國立臺灣大學電機資訊學院電信工程學研究所

博士論文

Graduate Institute of Communication Engineering
College of Electrical Engineering and Computer Science
National Taiwan University

Doctoral Dissertation

睡眠模式在綠能無線網路之建模、設計、與最佳化

Modeling, Design, and Optimization for Green Wireless Networks with Sleep Mode Operations

張正義

Chen-Yi Chang

指導教授：許大山 博士

共同指導教授：廖婉君 博士

中華民國 103 年 1 月

January 2014

國立臺灣大學博士學位論文
口試委員會審定書

睡眠模式在綠能無線網路之建模、設計、與最佳化

The Modeling, Design, and Optimization for Green Wireless Networks with Sleep Mode Operations

本論文係張正義 君（D95942022）在國立臺灣大學電信工程學研究所完成之博士學位論文，於民國 103 年 01 月 20 日承下列考試委員審查通過及口試及格，特此證明

口試委員：許大山 簽名（指導教授）　簡榮宏 簽名

高榮鴻

蘇炫榮　洪樂文

蔡育仁　謝志明

林永松

所　長　吳宗霖 簽名

誌謝

在經過這麼多年的歲月，終於來到了即將畢業這一刻，在漫長的博士班生涯裡，經歷了許多挫折以及孤獨的時光，但也在這樣的磨鍊與思考的過程當中，讓我有所成長領悟，而在這當中，培養出的態度與毅力，也對於我未來的人生有著長遠的影響與幫助。感謝一路上大家的支持與鼓勵。

首先，感謝我的指導教授許大山教授多年來的指導，對於我在研究上的啟發以及研究上的品味有著深遠的影響，教授對於一個博士的期許，讓我能夠朝著一個明確的方向邁進。感謝我的共同指導教授廖婉君教授的指導，慷慨的提供在各方面協助，與我分享研究上和生活上的經驗，並且時常鼓勵我，讓我能夠保持研究上的熱忱。感謝謝宏昀教授，花費很多時間與我討論我的研究和論文，並在研究成果的呈現上和論文寫作上給予我很大幫助。感謝我的其他口試委員，蘇炫榮教授、蔡育仁教授、林永松教授、洪樂文教授、高榮鴻教授，接受我的邀請並且給予我的論文諸多寶貴的建議。感謝電信所多位教授的鼓勵，特別是陳光禎教授、葉丙成教授、吳宗霖教授，時常勉勵我畢業就在不遠的前方。

另外，感謝電信所趙姐、欣梅、志豪、惠元、雅雯，多年來對於我特別的照顧與協助，忍受我的任性與嘮叨，讓我即使平常一個人在實驗室裡也不覺得孤單。感謝電信所很多的學長與學弟妹們的陪伴，感謝 BL524 實驗室，特別是蔡隆盛學長、宇鵬、存志、冠錡、家榮、孝賢，感謝 BL608 實驗室，特別是張晉嘉學長、孫哥、昆錚、易翰、博瀚、柏翰、昆霖、昱均、秀秀、筑涵，感謝 BL503 實驗室，特別是劉顏慶、郭軒豪、蔡華龍、林典育，感謝當中許多人常常陪我在 118 巷、女九、活大之間做選擇，感謝與我一起做計畫的學弟，家榮、厚昇、炳成、勝順、景凱、易翰、博瀚、柏翰、昆霖、昱均，感謝與我一起做研究的學弟妹，昆霖、裕恆、小晴，以及感謝 EE530 實驗室很多的學弟妹時常陪我打球。感謝我很多的好朋友時常約我出去散心與聊天，特別在我心情低落的時候，陪著我大吃大喝並且給予我很大的鼓勵。

最後感謝我的父母、我的妹妹們、我的家人們多年來的支持與鼓勵，在我當初決定要攻讀博士時，就對我的選擇無怨無悔的支持。在經過這麼多年的歲月，現在我終於可以把即將到手的博士學位獻給你們。

摘要

最小化耗能是綠色節能無線通訊的一個基本目標，其中最有效且最有潛力的方法之一是根據網路流量來適應性的開關無線電收發器，這樣的想法可以被廣泛的運用在各種無線通訊設備或是裝置上。在本論文中，我們將探討睡眠模式在無線網路中的運作。我們對於一個網路中的節點功能性是可以互相被取代而且願意一起合作來達成整體目標的不自私(altruistic)網路感興趣，為了綠化傳統無時無刻都是主動運作的網路，我們利用睡眠模式所帶來的額外自由度(degrees of freedom)來節省網路的能量，使其成為一個在必要時刻才主動運作的網路。

在行動網路中，根據網路中的流量來開關基地台是一個節省網路能量的有效方式，然而這樣的運作可能會產生網路覆蓋漏洞(coverage holes)。在這樣的問題中，我們嘗試在保持整體網路覆蓋率的情況下，藉由動態的開關基地台來最小化整體網路的耗能。我們推導出在最小化每單位覆蓋面積的耗能下，每個基地台的最佳的覆蓋細胞大小(cell size)。在滿足網路覆蓋率的前提下，我們提出一個多項式時間複雜度的流量察覺(load-aware)的基地台開啟演算法。除此之外，我們展示了為了增加網路中熱點吞吐量(throughput)，藉由佈建的小細胞(small cell)達成的網路密致化(network densification)也可以在低流量的期間改善整體網路耗能。此外，我們探討異質網路(heterogeneous networks)下有可能的網路架構與多模(multi-mode)基地台運作方式並且建議非對稱的基地台與移動裝置的連線將有潛力在低流量時大幅降低網路耗能。

在無線多重跳躍中繼網路(wireless multi-hop relay network)中，睡眠排程是一個有效的方法來降低網路耗能，但是通常會造成傳送訊息時的額外延遲。為了尋找在設計綠能多重跳躍中繼網路的洞見，我們建立了一個模型來分析和最佳化無線多重跳躍中繼網路在使用睡眠機制時，效能間的權衡(trade-offs)關係。進一步的說明，我們提出了一個隨機醒來的網路(random wakeup network)下使用投機式的路由(opportunistic routing)所建造的框架來分析。我們發現在最佳的參數設定時，整個網路用來偵測訊息的能量消耗占整體網路的$\frac{\alpha-1}{\alpha+2}$，其中$\alpha$是路徑損耗指數(path loss exponent)，在最佳的參數設定下，我們發現整個網路用來偵測訊息的耗能應該要和整個網路用來做實際通訊的耗能是差不多的，並且我們發現整個網

路在最佳參數設定下的最小耗能和傳送速度的$\frac{\alpha-1}{\alpha+2}$次方成正比。另一方面，我們研究在無線多重跳躍網路下訊息交換(information exchange)與訊息散播(information dissemination)的耗能最小化。我們提出了一個隨機散播(random gossip)下週期性醒來的分析模型。我們發現在最佳的參數設定下，每個週期醒來廣播的節點數量只和路徑損耗指數和網路維度有關，這說明了不論增加或減少網路的大小都不會影響整個網路的最佳參數設定，我們得到最佳的參數設定為擁有可拓展性(scalability)的性質。另外我們藉由模擬的方式展示我們提出的模型可以被很好的運用在效能的朔模(performance modeling)與最佳參數的預測。除了跨層最佳化之外，我們提出的模組可以在無線多重跳躍中繼網路下，藉由感知(cognition)與最佳化來巧妙的調整軟體可重組化(software-configurable)的功能用以達成整體網路的目標。

關鍵字：覆蓋漏洞(coverage holes)、異質網路(heterogeneous networks)、低流量(low traffic density)、多重跳躍中繼(multi-hop relay)、訊息偵測(message detection)、投機式路由(opportunistic routing)、耗能與延遲的權衡(energy-delay trade-offs)。

Abstract

Energy consumption minimization is a fundamental goal for green wireless communications. One of the most effective and promising way to save energy is to power off radio transceivers adaptively according to network traffic conditions. This idea can be utilized in a wide spectrum of wireless communication devices and equipments. In this dissertation, we explore the "sleep mode" operations in wireless networks. We are interested in altruistic networks in which nodes are interchangeable in functionalities and are willing to cooperatively achieve network level goals. To "greenize" traditional networks from always-on networks to necessary-on networks, we exploit the additional degrees of freedom by sleep mode operations for network energy conservation.

In cellular networks, switching on/off base stations (BSs) according to network traffic profile is an effective way to conserve network energy. However, such operations may create coverage holes in the network. We attempt to minimize the total power consumption of the network by switching on and off BSs adaptively while maintaining the network coverage. We derive the optimal cell size for minimizing BS power consumption per unit coverage area and propose a polynomial-time load-aware algorithm for energy-efficient BS activation while avoiding creating coverage holes. Besides, we demonstrate that network densification with small cells for bursting throughput in hot spot areas can also improve network energy savings during the low traffic load period. Furthermore, we explore potential network structures with multi-mode BS operations in heterogeneous networks (HetNets) and suggest that asymmetric BS-MS (mobile stations) connections with uplink by small cells and downlink by macro cells will be energy-efficient during the low traffic load periods for green HetNets.

For wireless multi-hop relay networks, while applying sleep-awake scheduling is an effective way to reduce network energy consumption, it usually comes at a price of additional delay for message delivery. To seek insights for the design of green wireless multi-hop relay networks, we develop a model for the analysis and optimization of performance trade-offs in wireless networks while sleep-awake mechanisms are applied. Specifically, we propose a random wakeup network with opportunistic relay as a framework for our analysis. Under optimal settings, we find that the proportion of power for message detection in the entire network is $\frac{\alpha-1}{\alpha+2}$, where α is the path loss exponent; the energy consumed for message detection should be of the same order as that for actual communications. Moreover, we find that the minimal network power consumption under optimal operations grows at $\frac{\alpha-1}{\alpha+2}$-th order of the delivery speed, meaning that the investment of network power can efficiently reduce delivery delay. Besides, we investigate network energy minimization for information exchange and information dissemination in wireless multi-hop relay networks. We propose a random gossip network with periodic listening as a framework for our analysis.

We find simple rules govern the optimal settings for network-wide information exchange in the random gossip network. One key relationship is that the optimal number of nodes broadcasting messages in a time epoch within the transmission range depends only on the path loss exponent and the network dimensionality. It is shown that neither increasing nor decreasing the physical network size will affect the optimal value of these design parameters. The optimal setting for information exchange is a scalable solution. We show through simulation results that our proposed frameworks are well applicable for performance modeling and parameter optimization in wireless multi-hop relay networks when sleep-awake operations are adopted. Beyond cross-layer optimizations, our proposed frameworks can facilitate the cognition and optimization for ingeniously adapting software-configurable functionalities to achieve network goals in wireless multi-hop relay networks.

Keywords– coverage holes, heterogeneous networks, low traffic density, multi-hop relay, message detection, opportunistic routing, energy-delay trade-offs.

Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

The evolution of wireless communication networks brings us fold-by-fold throughput increase last decades. To meet explosive growth of data volume requests in the future, potential directions for bursting capacity in next generation wireless communication networks can be peeked from pioneers. Since spatial frequency reuse serves as a simple and effective design technique for bursting network capacity [1], small-size cells are expected to densely deployed to further drain the capacity from frequency reuse. Recently, massive MIMO is proposed to exploit the fertile channel diversification in wireless environments [2, 3] by deploying a large excess of antenna at BS side. By utilizing the property that channel vectors of different users will tend to be more orthogonal as the increase of antenna, massive MIMO is expected to scale up cell capacity 10 times or more. Besides, it is known that the network capacity will increase but per node capacity will decrease as the increase of nodes in the network [4]. Putting dumb nodes in the network with multi-hop relaying can also enjoy extra capacity gain. In conclusion, no matter what kind of approaches will be adopted, extensive network densification (e.g., BSs, relays, antenna) is expected in the future. An illustration of our vision for wireless networks in the future is given in Figure 1-1.

It is recently pointed out that carbon emission and energy consumption of wireless networks (10-20% compounded annual growth rate (CAGR)) reach a level which can not be ignored [5, 6]. If we are not aware for what is paid for the development of advanced wireless technologies, the potentially exponential increase of electricity

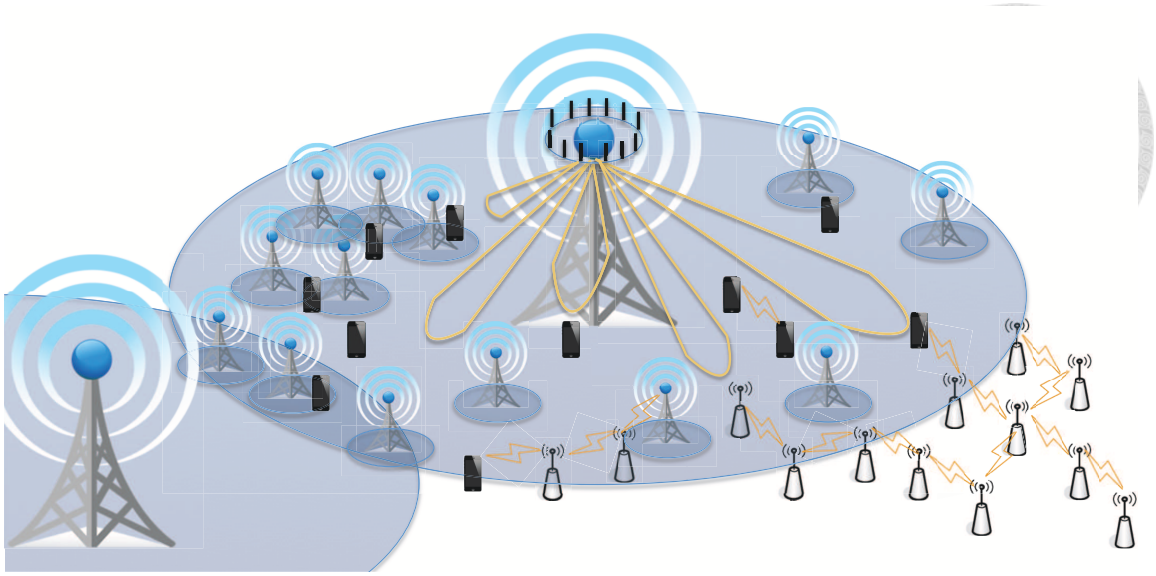


Figure 1-1: The vision of wireless networks in the future. To increase network capacity and support diverse applications, a large amount of equipment is expected to be deployed in the networks. Extensive network densification (e.g., BSs, relays, antenna) is expected in future wireless networks to support advanced wireless technologies.

consumption will exacerbate the energy crisis and global warming. "Greenizing" now becomes a key design objective to mitigate energy crisis and suppress carbon emissions in wireless networks. From economic perspectives, the growth of data volume does not directly reflect to operator's revenue. The data rate demands will grow with 92% CAGR in 2012-2015 (estimated by Cisco [7]) but revenue of network operators is growing only at 13.4% in 2013 (estimated by ABI Research). On the other hand, reducing energy cost will directly contribute to the reduction of network operational cost. In view of the issues discussed above, we believe that how to greenize the ever dense wireless networks will be important and challenging in the near future.

Network traffic variations provide great opportunities to save network energy dynamically. Traditionally, most of network designers and operators focus on how to plan and maintain a network in peak load conditions. However, a network is not always at such high load conditions. Based on the measurement results in wireless networks, the network energy consumption without adapting to network traffic loading causes a lot of energy waste. With more diversification of wireless network applications (e.g., smart metering versus 3D video streaming) in the future, traffic

demands is expected to vary more dramatically in various domain (e.g., space, time, frequency, and application). While a network shifts from high traffic load conditions to low traffic load conditions, we believe lots of network energy can be saved.

The aggregated power consumption for detecting network traffic may easily surpass the power consumption for actual transmissions in a network with low traffic load density. An intuitive approach to conserve energy in low traffic density network is to put "idle" nodes into so called "sleep mode". The term "sleep" generally refers to the practice where a communication terminal disengages itself for a period of time from the network, during which the transceiver is shut off to greatly reduce the power consumption. With the inclusion of the sleep state operation, in this dissertation, we desire to seek for general design principles and capture key network characteristics for green wireless networks while sleep-awake scheduling is adopted in the networks.

1.1 Motivation

The rapid exacerbation of energy crisis and the dramatic rise of carbon emissions caused by wanton abuse of energy resources gather our attention to green technologies. In wireless networks, the explosive access demands results in the exponential growth of energy consumption during last decades. Technology trends toward extensive network densification reveal that network energy consumption will continuously expand in the future. Nowadays most of human life is inseparable of wireless networks. The traffic variations resulting from human daily routine and diverse applications in wireless networks motivate us to investigate the intuitive "sleep mode" operation for energy conservation in wireless networks.

1.2 Summary of Contributions

The high level contribution of this dissertation is to provide general design principles of sleep mode operations for green wireless networks. We build analytical frameworks with sleep mode operations for quantifying network performance trade-offs and seek-

ing key properties in an optimized network. Applying the obtained design insights, we propose energy-efficient algorithms with sleep mode operations and suggest promising network structure for network energy conservation. Furthermore, we demonstrate our results can be applicable for the optimization of software-configurable functionalities in various wireless networks with different applications. Our proposed frameworks are expected to facilitate the cognitive processes in wireless relay networks while satisfying network level goals [8]. The specific contributions of this dissertation are highlighted as follows.

1.2.1 Sleep Mode Operations for Green Cellular Networks

To conserve network power in cellular networks, a promising approach is switching off BSs at low traffic load. However, the switching-off operations may induce network coverage holes. We attempt to minimize the total power consumption of the network by switching on and off BSs adaptively while maintaining the network coverage. We find that BS activation problem for minimal network power consumption with full network coverage preservation is an NP-hard problem. We derive the optimal cell size for minimizing area power consumption and and propose a polynomial-time load-aware algorithm for energy-efficient BS activation in HetNets. The simulation results show that our algorithm can approach the minimal network power consumption. Besides, we demonstrate that network densification by small cells for bursting throughput in hot spot areas can also be beneficial in saving network energy during the low traffic load period. Finally, we explore potential network structures with multi-mode BS operations in HetNets and suggest that uplink by small cells and downlink by macro cells will be promising for network energy saving during low traffic load periods.

1.2.2 Sleep Mode Operations for Green Wireless Multi-Hop Relay Networks

Conservation of network energy is a critical design goal for wireless communications comprised of battery-powered devices. While applying sleep-awake scheduling is an

effective way to reduce network energy consumption, it usually comes at a price of additional delay for message delivery. For message delivery in wireless relay networks, we develop a model for the analysis and optimization of performance trade-offs in wireless networks while sleep-awake scheduling mechanisms are applied. Specifically, we propose a random wakeup network with opportunistic relay as a framework for our analysis. In the proposed framework, nodes wake up randomly to detect potential incoming messages in a duty-cycle period, and messages are opportunistically relayed by these wakeup nodes. We first derive the closed-form expressions for network power consumption in the random wakeup network, and then we apply the model in an optimization problem. The goal is to minimize the total network energy consumption with a requirement on the end-to-end delay by jointly selecting key design parameters including transmission range, duty-cycle period, and wakeup density. We find that for a network that employs the optimal routing and sleep parameters thus obtained, the proportion of power for message detection in the entire network is $\frac{\alpha-1}{\alpha+2}$, where α is the path loss exponent. That is, under the optimal settings, the energy consumed for message detection should be of the same order as the energy consumed for actual communications. Moreover, we find that the minimal network power consumption under optimal operations grows at a rate of the $\frac{\alpha-1}{\alpha+2}$-th order of the message delivery speed, meaning that the investment of network power can efficiently reduce message delivery delay. We show through simulation results the proposed framework is well applicable for performance modelling and operational parameter optimization in wireless relay networks when sleep mode operation is adopted.

For information exchange and information dissemination in green wireless relay networks, we propose a "random gossip" wireless-relay network (RGN) as a framework for our analysis. In the RGN, we consider that each node initiates a message and wishes to exchange its message with all the other nodes in the network. We will demonstrate that our results in the RGN can indeed predict the optimal value of operational parameters for information exchange as well as information dissemination in actual WSNs. We find key properties for the three design parameters in the optimized RGNs. First, a key relationship is that the optimal number of nodes

broadcast messages within the transmission range in a sleep period depends only on the path loss exponent and the network dimensionality. It does not depend on factors such as the physical network size, the reception mode energy, and the message origination rate. Next, the existing optimal number of broadcasting nodes in a time epoch within the transmission range shows that either shortest hop transmission or blind flooding is rarely optimal for power consumption minimization. Last but not the least, it is shown that neither increasing nor decreasing the physical network size will affect the optimal value of these design parameters. The optimal setting for network power minimization is a scalable solution.

1.3 Outline

The rest of this thesis is organized as follows. In Chapter 2, we explore the sleep mode operation in cellular networks. We investigate how to switch off BSs in an energy-efficient manner with the consideration of network traffic variation and network coverage preservation. We formulate the green network coverage preservation problems and propose energy-efficient BS activation algorithms in HetNets. Besides, we evaluate potential network structures with multi-mode operations of BSs for network energy saving. In Chapter 3 and Chapter 4, we explore the sleep mode operations in wireless relay networks. We investigate the minimization of network energy consumption while satisfying network level goals. In Chapter 3, we investigate message delivery with sleep mode operation for green wireless relay networks. We propose the random wakeup model with opportunistic routing and demonstrate our derived results can be well applicable for message delivery in many proposed wireless relay networks. In Chapter 4, we investigate information exchange with sleep mode operations in green wireless relay networks. We propose the random gossip network with periodic listening and demonstrate that the derived results can be applicable for both information exchange and information dissemination in wireless relay networks. Finally, concluding remarks are given in Chapter 5.

Chapter 2

Sleep Mode Operations for Green Cellular Networks

Energy consumption minimization is a fundamental goal for green communications [6]. It is estimated that the cost spent in electricity globally for mobile networks is more than $10 billion dollars annually, and among all the network components, radio base stations (BSs) may consume over 80% of the energy in the network [9]. Therefore, it is very important for mobile operators to design energy-efficient mechanisms for BSs.

One effective way for energy saving in cellular access network is to turn on and off BSs adaptively. The daily traffic profile in [9] shows that the long-term behavior of BSs is predictable (e.g., day versus night), and for almost half of a day, the traffic load of a BS is low. However, a BS with low or no traffic load still consumes more than 90% of its peak power. As a result, switching off the cells which are under low traffic condition is one of the most effective ways for energy saving. A similar approach is also taken in the standards of next generation wireless networks (e.g., 3GPP LTE-Advanced), in which BSs are switched off to conserve energy [10].

Energy conservation by switching off BSs may encounter several challenges. An important problem, identified in many studies [9,11–15], is the traffic sharing problem induced by the switched off BSs. Users in the switched off cells shall be handed over to other nearby cells. The work in [9] points out that energy can be saved by adaptively switching on and off BSs with the consideration of daily traffic fluctuations. Based on

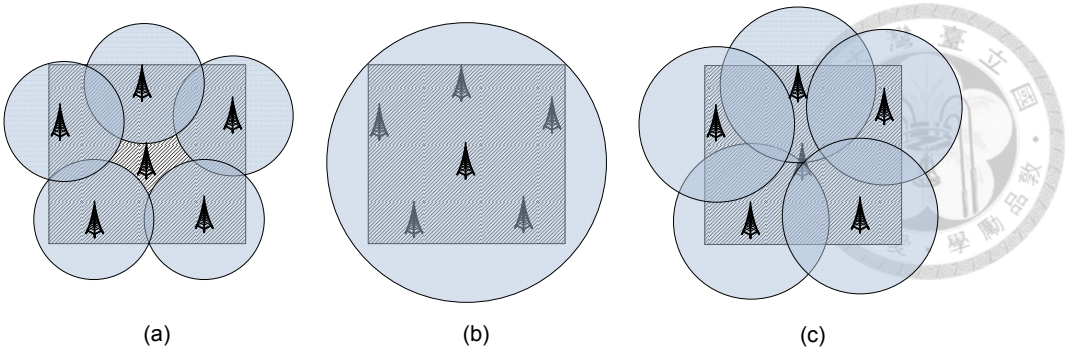


Figure 2-1: Illustration of the coverage hole problems due to switching on/off BSs. The rectangular region (shaded) is the targeted covering region. Five BSs are deployed in the region. The disk centered at a BS is the coverage of the BS. (a) Coverage hole present when a BS is switched off. (b) A BS activation strategy with high power level. The coverage of target area is ensured. (c) A BS activation strategy with low power level. The coverage of target area is ensured.

reassociating users among BSs with the available location information, the authors in [13] try to find a minimum set of active BSs with which users can be associated. In [14], the authors propose centralized and decentralized energy-efficient BS on-off strategies according to network traffic variations and show the trade-off between outage performance and energy efficiency. Taking the dynamics of user population into consideration, the authors in [15] propose an energy-efficient BS operation mechanism with user association policies for the trade-off between network energy saving and flow-level performance. According to the variations of daily traffic demand and inhomogeneous user population, these studies show that load balancing and user association by switching on/off BSs can significantly reduce the network energy consumption. However, these studies all assume that the network coverage will not be affected by switching off BSs, which may not always be the case as indicated in [10] (see Fig. 1 (a) for example). The work in [16] shows that the coverage maintenance for mobile cellular networks is a key challenge in energy-saving operations of BSs. By turning off a BS in the low traffic condition to save energy, the cells in the neighborhood of shut-down cells may need to extend their coverage (by increasing their power levels) so as to avoid creating coverage holes (see Figs. 1 (b) and (c) for example). However, in the literature, this coverage hole problem induced by the power-off BSs is rarely investigated.

The coverage problem has been treated in great details in wireless sensor networks [17, 18]. Many research studies (e.g., [19]) show that to find the minimum number of nodes to cover an area is often an NP-hard problem. In such networks, the key design objective is to maximize the lifetime of the network by uniformly distributing the residual energy in the whole network. A decentralized node scheduling scheme with limited local information is the typical technique used for practical applications (e.g., Optimal Geographical Density Control (OGDC) in [20]). However, in wireless sensor networks, the coverage range may not be adjustable and full network coverage may not be ensured (e.g., only coverage for nodes without mobility). In cellular networks, we must avoid the signal loss of users who turn on at uncovered region or who move through coverage holes. Therefore, existing solutions for wireless sensor networks may not be directly applicable for green cellular networks. Although the coverage and traffic rate requirements can be satisfied in traditional network planning problems, the goal of most network planning problems is to reduce the network deployment cost at peak load conditions instead of saving network power while network traffic varies. Most recent works are focused on energy-efficient network planning for coverage extension via BS cooperations (e.g., in [21]). However, the potential benefits of dynamic power control for cell breathing are still not explored. On the other hand, to maintain network coverage when traffic varies, pre-configured BS on/off patterns for hexagonal BS topologies are analyzed to show the potential energy saving capabilities by joint power control and cell activation in [22]. However, these analytical results may not be easily applicable to heterogeneous networks (HetNets).

HetNets with high capacity small cells [23, 24] (e.g., picocells or femtocells) are promising for bursting throughput at hotspot area in next generation cellular systems. To meet the future demands on data access rates (e.g., Cisco [7] estimates that Compounded Annual Growth Rate (CAGR) of network traffic reaches 92% in 2010-2015), Nokia Siemens Networks estimates that the number of BSs will increase by 10-fold in 2020 [25]. Thus, we can reasonably expect that a large number of small cells will be deployed to cooperatively provide service with macro cells in the future. In addition to increasing network capacity, these small cells may also improve the

network power savings during the low load periods by maintaining the coverage of certain small areas.

The diversification of BSs in HetNets (e.g., macro BSs, femtocells, small cells, etc.) will bring new opportunities and challenges in the traditional only macro cell scenario. The idea of providing exactly the same functionality of various BSs all the time may need to be re-examined. Recently, many proposed ideas shift the design paradigms such as coverage extension for mitigating asymmetric transmit power between macro cells and small cells [26], dual connectivity between macro cells and small cell to utilize the inherent asymmetric uplink and downlink coverage in HetNets [27], and umbrella cells to alleviate handover of mobility users [24]. Furthermore, control signalling reduction via the functionality separation of BSs is proposed in [28] to increase the network energy efficiency. The network diversification in conjunction with "flexible BS operations" will be promising in the future.

In view of the issues discussed above, in this chapter, we will explore the green network coverage problem in HetNets, in which we attempt to minimize the total energy consumption in cellular systems by joint BS activation and power control given the requirements that the network coverage and traffic load must be ensured. To follow the shifted design paradigms which break the traditional one-to-one association between users and BSs, we propose the "reception mode" design and demonstrate the potential gain of asymmetric uplink and downlink connections with the support of multi-mode operation of BSs for green HetNets. The major contributions are summarized as follows.

- We formulate the mathematical green network coverage problem in cellular networks. The switching on/off operation and power control of BSs are explored. We decompose the green network coverage problem into two sub-problems: BS operational power optimization problem and minimum-power BS activation problem.

- We derive the optimal power and optimal cell size of each active BS for minimizing BS power consumption per unit coverage area. We also derive a lower

bound for the minimal power consumption of the entire network. This bound can be jointly used with any BS deployment strategy.

- We show that the minimum-power BS activation problem is NP-hard and propose a polynomial-time load-aware algorithm for the activation of BSs while avoiding creating coverage holes in the network. We demonstrate that the performance of the proposed mechanism is indeed a very efficient power saving mechanism in various BS deployment strategies via extensive simulations. Compared with the proposed algorithm in [14], 44.5% more power will be consumed by their approach during the low traffic load period.

- We propose the "reception mode" design of BSs in conjunction with asymmetric MS-BS uplink and downlink connections to further explore the opportunities of network energy saving. We suggest that uplink by small cells and downlink by macro cells is promising for network energy saving during the low traffic load period for green HetNets. Compared with traditional symmetric connection by macro cells, this approach can save network power more than 6 dB in HetNets by simulations.

2.1 System Model and Problem Description

2.1.1 BS Operation Model

We consider a set of BSs, denoted by $\mathcal{B} = \{B_1, B_2, \ldots, B_M\}$, in a cellular network. We assume that the set of BSs is deployed in such a way that the transmission rate requirements induced by the service level agreement (SLA) to mobile stations (MSs) are satisfied. This assumption is reasonable as mobile operators deploy their BSs to cover the service area so as to satisfy the SLAs of their MSs.

In the network, each BS can be switched on/off according to the network condition. A BS which serves users in the network is referred to as an *active* BS. We assume that the long-term behavior of traffic in the network is predictable, and the daily traffic profiles in different BSs are highly correlated (e.g., BSs tend to have low traffic

load at midnight). Therefore, the focus of this chapter will be on the optimization of energy saving during off-peak hours.

Note that the switching-on/off operation of a BS we consider here is based on the average traffic load (e.g., in each hour). It is different from the sleep mode operation of radio transceivers which may be turned on and off more dynamically according to the load of instantaneous traffic.

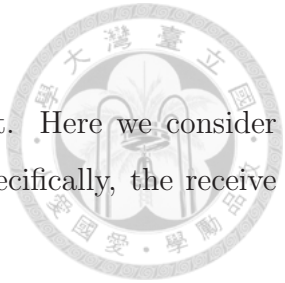
2.1.2 BS Power Consumption Model

The studies in [29, 30] show that the power consumption in a cellular BS (e.g., GSM, WCDMA, LTE, etc.) can be modelled by a simple but surprisingly accurate model:

$$P_m = \eta_m^{-1} P_{t,m} + P_{c,m} + \eta_m^{-1} \sum_k P_{k,m}, \tag{2.1}$$

where P_m represents the total power consumption of B_m, $P_{t,m}$ denotes the maximal transmit power of B_m, η_m and $P_{c,m}$ are the parameters determined by the system hardware and radio access techniques, and $P_{k,m}$ stands for the link level power consumption from B_m to MS k. Specifically, η_m depends highly on the efficiency of the power amplifier, $P_{c,m}$ models the power consumption which is independent of the radiation power such as the power consumed in the cooling system, and $\sum_k P_{k,m}$ models the aggregated MS-BS link power consumption which depends on the number of active MSs and traffic loads. In this chapter, we call the parameter η_m as the *transmit power efficiency* and the parameter $P_{c,m}$ as the *constant operational power* of a BS. In general, the constant operational power is a large constant which is non-negligible [11, 12]. Based on the measurement results of deployed BSs, it is showed that the power consumption of a BS with low load is almost the same as the power consumption of a BS with peak load [29] (e.g., the power consumption varies about 3% in a UMTS BS and about 2% in a GSM BS during a period of several days). For simplification, we omit the link level power consumption $P_{k,m}$ in our analysis. We will show that this simplification is reasonable via simulations in Section 2.3.4.

2.1.3 Channel Model



The propagation channel depends on the physical environment. Here we consider a channel model with path loss and shadowing effects [31]. Specifically, the receive power of an MS, denoted by P_r, is modelled by

$$P_r = K\Psi d^{-\alpha} P_{tx}, \tag{2.2}$$

where P_{tx} represents the transmit power, d is the separation between the BS and the MS, α is the path loss exponent, K is used to model the impacts from antenna heights, carrier frequency, and antenna patterns [32], and Ψ models the shadowing effects in the environment. Note that Ψ is a random variable with a log normal distribution. Therefore, the random variable $10\log_{10}(\Psi)$ follows the normal distribution with zero mean and standard deviation σ_Ψ. Here distance d is a normalized value with respect to a reference distance [31].

2.1.4 BS Signal Coverage Model

The coverage of a BS, denoted by $\mathcal{C}_m$, shall satisfy the received power requirements for both downlink and uplink direction as follows.

$$\mathbb{P}\left(P_r^d(x,y) < P_{min}^d\right) \leq p_{out}^d, \forall (x,y) \in \mathcal{C}_m, \tag{2.3}$$

$$\mathbb{P}\left(P_r^u(x,y) < P_{min}^u\right) \leq p_{out}^u, \forall (x,y) \in \mathcal{C}_m, \tag{2.4}$$

where $\mathbb{P}(\cdot)$ represents the probability function, $P_r^d(x,y)$ denotes the received power of an MS at location (x,y) from B_m, P_{min}^d is the minimum required received power of an MS determined by receiver sensitivity [33], and p_{out}^d is the outage probability of the received signal power of an MS. On the other hand, $P_r^u(x,y)$ denotes the received power of B_m from an MS at location (x,y), P_{min}^u is the minimum required received power of a BS, and p_{out}^u is the outage probability of the received signal power of a BS.

Note that the area of $\mathcal{C}_m$ is denoted by $A_{\mathcal{C}_m}$. For example, P^d_{min} can be computed by

$$P^d_{min}(\text{dBm}) = N_0 B(\text{dBm}) + \text{SINR(dB)} + \text{NF(dB)} + \text{IM(dB)}, \quad (2.5)$$

where N_0 is the thermal noise power density, B is the specified noise bandwidth, SINR is the requirement on signal to interference plus noise ratio for a specific coding and modulation scheme, NF is a measure of loss of SINR caused by RF components, IM is the implementation margin to account the SINR differences between practice and theory. To handle the potential interference caused by the switching-on/off operations, we set an interference margin to mitigate the interference from neighboring cells in the SINR requirement [31,33,34]. In LTE, the thermal noise density is specified as -174dBm/Hz, NF is specified as 9dB, SINR and IM depends on specific modulation and coding schemes [33].

2.1.5 Green Network Coverage Problem

The design objective of green network coverage problem is to find the set of active BSs such that the total power consumption of all BSs in the network is minimized while the network-wide coverage is still maintained. Without loss of generality, we assume that each BS is either in an ON (i.e., active) state or in an OFF (i.e., power-off) state during which there is zero power consumption. Denote the state of B_m by indicator variable I_m (" $I_m = 1$" means ON, and "$I_m = 0$" means OFF). To switch off low-load BSs, both downlink coverage and uplink coverage must be ensured by other neighboring active BSs. Also, the network traffic load requirements shall still be satisfied after the BS switching-off processes. An optimization problem for green network coverage problem can be formulated as:

$$\min \sum_{m=1}^{M} I_m \left(\eta_m^{-1} P_{t,m} + P_{c,m}\right) \quad (2.6)$$

$$\text{subject to } (2.7) - (2.15).$$

We specifically explain the requirements and constraints from (2.7) to (2.15) as follows.

$$\mathcal{A} \subseteq \bigcup_{B_m \in \mathcal{B}} \mathcal{C}_m. \tag{2.7}$$

In (2.7), $\mathcal{C}_m$ is the region covered by B_m, and $\mathcal{A}$ is the target coverage region. With the two requirements, an MS must be covered by at least one active BS in the target coverage area. The coverage of a BS shall satisfy the minimum receive signal requirements below.

$$\mathbb{P}(P_r^d(x,y)I_m < P_{min}^d) \leq p_{out}^d, \forall (x,y) \in \mathcal{C}_m, \forall m \in \{1,\ldots,M\}, \tag{2.8}$$

$$\mathbb{P}(P_r^u(x,y)I_m < P_{min}^u) \leq p_{out}^u, \forall (x,y) \in \mathcal{C}_m, \forall m \in \{1,\ldots,M\}. \tag{2.9}$$

In (2.8) and (2.9), the received power requirement for the downlink and uplink coverage of B_m at a point (x,y) is given, respectively. Note that no location will be covered by B_m (i.e., $\mathcal{C}_m = \emptyset$) if B_m is at off state (i.e., $I_m = 0$). To satisfy the downlink traffic requirement in the coverage of $\mathcal{C}_m$, the bandwidth constraint and the transmit power constraint are given in (2.10) and (2.11), respectively.

$$\sum_{\forall (x_k,y_k) \in \mathcal{C}_m} w_{k,m}^d \leq (1-\delta_m^d) W_m^d, \forall m \in \{1,\ldots,M\}, \tag{2.10}$$

$$\sum_{\forall (x_k,y_k) \in \mathcal{C}_m} P_{t,m,k} \leq (1-\delta_m^p) P_{t,m}, \forall m \in \{1,\ldots,M\}. \tag{2.11}$$

In the two equations, W_m^d is the total available bandwidth of B_m for downlink, $w_{k,m}^d$ is the required bandwidth of B_m for MS k, $P_{t,m,k}$ is the required transmit power of B_m for MS k, and (x_k, y_k) is the position of MS k. Note that δ_m^d and δ_m^u are the fraction of downlink and uplink bandwidth which is not used for data transmission (e.g., control signalling, pilots). Similarly, δ_m^p is the fraction of power not used for

data transmission.

$$\sum_{\forall(x_k,y_k)\in\mathcal{C}_m} w_{k,m}^u \leq (1-\delta_m^u)W_m^u, \forall m \in \{1,\ldots,M\}. \quad (2.12)$$

Similarly, to satisfy the uplink traffic requirements, the aggregated bandwidth allocated for uplink MS k by B_m, denoted as $w_{k,m}^u$, is constrained by the total available uplink bandwidth W_m^u in (2.12). Besides, the maximum transmit power constraints and the BS operation state constraints are given as follows.

$$0 \leq P_{t,m} \leq I_m P_{t,m}^{max}, \forall m \in \{1,\ldots,M\}, \quad (2.13)$$

$$0 \leq P_{t,k} \leq P_{t,max}, \forall k \in \{1,\ldots,K\}, \quad (2.14)$$

$$I_m \in \{0,1\}, \forall m \in \{1,\ldots,M\}. \quad (2.15)$$

In (2.13), $P_{t,m}^{max}$ is the constraint on the maximum transmit power of B_m due to the radiation power regulations or hardware limits. Note that the maximal transmit power of B_m is equal to zero if B_m is switched off (i.e., $I_m = 0$). In (2.14), $P_{t,max}$ stands for the maximum transmit power constraint of MS k.

Depending on underlying communication systems and network environments, BS coverage may be uplink dominated (e.g., WCDMA networks [33, 35]) or downlink dominated (e.g., small cells in HetNets [23]). In this chapter, we investigate the uplink coverage-limited problem and downlink coverage-limited problem, respectively. Finally, we consider a joint uplink and downlink network coverage preservation problem in HetNets in which cells can either uplink-dominated or downlink-dominated. To further explore the opportunities for network energy saving, we propose "reception mode" design of BSs and then discuss the potential energy-efficient network architectures in HetNets during the low traffic load period.

2.2 Minimum Power BS Activation Problem for Uplink Coverage-Limited Networks

We first investigate the green network coverage problem in uplink direction. As we know that BS coverage is usually dominated by the downlink traffic load during high load periods. When network traffic load is reduced, most of BS coverage will be uplink limited by the maximum transmit power of MSs. Here we desire to save network energy by powering-off redundant BSs. Here we assume that the maximum transmit power of MSs are identical and BS coverage is restricted by the uplink direction. Then, an optimization for uplink network coverage preservation can be formulated as:

$$\min \sum_{m=1}^{M} I_m P_{c,m}, \tag{2.16}$$

subject to

$$\mathcal{A} \subseteq \bigcup_{m=1}^{M} \mathcal{C}_m, \tag{2.17}$$

$$\mathbb{P}(P_r^u(x, y, P_{t,k}) I_m < P_{min}^u) \leq p_{out}^u, \forall (x, y) \in \mathcal{C}_m, \forall m \in \{1, \ldots, M\}, \tag{2.18}$$

$$0 < P_{t,k} \leq P_{t,max}, \forall k \{1, \ldots, K\}, \tag{2.19}$$

$$I_m = \{0, 1\}, \forall m \in \{1, \ldots, M\}. \tag{2.20}$$

In the problem, an MS shall satisfy the uplink received signal requirement and the total bandwidth constraint for uplink in a BS must be satisfied. Specifically, $\mathcal{C}_m$ is the region covered by B_m, and $\mathcal{A}$ is the target coverage region. In (2.18), the received power requirement for uplink coverage of B_m at a point (x, y) is given. In (2.19), $P_{t,k}$ is the transmit power of MS k and $P_{t,max}$ stands for the maximum transmit power constraint of MS k. Here we want to determine the BS activation set for minimizing

network power consumption while avoiding coverage holes in the network. Note that uplink coverage can be extended by TTI (Transmission Time Interval) bundling but only approximately 1 dB gain on SINR (e.g., [36]) can be obtained. Here we focus on the gain obtained by switching off BSs and controlling power of BSs in low traffic conditions. Other potential techniques (e.g., [36]) for uplink coverage extension can be jointly used with our approach.

Let R_m denote as the maximum coverage range of B_m. To satisfy the received signal power requirement at BS side in (2.18) for B_m at active state (i.e., $I_m = 1$), the MS transmit power shall follow the inequality as follows:

$$P_{t,k} \geq K^{-1} P^u_{min} R^{\alpha}_m 10^{\frac{\sigma_\Psi}{10} Q^{-1}(p^u_{out})}, \tag{2.21}$$

where $Q^{-1}(\cdot)$ represents the inverse Q-function. Given the maximum transmit power of an MS, the uplink BS coverage for $I_m = 1$ can be upper bounded by

$$R_m \leq \left(\frac{P_{t,max} K}{P^u_{min} 10^{\frac{\sigma_\Psi}{10} Q^{-1}(p^u_{out})}} \right)^{\frac{1}{\alpha}}. \tag{2.22}$$

The maximum uplink coverage range for satisfying the received signal requirement is given above. Given resource allocation strategies in BSs, we can obtain the maximum uplink coverage of a BS. On the other hand, the coverage of B_m is $\mathcal{C}_m$ is an empty set if $I_m = 0$.

Then, we attempt to find an optimal BS set which minimizes the total power consumption while avoiding coverage holes in the network with an arbitrary deployment of BSs from the obtained maximal coverages. Here we first prove that the minimum power BS activation problem is NP-hard and then propose a heuristic algorithm called Cell Overlap Minimization with Intersection Covered (COMIC) to determine the set of active BSs. Later in Section 2.2.4, we will show via simulations that our proposed COMIC algorithm in this section can approach the performance upper bound derived in this section whenever the network size is much larger than the coverage of a BS.

2.2.1 Complexity Analysis of Minimum Power BS Activation Problem

The minimum power BS activation problem for minimizing total power by activating a set of BSs while preserving coverage can be closely related to the *minimum-weight disk cover* problem in which the goal is to minimize the total cost to cover a set of points in the plane by selecting a subset of disks [37–39]. It is known to be NP-hard [39]. Recently, a polynomial time approximation scheme (PTAS) is proposed in [37] for a minimum-weight disk cover problem. However, the authors in [38] show that PTAS for set cover does not exist (unless $P = NP$) for circles of roughly the same size. To cover a set of points by unit disks, an algorithm with constant approximation ratio is proposed for minimizing the summation of the weight of disks in [39]. Since the design objective is to cover a set of points instead of covering the entire region, the authors in [39] indicate that one hole or many holes will be present with the proposed algorithm. From the above, we can find that the results in the minimum-weight disk cover problem cannot be directly applicable in our green network coverage problem in which coverage holes shall be prevented.

We first examine the complexity of the problem and derive key properties for preserving network coverage which inspire us to design the BS activation algorithm. To analyze the complexity of this problem, we look at a network in which 1) the entire network can be covered when all BSs are activated, 2) the entire network is a continuous region within which an MS can move from any position to another position, and 3) no BSs will be placed at the same position.

Given a target covering region $\mathcal{A}$ and a set of BSs $\mathcal{B}$, we have the following definitions for the coverage problem.

Definition 2.1. *A point $i \in \mathcal{A}$ is said to be covered by an active BS B_m if it is within the coverage range of B_m, $D(i, B_m) < R_m$ in which $D(\cdot, \cdot)$ is the Euclidean distance between two positions. Otherwise, the point i is said to be not covered by B_m.*

Definition 2.2. *A boundary intersection is a crossing point between the boundary of two sets. The set can be the coverage of a BS or the target coverage region. Formally,*

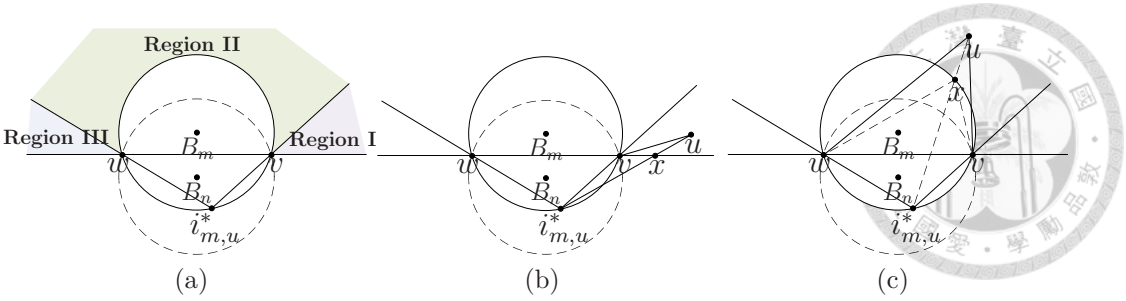


Figure 2-2: The relative positions between an uncovered point and the active BS closest to the uncovered point. (a)The regional classification of an uncovered point. (b)The position of an uncovered point in region I. (c)The position of an uncovered point in the region II.

the set of all boundary intersections in the target area $\mathcal{A}$, denoted by $\mathcal{I}_{\mathcal{A}}$, can be represented as

$$\mathcal{I}_{\mathcal{A}} \equiv \{i \in \mathcal{I}_{\mathcal{A}} | \bigcup_{B_m, B_n \in \mathcal{B}, B_m \neq B_n} \partial\mathcal{C}_m \cap \partial\mathcal{C}_n \bigcup_{B_m \in \mathcal{B}} \partial\mathcal{C}_m \cap \partial\mathcal{A}\},$$

where the boundary of coverage $\mathcal{C}_m$ is denoted by $\partial\mathcal{C}_m$ and the boundary of the target region is denoted by $\partial\mathcal{A}$.

We first derive the sufficient conditions for an activated BS set, denoted by $\mathcal{B}_{on}$, to maintain network coverage and then show the minimum BS activation problem is NP-hard.

Lemma 2.3. *If all boundary intersections of an activated BS set $\mathcal{B}_{on}$ in the target area are covered by $\mathcal{B}_{on}$ itself, all positions in the target region will be covered.*

Proof. We prove this lemma by contradiction. Specifically, we will prove that if there exists a point in the network not covered by the set of active BSs, there must exist a boundary intersection which is not covered by the set of active BSs. Suppose that point $u \in \mathcal{A}$ in the network is not covered by any active BSs. That is, $\{u \in \mathcal{A} | u \notin \bigcup_{B_n \in \mathcal{B}_{on}} \mathcal{C}_n\}$. If $B_m \in \mathcal{B}_{on}$ is the activated BS closest to the uncovered (not covered) point $u \in \mathcal{A}$, we will prove that the boundary intersection closest to u at the boundary of $\mathcal{C}_m$ will not be covered by any active BSs. Denote $i^*_{m,u} \in \mathcal{A}$ as the boundary intersection between the coverage of B_m and the coverage of other BS in $\mathcal{B}_{on}$ closest to u. We will prove that $i^*_{m,u}$ will not be covered by any BSs in $\mathcal{B}_{on}$ if u

is an uncovered point in the network. It will be a contradiction to our assumption.

Suppose that the boundary intersection $i^*_{m,u}$ is covered by an active BS $B_n \in \mathcal{B}_{on}$. Since B_n covers $i^*_{m,u}$, there will be two intersections between boundary intersection $\partial\mathcal{C}_m$ and $\partial\mathcal{C}_n$. See Figure 2-2 (a) as an example. In the figure, v and w are the intersections between $\partial\mathcal{C}_m$ and $\partial\mathcal{C}_n$. Because B_m is the active BS closest to the uncovered point, the uncovered point will be at the upper side of $\overline{vw}$. Otherwise, B_n will be closer to the uncovered point than B_m. Here we classify the possible positions of the uncovered point u into three regions and prove that if $i^*_{m,u}$ is covered by $\mathcal{C}_n$, we will prove that $i^*_{m,u}$ will not be the intersection on the boundary of $\mathcal{C}_m$ closest to u.

We first consider the uncovered position u located in Region I, see Figure 2-2 (b) for example. In this region, because $\angle wi^*_{m,u}v$ is an inscribed angle centred at B_m outside chord $\overline{vw}$, $\angle wi^*_{m,u}v > \frac{\pi}{2}$ and $\angle xvi^*_{m,u} = \angle vwi^*_{m,u} + \angle vi^*_{m,u}w$ will be larger than $\frac{\pi}{2}$. Moreover, because $D(u, B_n) \geq D(u, B_m)$, $\angle uvx \geq 0$. The equality holds when u is at $\overrightarrow{wv}$. Thus, $\angle uvi^*_{m,u} = \angle xvi^*_{m,u} + \angle uvx > \frac{\pi}{2}$. By law of sines in $\triangle uvi^*_{m,u}$, we can obtain $\overline{ui^*_{m,u}} > \overline{uv}$. Similarly, if $i^*_{m,u}$ is in Region III, we can prove $\overline{ui^*_{m,u}} > \overline{uw}$. On the other hand, if the uncovered position u is at $\overline{i^*_{m,u}v}$ or $\overline{i^*_{m,u}w}$, $\overline{ui^*_{m,u}} \geq \overline{uv}$, because u is either outside $\partial\mathcal{C}_m$ or on $\partial\mathcal{C}_m$. A contradiction to $i^*_{m,u}$ is a boundary intersection on $\partial\mathcal{C}_m$ closest to u.

Finally, consider that an uncovered point u in Region II, see Figure 2-2 (c) as an example. In the figure, because $\angle xvi^*_{m,u}$ and $\angle xwi^*_{m,u}$ are two opposite angles of a cyclic quadrilateral on $\partial\mathcal{C}_m$, $\angle xvi^*_{m,u} + \angle xwi^*_{m,u} = \pi$. Because u is on $\overrightarrow{i^*_{m,u}x}$ either outside the circle or on the circle, x is an interior point either in $\triangle uwv$ or x is at the same position of u such that $\angle uvi^*_{m,u} + \angle uwi^*_{m,u} = \angle xvi^*_{m,u} + \angle xwi^*_{m,u} + \angle uwx + \angle uvx \geq \pi$. Equality holds when u is on the circle. Thus, we can find that at least $\angle uvi^*_{m,u}$ or $\angle uwi^*_{m,u}$ will be larger than $\frac{\pi}{2}$. By law of sines, at least $\overline{ui^*_{m,u}} \geq \overline{uv}$ or $\overline{ui^*_{m,u}} \geq \overline{uw}$. As a result, $i^*_{m,u}$ is not the boundary intersection on B_m closest to u. This leads to a contradiction.

From the above, we have proven that if there exists a point in the network not covered by the set of active BSs, there must exist a boundary intersection between two active BSs which is not covered by the set of active BSs. It is a contradiction to

our assumption.

□

Lemma 2.3 can be easily realized by observations as follows. If a boundary intersection between two active BSs can be covered by another active BS, there will be no coverage hole in the interior triangle region of the three active BSs. By recursively examining the intersections between two active BSs, we can know whether coverage holes exist among the active cells.

Lemma 2.4. *If all boundary intersections in the target region $\mathcal{I}_{\mathcal{A}}$ are covered by an activated BS set $\mathcal{B}_{on}$, the entire target region $\mathcal{A}$ is covered by $\mathcal{B}_{on}$.*

Proof. In Lemma 2.3, we prove that the sufficient condition to cover the network is that all boundary intersections of the activated BSs in the target region are covered. The set of all boundary intersections of BSs in the target region is a super set of the set of boundary intersections of activated BSs in the target region. If the super set is covered, we can obtain the network will be covered by the activated BS set $\mathcal{B}_{on}$ from Lemma 2.3.

□

From Lemma 2.4, if the target coverage region contains boundary intersections which are not covered by active BSs, the target coverage region is not covered. One can easily examine whether coverage holes exist in the target coverage region by checking only the boundary intersections of all BSs instead of checking every position in the target region.

Theorem 2.5. *To find the minimum activated BS set to cover the network is NP-hard.*

Proof. From Lemma 2.4, we can reduce the minimum BS activation problem from a continuous region covering problem to a discrete point covering problem. By this transformation, this problem is equivalent to the *minimum disk cover* problem [37,40]. Specifically, the minimum disk cover problem asks to cover a set of points $\mathcal{P}$ in the plane with a subset of disks $\mathcal{D}$ with minimum cardinality. This minimum disk cover

problem is known to be NP-hard by a reduction from the planar 3SAT problem [41]. In our green cellular coverage problem, the set of points $\mathcal{P}$ to cover is all boundary intersections in the target area region $\mathcal{I}_{\mathcal{A}}$. We try to find a subset of BS coverages $\mathcal{D} = \{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$ to cover $\mathcal{I}_{\mathcal{A}}$ with minimum cardinality. Henceforth, the minimum BS activation problem corresponds to the minimum disk cover problem and is thus an NP-hard problem. □

2.2.2 Cell Activation Algorithm– Cell Overlap Minimization with Intersection Covered (COMIC)

Here our goal is to activate a specific set of BSs which can minimize the network power consumption with network-wide coverage and can also support the network traffic loads. Based on the current standards, traffic volume information is usually collected in a group unit (e.g., RNC in UMTS). Here we propose a centralized algorithm which may be more fitting with the current standardization directions. Decentralized mechanisms may need the development of extra signalling mechanisms and may easily cause ping-pong effects (a BS is switching on and off repeatedly) [42] with only local information.

We propose a load-aware algorithm called Cell Overlap Minimization with Intersection Covered (COMIC) for activating an energy-efficient BS set while maintaining network coverage. With the obtained optimal coverages for minimizing area power consumption, we try to find an optimal set which minimizes the overlaps while avoiding coverage holes in the target area. Here our solution is designed for the network in which the location of BSs is available. As shown in [20], the optimal position of a BS to minimize the overlap of the coverage of two other activated BSs shall be located such that the distances from the joint boundary intersection point to the other two boundary intersections are exactly the same. Specifically, if $i \in \partial\mathcal{C}_m \cap \partial\mathcal{C}_n$, the optimal coverage to cover i (denoted by $\mathcal{C}_i^*$) shall satisfy 1) $i \in \partial\mathcal{C}_i^*$ and 2) $D(i,j) = D(i,k)$ for $j \in \partial\mathcal{C}_m \cap \partial\mathcal{C}_i^*, j \neq i$ and $k \in \partial\mathcal{C}_n \cap \partial\mathcal{C}_i^*, k \neq i$.

The steps of the COMIC algorithm are as follows.

1. Choose a BS which has the maximal coverage range from $\mathcal{B}$ (say B_1) as the initial active BS, and add it as the first element of $\mathcal{B}_{on}$ (i.e., $\mathcal{B}_{on} = \{B_1\}$).

2. Find B_k which has the maximal coverage range from $\mathcal{B} \setminus \{B_1\}$ and $\mathcal{C}_k \cap \mathcal{C}_1 \neq \emptyset$. Set B_k as the new active BS. Add the new active BS B_k to $\mathcal{B}_{on}$. Denote the set of uncovered boundary intersections between activated BSs as $\mathcal{I}$ and initiate $\mathcal{I} = \emptyset$.

3. For each $B_m \in \mathcal{B}_{on}$, find boundary intersection i between B_k and B_m which is not covered by any BS in the set of $\mathcal{B}_{on} \setminus \{B_m\}$. Add boundary intersection i to $\mathcal{I}$. Repeat until no such intersections exist for B_m, i.e.,

$$\mathcal{I} = \mathcal{I} \bigcup \{i | i \in \bigcap_{B_j \in \{\mathcal{B}_{on} \setminus \{B_m\}\}} \mathcal{C}'_j \bigcap \partial\mathcal{C}_m \bigcap \partial\mathcal{C}_k \bigcap \mathcal{A}\},$$

   where $\mathcal{C}'_j$ represents the complement of $\mathcal{C}_j$, $\partial\mathcal{C}_m$ is the boundary of $\mathcal{C}_m$, and $\mathcal{A}$ is the target covering region.

4. For each new $i \in \mathcal{I}$, find the optimal BS for i. The optimal BS for i is the BS which is closest to the optimal BS position of i and can cover the boundary intersection i. Set these BSs as candidate BSs.

5. Choose B_k which minimizes the distance from its position to the optimal position of the corresponding boundary intersection from the candidate BSs as the new active BS. Add the new active BS B_k to $\mathcal{B}_{on}$.

6. For each $i \in \mathcal{I}$, if i is covered by B_k, remove it from $\mathcal{I}$.

7. Go to Step 3 until $\mathcal{I} = \emptyset$.

8. Find uncovered boundary intersections at network boundaries, i.e.,

$$\mathcal{I}_{\partial\mathcal{A}} = \bigcup_{B_m \in \mathcal{B}_{on}} \partial\mathcal{C}_m \cap \mathcal{C}'_{on} \cap \partial\mathcal{A}, \tag{2.23}$$

   where $\mathcal{I}_{\partial\mathcal{A}}$ is the set of uncovered boundary intersections between active cells and the target area $\mathcal{A}$, $\mathcal{C}_{on} = \bigcup_{B_m \in \mathcal{B}_{on}} \mathcal{C}_m$, and $\mathcal{C}'_{on}$ is the complement of $\mathcal{C}_{on}$.

9. If $\mathcal{I}_{\partial\mathcal{A}} \neq \emptyset$, choose B_k such that $\mathcal{C}_k$ minimize the overlap with $\mathcal{C}_{on}$ and can cover a boundary intersection $j \in \mathcal{I}_{\partial\mathcal{A}}$. For each $j \in \mathcal{I}_{\partial\mathcal{A}}$, if j is covered by B_k, remove it from $\mathcal{I}_{\partial\mathcal{A}}$. Add the new active BS B_k to $\mathcal{B}_{on}$.

10. Go to Step 3 until $I_{\partial\mathcal{A}} = \emptyset$.

The rationale behind COMIC is to activate a BS which can cover a boundary intersection of any two active BSs with minimum overlap until no uncovered boundary intersection is detected in the target region. Other BSs which are not activated by COMIC can be switched off to save power. Practically, more BSs around a hot spot area may need to power on to satisfy throughput requirements. Alternatively, low-power picocells or femtocells can be deployed to support the capacity in hot spot areas.

2.2.3 Complexity analysis for COMIC

Theorem 2.6. *The time complexity of COMIC with M BSs deployed in the network is in $O(M^3)$.*

Proof. The time complexity of COMIC is dominated by finding exterior intersections which is the boundary intersections not covered by any BS in $\mathcal{B}_{on}$ in Step 3 and the search of the optimal BS for all new exterior intersections in Step 4. The complexity for finding exterior intersections depends on the number of activated BSs. Denote the number of active BSs in $\mathcal{B}_{on}$ by $|\mathcal{B}_{on}|$. Assume n BSs are already activated in the network, $|\mathcal{B}_{on}| = n$. The time complexity to determine whether a new boundary intersection between $B_m \in \mathcal{B}_{on}$ and a newly activated BS B_k belongs to an exterior intersection is $O(n)$ and at most $2n$ intersections need to be examined. On the other hand, to find the optimal BS for an exterior intersection, all BSs which is not activated shall be examined and its time complexity is $O(M - n - 1)$. In Step 3, the number of activated BSs is $|\mathcal{B}_{on}| \leq M - 1$. As a result, the complexity of COMIC is

$$\sum_{n=1}^{M-1} \Big(2n \cdot O(n) + 2n \cdot O(M - n - 1) \Big) = O(M^3). \tag{2.24}$$

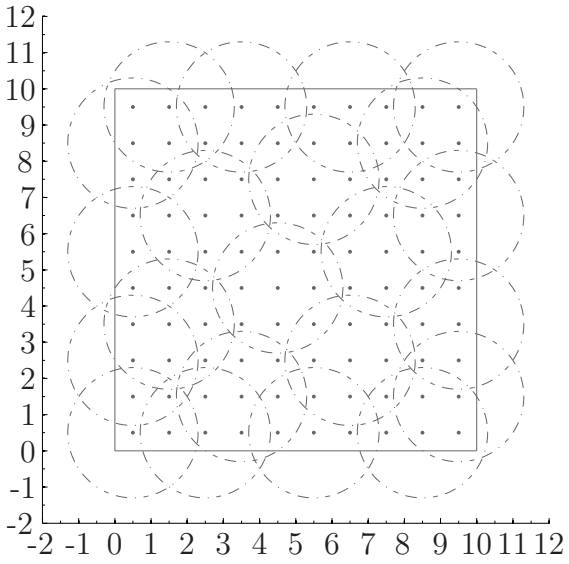


Figure 2-3: The coverage of activated BSs by COMIC to cover a 10-by-10 planar network. BSs are deployed in a grid structure from position (0.5,0.5) to (9.5,9.5). The coverage range $R = 1.8$. Only 22 out of 100 BSs are activated by COMIC to cover the network.

□

2.2.4 Performance Evaluation for Network Coverage Preservation

We examine our proposed algorithm for maintaining network coverage. In the simulation, 100 BSs are deployed regularly in a 10 by 10 planar network. Figure 2-3 shows the coverage of the selected active BSs, each with a coverage range $R = 1.8$. With our proposed algorithm, 22 out of 100 BSs are selected to be active. It can be observed that the network can be covered by the selected BSs except the certain area around the network corners. To ensure network-wide coverage, our algorithm will activate the BSs at the uncovered network boundaries. Only a few BSs near the uncovered network corners shall be activated. For example, the BSs at (9.5,9.5) in Figure 2-3 will be activated to ensure that the entire network area is covered in the proposed algorithm.

We next examine the proposed algorithm for various BS deployment strategies. In the simulation, the BSs are deployed in a 30-by-30 plane. Specifically, to ensure

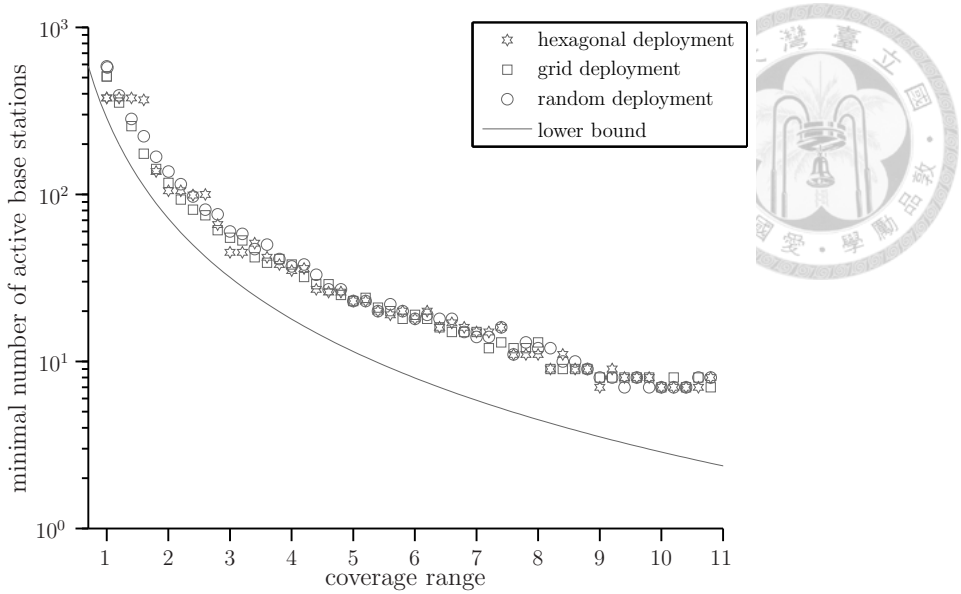


Figure 2-4: The minimal number of active BSs as a function of coverage range to cover a 30-by-30 planar network. Three BS deployment strategies are examined (hexagonal, grid, and random).

that the network can be covered by the deployed BSs each with a unit disk coverage, we consider the following three deployment strategies: 368 BSs are deployed in a hexagonal topology, 900 BSs are deployed in a grid structure, and 3000 BSs are deployed randomly in the network.

Figure 2-4 shows the minimal number of BSs in the network which are selected by our algorithm as a function of the coverage range. As can be seen, the number of selected active BSs can approach the lower bound of the minimal number of active BSs as long as the network size is much larger than the cell coverage (e.g. 10 active BSs in the network). Moreover, it shows that this algorithm performs well in various network deployment strategies. The total number of selected active BSs is only slightly affected by the BS deployment strategies and the number of deployed BSs.

2.3 Green Network Coverage Problem for Downlink Coverage-Limited Networks

We now consider a network in which coverage of a BS is limited by the requirements on the downlink direction. The objective is to minimize the aggregation of active BS power consumption while avoiding coverage holes in the network. Mathematically,

$$\min \sum_{m=1}^{M} I_m \left(\eta_m^{-1} P_{t,m} + P_{c,m}\right) \tag{2.25}$$

subject to

$$\mathcal{A} \subseteq \bigcup_{B_m \in \mathcal{B}} \mathcal{C}_m, \tag{2.26}$$

$$\mathbb{P}(P_r^d(x,y) < P_{min}^d) \leq p_{out}^d, \forall (x,y) \in \mathcal{C}_m, \forall m \in \{1,\ldots,M\}, \tag{2.27}$$

$$\sum_{\forall (x_k,y_k) \in \mathcal{C}_m} w_{k,m}^d \leq (1-\delta_m^d) W_m^d, \forall m \in \{1,\ldots,M\}, \tag{2.28}$$

$$\sum_{\forall (x_k,y_k) \in \mathcal{C}_m} P_{t,m,k} \leq (1-\delta_m^p) P_{t,m}, \forall m \in \{1,\ldots,M\}, \tag{2.29}$$

$$0 \leq P_{t,m} \leq I_m P_{t,m}^{max}, \forall m \in \{1,\ldots,M\}, \tag{2.30}$$

$$I_m \in \{0,1\}, \forall m \in \{1,\ldots,M\}. \tag{2.31}$$

In (2.26), $\mathcal{C}_m$ is the region covered by B_m, and $\mathcal{A}$ is the target coverage region. In (2.8), the received power requirement for the downlink and uplink coverage of B_m at a point (x,y) is given, respectively. To satisfy the downlink traffic requirement in the coverage of $\mathcal{C}_m$, the bandwidth constraint and the transmit power constraint are given in (2.28) and (2.29), respectively. In the two equations, W_m^d is the total available bandwidth of B_m for downlink, $w_{k,m}^d$ is the required bandwidth of B_m for MS k, $P_{t,m,k}$ is the required transmit power of B_m for MS k, and (x_k, y_k) is the position

of MS k. Besides, δ_m^d and δ_m^u are the fraction of downlink and uplink bandwidth which is not used for data transmission (e.g., control signalling, pilots). Similarly, δ_m^p is the fraction of power not used for data transmission. In (2.30), $P_{t,m}^{max}$ is the constraint on the maximum value of BS k transmit power due to the radiation power regulations or hardware limits. Note that the optimization problem is expected to be performed during low traffic load period with the available traffic volume information from daily traffic profiles or the information collected by central control unit (e.g., RNC in UMTS). According to the measurement results in operational BSs (e.g., [11]), traffic load is highly correlated in neighboring cells during specific time duration. The switching-off cell can often distribute its traffic load to the neighboring cells without violating their resource constraint. Practically, extended cell coverage may still be restricted by the traffic load at hotspots during low traffic load period. We will deal with the potentially capacity-limited cells in our proposed algorithm in Section 2.3.3. In this optimization problem, the target region $\mathcal{A}$ is the union of regions under low traffic load conditions such that SLAs of MSs can still be satisfied after BS power level adjustment. According to the network traffic conditions, the target region can be set adaptively for the optimization of green cellular networks. Without loss of generality, we set the target coverage region as the entire network area in the chapter.

Unfortunately, we cannot efficiently find the optimal settings for the mixed-integer optimization problem. From (2.25), for any given BS power settings $\{P_1, \ldots, P_M\}$ with which the target area can be covered (i.e., $\mathcal{A} \subseteq \bigcup_{B_m \in \mathcal{B}} \mathcal{C}_m$ for the given $\{P_1, \ldots, P_M\}$), to minimize the network power consumption is equivalent to finding the minimum-power set of active BSs. Thus, this problem can be further decomposed and solved as two sub-problems. (1) optimal BS operational power problem and (2) minimum-power BS activation problem. In the optimal BS operational power problem, we derive the optimal P_m and the corresponding optimal cell size (in Section 2.3.1). In the minimum-power BS activation problem, we first show the NP-hardness of the problem and determine the activation stet for minimizing aggregated BS power so that the network-wide coverage is still maintained (in Section 2.3.3).

With our algorithm, we could solve this green network coverage problem, as il-

lustrated in Figure 2-1. To cover the target area indicated in the shaded region, one can either activate the BS at the center of the network with a high power to ensure network-wide coverage (Fig. 1 (b)) or activate the other five BSs each with a low power level (Fig. 1 (c)). We want to find which combination of BS power consumption and active BS set is optimal in terms of minimizing the total network power consumption while satisfying the network coverage requirement. Furthermore, to quantify possible performance loss by decomposing the optimization problem into two sub-problems, we derive a lower bound for the minimal network power consumption with network coverage preserved in Section 2.3.2. We will compare our proposed algorithm and the derived lower bound in Section 2.3.4.

2.3.1 BS Operational Power Optimization Problem

Here our goal is to find the optimal operation power and the optimal cell size of an active BS for minimizing the total power consumption of the network. Specifically, to maintain the network coverage in HetNets in an energy-efficient manner, we derive the optimal cell size for minimizing area power consumption for BSs. To quantify the performance loss from the original mixed-integer optimization problem, we also derive lower bounds for the minimal network power consumption which can be applicable for arbitrary BS deployment strategies.

For bursting throughput, high capacity cells (e.g., picocells or femtocells) may be deployed at hot spot areas in the network. When the hardware parameters of BSs are not exactly the same, maintaining BS coverages with different size may require less power than having all activated BSs maintain the same coverage. To evaluate the power efficiency of a BS in covering the network, here we introduce the concept of area power consumption [31] for the optimization problem.

Definition 2.7. *Define area power consumption as the power consumption of an active BS divided by its coverage area. Specifically, the area power consumption of*

B_m with coverage area $A_{\mathcal{C}_m}$ and power consumption P_m is

$$\zeta_m \equiv \frac{P_m}{A_{\mathcal{C}_m}}. \tag{2.32}$$

An optimization problem for area power consumption minimization while preserving network coverage can be represented by

$$\min \zeta, \text{ s.t. } \mathcal{A} \subseteq \bigcup_{m=1}^{M} \mathcal{C}_m, \tag{2.33}$$

$$P_m = I_m(\eta^{-1}P_{t,m} + P_{c,m}), I_m = \{0,1\},$$

$$P_{t,m} = \begin{cases} \eta_m(\zeta A_{\mathcal{C}_m} - P_{c,m}), & \text{if } I_m = 1, 0 \le P_{t,m} \le I_m P_{t,m}^{max}, \\ 0, & \text{otherwise,} \end{cases}$$

$$\mathbb{P}(P_r(x,y) < P_{min}) \le p_{out}, \forall (x,y) \in \mathcal{C}_m, \forall m \in \{1,\ldots,M\}.$$

In this optimization problem, we try to minimize the area power consumption of the network, denoted by ζ, such that each activated BS can efficiently maintain its coverage. Note that the power consumption caused by traffic load is only a small part of BS power consumption (the BS power consumption increase from low load to peak load is less than 10%) [9, 43] and a BS during low traffic load period can often hand off its MSs to neighboring cells which are usually in low traffic load condition [11]. To facilitate our analysis, here we omit the effects of traffic load on area power consumption. The effects of network traffic load will be evaluated in Section 5.

Let R_m denote the maximum range which can be covered by active B_m. To satisfy the coverage requirement in (2.3), the transmit power of B_m for the adopting R_m shall satisfy the following condition:

$$P_{t,m} \ge K^{-1} P_{min} R_m^{\alpha} 10^{\frac{\sigma_\Psi}{10} Q^{-1}(p_{out})}, \tag{2.34}$$

where $Q^{-1}(\cdot)$ represents the inverse Q-function. The area power consumption of B_m

has the relationship as follows:

$$\zeta_m = \frac{P_m}{A_{\mathcal{C}_m}} = \frac{\eta_m P_{t,m} + P_{c,m}}{\pi R_m^2} \geq \frac{\eta_m^{-1} K^{-1} P_{min} R_m^{\alpha} 10^{\frac{\sigma_\Psi}{10} Q^{-1}(p_{out})} + P_{c,m}}{\pi R_m^2}$$
$$\geq \frac{1}{\pi} \left(\frac{P_{c,m}}{\alpha - 2} \right)^{\frac{\alpha-2}{\alpha}} \left(\frac{P_{min} 10^{\frac{\sigma_\Psi}{10} Q^{-1}(p_{out})}}{2\eta_m K} \right)^{\frac{2}{\alpha}}. \quad (2.35)$$

In (2.35), the minimal area power consumption of B_m, denoted by ζ_m^*, can be achieved by the optimal coverage range:

$$R_m^* = \left(\frac{2\eta_m K P_{c,m}}{P_{min} 10^{\frac{\sigma_\Psi}{10} Q^{-1}(p_{out})} (\alpha - 2)} \right)^{\frac{1}{\alpha}}. \quad (2.36)$$

Note that (2.36) is valid only when $\alpha > 2$ and $P_{t,m} \leq P_{t,m}^{max}$. For $\alpha \leq 2$, one shall try to maximize the coverage range of active BSs for minimizing area power consumption. Besides, we can observe that a BS with inappropriate hardware parameters (i.e., a low transmit power efficiency or a large constant operational power) will result in larger minimal area power consumption of a BS. To satisfy a target network area power consumption ζ, a BS with inappropriate hardware parameters may never satisfy the target area power consumption for whatever coverage range adopted. That is, a BS with $\zeta_m^* > \zeta$ will never be activated for minimizing network area power consumption.

To minimize the area power consumption for network coverage preservation, we derive the following lemma for the optimal coverage range of an active BS.

Lemma 2.8. *The optimal coverage range of an active BS for achieving a targeted area power consumption with network coverage preservation is the maximal coverage range that can achieve the targeted area power consumption.*

Proof. Suppose that there exists a set of N active BSs, with coverage range $\{R_1^a, \ldots, R_N^a\}$ and corresponding coverages $\{\mathcal{C}_1^a, \ldots, \mathcal{C}_N^a\}$, which can achieve the targeted area power consumption and also satisfy the network coverage requirement $\mathcal{A} \subseteq \bigcup_{n=1}^{N} \mathcal{C}_n^a$. Assume that $R_{n,max}^a$ is the maximal coverage range, and $\mathcal{C}_{n,max}^a$ the corresponding coverage that the n-th BS can achieve for the targeted area power consumption. Since $\mathcal{C}_n^a \subseteq \mathcal{C}_{n,max}^a$ for $i = 1, \ldots, N$, we can have $\mathcal{A} \subseteq \bigcup_{n=1}^{N} \mathcal{C}_{n,max}^a$. The network coverage

can be maintained by maximizing BS coverage ranges for the targeted area power consumption. □

For a target area power consumption ζ, the maximal coverage range which B_m can adopt, denoted by $R_m^\dagger$, is:

$$R_m^\dagger = \max R_m, \text{ s.t. } \frac{\eta_m^{-1} P_{t,m} + P_{c,m}}{\pi R_m^2} = \zeta,$$
$$P_{t,m} \geq K^{-1} P_{min} R_m^\alpha 10^{\frac{\sigma_\Psi}{10} Q^{-1}(p_{out})}, 0 \leq P_{t,m} \leq P_{t,m}^{max}.$$

With the given hardware parameters of B_m, the maximal coverage range $R_m^\dagger$ to achieve area power consumption ζ can be easily obtained. Unfortunately, we cannot obtain closed-form expressions for the minimal area power consumption while maintaining the network coverage. The numerical results for the maximal coverage range will be demonstrated in Section 2.3.4.

2.3.2 Lower Bound on Network Power Consumption

To learn the optimal performance in the network, we consider a network with no overlap between the coverage of any two BSs in this section. The minimal power consumption under this condition can be regarded as the performance upper bound for any deployment strategies of BSs.

Given the hardware parameters of all BSs to cover a target region $\mathcal{A}$ with area A, we can lower bound the network area power consumption ζ in (2.25) by replacing the coverage constraint from $\mathcal{A} \subseteq \bigcup_{m=1}^{M} \mathcal{C}_m$ to $A \leq \sum_{m=1}^{M} A_{\mathcal{C}_m}$. This bound is obtained by relaxing the constraint on the fixed BS locations and assuming that there are no overlaps between the coverages of activated BSs. We will demonstrate the lower bounds on minimal network power consumption by numerical computation in Section 2.3.4.

Furthermore, if a network has the same BS settings (i.e., $\eta_m = \eta$, $P_{c,m} = P_c$, and $P_{t,m} = P_t$ for $m = \{1, \ldots, M\}$), we can derive a lower bound of network power consumption for the original optimization problem in Section 2.3.

With the given BS setting above, the power consumption of the network, denoted by P_N, can be obtained by substituting the power consumption model in (2.1) into (2.25):

$$P_N = \sum_{m=1}^{M} I_m(\eta^{-1}P_t + P_c). \tag{2.37}$$

Given the coverage area of an active BS $A_{\mathcal{C}}$ and the area of target region A, the number of activated BSs can be lower bounded by:

$$\sum_{m=1}^{M} I_m \geq \frac{A}{A_{\mathcal{C}}}. \tag{2.38}$$

In (2.38), the equality holds only in the ideal case with no overlap between the coverage of any two BSs. Under the ideal condition, for a cell with coverage range R, the objective function for minimizing network power consumption while satisfying coverage requirement can then be obtained as follows:

$$\min_{R} \left\{ \frac{A}{\pi R^2} \left(\eta^{-1}K^{-1}P_{min}R^{\alpha}10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out})} + P_c \right) \right\}, \tag{2.39}$$

From (2.39), we observe that the problem will be equivalent to minimize the area power consumption in (2.33) in Section 2.3.1 and we will obtain the same optimal setting as in (2.36). That is, the optimal value of coverage range for the minimizing the network power consumption in the ideal condition is exactly the same as that for minimizing area power consumption. This observation also motivates us to formulate the area power consumption minimization problem.

By substituting (2.36) into (2.39), we can obtain the minimal network power consumption while maintaining the network coverage, denoted by P_N^*, as:

$$P_N^* = \frac{\alpha A}{\pi} \left(\frac{P_c}{\alpha - 2} \right)^{\frac{\alpha-2}{\alpha}} \left(\frac{P_{min}10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out})}}{2\eta K} \right)^{\frac{2}{\alpha}}. \tag{2.40}$$

Specifically, the minimal network power consumption of the entire network is propor-

tional to $P_c^{\frac{\alpha-2}{\alpha}}$ and $\eta^{\frac{-2}{\alpha}}$. Under the optimal setting, the network power consumption will never increase as a linear function of the hardware parameters. Under ideal conditions, the minimal network power consumption in (2.40) can serve as a lower bound on network power consumption for (2.25) with any BS deployment strategy.

2.3.3 Load-Aware COMIC

Due to the fact that the coverage of a BS may be restricted by the underlying traffic loading instead of the received signal strength, here we take the underlying BS traffic load into consideration in the proposed algorithm to ensure that both the coverage requirements and traffic rate requirements can be satisfied. We define *load supporting range* of a BS, denoted by r_m, as the maximum range in which all MSs can be served by the BS under the adopted radio resource allocation mechanism. Specifically, r_m shall satisfy the following resource constraints in B_m: $\sum_{\forall D(k,m)\leq r_m} w_{k,m}^d \leq W_m^d$ and $\sum_{\forall D(k,m)\leq r_m} p_{k,m} \leq P_{t,m}$ where W_m^d is the available bandwidth of B_m, $P_{t,m}$ is the maximal transmit power of B_m, $w_{k,m}^d$ is the required bandwidth of B_m for MS k, and $p_{k,m}$ is the required transmit power of B_m for MS k. Note that $D(k,m)$ stands for the distance between MS k and B_m. Depending on the traffic conditions, the load supporting range r_m may be smaller than the signal coverage range R_m even during the low traffic load period. To satisfy both the coverage requirements and traffic requirements, we set $\mathcal{C}_m$ as the intersection region enclosed by R_m and r_m in the proposed COMIC algorithm. We will demonstrate that the load-aware COMIC can adaptively activate BSs for a network with non-uniformly distributed traffic loads in next section.

2.3.4 Performance Evaluation for Load-Aware COMIC

In this section, we evaluate our proposed algorithm by simulations. We consider various BS placements to show that our algorithm can approach the lower bound of activated BSs in different deployment strategies. We examine non-uniform traffic load distributions to show that our proposed algorithm can adapt to network traffic

Table 2.1: Simulation parameters of COMIC for downlink.

Parameter	**Value**
channel model	path loss+shadowing
shadowing standard deviation σ_Ψ	6 dB
minimum received power P_{min}	-100 dBm
outage probability p_{out}	0.1
path loss exponent α	4
network size	30 Km by 30 Km
K (reference propagation distance = 1m)	10^{-1}
N_0 (noise spectral density)	-174 dBm/Hz
system bandwidth	10 MHz
subcarrier spacing	15 KHz
interference margin	3 dB

load. We also compare our proposed algorithm with another load-aware cell activation algorithm in various traffic load conditions to show the improvement on energy-saving capability. Finally, we will show the effectiveness on power saving by the deployed small cells.

Network Energy Consumption under Non-Uniform Traffic Load Distributions

We examine the performance of COMIC for saving network operational energy. We evaluate the performance of COMIC for non-uniform traffic load distributions to show that COMIC can adaptively activate BS according to network traffic load conditions. We also compare COMIC with another load-aware cell activation algorithm and examine how much network power consumption is due to link-level power consumption.

In the simulation, we consider an OFDMA downlink system based on 3G-LTE [44]. The total bandwidth of a BS is 10 MHz and the subcarrier spacing is 15 KHz. Assume that subcarriers are dynamically allocated among users and power is equally allocated among subcarriers [45]. Users are associated to active BSs with highest SINR [14]. Given the average SINR of an MS-BS link Γ, the average spectral efficiency with path

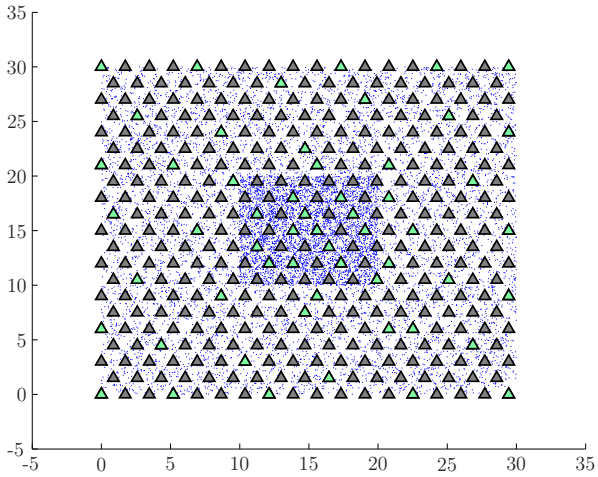


Figure 2-5: The activated BSs (green) by COMIC with a hot spot area at the center in the network. Only 56 out of 368 BSs (gray) are activated. In the network, 5000 users (blue) are uniformly distributed in the hot spot region and 8000 users (blue) are uniformly distributed in the other region.

loss and log-normal shadowing effects can be approximated by [46]:

$$\bar{C} \approx \ln\left(\frac{(1+\Gamma)^2}{\sqrt{1+2\Gamma+e^{\frac{\sigma_\Psi^2}{\xi^2}}\Gamma^2}}\right) \text{nats/s/Hz},$$

where ξ is a constant $\frac{10}{\ln(10)}$. Note that the shadowing standard deviation σ_Ψ is expressed in dB. To mitigate the potential interference from neighboring cells, we set an interference margin for the MS-BS link budget in the simulations. The network simulation parameters are given in Table 2.1.

We first evaluate the performance of COMIC during the low traffic load period. Due to the non-uniform user population, hot spot area may be present even in the low traffic load period. Here we consider a network of 368 BSs grid-deployed in the 30 Km by 30 Km region with a hot spot region of 10 Km by 10 Km at the center. The population density is 50 active users per square kilometer in the hot spot area and 10 active users per square kilometer in the other region [47] and each generates an average data rate 10 Kbps during the low traffic load period. A typical BS power profile $P_c = 1000$ W and $\eta = 0.5$ [29] is given.

The BSs activated by COMIC are shown in Figure 2-5. From the figure, we can

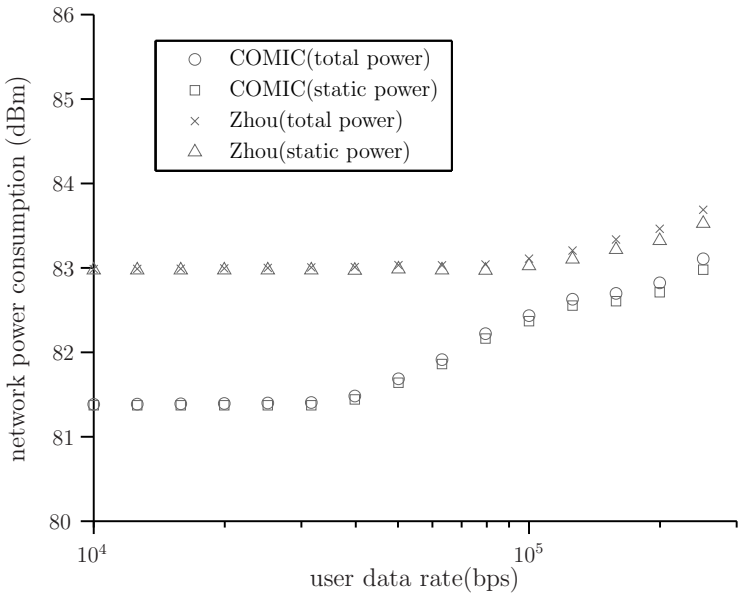


Figure 2-6: The minimal network power consumption as a function of average user data rate. The static part of network power consumption is also shown. The performance is compared with the algorithm by Zhou et al.

observe that the cell size is determined by the traffic load demands around the hot spot area and the cell size is determined by the coverage requirement in other low traffic regions. From the figure, we can observe that COMIC can adaptively activate BSs according to the traffic load conditions and greatly reduce the number of activated BSs in the low traffic regions.

We then compare our algorithm with another load-aware cell activation algorithm [14] when traffic load varies. In the simulations, the BS deployment and active user distribution in Figure 2-5 are considered. The maximum transmit power 46 dBm is given and the performance is averaged over 20 random active user distributions.

The total network power consumption as a function of user traffic rates is shown in Figure 2-6. Also, the static part of network power consumption which is independent of traffic loading is also shown. The performance of COMIC is compared with that of the power saving algorithm in [14] proposed by Zhou et al. In the figure, we can observe that the static power consumption will dominate the network power consumption which verifies our simplifications on BS power consumption. Besides, compared with the power saving algorithm, COMIC can gain more than 1.5 dB

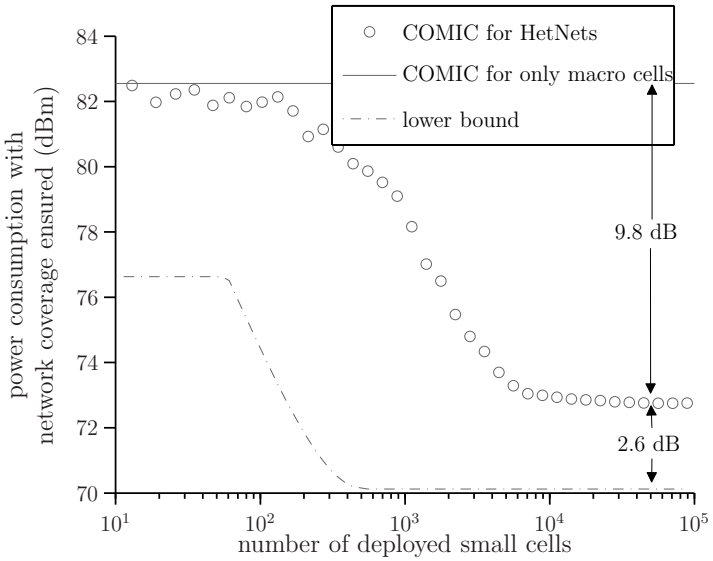


Figure 2-7: The minimal network power consumption for maintaining network coverage as a function of the number of deployed small cells in the network.

performance for low user traffic rates. That is, approximately 44.5% more power will be consumed by the power saving algorithm than COMIC. Even for higher user traffic rates, COMIC can still maintain lower network power consumption than the power saving algorithm. The reason is that the proposed COMIC algorithm is not only load-aware but also topology-aware.

Network Energy Saving by Network densification with Small Cells

Finally, we examine the impacts of the number of small cells in the network. A question of interest is whether the network power consumption can be further reduced with the deployment of small cells during the low traffic load period. In the simulations, 100 macro cells and a certain number of small cells are randomly deployed in the network. To facilitate our simulations, we assume the traffic is low enough for us to omit its effects during the low traffic load period. The network performance is averaged over 20 randomly generated networks. The hardware parameters are: $P_{c,macro} = 1000$ W, $\eta_{macro} = 0.5$, $P_{c,small} = 10$ W, and $\eta_{small} = 0.1$. The other simulation parameters are given in Table 2.1.

We examine the network power consumption as a function of the number of small cells deployed in the network. Figure 2-7 shows the minimal network power con-

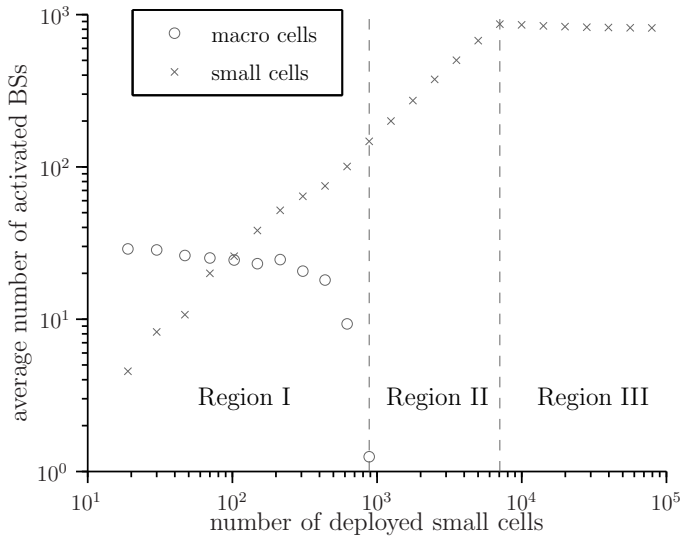


Figure 2-8: The average number of activated BSs for minimizing network power consumption while maintaining network coverage as a function of the number of deployed small cells in the network.

sumption for network coverage preservation as a function of the number of small cells deployed in the network. In the figure, we can observe that the network power consumption can be further reduced if we jointly take the macro cells and small cells into COMIC activation processes even though only a few small cells are deployed in the network. In general, more power can be conserved when more small cells are deployed in the network. From the figure, compared with the performance of only activating macro cells, almost 10 dB performance gain can be achieved if we jointly activate macro cells and small cells. Besides, compared with the lower bound of minimal network power consumption, our scheme only incurs approximately 2.6 dB performance gap.

The number of macro cells and small cells activated in the entire network is shown in Figure 2-8. In the figure, the number of activated BSs can be divided into three regions. In Region I, macro cells and small cells are jointly activated for minimizing the network power consumption of the entire network. In Region II, deploying more small cells can effectively improve the performance of network power consumption. Activating macro cells is no more energy-efficient. In Region III, the number of activated small cells is almost the same such that deploying more small cells cannot

significantly reduce more power consumption. From the observation in the figure, the simple transition behavior can be easily applied by a network operator to evaluate the effectiveness of deploying more small cells on network power consumption reduction. Furthermore, to achieve an energy-efficient green cellular network in the low load period, it can be observed that maintaining network coverage only with macro BSs may not always be the optimal strategy especially when a large number of small cells are deployed in the network.

2.4 Joint Uplink and Downlink Green Network Coverage Problem

Finally, we investigate a general problem that BS coverages can be either downlink-limited or uplink-limited in the network. To explore opportunities for network energy saving, we introduce the "reception mode" operation of a BS in conjunction with the asymmetric connections between BSs and MSs. The asymmetric uplink and downlink between MSs and BSs means that an MS can uplink from one BS and downlink from another BS. Here we suppose that a BS in reception mode will listen the uplink spectrum for potential incoming traffic from MSs and the power amplifier and other circuit for downlink operations are switched off to conserve energy. The uplink signalling for a BS in reception mode will be accomplished by other cells which support downlink operations. Specifically, we define the following four modes of BS operations for asymmetric uplink and downlink: active mode, downlink mode, uplink mode, and sleep mode. In active mode, a BS operates normally for both uplink and downlink direction. In downlink mode, a BS transmits data, control signals, and training symbols without reception functionalities. In uplink mode, a BS only listen to the uplink spectrum to detect possible incoming messages. In sleep mode, a BS switch off all parts of functionalities for radio access. Without loss of generality, we assume that a BS in sleep mode will not consume any energy.

Specifically, we investigate the minimization of network power consumption while

maintaining both uplink network coverage and downlink network coverage by asymmetric uplink and downlink. We assume that a BS will operate in four states: active mode , downlink mode, uplink mode, and sleep mode. A general optimization problem can be formulated as:

$$\min \sum_{m=1}^{M} \left\{ I_m^d \left(\eta_m^{-1} P_{t,m} + P_{c,m}^d \right) + I_m^u P_{c,m}^u + I_m^{on} P_{c,m}^{on} \right\} \tag{2.41}$$

subject to

$$\mathcal{A} \subseteq \bigcup_{m=1}^{M} \mathcal{C}_m^d, \tag{2.42}$$

$$\mathcal{A} \subseteq \bigcup_{m=1}^{M} \mathcal{C}_m^u, \tag{2.43}$$

$$\mathbb{P}(P_r^d(x,y) I_m^d < P_{min}^d) \leq p_{out}^d, \forall (x,y) \in \mathcal{C}_m^d, \forall m \in \{1,\ldots,M\}, \tag{2.44}$$

$$\mathbb{P}(P_r^u(x,y) I_m^u < P_{min}^u) \leq p_{out}^u, \forall (x,y) \in \mathcal{C}_m^u, \forall m \in \{1,\ldots,M\}, \tag{2.45}$$

$$\sum_{\forall k \in C_m^d} w_{k,m}^d \leq (1-\delta_m^d) W_m^d, \forall m \in \{1,\ldots,M\}, \tag{2.46}$$

$$\sum_{\forall k \in C_m^d} p_{k,m}^d \leq (1-\delta_m^p) P_{t,m}, \forall m \in \{1,\ldots,M\}, \tag{2.47}$$

$$\sum_{k \in \mathcal{C}_m^u} w_{m,k}^u \leq (1-\delta_m^u) W_m^u, \forall m \in \{1,\ldots,M\}, \tag{2.48}$$

$$0 \leq P_{t,m} \leq I_m^d P_{t,m}^{max}, \forall m \in \{1,\ldots,M\}, \tag{2.49}$$

$$0 \leq P_{t,k} \leq P_{t,max}, \forall k \in \{1,\ldots,K\}, \tag{2.50}$$

$$I_m^{on} = I_m^u \vee I_m^d, \forall m \in \{1,\ldots,M\}, \tag{2.51}$$

$$I_m^u, I_m^d, I_m^{on} \in \{0,1\}, \forall m \in \{1,\ldots,M\}. \tag{2.52}$$

In (2.41), I_m^d denotes the state of B_m for downlink, I_m^u denotes the state of B_m for uplink, I_m^{on} denotes the state of B_m which is active either for downlink or for uplink, $P_{c,m}^{on}$ stands for the power consumption of B_m for $I_m^{on} = 1$, $P_{c,m}^d$ stands for the additional power consumption of B_m for $I_m^d = 1$ compared with $I_m^{on} = 1$, and $P_{c,m}^u$ stands for the additional power consumption of B_m for $I_m^u = 1$ compared with $I_m^{on} = 1$. In (2.42), $\mathcal{C}_m^d$ is the region covered by B_m in the uplink direction and $\mathcal{A}$ is the target coverage region. Similarly, $\mathcal{C}_m^u$ is the region covered by B_m in the downlink direction in (2.43). In (2.44) and (2.45), the received power requirement for the downlink and uplink coverage of B_m at a point (x, y) is given, respectively. To satisfy the downlink traffic requirement in the coverage of $\mathcal{C}_m$, the bandwidth constraint and the transmit power constraint are given in (2.46) and (2.47), respectively. In the two equations, W_m^d is the total available bandwidth of B_m for downlink, $w_{k,m}^d$ is the required bandwidth of B_m for MS k, $P_{t,m,k}$ is the required transmit power of B_m for MS k, and (x_k, y_k) is the position of MS k. Besides, δ_m^d and δ_m^u are the fraction of downlink and uplink bandwidth which is not used for data transmission (e.g., control signalling, pilots). Similarly, δ_m^p is the fraction of power not used for data transmission. In (2.48), the aggregated bandwidth allocated for uplink MS k by B_m, denoted as $w_{k,m}^u$, is constrained by the total available uplink bandwidth W_m^u. In (2.49), $P_{t,m}^{max}$ is the constraint on the maximum value of BS k transmit power due to the radiation power regulations or hardware limits. In (2.50), $P_{t,max}$ stands for the maximum transmit power constraint of MS k. In (2.51), the relation of BS states is shown.

2.4.1 Lower Bounds of Network Power Consumption

Here we derive lower bounds of minimal network power consumption while avoiding uplink and downlink network coverage holes for both symmetric BS activation and asymmetric BS activation. Suppose that a network has the same BS settings (i.e., $\eta_m = \eta$, $P_{c,m}^u = P_c^u$, $P_{c,m}^d = P_c^d$, $P_{c,m}^{on} = P_c^{on}$, and $P_{t,m} = P_t$ for $m = \{1, \ldots, M\}$), we derive lower bounds of network power consumption. The lower bounds can be applicable for any kind of BS deployment strategies.

The power consumption of the network, denoted by P_N, can be obtained by (2.41):

$$P_N = \sum_{m=1}^{M} \left\{ I_m^d(\eta^{-1}P_t + P_c^d) + I_m^u P_c^u + I_m^{on} P_c^{on} \right\}. \tag{2.53}$$

Given the uplink and downlink coverage area of an active BS $A_{\mathcal{C}^u}$ and $A_{\mathcal{C}^d}$, and the area of target region A, the number of activated BSs can be lower bounded by:

$$\sum_{m=1}^{M} I_m^{on} \geq \max\left\{ \frac{A}{A_{\mathcal{C}^u}}, \frac{A}{A_{\mathcal{C}^d}} \right\}. \tag{2.54}$$

For asymmetric uplink and downlink, if $A_{\mathcal{C}^u} \leq A_{\mathcal{C}^d}$ with cell uplink coverage range R_u and downlink coverage R_d given, the network power consumption can be lower bounded by

$$\min_R \left\{ \frac{A}{\pi R_d^2}\left(\eta^{-1}K^{-1}P_{min}^d R_d^\alpha 10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out}^d)} + P_c^d\right) + \frac{A(P_c^{on}+P_c^u)}{\pi R_u^2} \right\}. \tag{2.55}$$

Furthermore, we can have the following inequality for R_u:

$$P_{t,max} \geq K^{-1}P_{min}^u R_u^\alpha 10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out}^u)}, \tag{2.56}$$

where $P_{t,max}$ is the maximum transmit power of MS. By substituting (2.56) and the optimal value of R_d into (2.55), we can obtain the minimal network power consumption while maintaining the network coverage, denoted by P_N^*, as:

$$P_N^* > \frac{\alpha A}{\pi}\left(\frac{P_c^d}{\alpha-2}\right)^{\frac{\alpha-2}{\alpha}}\left(\frac{P_{min}10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out})}}{2\eta K}\right)^{\frac{2}{\alpha}} + \frac{A(P_c^{on}+P_c^u)}{\pi}\left(\frac{P_{min}^u 10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out}^u)}}{P_{t,max}K}\right)^{\frac{2}{\alpha}}.$$

On the other hand, when $A_{\mathcal{C}^u} > A_{\mathcal{C}^d}$ with cell uplink coverage range R_u and downlink coverage R_d given, the network power consumption can be lower bounded by

$$\min_R \left\{ \frac{A}{\pi R_d^2}\left(\eta^{-1}K^{-1}P_{min}^d R_d^\alpha 10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out}^d)} + P_c^d + P_c^{on}\right) + \frac{AP_c^u}{\pi}\left(\frac{P_{min}^u 10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out}^u)}}{P_{t,max}K}\right)^{\frac{2}{\alpha}} \right\}.$$

Similarly, we can obtain the lower bound of minimal network power consumption while maintaining the network coverage, denoted by P_N^*, as:

$$P_N^* > \frac{\alpha A}{\pi}\left(\frac{P_c^d + P_c^{on}}{\alpha - 2}\right)^{\frac{\alpha-2}{\alpha}}\left(\frac{P_{min}10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out})}}{2\eta K}\right)^{\frac{2}{\alpha}} + \frac{AP_c^u}{\pi}\left(\frac{P_{min}^u 10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out}^u)}}{P_{t,max}K}\right)^{\frac{2}{\alpha}}.$$

For symmetric uplink and downlink, cell coverage $\mathcal{C}_m$ is the intersection of the uplink coverage and downlink coverage (i.e., $\mathcal{C}_m = \mathcal{C}_m^d \bigcap \mathcal{C}_m^u$) and $I_m^u = I_m^d = I_m^{on}$.

If $A_{\mathcal{C}^u} \leq A_{\mathcal{C}^d}$, the network power consumption can be lower bounded by

$$\frac{A}{\pi}\left(\left(\frac{P_{min}^u 10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out}^u)}}{P_{t,max}K}\right)^{\frac{2-\alpha}{\alpha}}\frac{P_{min}^d 10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out}^d)}}{\eta K}\right) + \frac{A(P_c^{on} + P_c^u + P_c^d)}{\pi}\left(\frac{P_{min}^u 10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out}^u)}}{P_{t,max}K}\right)^{\frac{2}{\alpha}}.$$

On the other hand, if $A_{\mathcal{C}^u} > A_{\mathcal{C}^d}$, the network power consumption can be lower bounded by

$$P_N^* > \frac{\alpha A}{\pi}\left(\frac{P_c^d + P_c^u + P_c^{on}}{\alpha - 2}\right)^{\frac{\alpha-2}{\alpha}}\left(\frac{P_{min}10^{\frac{\sigma_\Psi}{10}Q^{-1}(p_{out})}}{2\eta K}\right)^{\frac{2}{\alpha}}$$

2.4.2 COMIC for Joint Uplink and Downlink Coverage Preservation

We extend the proposed COMIC algorithm for joint uplink and downlink network coverage preservation. For symmetric uplink and downlink, COMIC algorithm can be applied with $\mathcal{C}_m = \mathcal{C}_m^u \bigcap \mathcal{C}_m^d$ in which $\mathcal{C}_m^u$ is the uplink coverage satisfying the uplink coverage requirements, $\mathcal{C}_m^d$ is the downlink coverage satisfying the uplink coverage requirements.

To ensure the coverage of a network with both uplink coverage-limited BSs and downlink coverage-limited BSs, we propose the following asymmetric COMIC algorithm for the asymmetric uplink and downlink. Specifically, in the asymmetric COMIC algorithm, we first ensure the network uplink coverage and then minimize the network are power consumption for maintaining network downlink coverage. In this way, the extra power consumption to activate a BS for maintaining downlink

coverage will be smaller for a BS which is already activated for maintaining uplink coverage. With the preferences in already activated BSs in asymmetric COMIC, more constant operational power of BSs can be saved.

The steps of the asymmetric COMIC algorithm are as follows.

1. Choose a BS which has the maximal coverage range from $\mathcal{B}$ (say B_1) as the initial active BS, and add it as the first element of the activated uplink BS set, denoted by $\mathcal{B}^u_{on}$, (i.e., $\mathcal{B}^u_{on} = \{B_1\}$).

2. Find B_k which has the maximal coverage range from $\mathcal{B} \setminus \{B_1\}$ and $\mathcal{C}^u_k \cap \mathcal{C}^u_1 \neq \emptyset$. Set B_k as the new active BS. Add the new active BS B_k to $\mathcal{B}_{on}$. Denote the set of uncovered boundary intersections between activated BSs as $\mathcal{I}^u$ and initiate $\mathcal{I}^u = \emptyset$.

3. For each $B_m \in \mathcal{B}^u_{on}$, find boundary intersection i between B_k and B_m which is not covered by any BS in the set of $\mathcal{B}^u_{on} \setminus \{B_m\}$. Add boundary intersection i to $\mathcal{I}$. Repeat until no such intersections exist for B_m, i.e.,

$$\mathcal{I}^u = \mathcal{I}^u \bigcup \{i | i \in \bigcap_{B_j \in \{\mathcal{B}^u_{on} \setminus \{B_m\}\}} \mathcal{C}'_j \bigcap \partial\mathcal{C}^u_m \bigcap \partial\mathcal{C}^u_k \bigcap \mathcal{A}\},$$

   where $\mathcal{C}'_j$ represents the complement of $\mathcal{C}^u_j$, $\partial\mathcal{C}^u_m$ is the boundary of $\mathcal{C}^u_m$, and $\mathcal{A}$ is the target covering region.

4. For each new $i \in \mathcal{I}^u$, find the optimal BS for i. The optimal BS for i is the BS which is closest to the optimal BS position of i and can cover the boundary intersection i. Set these BSs as candidate BSs.

5. Choose B_k which minimizes the distance from its position to the optimal position of the corresponding boundary intersection from the candidate BSs as the new active BS. Add the new active BS B_k to $\mathcal{B}^u_{on}$.

6. For each $i \in \mathcal{I}^u$, if i is covered by B_k, remove it from $\mathcal{I}^u$.

7. Go to Step 3 until $\mathcal{I}^u = \emptyset$.

8. Find uncovered boundary intersections at network boundaries, i.e.,

$$\mathcal{I}^u_{\partial\mathcal{A}} = \bigcup_{B_m \in \mathcal{B}^u_{on}} \partial\mathcal{C}^u_m \cap \mathcal{C}'_{on} \cap \partial\mathcal{A} \tag{2.57}$$

where $\mathcal{I}^u_{\partial\mathcal{A}}$ is the set of uncovered boundary intersections between active cells and the target area $\mathcal{A}$, $\mathcal{C}^u_{on} = \bigcup_{B_m \in \mathcal{B}^u_{on}} \mathcal{C}^u_m$ and $\mathcal{C}'_{on}$ is the complement of $\mathcal{C}^u_{on}$.

9. If $\mathcal{I}^u_{\partial\mathcal{A}} \neq \emptyset$, choose B_k such that $\mathcal{C}^u_k$ minimize the overlap with $\mathcal{C}^u_{on}$ and can cover a boundary intersection $j \in \mathcal{I}^u_{\partial\mathcal{A}}$. For each $j \in \mathcal{I}^u_{\partial\mathcal{A}}$, if j is covered by B_k, remove it from $\mathcal{I}^u_{\partial\mathcal{A}}$. Add the new active BS B_k to $\mathcal{B}^u_{on}$.

10. Go to Step 3 until $I^u_{\partial\mathcal{A}} = \emptyset$.

11. For the target area power consumption ζ, find the maximal downlink coverage for each BS $\{\mathcal{C}^d_1, \ldots, \mathcal{C}^d_M\}$.

12. Choose a BS which has maximal downlink coverage range from $\mathcal{B}$ (say B_1) as the initial active BS, and add it as the first element of the activated downlink BS set, denoted by $\mathcal{B}^d_{on}$, (i.e., $\mathcal{B}^d_{on} = \{B_1\}$).

13. Find B_k which has maximal uplink coverage range from $\mathcal{B} \backslash \{B_1\}$ and $\mathcal{C}^d_k \cap \mathcal{C}^d_1 \neq \emptyset$. Set B_k as the new active BS. Denote the set of uncovered boundary intersections between activated BSs as $\mathcal{I}^d$ and initiate $\mathcal{I}^d = \emptyset$.

14. For each $B_m \in \mathcal{B}^d_{on}$, find boundary intersection i between B_k and B_m which is not covered by any BS in the set of $\mathcal{B}^d_{on} \setminus \{B_m\}$. Add boundary intersection i to $\mathcal{I}^d$. Repeat until no such intersections exist for B_m, i.e.,

$$\mathcal{I}^d = \mathcal{I}^d \bigcup \{i | i \in \bigcap_{B_j \in \{\mathcal{B}^d_{on} \backslash \{B_m\}\}} \mathcal{C}''_j \bigcap \partial\mathcal{C}^d_m \bigcap \partial\mathcal{C}^d_k \bigcap \mathcal{A}\},$$

where $\mathcal{C}''_j$ represents the complement of $\mathcal{C}^d_j$ and $\partial\mathcal{C}^d_m$ is the boundary of $\mathcal{C}^d_m$. Finally, add the new active BS B_k to $\mathcal{B}^d_{on}$.

15. For each new $i \in \mathcal{I}^d$, find the optimal BS for i. The optimal BS for i is the BS which is closest to the optimal BS position of i and can cover the boundary intersection i. Set these BSs as candidate BSs. If no BS can cover the boundary intersection i, set $\zeta = \zeta + \Delta\zeta$ and go to Step 11.

16. Choose B_k which minimizes the distance from its position to its optimal position of the corresponding boundary intersection from the candidate BSs as the new active BS.

17. For each $i \in \mathcal{I}^d$, if i is covered by B_k, remove it from $\mathcal{I}^d$.

18. Go to Step 11 until $\mathcal{I}^d = \emptyset$.

19. Find uncovered boundary intersections at network boundaries, i.e.,

$$\mathcal{I}^d_{\partial\mathcal{A}} = \bigcup_{B_m \in \mathcal{B}^d_{on}} \partial\mathcal{C}^d_m \cap \mathcal{C}''_{on} \cap \partial\mathcal{A} \tag{2.58}$$

where $\mathcal{I}^d_{\partial\mathcal{A}}$ is the set of uncovered boundary intersections between active cells and the target area $\mathcal{A}$, $\mathcal{C}^d_{on} = \bigcup_{B_m \in \mathcal{B}^d_{on}} \mathcal{C}^d_m$ and $\mathcal{C}''_{on}$ is the complement of $\mathcal{C}^d_{on}$.

20. If $\mathcal{I}^d_{\partial\mathcal{A}} \neq \emptyset$, choose B_k such that $\mathcal{C}^d_k$ minimize the overlap with $\mathcal{C}^d_{on}$ and can cover a boundary intersection $j \in \mathcal{I}^d_{\partial\mathcal{A}}$. For each $j \in \mathcal{I}^d_{\partial\mathcal{A}}$, if j is covered by B_k, remove it from $\mathcal{I}^d_{\partial\mathcal{A}}$. Add the new active BS B_k to $\mathcal{B}^d_{on}$.

21. Go to Step 11 until $I^d_{\partial\mathcal{A}} = \emptyset$.

2.4.3 Performance Evaluation for Various Network Structures

In this section, we evaluate the performance of various potential network architectures in HetNets for energy saving during the low traffic load period. In the simulations, we consider a network with regularly deployed macro BSs and randomly deployed small cell BSs. Specifically, we will compare the minimal network power consumption

Table 2.2: Simulation parameters of COMIC for joint uplink and downlink.

Parameter	**Value**
channel model	path loss+shadowing
shadowing standard deviation σ_Ψ	6 dB
minimum received power of macro BS	-120 dBm
minimum received power of small cell BS	-108 dBm
minimum received power of MS	-100 dBm
antenna gain of macro BS	15 dBi
antenna gain of small cell BS	2 dBi
transmit power efficiency of macro BS	0.5
transmit power efficiency of small cell BS	0.5
outage probability p_{out}	0.1
path loss exponent α	4
network size	30 Km by 30 Km
K (reference distance = 1 m) for macro BS	10^{-1}
ratio of uplink constant operational power	0.3
ratio of downlink constant operational power	0.3
ratio of active constant operational power	0.4

among only power control, cell activation with symmetric uplink and downlink, and cell activation with asymmetric uplink and downlink in HetNets. Here we compare only the network power consumption to avoid signal coverage holes during the low traffic load periods. The simulation parameters are given in Table 2.2.

Figure 2-9 and Figure 2-10 show the aggregated power consumption of BSs and the corresponding lower bounds for grid BS deployment and hexagonal BS deployment in a network, respectively. In the figures, compared with only power control, we can observe that the gain of cell activation with symmetric connection depends highly on the network topologies. The reason is that the uplink coverage restricted by MS maximum transmit power will limit the opportunities for a cell to extend its coverage with the requirements on symmetric uplink and downlink connections between MSs and BSs. By decoupling uplink connection and downlink connection by the proposed reception mode design for BSs, we demonstrate that the network power consumption can be further reduced. In the figure, we can observe that 1-2 dB gain can be obtained by asymmetric uplink and downlink.

Besides, with the exploration of data rate demands in the future, small cells

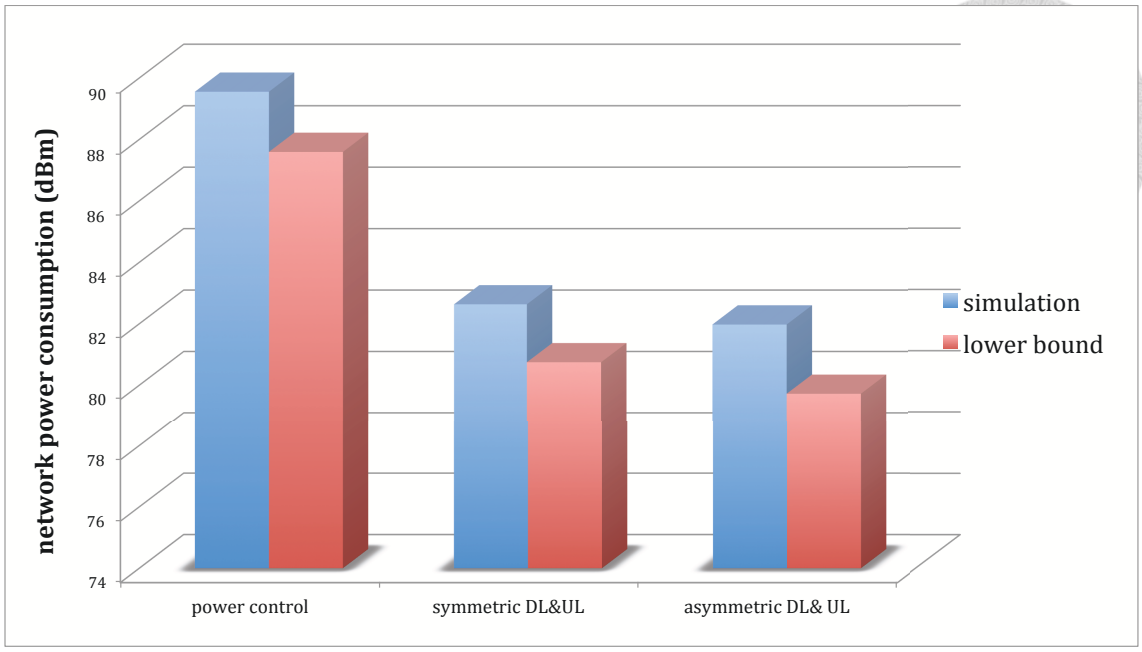


Figure 2-9: The network power consumption of grid deployment of 900 BSs for different power saving strateiges. Both the simulation results and the lower bounds of network power consumption are shown.

are expected to be densely deployed in the network for bursting network capacity. An interesting question is whether network densification by small cells for bursting network capacity can further reduce network energy consumption during the low traffic load period. In the following, we evaluate various potential network structures in HetNets for network energy saving.

Figure 2-11 and Figure 2-12 show the network power consumption as a function of the constant operational power of a small cell with various network structures. Taking a glance in the figures, we can observe that not all schemes with small cells can achieve a better performance than activating only macro BSs. Compared with asymmetric uplink and downlink connections by only macro BSs, a promising approach which can achieve lower network power consumption with 4000 small cells in the network is the approach of uplink by small cells and downlink by macro cells. From Figure 2-11, we can find that uplink by macro BSs and downlink by small cells may be our least choice for network energy saving. Due to the fact that uplink coverage of a small cell is similar to uplink coverage of a macro cell, more network energy can be conserved

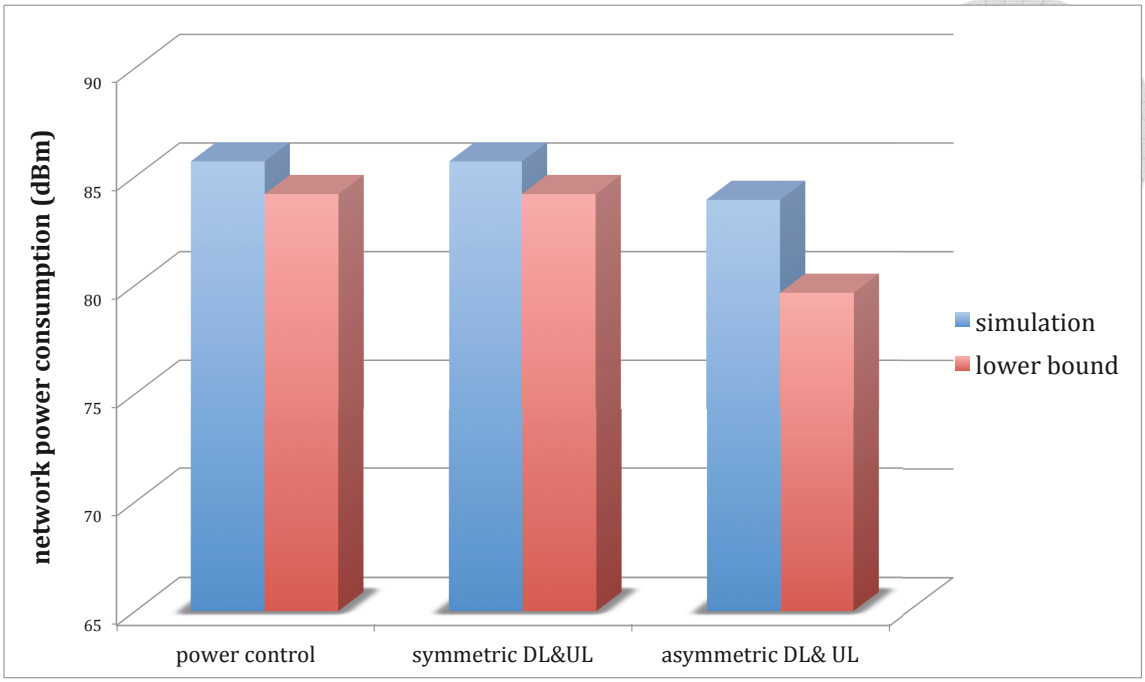


Figure 2-10: The network power consumption of hexagonal deployment of 368 BSs for different power saving strateiges. Both the simulation results and the lower bounds of network power consumption are shown.

by activating a small cell which has a much lower uplink power consumption than that of a macro BS. On the other hand, due to the better transmit power efficiency and antenna gain of macro cells, the approach of downlink by macro BSs will be more energy-efficient than the approach of downlink by small cells. Therefore, the approach that uplink by small cells and downlink by macro cells has a much better performance than other approaches. From Figure 2-11, we can observe that this approach is better than all the other approaches as long as the constant operational power of a small cell is sufficiently small. Compared with network coverage preservation by only macro BSs, around 6 dB more power can be conserved by the approach (uplink by small cells and downlink by macro cells). To evaluate the network energy saving by network densifcation with small cells, we also show the results for a large amount of small cells deployed in the network. In this case, the increase of the deployed small cells will hardly affect the network power consumptions. Figure 2-12 shows the network power consumption as a function of the constant operational power of a small cell. In Figure 2-12, we can find that the network power consumption will be significantly reduced

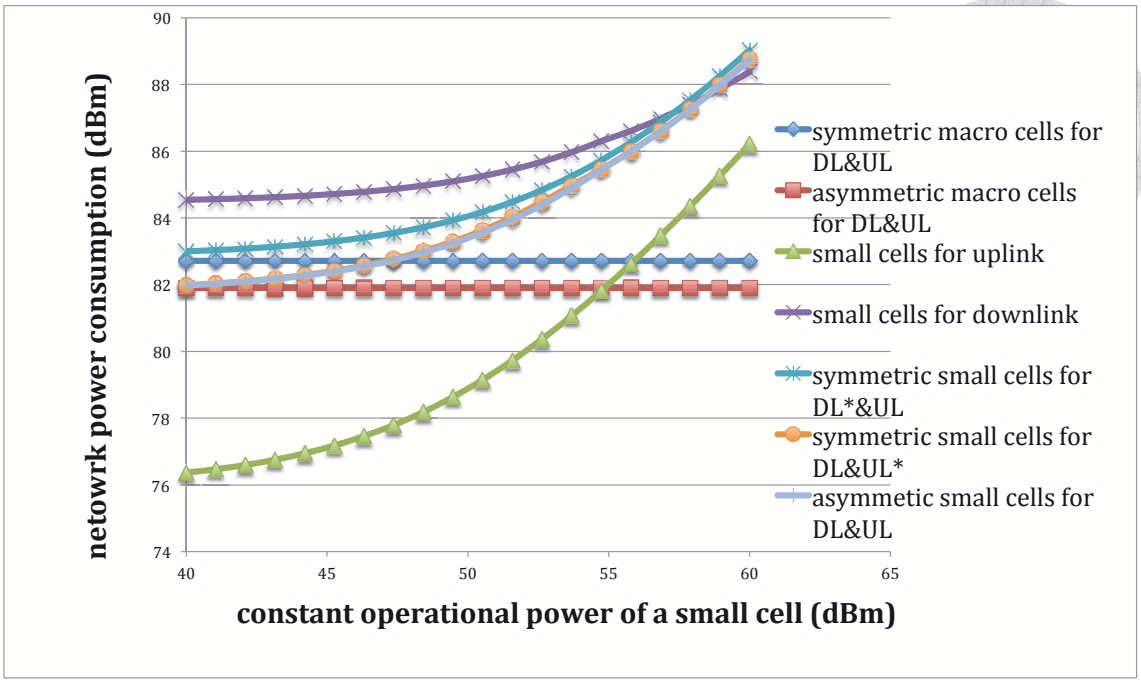


Figure 2-11: The network power consumption as a function of the constant operational power of a small cell. Various network structures in HeNets are examined with 368 hexagonally deployed macro-cell BSs and 4000 randomly deployed small-cell BSs.

in the approach of maintaining coverage by only small cells compared with only 4000 small cells deployed in the network. With the state of art of power consumption in small cells (e.g., at the level of 10 Watts), maintaining network coverage by only small cells seems not the most energy-efficient approach. The reason is that the antenna gain provided by macro cells (e.g., by sectorization) can greatly reduce the transmit power consumption for maintaining downlink coverage. The lower gain omni-directional antenna adopted by small cell results in more power consumption by maintaining downlink coverage. The approach that maintaining uplink network coverage by small cells and downlink network coverage by macro cells will be most energy-efficient. Besides, it will be fitted into the proposed structure that an umbrella cell and multiple small cell to form a two-tier network [1]. It is intuitive that the umbrella cell serves as a downlink coverage preservation cell and the other small cells serve as uplink coverage preservation cells in the two-tier networks. Furthermore, to reduce signalling latency, the umbrella cell can transmit the control signalling to the small cells through the downlink channels during the low traffic load period under

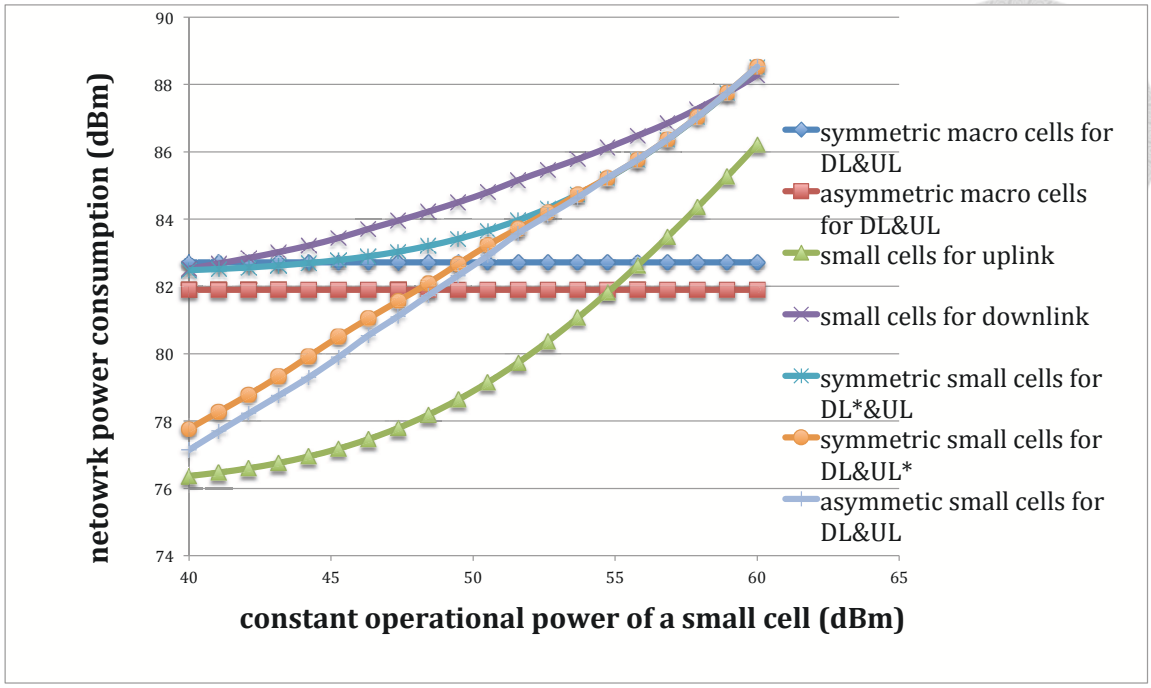
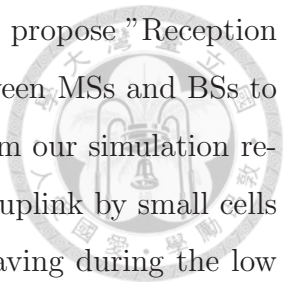


Figure 2-12: The network power consumption as a function of the constant operational power of a small cell. Various network structures in HeNets are examined with 368 hexagonally deployed macro-cell BSs and 40000 randomly deployed small-cell BSs.

the two-tier network structure with downlink by only umbrella cells.

2.5 Summary

In this chapter, we study the effectiveness of powering off BSs to minimize network power consumption while satisfying the network-wide coverage requirement for various BS deployment strategies. We analyze the power consumption of the whole network and determine the optimal power consumption and the cell size of an active BS for minimizing the area power consumption. With the obtained optimal cell coverage, we show that the minimum-power BS activation problem is NP-hard and we propose a polynomial-time load-aware algorithm called COMIC for minimum-power BS activation while maintaining network-wide coverage. The simulation results show that the performance of COMIC can approach the performance upper bound for various kinds of BS deployment strategies. Besides, we show that the deployment of small cells can not only burst network throughput during the peak load periods but



also save energy in the low traffic load periods. Furthermore, we propose "Reception Mode" design in conjunction with asymmetric connections between MSs and BSs to further explore the opportunities of network energy saving. From our simulation results, we suggest that the asymmetric BS-MS connections with uplink by small cells and downlink by macro cells is promising for network energy saving during the low traffic load periods for green HetNets.

Chapter 3

Sleep Mode Operations for Message Delivery in Green Wireless Multi-Hop Relay Networks

The explosive growth of battery-powered wireless devices in recent years makes multi-hop relaying a promising approach for energy-efficient communications [48, 49]. Instead of connecting directly to the remote sink or gateway, delivering messages through multiple small hops allows wireless devices to use small transmission powers for improving spatial reuse and energy efficiency. Many emerging applications such as Machine-to-Machine (M2M) networks [50] and Device-to-Device (D2D) communications [51] have also considered the adoption of multi-hop relaying in mainstream wireless networks.

To conserve energy in wireless networks, an alternative approach that has been popularly employed is to put the nodes into a so-called "sleep" mode. The term "sleep" generally refers to the practice where a communication terminal disengages itself for a short period of time from the network, during which the transceiver is shut off to greatly reduce the power consumption. Many energy-efficient sleep-awake

scheduling mechanisms have been proposed for energy conservation of wireless devices in the literature [52–56].

One fundamental challenge when sleep-awake mechanisms are applied in wireless relay networks (WRNs) is how to maximize the energy efficiency while satisfying the requirement on data delivery delay time [57–59]. In general, delivery delay curtailment and network energy conservation can be mutually conflicting goals, so routing and sleep parameters should be chosen carefully to achieve the best balance between the two goals. In [60], the authors identify that most existing works treat energy efficient routing [61–63] and sleep scheduling [57, 64–66] as two separate tasks, assuming that one is pre-determined while optimizing the other. They propose an algorithm to demonstrate the importance of joint routing and sleep scheduling for network lifetime maximization. The authors in [59] further investigate the joint optimization of the wakeup rate and the forwarding set for minimizing the energy consumption of wakeup radios under the consideration of delivery delay constraint. Due to the hardness of modeling power consumption in wireless multi-hop networks, however, *existing works for the lifetime maximization problems are often numerically solved* [58–60]. The drawbacks of using only numerical results include the potential lack of insights for profiling the trade-offs among different network design parameters and the difficulty in applying these results in large-scale wireless networks.

Ideally, when sleep-awake mechanisms are applied in a wireless relay network, several questions that should be answered for optimal network operations include: 1) What is the optimal relationship between network energy consumption and message delivery delay time such that a network designer can strike a balance between these two metrics? 2) How to jointly select routing parameters and sleep parameters to achieve the optimal trade-offs? 3) How network parameters (e.g., network size or message origination rate) affect the optimal values of routing parameters and sleep parameters? We note that prior analytical models usually focus on how to precisely model the network performance when network operational parameters (e.g., transmission range and duty-cycle period) are all determined in advance [67,68]. The lack of a model that can be used to jointly select routing parameters and sleep parameters for

achieving optimal performance trade-offs in energy-efficient wireless relay networks motivates our work towards this direction.

In this chapter, our goal is to determine the optimal relationship between network energy consumption and the message delivery delay for wireless relay networks with sleep-awake schedule. To proceed, we first propose the general concept of a *random wakeup network* (RWN) with opportunistic relay as a framework for our analysis. Three key parameters have been considered in the proposed framework to capture the energy costs of routing, media access, and sleep mechanisms, including transmission range, duty-cycle period, and wakeup density. We then derive *closed-form relationships* between the three design parameters and the network power consumption in the random wakeup network with opportunistic relay. Based on the analytical expressions, we consider the optimization problem of minimizing network energy consumption under the requirement of end-to-end message delivery time. Through simulation of a wireless relay network based on geographic routing, we demonstrate that the proposed analytical framework can be applied reasonably well for profiling the performance trade-offs between network energy consumption and message forwarding delay.

Based on the analytical model derived in this chapter, we have found several interesting yet surprising results regarding the optimal values of transmission range, wakeup density, and duty-cycle period in planar networks. Firstly, there does exist an optimal number of nodes participating in the relay activities within the transmission range of the transmitter in an optimized RWN. In typical scenarios where the transmit energy consumption induced by a transmitter is larger than that by a receiver and the path loss exponent is in the regime of $2 \leq \alpha \leq 4$, this optimal number shall be larger than one and small than ten on average. Secondly, the optimal power consumed for message detection is of the same order of the power consumed for actual communications from the perspective of the entire network. The optimal ratio between the two quantities is $\frac{\alpha-1}{3}$. That it is optimal to dedicate such a large portion of power consumption to incoming message detection (carrier sensing) for optimal network operations could be a surprise to some. Finally, the investment of network

power can efficiently reduce the message delivery delay when all design parameters can be jointly optimized. The minimal network power consumption under optimal operations grows at a rate of the $\frac{\alpha-1}{\alpha+2}$-th order of the message delivery speed. One needs to invest less than an α-fold increase in total power in order to increase the mean delivery speed by a factor of 10.

Therefore, the main contributions in this chapter include:

1. We propose an analytical framework for modeling the trade-off between network energy consumption and message delivery delay while a random sleep-awake mechanism is adopted.

2. We derive closed-form relationships between routing and sleep parameters and relevant network metrics in the optimized RWN to provide insights for network design.

3. We provide closed-form bounds and approximations for profiling the trade-offs between network energy consumption and message delivery delay in the RWN.

4. We show through a case study that our model can be applicable for both network performance modeling and operational parameter optimization in wireless relay networks with a sleep mode operation.

The rest of this chapter is organized as follows. In Section 3.1, we present the framework of random wakeup networks used in our analysis and in Section 3.2, we analyze the power consumption for message delivery in an RWN. Based on the closed-form approximations, in Section 3.3, we derive optimal settings for the transmission range, duty-cycle period, and wakeup density with a given delivery time requirement. In Section 3.4, we generalize and extend our results in 1-D, 2-D, and 3-D networks. In Section 3.5, we evaluate the proposed analytical models by simulating wireless relay networks with a sleep mode operation. Finally, summary of the chapter is given in Section 3.6.

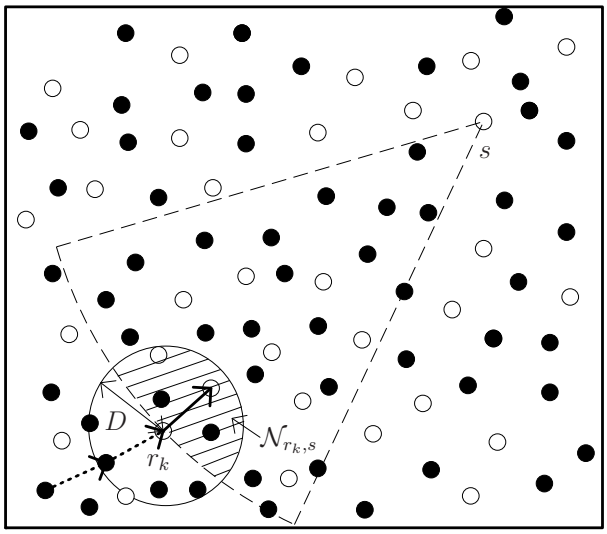


Figure 3-1: An example of a message traversing from a source node to the sink node in the RWN. A message originated from a source node is to be delivered to the sink node by multi-hop transmissions. In epoch k, some nodes (white nodes) wake up to listen to potential incoming traffic. If node r_k currently relays a message to sink node s, the effective neighborhood of node r_k, $\mathcal{N}_{r_k,s}$, is the shaded region. Awake nodes (two white nodes) in the effective neighborhood of node r_k serve as candidates for the next forwarding node. The wakeup node in the effective neighborhood of node r_k closest to the sink node will be selected as the next forwarding node.

3.1 Random Wakeup Framework for Message Delivery

We consider a network where N nodes, each with transmission range D, are randomly deployed within an area A. Each node in the network can serve as an originator or a relay node of messages destined to the sink node (e.g., gateway or base station). As shown in Fig. 3-1, a message is relayed to the sink node by multi-hop transmissions. In the following, we first describe the sleep and relay mechanisms in the network that we refer to as random wakeup network (RWN), and then describe the role-based energy consumption model for performance analysis. Key parameters used in the proposed framework are summarized in Table 3.1.

3.1.1 Random Sleep-Awake Schedule

In the RWN, the time is partitioned into contiguous same-sized epochs of common duration denoted by T_d (*duty-cycle period*). As shown in Fig. 3-2, all nodes follow a random sleep-awake schedule where nodes can participate in wireless communication only at the start of a period; for the rest of the period, the transceivers are shut down to save power. Any node that does not have a message to transmit would unilaterally determine whether it will wake up to detect the presence of messages with a given probability. On the other hand, any node that wishes to transmit a message, either as an originator or as a relay, will wake up to transmit the message at the start of a period. We define the wakeup density p as the geographical density of nodes awakened to listen to potential incoming traffic. It is equal to the expected number of awake nodes for message detection per unit area.

Since a node in a duty-cycle period can transmit messages to other nodes, wake up to listen to potential incoming traffic, or stay in sleep mode to conserve energy, we divide a duty-cycle period into the following four intervals: listening interval, data transmission-reception interval, route negotiation interval and sleep interval. Nodes with different roles will take different actions in these intervals. For example, in Fig. 3-3, node A transmits a message in the duty-cycle period, node B and node C wake up to listen to potential incoming traffic, and node C switches its radio transceiver while no interested message is detected.

3.1.2 Opportunistic Relay with Sleep-Awake Nodes

For a message to traverse from a source node to the sink node, we apply a simple opportunistic relay mechanism [64, 67]. In the following, we describe the behavior of nodes for opportunistic relay in each of the three active intervals of a duty-cycle period.

In the listening interval, a node that has a message to transmit can send certain request signals (e.g., RTS or ATIM) to notify its neighbors. For opportunistic relaying, this message can usually be a *broadcast* message specifying the priorities of the

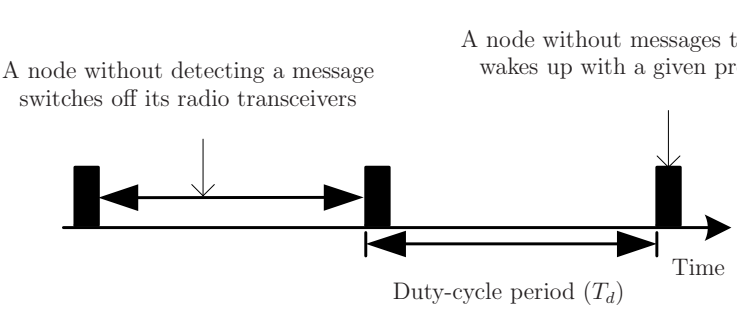


Figure 3-2: An illustration of the sleep-awake mechanism in the random wakeup network. A node without a message to transmit wakes up randomly in a duty-cycle period. If it does not detect any message during the listening interval, it will go back to sleep mode immediately.

neighbors to be the next forwarding node. On the other hand, a node without any message to transmit will wake up randomly to detect messages and return to sleep mode at the end of the listening interval. If no interested signals are detected, an awake node will switch into sleep mode during the rest of the duty-cycle period. If a request signal is detected, nodes will reply to the request signals in the listening interval.

In the data transmission-reception interval, the transmitter will broadcast the message to its neighbors and the neighboring nodes that reply to the transmitter in the listening interval will wake up to receive it. Define the *effective neighborhood* $\mathcal{N}_{r_k,s}$ of transmitting node r_k as a region in which a node closer to sink node s can successfully receive messages transmitting from node r_k. The shaded area in Fig. 3-1 shows the effective neighborhood of node r_k.

In the route negotiation interval, the transmitter will negotiate with neighboring nodes that have successfully received the message to determine the next transmitter [69]. In the RWN, only awake nodes in the effective neighborhood of the current forwarding node will be considered to forward the message during route negotiation interval. These nodes will serve as *candidate forwarders* for the next transmission. While it is possible to choose the node closest to the sink as the next forwarding node, our analytical framework does not dictate a particular mechanism for relay node selection. Note that in one duty-cycle period, a message can complete *at most*

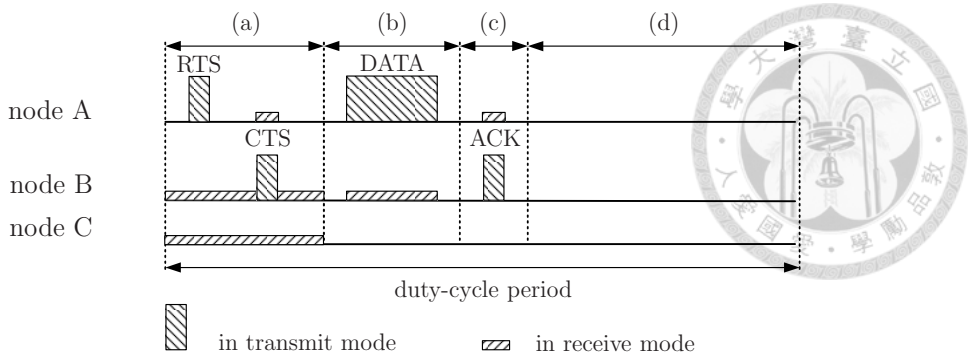


Figure 3-3: An example of node's behavior in a duty-cycle period. A duty-cycle period can be divided into four intervals: (a) listening interval (b) data transmission-reception interval (c) route negotiation interval (d) sleep interval. These four intervals may be reordered for different scenarios.

one hop, i.e., moving from the transmitting node to the resolved forwarding node. If, unfortunately, none of the nodes in the effective neighborhood of the current forwarding node wakes up to relay the message, the transmitting node should repeat the process during the next period.

Note that depending on the particular medium-access mechanism used, *these intervals may be collocated or disordered.* For example, if the message size is small, a network designer may adopt a MAC protocol in which a transmitter sends messages directly in the listening interval instead of sending control frames (e.g., RTS in the 802.11 basic access method). In this case, the listening interval and the data transmission-reception interval are collocated. Our analytical framework does not dictate a particular handshake procedures between the transmitter and forwarder as evident from the role-based energy consumption discussed in the following.

3.1.3 Role-Based Energy Consumption Model

Different from various energy consumption models proposed in the literature based on the modes of operation such as transmit, receive, sense (detect), and sleep, in this chapter we consider a role-based energy consumption model that puts together all energy components consumed for acting a particular role during opportunistic relay. In particular, we consider the energy consumed for transmitters, candidate forwarders, and awake nodes without detecting any message. Since the power consumption in

sleep mode is usually order-of-magnitude smaller than the other types of power consumption, we omit those nodes that do not wake up to detect incoming traffic in the framework.

To better understand the role-based energy consumption model, we consider a typical message handshake as shown in Fig. 3-3. The total energy consumed by a transmitter (e.g. node A in Fig. 3-3) in a duty-cycle period to send RTS, send data, and listen for CTS and ACK is modeled by the quantity E_t. The total energy consumed by a candidate forwarder (e.g. node B in Fig. 3-3) in a duty-cycle to detect incoming message, receive data, and transmit CTS and ACK is modeled by the quantity E_c. The energy consumed by a node awakened to detect incoming message in the listening interval is modeled by the quantity E_d. It is also the total energy consumed for an awake node without detecting any message (e.g. node C in Fig. 3-3). In the following, we describe how we model different energy consumption quantities.

The energy consumption of a transmitter in a duty-cycle period, denoted by E_t, is a function of the transmission range. For a path loss exponent α, E_t is modeled as $E_t^1 D^\alpha + E_t^0$, where E_t^1 stands for the range-dependent energy consumption for a transmitter to reach a node at unit distance in the entire duty-cycle period and E_t^0 stands the range-independent energy consumed by the transmitter in the entire duty-cycle period (e.g. the energy consumed to receive CTS or ACK from receivers or circuit energy consumption while transmitting). The energy consumption for a node awakened to detect potential incoming messages during a duty-cycle period is denoted by E_d (detection energy). The energy consumption for a node that serves as a candidate forwarder to receive a message during a duty-cycle period is denoted by E_c. Since a candidate forwarder will need to transmit packets and negotiate with the transmitting node, E_c can be modeled by $E_c^1 D^\alpha + E_c^0$, where E_c^1 stands for the range-dependent energy consumed to negotiate with the transmitter at unit distance and E_c^0 stands for the range-independent energy consumed in the duty-cycle period. Note that a node that does not wake up in a duty-cycle period consumes no energy in our model.

Table 3.1: Parameters used in the RWN. The parameters are categorized into environmental parameters, application parameters, energy parameters, and design parameters.

Environmental Parameters	
A:	area of the network
α:	path loss exponent
Application Parameters	
T_m:	mean message origination period
T_1:	maximum delay time to advance unit distance
Energy Parameters	
E_t:	energy consumed of a transmitter in a duty-cycle period
E_c:	energy consumed of a candidate forwarder in a duty-cycle period
E_d:	energy consumed of a wakeup node for message detection in a duty-cycle period
Design Parameters	
T_d:	duration of the duty-cycle period
p:	wakeup density of nodes
D:	transmission range

Based on the energy consumption model for each node in the network, we define in the following the quantities for capturing the energy consumption of *all nodes* in the entire network.

Definition 3.1. *The energy consumed for listening (ECL) is defined as the energy consumption for all nodes in the network to wake up to listen to potential incoming messages. On the other hand, the additional energy consumed over listening (AECOL) is defined as the energy consumption for all nodes to transmit, receive, and forward messages on top of the energy consumed for listening.*

Briefly, if a network does not have any message in propagation, nodes in the network would still consume energy because they do wake up to detect messages. We refer to such detection energy consumption as ECL. In the case of no messages, AECOL for the network is zero. While ECL is usually a simple function of the sleep parameters, AECOL depends heavily on routing and sleep parameters that we show how it can be derived in the following.

3.2 Energy Consumption in RWNs

In this section, we compute the network power consumption by deriving AECOL and ECL in the proposed framework. Since we consider a large-sized RWN with low traffic density, occurrences of events such as buffer overflow, message queueing, channel contention, or data collision are rare. Hence, we do not consider these rare events in the following analysis. In addition, since the sink node is more powerful and has more energy resources than all the other nodes in the network, we exclude the energy consumption of the sink node in the analysis.

3.2.1 Expected AECOL for a Message

To start, denote the AECOL incurred by a message traversing from node i to node s by $W_{i,s}$ and its expected value by $\mathbb{E}[W_{i,s}]$. We have the following theorem for finding the expected value of AECOL:

Theorem 3.2. *The expected value of AECOL incurred by a message traversing from node i to node s can be bounded and approximated by the product of the expected number of epochs for a message to traverse from node i to node s and the expected value of AECOL in an epoch as follows:*

$$\mathbb{E}[H_{i,s}]\mathbb{E}[W_{i,s}^k]^- \leq \mathbb{E}[W_{i,s}] \approx \mathbb{E}[H_{i,s}]\mathbb{E}[W_{i,s}^k]' \leq \mathbb{E}[H_{i,s}]\mathbb{E}[W_{i,s}^k]^+,$$

where $H_{i,s}$ stands for the number of epochs it takes for a message traversing from node i to node s and $W_{i,s}^k$ represents the AECOL of the message during epoch k for a message traversing from node i to node s. $\mathbb{E}[W_{i,s}^k]^-$ and $\mathbb{E}[W_{i,s}^k]^+$ are the minimum and maximum values of $\mathbb{E}[W_{i,s}^k]$ respectively, while $\mathbb{E}[W_{i,s}^k]'$ is the approximation of $\mathbb{E}[W_{i,s}^k]$ such that $\mathbb{E}[W_{i,s}^k]^- \leq \mathbb{E}[W_{i,s}^k]' \leq \mathbb{E}[W_{i,s}^k]^+$.

Proof. The expected value of AECOL is the sum of the expected values of AECOL

consumed over all epochs:

$$\begin{aligned}\mathbb{E}[W_{i,s}] &= \mathbb{E}\left[\sum_{k=1}^{\infty} W_{i,s}^k\right] \\ &= \sum_{k=1}^{\infty} \mathbb{E}\left[P(H_{i,s} \geq k)W_{i,s}^k + P(H_{i,s} < k)W_{i,s}^k\right] \quad (3.1) \\ &= \sum_{k=1}^{\infty} P(H_{i,s} \geq k)\mathbb{E}[W_{i,s}^k | H_{i,s} \geq k]. \quad (3.2)\end{aligned}$$

In (3.1), $H_{i,s} < k$ means that the message originated from node i is arrived at node s before time epoch k. When this happens, the AECOL consumed by the message will be zero at time epoch k and $P(H_{i,s} < k)W_{i,s}^k$ is equal to zero. In addition, since $\sum_{k=1}^{\infty} P(H_{i,s} \geq k)$ in (3.2) is equal to $\mathbb{E}[H_{i,s}]$, we can have bounds of $\mathbb{E}[W_{i,s}]$ as follows.

$$\mathbb{E}[H_{i,s}]\mathbb{E}[W_{i,s}^k]^- \leq \mathbb{E}[W_{i,s}] \leq \mathbb{E}[H_{i,s}]\mathbb{E}[W_{i,s}^k]^+, \quad (3.3)$$

where $\mathbb{E}[W_{i,s}^k]^- = \min\{\mathbb{E}[W_{i,s}^1], \ldots, \mathbb{E}[W_{i,s}^{H_{i,s}}]\}$ and $\mathbb{E}[W_{i,s}^k]^+ = \max\{\mathbb{E}[W_{i,s}^1], \ldots, \mathbb{E}[W_{i,s}^{H_{i,s}}]\}$. To satisfy the bounds in (3.3), we can approximate $\mathbb{E}[W_{i,s}^k] \approx \mathbb{E}[W_{i,s}^k]'$ for $k = 1, \ldots, H_{i,s}$ with $\mathbb{E}[W_{i,s}^k]^- \leq \mathbb{E}[W_{i,s}^k]' \leq \mathbb{E}[W_{i,s}^k]^+$. □

To derive the expected number of epochs for a message traversing from node i to node s, we assume that the sink node will wake up to listen to potential incoming traffic in every epoch. We then have the following theorem:

Theorem 3.3. *To transmit a message from source node i to sink node s, the expected number of time epochs it takes, $\mathbb{E}[H_{i,s}]$, can be bounded and approximated by:*

$$\frac{L_{i,s} - D}{\mathbb{E}[X_{i,s}^k]^+} + 1 \leq \mathbb{E}[H_{i,s}] \approx \frac{L_{i,s}}{\mathbb{E}[X_{i,s}^k]'} < \frac{L_{i,s}}{\mathbb{E}[X_{i,s}^k]^-} + 1, \quad (3.4)$$

where $L_{i,s}$ is the separation between node i and node s and $X_{i,s}^k$ is the expected forwarding distance in epoch k. $\mathbb{E}[X_{i,s}^k]^+$ and $\mathbb{E}[X_{i,s}^k]^-$ are the maximal and minimal values of the expected forwarding distance in an epoch respectively, while $\mathbb{E}[X_{i,s}^k]'$ is the approximation of the expected forwarding distance in an epoch such that $\mathbb{E}[X_{i,s}^k]^- \leq \mathbb{E}[X_{i,s}^k]' \leq$

$\mathbb{E}[X_{i,s}^k]^+$.

Proof. Suppose that a message is transmitted from node i to sink node s in epoch 1. We define the forwarding distance in epoch k, denoted by $X_{i,s}^k$, as the distance advanced in the epoch such that $X_{i,s}^k = L_{r_k+1,s} - L_{r_k,s}$, where r_k is the relay node in epoch k. Note that $L_{r_k,s}$ is the remaining distance between relay node r_k and sink node s.

With the notation of $X_{i,s}^k$, the total distance advanced from epoch 1 to epoch t, denoted by $Y_{i,s}^t$, can be written as

$$Y_{i,s}^t = \sum_{k=1}^{t} X_{i,s}^k. \tag{3.5}$$

Since a message in the transmission range of the sink node can reach the sink node with exactly one hop, we can have the following boundary conditions for a message traversing from node i to sink node s:

$$L_{i,s} - D \leq Y_{i,s}^{H_{i,s}-1} < L_{i,s}. \tag{3.6}$$

By computing the expected value of $Y_{i,s}^{H_{i,s}-1}$, we can obtain bounds for $\mathbb{E}[H_{i,s}]$ as:

$$(\mathbb{E}[H_{i,s}] - 1)\mathbb{E}[X_{i,s}^k]^- \leq \mathbb{E}[Y_{i,s}^{H_{i,s}-1}] \leq (\mathbb{E}[H_{i,s}] - 1)\mathbb{E}[X_{i,s}^k]^+,$$

where $\mathbb{E}[X_{i,s}^k]^- = \min\{\mathbb{E}[X_{i,s}^1], \ldots, \mathbb{E}[X_{i,s}^{H_{i,s}-1}]\}$ and $\mathbb{E}[X_{i,s}^k]^+ = \max\{\mathbb{E}[X_{i,s}^1], \ldots, \mathbb{E}[X_{i,s}^{H_{i,s}-1}]\}$. Applying (3.6) in the equation above, we have

$$\frac{L_{i,s} - D}{\mathbb{E}[X_{i,s}^k]^+} + 1 \leq \mathbb{E}[H_{i,s}] < \frac{L_{i,s}}{\mathbb{E}[X_{i,s}^k]^-} + 1. \tag{3.7}$$

It can be easily verified that (3.7) holds for $L_{i,s} \leq D$. Furthermore, if we approximate $\mathbb{E}[H_{i,s}] \approx \frac{L_{i,s}}{\mathbb{E}[X_{i,s}^k]'}$ with $\mathbb{E}[X_{i,s}^k]^- \leq \mathbb{E}[X_{i,s}^k]' \leq \mathbb{E}[X_{i,s}^k]^+$, it can be verified that the approximation will follow the bounds in (3.7). □

From Theorem 3.2 and Theorem 3.3, the AECOL for a message traversing from

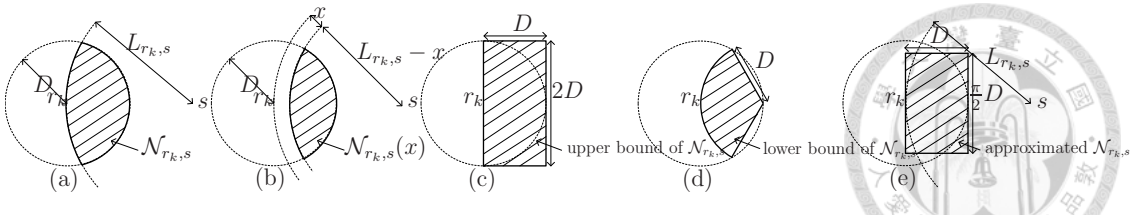


Figure 3-4: An illustration of the effective neighborhood of a transmitter. (a) The effective neighborhood $\mathcal{N}_{r_k,s}$ of transmitter r_k in the RWN in which the separation between the transmitter r_k and sink node s is $L_{r_k,s}$. (b) $\mathcal{N}_{r_k,s}(x)$ is the region where a message can advance a distance more than x toward sink node s from r_k. (c) An upper bound of $\mathcal{N}_{r_k,s}$. (d) A lower bound of $\mathcal{N}_{r_k,s}$. (e) An approximation for $\mathcal{N}_{r_k,s}$.

node i to node s can be approximated as follows:

$$\mathbb{E}[W_{i,s}] \approx \mathbb{E}[H_{i,s}]\mathbb{E}[W_{i,s}^k]' \approx \frac{L_{i,s}}{\mathbb{E}[X_{i,s}^k]'}\mathbb{E}[W_{i,s}^k]'. \tag{3.8}$$

We explain in the following how approximations of $\mathbb{E}[X_{i,s}^k]$ and $\mathbb{E}[W_{i,s}^k]$ can be obtained.

3.2.2 Sensor Field Approximation

To derive expressions for $\mathbb{E}[X_{i,s}^k]$ and $\mathbb{E}[W_{i,s}^k]$, we note that while many studies address the hop count and hop distance problems in wireless multi-hop networks [70–72], these results cannot be directly applied due to the lack of closed-form expressions and the lack of consideration for random sleep-awake mechanisms. We show in this section how we resort to the concept of the sensor field for finding approximations to $\mathbb{E}[X_{i,s}^k]$ and $\mathbb{E}[W_{i,s}^k]$.

Definition 3.4. *A uniform random wakeup sensor field is a network with infinitely many sensors uniformly placed in the network. Over any contiguous area δ, the expected number of nodes that wake up during a time epoch is $p\delta$. Alternatively, the expected number of awake nodes to participate in the relay activities within the transmission range D is $\gamma \equiv \pi p D^2$.*

We first derive the expected forwarding distance per epoch $\mathbb{E}[X_{i,s}^k]$ in a random wakeup sensor field in the following. The boundary effect is neglected in consideration of the fact that messages are rarely routed at network corners for a network whose size is order-of-magnitude larger than the transmission range.

Lemma 3.5. *For a uniform random wake-up sensor field, the expected forwarding distance in an epoch* $\mathbb{E}[X_{i,s}^k]$ *for* $k = 1, \ldots, H_{i,s} - 1$ *with the transmission range* D *and the wakeup density* p *is*

$$\mathbb{E}[X_{i,s}^k] = \int_0^D \left(1 - e^{-p|\mathcal{N}_{r_k,s}(x)|}\right) dx, \tag{3.9}$$

where r_k *is the relay node in epoch* k, $\mathcal{N}_{r_k,s}(x)$ *is the region in the effective neighborhood of* r_k *where a message can advance a distance more than* x *toward sink node* s, *and* $|\mathcal{N}_{r_k,s}(x)|$ *is its area.*

Proof. Consider a message transmitted from node r_k to the designated sink node s at $L_{r_k,s}$ distance away in a uniform random wakeup sensor field. As shown in Fig. 3-4(a), the effective neighborhood of node r_k toward node s is the intersection region $\mathcal{N}_{r_k,s}$. Define $\mathcal{N}_{r_k,s}(x) = \{j \in \mathcal{N}_{r_k,s} : D(r_k, s) - D(j, s) > x\}$ in which $D(a, b)$ represents the separation between node a and node b. As shown in Fig. 3-4(b), to relay a message from node r_k to sink node s, $\mathcal{N}_{r_k,s}(x)$ is a region in the effective neighborhood of r_k where a message can advance a distance more than x toward sink node s.

To find $\mathbb{E}[X_{i,s}^k]$, denote $F_{X_{i,s}^k}(x) \equiv P(X_{i,s}^k \leq x)$ as the distribution function for the distance gains toward sink node s in time epoch k. Clearly, $F_{X_{i,s}^k}(x)$ is equal to the probability that no nodes wake up to listen to the potential incoming messages in region $\mathcal{N}_{r_k,s}(x)$. In the uniform random wakeup field with wakeup density p, $F_{X_{i,s}^k}(x)$ is a truncated exponential distribution as follows:

$$F_{X_{i,s}^k}(x) = e^{-p|\mathcal{N}_{r_k,s}(x)|},\ 0 \leq x \leq D. \tag{3.10}$$

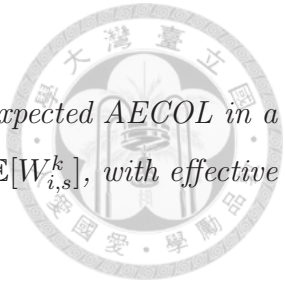
The expected forwarding distance in a time epoch thus can be obtained by

$$\mathbb{E}[X_{i,s}^k] = \int_0^D \left(1 - F_{X_{i,s}^k}(x)\right) dx. \tag{3.11}$$

□

To derive the expected AECOL in a time epoch in the random wakeup sensor

field, we have the following lemma for $\mathbb{E}[W_{i,s}^k]$:



Lemma 3.6. *For a uniform random wake-up sensor field, the expected AECOL in a time epoch for a message traversing from node i to sink node s, $\mathbb{E}[W_{i,s}^k]$, with effective neighborhood $\mathcal{N}_{i,s}$ is*

$$\mathbb{E}[W_{i,s}^k] = E_t + p|\mathcal{N}_{i,s}|(E_c - E_d), \forall k = 1, \ldots, H_{i,s} - 1. \tag{3.12}$$

The expected AECOL for a message in a time epoch, $\mathbb{E}[W_{i,s}^k]$, is the sum of the energy consumed by the transmitter and the extra energy consumed for message reception by the expected number of awake nodes in the effective neighborhood.

Proof. The expected AECOL for a message traversing from node i to node s in epoch k can be represented as:

$$\mathbb{E}[W_{i,s}^k] = P_0^k \mathbb{E}[W_{i,s}^k(0)] + \int_{0^+}^{D} f_{X_{i,s}^k}(x) \mathbb{E}[W_{i,s}^k(x)] dx, \tag{3.13}$$

where P_0^k represents the probability that a message does not find a new forwarding node in epoch k, $f_{X_{i,s}^k}(x)$ represents the probability density function of a message advancing over length x in epoch k, and $W_{i,s}^k(x)$ represents the AECOL for a message advancing distance x in epoch k.

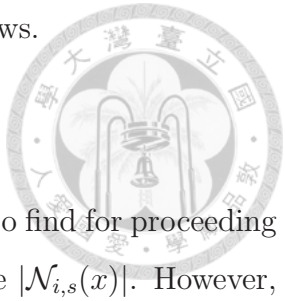
In (3.40), the average AECOL in epoch k, given that the message eventually advances distance x, is

$$\mathbb{E}[W_{i,s}^k(x)] = E_t + p|\mathcal{N}_{i,s} - \mathcal{N}_{i,s}(x)|E_c + E_c - p|\mathcal{N}_{i,s}|E_d, \quad 0 < x \leq D.$$

If no nodes wake up in the effective neighborhood of the current forwarding node, the average AECOL is $\mathbb{E}[W_{i,s}^k(0)] = E_t - p|\mathcal{N}_{i,s}|E_d$. By applying integration by parts for the second term in (3.40), a simple closed-form equation as shown in (3.41) for the expected AECOL of a message traversing from node i to sink node s in epoch k can be obtained. □

Using Lemma 3.5 and Lemma 3.6, we can proceed to obtain the expected for-

warding distance and the expected AECOL in an epoch as follows.



3.2.3 Expected AECOL in a Closed Form

It can be observed from (3.9) and (3.41) that one key parameter to find for proceeding with the derivation is the effective neighborhood of a relay node $|\mathcal{N}_{i,s}(x)|$. However, it is difficult to get closed-form expressions for arbitrary $|\mathcal{N}_{i,s}(x)|$. Instead, we derive closed-form bounds and approximations for the expected forwarding distance and the expected AECOL in an epoch as follows.

As shown in Fig. 3-4(c), the effective neighborhood of relay node r_k in epoch k can be upper bounded by the 2D-by-D rectangular region. Similarly, as shown in Fig. 3-4(d), the effective neighborhood of relay node r_k can be lower bounded by the circular sector with radius D and angle $\frac{\pi}{3}$. By bounding the effective neighborhood of a relay, we have the following lemma for $\mathbb{E}[X_{i,s}^k]$ and $\mathbb{E}[W_{i,s}^k]$.

Lemma 3.7. *In a uniform random wakeup sensor field, the expected forwarding distance in epoch k ($k = 1, \ldots, H_{i,s} - 1$) can be bounded by*

$$D\left(1 - \frac{1 - 2Q(\sqrt{\frac{2}{3}\gamma})}{\sqrt{\frac{4}{3\pi}\gamma}}\right) < \mathbb{E}[X_{i,s}^k] < D\left(1 - \frac{1 - e^{-\frac{2}{\pi}\gamma}}{\frac{2}{\pi}\gamma}\right)$$

and the AECOL in epoch k ($k = 1, \ldots, H_{i,s} - 1$) can be bounded by

$$E_t + \frac{1}{3}\gamma(E_c - E_d) < \mathbb{E}[W_{i,s}^k] < E_t + \frac{2}{\pi}\gamma(E_c - E_d).$$

Note that $\gamma = \pi p D^2$ is the expected number of awake nodes within the transmission range of the transmitter.

Proof. The upper bound (refer to Fig. 3-4(c)) and lower bound (refer to Fig. 3-4(d)) for the expected forwarding distance and the expected AECOL in a duty-cycle period can be obtained by applying Lemma 3.5 and Lemma 3.6, respectively. □

Note that when the sink node is in the effective neighborhood of the transmitter, the message can be relayed to the sink node with exactly one hop. Therefore, $\mathbb{E}[W_{i,s}^{H_{i,s}}]$

for the last hop can be lower bounded by the energy consumption of the transmitter next to the sink node as follows:

$$E_t \leq \mathbb{E}[W_{i,s}^{H_{i,s}}] < E_t + \frac{2}{\pi}\gamma(E_c - E_d). \tag{3.14}$$

Based on the Theorem 3.2, Theorem 3.3, and Lemma 3.7, we can now obtain a lower bound for the expected AECOL of a message traversing from a source node to the sink node as follows:

$$\mathbb{E}[W_{i,s}] > \left(\frac{(E_t + \frac{\gamma}{3}(E_c - E_d))(L_{i,s} - D)}{D\left(1 - \frac{1-e^{-\frac{2}{\pi}\gamma}}{\frac{2}{\pi}\gamma}\right)} \right) + E_t.$$

Similarly, an upper bound for the expected AECOL of a message traversing from a source node to the sink node can also be obtained as follows:

$$\mathbb{E}[W_{i,s}] < \left(\frac{(E_t + \frac{2}{\pi}\gamma(E_c - E_d))L_{i,s}}{D\left(1 - \frac{1-2Q(\sqrt{\frac{2}{3}\gamma})}{\sqrt{\frac{4}{3\pi}\gamma}}\right)} \right) + E_t + \frac{2}{\pi}\gamma(E_c - E_d).$$

To obtain the closed-form approximation (instead of bounds) for $\mathbb{E}[W_{i,s}^k]$, note that in a large network, the number of hops from a source node to the sink node is usually large. When the sink node is *infinitely far away* from the relay node, the effective neighborhood for a relay node is a half-circle with area $\frac{\pi}{2}D^2$. As shown in Fig. 3-4(e), therefore, we can approximate the effective neighborhood of a relay node by a $\frac{\pi}{2}D$-by-D rectangular region and arrive at the following proposition for closed-form approximations of $\mathbb{E}[X_{i,s}^k]$ and $\mathbb{E}[W_{i,s}^k]$.

Proposition 3.8. *For a uniform random wakeup sensor field, the expected forwarding distance and the expected AECOL in a time epoch for a message traversing from node*

i to sink node s can be approximated by

$$\mathbb{E}[X_{i,s}^k] \approx D\left(1 - \frac{1 - e^{-\frac{1}{2}\gamma}}{\frac{1}{2}\gamma}\right), \tag{3.15}$$

$$\mathbb{E}[W_{i,s}^k] \approx E_t + \frac{1}{2}\gamma(E_c - E_d), \tag{3.16}$$

where D is the transmission range and $\gamma = \pi p D^2$ is the expected number of wakeup nodes within transmission range D.

Proof. The expected forwarding distance and the AECOL in an epoch with an effective neighborhood of $\frac{\pi}{2}D$-by-D rectangular region, as shown in Fig. 3-4(e), can be easily obtained by applying Lemma 3.5 and Lemma 3.6, respectively.

We now show that the approximations follow the upper and lower bounds set forth in Lemma 3.7. For the approximated AECOL per time epoch in (3.32), the verification is trivial. For the approximated forwarding distance in a time epoch, to prove that (3.15) is smaller than the upper bound in Lemma 3.7, we need to show that the following equation will be greater than zero for $\gamma > 0$.

$$f(\gamma) = \left(1 - \frac{1 - e^{-\frac{2}{\pi}\gamma}}{\frac{2}{\pi}\gamma}\right) - \left(1 - \frac{1 - e^{-\frac{1}{2}\gamma}}{\frac{1}{2}\gamma}\right).$$

By differentiating γ for the numerator of $f(\gamma)$ with common denominator, we can find $f(\gamma) > 0$ for $\gamma > 0$. Therefore, satisfaction of the upper bound is verified. On the other hand, to prove that (3.15) is larger than the lower bound in Lemma 3.7, we need to show that the following equation will be greater than zero for $\gamma > 0$.

$$g(\gamma) = \left(1 - \frac{1 - e^{-\frac{1}{2}\gamma}}{\frac{1}{2}\gamma}\right) - \left(1 - \frac{1 - 2Q(\sqrt{\frac{2}{3}\gamma})}{\sqrt{\frac{4}{3\pi}\gamma}}\right).$$

By using the inequalities in [73] for $Q(\sqrt{\frac{2}{3}\gamma}) < \sqrt{\frac{3}{4\pi}\gamma}e^{\frac{-\gamma}{3}}$ for $\gamma > 3$ and $(1 - 2Q(\sqrt{\frac{2}{3}\gamma}) \leq \sqrt{\frac{4\gamma}{3\pi}}e^{\frac{-\gamma}{9}}$ for $\gamma \leq 3$, we can verify $g(\gamma) > 0$ for $\gamma > 0$. Therefore, satisfaction of the lower bound is also verified. Based the analysis above, we prove that the approximation in (3.15) also follows the bounds in Lemma 3.7. □

Based on (3.8) and Proposition 3.8, we can now obtain the closed-form approximation of the expected AECOL of a message traversing from a source node i to the sink node s as follows:

$$\mathbb{E}[W_{i,s}] \approx \left(\frac{L_{i,s}}{D\left(1 - \frac{1-e^{-\frac{1}{2}\gamma}}{\frac{1}{2}\gamma}\right)} \right) \left(E_t + \frac{1}{2}\gamma(E_c - E_d) \right), \tag{3.17}$$

where $L_{i,s}$ is the separation between the source and sink nodes.

Before we wrap up the derivation for $\mathbb{E}[W_{i,s}]$, we show in Fig. 3-5 the tightness of the bounds and approximations. In particular, we show the numerical results of the expected forwarding distance in a time epoch. It can be observed that the closed-form expression satisfies the bounds and is a good approximation for a wide range of wakeup density p.

3.2.4 Power Consumption of the Network

After the closed-form expression of the expected AECOL for a message traversing from a source node to the sink node is obtained, we can now proceed to find the power consumption of the entire network. The network power consumption can be obtained by summing the power consumption of ECL and AECOL for all nodes in the network. Since ECL in a duty-cycle period depends only on the detection energy and the number of awake nodes in the network, it can be written as pAE_d, where A is the area of the network and p is the wakeup density p. Denoting the expected AECOL for a message averaged over all random source-to-sink pairs in the network as $\mathbb{E}[\mathbb{E}[W_{i,s}]]$, the network power consumption $\overline{P_N}$ can be represented by

$$\overline{P_N} = \frac{pAE_d}{T_d} + \frac{\mathbb{E}[\mathbb{E}[W_{i,s}]]}{T_m}, \tag{3.18}$$

where T_d is the duty-cycle period and T_m is the average message originated period. By using the result in (3.17), the expected AECOL for any message in the network

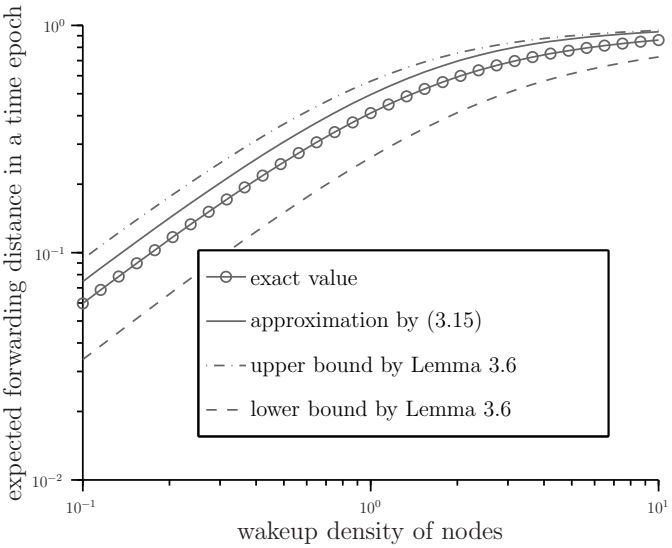


Figure 3-5: The expected forwarding distance in a time epoch as a function of the wakeup density of nodes with transmission range $D = 1$ and distance toward the sink node $L_{r_k,s} = 5$.

can be written as

$$\mathbb{E}[\mathbb{E}[W_{i,s}]] \approx \frac{\mathbb{E}[L_{i,s}]}{D}\left(1 - \frac{1 - e^{-\frac{1}{2}\gamma}}{\frac{1}{2}\gamma}\right)^{-1}\left(E_t + \frac{1}{2}\gamma(E_c - E_d)\right),$$

where $\mathbb{E}[L_{i,s}]$ represents the expected value of separation between node i and node s.

Note that from (3.8), $\mathbb{E}[W_{i,s}]$ can also be expressed in terms of $\mathbb{E}[H_{i,s}]$, the expected number of epochs for a message traversing from source node i to sink node s. Therefore, we can express the power consumption of the whole network in terms of the expected number of time epochs for any message as follows:

$$\overline{P_N} \approx \frac{pAE_d}{T_d} + \mathbb{E}[\mathbb{E}[H_{i,s}]]\frac{E_t + \frac{1}{2}\gamma(E_c - E_d)}{T_m}, \tag{3.19}$$

where $\mathbb{E}[\mathbb{E}[H_{i,s}]]$ is the expected number of time epochs for a message averaged over all source-to-sink pairs and it can be approximated as follows:

$$\mathbb{E}[\mathbb{E}[H_{i,s}]] \approx \frac{\mathbb{E}[L_{i,s}]}{D}\left(1 - \frac{1 - e^{-\frac{1}{2}\gamma}}{\frac{1}{2}\gamma}\right)^{-1}. \tag{3.20}$$

3.3 Energy-Delay Trade-offs in RWNs

We have so far expressed the power consumption of the whole network in terms of the energy consumed by each node participating in opportunistic relaying (i.e. E_t, E_c, and E_d) as well as several design parameters, including transmission range D, duty-cycle period T_d, and wakeup density p. In this section, we use the power consumption model thus derived to investigate the performance trade-offs between network energy consumption and message forwarding delay. We show that the ECL and the AECOL must be balanced carefully for minimizing network power consumption.

To proceed, we consider an optimization problem with a goal of minimizing the network power consumption under the constraint of the message delay requirement. Specifically, the objective function is

$$\min_{p,D,T_d} \left\{ \frac{pAE_d}{T_d} + \mathbb{E}[\mathbb{E}[H_{i,s}]] \frac{E_t + \frac{\pi}{2} pD^2(E_c - E_d)}{T_m} \right\}, \tag{3.21}$$

where wakeup density p, transmission range D, and duty-cycle period T_d are optimization variables. The delay constraint is

$$\frac{\mathbb{E}[\mathbb{E}[H_{i,s}]]T_d}{\mathbb{E}[L_{i,s}]} \leq T_1, \tag{3.22}$$

where T_1 is the maximal delay time for a message to advance over unit distance. We describe in the following how the optimization problem can be solved and the properties that can be observed in an optimized RWN.

3.3.1 Jointly Solving for Design Parameters

Under the delay requirement, it can be shown that the combination of (p^*, D^*, T_d^*) that minimizes the power consumption must satisfy $T_d^* = \mathbb{E}[L_{i,s}]T_1/\mathbb{E}[\mathbb{E}[H_{i,s}]]$. Substituting this equality into (3.21), we can express the network power consumption as

follows:

$$\begin{aligned}\overline{P_N} &\approx \mathbb{E}[\mathbb{E}[H_{i,s}]]\left(\frac{pAE_d}{\mathbb{E}[L_{i,s}]T_1}+\frac{E_t+\frac{\pi}{2}pD^2(E_c-E_d)}{T_m}\right)\\ &\approx \left(\frac{1}{T_m}\right)\times\left(\frac{\mathbb{E}[L_{i,s}]}{D}\left(1-\frac{1-e^{-\frac{\pi}{2}pD^2}}{\frac{\pi}{2}pD^2}\right)^{-1}\right)\\ &\times\left(\frac{pA}{\frac{\mathbb{E}[L_{i,s}]T_1}{T_m}}E_d+E_t^1D^\alpha+E_t^0+\frac{\pi}{2}pD^2(E_c^1D^\alpha+E_c^0-E_d)\right),\end{aligned} \tag{3.23}$$

where the product of the last two terms can be considered as the average energy spent in delivering a message: the second term is the average number of epochs to deliver the message and the last term is the average energy consumption for a message per epoch.

We note that the range-independent portion of the energy consumed by a transmitter E_t^0 is often small in the proportion of the entire network power consumption [48, 74]. In addition, the range-independent component of the energy consumed by a candidate forwarder E_c^0 includes the energy consumed for message detection E_d and is often small as far as the network power consumption is concerned [75]. In light of these, in the following analysis we omit E_t^0 and $(E_c^0 - E_d)$ in (3.23) for finding the optimal design parameters that can minimize network power consumption. We show the validity of the assumptions in Section 3.5 when these constant energy components are included.

To find the optimal values of of design parameters p^*, D^*, and T_d^*, we can solve for $\frac{\partial \overline{P_N}}{\partial p}=0$ and $\frac{\partial \overline{P_N}}{\partial D}=0$ by using the approximation expression of $\overline{P_N}$ in (3.23). It can be thus derived that the optimal values must satisfy the following two equations:

$$\begin{aligned}&(\alpha+2)\left((e^{-\frac{\gamma}{2}}+1)(\gamma)^2+4(e^{-\frac{\gamma}{2}}-1)\gamma\right)\frac{E_c^1}{E_t^1}\\ &+2\left((\alpha+2)e^{-\frac{\gamma}{2}}+\alpha-1\right)\gamma+4(2\alpha+1)e^{-\frac{\gamma}{2}}\\ &=8\alpha+4,\end{aligned} \tag{3.24}$$

$$D = \left(\frac{\pi}{2}(\frac{E_c^1}{E_t^1} + \frac{2}{\gamma})\frac{\alpha - 1}{3}\frac{\mathbb{E}[L_{i,s}]T_1}{T_m}\frac{1}{A}\frac{E_t^1}{E_d}\right)^{\frac{-1}{\alpha+2}}. \quad (3.25)$$

We propose the procedures to solve for the optimal values of p, D, and T_d as follows:

(i) The optimal value of γ is first solved in (3.24), where the path loss exponent α and the ratio between E_c^1 and E_t^1 are given.

(ii) The optimal transmission range D is then solved in (3.25) based on the optimal value of γ and other given parameters. Since $\gamma = \pi p D^2$, the optimal wakeup density p is also solved.

(iii) Finally, the optimal duration of the duty-cycle period T_d is solved in (3.22) to meet the delivery delay requirement.

3.3.2 Key Properties in Optimized RWNs

We explore in this section several important properties of an optimized RWN where the three design parameters (p, D, T_d) are set to the optimal solution for (3.21). These properties provide insights for trade-offs between network energy consumption and message forwarding delay in RWNs.

The first property regards the optimal number of nodes that should wake up within the transmission range of a relay node to participate in the relay activities.

Theorem 3.9. *The optimal expected number of nodes γ awakened within the transmission range in a time epoch to participate in message forwarding depends only on the path loss exponent α and the response-to-transmit energy ratio E_c^1/E_t^1. In addition, γ can be lower bounded by*

$$\gamma^* > \frac{1 - \alpha + \sqrt{(\alpha - 1)^2 + 12(\alpha + 2)\frac{E_c^1}{E_t^1}}}{(\alpha + 2)\frac{E_c^1}{E_t^1}}. \quad (3.26)$$

Proof. The property can be obtained by substituting the inequality of $e^{-\frac{1}{2}\gamma} < (1 + \frac{1}{2}\gamma + \frac{1}{8}\gamma^2)^{-1}$ into (3.24). □

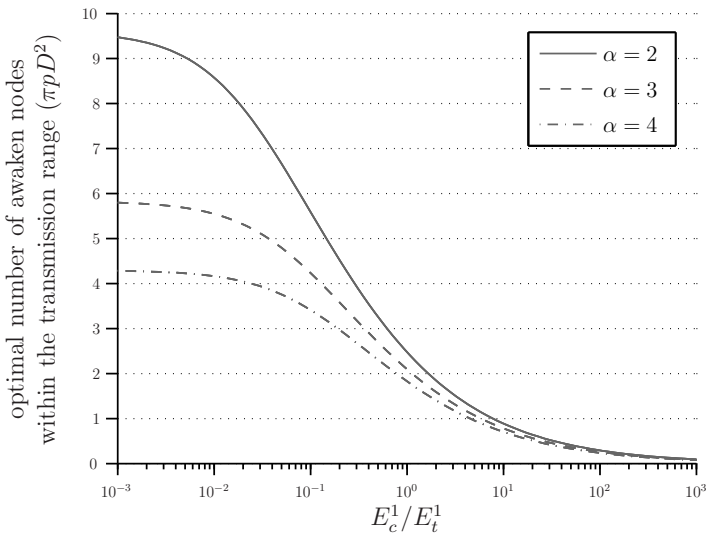


Figure 3-6: Optimal number of nodes awakened to forward messages in a time epoch within the transmission range as a function of response-to-transmit energy ratio E_c^1/E_t^1.

Surprisingly, the optimal number of nodes to wake up within the transmission range does not depend on factors such as the message origination rate $1/T_m$, the required delivery speed $1/T_1$, and the detection energy E_d. From (3.24), it can be observed that the optimal number of participating nodes decreases monotonically as the response-to-transmit energy ratio increases and as the path loss exponent increases. As shown in Fig. 3-6, for a typically region where $0 < E_c^1 \leq E_t^1$, *on average* at least one node shall wake up in the transmission range in a time epoch. It can also be observed that the number of nodes awakened to participate in the relay activities within the transmission range shall be *smaller than 10* for minimizing network power consumption in most network environments of different response-to-transmit energy ratios.

Note that γ is a function of p and D, and from (3.25) one can know $D^* \propto (E_t^1/E_d)^{\frac{-1}{\alpha+2}}$ and $p^* \propto (E_t^1/E_d)^{\frac{2}{\alpha+2}}$. In other words, the optimal transmission range is expected to rise as the ratio of E_t^1/E_d reduces. A corollary to this finding is that *always taking the shortest hop* or *persistent listening to potential incoming traffic* cannot be the optimal strategy for minimizing network energy consumption in RWNs, as long as the detection energy is nonzero. Another corollary that can be directly obtained is that *the optimal p can serve as a lower bound of the node density deployed in the*

network for minimizing network power consumption.

The second property regards the optimal proportion of battery energy that should be spent in detecting whether a message needs to be forwarded.

Theorem 3.10. *Under the optimal settings of p, D, and T_d, the ratio between ECL and AECOL is $\frac{\alpha-1}{3}$.*

Proof. This property can be obtained by substituting p^*, D^*, and T_d^* into (3.23). □

In this theorem, the optimal ratio depends only on the path loss exponent. The higher the path loss exponent, the less proportion of energy shall be spent in transmissions. Given that α is typically between 2 to 4, *the energy used for detecting messages is surprisingly of the same order of the energy used for transmitting and receiving messages.* What is equally surprising is that such an optimal proportion does not depend on any of the node energy parameters E_d, E_t^1, E_c^1 nor on the application parameters T_1 and T_m.

The third property regards the impact of the message delivery delay requirement and the message origination rate on the overall power consumption.

Theorem 3.11. *The power consumption under optimal setting is proportional to the $\frac{\alpha-1}{\alpha+2}$-th power of the required mean delivery speed and proportional to the $\frac{3}{\alpha+2}$-th power of the mean message origination rate.*

Proof. This property can be obtained by substituting p^*, D^*, and T_d^* into (4.7) such that

$$\overline{P_N} \approx f(\alpha, A, E_c^1, E_t^1, E_d)\mathbb{E}[L_{i,s}]^{\frac{3}{\alpha+2}} \left(\frac{1}{T_1}\right)^{\frac{\alpha-1}{\alpha+2}} \left(\frac{1}{T_m}\right)^{\frac{3}{\alpha+2}}, \tag{3.27}$$

where the function $f(\alpha, A, E_c^1, E_t^1, E_c)$ depends only on the environmental parameters and energy parameters. □

In this theorem, a rule of thumb for the optimal energy-delay trade-off in RWNs is that one needs to invest less than an α-fold increase in total power in order to increase the mean delivery speed by a factor of 10 when the path loss exponent is α. The investment of network energy thus can be quite effective for reducing the message delivery delay in an optimized RWN.

3.4 Generalizations and Extensions for n-Dimensional Networks

In the section, we extend our framework into n-Dimensional (n-D) networks generally. To keep our framework generally, we first derive the following close approximations for an n-D uniform random wake-up field. Then, we derive the optimal settings with different optimize-able parameters generally for 1-D, 2-D, and 3-D networks.

Proposition 3.13. *For an n-D uniform random wake-up sensor field, the expected forwarding distance and the average AECOL in a time epoch can be approximated by:*

$$\mathbb{E}[X_{i,s}^k] \approx D\left(1 - \frac{1 - e^{-p^n D^n}}{p^n D^n}\right), \tag{3.28}$$

$$\mathbb{E}[W_{i,s}^k] \approx E_t + p^n D^n (E_c - E_d) \tag{3.29}$$

Proof. We first consider a 1-D RWN. We compute the expected forwarding distance in a time epoch, $\mathbb{E}[X_{i,s}^k]$, and the corresponding average AECOL. In a 1-D uniform random wake-up sensor field, the distance a message gains during a sleep epoch is a random variable with a truncated exponential distribution:

$$P(X_{i,s}^k \le x) = \begin{cases} 1, & x > D, \\ e^{-p(D-x)}, & 0 < x \le D, \\ 0, & x < 0. \end{cases} \tag{3.30}$$

Note that there is a finite probability of e^{-pD} that a message does not advance to a new forwarding node at all during a sleep epoch. By (3.30), the expected forwarding distance experienced by a message during a sleep epoch, $\mathbb{E}[X_{i,s}^k]$, is:

$$\mathbb{E}[X_{i,s}^k] = D\left(1 - \frac{1 - e^{-pD}}{pD}\right). \tag{3.31}$$

If there exists only one message in the network, the average AECOL of the network

for epoch k, given that the message eventually advances a nonzero distance x, is

$$\mathbb{E}[W_{i,s}^k(x)] = E_t + pxE_c + E_c - pDE_d, \quad 0 < x \leq D. \tag{3.32}$$

On the other hand, when the message does not advance to any new forwarding node during this time epoch, the average AECOL is $E_t - pDE_d$.

In (3.32), for convenience, we reuse the notation $W_{i,s}^k(x)$ to represent the AECOL for a message advancing distance x at time epoch k in a random wakeup sensor field. Combining the conditional means from the analysis above, one can obtain the expression for the average AECOL in a time epoch:

$$\mathbb{E}[W_{i,s}^k] = E_t + pD(E_c - E_d). \tag{3.33}$$

In (3.33), the average AECOL in a time epoch is the sum of the transmission energy consumed by the forwarding node and the extra energy consumed for message detection by the average number of nodes awaken in the effective neighborhood.

We then derive bounds for the expected forwarding distance in a time epoch for a 2-D or a 3-D uniformly random wakeup field. First, consider a 2-D network. On one hand, one can obtain the maximal expected forwarding distance in a time epoch when the sink node is infinite far away from the current forwarding node. Shown in Figure 3-4-(b), the effective neighborhood in a 2-D RWN can be upper bounded by the rectangular region. A lower bound for the distribution function $F_X(x)$ can be obtained:

$$F_{X_{i,s}^k}(x) > e^{-2p^2D(D-x)}, \quad 0 \leq x < D. \tag{3.34}$$

Substituting equation (3.34) into equation (3.11), the expected forwarding distance per epoch in a 2-D random wakeup field can be upper bounded by:

$$\mathbb{E}[X_{i,s}^k] < D\left(1 - \frac{1 - e^{-2p^2D^2}}{2p^2D^2}\right). \tag{3.35}$$

On the other hand, the expected forwarding distance is minimal when the sink

node is just outside the transmission range of current forwarding node. Shown in Figure 3-4-(c), the effective neighborhood in a 2-D RWN can be lower bounded by the circular sector. An upper bound for the distribution function $F_{X_{i,s}^k}(x)$ is:

$$F_{X_{i,s}^k}(x) < e^{-\frac{\pi}{3}p^2(D-x)^2}, \quad 0 \leq x < D. \tag{3.36}$$

Substituting equation (3.36) into equation (3.11), the expected forwarding distance per epoch in a 2-D random wakeup field can be lower bounded by:

$$\mathbb{E}[X_{i,s}^k] > D - \frac{\sqrt{3}}{2p}\left(1 - 2Q(\sqrt{\frac{2\pi}{3}}pD)\right). \tag{3.37}$$

In equation (3.37), $Q(\cdot)$ represents the Q function.

Similarly, the expected forwarding distance in a 3-D random wakeup field can also be bounded. The expected forwarding distance per epoch in a 3-D random wakeup field can be upper bounded by

$$\mathbb{E}[X_{i,s}^k] < D\left(1 - \frac{1 - e^{-4p^3D^3}}{4p^3D^3}\right). \tag{3.38}$$

On the other hand, the expected forwarding distance per epoch in a 3-D random wakeup field can be lower bounded by:

$$\mathbb{E}[X_{i,s}^k] > D - \int_0^D e^{-\frac{\pi}{3}p^3(D-x)^3}dx = D - \sqrt[3]{\frac{1}{9\pi}}\frac{1}{p}\left(\Gamma(\frac{1}{3}) - \Gamma(\frac{1}{3}, \frac{\pi}{3}p^3D^3)\right). \tag{3.39}$$

In equation (3.39), $\Gamma(\cdot)$ represents the gamma function and $\Gamma(\cdot,\cdot)$ represents the upper incomplete gamma function.

Now let us consider the expected AECOL per epoch in a multi-dimensional network. If there exists only one message in the network, the average AECOL of the network, the expected AECOL in epoch k can be represented as:

$$\mathbb{E}[W_{i,s}^k] = P_0\mathbb{E}[W_k(0)] + \int_{0^+}^D f_X(x)\mathbb{E}[W_{i,s}^k(x)]dx \tag{3.40}$$

Note that P_0 represents the probability that a message does not find a new forwarding node in a time epoch and $f_{X_{i,s}^k}(x)$ represents the probability density function of a message advancing over length x in a time epoch.

In equation (3.40), the average AECOL of the network, given that the message eventually advances distance x at time epoch k, is

$$\mathbb{E}[W_{i,s}^k(x)] = E_t + p^n|\mathcal{N}_{i,s} - \mathcal{N}_{i,s}(x)|E_c + E_c - p^n|\mathcal{N}_{i,s}|E_d, \quad 0 < x \leq D.$$

Otherwise, if the message does not advance to any new forwarding node during this time epoch, the average AECOL is $\mathbb{E}[W_{i,s}^k(0)] = E_t - p^n|\mathcal{N}_{i,s}|E_d$.

Applying integration by parts for the second term of equation (3.40), a simple closed-form equation for the expected AECOL in epoch k can be obtained.

$$\mathbb{E}[W_{i,s}^k] = E_t + p^n|\mathcal{N}_{i,s}|(E_c - E_d). \tag{3.41}$$

Based on the results above, for the ease of analysis, we approximate the expected forwarding range per epoch in a multi-dimensional random wakeup field as

$$\mathbb{E}[X_{i,s}^k] \approx D\left(1 - \frac{1 - e^{-p^n D^n}}{p^n D^n}\right) \tag{3.42}$$

and the corresponding AECOL per epoch within the n-dimensional cube of side D as

$$\mathbb{E}[W_{i,s}^k] \approx E_t + p^n D^n (E_c - E_d). \tag{3.43}$$

□

Finally, the power consumption of the whole network can be approximated as:

$$\frac{p^n V E_d}{T_d} + \mathbb{E}[H]\frac{E_t + p^n D^n (E_c - E_d)}{T_m}. \tag{3.44}$$

In an n-D network, V represents the physical size of the network. It represents length,

area, and volume for 1-D, 2-D, and 3-D networks, respectively.

Now we derive the optimal values for the design parameters including the transmission range, the duty-cycle period, and the participation density to minimize the power consumption of the whole network. Specifically, we consider that source nodes are randomly present in the network to deliver messages to the sink node. Denote the expected number of time epochs for a message averaged over these source-to-sink pairs as $\mathbb{E}[\mathbb{E}[H_{i,s}]]$. The objective function of this network for optimization can be represented as:

$$\frac{p^n V E_d}{T_d} + \mathbb{E}[\mathbb{E}[H_{i,s}]] \frac{E_t + p^n D^n (E_c - E_d)}{T_m} \tag{3.45}$$

$$\approx \frac{p^n V E_d}{T_d} + \mathbb{E}[\mathbb{E}[H_{i,s}]] \frac{E_t + p^n D^n (E_c^1 D^\alpha)}{T_m} . \tag{3.46}$$

In the optimization problems, the environmental parameters, the node parameters, and the application parameters are assumed as given constants. The optimization is performed with a minimal velocity constraint. This constraint is formally defined as follows:

$$\frac{1}{T_1} \leq \frac{\mathbb{E}[L_{i,s}]}{\mathbb{E}[\mathbb{E}[H_{i,s}]] T_d} . \tag{3.47}$$

In equation (3.47), $L_{i,s}$ represents the separation between a source node and the sink node, and $\mathbb{E}[L_{i,s}]$ is its expectation.

We first examine the situation where all three design parameter p, D, and T_d can be jointly optimized in 3.4.1. We then investigate cases where one of the design parameters is given constant in 3.4.2–3.4.4.

3.4.1 Optimize-able Parameters are p, D, and T_d

Under the delay requirement, it can be shown that the combination of (p^*, D^*, T_d^*) that minimizing the power consumption must satisfy $T_d^* = \mathbb{E}[L_{i,s}]T_1 / \mathbb{E}[\mathbb{E}[H_{i,s}]]$. Substituting this equality into (3.46), one can express the average total energy consumption

rate as:

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[H_{i,s}]]\left(\frac{p^nVE_d}{\mathbb{E}[L_{i,s}]T_1}+\frac{E_t^1D^\alpha+p^nD^nE_c^1D^\alpha}{T_m}\right)\\
&\approx\left(\frac{1}{T_m}\right)\times\left(\frac{\mathbb{E}[L_{i,s}]}{D}\left(1-\frac{1-e^{-p^nD^n}}{p^nD^n}\right)^{-1}\right)\\
&\qquad\times\left(\frac{p^nV}{\frac{\mathbb{E}[L_{i,s}]T_1}{T_m}}E_d+E_t^1D^\alpha+p^nD^nE_c^1D^\alpha\right).
\end{aligned}\tag{3.48}$$

One can find that the average energy spent to deliver a message is the product of the last two terms in equation (3.48). The second term is the average number of hops to deliver the message and the last term is the average energy consumption for a message to advance one hop.

It can be shown that the optimal setting satisfies the following two equations:

$$\begin{aligned}
&(\alpha+n)\left((e^{-p^nD^n}+1)(p^nD^n)^2+2(e^{-p^nD^n}-1)p^nD^n\right)\frac{E_c^1}{E_t^1}\\
&+\left((\alpha+n)e^{-p^nD^n}+\alpha-1\right)p^nD^n+(2\alpha+n-1)e^{-p^nD^n}\\
&=2\alpha+n-1
\end{aligned}\tag{3.49}$$

and

$$D=\left((p^{-n}D^{-n}+\frac{E_c^1}{E_t^1})\frac{\alpha-1}{n+1}\frac{\mathbb{E}[L_{i,s}]T_1}{T_m}\frac{1}{V}\frac{E_t^1}{E_d}\right)^{\frac{-1}{\alpha+n}}.\tag{3.50}$$

In (3.49), note that the pD-product stands for the expected number of nodes awaken in a time epoch within the transmission range for message relay per dimension. As it turns out, it is a critical quantity for the optimized random wakeup relay network. From (3.49), we observe the following important property.

Proposition 3.14. *The optimal expected number of nodes awaken in a time epoch to participate in message forwarding in the transmission range depends only on the path loss exponent, the network dimensionality, and the response-to-transmission energy ratio.*

It is quite surprising to note that this optimal quantity does not depend on factors

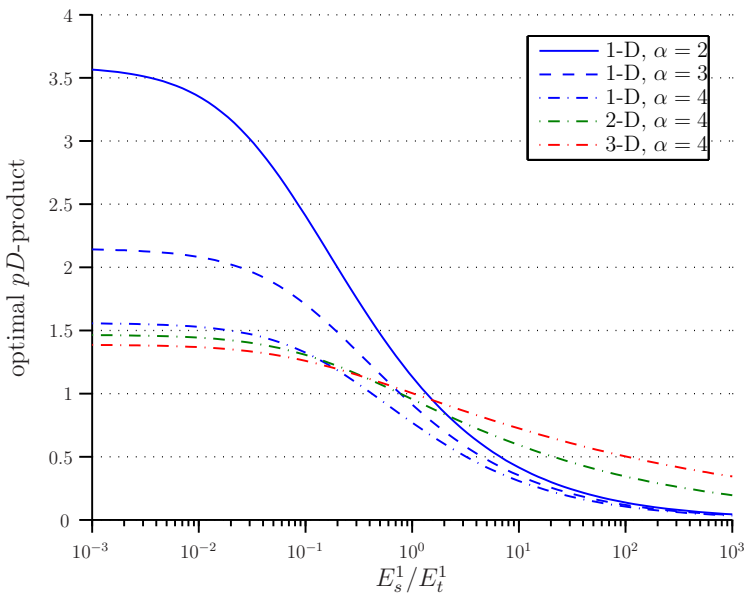


Figure 3-7: The optimal number of nodes awaken to forward messages in a time epoch within the transmission range per dimension as a function of response-to-transmission energy ratio E_c^1/E_t^1.

such as the message origination rate, the required delivery speed, and the message detection energy.

From (3.49), it can be derived that the optimal number of participating nodes decreases monotonically as the *response-to-transmission* energy ratio increases. Also, the optimal number of participating nodes decreases as the path loss exponent increases. The optimal number of participating nodes within the transmission range increases as the network dimensionality increases.

The optimal pD-product for the overall power consumption with the required delivery speed satisfied is shown in Figure 3-7. For a typically regime where $0 < E_c^1 \leq E_t^1$, *on average* at least one node shall wake up in the transmission range in a time epoch per dimension. Under the optimal settings in this typical regime, a message advances some distance toward the sink node in a time epoch with high probability.

Substituting the optimal pD-product into (3.50), one can determine the values of p^* and D^*. We note specifically that $D^* \propto (E_t^1/E_d)^{\frac{-1}{\alpha+n}}$ and $p^* \propto (E_t^1/E_d)^{\frac{1}{\alpha+n}}$. This result is quite intuitive; the range is expected to rise as the ratio of transmission energy

per unit length to detection energy reduces. A corollary to this finding is that *always taking the shortest hop* or *persistent listening to potential incoming traffic* cannot be the optimal strategy for minimum energy routing in WRNs, as long as the message detection energy is nonzero.

Let us summarize how the optimal values of p, D, and T_d are derived:

- The pD-product shall first be matched to the response-to-transmission energy ratio.
- The range D (and hence p) shall then be chosen to reflect the ratio of unit-length transmission energy to detection energy.
- Finally, T_d is chosen to meet the delivery speed requirement.

Armed with the optimal setting of the three design parameters, we can investigate the properties of an optimized network. The first property we seek to understand is the right proportion of battery energy to be spent on (mostly fruitlessly) detecting that a neighbor node is requesting a message to be forwarded. The answers for the first question is quite clear!

Proposition 3.15. *Under the optimal settings of p, D, and T_d, the ratio between ECL and AECOL is $\frac{\alpha-1}{n+1}$.*

In the proposition, the optimal ratio depends only on the path loss exponent and the network dimensionality. The higher the path loss exponent, the less proportion of energy shall be spent on transmission. Given that α is typically between 2 to 4, the total detection power is at the same order of total transmission power. This optimal proportion can be quite a lot higher than conventional thinking. What may be more surprising is that such optimal proportion does not depend on any of the node energy parameters E_d, E_t^1, E_c^1 nor on the application parameters T_1 and T_m.

The second property we seek to understand is how do factors such as the message delivery speed requirement and the message origination rate impact the overall power consumption? Note that the overall power consumption under the optimal setting

can be found by substituting p^*, D^*, and T_d^* into equation (3.48). The minimal power consumption of the whole network is

$$f(\alpha, n, V, E_c^1, E_t^1, E_d)\mathbb{E}[L_{i,s}]^{\frac{n+1}{\alpha+n}} \left(\frac{1}{T_1}\right)^{\frac{\alpha-1}{\alpha+n}} \left(\frac{1}{T_m}\right)^{\frac{n+1}{\alpha+n}}. \quad (3.51)$$

In (3.51), the function $f(\alpha, n, V, E_c^1, E_t^1, E_d)$ depends only on the environmental parameters and the node parameters. From (3.51), we can observe the following property.

Proposition 3.16. *The power consumption under optimal setting is exactly proportional to the $\frac{\alpha-1}{\alpha+n}$-th power of required mean delivery speed and is exactly proportional to the $\frac{n+1}{\alpha+n}$-th power of mean message origination rate.*

From the proposition, a rule of thumb for the optimal power-delay trade-off in a 1-D network is that one needs only to invest a 2-, 3-, or 4-fold in total power in order to increase the mean delivery speed by a factor of 10 given that the path loss exponent is 2, 3, or 4, respectively. Another observation is that, interestingly, the higher the network dimensionality is, the less sensitive the network power is to the velocity requirement.

3.4.2 Optimize-able Parameters are D and T_d

The problem here is to find the combination of range D and sleep period T_d to minimize the total power consumption while the participation density p^n is given. Under the delay requirement, the optimal duty-cycle period T_d^* shall satisfy $T_d^* = \mathbb{E}[L_{i,s}]T_1/\mathbb{E}[\mathbb{E}[H_{i,s}]]$. Substituting the equality to equation (3.46), the objective function for optimization can be represented the same as equation (3.48). The optimal transmission range D^* can be solved by equation (3.48), and the T_d^* can be obtained by substituting D^* into $T_d^* = \mathbb{E}[L_{i,s}]T_1/\mathbb{E}[\mathbb{E}[H_{i,s}]]$.

There is no simple closed-form solution for D^*. Here we derive close approximations for the two regimes. In one, the energy to transmit a message dominates the energy to negotiate the next forwarding node; in the other, the opposite is considered.

When the transmission energy dominates the expected total response energy, the optimal transmission range can be approximated as:

$$\left(\frac{\alpha-1}{p^n}\frac{\mathbb{E}[L_{i,s}]T_1}{T_m}\frac{1}{V}\frac{E_t^1}{E_d}\right)^{\frac{-1}{\alpha}}. \tag{3.52}$$

On the other hand, if the expected total response energy dominates, the optimal transmission range D^* can be well approximated by:

$$\left((\alpha+n-1)\frac{\mathbb{E}[L_{i,s}]T_1}{T_m}\frac{1}{V}\frac{E_c^1}{E_d}\right)^{\frac{-1}{\alpha+n}}. \tag{3.53}$$

In this regime, the D^* does not depend on the given participation density.

Under the optimal setting obtained above, we examine the proportion of network power spent in the reception mode. In the regime that the transmission energy dominates, the ratio between ECL and AECOL is $\alpha g - 1$. On the other hand, in the regime that the response energy dominates, the ratio between ECL and AECOL is $(\alpha+n)g-1$. The ratio between ECL and AECOL can be bounded by the factor g, $\frac{1}{n+1} < g < 1$. The factor g approaches to 1 monotonically, as the optimal transmission range increases.

In the regime that the transmission energy dominates, the minimal power consumption is proportional to the $\frac{\alpha-1}{\alpha+n}$-th order of the mean delivery speed and the $\frac{n+1}{\alpha+n}$-th order of the mean message origination rate. On the other hand, in the regime that the response energy dominates, the minimal power consumption is proportional to the $\frac{\alpha+n-1}{\alpha+n}$-th order of the mean delivery speed and the $\frac{1}{\alpha+n}$-th order of the mean message origination rate.

3.4.3 Optimize-able Parameters are p and T_d

The problem here is to minimize power consumption using the duty-cycle period T_d and the participation density p^n while the transmission range D is given.

We first define $\beta \equiv E_t D^n / (\frac{T_m}{\mathbb{E}[L_{i,s}]T_1} V E_d + (E_c - E_d)D^n)$. The parameter β is the key engineering parameter for this optimization problem.

Again it can be shown that the optimal duty-cycle period T_d^* must satisfy $T_d^* = \mathbb{E}[L_{i,s}]T_1/\mathbb{E}[\mathbb{E}[H_{i,s}]]$. Substituting this equality into (3.45), one can approximate the average total energy consumption rate as:

$$\frac{\mathbb{E}[L_{i,s}]}{T_m}\frac{\frac{T_m}{\mathbb{E}[L_{i,s}]T_1}VE_d + (E_c - E_d)D^n}{D^{n+1}}\left(1 - \frac{1 - e^{-p^n D^n}}{p^n D^n}\right)^{-1}(p^n D^n + \beta).$$

In the equation, only the last two terms depend on the optimize-able parameter p^n. The optimal participation density $(p^*)^n$ can be solved from the equation above. Once p^* is obtained, T_d^* can be obtained by $T_d^* = \mathbb{E}[L_{i,s}]T_1/\mathbb{E}[\mathbb{E}[H_{i,s}]]$.

Unfortunately, there is no simple closed-form expression for p^*. Instead, we find close approximations. The optimal number of nodes awaken within the given n-dimensional cube D^n is upper bounded by:

$$(p^*D)^n < 1 + \sqrt{1+\beta}. \tag{3.54}$$

The above upper bound is a very good approximation when $\beta > 5$. When $\beta \leq 5$, the participation density can be well approximated by

$$(p^*)^n \approx (p_1^*)^n\sqrt{\beta}. \tag{3.55}$$

where $(p_1^*)^n \approx 1.5/D^n$ is the optimal participation density corresponding to $\beta = 1$. The optimal participation density, its upper bound in (3.54), and its approximation in (3.55), are shown in Figure 3-8.

When the optimal setting is applied, the proportion of power consumption spent for message detection can also be characterized below. In the regime $\beta > 5$, the proportion of overall network power spent for detection is upper bounded by:

$$\frac{1 + 1/\sqrt{1+\beta}}{1 + \sqrt{1+\beta}}. \tag{3.56}$$

On the other hand, in the regime $\beta \leq 5$, the proportion of network power spent for

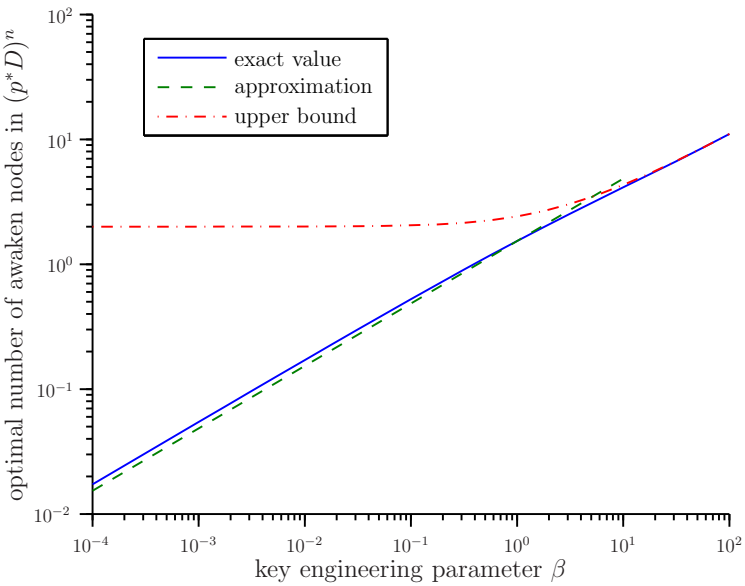


Figure 3-8: The optimal number of nodes awaken to participate the relay activities in a time epoch within the given n-dimensional cube D^n as a function of the key engineering parameter $\beta \equiv E_t D^n / (\frac{T_m}{\mathbb{E}[L_{i,s}]T_1} V E_d + E_c D^n)$.

detection is upper bounded by:

$$\frac{1.5}{1.5 + \sqrt{\beta}}. \tag{3.57}$$

In (3.56) and (3.57), if the message detection power dominates the response power, the two bounds are very good approximations.

In a WRN where the network size is much larger than the range of single hop transmission, the parameter β is usually much smaller than one. In this regime, the minimal power consumption almost grows linearly as a function of the mean delivery speed. Interestingly, as a contrast to the previous cases, as long as β is much smaller than one, the minimal power consumption hardly depends on the message origination rate.

3.4.4 Optimize-able Parameters are p and D

Here the problem is to minimize power consumption by adjusting D and p^n while the duty-cycle period T_d is a predetermined factor.

Under the delay requirement, the optimal participation density $(p^*)^n$ and the optimal transmission range D^* should satisfy the the expected number of hops for

a message from a source node to the sink node $\mathbb{E}[\mathbb{E}[H_{i,s}]] \leq \mathbb{E}[L_{i,s}]T_1/T_d$. With the approximation of $\mathbb{E}[H_{i,s}] \approx L_{i,s}/\mathbb{E}[X^k_{i,s}]$, the expected forwarding distance in a time epoch for a message shall satisfy $\mathbb{E}[X^k_{i,s}] \geq T_d/T_1$ and then the power consumption of the whole network can be represented as

$$\frac{p^n V E_d}{T_d} + \frac{\mathbb{E}[L_{i,s}]}{\mathbb{E}[X^k_{i,s}]} \frac{E^1_t D^\alpha + p^n D^n E^1_c D^\alpha}{T_m} . \tag{3.58}$$

Approximating $\mathbb{E}[X^k_{i,s}]$ by equation (3.28), for path loss exponent $\alpha \geq n+1$, it can be shown that the combination of (p^*, D^*) which minimizes power consumption in (3.58) satisfies $\mathbb{E}[X^k_{i,s}] = T_d/T_1$.

Although there is no simple closed-form solution for p^* and D^*, we indeed can derive the optimal pD-product for power consumption minimization with a given T_d. The p^* and D^* can be derived individually with the obtained optimal pD-product.

Consider the regime that the expected forwarding distance per epoch is much larger than unit length. If the detection power consumption dominates the response power consumption in the whole network, the $(p^*D^*)^n$ can be approximated to

$$\sqrt{\alpha \frac{\mathbb{E}[L_{i,s}]T_1}{T_m} \frac{1}{V} \frac{E^1_t}{E_d} (\frac{T_d}{T_1})^{\alpha+n}} . \tag{3.59}$$

On the other hand, when the response power consumption dominates, the $(p^*D^*)^n$ can be approximated to

$$\frac{\alpha+1}{2} + \sqrt{\left(\frac{\alpha+1}{2}\right)^2 + \alpha \frac{E^1_t}{E^1_c}} . \tag{3.60}$$

In (3.60), the optimal number of awaken nodes in the transmission range depends only on the path loss exponent and the response-to-transmission energy ratio. Also, interestingly, this quantity is lower bounded by the path loss exponent α.

Now let's turn our attention to the regime where the expected forwarding distance is much smaller than unit length. If the transmission power consumption dominates

the response power consumption, the $(p^*D^*)^n$ can be approximated as

$$2\left(\frac{\alpha}{n+1}\frac{\mathbb{E}[L_{i,s}]T_1}{T_m}\frac{1}{V}\frac{E_t^1}{E_d}(\frac{T_d}{T_1})^{\alpha+n}\right)^{\frac{1}{\alpha+n+1}}. \tag{3.61}$$

On the other hand, if the response power consumption dominates, the $(p^*D^*)^n$ can be approximated by

$$2\frac{T_d}{T_1}\left(\frac{\alpha-1}{n+1}\frac{\mathbb{E}[L_{i,s}]T_1}{T_m}\frac{1}{V}\frac{E_c^1}{E_d}\right)^{\frac{1}{\alpha+n}}. \tag{3.62}$$

In (3.62), the optimal number of nodes participating the relay activities is proportional to the given expected forwarding distance in a time epoch.

With the optimal setting applied, the ratio between ECL and AECOL of the whole network can be obtained. In the regime that the expected forwarding distance is much larger than unit length, if the detection power dominates, the ratio is $\alpha(p^*D^*)^{-n}$; if the response power dominates, the ratio obviously approaches zero. On the other hand, in the regime that the expected forwarding distance is much smaller that unit length, if the message transmission power dominates, the ratio is $\frac{2\alpha}{n+1}$; if the response power dominates, the ratio is $\frac{\alpha-1}{n+1}$.

3.5 Case Studies

To substantiate the random wakeup framework and to validate the closed-form approximations derived in this chapter, we consider a wireless relay network with the sleep mode operation and simulate the network performance under varying network parameters. Through comparing analytical and simulation results, we show that the proposed framework indeed can be used to model the network performance and predict the optimal parameters with good accuracy.

3.5.1 Geographical Routing with Sleep Mode Operation

We consider a network that uses a geographic routing protocol for message relay, where the positions of neighbors around a forwarding node are used to help determine

the next forwarding node [76–78]. Specifically, we implement the Greedy Perimeter Stateless Routing (GPSR) [77] for our simulation. In GPSR, normally the neighbor within the broadcasting range that is closest to the sink node becomes the next forwarding node. When the sink node is unreachable using this greedy criterion, perimeter routing is employed.

The original GPSR paper does not include sleep mode operation, so we add the random sleep-awake mechanism in Section 3.1 to it. If a node has a message to transmit, it wakes up in every epoch to attempt to send the message to some next forwarding node. On the other hand, if a node does not have any message to transmit, it wakes up to listen for messages with probability q. A message under such opportunistic relay thus can advance either zero or one hop per epoch. The goal of a network designer is to determine the optimal values for wakeup probability q, transmission range D, and duty-cycle period T_d. Note that the wakeup density p defined in Section 3.1 can be written as the geographic node density multiplied by the wakeup probability q.

Simulation Setup

To compare the analytical results obtained in this chapter, we implement GPSR and IEEE 802.11 MAC using C++. We simulate a network with an area of 600 meter by 3000 meter, where 200 nodes are randomly deployed in the network. Messages are originated from 30 randomly selected nodes in the network to be delivered to the sink node located at the center of the network. We consider a network of low traffic density, and the average message origination period T_m is set to 1 second. Since collisions of messages rarely happen at such a low traffic density, all transmitted data packets or control packets thus can be assumed to be successfully received within the transmission range in our MAC-layer implementation.

For the power profile, we adopt the settings of MAXIM MAX2830 in the simulation. Specifically, MAX2830 is based on 802.11 and it consumes 636 mW in transmit mode, 186 mW in receive mode, and 60 µW in sleep mode with a 3V power supply. The data rate is set to 1Mbps, the data frame size is 125 bytes, and the ACK frame

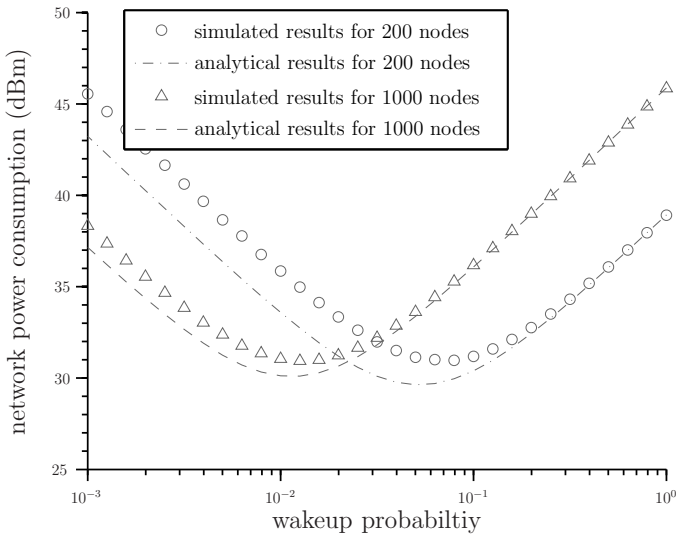


Figure 3-9: Network power consumption and delivery delay as a function of the wakeup probability of a node. Both the simulation results and analytical results are shown.

size is 14 bytes. The listening interval of a node is set to 2 ms, while the range-independent energy consumption for a node to receive messages $E_d = E_c^0 = 372$ µJ and the range-independent energy consumption for transmitter $E_t^0 = 20.83$ µJ. For a given transmission range $D = 100$ meter, the unit-length transmit energy E_t^1 and unit-length response energy E_c^1 for path loss $\alpha = 4$ are 6.36 pJ/ m^4 and 0.71 pJ/ m^4, respectively. All simulation results shown are averaged over 20 networks of random topologies.

Network Power Consumption

To show the validity of the analytical model proposed in this chapter, we obtain the total power consumption in the network through simulations and compare it against the analytical results obtained by using using (3.19). We simulate two network sizes of 200 and 1000 nodes; the transmission range D is set to 250 meter and the duty-cycle period T_d is set to 10 milliseconds. The wakeup probability q is varied for different simulations.

Fig. 3-9 shows the network power consumption as a function of the wakeup probability of nodes. From the figure, it can be observed that the analytical result reasonably models the performance of a GPSR-based wireless relay network with sleep

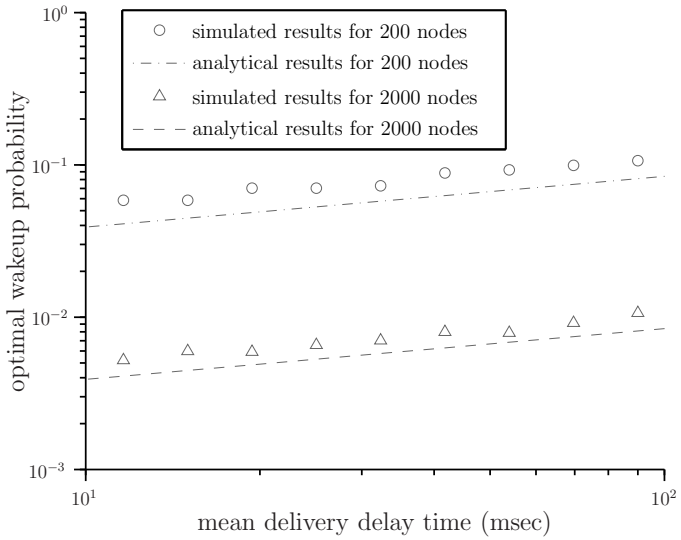


Figure 3-10: Optimal wakeup probability for minimizing network power consumption as a function of mean delivery delay time. Simulation results as well as analytical results are shown.

mode operation, especially when the node density is high. This result is intuitive since we employ the sensor field to derive network power consumption in an RWN. In addition, the result substantiates the approximations made in this chapter for deriving the closed-form expression of the expected AECOL for message relay. It can be observed from the figure that there does exist an optimal value of the wakeup probability that can minimize the network power consumption.

Optimization of Network Design Parameters

Different applications may have different delay time requirements for a relay network with sleep mode operation. Therefore, we vary the requirement of the message delay time T_1 introduced in Section 4.3 and find optimized network design parameters for different scenarios. In particular, we focus on the optimal wakeup density that can minimize the total network power consumption under the delay requirement. For the analytical model, the optimal wakeup density is obtained through (3.24) and (3.25); for network simulation, on the other hand, the optimal wakeup density is found by exhaustively searching for the value (used as the simulation parameter) that can minimize the power consumption of the simulated network. Fig. 3-10 thus shows

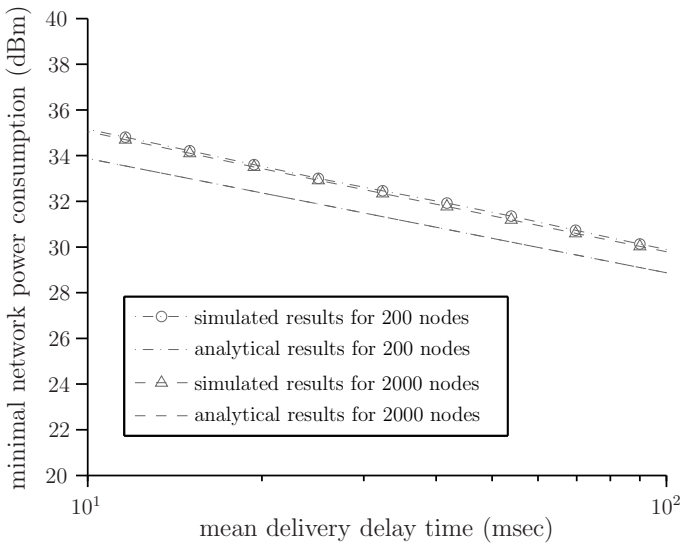


Figure 3-11: Minimal power consumption under the jointly optimal values of the wakeup probability, transmission range, and sleep period as a function of the mean delivery delay constraint.

the optimal wakeup probability as a function of mean message delay time. It can be observed that the case with a network of 2000 nodes shows better match compared to the case of 200 nodes due to our assumption of the sensor field. Nonetheless, the analytical model reasonably predicts the optimal values of the wakeup probability obtained through simulations, even for the case of 200 nodes.

To further investigate the performance trade-offs between network power consumption and message forwarding delay, Fig. 3-11 shows the minimal network power consumption as a function of the required message delay time under optimal design parameters (transmission range, duty-cycle period, and wakeup probability). While the results for 2000 nodes show better match than the results for 200 nodes, it can be observed that there is approximately only 1dB performance loss for the case of 200 nodes. Another interesting result that can be observed from the figure is that the optimal trade-offs between network power consumption and the message delay time follows Theorem 4.3 where the power consumption decreases at a rate of the $\frac{1}{2}$-th order of message delivery delay for path loss exponent $\alpha = 4$. This result further substantiates the validity of the analytical model proposed in this chapter.

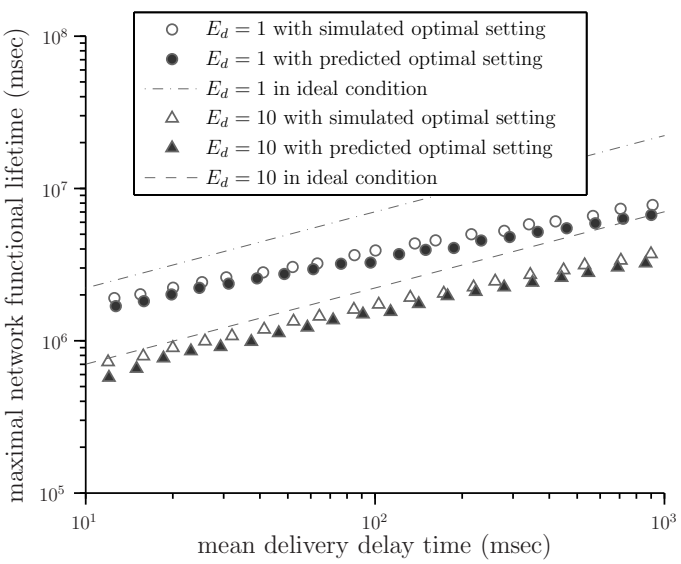


Figure 3-12: Optimal value of the wakeup probability and the maximal network lifetime as a function of mean delivery delay time. Energy parameters E_t and E_c are normalized to 1 for transmission range $D = 100$ m. Initial battery energy of a node is set to 10^5.

Maximization of Network Lifetime

Finally, we study the impact of optimal network parameters on maximization of the lifetime of a random wakeup network. The network lifetime is defined as the time until an originated message can not be delivered to the sink node, meaning that the network is unfunctional. We assume that small data packets are generated in the network where a wide range of M2M applications belongs to [50]. In the following simulation, the data size is set to the size of acknowledgements and $E_t^1 = E_c^1$.

To find the maximal network lifetime under different delay requirements by using network simulation, we vary the values of the design parameters (i.e. transmission range, duty-cycle period, and wakeup probability) as different simulation settings and exhaustively search for the optimal values. The result is labeled as "simulated optimal setting" in Fig. 3-12. In comparison, we also find the optimal values of the design parameters based on (3.22), (3.24), and (3.25) and use the optimal setting for simulating the GPSR-based relay network. The network lifetime thus obtained is labeled as "predicted optimal setting" in Fig. 3-12. From the figure, we can observe that there is only a small performance loss for maximizing network lifetime when the network uses the predicted optimal settings as the operational parameters. This is

true for different energy profiles of nodes.

Fig. 3-12 also shows the ideal network lifetime obtained as the total energy divided by the minimal network power consumption in (3.27). While the network lifetime thus obtained is not a practical value in the sense that it ignores the inherent non-uniformity of node locations and routing paths. Nonetheless, it serves as an upper bound that shows how *balanced* the underlying routing mechanism is. For nodes with high wakeup probability, GPSR tends to select the same relay node more often such that the network lifetime is constrained by the nodes in the hot-spot area. Therefore, the actual network lifetime deviates more from the ideal value. On the other hand, with the inherent node diversity for route selection in the low wakeup probability region, the actual network lifetime can approach the ideal network lifetime.

3.5.2 A Cluster-Based Network with Sleep Mode Operation

Last but not the least, we consider a cluster-based network. Cluster-based routing protocols have been studied in great detail; e.g., [79–82]. Generally, in a cluster-based network, nodes in a local area elect a "community center". Leaf nodes forward data to the center nodes. A center node summarizes its local data and forms a condensed packet to forward on. Conversely, a center node is also responsible for distributing remote messages to its leaf nodes.

Here we consider a cluster-based network whose structure was determined using the Energy-aware Virtual Backbone Tree (EVBT) [82] algorithm. The EVBT algorithm is proposed under the following node energy consumption model. Given a radio transmission range d, the energy spent to transmit a message, E_{tx}, and the energy spent to positively receive a message, E_{rx}, are the following, respectively:

$$E_{tx} = \alpha_{11} + \alpha_2 d^2,$$
$$E_{rx} = \alpha_{12}.$$

In [82], nodes are classified into tree nodes and leaf nodes. A leaf node associates itself with a tree node closest to it. A message originated from a leaf node is first

Table 3.2: The node energy profile for EVBT simulation.

Cause	Symbol	Expression
Energy to transmit a packet	E_{tx}	$\alpha_{11} + \alpha_2 d^2$
Energy for negative detection	E_{mx}	α_3
Energy for positive detection	E_{rx}	α_{12}

Table 3.3: The node energy profile for the EVBT-approximate RWN simulation.

Cause	Symbol	Expression
Energy to transmit a packet	E_t	$\alpha_2 d^2$
Energy for negative detection	E_r	α_3
Energy for positive detection	E_s	0

transmitted to the corresponding tree node. Then the message is progressively relayed among tree nodes until it reaches the root, which is the sink node.

The originally proposed EVBT network does not include a sleep mode operation, so we add the following sleep mechanism to it. In the network, the wakeup period is T_d. Tree nodes wake up periodically to detect incoming traffic either from a leaf node or from a neighboring tree node. If a leaf node has an original message to send, it wakes up during a sleep epoch to send the message to its tree node. Otherwise, a leaf node never wakes up. A message can advance one hop per wakeup period. The goal for a network designer is to pick the optimal values for the radio transmission range d, the number of tree nodes, and the wakeup period T_d.

For this network, an analogous RWN has the following parameters:

- transmission range d
- duty-cycle period T_d
- the wakeup density equals the geographical density of tree nodes

Under this transformation, the original optimization problem corresponds to the case where p, D, and T_d are all optimize-able in the analogous RWN.

We simulate a network which, except for the sleep behavior, is identical to the one in [82]. Specifically, we simulate a network with 10,000 nodes uniformly placed in a 600 meter-by-600 meter area. Every node originates a message to transport to

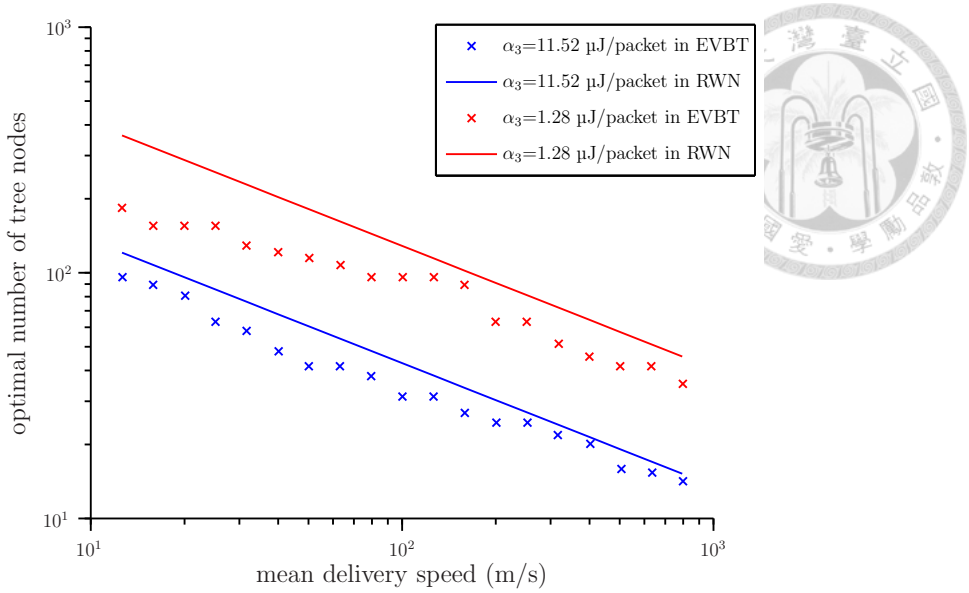


Figure 3-13: The optimal number of tree nodes as a function of mean delivery speed in the EVBT network with sleep mechanism adopted. $\alpha_{11} = 12.8$ µJ/packet. $\alpha_{12} = 12.8$ µJ/packet. $\alpha_2 = 16$nJ/packet/ m^2.

the sink node at the center of the network. The average message origination period is 1 second. The node parameters for both the EVBT network and the analogous RWN are shown in Table 3.2 and Table 3.3 , respectively. Please note that, in order to render this optimization problem meaningful, a small positive energy is consumed when a tree node detects no messages.

The optimal number of tree nodes that minimizes network energy consumption as a function of mean delivery speed is shown in Figure 3-13. For comparison purposes, the counterpart for the analogous RWN, assuming that $\alpha_{11} = 0$ and $\alpha_{12} = 0$, is shown as well. From the figure, we can observe that, despite the mismatch in node power consumption model, the RWN predicts well the optimal settings for this cluster-based WRN.

3.6 Summary

In this chapter, we have investigated the trade-offs between network energy consumption and message forwarding delay in random wakeup wireless networks. We have proposed the random wakeup network with opportunistic relay as a framework in our

analysis. We first derived the analytical model for network power consumption in the random wakeup network, and then we apply the model in an optimization problem to find the optimal design parameters (transmission range, duty-cycle period, and wakeup density) that can minimize the network power consumption while meeting the message delay requirement. Finally, we presented a case study to show that the proposed framework and analysis reasonably models the power consumption behavior of wireless relay networks with random sleep-awake schedule.

Several interesting properties that we have found in this chapter include: (1) The optimal number of nodes participating in the relay activities in a duty-cycle period within the transmission range is not a function of the allowable delivery delay nor of how often a message is originated. Instead, it depends only on the path loss exponent and the response-to-transmit energy ratio. (2) When the optimal routing and sleep parameters are applied, the proportion of power consumption for detecting potential incoming messages in the whole network is $\frac{\alpha-1}{\alpha+2}$, indicating that the power consumed for detection should be of the same order as the power spent in transmissions. (3) The minimal network power consumption grows at the $\frac{\alpha-1}{\alpha+2}$-th power of the factor by which the minimal required delivery speed increases. That is, to reduce the message delivery delay in RWNs, the investment of network energy can be quite efficient.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Sleep Mode Operations for Information Exchange and Information Dissemination in Green Wireless Multi-Hop Relay Networks

A wireless relay network (WRN) [83] is an aggregation of nodes with the ability to transmit and receive wirelessly. To serve the underlying application, nodes may wish to communicate with each other. Whenever necessary, nodes serve as relays to transport messages for others.

For a WRN with battery-powered nodes, the ideal use of power is clearly of great significance. In many sensing applications, the node energy used for sensing is minuscule compared to that used for wireless communication [48]. As a result, the way nodes communicate with each other shall be carefully designed.

Communication between nodes requires three actions: transmission, detection, and reception. Traditionally, network power consumption is optimized to minimize the total transmission energy. However, the reception mode power which we define

as the sum of energy to detect messages and receive such messages may easily rival the transmission energy [84], especially in low traffic density WRNs.

To reduce the energy consumption in reception mode, a popular approach is to switch the node into sleep mode in which a node shuts off its radio transceiver to disengage itself from the network for a period of time [80]. Compared with the power consumption in transmission mode or reception mode, the power consumption in sleep mode is often negligible.

In general, communication paradigms for energy-efficient WRNs can be characterized into: one-to-one [85], many-to-one [86, 87], one-to-many [88, 89], and many-to-many [90]. The optimization of power consumption often highly depends on these paradigms which the underlying network applications belong to. In these paradigms, information exchange in which many nodes wish to exchange its own message with other nodes in the network is essential for various WRN applications, e.g., consensus achievement [91]. Such a communication pattern often depletes a large amount of network energy, however, it is rarely investigated.

In this chapter, we investigate the optimization of power consumption minimization for information exchange which is a dual problem for message delivery in Chapter 3. We desire to seek for general guidelines in the design of energy-efficient WRNs. In the optimization problems, the power profile of nodes, the path loss exponent, and the message origination period are given and the position of nodes are determined to meet the sensing needs. Three key parameters are under our control: the transmission range, the duty-cycle period, and the participation density. The optimization requirement is that all interested messages shall be exchanged in a targeted region within the allowable delay time.

We propose a "random gossip" wireless-relay network (RGN) as a framework for our analysis. In the RGN, we consider that each node initiates a message and wishes to exchange its message with all the other nodes in the network. We will demonstrate that our results in the RGN can indeed predict the optimal value of operational parameters for information exchange as well as information dissemination in actual WRNs.

We find a number of properties for the three design parameters in the optimized RGNs. First, a key relationship is that the optimal number of nodes particiating to broadcast messages within the transmission range in a duty-cycle period depends only on the path loss exponent and the network dimensionality. It does not depends on factors such as the physical network size, the detection energy, and the message origination rate. Next, the existing optimal number of broadcasting nodes in a time epoch within the transmission range shows that either shortest hop transmission or blind flooding is rarely optimal for power consumption minimization. Last but not the least, it is shown that neither increasing nor decreasing the physical network size will affect the optimal value of these design parameters. The optimal setting for power consumption minimization is a scalable solution.

The rest of this chapter is organized as follows. In Section 4.1, we present the random gossip framework with periodic listening. In Section 4.2, we analyze the power consumption in the random gossip WRN. In Section 4.3, we derive optimal value of the design parameters for minimizing overall network power consumption. In section 4.4, we apply our results for information exchange as well as information dissemination in proposed WRNs. Summary of the chapter is given in Section 4.5.

4.1 Framework for Power Consumption Analysis

4.1.1 Random Gossip WRN

In the random gossip network, all nodes are synchronized by a common clock. Time is partitioned into contiguous same-sized epochs of duration T_d. This duration is called the *duty-cycle period*. Nodes carry out communication at the start of a duty-cycle period. For the rest of the period, the transceivers are shutdown to save power.

In the RGN, the action that a node exchanges its own message within its neighborhood is defined as *gossip*. In a duty-cycle period, any node wishes to exchange information with other nodes would unilaterally determine whether or not to gossip. A node which decides to gossip in a duty-cycle period will broadcast its own

information to its neighbors and listen to potential incoming messages from its neighbors. Other non-gossiping nodes will wake up to only listen to the potential incoming traffic. It is possible for a node to detect multiple messages during the same epoch. We assume that an awaken node can successfully receive a message from a gossiping node as long as it is in the transmission range of the currently gossiping nodes. The received messages and the node's original message will be fused into a single message for gossiping in next epochs.

Clearly, one can directly investigate the effect of parameters such as the sleep period and the transmission range in this framework. What may not be immediately apparent is that, due to the fact that any node that gossips can be thought of as a volunteered center node, the optimal probability for gossiping hints at the corresponding level of optimal degrees for a hierarchical network.

In order to keep our results can be generally applied in WRNs, we decouple media access control (MAC) mechanism in the analysis of our framework. Many proposed MAC techniques [53, 87] are potential approaches for energy-efficient WRNs.

4.1.2 Parameter Notations in RGNs

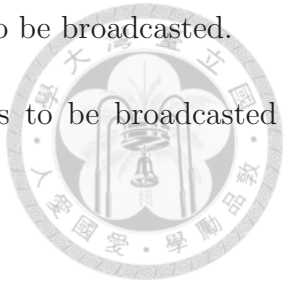
We define the following parameters in the random gossip network. The parameters are classified into: environmental parameters, application parameters, design parameters, and node parameters.

Environmental parameters:

- n: the network dimensionality. We consider 1-D, 2-D, and 3-D networks.
- V: the physical size of the network. It represents the length, area, and volume for a 1-D, 2-D, and 3-D network, respectively.
- α: the path loss exponent.
- ρ: the geographical node density of the network.

Application parameters:

- T_m: the average period that all nodes initiate a message to be broadcasted.
- T_{max}: the maximal allowable delay time for all messages to be broadcasted within the whole network.



Design parameters:

- D: the transmission range. The maximal distance at which a message can be successfully received. On the other hand, a node can never receive a message out of the range. Even though this simplistic assumption indeed does not fully model the nature of wireless transmission between transmission range and error probability, we believe this simplification will not significantly affect the validity of our work.
- p^n: the participation density. It represents the geographical density of gossiping nodes in a RGN.
- T_d: the duration of duty-cycle period.

Node parameters:

- E_t: the transmission energy. The energy consumed by a node to transmit a message in a duty-cycle period. For a path loss exponent α, E_t is often modelled as $E_t^1 D^\alpha$ where E_t^1 stands for the transmission energy when the reach D is set to 1.
- E_d: the detection energy. The energy consumed by a node to detect a message in a duty-cycle period. We assume that the detection energy of a node does not depend on whether or not a message is received [75].

Note that consensus achievement [91] by simple data processing function such as average or maximize the received interests is considered in various WRN applications. In these applications, the size of a broadcasted message is usually independent on how many times the message is being processed or being relayed. For simplicity, we assume that the same energy is spent for a node to broadcast once in our analysis.

It is shown in case studies, by removing this assumption, our framework can still reasonably predict the optimal value of operational parameters in proposed WRNs.

4.2 Power Consumption Analysis in RGN

In this section, we consider the network power consumption for all-to-all broadcast in the RGN. Specifically, each node in the network initiates a message to be broadcasted to all the other nodes with a given message origination period.

4.2.1 The Notion of Additional Energy Consumption over Listening

If there is no message in the network, nodes in the network wake up periodically to listen to potential incoming messages. We refer to such energy consumption as the energy consumption for listening (ECL). This function is usually a simple function of sleep mechanism.

On the other hand, the additional energy consumed over listening (AECOL) is defined as the energy used for gossiping on the top of the energy consumed for listening. According to this definition, if there is no traffic in the network, the average AECOL of the network is zero. The AECOL depends heavily on the gossiping mechanism.

To minimize the energy consumption of the network, the ECL and AECOL must be balanced carefully.

4.2.2 Mean ECL and AECOL for Network-Wide Broadcast in a Linear RGN

Consider a 1-D regular RGN in which nodes are placed with unit-length apart, shown in Figure 4-1. In the network, each node has a message to broadcast to all the other nodes. Suppose that all nodes starts to gossip randomly at time epoch 1.

First, we compute the energy consumption of the whole network in a duty-cycle period. Define the random variable W_k as the AECOL of the network in epoch k.

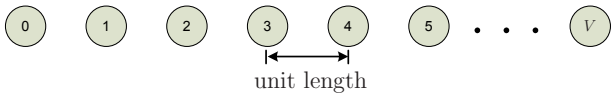


Figure 4-1: One-dimensional network with unit-length spaced nodes.

In an epoch, each gossiping node consumes exactly energy $E_t + E_d$. The AECOL of a gossiping node is E_t and the AECOL of a non-gossiping node is certainly zero. As a result, W_k depends only the number of gossiping nodes and the transmission energy. The expected value of AECOL spent in epoch k is $\mathbb{E}[W_k] = p^n V E_t$. On the other hand, the ECL of the whole network depends only on the number of nodes in the network and the detection energy. The ECL of the whole network in an epoch is $\rho V E_d$.

Next, we consider the AECOL for all messages to be broadcasted in the whole network. Define a random variable H as the number of time epochs it takes for all the messages to be broadcasted in the whole network. One can obtain the following theorem for AECOL of the whole network.

Theorem 4.1. *The expected value of AECOL incurred by an all-to-all broadcast in the RGN is* $\mathbb{E}[H]\mathbb{E}[W]$ *where* $\mathbb{E}[H]$ *represents the expected number of time epochs for a single all-to-all broadcast and* $\mathbb{E}[W]$ *stands for the expected value of AECOL during a time epoch.*

Finally, given the average rate of message origination T_m, the average power consumption of the whole network can be represented as:

$$\frac{\rho V E_d}{T_d} + \mathbb{E}[H]\frac{p^n V E_t}{T_m} \quad . \tag{4.1}$$

4.2.3 Sensor Field Approximation for the Linear RGN

We desire expressions to gain better insights for the optimization of the design parameters. Unfortunately, we can not find simple closed-form expression for $\mathbb{E}[H]$. In this section, we approximate the discrete RGN as a continuous one to obtain close approximations. Specifically, a 1-D uniform random gossip *sensor filed* is a network

in which infinite many nodes are uniformly placed on a segment $[0, V]$. Over any continuous segment of length δ, there exists $p\delta$ nodes gossiping on average in a duty-cycle period.

In the 1-D random gossip sensor filed, consider a message originated from one end to be broadcasted to the other end. Define the random variable H' as the number of time epochs for the message to be broadcasted over a distance L away. Given the operational assumption that the message is broadcasted at time epoch 1, the expected value of the H' can be bounded by the following proposition.

Proposition 4.2. *The expected value of the number of time epochs for a message to be broadcasted over a distance L away in the random gossip sensor field can be bounded as:*

$$\frac{L-2D}{\mu_1} + 2 \leq \mathbb{E}[H'] < \frac{L-D}{\mu_1} + \frac{1}{1-P_0} + 1. \tag{4.2}$$

In (4.2), μ_1 stands for the expected forwarding distance per epoch and P_0 represents the probability that a message does not advance to any new node in a sleep epoch. Note that in a large-size WRN in which the network size is much larger than a single hop transmission, the extra cost in epoch 1 can be reasonably omitted.

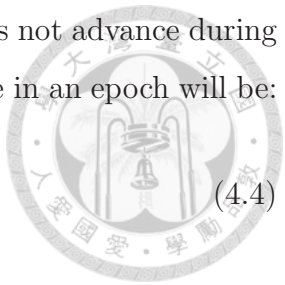
From Proposition 4.2.3, it can be observed that the expected number of time epochs for a message to be broadcasted in the whole network in the random gossip sensor field depends highly on the traveling distance of a message. Denote the maximal distance between any two nodes in the whole network as L_{max}. Henceforth, for the ease of analysis, we approximate $\mathbb{E}[H] \approx L_{max}/\mu_1$.

Now let us compute the expected forwarding distance in the 1-D random gossip sensor field. Define the random variable X as the distance a message gains in a time epoch. In the 1-D uniform random gossip sensor field, except the first epoch, the distance gains for a message in a time epoch is a truncated exponential distribution:

$$P(X \leq x) = \begin{cases} 1, & x > D, \\ e^{-p(D-x)}, & 0 < x \leq D, \\ 0, & x < 0. \end{cases} \tag{4.3}$$

Note that there is a finite probability of e^{-pD} that a message does not advance during an epoch. From equation (4.3), the expected forwarding distance in an epoch will be:

$$\mu_1 = D - \frac{1 - e^{-pD}}{p} \quad . \tag{4.4}$$



4.2.4 Multi-Dimensional RGNs

The results in the 1-D RGN can not be directly applied in a multi-dimensional RGN. Why? One significant difference between the 1-D RGN and a multi-dimensional RGN is the presence of network boundaries. In a 2-D or a 3-D RGN, the expected number of hops for a message to be propagated within the whole network indeed depends on the shape of the network. As a result of this fact and other minor reasons, we can not find closed-form expressions of the expected forwarding distance in a higher-dimensional network.

Instead, we derive the following approximations for an n-dimensional (n-D) random gossip sensor field.

Proposition 4.3. *For an n-D uniform random gossip sensor field, the expected forwarding distance in a time epoch can be approximated as:*

$$\mu_1 \approx D\left(1 - \frac{1 - e^{-p^n D^n}}{p^n D^n}\right). \tag{4.5}$$

As it turns out, the quantity $p^n D^n$ is a key parameter in the proposition. This quantity is exactly the expected number of nodes gossiping in a time epoch within an n-dimensional cube of side D.

4.3 Design Parameter Optimization

In this section, we derive the optimal value of the design parameters for minimizing power consumption of all-to-all broadcast in the RGN. We consider three design parameters: the transmission range, the duty-cycle period, and the participation density are free to be adjusted. Due to the space limitations, we show only the

results that all three free parameters can be jointly adjusted in the following. In the optimization problems, the environmental parameters, the application parameters and the node parameters are given constants. The optimization is performed with a delay time requirement:

$$\mathbb{E}[H]T_d \leq T_{max.} \tag{4.6}$$

It can be shown that the combination of (p^*, D^*, T_d^*) that the network power consumption minimized must satisfy $T_d^* = T_{max}/\mathbb{E}[H]$. By substituting $T_d = T_{max}/\mathbb{E}[H]$ into (4.1), one can express the average total energy consumption rate as:

$$\mathbb{E}[H](\frac{\rho V E_d}{T_{max}} + \frac{p^n V E_t}{T_m}) \approx \frac{L_{max}}{D}(1 - \frac{1 - e^{-p^n D^n}}{p^n D^n})^{-1}(\frac{\rho V E_d}{T_{max}} + \frac{p^n V E_t^1 D^\alpha}{T_m}). \tag{4.7}$$

The optimal setting for power consumption minimization for all-to-all broadcast satisfies the following two equations:

$$p^n D^n + (\alpha - n + 1)e^{-p^n D^n} + (\alpha - n)p^n D^n e^{-p^n D^n} = \alpha - n + 1 \tag{4.8}$$

and

$$D = \left((\alpha - n - 1)\rho^{-1} p^n D^n \frac{E_t^1}{E_d}\frac{T_{max}}{T_m}\right)^{\frac{-1}{\alpha - n}}. \tag{4.9}$$

Note that (4.8) and (4.9) are valid only for path loss exponent $\alpha > n + 1$. In (4.8), pD-product stands for the expected number of nodes gossiping in a duty-cycle period within the transmission range per dimension. It is a critical quantity for an optimized RGN.

From (4.8), we can observe the following important property.

Proposition 4.4. *The optimal number of nodes gossiping within the transmission range per sleep epoch depends only on the path loss exponent and the network dimensionality.*

It is quite surprising to note that this quantity does not depend on factors such as the physical network size, the message origination rate, and the detection energy. Especially, neither increasing nor decreasing the physical network size will affect the

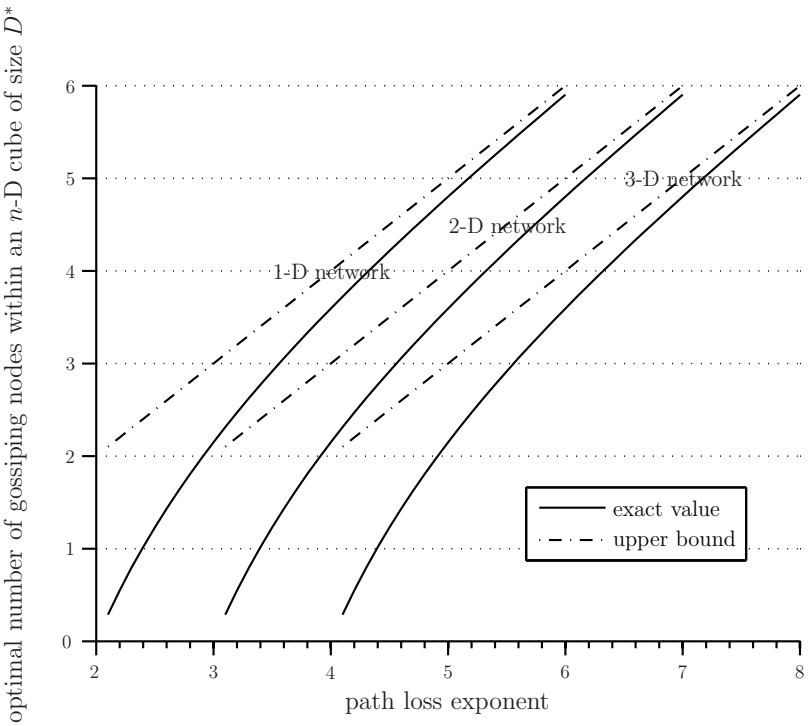


Figure 4-2: The optimal number of nodes gossiping within an n-D cube of size D^* in a time epoch. Both the optimal values and the lower bounds for 1-D 2-D and 3-D networks are shown.

optimal value of the design parameters. The optimal setting for power consumption minimization of all-to-all broadcast is inherent a scalable solution.

From equation (4.8), it can be derived that the optimal number of nodes gossiping within the transmission range increases monotonically as the path loss exponent increases. Also, it can be obtained that the optimal number of nodes gossiping within the transmission range can upper bounded by $(p^*D^*)^n < \alpha - n + 1$.

Figure 4-2 shows the optimal number of nodes gossiping within an n-D cube and its upper bound for power consumption minimization as a function of the path loss exponent. It can be observed that more nodes shall wake up to gossip within the transmission range in a higher dimensional network. What may be surprise is that, the optimal expected number of nodes awaken to broadcast messages in the transmission range increases monotonically as a function of the path loss exponent.

Given the optimal pD-product, from (4.9), one can determine the trend of p^* and D^* as the transmission-to-reception energy ratio varies. Specifically, $D^* \propto (E_t^1/E_d)^{\frac{-1}{\alpha-n+1}}$ and $p^* \propto (E_t^1/E_d)^{\frac{1}{\alpha-n+1}}$. The result is quite intuitive; the transmis-

sion range should decrease as the transmission energy per unit length normalized to the detection energy grows. A corollary to this finding is that *either shortest hop transmission or blind flooding can never be the optimal strategy as long as the detection energy is nonzero.*

Although Proposition 4.3 can not be applied in the regime $\alpha \leq n+1$, it can also be found that either minimal transmission energy broadcasting or always flooding the received messages is rarely the optimal strategy. In a large size RGN in which the network size is much larger than a single hop transmission, a surprising result is that the network power consumption is usually minimized by adopting a maximum feasible transmission range.

Now let us examine properties under the optimal settings above. One interesting question is how much of energy nodes is spent for listening to the potential incoming traffic? Under the optimal setting obtained from equation (4.8) and (4.9), the ratio between ECL and AECOL can be characterized by the following proposition.

Proposition 4.5. *The ratio between ECL and AECOL is exactly* $\alpha - n - 1$ *in the optimized RGN.*

In the proposition, the optimal ratio depends only on the path loss exponent and the network dimensionality. The network power consumed to detect potential incoming messages is usually at the same order of power consumption for gossiping. The higher the path loss exponent, the less energy spent in transmission. Note that this value does not depend on factors such as the the network size, the detection energy, and the message origination rate.

Another interesting question is how the application parameters such as the maximal allowable delay time and message origination rate affect the network power consumption? The power consumption under optimal setting can be obtained by substituting p^*, D^*, T_s^* into (4.7): The minimal power consumption of the whole network is

$$f(\alpha, n, \rho, E_t^1, E_d) V L_{max} \left(\frac{E_d}{T_{max}}\right)^{\frac{\alpha-n-1}{\alpha-n}} \left(\frac{E_t^1}{T_m}\right)^{\frac{1}{\alpha-n}} \tag{4.10}$$

In equation (4.10), the function $f(\alpha, n, \rho, E_t^1, E_d)$ depends only on the environmental

parameters and the node parameters. The minimal power consumption is proportional to the physical size of the network and the maximal separation of two nodes in the network. From (4.10), we can obtain the following proposition.

Proposition 4.6. *The power consumption is exactly proportional to the* $\frac{\alpha-n-1}{\alpha-n}$*-th power of the inverse delay time and is exactly proportional to the* $\frac{1}{\alpha-n}$*-th power of the message origination rate.*

In the proposition, we can observe that the network power will never increase linearly for both of the application parameters T_{max} and T_m. While the value of $\alpha - n - 1$ approaches zero, the network power consumption will almost be irrelevant to the delay time T_{max}. On the other hand, while the value of $\alpha - n - 1$ approaches infinity, the network power consumption will almost be irrelevant to the delay time T_m.

4.4 Case Studies

To illustrate how the core concept derived in the prior Sections can be applied, in this section we examine a WRN already proposed in the literature. We will show that our results can indeed predict the optimal setting with good accuracy not only for information exchange but also for information dissemination in the proposed WRN.

For energy-efficient broadcasting, especially in large-size WRNs, decentralized algorithms have been investigated in great detail, e.g. [88, 91, 92]. In these scalable designs, a node often determines whether or not to propagate its information via only partial neighborhood information.

Here we consider a network using $GOSSIP1(p)$ proposed in [88]. In $GOSSIP1(p)$, a node at which a message is originated broadcasts the message with probability 1. A node which receives the message at the first time will broadcast the message with probability p. A node will never broadcast a message that is received more than once. The broadcasting processes will be terminated by itself. In [88], the parameter p is called the gossip probability.

As the proposed $GOSSIP1(p)$ does not specify any sleep mechanism, here we add a simple sleep mechanism for this network. In the network, nodes are synchronized to wake up to listen to the potential incoming traffic at the start of an epoch and switch to sleep mode at the rest of the period. In an epoch, a node can broadcast once.

In this case, our goal is to find the optimal settings for $GOSSIP1(p)$ with the given sleep mechanism. We assume that the transmission range, the gossip probability, and the sleep period are all free to be adjusted for power consumption minimization.

For this network, an analogous random gossip network has the parameters as follows. The transmission range is the maximal range a message can be received in one hop. The duty-cycle period is the duration of an epoch. The participation density is the geographical node density multiplied by the gossip probability. By the simple parameter transformation, this problem corresponds to the optimization of the three design parameters in the analogous RGN.

Both information exchange and information dissemination are considered in our simulation. In [88], $GOSSIP1(p)$ is originally proposed for information dissemination. We first consider $GOSSIP1(p)$ for information dissemination and then apply it for information exchange. In the simulation, a message is originated from a randomly selected node in the whole network within the message origination period for information dissemination. On the other hand, each node in the network initiates a message within the message origination period and starts to broadcast its message concurrently for information exchange.

Except the given sleep mechanism, a network exactly the same as the one in [88] is simulated. Specifically, 1000 nodes are randomly deployed in 7500 meter-by-3000 meter area. The decentralized algorithm does not guarantee that a message will be received by all nodes in the network. In the simulation, we put a requirement on the reception probability, denoted as P_r, which is defined as the average percentage of messages received by a node. In [88], the node energy profile is not specified in the simulations. To show that our results can be applied in nodes with various energy profiles, we examine various combination of node energy profiles in our simulation.

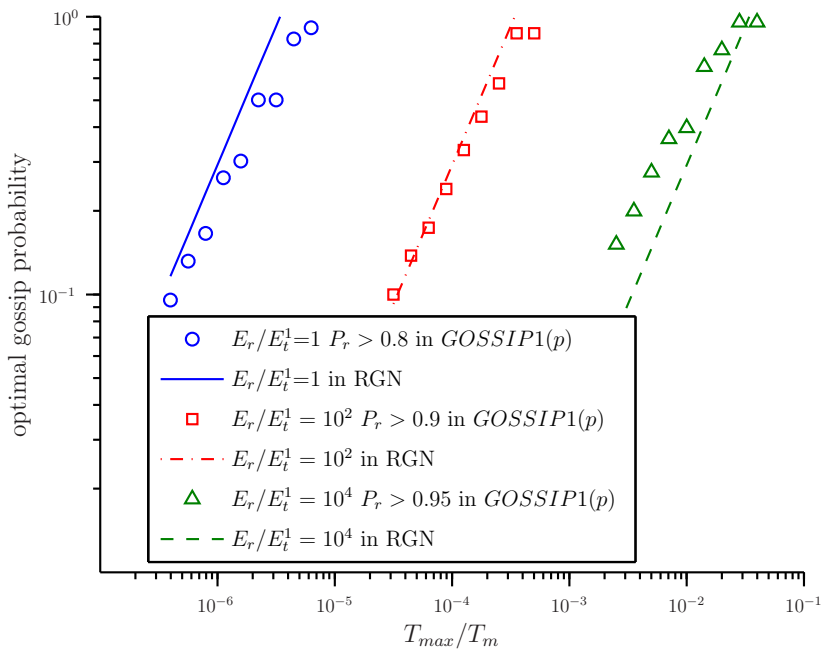


Figure 4-3: The optimal gossip probability as a function of the average delay time for information dissemination. The delay time is normalized by the message origination period. Both the simulation results and the corresponding optimal parameter in the analogous RGN are shown. The path loss exponent $\alpha = 3$ is given. Different average reception probability requirements are examined.

4.4.1 Information Dissemination in a Planar Network

For information dissemination, network parameters shall be further transformed to match the analogous RGN. Specifically, if most of nodes receive a broadcasted message, the total number of nodes to broadcast the message by $GOSSIP1(p)$ can be approximated by the number of nodes gossiping in a single hop in the analogous RGN. For simplicity, we normalize transmission energy to broadcast a message as $\frac{D}{L_{max}}E_t = \frac{E_t^1}{L_{max}}D^{\alpha+1}$ for the analogous RGN. In other words, the network can be modelled by increasing the path loss exponent by one and adjusting the transmission energy per unit length as E_t^1 divided by L_{max} in the analogous RGN.

The optimal gossip probability for power consumption minimization of information dissemination by $GOSSIP1(p)$ given the sleep mechanism is shown in Figure 4-3. The corresponding optimal parameter in the analogous RGN is also shown. It can be observed that the optimal gossip probability can be well predicted by the analogous RGN with a given high reception probability requirement.

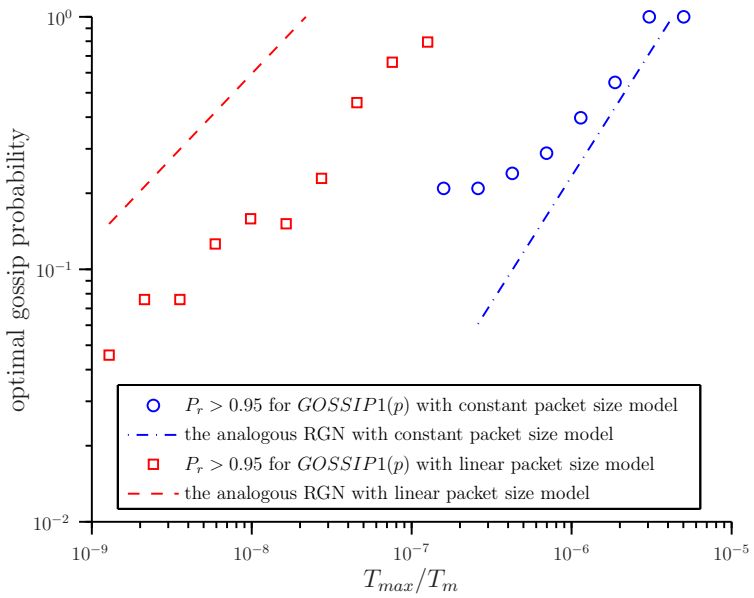


Figure 4-4: The optimal gossip probability as a function of the average delay time for information exchange. The delay time is normalized by the message origination period. Both the simulation results and the corresponding optimal parameter in the analogous RGN are shown. The path loss exponent $\alpha = 4$ and $E_d/E_t^1 = 10^4$ are given. Different data fusion models are examined.

4.4.2 Information Exchange with Data Fusion Techniques

Two extreme data fusion models are considered for information exchange in the simulation. One is the constant packet size model and the other is the linear packet size model. In the constant packet size model, newly received messages are fused into a constant-size packet for transmission. In the linear packet size model, newly received messages are concatenated into a single large packet for transmission. The size of a broadcasted message is the summation of the size of the newly received messages.

In the constant size model, the simulated network can be directly modelled by the analogous RGN without any further parameter transformation. On the other hand, for the linear packet size model, network parameters shall be further transformed to match the analogous RGN. Specifically, if a high reception probability requirement is given, almost every node will receive all broadcasted messages. The total transmission energy for a single all-to-all broadcast can be approximated as $p^n V \rho V E_t$ in the analogous RGN. For simplicity, we normalize transmission energy to broadcast once as $\frac{D}{L_{max}} \rho V E_t = \frac{\rho V E_t^1}{L_{max}} D^{\alpha+1}$ in the analogous RGN. In other words, the network can be

modelled by increasing the path loss exponent by one and adjusting the transmission energy per unit length as $\rho V E_t^1$ divided by L_{max} in the analogous RGN.

Figure 4-4 shows the optimal gossip probability for power consumption minimization of information exchange by $GOSSIP1(p)$ given the sleep mechanism. The corresponding optimal parameter in the analogous RGN is also shown. Except the condition that the network power consumption is dominated by the first hop transmission, the optimal gossip probability for both data fusion models can be reasonably predicted by the analogous RGN.

4.5 Summary

In this chapter, we investigate the power consumption for information exchange in a wireless-relay network. In light of the fact that it is typical for a low traffic density network to consume most of its energy on detecting the presence of incoming packets, we specifically take into account the reception mode power entailed by the sleep operation.

We propose the random gossip network as a framework for the analysis of power consumption in WRNs. It is shown by case studies that this framework can reasonably represent the power consumption behavior of proposed WRNs. Also, it is shown that our results can predict the optimal value of operational parameters for information exchange as well as information dissemination in these WRNs.

In the random gossip network, we consider that a network designer can manipulate the transmission range, the duty-cycle period, and the participation density. We find that surprisingly simple rules govern the optimal value of these three parameters. One key relationship is that the optimal number of nodes broadcasting messages in an epoch within the transmission range is determined only by the path loss exponent and the network dimensionality. This optimal value hints at the level of hierarchy of actual WRNs. It does not depend on factors such as the network size, the detection energy, the average message origination rate, and the maximal allowable delay time. Furthermore, the optimal value of these design parameters is not affected by the

physical size of a WRN. The optimal setting for power consumption minimization is inherent a scalable solution.

Chapter 5

Concluding Remarks

In this dissertation, we investigate the energy saving problems in wireless networks. Due to the dramatic variations of network traffic load caused by diverse applications and human daily routine, we intuitively explore the sleep mode operation for network energy saving for both cellular networks and multi-hop relay networks. We demonstrate that network energy consumption can be greatly reduced with the additional degrees of freedom provided by sleep-awake operations in wireless networks.

Specifically, we investigate green network coverage preservation problem in cellular networks. We find that BS activation problem for minimal network power consumption while maintaining network coverage is an NP-hard problem. We derive the optimal cell size for minimizing area power consumption and and propose a near-optimal polynomial-time load-aware algorithm for energy-efficient BS activation in HetNets. The simulation results show that our algorithm can approach the minimal network power consumption. Besides, we demonstrate that network densification by small cells for bursting throughput in hot spot areas can also be beneficial in saving network energy during the low traffic load period. Furthermore, we explore potential network structures with multi-mode BS operations in HetNets and suggest that uplink by small cells and downlink by macro cells will be promising during low traffic load periods for green HetNets.

For wireless multi-hop relay networks, we investigate network energy saving for two dual problems: message delivery and information exchange. We seek for general

design insights for energy-efficient operations in wireless multi-hop relay networks. For message delivery in wireless multi-hop relay networks, we develop a model for the analysis and optimization of performance trade-offs in wireless networks while sleep-awake scheduling mechanisms are applied. Specifically, we propose a random wakeup network with opportunistic relay as a framework for our analysis. We first derive network power consumption for the proposed framework and then find the optimal design parameters that can minimize the network power consumption while meeting the message delay requirement. We find that for a network that employs the optimal routing and sleep parameters thus obtained, the proportion of power for message detection in the entire network is $\frac{\alpha-1}{\alpha+2}$, where α is the path loss exponent. That is, under the optimal settings, the energy consumed for message detection should be of the same order as the energy consumed for actual communications. Moreover, we find that the minimal network power consumption under optimal operations grows at a rate of the $\frac{\alpha-1}{\alpha+2}$-th order of the message delivery speed, meaning that the investment of network power can efficiently reduce message delivery delay. We show through simulation results the proposed framework is well applicable for performance modelling and operational parameter optimization in wireless relay networks when sleep mode operation is adopted.

For information exchange and information dissemination in green wireless multi-hop relay networks, we propose a "random gossip" wireless-relay network (RGN) as a framework for our analysis. In the RGN, we consider that each node initiates a message and wishes to exchange its message with all the other nodes in the network. We find key properties for the three design parameters in the optimized RGNs. First, a key relationship is that the optimal number of nodes broadcast messages within the transmission range in a sleep period depends only on the path loss exponent and the network dimensionality. It does not depend on factors such as the physical network size, the reception mode energy, and the message origination rate. Next, the existing optimal number of broadcasting nodes in a time epoch within the transmission range shows that either shortest hop transmission or blind flooding is rarely optimal for power consumption minimization. Last but not the least, it is shown that neither

increasing nor decreasing the physical network size will affect the optimal value of these design parameters. The optimal setting for information exchange is a scalable solution. We demonstrate that our results in the RGN can indeed predict the optimal value of operational parameters for information exchange as well as information dissemination in actual WRNs.

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Vehicular Technology Magazine*, vol. 6, no. 1, pp. 37–43, 2011.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[3] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for Next Generation Wireless Systems," *arXiv preprint arXiv:1304.6690*, 2013.

[4] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[5] G. Fettweis and E. Zimmermann, "ICT energy consumption-trends and challenges," in *Proceedings of the 11th International Symposium on Wireless Personal Multimedia Communications*, 2008.

[6] W. Vereecken, W. Van Heddeghem, D. Colle, M. Pickavet, and P. Demeester, "Overall ICT footprint and green communication technologies," in *4th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, 2010.

[7] J. Hoadley and P. Maveddat, "Enabling small cell deployment with HetNet," *IEEE Wireless Communications*, vol. 19, no. 2, pp. 4–5, 2012.

[8] R. W. Thomas, D. H. Friend, L. A. DaSilva, and A. B. MacKenzie, "Cognitive networks: adaptation and learning to achieve end-to-end performance objectives," *IEEE Communications Magazine*, vol. 44, no. 12, pp. 51–57, 2006.

[9] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," in *IEEE ICC'09 Workshops*.

[10] 3GPP TSG-RAN WG3, "Potential solutions for energy saving for E-UTRAN," 3GPP TR 36.927.

[11] C. Peng, S. Lee, S. Lu, H. Luo, and H. Li, "Traffic-driven power saving in operational 3G cellular networks," in *ACM MobiCom'11*.

[12] S. Bhaumik, G. Narlikar, S. Chattopadhyay, and S. Kanugovi, "Breathe to stay cool: adjusting cell sizes to reduce energy consumption," in *first ACM SIGCOMM workshop on Green networking*, 2010.

[13] K. Dufková, M. Bjelica, B. Moon, L. Kencl, and J. Le Boudec, "Energy savings for cellular network with evaluation of impact on data traffic performance," in *12th European Wireless Conference (EW)*, 2010.

[14] S. Zhou, J. Gong, Z. Yang, Z. Niu, and P. Yang, "Green mobile access network with dynamic base station energy saving," in *ACM MobiCom'09*.

[15] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "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.

[16] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 56–61, 2011.

[17] N. Ahmed, S. S. Kanhere, and S. Jha, "The holes problem in wireless sensor networks: a survey," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 9, no. 2, pp. 4–18, 2005.

[18] S. Megerian, F. Koushanfar, M. Potkonjak, and M. B. Srivastava, "Worst and best-case coverage in sensor networks," *IEEE Transactions on Mobile Computing*, vol. 4, no. 1, pp. 84–92, 2005.

[19] D. Hochba, "Approximation algorithms for NP-hard problems," *ACM SIGACT News*, vol. 28, no. 2, pp. 40–52, 1997.

[20] H. Zhang and J. Hou, "Maintaining sensing coverage and connectivity in large sensor networks," *Ad Hoc & Sensor Wireless Networks*, vol. 1, no. 1-2, pp. 89–124, 2005.

[21] Z. Niu, S. Zhou, Y. Hua, Q. Zhang, and D. Cao, "Energy-aware network planning for wireless cellular system with inter-cell cooperation," *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, pp. 1412–1423, 2012.

[22] F. Han, Z. Safar, W. S. Lin, Y. Chen, and K. Liu, "Energy-efficient cellular network operation via base station cooperation," in *IEEE ICC'12*.

[23] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10–21, 2011.

[24] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. A. Thomas, J. G. Andrews, P. Xia, H. S. Jo, *et al.*, "Heterogeneous cellular networks: From theory to practice," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 54–64, 2012.

[25] Nokia Siemens Networks, "2020: Beyond 4G – radio evolution for the gigabit experience," *White Paper*, 2011.

[26] 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.

[27] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs, "LTE release 12 and beyond," *IEEE Communications Magazine*, vol. 51, no. 7, 2013.

[28] X. Xu, G. He, S. Zhang, Y. Chen, and S. Xu, "On functionality separation for green mobile networks: concept study over LTE," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 82–90, 2013.

[29] O. Arnold, F. Richter, G. P. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *IEEE Future Network and Mobile Summit*, 2010.

[30] P. Frenger, P. Moberg, J. Malmodin, Y. Jading, and I. Godor, "Reducing energy consumption in LTE with cell DTX," in *IEEE VTC 2011-Spring*.

[31] F. Richter, A. Fehske, and G. P. Fettweis, "Energy efficiency aspects of base station deployment strategies for cellular networks," in *IEEE VTC 2009-Fall*.

[32] IST-4-02-7756 WINNER II *D1.1.2 V1.2 WINNER II Channel Models*, September 2007.

[33] S. Sesia, I. Toufik, and M. Baker, *LTE: the UMTS long term evolution*. Wiley Online Library, 2009.

[34] B. Han, M. Peng, C. Yang, and W. Wang, "Interference margin analysis for OFDMA cellular networks planning," in *IEEE PIMRC'10*.

[35] C. Brunner, A. Garavaglia, M. Mittal, M. Narang, and J. V. Bautista, "Inter-system handover parameter optimization," in *IEEE VTC-Fall'06*.

[36] R. Susitaival and M. Meyer, "LTE coverage improvement by TTI bundling," in *IEEE VTC-Spring'09*.

[37] C. Liao and S. Hu, "Polynomial time approximation schemes for minimum disk cover problems," *Journal of combinatorial optimization*, vol. 20, no. 4, pp. 399–412, 2010.

[38] S. Har-Peled, "Being fat and friendly is not enough," *CoRR*, vol. abs/0908.2369, 2009.

[39] C. Ambühl, T. Erlebach, M. Mihalák, and M. Nunkesser, "Constant-factor approximation for minimum-weight (connected) dominating sets in unit disk graphs," *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pp. 3–14, 2006.

[40] D. Hochbaum and W. Maass, "Approximation schemes for covering and packing problems in image processing and VLSI," *Journal of the ACM*, vol. 32, no. 1, pp. 130–136, 1985.

[41] S. Masuyama, T. Ibaraki, and T. Hasegawa, "The computational complexity of the m-center problems on the plane," *IEICE Transactions (1976-1990)*, vol. 64, no. 2, pp. 57–64, 1981.

[42] L. Saker and S.-E. Elayoubi, "Sleep mode implementation issues in green base stations," in *IEEE PIMRC'10*.

[43] A. Corliano and M. Hufschmid, "Energieverbrauch der mobilen kommunikation," *Bundesamt für Energie, Ittigen, Switzerland, Tech. Rep*, 2008.

[44] 3GPP TSG-RAN, "Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA)," 3GPP TR 25.814.

[45] I. C. Wong and B. L. Evans, "Optimal downlink OFDMA resource allocation with linear complexity to maximize ergodic rates," *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, pp. 962–971, 2008.

[46] A. Laourine, A. Stephenne, and S. Affes, "On the capacity of log-normal fading channels," *IEEE Transactions on Communications*, vol. 57, no. 6, pp. 1603–1607, 2009.

[47] 3GPP TSG-RAN, "Further advancements for E-UTRA physical layer aspects," 3GPP TR 36.814.

[48] V. Raghunathan, C. Schurgers, S. Park, and M. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 40–50, 2002.

[49] A. Carroll and G. Heiser, "An analysis of power consumption in a smartphone," in *USENIX conference on USENIX annual technical conference*, 2010.

[50] R. Hu, Y. Qian, H. Chen, and A. Jamalipour, "Recent progress in machine-to-machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, 2011.

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

[52] L. Wang and Y. Xiao, "A survey of energy-efficient scheduling mechanisms in sensor networks," *Mobile Networks and Applications*, vol. 11, no. 5, pp. 723–740, 2006.

[53] I. Demirkol, C. Ersoy, and F. Alagoz, "MAC protocols for wireless sensor networks: a survey," *IEEE Communications Magazine*, vol. 44, no. 4, pp. 115–121, 2006.

[54] C. Shanti and A. Sahoo, "DGRAM: a delay guaranteed routing and MAC protocol for wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 10, pp. 1407–1423, 2010.

[55] R. Cohen and B. Kapchits, "An optimal wake-up scheduling algorithm for minimizing energy consumption while limiting maximum delay in a mesh sensor network," *IEEE/ACM Transactions on Networking*, vol. 17, no. 2, pp. 570–581, 2009.

[56] S. Lim, C. Yu, and C. Das, "RandomCast: an energy-efficient communication scheme for mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 8, pp. 1039–1051, 2009.

[57] Q. Cao, T. Abdelzaher, T. He, and J. Stankovic, "Towards optimal sleep scheduling in sensor networks for rare-event detection," in *ACM/IEEE IPSN*, 2005.

[58] S. Cui, R. Madan, A. Goldsmith, and S. Lall, "Cross-layer energy and delay optimization in small-scale sensor networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 10, pp. 3688–3699, 2007.

[59] J. Kim, X. Lin, N. Shroff, and P. Sinha, "Minimizing delay and maximizing lifetime for wireless sensor networks with anycast," *IEEE/ACM Transactions on Networking*, vol. 18, no. 2, pp. 515–528, 2010.

[60] F. Liu, C. Tsui, and Y. Zhang, "Joint routing and sleep scheduling for lifetime maximization of wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 7, pp. 2258–2267, 2010.

[61] H. Kim, T. Abdelzaher, and W. Kwon, "Minimum-energy asynchronous dissemination to mobile sinks in wireless sensor networks," in *ACM Sensys*, 2003.

[62] K. Akkaya and M. Younis, "Energy-aware delay-constrained routing in wireless sensor networks," *International Journal of Communication Systems*, vol. 17, no. 6, pp. 663–687, 2004.

[63] H. Ammari and S. Das, "A trade-off between energy and delay in data dissemination for wireless sensor networks using transmission range slicing," *Computer Communications*, 2007.

[64] M. Zorzi and R. Rao, "Geographic random forwarding (GeRaF) for ad hoc and sensor networks: energy and latency performance," *IEEE Transactions on Mobile Computing*, vol. 2, no. 4, pp. 349–365, 2003.

[65] X. Yang and N. Vaidya, "A wakeup scheme for sensor networks: Achieving balance between energy saving and end-to-end delay," in *IEEE RTAS*, 2004.

[66] O. Dousse, P. Mannersalo, and P. Thiran, "Latency of wireless sensor networks with uncoordinated power saving mechanisms," in *5th ACM international symposium on Mobile ad hoc networking and computing*, 2004.

[67] R. Shah, S. Wietholter, and A. Wolisz, "Modeling and analysis of opportunistic routing in low traffic scenarios," in *WiOpt*, 2005.

[68] C. Chiasserini and M. Garetto, "An analytical model for wireless sensor networks with sleeping nodes," *IEEE Transactions on Mobile Computing*, vol. 5, no. 12, pp. 1706–1718, 2006.

[69] S. Biswas and R. Morris, "Opportunistic routing in multi-hop wireless networks," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 1, pp. 69–74, 2004.

[70] S. De, A. Caruso, T. Chaira, and S. Chessa, "Bounds on hop distance in greedy routing approach in wireless ad hoc networks," *International Journal of Wireless and Mobile Computing*, vol. 1, no. 2, pp. 131–140, 2006.

[71] S. Vural and E. Ekici, "On multihop distances in wireless sensor networks with random node locations," *IEEE Transactions on Mobile Computing*, vol. 9, no. 4, pp. 540–552, 2010.

[72] G. Rahmatollahi and G. Abreu, "Closed-form hop-count distributions in random networks with arbitrary routing," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 429–444, 2012.

[73] E. Neuman, "Inequalities and bounds for the incomplete gamma function," *Results in Mathematics*, 2012.

[74] S. Cui, A. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, 2005.

[75] M. Stemm and R. Katz, "Measuring and reducing energy consumption of network interfaces in hand-held devices," *IEICE Transactions on Communications*, vol. 80, no. 8, pp. 1125–1131, 1997.

[76] V. Rodoplu and T. Meng, "Minimum energy mobile wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1333–1344, 1999.

[77] B. Karp and H. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in *ACM MobiCom*, 2000.

[78] Y. Sun, Q. Jiang, and M. Singhal, "An edge-constrained localized delaunay graph for geographic routing in mobile ad hoc and sensor networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 4, pp. 479–490, 2010.

[79] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *33rd Annual Hawaii International Conference.*

[80] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed energy conservation for ad hoc routing," in *7th Annual International Conference on Mobile Computing and Networking*, 2001.

[81] A. Manjeshwar and D. Agrawal, "TEEN: a routing protocol for enhanced efficiency in wireless sensor networks," in *Parallel and Distributed Processing Symposium., Proceedings 15th Internationa1, IEEE*, pp. 2009–2015, 2001.

[82] B. Zhou, A. Marshall, and T.-H. Lee, "An energy-aware virtual backbone tree for wireless sensor networks," in *IEEE GLOBECOM '05.*

[83] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[84] L. Feeney and M. Nilsson, "Investigating the energy consumption of a wireless network interface in an ad hoc networking environment," in *INFOCOM 2001.*

[85] K. Akkaya and M. Younis, "A survey on routing protocols for wireless sensor networks," *Ad hoc networks*, vol. 3, no. 3, pp. 325–349, 2005.

[86] B. Krishnamachari, D. Estrin, and S. Wicker, "The impact of data aggregation in wireless sensor networks," in *22nd IEEE International Conference on Distributed Computing Systems*, 2002.

[87] N. Burri, P. Von Rickenbach, and R. Wattenhofer, "Dozer: ultra-low power data gathering in sensor networks," in *6th ACM international conference on Information processing in sensor networks*, pp. 450–459, 2007.

[88] Z. Haas, J. Halpern, and L. Li, "Gossip-based ad hoc routing," *IEEE/ACM Transactions on Networking*, vol. 14, no. 3, pp. 479–491, 2006.

[89] C. Chang and D. Shiu, "Transmit and receive power optimization for source-initiated broadcast in wireless-relay sensor networks," in *the World Congress on Engineering*, 2009.

[90] S. Lindsey and C. Raghavendra, "Energy efficient broadcasting for situation awareness in ad hoc networks," in *International Conference on Parallel Processing*, 2001.

[91] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[92] I. Stojmenovic, M. Seddigh, and J. Zunic, "Dominating sets and neighbor elimination-based broadcasting algorithms in wireless networks," *IEEE transactions on Parallel and Distributed Systems*, pp. 14–25, 2002.