p}, \boldsymbol{\sigma}/\alpha)$ have the same fixed-point, i.e., $\boldsymbol{\rho} = \mathbf{H}(\alpha\mathbf{p}, \boldsymbol{\sigma}) = \mathbf{H}(\mathbf{p}, \boldsymbol{\sigma}/\alpha)$. ☐

6.1.2. Load Interference Function

We define $\mathbf{I}(\mathbf{p}) = [I_1(\mathbf{p}), \ldots, I_{M+K}(\mathbf{p})] : \mathbb{R}_+^{M+K} \to \mathbb{R}_{++}^{M+K}$ to be the vector of the *Load Interference Functions*:

$$I_i(\mathbf{p}) = \begin{cases} p_i \rho_i(\mathbf{p}) & \text{for} \quad p_i > \mathbf{0}, \\ \sum_{j \in \mathcal{U} \bigcup \mathcal{R}} r_j x_{i,j} \frac{\sum_{d \in \mathcal{B} \bigcup \mathcal{R}}(p_d g_{d,j} s_{d,j} \rho_d((\mathbf{p})) + \sigma_j)}{b_i g_{i,j}} & \text{otherwise.} \end{cases} \tag{6.5}$$

where $\rho_i(\mathbf{p})$ is the load function defined by the static or the dynamic interference model (if exists).

Lemma 6.2. $I_i(\mathbf{p})$ is a well defined *positive continuous* function: $\mathbb{R}_+^n \to \mathbb{R}_{++}$.

Proof. For the static interference model, it is obvious that $I_i(\mathbf{p})$ is positive for $\mathbf{p} > \mathbf{0}$. Furthermore, it is continuous since both $\mathbf{p}_i$ and $\rho_i(\mathbf{p})$ are continuous functions. For the dynamic interference model, we know from Lemma 6.1 and Proposition 4.4 that if $\mathbf{H}(\mathbf{p})$ exists, it is positive and continuous differentiable. Therefore, $I_i(\mathbf{p})$ is positive and continuous for $\mathbf{p} > \mathbf{0}$. Then, we follow L' Hospital's rule for $p_i = 0$:

$$\lim_{p_i \to 0^+} I_i(\mathbf{p}) = \lim_{p_i \to 0^+} \sum_{j \in \mathcal{U}_i \bigcup \mathcal{R}_i} \frac{r_j(1 + \frac{g_{i,j}p_i}{\sum_{d \in \mathcal{I}_i} p_d g_{d,j} \rho_d(\mathbf{p}) + \sigma_j})}{b_i \frac{g_{i,j}}{(\sum_{d \in \mathcal{I}_i} p_d g_{d,j} \rho_d(\mathbf{p}) + \sigma_j)}} = I_i(\mathbf{p})|_{p_i = 0} > 0.$$

Thus, $I_i(\mathbf{p})$ is a positive continuous function for $\mathbf{p} \geq \mathbf{0}$. $\qquad\square$

Lemma 6.3. If $I_i(\mathbf{p})$ is an SIF as in Definition 4.5 for $\mathbf{p} > \mathbf{0}$, it is also an SIF for $\mathbf{p} \geq \mathbf{0}$.

Proof. Suppose $\mathcal{C}$ is the set of cells with zero transmission power, i.e. $p_i = 0$, $\forall i \in \mathcal{C}$. For monotonicity in Definition 4.1, we have $\forall \mathbf{p}' \geq \mathbf{p}$ and $\forall i \in \mathcal{C}$ that

$$I_i(\mathbf{p}') \geq \lim_{p_i \to 0^+, \forall i \in \mathcal{C}} I_i(\mathbf{p}) = I_i(\mathbf{p}),$$

since I_i is monotonically increasing on $\mathbf{p} > \mathbf{0}$. For scalability in Definition 4.3, we have due to the scalability on $\mathbf{p} > \mathbf{0}$ that for $i \in \mathcal{C}$:

$$I_i(\alpha \mathbf{p}) = \lim_{p_i \to 0^+} I_i(\alpha \mathbf{p}) < \alpha \lim_{p_i \to 0^+} I_i(\mathbf{p}) = \alpha I_i(\mathbf{p}).$$

Furthermore, since the function $I_i(\mathbf{0}) > 0$, I_i satisfies all the conditions for Definition 4.5 which completes the proof. $\qquad\square$

Proposition 6.2. $\mathbf{I}(\mathbf{p})$ is an SIF in $\mathbf{p} \geq \mathbf{0}$ assuming the static interference model.

Proof. Since r_j, $x_{i,j}$, $s_{i,j}$, b_i and $g_{i,j}$ are all positive constants, it suffices to investigate $\tilde{I}_i(\mathbf{p}) = p_i / \log(1 + \frac{p_i}{\sum_{d \in \mathcal{I}_i} p_d + \sigma_j})$. Due to Lemma 6.3, we only need to prove that $\tilde{I}_i(\mathbf{p})$ satisfies Definition 4.5 for $\mathbf{p} > \mathbf{0}$.

Let $g = \log(1 + \frac{p_i}{\sum_{d \neq i} p_d + \sigma_j}) > 0$, then the gradient of $\tilde{I}_i$ can be formulated as:

$$\mathbf{J}_{\tilde{I}_i}^{p_{i'}}(\mathbf{p}) = (g \cdot \frac{\partial p_i}{\partial p_{i'}} - p_i \cdot \frac{\partial g}{\partial p_{i'}})/g^2.$$

For $i' = i$, let $h = g - p_i \cdot \frac{\partial g}{\partial p_i}$, then $\frac{\partial h}{\partial p_i} = \frac{\partial g}{\partial p_i} - \frac{\partial g}{\partial p_i} - p_i \cdot \frac{\partial^2 g}{\partial p_i^2} = \frac{p_i}{(p_i + \sum_{d \neq i} p_d + \sigma_j)^2} > 0$, $\forall p_i > 0$. Therefore, $h > \lim_{p_i \to 0} h = \lim_{p_i \to 0} g = 0$ and $\mathbf{J}_{\tilde{I}_i}^{p_{i'}}(\mathbf{p}) = \frac{h}{g^2} > 0$. For $i' \neq i$, $\mathbf{J}_{\tilde{I}_i}^{p_{i'}}(\mathbf{p}) = \frac{-p_i \cdot \frac{\partial g}{\partial p_{i'}}}{g^2} =$

$\frac{1}{\log^2(1+\frac{p_i}{\sum_{d\neq i}p_d+\sigma_j})}\frac{p_i^2}{(\sum_{d\neq i}p_d+\sigma_j)^2} > 0$. Hence, $\mathbf{J}_{\tilde{I}_i}^{p_{i'}}(\mathbf{p}) > 0$ for every i' and i, which results in the monotonicity of $\tilde{I}_i(\mathbf{p})$.

For scalability, it can verified that $\mathbf{J}_{\tilde{I}_i}^{\sigma_j}(\mathbf{p}) = \frac{1}{g(1+\frac{p_i}{\sum_{d\neq i}p_d+\sigma_j})(\sum_{d\neq i}p_d+\sigma_j)} > 0$. Thus,

$$\tilde{I}_i(\alpha\mathbf{p}) = \frac{\alpha p_i}{\log(1+\frac{p_i}{\sum_{d\in\mathcal{I}_i}p_d+\sigma_j/\alpha})} < \frac{\alpha p_i}{\log(1+\frac{p_i}{\sum_{d\in\mathcal{I}_i}p_d+\sigma_j})} = \alpha\tilde{I}_i(\mathbf{p}).$$

Therefore, $\tilde{I}_i(\mathbf{p})$ satisfies Definition 4.5 for $\mathbf{p} > \mathbf{0}$, indicating that $\mathbf{I}(\mathbf{p})$ is an SIF for $\mathbf{p} \geq \mathbf{0}$. $\quad\square$

Proposition 6.3. $\mathbf{I}(\mathbf{p})$ is an SIF on $\mathcal{P} : \{\mathbf{p} \in \mathbb{R}_+^{M+K} | \exists_{\boldsymbol{\rho}\geq\mathbf{0}}, \boldsymbol{\rho} \geq \mathbf{F}(\boldsymbol{\rho},\mathbf{p})\} \neq \emptyset$ assuming the dynamic interference model.

Proof. According to Lemma 6.3, it suffices to prove the monotonicity and scalability of $I_i(\mathbf{p}) = p_i\mathbf{H}_i(\mathbf{p})$ for $p_i > 0$. For monotonicity, we use the implicit function defined in (6.1) and calculate the Jacobian with respect to $\mathbf{p}$ as

$$\mathbf{J}_{\mathbf{H}}^{\mathbf{p}}(\boldsymbol{\rho},\mathbf{p},\boldsymbol{\sigma}) = -\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\rho}}(\boldsymbol{\rho},\mathbf{p},\boldsymbol{\sigma})^{-1}\mathbf{J}_{\tilde{\mathbf{F}}}^{\mathbf{p}}(\boldsymbol{\rho},\mathbf{p},\boldsymbol{\sigma}). \tag{6.6}$$

We know that $\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\rho}}(\boldsymbol{\rho},\mathbf{p},\boldsymbol{\sigma})$ is a GDM over $\mathcal{P}$ and hence $\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\rho}}(\boldsymbol{\rho},\mathbf{p},\boldsymbol{\sigma})^{-1}$ is element-wise non-positive, i.e., $\mathbf{J}_{\tilde{\mathbf{F}}}^{\boldsymbol{\rho}}(\boldsymbol{\rho},\mathbf{p},\boldsymbol{\sigma})^{-1} \leq \mathbf{0}$. Furthermore, $\mathbf{J}_{\tilde{\mathbf{F}}}^{\mathbf{p}}(\boldsymbol{\rho},\mathbf{p},\boldsymbol{\sigma})$ is non-negative from the proof of Proposition 6.2, and therefore $\mathbf{J}_{\tilde{\mathbf{F}}}^{\mathbf{p}}(\boldsymbol{\rho},\mathbf{p},\boldsymbol{\sigma})$ is element-wise non-negative, i.e., $\mathbf{J}_{\tilde{\mathbf{F}}}^{\mathbf{p}}(\boldsymbol{\rho},\mathbf{p},\boldsymbol{\sigma}) \geq \mathbf{0}$. Thus, $\mathbf{J}_{\mathbf{H}}^{\mathbf{p}}(\boldsymbol{\rho},\mathbf{p},\boldsymbol{\sigma})$ exists as a matrix with only non-negative elements for all $\mathbf{p} > 0$, which satisfies the monotonicity condition according to Lemma 4.1. For scalability, if $\alpha > 1$, using Lemma 6.1 we have easily $I_i(\alpha\mathbf{p}) = \alpha p_i\mathbf{H}(\alpha\mathbf{p}) < \alpha p_i\mathbf{H}(\mathbf{p}) = \alpha I_i(\mathbf{p})$. $\quad\square$

6.1.3. Dynamic Energy Saving Optimization

Having the definition of load interference function in hand and assuming an identical dynamic energy consumption factor (same d_i for all i), we can reformulate the energy saving problem in (3.5) as

$$\min_{\mathbf{p}} \quad \mathbf{p}^T\boldsymbol{\rho} \tag{6.7a}$$

$$\text{subject to} \quad \mathbf{0} \leq \mathbf{p} \leq \hat{\mathbf{p}} \tag{6.7b}$$

$$\boldsymbol{\rho} = \mathbf{F}(\boldsymbol{\rho},\mathbf{p}) \leq \mathbf{1} \tag{6.7c}$$

This is indeed the same optimization problem as in [64], except the constraint (6.7c) is replaced with the condition that a fixed point load exists, i.e., $\boldsymbol{\rho} \geq \mathbf{F}(\boldsymbol{\rho}, \mathbf{p})$. In this work, we consider a stricter and more practical constraint where no overload is allowed in the system.

6.2. Optimal Power Control

In this section, we discuss the feasibility and optimality condition for the power control problem for dynamic energy saving. We first give some preliminary results from the SIR balancing studies and then derive based on the load interference function the conditions for our problem.

6.2.1. Power Control for SINR Balancing

Assume there are L links or L transceiver pairs in an ad-hoc or uplink scenario. Denote link (i, i) as the desired link between transmitter i and receiver i and link (i, j), $i \neq j$ as a interfering link. The power control works in the early time focused on the following SINR balancing problem in single frequency networks:

$$\min_{\mathbf{p}} \quad \mathbf{1}^T \mathbf{p} \tag{6.8a}$$

$$\text{subject to} \quad \tau_i \geq \gamma_i, \quad \text{for } i \in 1, \dots L \tag{6.8b}$$

where γ_i is the SINR target of link (i, i) and the achieved SINR τ_i is computed by:

$$\tau_i = \frac{p_i g_{i,i}}{\sum_{d=1,\dots,L,d \neq i} p_d g_{d,j} + \sigma_j}, \tag{6.9}$$

where $g_{d,j}$ is the channel gain of link (d, j) and σ_j is the receiver side noise at node j.

The condition in (6.8b) can be further written in matrix form as [77]:

$$(\mathbf{I} - \mathbf{D})\mathbf{p} \geq \boldsymbol{\eta}, \tag{6.10}$$

where $\eta_i = \gamma_i \sigma_i / g_{i,i}$ and $\mathbf{D}$ has only non-zero off-diagonal elements which are defined to be $d_{i,j} = g_{i,j} \gamma_i / g_{i,i}$. According to the Perron-Frobenius theory [93], the feasibility condition that a

positive solution exists can be written as

$$r(\mathbf{D}) < 1, \tag{6.11}$$

where $r(\cdot)$ denotes the spectral radius of a matrix, which is defined as the maximum modulus eigenvalue $\max\{|\lambda| : \mathbf{D} - \lambda\mathbf{I} \text{ is singular}\}$). Then, (6.10) has a nonnegative solution:

$$\mathbf{p}^* = (\mathbf{I} - \mathbf{D})^{-1} \cdot \boldsymbol{\eta}. \tag{6.12}$$

The condition in (6.11) ensures the invertability of $(\mathbf{I} - \mathbf{D})$. Furthermore, $\mathbf{p}^*$ is Pareto efficient in the sense that any other solution $\mathbf{p}$ satisfying (6.10) needs at least as much power component-wise [78]. It means, if exists, i.e., (6.11) is fulfilled, $\mathbf{p}^*$ is the unique optimal solution to the power control problem in (6.8).

6.2.2. Energy Saving Load Power Balancing

In the network point of view, the SINR balancing problem turns to be the load balancing problem. In [64, 83], the feasibility condition and optimality condition are formulated based on the existence of the fixed point iteration. We reformulate and extend the theorems to satisfy the load constraint defined in our problem.

Lemma 6.4. Suppose $\mathcal{F} : \mathbb{R}_+^n \to \mathbb{R}_{++}^n$ is concave, continuously differentiable and monotonically increasing, then $\exists \mathbf{x} > \mathbf{0}, \mathcal{F}(\mathbf{x}) = \mathbf{x}$ if $\exists \mathbf{x}^\circ > 0$ such that $r(\mathbf{J}_{\mathcal{F}}^{\mathbf{x}}(\mathbf{x}^\circ)) < 1$.

Proof. Due to the concavity, we have for $\mathbf{x}^\circ > 0$ that

$$\mathcal{F}(\mathbf{x}^\circ) - \mathbf{x}^\circ \mathbf{J}_{\mathcal{F}}^{\mathbf{x}}(\mathbf{x}^\circ) \geq \mathcal{F}(\mathbf{0}) > 0. \tag{6.13}$$

Applying the Perron-Frobenius theory [93], it can be concluded that $\exists \mathbf{x} > 0$, such that $(\mathbf{I} - \mathbf{J}_{\mathcal{F}}^{\mathbf{x}}(\mathbf{x}^\circ))\mathbf{x} = \mathcal{F}(\mathbf{x}^\circ) - \mathbf{x}^\circ \mathbf{J}_{\mathcal{F}}^{\mathbf{x}}(\mathbf{x}^\circ) \geq \mathcal{F}(\mathbf{0}) > 0$. Since $\mathcal{F}$ is concave, we can further conclude that $\exists \mathbf{x} > 0$, such that

$$\mathbf{x} = \mathcal{F}(\mathbf{x}^\circ) + \mathbf{J}_{\mathcal{F}}^{\mathbf{x}}(\mathbf{x}^\circ)(\mathbf{x} - \mathbf{x}^\circ) \geq \mathcal{F}(\mathbf{x}). \tag{6.14}$$

From [38], we know that a concave increasing function is an SIF if it is defined over $\mathbb{R}_+^n \to \mathbb{R}_{++}^n$. Therefore, $\mathcal{F}$ is an SIF, for which the condition in (6.14) implies the existence of a positive fixed point $\mathbf{x}^* = \mathcal{F}(\mathbf{x}^*)$.

$\square$

Proposition 6.4 (Feasibility)**.** Assume the worst-case or the dynamic interference model. Given $\mathbf{p} > \mathbf{0}$ and $\mathbf{x} \geq \mathbf{1}$, $\mathbf{r} \geq \mathbf{0}$, the load constraints $\boldsymbol{\rho} \leq \mathbf{1}$ of the optimization problem in (3.4c) can be fulfilled, if $\exists \mathbf{p} > \mathbf{0}$ such that

$$r(\mathbf{J}_{\mathbf{I}}^{\mathrm{p}}(\mathbf{p})) < 1, \qquad (6.15)$$

where $\mathbf{I}$ is the load interference function defined in (6.5) under the worst-case interference model, i.e., $\mathbf{I} = \mathbf{p} \circ \mathbf{F}(\mathbf{1}, \mathbf{p})$.

Proof. Since $\mathbf{F}(\boldsymbol{\rho}, \mathbf{p})$ is an SIF in $\boldsymbol{\rho}$, it is equivalent to prove that $\exists \mathbf{p} > \mathbf{0}$ such that $\mathbf{F}(\mathbf{1}, \mathbf{p}) \leq \mathbf{1}$ for the existence of a fixed point $\boldsymbol{\rho} = \mathbf{F}(\boldsymbol{\rho}, \mathbf{p}) \leq \mathbf{1}$. Multiplying both sides with positive vector $\mathbf{p}$, it suffices and necessitates to prove that $\exists \mathbf{p} > \mathbf{0}$ satisfying $\mathbf{I}(\mathbf{p}) = \mathbf{p} \circ \mathbf{F}(\mathbf{1}, \mathbf{p}) < \mathbf{p}$. This is also equivalent to prove that $\exists \mathbf{p} > \mathbf{0}$ such that $\mathbf{I}(\mathbf{p}) = \mathbf{p}$. Therefore, applying Lemma 6.4, the proof is complete. $\square$

Proposition 6.5 (Optimality)**.** The optimal solution for Problem 6.7 yields when the load vector satisfies $\boldsymbol{\rho} = \mathbf{1}$.

Proof. Since $\mathbf{I}(\mathbf{p})$ is an SIF in $\mathbf{p}$, it following from [81, Lemma 1] that for any feasible power vector $\mathbf{p}$, $\mathbf{p}* = \mathbf{I}(\mathbf{p}^*) \leq \mathbf{I}(\mathbf{p})$. Therefore, at the optimum, $\mathbf{p}* = \mathbf{I}(\mathbf{p}^*) = \mathbf{p}^* \circ \boldsymbol{\rho}$ and hence $\boldsymbol{\rho} = \mathbf{1}$. $\square$

Based on the feasibility and optimality, we define in the following the δ-Feasibility and δ-Optimality.

Definition 6.1. A system is called δ-Feasibility, if $\exists \mathbf{p} \geq 0$ such that $\boldsymbol{\rho} \leq \frac{1}{\delta}$ and $\delta \geq 1$.

Definition 6.2. A power vector is called δ-Optimality, if $\exists \mathbf{p} \geq 0$ such that $\boldsymbol{\rho} = \frac{1}{\delta}$ and $\delta \geq 1$.

6.3. Distributed Power Control Algorithm

6.3.1. Active Cell Protection

With the load interference function in hand, we proceed along similar lines as [78, 80] and formulate the distributed power control algorithm with active cell protection as follows:

$$p_i(\mathrm{t}+1) = \begin{cases} \delta I_i(\mathbf{p}(\mathrm{t})) & \text{for} \quad i \in \mathcal{A}_t, \\ \delta p_i(\mathrm{t}) = \delta^{(\mathrm{t}+1)} p_i(0) & \text{for} \quad i \in \mathcal{D}_t, \end{cases} \qquad (6.16)$$

where we denote the active ($\rho_i \leq 1$) and inactive ($\rho_i > 1$) sets of cells at time instance t as $\mathcal{A}_\mathrm{t}$ and $\mathcal{D}_\mathrm{t}$, respectively. If $\mathbf{p}_i > \mathbf{0}$, this can be written in a more compact way as

$$p_i(\mathrm{t} + 1) = \delta p_i(\mathrm{t})\bar{\rho}_i(\mathbf{p}(\mathrm{t})), \tag{6.17}$$

where $\bar{\rho}_i$ is the real load as defined in (2.27). The algorithm can be understood as first scaling transmission power by the real load and then multiplying with a power incremental δ. Now, we explain and prove the concept of ACP using Proposition 6.6 and Proposition 6.7. Note that the proof for the case of ALP is given in [78, 80].

Proposition 6.6. If $i \in \mathcal{A}_t$ and $\delta > 1$, $i \in \mathcal{A}_{t+1}$.

Proof. First, it can be easily concluded from (6.17) that $p_i(\mathrm{t} + 1) \leq \delta p_i(\mathrm{t})$, since $\bar{\rho}_i(\mathbf{p}(\mathrm{t})) \leq 1$. Then, due to the monotonicity and scalability of $I_i(\mathbf{p})$, we have

$$I_i(\mathbf{p}(\mathrm{t} + 1)) \leq I_i(\delta \mathbf{p}(\mathrm{t})) < \delta I_i(\mathbf{p}(\mathrm{t})) \tag{6.18}$$

Therefore,

$$\rho_i(\mathrm{t} + 1) = \frac{I_i(\mathrm{t} + 1)}{p_i(\mathrm{t} + 1)} = \frac{I_i(\mathbf{p}(\mathrm{t} + 1))}{\delta I_i(\mathbf{p}(\mathrm{t}))} \leq \frac{I_i(\delta \mathbf{p}(\mathrm{t}))}{\delta I_i(\mathbf{p}(\mathrm{t}))} < \frac{\delta I_i(\mathbf{p}(\mathrm{t}))}{\delta I_i(\mathbf{p}(\mathrm{t}))} = 1.$$

$\square$

Proposition 6.7. If $i \in \mathcal{D}_t$ and $\delta > 1$, $\rho_i(\mathrm{t} + 1) < \rho_i(\mathrm{t})$.

Proof. Using the two facts for proving the Proposition 6.6, we have

$$\rho_i(\mathrm{t} + 1) = \frac{I_i(\mathrm{t} + 1)}{p_i(\mathrm{t} + 1)} = \frac{I_i(\mathbf{p}(\mathrm{t} + 1))}{\delta p_i(\mathrm{t})} \leq \frac{I_i(\delta \mathbf{p}(\mathrm{t}))}{\delta p_i(\mathrm{t})} < \frac{\delta I_i(\mathbf{p}(\mathrm{t}))}{\delta p_i(\mathrm{t})} = \rho_i(\mathrm{t}).$$

$\square$

Proposition 6.8. The ALP/ACP is also valid for $\delta = 1$, however, indicating a power reduction control scheme.

Proof. For $\delta = 1$, (6.18) holds with weak inequality, i.e., $I_i(\mathbf{p}(\mathrm{t} + 1)) \leq I_i(\delta \mathbf{p}(\mathrm{t})) \leq \delta I_i(\mathbf{p}(\mathrm{t}))$. This indicates both Proposition 6.6 and Proposition 6.7 can be further formulated for $\delta = 1$ as: If $i \in \mathcal{A}_t$, $\rho_i(\mathrm{t} + 1) \leq 1$, while if $i \in \mathcal{D}_t$, $\rho_i(\mathrm{t} + 1) \leq \rho_i(\mathrm{t})$. $\square$

Remark 6.1. The proposed algorithm becomes the ALP algorithm in [80], if only one UE is associated with each BS. In this case, cell i is active if: $\frac{r_j}{b_i \log(1+\tau_{i,j})} \leq 1$, which is equivalent to the SINR threshold: $\tau_{i,j} \geq e^{r_j/b_i} - 1$.

6.3.2. Admissibility and Convergence

Based on the definitions in Section 6.2.2, we distinguish similarly to [80] three different levels for admissibility. Let $\mathcal{P}$ denote the feasible power region of the system and first consider the unlimited case, i.e., $\mathcal{P} = \mathbb{R}_+^{M+K}$.

(C.1) Fully admissible: if the δ-Feasibility condition is fulfilled, i.e., there exists $\mathbf{p} \in \mathcal{P}$ such that $\mathbf{0} \leq \boldsymbol{\rho}(\mathbf{p}) \leq \mathbf{1}/\delta$;

(C.2) δ-incompatible: (C.1) is not feasible but the feasibility condition is fulfilled, i.e., there exists $\mathbf{p} \in \mathcal{P}$ such that $\mathbf{0} \leq \boldsymbol{\rho}(\mathbf{p}) \leq \mathbf{1}$;

(C.3) Not fully admissible: (C.1) and (C.2) are not feasible, i.e., there exists no positive power vector $\mathbf{p} > 0$ satisfying $\boldsymbol{\rho}(\mathbf{p}) \leq \mathbf{1}$.

Note that $\boldsymbol{\rho}(\mathbf{p})$ is the load vector induced by the power vector $\mathbf{p}$ for two different interference models. For the worst-case interference model $\boldsymbol{\rho}(\mathbf{p}) = \mathbf{F}(\mathbf{1}, \mathbf{p})$, while $\boldsymbol{\rho}(\mathbf{p}) = \mathbf{H}(\mathbf{p})$ is used for the dynamic interference system.

Proposition 6.9. In case of (C.1), for every cell i, $\rho_i^* = \lim_{t \to \infty} \rho_i = 1/\delta$ and $p_i^* = \lim_{t \to \infty} p_i < \infty$, where the vector $\mathbf{p}^*$ is the optimal power vector that minimizes the energy consumption under the constraints $\boldsymbol{\rho} \leq \mathbf{1}/\delta$.

Proof. It has been shown in [80] that all cells become active in this case while the algorithm converges to $p_i^* = \delta I_i(\mathbf{p}^*) = \delta p_i^* \rho_i^*$. Therefore, we can conclude that $\rho_i^* = 1/\delta$. Furthermore, $\mathbf{p}^*$ minimizes the energy consumption, since the maximum load vector $\mathbf{1}/\delta$ is reached by $\mathbf{p}^*$. $\qquad\square$

Proposition 6.10. If (C.1) holds and $\delta > 1$, then the worst-case interference model requires a higher power at the convergence than the dynamic interference model.

Proof. Let $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$ denote the power convergence in case of (C.1) for the worst-case and the dynamic interference model, respectively. Due to the monotonicity of $\mathbf{F}$ in $\boldsymbol{\rho}$, we have $\mathbf{p}^{(1)} = \delta\mathbf{F}(\mathbf{1}, \mathbf{p}^{(1)}) \circ \mathbf{p}^{(1)} \geq \delta\mathbf{F}(\mathbf{1}/\delta, \mathbf{p}^{(1)}) \circ \mathbf{p}^{(1)}$. Furthermore, $\mathbf{p}^{(2)} = \delta\mathbf{F}(\mathbf{1}/\delta, \mathbf{p}^{(2)}) \circ \mathbf{p}^{(2)}$ is the fixed point of $\mathbf{p} = \delta\mathbf{F}(\mathbf{1}/\delta, \mathbf{p}) \circ \mathbf{p}$. In a static interference model with interference $\boldsymbol{\rho}'$, $\mathbf{I}(\mathbf{p}, \boldsymbol{\rho}')$ is a vector of SIFs, and therefore $\mathbf{I}'(\mathbf{p}) = \delta\mathbf{I}(\mathbf{p}, \boldsymbol{\rho}')$ is also a vector of SIFs. Hence, the fixed-point of $\mathbf{I}'(\mathbf{p}, \mathbf{1}/\delta)$, which equals to $\mathbf{p}^{(2)}$ is smaller than $\mathbf{p}^{(1)}$ element-wise. $\qquad\square$
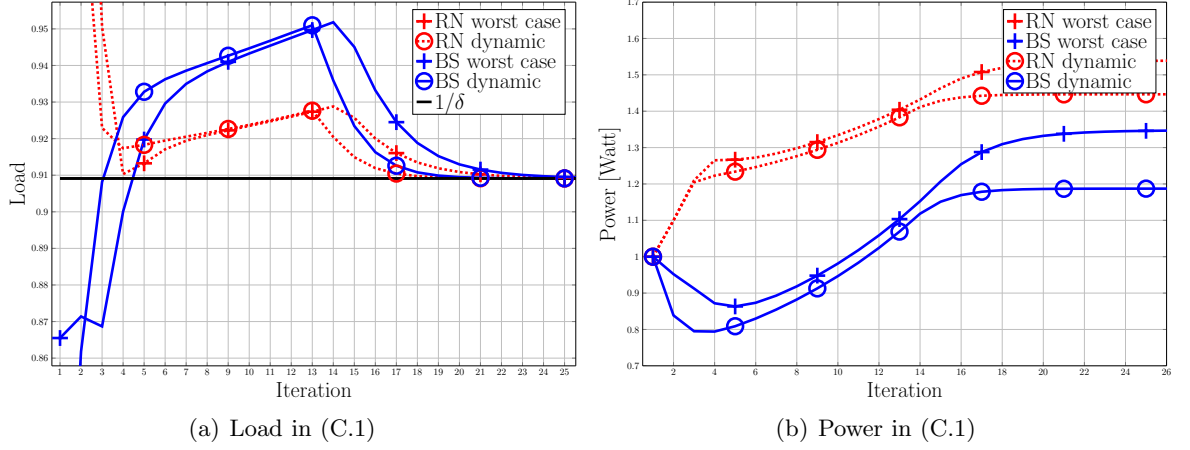
(a) Load in (C.1)

(b) Power in (C.1)

Figure 6.1.: Load and power performance in (C.1).

In Fig. 6.1, the behavior of load and power of a BS and a RN in case of (C.1) is depicted along the iterations. We simulate an exemplary network with 7 BSs, 50 RNs and 50 UEs. The other system parameters and details are listed in Table 6.1 in Section 6.5.

It can be clearly seen from Fig. 6.1 that the load of both the BS and the RN iteratively converges to $1/\delta$. In accordance with Proposition 6.9, the same convergence value is achieved in both the static and the dynamic interference model without violating the load constraints. Furthermore, Proposition 6.10 can be also justified since more power is required for the worst-case interference model.

Assume $\{\boldsymbol{\rho}'(t)\}_t$ ($\{\mathbf{p}'(t)\}_t$) and $\{\boldsymbol{\rho}''(t)\}_t$ ($\{\mathbf{p}''(t)\}_t$) to be the vector of load (power) sequences generated by the algorithm in (6.16) under the worst-case and the dynamic interference model, respectively. The convergence in case of (C.2) is summarized in Proposition 6.11-6.12.

Proposition 6.11. In (C.2) holds and $\delta > 1$, for every cell i, $1/\delta < \lim_{t\to\infty} \rho_i < 1$. Furthermore, the load under the dynamic interference model converges to a lower value compared with the worst-case interference model, i.e., as $t \to \infty$,

$$\mathbf{1} > \boldsymbol{\rho}'(t) > \boldsymbol{\rho}''(t) > \mathbf{1}/\delta. \tag{6.19}$$

Proof. Following [80, Prop.6], we know that $p_i \to \infty$ and all cells become active as $t \to \infty$, i.e. for every cell i, $\lim_{t\to\infty} \rho_i < 1$. This means, $\mathbf{p}'(t+1)/\mathbf{p}'(t) = \delta\boldsymbol{\rho}'(t) > 1$ and $\mathbf{p}''(t+1)/\mathbf{p}''(t) = \delta\boldsymbol{\rho}''(t) > 1$, indicating that $\boldsymbol{\rho}'(t) > 1/\delta$ and $\boldsymbol{\rho}''(t) > 1/\delta$, as $t \to \infty$. Furthermore, we have $\boldsymbol{\rho}'(t) = \mathbf{F}(\mathbf{1}, \mathbf{p}'(t)) = \mathbf{F}(\mathbf{1}, \mathbf{p}''(t)) > \mathbf{F}(\boldsymbol{\rho}''(t), \mathbf{p}''(t)) = \boldsymbol{\rho}''(t)$. □

Proposition 6.12. If (C.2) holds and $\delta > 1$, then $p_i \to \infty$ and the worst-case interference

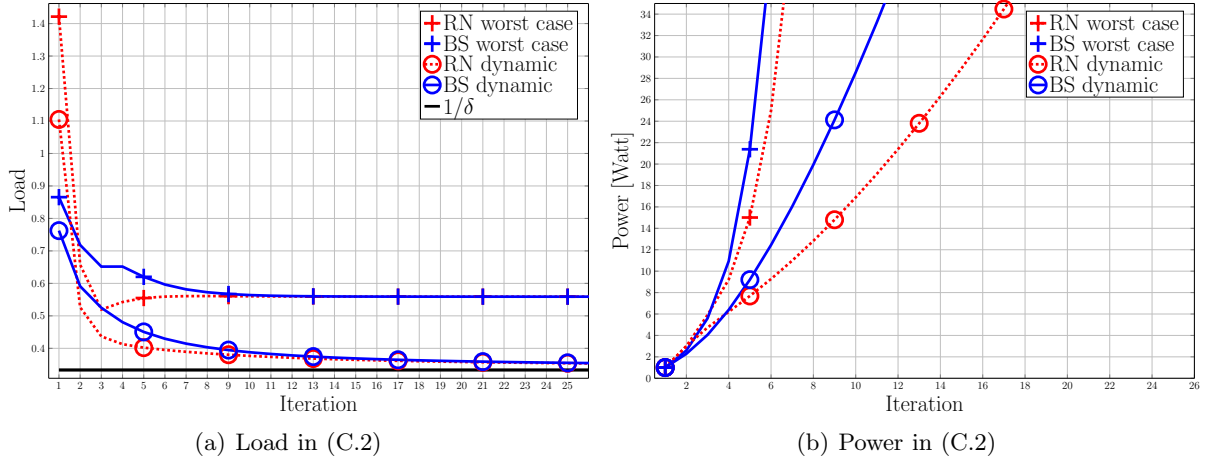(a) Load in (C.2)   (b) Power in (C.2)

Figure 6.2.: Load and power performance in (C.2).

model requires a higher power increase rate at the convergence, i.e., as $t \to \infty$,

$$\mathbf{p}'(t+1)/\mathbf{p}'(t) > \mathbf{p}''(t+1)/\mathbf{p}''(t). \tag{6.20}$$

Proof. The proof can be easily done by using the results of Proposition 6.11:

$$\mathbf{p}'(t+1)/\mathbf{p}'(t) = \delta\boldsymbol{\rho}'(t) > \delta\boldsymbol{\rho}''(t) = \mathbf{p}''(t+1)/\mathbf{p}''(t) > 1. \tag{6.21}$$

$\square$

In Fig. 6.2, the same scenario is simulated as for Fig. 6.1. It can be concluded that both power and load behave under (C.2) as Proposition 6.11-6.12 state. Diverging transmission powers are observed in both the worst-cases interference and the static interference model. Note that in practice, the growth of power to infinity is not possible, and therefore the algorithm cannot guarantee ACP in (C.2) with power constraints. We will discuss the case in detail in Section 6.4.

Proposition 6.13. In (C.3) holds, $\lim_{t \to \infty} \rho_i = 1$ for $i \in \mathcal{A}_t$, whereas $\lim_{t \to \infty} \rho_i > 1$ for $i \in \mathcal{D}_t$. Further, for all i, $p_i \to \infty$ and both models have the same behavior in load and power.

Proof. The conclusion follows directly [80, Prop.4]. Since the load converges to a value that is larger than or equal to one, the dynamic model becomes at the convergence indeed the worst-case model. Hence, the same power and load convergence will be observed. $\square$

The simulation results on the behavior of load and power in (C.3) are shown in Fig 6.3 for the same power unlimited scenario as for (C.1) and (C.2). It is clearly that in both cases the load and
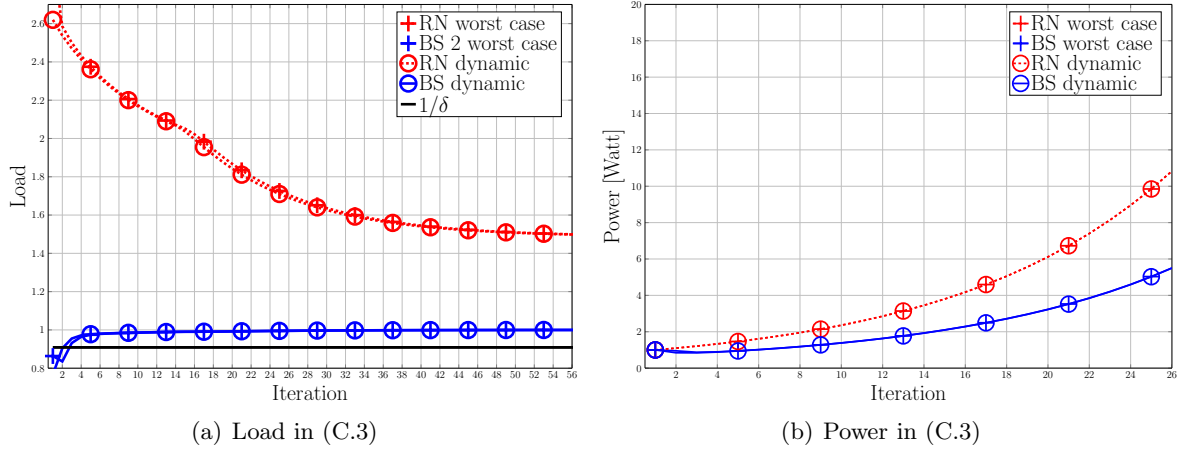
(a) Load in (C.3)

(b) Power in (C.3)

Figure 6.3.: Load and power performance in (C.3).

power curves differ only slightly in the first iteration. After, the load of the dynamic interference model reaches $\mathbf{1}$, exactly the same performance can be expected then for both models.

Now that there are two different interference models, we point out the performance differences of the two models with respect to admissibility and convergence.

Proposition 6.14. If (C.1) holds for worst-case interference, it also holds for dynamic interference assumption. If (C.1) or (C.2) holds for worst-case interference, (C.1) or (C.2) also holds for dynamic interference assumption.

Proof. If (C.1) holds for the worst-case model, then there exists $\mathbf{p}^{(1)}$ such that $\mathbf{H}(\mathbf{p}^{(1)}) = \mathbf{F}(\mathbf{H}(\mathbf{p}^{(1)}), \mathbf{p}^{(1)}) \le \mathbf{F}(\mathbf{1}, \mathbf{p}^{(1)}) \le 1/\delta$; if (C.1) or (C.2) holds for the worst-case model, then there exists $\mathbf{p}^{(2)}$ satisfying $\mathbf{H}(\mathbf{p}^{(2)}) = \mathbf{F}(\mathbf{H}(\mathbf{p}^{(2)}), \mathbf{p}^{(2)}) \le \mathbf{F}(\mathbf{1}, \mathbf{p}^{(2)}) \le \mathbf{1}$. $\qquad\square$

Proposition 6.15. If the system is not fully admissible for dynamic interference (i.e., (C.3)), it is also not fully admissible for the worst-case interference model.

Proof. The proof is simple since Proposition 6.15 is contrapositive of Proposition 6.14. $\qquad\square$

Proposition 6.14 and Proposition 6.15 indicate that the two interference models differ only when it is (C.2) for the worst-case model and (C.1) for the dynamic interference model. This difference vanishes as the power incremental δ goes to 1, indicating the equivalence of (C.1) and (C.2).

Fig 6.4 illustrates the comparison of load behavior between the worst-case and dynamic load assumption when the average rate requirement is steadily increased, where the other simulation parameters are identical with the simulations for (C.1)-(C.3). The two horizontal lines separate the cases in which the system ends up with under a certain average rate requirement. For
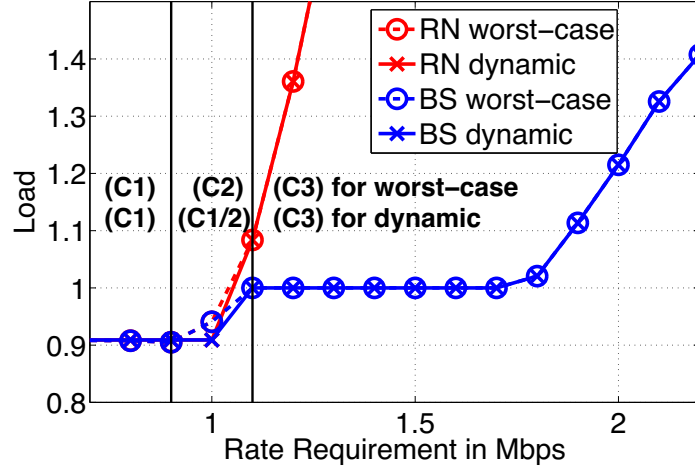
Figure 6.4.: Comparison of worst-case and dynamic interference model

instance, the system is fully admissible under both static and dynamic interference for rate requirements lower than 0.9 Mbps. By further increasing the rate requirement, the system under the worst-case interference model turns to be (C.2) whereas it is either (C.1) or (C.2) for the dynamic interference model. This can be observed by the different load convergence values for rate requirements between 0.9 Mbps and 1.1 Mbps, since according to Proposition 6.13, the two models have the same load convergence in (C.1), but different load convergences in (C.2). Furthermore, both scenarios are not fully admissible for rates equal or greater than 1.1 Mbps, since the loads converge to a value that is larger than 1.

6.4. Power Constraints and Implementation

In practical systems, the transmission powers are limited. To take the power limit into account for (C.1)-(C.3), we denote the power limit by $\hat{\mathbf{p}} \in \mathbb{R}_+^{M+K}$ and denote $\mathcal{P} = \{\mathbf{p} | \mathbf{0} \leq \mathbf{p} \leq \hat{\mathbf{p}}\}$. In this paragraph, we still refer to (C.1)-(C.3) as the three levels of admissibility but under power constraints.

It has been pointed out in [80] that full admission cannot be guaranteed for (C.2) in power constrained cases since it requires infinite power for convergence. This requires a *smaller* δ to avoid (C.2) in a limited power scenario. In particular, from Proposition 6.14 and Proposition 6.15, we know that (C.2) happens more frequently under the worst-case interference model. On the other hand, in [79], it is well understood that larger δ trades-off convergence speed with energy consumption. In this work, energy consumption is not the optimization objective. Hence, we
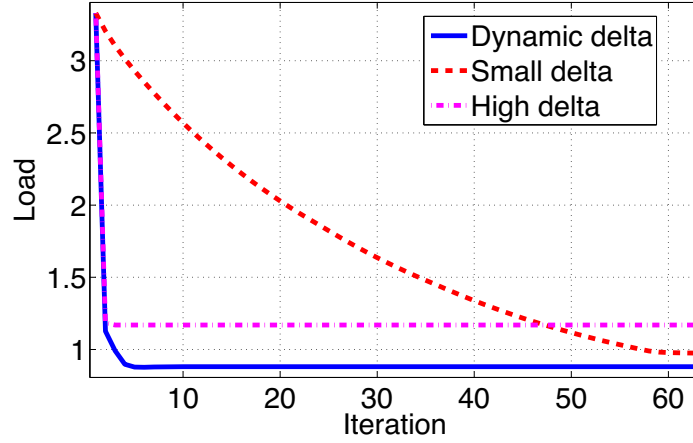
Figure 6.5.: Load behavior of various settings of δ

can optimize the convergence speed of the algorithm by using *large* values of δ as long as the power constraint is not violated. Based on all these observations, we propose an algorithm that dynamically controls δ as

$$p_i(\text{t}+1) = \delta(t)p_i(\text{t})\bar{\rho}_i(\mathbf{p}(\text{t})), \tag{6.22}$$

where positive initial powers, i.e. $\mathbf{0} < \mathbf{p}(0) \leq \hat{\mathbf{p}}$ and

$$\delta(\text{t}) = \min_i \frac{\hat{p}_i}{p_i(\text{t})\bar{\rho}_i(\text{t})}, \quad \text{for } i \in \mathcal{B}\bigcup\mathcal{R}. \tag{6.23}$$

Remark 6.2. Suppose the system power is bounded by $\mathbf{0} \leq \mathbf{p} \leq \hat{\mathbf{p}}$. First, $p_k(\text{t}+1) = \min_i \frac{\hat{p}_i}{p_i(\text{t})\bar{\rho}_i(\text{t})}p_k(\text{t})\bar{\rho}_k(\mathbf{p}(\text{t})) \leq \frac{\hat{p}_k p_k(\text{t})\bar{\rho}_k(\mathbf{p}(\text{t}))}{p_k(\text{t})\bar{\rho}_k(\text{t})} = \hat{p}_k$. Then, the algorithm in (6.23) chooses the largest $\delta(t)$ at time t. This can be proven by formulating the feasible set of $\delta(t)$ which is $\bigcap_{i\in\mathcal{B}\bigcup\mathcal{R}}\{\delta|1 \leq \delta \leq \frac{\hat{p}_i}{p_i(\text{t})\bar{\rho}_i(\text{t})}\}$. Choosing the largest $\delta(t)$ leads to best robustness and speed of convergence according to the trade-off study in [79] to. Moreover, $\delta(\text{t}) = \min_i \frac{\hat{p}_i}{p_i(\text{t})\bar{\rho}_i(\text{t})} \leq \frac{\hat{p}_k}{p_k(\text{t})\bar{\rho}_k(\text{t})} = \frac{\hat{p}_k}{p_k(\text{t})}$, for $k \in \mathcal{D}_t$. As long as $\delta(\text{t}) > 1$, $p_k(\text{t})$ will increase until reaching the power bound $\hat{p}_k$ and $\delta(\text{t})$ goes to 1. This means, as long as a cell stays inactive, $\delta(\text{t})$ converges to 1. Thus, if the system is not (C.3), the algorithm in (6.22) ensures full admission.

The algorithm (6.17) or (6.22) is easy to implement in real systems, since both current power $p_i(t)$ and current load $\rho_i(t)$ are known or can be easily estimated at the cell i. For instance, in LTE, the UEs are able to measure the RSRP, which is $p_i g_{i,j}$, and RSSI, which can be seen as $\sum_{d\in\mathcal{I}_i} p_d g_{d,j} + \sigma_j$. These can be fed back to the cells such that ρ_i can be computed and the power control algorithm can be carried out. One critical point for implementation is that the value of δ should be synchronized among all cells. Especially, dynamical updating of δ requires to exchange information between the BSs.
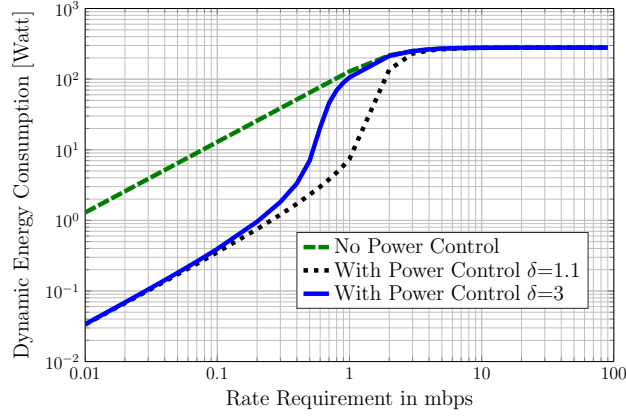
Figure 6.6.: Dynamic energy saving performance versus UE rate density.

Finally, we compare the ACP algorithm in (6.22)-(6.23) with a fixed delta algorithm in a power constrained scenario such that $p_i(t + 1) = \min\{p_i(t + 1), \hat{p}_i\}$. Fig 6.5 shows the importance of choosing a suitable δ: A high δ is preferable for fast convergence but can lead to (C.3). In contrast, a small δ implies a slow convergence speed. The proposed varying δ algorithm guarantees admission and achieves a high speed of convergence.

6.5. Energy Saving Performance Evaluation

The ACP properties have been intensively analyzed in the last sections. Now, we assume (C.1) holds and evaluate the benefit in terms of dynamic energy savings by performing the iterative power control scheme. A nomadic relay network with 7 hexagon layout BSs and 50 UEs is simulated. A simple cell selection scheme is used such that the UEs select an RN-BS or directly connect to a BS on the best end-to-end SINR basis. The other system parameters are listed in Table 6.1.

Table 6.1.: simulation configurations

Transmission Parameters	
initial transmission power	46 dBm for BS & 30 dBm for RN
available bandwidth	10 MHz for BS & 10 MHz for RN
antenna configuration	2 antennas for BSs, RNs and UEs
Channel and Noise Parameters in [dB]	
path loss model for all links	as in Table A.2.1.1.2-3 in [11]
noise figure	5 dB at UE & RN

Assuming 100 RNs in the network and a varying user data rate requirements of [0.01, 0.1, 1, 10,
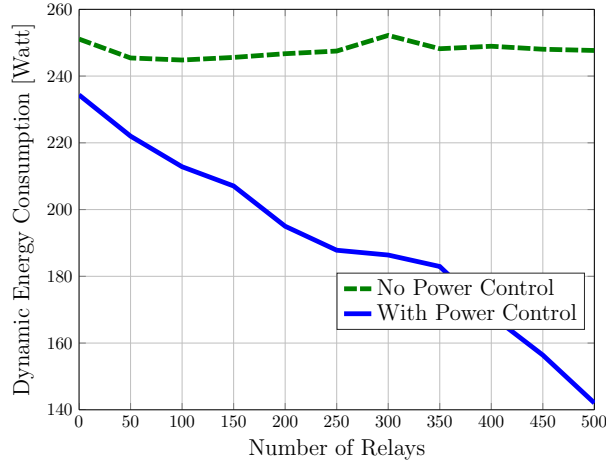
Figure 6.7.: Dynamic energy saving performance versus RN density.

100] mbps, the performance in terms of dynamic energy consumption is evaluated and displayed in Fig. 6.7. As comparison, we choose three algorithms: (i) No power control, i.e., all the BSs and RNs are transmitting at full power. (ii) Applying the power control algorithm with $\delta = 3$ and (iii) Applying the power control algorithm with $\delta = 1.1$.

If the average user rate requirement is lower than 1 Mbps, the proposed algorithm has the potential of reducing 90% of the total dynamic energy. Furthermore, it can be also concluded that a lower δ results in lower dynamic energy consumption. In the low rate region, however, the energy saving difference is not significant compared with a higher δ. As the average user rate increases, more benefit in terms of dynamic energy saving can be achieved by using smaller δ. The gain disappears and converges to the full power scheme if the total user rate is too high and the system is overloaded, since full load and full power is required for all algorithms.

In Fig. 6.7, we assume an average user data rate of 0.5 mbps and analyze the energy saving potentials with respect to varying amount of RNs. If no power control scheme is applied, the total dynamic energy consumption does not change too much when the number of RNs increases. In case of high number of RNs, more RNs are activated and a lower resource utilization is achieved if no power control scheme is applied. However,the RNs transmit at full power, and therefore the total dynamic energy consumption remains at the similar level. Significant reduction of dynamic energy consumption is achieved with more RNs by using the proposed power control scheme. This can be attributed to the fact that the overall spectral efficiency increases, which in turn requires lower total energy for delivering a certain amount of information.

Chapter 7.

Conclusion and Outlook

In this thesis, we have investigated the energy-saving potentials of nomadic relaying networks. Firstly, a mathematical model has been designed for the nomadic relaying network, taking into account the user requirement, rate assignment and the network load. Furthermore, we have formulated a generic interference coordination model and load interference coupling model for the nomadic relaying network. Based on these models, we have developed an optimization framework for the nomadic relaying network such that the user requirements are satisfied by the available resources at the cells. Subsequently, energy-saving optimizations have been carried out under this optimization framework, where the energy consumption is modeled as the sum of the l_0-norm and the l_1-norm of the product of load and the power.

We have divided the energy-saving problem into two sub-problems: a problem for assignment optimization and a problem for power control optimization. In the assignment-optimization problem, the properties of load-coupling function have been intensively discussed. For the static interference model, we have proven that the constraints are non-convex and monotonically increasing. For the dynamic interference model, we have proven the existence of an explicit load function that is monotonically increasing. Based on the properties, we have reformulated and relaxed the load function to allow for heuristically solving the problem. The l_0-norm objective is approximated as a strictly concave function that is then further iteratively optimized by LPs in the framework of the MM-algorithm. Extensive simulation results have been carried out to confirm the energy-saving potentials of the nomadic relaying network. In the low-traffic period, more than 50% of energy savings can be achieved by the proposed algorithms. The gain increases as the density and antenna capability of the nomadic relays can be further boosted. This means that an enhanced energy-saving performance can be expected if more vehicles are equipped with high-performance antennas.

In order to enable a more practical implementation, distributed algorithms have been proposed, whereby the available LTE measurements are used for the nodes to perform cell selection. Admission control and sleeping/awaking mechanisms have also been proposed in order to avoid overloading and to achieve energy savings. Simulation results have shown that the distributed algorithms achieve significant gains compared with conventional cell selection algorithms. Moreover, in most cases, a similar energy-saving performance is achieved as the centralized algorithm.

Finally, we have focused on the power control problem in order to further reduce dynamic energy consumption. We have derived the feasibility and optimality condition for power control algorithms under fixed assignments. Then, the power load coupling function has been investigated analogously as the (assignment) load coupling function. We have proven the existence of an SIF in both the static and the dynamic interference model. Based on this function, a distributed power control algorithm has been proposed such that the active cells are protected during the power ramping procedure. The power control algorithm is proven to converge to the optimal power vector that optimized the total dynamic energy under certain load constraints. While taking the power constraints into account, we have proposed an adaptive algorithm to achieve both optimal convergence speed and optimal system capacity.

Significant energy-saving potentials have been identified in this thesis. However, in order to fully exploit the benefit of the nomadic relaying network, both technical and business analysis need to be carried out. Firstly, by jointly optimizing assignments and power, further energy savings can be achieved, although this implies high complexity in the meantime. Then, the performance boost and modeling aspects due to application of multi-hop and advanced relaying techniques can be further investigated. Other objectives, such as load balancing or coverage enhancement, can also be studied in order to achieve gains in aspects other than energy savings. Furthermore, the dynamic network performance can be studied in order to jointly consider the traffic fluctuation in the optimization.

Apart from the aforementioned research directions, the signaling design for the management of the nomadic nodes is another key topic for practical implementations. Security concepts, energy supply management and a detailed architecture design are needed for the automotive manufacturers to integrate the nomadic relaying concept into vehicle architectures.

Appendix A.

Basic Matrix Operation Rules

We give in this section useful matrix equalities for the thesis. In Appendix A.1, we focus on the transformations involving Kronecker product and matrix vectorization, whereas matrix equalities including Hardamard product are discussed in Appendix A.2.

A.1. Kronecker and Vectorization

For any matrice $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{s \times t}$, and $\mathbf{X} \in \mathbb{R}^{n \times s}$ with suitable sizes for matrix operations, it holds according to [93, Th. 13.26] that:

$$(\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{A}\mathbf{X}\mathbf{B}). \tag{A.1}$$

Then, by assuming respectively $\mathbf{A} = \mathbf{I}$ and $\mathbf{B} = \mathbf{I}$, the following variants can be derived:

$$(\mathbf{B}^T \otimes \mathbf{I})\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{X}\mathbf{B}), \tag{A.2}$$

$$(\mathbf{I} \otimes \mathbf{A})\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{A}\mathbf{X}). \tag{A.3}$$

Furthermore, if $\mathbf{A}$ is a row vector, i.e., $m = 1$ and $\mathbf{A} = \mathbf{a}^T$, with suitable size for performing $\mathbf{A}\mathbf{X} = \mathbf{a}^T\mathbf{X}$, we have from (A.3):

$$(\mathbf{I} \otimes \mathbf{a}^T)\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{a}^T\mathbf{X}) = (\mathbf{a}^T\mathbf{X})^T = \mathbf{X}^T\mathbf{a}. \tag{A.4}$$

A.2. Kronecker, Hardmard and Vectorization

For two vectors, $\mathbf{x}$ and $\mathbf{y}$, of the same size, Hardmard production can be transformed into normal matrix product as

$$\mathbf{x} \circ \mathbf{y} = \operatorname{diag}(\mathbf{x}) \cdot \mathbf{y} = \operatorname{diag}(\mathbf{y}) \cdot \mathbf{x}. \tag{A.5}$$

Given any matrices $\mathbf{A}$, $\mathbf{B}$ of the same size and a vector $\mathbf{x}$ that has the suitable size for performing $\mathbf{Bx}$, we can apply (A.5) and (A.2) to $(\mathbf{A} \circ \mathbf{B})\mathbf{x}$ such that

$$(\mathbf{A} \circ \mathbf{B})\mathbf{x} = (\mathbf{x}^T \otimes \mathbf{I})\operatorname{vec}(\mathbf{A} \circ \mathbf{B}) = (\mathbf{x}^T \otimes \mathbf{I})\operatorname{diag}(\mathbf{A})\operatorname{vec}(\mathbf{B}). \tag{A.6}$$

It can be easily verified from the definition of the matrix product that

$$\mathbf{A}\mathbf{x} = (\mathbf{A} \circ (\mathbf{1} \otimes \mathbf{x}^T))\mathbf{1}. \tag{A.7}$$

Therefore $(\mathbf{A} \circ \mathbf{B})\mathbf{x}$ can be also written as

$$(\mathbf{A} \circ \mathbf{B})\mathbf{x} = (\mathbf{A} \circ \mathbf{B} \circ (\mathbf{1} \otimes \mathbf{x}^T))\mathbf{1}. \tag{A.8}$$

Furthermore, consider vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ with the same size as. Then, by (A.6), we have

$$
\begin{aligned}
\mathbf{z}^T \cdot (\mathbf{y} \circ \mathbf{x}) &= (\mathbf{y}^T \circ \mathbf{z}^T) \cdot \mathbf{x} \\
&= (\mathbf{x}^T \otimes \mathbf{I}) \cdot \operatorname{diag}(\mathbf{z}^T) \cdot \operatorname{vec}(\mathbf{y}^T) \\
&= \mathbf{x}^T \cdot \operatorname{diag}(\mathbf{z}) \cdot \mathbf{y},
\end{aligned} \tag{A.9}
$$

whereby the position of $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ can be arbitarily switched.

Without giving details of proof, we further state the following matrix equality:

$$([\mathbf{1}]^T \otimes \mathbf{a}) \circ \mathbf{B})\mathbf{r} = \mathbf{a} \circ (\mathbf{B} \cdot \mathbf{r}), \tag{A.10}$$

which can be generalized as

$$([\mathbf{1}]^T \otimes \mathbf{A}) \circ \mathbf{B})(\mathbf{r} \otimes \mathbf{I}) = \mathbf{A} \circ (\mathbf{B}(\mathbf{r} \otimes \mathbf{I})). \tag{A.11}$$

Appendix B.

Quadratic Forms

B.1. Quadratic and Bilinear Form

Following [84], we summarize the main properties of the quadratic form.

Definition B.1. A quadratic form of the variables $\mathbf{x} \in \mathbb{R}^n$ is a polynomial function $Q : \mathbb{R}^n \to \mathbb{R}$, where all terms in the functional expression $Q(\mathbf{x})$ have order two.

Lemma B.1. A function $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$ is a quadratic form if and only if it can be written as

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}, \tag{B.1}$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a symmetric matrix.

Definition B.2. Let $Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, $\mathbf{x} \in \mathbb{R}^n$ be a quadratic form, with associated symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. Then, $\mathbf{A}$ is

- positive (semi)-definite, if $Q(\mathbf{x}) > 0 (Q(\mathbf{x}) \geq 0)$ when $\mathbf{x} \neq 0$;

- negative (semi)-definite, if $Q(\mathbf{x}) < 0 (Q(\mathbf{x}) \leq 0)$ when $\mathbf{x} \neq 0$;

- indefinite, if $Q(\mathbf{x})$ takes both positive and negative values.

The convexity of the quadratic form can be decided by the Proposition B.1.

Proposition B.1. Let $Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, $\mathbf{x} \in \mathbb{R}^n$ be a quadratic form, with associated symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. Then we have:

- Q is strictly convex (or convex), if and only if $\mathbf{A}$ is positive (semi)-definite;

- Q is strictly concave (or concave) if and only if $\mathbf{A}$ is negative (semi)-definite.

B.2. Semidefinite Programming Relaxation

We explain in this section according to [94] the SDP relaxation of Quadratically Constrained Quadratic Programming (QCQP). Consider the following QCQP

$$
\begin{aligned}
\text{QCQP:} \qquad \min \qquad & \frac{1}{2}\mathbf{x}^T\mathbf{Q}_0\mathbf{x} + \mathbf{a}_0^T\mathbf{x} \\
\text{subject to} \qquad & \frac{1}{2}\mathbf{x}^T\mathbf{Q}_i\mathbf{x} + \mathbf{a}_i^T\mathbf{x} \leq b_i \quad i \in \mathcal{I} \\
& \frac{1}{2}\mathbf{x}^T\mathbf{Q}_i\mathbf{x} + \mathbf{a}_i^T\mathbf{x} = b_i \quad i \in \mathcal{E} \\
& \mathbf{l} \leq \mathbf{x} \leq \mathbf{u},
\end{aligned}
$$

where $\mathbf{x}, \mathbf{a}_i \in \mathbb{R}^n$, $\mathbf{Q}_i \in \mathbb{R}^{n \times n}$, while $\mathcal{I}$ and $\mathcal{E}$ are sets of inequalities and equalities, respectively. Furthermore, we assume that the bounds are limited, i.e., $-\infty < \mathbf{l} < \mathbf{u} < \infty$, and the matrices $\mathbf{Q}_i$ are all symmetric.

The problem is convex if and only if all matrices are positive-semidefinite, i.e., $\mathbf{Q}_i \succeq 0$. For non-covex problem, SDP relaxation can be applied, where the idea is to impose a convex constration $\mathbf{X} \succeq \mathbf{x}\mathbf{x}^T$ instead the non-convex constraints $\mathbf{X} = \mathbf{x}\mathbf{x}^T$

$$
\begin{aligned}
\text{SDP:} \qquad \min \qquad & \frac{1}{2}\text{tr}\left(\mathbf{X}\mathbf{Q}_0\right) + \mathbf{a}_0^T\mathbf{x} \\
\text{subject to} \qquad & \frac{1}{2}\text{tr}\left(\mathbf{X}\mathbf{Q}_i\right) + \mathbf{a}_i^T\mathbf{x} \leq b_i \quad i \in \mathcal{I} \\
& \frac{1}{2}\text{tr}\left(\mathbf{X}\mathbf{Q}_i\right) + \mathbf{a}_i^T\mathbf{x} = b_i \quad i \in \mathcal{E} \\
& \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \\
& \mathbf{X} - \mathbf{x}\mathbf{x}^T \succeq 0.
\end{aligned}
$$

This can be more compactly written as:

$$
\begin{aligned}
\text{SDP:} \qquad \min \qquad & \text{tr}\left(\tilde{\mathbf{X}}\tilde{\mathbf{Q}}_0\right) \\
\text{subject to} \qquad & \text{tr}\left(\tilde{\mathbf{X}}\tilde{\mathbf{Q}}_i\right) \leq 0 \quad i \in \mathcal{I} \\
& \text{tr}\left(\tilde{\mathbf{X}}\tilde{\mathbf{Q}}_i\right) = 0 \quad i \in \mathcal{E} \qquad\qquad (\text{B.2}) \\
& \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \\
& \tilde{\mathbf{X}} \succeq 0.
\end{aligned}
$$

where

$$\tilde{\mathbf{X}} = \begin{pmatrix} 1 & \mathbf{x}^T \\ \mathbf{x} & \mathbf{X,} \end{pmatrix}$$

and

$$\tilde{\mathbf{Q}} = \begin{pmatrix} 2b_i & \mathbf{a}_i^T \\ \mathbf{a}_i & \mathbf{Q.} \end{pmatrix}$$

If the orginal problem is convex, the Problem B.2 is equivalent to the origninal problem. If the orginal problem is not convex, the Problem B.2 can be unbounded, but bounds can be easily formulated, e.g., by multiplying the bounds of $\mathbf{x}$, we have $\mathbf{l} \cdot \mathbf{l}^T \leq \mathbf{X} \leq \mathbf{u} \cdot \mathbf{u}^T$

B.3. Reformulation Linearization Techniques

Introduced in [95], RLT, as called McCormick convex relaxation, is another technique for solving non-convex QCQP. The basic concept of RLT is to multiply the inequality constraints of the orignal problem. For instance, by multiplying the two upper bounds we have

$$(\mathbf{x} - \mathbf{u})(\mathbf{x} - \mathbf{u})^T = \mathbf{X} - \mathbf{u}\mathbf{x}^T - \mathbf{x}\mathbf{u}^T + \mathbf{u}\mathbf{u}^T \geq 0.$$

Therefore the RLT reformulation of the problem can be written as, where :

$$
\begin{aligned}
\text{RLT:} \quad &\min & &\frac{1}{2}\text{tr}\,(\mathbf{X}\mathbf{Q}_0) + \mathbf{a}_0^T\mathbf{x} \\
&\text{subject to} & &\frac{1}{2}\text{tr}\,(\mathbf{X}\mathbf{Q}_i) + \mathbf{a}_i^T\mathbf{x} \leq b_i \quad i \in \mathcal{I} \\
& & &\frac{1}{2}\text{tr}\,(\mathbf{X}\mathbf{Q}_i) + \mathbf{a}_i^T\mathbf{x} = b_i \quad i \in \mathcal{E} \\
& & &\mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \\
& & &\mathbf{X} = \mathbf{X}^T \\
& & &\mathbf{X} - \mathbf{u}\mathbf{x}^T - \mathbf{x}\mathbf{u}^T + \mathbf{u}\mathbf{u}^T \geq 0 \\
& & &\mathbf{X} - \mathbf{l}\mathbf{x}^T - \mathbf{x}\mathbf{l}^T + \mathbf{l}\mathbf{l}^T \geq 0 \\
& & &\mathbf{X} - \mathbf{u}\mathbf{x}^T - \mathbf{x}\mathbf{l}^T + \mathbf{u}\mathbf{l}^T \leq 0 \\
& & &\mathbf{X} - \mathbf{l}\mathbf{x}^T - \mathbf{x}\mathbf{u}^T + \mathbf{l}\mathbf{u}^T \leq 0
\end{aligned}
$$

Note that the last two constraints are in fact identical. Furthermore, the condition that $\mathbf{X}$ is symetric is also included in the reformulation, resulting in an ordinary LP with $n(n+3)/2$ variables and a total of $m + n(2n+3)$ constraints.

B.4. Reduced Reformulation Linearization Techniques

RRLT can be seen as the application of RLT to the equality constraints, resulting in the cancelation of the higher order constraints. The Concept of RRLT is explained in [87] and we explain here the basic principles. Consider a QCQP with linear equality constraints:

$$
\begin{aligned}
\text{QCQP2:} \qquad \min \quad & \frac{1}{2}\mathbf{x}^T\mathbf{Q}_0\mathbf{x} + \mathbf{a}_0^T\mathbf{x} \\
\text{subject to} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{Q}_i\mathbf{x} + \mathbf{a}_i^T\mathbf{x} \le b_i \quad i \in \mathcal{I} \\
& \mathbf{a}_j^T\mathbf{x} = b_j \quad j \in \mathcal{E} \\
& \mathbf{l} \le \mathbf{x} \le \mathbf{u},
\end{aligned}
$$

where $\mathbf{x}, \mathbf{a}_i \in \mathbb{R}^n$, $\mathbf{Q}_i \in \mathbb{R}^{n \times n}$, while $\mathcal{I}$ and $\mathcal{E}$ are sets of inequalities and equalities, respectively. We can make use of the equalities by multipling the inequalities with the coefficient in the linear equalities, yielding:

$$
\mathbf{a}_j(\frac{1}{2}\mathbf{x}^T\mathbf{Q}_i\mathbf{x} + \mathbf{a}_i^T\mathbf{x}) \le \mathbf{a}_j b_i, \quad i \in \mathcal{I}, \quad j \in \mathcal{E}.
$$

By noting the fact that $\mathbf{a}_j\mathbf{x}^T = \mathbf{b}_j^T$, we can simplify it as

$$
\frac{1}{2}\mathbf{b_j}^T\mathbf{Q}_i\mathbf{x} + \mathbf{a}_j\mathbf{a}_i^T\mathbf{x} \le \mathbf{a}_j b_i \quad i \in \mathcal{I}, \quad j \in \mathcal{E},
$$

or more campactly as:

$$
\frac{1}{2}\mathbf{B}^T\mathbf{Q}_i\mathbf{x} + \mathbf{A}\mathbf{a}_i^T\mathbf{x} \le \mathbf{A}b_i \quad i \in \mathcal{I},
$$

where $\mathbf{A}$ and $\mathbf{B}$ are matrices containing $\mathbf{a}_j$ and $\mathbf{b}_j$ as column vectors. It has been shown in [87] that this linear constraint is redundant with respect to the original constraints.

Especially, we develop a special case of Bilinear Programm (BLP):

$$
\begin{aligned}
\text{BLP:} \qquad \min \quad & \frac{1}{2}\mathbf{x}^T\mathbf{Q}_0\mathbf{y} + \mathbf{a}_0^T\mathbf{x} \\
\text{subject to} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{Q}_i\mathbf{y} + \mathbf{a}_i^T\mathbf{x} \le b_i \quad i \in \mathcal{I} \\
& \mathbf{a}_j^T\mathbf{y} = b_j \quad j \in \mathcal{E} \\
& \mathbf{l_x} \le \mathbf{x} \le \mathbf{u_x}, \quad , \mathbf{l_y} \le \mathbf{y} \le \mathbf{u_y}
\end{aligned}
$$

Assume $\mathbf{Q}_i$ is separatable, i.e., $\mathbf{Q}_i = \mathbf{h}_i\mathbf{g}_i^T$. Therefore, by assuming $\mathbf{Z} = \mathbf{x}\mathbf{y}^T$, we have an extra

reformulation of the problem as

$$
\begin{aligned}
\text{BLP-R:} \qquad \min \qquad & \frac{1}{2}\mathbf{h}_i^T \mathbf{Z}_0 \mathbf{g} + \mathbf{a}_0^T \mathbf{x} \\
\text{subject to} \qquad & \frac{1}{2}\mathbf{h}_i^T \mathbf{Z}_i \mathbf{g} + \mathbf{a}_i^T \mathbf{x} \le b_i \quad i \in \mathcal{I} \\
& \mathbf{a}_j^T \mathbf{y} = b_j \quad j \in \mathcal{E} \\
& \mathbf{l_x} \le \mathbf{x} \le \mathbf{u_x}, \quad , \mathbf{l_y} \le \mathbf{y} \le \mathbf{u_y}.
\end{aligned}
$$

Therefore, we end up with the equalites constraint $\mathbf{Z}\mathbf{a}_j = \mathbf{x}b_j^T$.

Appendix C.

Load Power Coupling Function

C.1. Implicit Function Theorem

Though there are many variations, we formulate for the thesis the *differentiable implicit function theroem* the based on [96, Theorem 1].

Theorem C.1 (Differentiable Implicit Function Theorem)**.** Suppose a function $\mathbf{F} : \mathcal{X} \times \mathcal{Y} \to \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are open sets on $\mathbb{R}^x$ and $\mathbb{R}^y$, respectively.

If $\mathbf{F}(\mathbf{x}^*, \mathbf{y}^*) = \mathbf{0}$, $\mathbf{F}$ is differentiable on $\mathcal{X} \times \mathcal{Y}$ and $\mathbf{F}$ is surjective with respect to $\mathbf{y}$ and , i.e., $\mathbf{J}_{\mathbf{y}}^{\mathbf{F}}(\mathbf{x}, \mathbf{y})$ is invertable on $\mathcal{X} \times \mathcal{Y}$,

Then, there exists a unique differentiable function $\mathbf{G} : \mathcal{X} \to \mathcal{Y}$ in the neighborhood of $(\mathbf{x}^*, \mathbf{y}^*)$ such that $\mathbf{y} = \mathbf{G}(\mathbf{x})$ with

$$\mathbf{J}_{\mathbf{x}}^{\mathbf{G}}(\mathbf{x}^*) = -(\mathbf{J}_{\mathbf{F}}^{\mathbf{y}}(\mathbf{x}^*, \mathbf{G}(\mathbf{x}^*)))^{-1} \mathbf{J}_{\mathbf{F}}^{\mathbf{x}^*}(\mathbf{x}, \mathbf{G}(\mathbf{x}^*)).$$

C.2. Generalized Diagonal Dominated Matrix

Definition C.1 (GDM [86])**.** A matrix $\mathbf{H}$ is a GDM if

- $\mathbf{H}$ has non-negative elements on the diagonal and non-positive elements elsewhere;

- there exists an all positive vector $\mathbf{s}$, such that $\mathbf{H} \cdot \mathbf{s}$ is an all positive vector, i.e., $\mathbf{H} \cdot \mathbf{s} > 0$.

Furthermore, $\mathbf{H}$ is *non-singular* and $\mathbf{H}^{-1}$ has only non-negative elements, i.e. $\mathbf{H}^{-1} \geq \mathbf{0}$.

Acronyms

ABS   Antilock Braking System

ACP   Active Cell Protection

ALP   Active Link Protection

AMC   Adaptive Modulation and Coding

BLP   Bilinear Programm

BS   Base Station

CAN   Control Area Network

CCS   Closest Cell Selection

CDMA   Code Division Multiple Access

CE   Consumer Electronics

CMU   Centralized Management Unit

CoMP   Coordinated Multipoint Transmission and Reception

CSI   Channel State Information

DCAG   Distributed Cell Selection and Admission Control for Generic Energy Saving

DCAS   Distributed Cell Selection and Admission Control for Static Energy Saving

DSRC   Dedicated Short-Range Communications

DSP   Digital Signal Processing

DPC   Distributed Power Control

DPM   Dynamic Power Management

DRRLT   Dynamic Reduced Reformulation-Linearization Technique

DTX   discontinuous transmission

D2D   Device to Device

ECR	Energy Consumption Ratio
ECU	Electronical Control Unit
EDGE	Enhanced Data Rate for GSM Evolution
eNodeB	evolved Node B
EPC	Evolved Packet Core
ESP	Electronic Stability Program
E-UTRA	Evolved-UMTS Terrestrial Radio Access
GDCA	Generic Distributed Cell Selection and Admission Control
GDM	Generalized Diagonally Dominant Matrix
GIF	General Interference Function
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HSPA	High Speed packet Access
IBU	Iterative Backhaul Updating
ICT	Information and Communication Technology
IoT	Internet of Things
IMT	International Mobile Telecommunications
ISD	Inter Site Distance
ITS	Intelligent Transportation System
LIN	Local Interconnect Network
LP	Linear Program
LTE	Long Term Evolution
MIMO	Multiple Input Multiple Output
MIB	Master Information Block
MM	Majorization Minimization
MME	Mobility Management Entity
MMS	Multi Media Messages
MNO	Mobile Network Operator
MOST	Media Oriented Systems Transport

M2M	Machine to Machine
NMT	Nordic Mobile Telephone
NP	Non-deterministic Polynomial-time
NTT	Nippon Telegraph and Telephone
OEM	Original Equipment Manufacturer
OFDM	Orthogonal Frequency Division Multiplexing
OPEX	Operational Expenditure
PA	Power Amplifier
PBCH	Physical Broadcast Channel
PCRF	Policy and Charging Rules Function
PCU	Policy Control Unit
PRB	Physical Resource Block
PSS	Primary Synchronization Signal
P-GW	Packet Gateway
QCQP	Quadratically Constrained Quadratic Programming
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RDPC	Robust Distributed Power Control
RF	Radio Frequency
RLT	Reformulation-Linearization Technique
RN	Relay Node
RRLT	Reduced Reformulation-Linearization Technique
RRM	Radio Resource Management
RS	Relay Selection
RSRP	Reference Signal Received Power
RSSI	Received Signal Strength Indicator
RSRQ	Reference Signal Received Quality
SAE	System Architecture Evolution

SDP	Semi-Definite Programming
SE	Spectral Efficiency
SIB	System Information Block
SINR	Signal-to-Interference-plus-Noise-Ratio
SIF	Standard Interference Function
SIR	Signal to Interference Ratio
SLR	Sequential Linear Reformulation
SMS	Short Message Services
SON	Self-Organizing Network
SRR	SDP and RLT Relaxation
SSS	Secondary Synchronization Signal
S-GW	Service Gateway
TDMA	Time Division Multiple Access
TTT	Time to Trigger
UA	User Association
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
V2V	Vehicle-to-Vehicle
V2I	Vehicle-to-Infrastructure
VANET	Vehicular Ad Hoc Network
WAVE	Wireless Communication in Vehicular Environments
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless Local Area Network
1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Program
4G	Fourth Generation
5G	Fifth Generation

List of Figures

List of Tables

Bibliography

[1] N. Navet, Y. Song, F. Simonot-Lion, and C. Wilwert, "Trends in Automotive Communication Systems," *Proceedings of the IEEE*, vol. 93, no. 6, June 2005.

[2] *BLUETOOTH SPECIFICATION: Hands-Free Profile (HFP) 1.5*, Car Working Group, Bluetooth Special Interest Group (SIG), Std. V10r00, 2005.

[3] H. Hartenstein and K. P. Laberteaux, "A Tutorial Survey on Vehicular Ad Hoc Networks," *IEEE Communications Magazine*, vol. 46, no. 6, June 2008.

[4] *Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Definitions*, ETSI Std. TR 102.638, v1.1.1, 2009.

[5] L. Xu, W. He, and S. Li, "Internet of Things in Industries: A Survey," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, November 2014.

[6] G. Americas, "4G Mobile Broadband Evolution: 3GPP Release 10 and Beyond," 4G Americas, White Paper, Februray 2011.

[7] ——, "4G Mobile Broadband Evolution: 3GPP Release 11 & Release 12 and Beyond," 4G Americas, White Paper, Februray 2014.

[8] P. Popovski, V. Braun, H. P. Mayer, P. Fertl *et al.*, "ICT-317669-METIS/D1.1 Scenarios, Requirements and KPIs for 5G Mobile and Wireless Systems," EU-Project METIS (ICT-317669), Deliverable, 2013.

[9] P. Popovski, V. Braun, G. Mange, P. Fertl *et al.*, "ICT-317669-METIS/D6.2 Initial Report on Horizontal Topics, First Results and 5G System Concept," EU-Project METIS (ICT-317669), Deliverable, 2014.

[10] E. Lang, S. Redana, and R. Bernhard, "Business Impact of Relay Deployment for Coverage Extension in 3GPP LTE-Advanced," in *Proceedings of IEEE International Conference on Communications (ICC) Workshops, Dresden, Germany*, June 2009, pp. 1–5.

[11] 3GPP, *Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects*, 3GPP Specification TR36.814 v9.0.0, 2010.

[12] ——, *Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-Configuring and Self-Optimizing Network (SON) Use Cases and Solutions*, 3GPP Specification TR36.902 v9.3.1, 2011.

[13] R. Combes, Z. Altman, and E. Altman, "Self-Organizing Relays: Dimensioning, Self-Optimization, and Learning," *IEEE Transactions on Network and Service Management*, vol. 9, no. 4, pp. 487–500, 2012.

[14] 3GPP, *Radio Planning Tool Access (RPTA) Integration Reference Point (IRP); Solution Set (SS) Definitions*, 3GPP Specification TS28.669, 2014.

[15] F. Haider, M. Dianati, and R. Tafazolli, "A Simulation Based Study of Mobile Femtocell Assisted LTE Networks ," in *Proceedings of the 7th International Wireless Communications and Mobile Computing Conference (IWCMC), Istanbul*, June 2011, pp. 2198–2203.

[16] H. Drost, G. Zimmermann *et al.*, "ICT-317669-METIS/D6.4 Final Report on 5G Architecture," EU-Project METIS (ICT-317669), Deliverable, 2015.

[17] Z. Ren, S. Stańczak, P. Fertl, and F. Penna, "Energy-Aware Activation of Nomadic Relays for Performance Enhancement in Cellular Networks," in *Proceedings of IEEE International Conference on Communications (ICC), Sydney, Australia*, June 2014, pp. 1–6.

[18] Z. Ren, S. Stańczak, and P. Fertl, "Activation of Nomadic Relays in Dynamic Interference Environment for Energy Savings," in *Proceedings of IEEE Global Conference on Communications (GLOBECOM), Austin, Texas*, December 2014, pp. 1–6.

[19] Z. Ren, S. Stańczak, M. Shabeeb, P. Fertl, and L. Thiele, "A Distributed Algorithm for Energy Saving in Nomadic Relaying Networks," in *Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA*, November 2014, pp. 1–5.

[20] Z. Ren, M. Jäger, S. Stańczak, and P. Fertl, "Distributed Power Control with Active Cell Protection in Future Cellular Systems," in *IEEE International Conference on Communications (ICC), London, UK*, June 2015, pp. 1–6.

[21] Z. Ren, S. Stańczak, and P. Fertl, "An Optimization Framework for Energy Saving in 5G Nomadic Relaying Networks," *Preprint*, 2015.

[22] Ömer Bulakci, Z. Ren, C. Zhou, J. Eichinger, P. Fertl, and S. Stańczak, "Dynamic Nomadic Node Selection for Performance Enhancement in Composite Fading/Shadowing Environments," in *Processing of IEEE Vehicular Technology Conference Spring (VTC-Spring), Soul, Korea*, May 2014, pp. 1–6.

[23] Ömer Bulakci, Z. Ren, C. Zhou, J. Eichinger, P. Fertl, D. Gozalvez-Serreno, and S. Stańczak, "Towards Flexible Network Deployment in 5G: Nomadic Node Enhancement to Heterogeneous Networks," in *Workshop of IEEE International Conference on Communications (ICC Workshop), London, UK*, June 2015, pp. 1–6.

[24] Z. Ren, P. Fertl, Q. Liao, F. Penna, and S. Stańczak, "Street Specific Handover Optimization in Future Cellular Networks," in *Processing of IEEE Vehicular Technology Conference Spring (VTC-Spring), Dresden, Germany*, June 2013, pp. 1–5.

[25] Q. Liao, F. Penna, S. Stańczak, Z. Ren, and P. Fertl, "Context-Aware Handover Optimization for Relay-Aided Vehicular Terminals," in *Proceedings of IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Darmstadt, Germany*, June 2013, pp. 1–5.

[26] F. Penna, S. Stańczak, Z. Ren, and P. Fertl, "MMSE Interference Identification in LTE Networks," in *Proceedings of IEEE International Conference on Communications (ICC), Sydney, Australia*, June 2014, pp. 1–6.

[27] A. Fehske, H. Klessig, J. Voigt, and G. Fettwei, "Concurrent Load-Aware Adjustment of User Association and Antenna Tilts in Self-Organizing Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 1974–1988, 2013.

[28] R. Combes, Z. Altman, and E. Altman, "Self-Organizing Relays: Queuing Analysis and Algorithms ," in *7th International Conference on Network and Service Management (CNSM), Paris, France*, October 2011, pp. 1–8.

[29] E. Pollakis, R. L. G. Cavalcante, and S. Stanczak, "Base Station Selection for Energy Efficient Network Operation with the Majorization-Minimization Algorithm," in *IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cesme, Turkey, June*, 2012, pp. 1–6.

[30] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, July, 623-656, October 1948.

[31] I. Siomina, A. Furuskar, and G. Fodor, "A Mathematical Framework for Statistical QoS and Capacity Studies in OFDM Networks," in *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, Tokyo, Japan, Sept.*, 2009, pp. 2772 – 2776.

[32] A. Fehske and G. Fettwei, "Aggregation of Variables in Load Models for Interference-Coupled Cellular Data Networks," in *IEEE International Conference on Communications (ICC), Ottawa, Canada, June*, 2012.

[33] K. Majewski and M. Koonert, "Conservative Cell Load Approximation for Radio Networks with Shannon Channels and its Application to LTE Network Planning," in *2010 Sixth Advanced International Conference on Telecommunications (AICT), May*, 2010, pp. 219–225.

[34] A. B. Saleh, O. Bulakci, S. Redana, B. Raaf, and J. Haemaelaeinen, "A Divide-and-Conquer Approach to Mitigate Relay-to-Relay Interference," in *IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Toronto, Canada*, September 2011, pp. 1–6.

[35] E. Pollakis, R. Cavalcante, and S. Stanczak, "Enhancing Energy Efficient Network Operation in Multi-RAT Cellular Environments through Sparse Optimization," in *The IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Darmstadt, Germany, June 2013.

[36] K. Son, H. Kim, Y. Yi, and B. Khrishnamachari, "Base Station Operation and User Association Mechanisms for Energy-Delay Tradeoffs in Green Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1525–1536, 2011.

[37] H. Kim, G. Veciana, X. Yang, and M. Venkatachalam, "Distributed α-Optimal User As-

sociation and Cell Load Balancing in Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, pp. 177–190, 2012.

[38] R. L. G. Cavalcante, S. Stanczak, M. Schubert, A. Aisenblätter, and U. Türke, "Toward Energy-efficient 5G Wireless Communication Technologies," *IEEE Signal Processing Magazine*, vol. 31, no. 6, 2014.

[39] I. Siomina and D. Yuan, "Analysis of Cell Load Coupling for LTE Network Planning and Optimization," *IEEE Transactions on Wireless Commununications*, vol. 11, no. 6, 2012.

[40] ——, "Optimizing Small Cell Range in Heterogeneous and Load-Coupled LTE Networks," *IEEE Transactions on Vehicular Technology*, vol. 99, July 2014.

[41] M. Nawrocki, H. Aghvami, and M. Dohler, *Understanding UMTS Radio Network Modelling, Planning and Automated Optimisation: Theory and Practice.* Wiley, April 2006.

[42] Z. Hasan, H. Boostanimehr, and V. Bhargava, "Green Cellular Networks: A Survey, Some Research Issues and Challenges," *IEEE Communications Surveys and Tutorials*, vol. 13, no. 4, pp. 524 – 540, 2011.

[43] L. M. Correia, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. V. der Perre, "Challenges and Enabling Technologies for Energy Aware Mobile Radio Networks," *IEEE Communications Magazine*, pp. 66 – 72, November 2010.

[44] D. Feng, C. Jiang, G. Lim, J. Leonard J. Cimini, G. Feng, and G. Y. Li, "A Survey of Energy-Efficient Wireless Communications," *IEEE Communications Surveys and Tutorials*, no. 1, pp. 167–178, 2013.

[45] H. Leem, S. Y. Baek, and D. K. Sung, "The Effects of Cell Size on Energy Saving, System Capacity, and Per-Energy Capacity," in *IEEE Wireless Communications and Networking Conference (WCNC), Sydney, NSW, Austrailia*, April 2010, pp. 1–6.

[46] B. Badic, T. O'Farrrell, P. Loskot, and J. He, "Energy Efficient Radio Access Architectures for Green Radio: Large versus Small Cell Size Deployment," in *IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall), Anchorage, AK, USA*, September 2009, pp. 1–5.

[47] L. del Apio, E. Mino, L. Cucala, O. Moreno *et al.*, "Energy Efficiency and Performance in Mobile Networks Deployments with Femtocells," in *IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Toronto, ON, Canada*, September 2011, pp. 107–111.

[48] F. Richter, A. Fehske, and G. Fettweis, "Energy Efficiency Aspects of Base Station Deployment Strategies for Cellular Networks," in *IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall), Anchorage, AK, USA*, September 2009, pp. 1–5.

[49] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental Trade-offs on Green Wireless Networks," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 46–54, 2011.

[50] J. Akhtman and L. Hanzo, "Power Versus Bandwidth-Efficiency in Wireless Communications: The Economic Perspective," September 2009.

[51] C. Khirallah, J. Thompson, and H. Rashvand, "Energy and Cost Impacts of Relay and Femtocell Deployments in Long-Term-Evolution Advanced," *IET Communications, Special Section: Green Technologies for Wireless Communications and Mobile Computing*, vol. 5, no. 18, pp. 2617–2628, 2011.

[52] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power Consumption Modeling of Different Base Station Types in Heterogeneous Cellular Networks ," in *Future Network and Mobile Summit*, 2010, pp. 1–8.

[53] M. Deruyck, E. Tanghe, W. Joseph, W. Vereecken, M. Pickavet, L. Martens, and B. Dhoedt, "Model for Power Consumption of Wireless Access Networks," *IET Communications*, vol. 5, no. 4, pp. 155–161, 2010.

[54] B. Debaillte, A. Giry, M. J. Gonzales, L. Dusspot *et al.*, "Opportunities for Energy Savings in Pico/Femto-cell Base-Stations ," in *Future Network and Mobile Summit*, 2011, pp. 1–8.

[55] S. Mclaughlin, P. M. Grant, J. Thompson, H. Haas, D. Laurenson, C. Khirallah, H. Ying, and R. Wang, "Techniques for Improving Cellular Radio Base Station Energy Efficiency," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 10 – 17, 2011.

[56] I. Ashraf, F. Boccardi, and L. Ho, "SLEEP Mode Techniques for Small Cell Deployments," *IEEE Communications Magazine*, pp. 72–79, August 2011.

[57] H. Congzheng, T. Harrold, S. Armour, I. Krikidis *et al.*, "Green Radio: Radio Techniques to Enable Energy-Efficient Wireless Networks," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 46–54, 2011.

[58] L. Benini, A. Bogliolo, and G. D. Micheli, "A Survey of Design Techniques for System-Level Dynamic Power Management," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 8, no. 3, pp. 299 – 316, 2000.

[59] K. Son, S. Chong, and G. de Veciana, "Dynamic Association for Load Balancing and Interference Avoidance in Multi-Cell Networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3566–3576, 2009.

[60] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell Association and Interference Coordination in Heterogeneous LTE-A Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1479 – 1489, 2010.

[61] B. Rengarajan and G. de Veciana, "Practical Adaptive User Association Policies for Wireless Systems With Dynamic Interference," *IEEE/ACM Transactions on Networking*, vol. 19, no. 6, pp. 1690 – 1703, 2011.

[62] A. Fehske, H. Klessig, J. Voigt, and G. Fettweis, "Flow-Level Models for Capacity Planning and Management in Interference-Coupled Wireless Data Networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 164 – 171, 2014.

[63] I. Siomina and D. Yuan, "On Constrained Cell Load Coupling With Applications in LTE Networks," *IEEE Communications Letters*, vol. 18, no. 10, 2014.

[64] C. K. Ho, D. Yuan, and S. Sun, "Data Offloading in Load Coupled Networks: A Utility Maximization Framework," *IEEE Transaction on Wireless Communications*, vol. 13, no. 4, pp. 1921–1932, 2014.

[65] H. Klessig, A. Fehske, G. Fettweis, and J. Voigt, "Cell Load-Aware Energy Saving Management in Self-Organizing Networks," *IEEE 78th Vehicular Technology Conference (VTC Fall), Las Vegas, NV, USA*, pp. 1 – 6, September 2013.

[66] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A Simple Cooperative Diversity Method Based on Network Path Selection," *IEEE Journal on Selected Areas in Communi-*

cations, vol. 24, no. 3, 2006.

[67] C.-Y. Hong and A.-C. Pang, "3-Approximation Algorithm for Joint Routing and Link Scheduling in Wireless Relay Networks," *IEEE Wireless Communication Letters*, vol. 2, no. 8, pp. 856–861, 2009.

[68] K. Sundaresan and S. Rangarajan, "On Exploiting Diversity and Spatial Reuse in Relay-enabled Wireless Networks," in *Proceedings of the 9th ACM international symposium on Mobile ad hoc Networking and Computing, New York, NY, USA*, November 2008, pp. 13–22.

[69] X. Guo, W. Ma, Z. Guo, X. Shen, and Z. Hou, "Adaptive Resource Reuse Scheduling for Multihop Relay Wireless Network Based on Multicoloring," *IEEE Communication Letters*, vol. 3, no. 12, pp. 176 – 178, 2008.

[70] S. Zhou, A. Goldsmith, and Z. Niu, "On Optimal Relay Placement and Sleep Control to Improve Energy Efficiency in Cellular Networks," in *Proceedings of IEEE International Conference on Communications (ICC), Kyoto, Japan*, June 2011, pp. 1 – 6.

[71] D. Lee, S. Zhou, and Z. Niu, "Multi-hop Relay Network for Base Station Energy Saving and Its Performance Evaluation," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM), 2011, Houston, TX, USA*, December 2011, pp. 1–5.

[72] Z. Yang, Q. Zhang, and Z. Niu, "Throughput Improvement by Joint Relay Selection and Link Scheduling in Relay-assisted Cellular Networks," *IEEE Transaction on Vehicular Technology*, vol. 61, no. 6, 2012.

[73] Z. Yang and Z. Niu, "Load Balancing by Dynamic Base Station Relay Station Association in Cellular Networks," *IEEE Wireless Communication Letters*, vol. 2, no. 2, pp. 155–158, 2013.

[74] ——, "Energy Saving in Cellular Networks by Dynamic RS-BS Association and BS Switching," *IEEE Transaction on Vehicular Technology*, vol. 62, no. 9, pp. 4602 – 4614, 2013.

[75] R. Nettleton and H. Alavi, "Power Control for a Spread Spectrum Cellular Mobile Radio System," in *33rd IEEE Vehicular Technology Conference*, 1983, pp. 242–246.

[76] J. Zander, "Performance of Optimum Transmitter Power Control in Cellular Radio Sys-

tems," *IEEE Transactions on Vehicular Technology*, vol. 41, pp. 57–62, February 1992.

[77] G. J. Foschini and Z. Miljanic, "A Simple Distributed Autonomous Power Control Algorithm and its Convergence," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, pp. 641–646, November 1993.

[78] N. Bambos, S. Chen, and G. Pottie, "Channel Access Algorithms with Active Link Protection for Wireless Communication Networks with Power Control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 583–597, 2000.

[79] C. W. Tan, D. P. Palomar, and M. Chiang, "Energy-Robustness Tradeoff in Cellular Network Power Control," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 912–925, June 2009.

[80] S. Stanczak, M. Kaliszan, and N. Bambos, "Decentralized Admission Control for Power-Controlled Wireless Links," *Preprint available at http://arxiv.org/abs/0907.2896*, 2010.

[81] R. D. Yates, "A Framework for Uplink Power Control in Cellular Radio Systems," *IEEE Journal of Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, 1995.

[82] H. Boche1 and M. Schubert, "Characterization of the Structure of General Interference Functions," in *International Symposia on Information Theory (ISIT), Nice, France*, June 2007.

[83] C. Ho, D. Yuan, L. Lei, and S. Sun, "Power and Load Coupling in Cellular Networks for Energy Optimization," *IEEE Transactions on Wireless Communications*, vol. 44, no. 1, pp. 509 – 519, 2014.

[84] O. O'Meara, *Introduction to Quadratic Forms*. New York: Springer-Verlag, 2000.

[85] S. Lang, *Introduction to Differentiable Manifolds. Revised 2nd Ed.* New York: Springer-Verlag, August 2002.

[86] A. Berman and R. Plemmons, "Nonnegative Matrices in the Mathematical Sciences, Classics in Applied Mathematics," *SIAM*, 1994.

[87] L. Liberti and C. C. Pantelides, "An Exact Reformulation Algorithm for Large Nonconvex

NLPs Involving Bilinear Terms," *Journal of Global Optimization*, no. 2, 2006.

[88] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing Sparsity by Reweighted $l_1 Minimization$," *Dec.*2008.

[89] D. R. Hunter and K. Lange, "A Tutorial on MM Algorithms," Feb. 2004.

[90] 3GPP, *Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode*, 3GPP Std. 36.304 v12.3.0, 2014.

[91] ——, *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer; Measurements*, 3GPP Specification TR36.214 v12.1.0, 2014.

[92] E. Ternon, P. Agyapong, H. Liang, and A. Dekorsy, "Context-Aware Handover Optimization for Relay-Aided Vehicular Terminals," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), Instanbul, Turkey, April*, 2014, pp. 2811 – 2816.

[93] A. J. Laub, *Matrix Analysis - for Scientists and Engineers.* SIAM, 2005.

[94] L.Vandenberghe and S.Boyd, "Semidefinite programming," *SIAM, Review*, vol. 38, no. 1, pp. 49–95, 1996.

[95] H. Sherali and W.P.Adams, "A reformulation-linearization technique for solving discrete and continuous nonconvex problems," *Kluwer, Dordrecht*, 1998.

[96] L. Hurwicz and M. K. Richter, "Implicit Functions and Diffeomorphisms without C^1," *Advances in Mathematical Economics*, 1995.