Power Control in Wireless Body Area Networks for Interference Mitigation

By

Ramtin Kazemi

A thesis submitted to Macquarie University
for the degree of
Doctor of Philosophy
Department of Engineering
May 2013

MACQUARIE
UNIVERSITY
FACULTY OF SCIENCE

Typeset in LaTeX 2_ε.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Ramtin Kazemi

Acknowledgments

The work presented in this thesis would not have been accomplished without the assistance of a number of people who deserve special mention.

First and foremost, I express my sincerest gratitude to my supervisor, Dr. Rein Vesilo. He has supported me with his knowledge, experience and patience throughout all the stages of the work. Definitely without his support, this thesis would have never finished. Thank you Rein, it has been a pleasure for me working with you.

I would also like to extend my gratefulness to Prof. Eryk Dutkiewicz who has been always very supportive, friendly and helpful with comments and suggestions as well as collaborating on the papers. You certainly made this research more fruitful with your support. Thank you Eryk. I am also thankful to Dr. Ren Ping Liu from CSIRO for his time, helpful comments and also for collaboration on a number of papers.

I like to thank Dr. Gengfa Fang for providing me with some codes and papers which were extremely helpful at the early stage of this research. I am also grateful to Dr. Lucian Busoniu who always generously provided me with his useful comments, help and support.

I also thank all my colleagues and friends at WCNL group for providing me with their help and support during my research. Their expertise on various fields, helped me get a deep insight into my research. Thanks for the time and support. In my workplace at the Department of Engineering, Macquarie University, I was honored to work next to a friendly and supportive group of PhD students. I would like to thank them all specially Boyd Murray and Quang Thai who have always provided help and support. I never forget the great time working next to you.

Last, but not least, I would like to thank my father and mother, who encouraged me to pursue a PhD degree and always mentally and emotionally supported me.

List of Publications

- Ramtin Kazemi, Rein Vesilo, Eryk Dutkiewicz, Ren Liu *Reinforcement Learning in Power Control Games for Inter-network Interference Mitigation in Wireless Body Area Networks.* (In the Proceedings of IEEE ISCIT 2012)
- Ramtin Kazemi, Rein Vesilo, Eryk Dutkiewicz, Ren Liu *Design Considerations of Reinforcement Learning Power Controllers in Wireless Body Area Networks.* (In the Proceedings of IEEE PIMRC 2012)
- Ramtin Kazemi, Rein Vesilo, Eryk Dutkiewicz, Ren Liu *Dynamic Power Control in Wireless Body Area Networks Using Reinforcement Learning With Approximation.* (In the Proceedings of IEEE PIMRC 2011)
- Ramtin Kazemi, Rein Vesilo, Eryk Dutkiewicz *Novel Genetic-Fuzzy Power Controller with Feedback for Interference Mitigation in Wireless Body Area Networks.* (In the Proceedings of IEEE VTC 2011)
- Ramtin Kazemi, Rein Vesilo, Eryk Dutkiewicz *Inter-network interference mitigation in WBANs Using Power Control Games.* (In the Proceedings of IEEE ISCIT 2010)

Abstract

Wireless Body Area Network (WBAN) is a new advanced technology of wireless networking inside and around the human body which has the potential to provide ubiquitous and continuous health monitoring, and reduce the cost of health-care services. This technology benefits from the recent advances in electronics and telecommunications brought about tiny sensor devices which can be implanted inside or attached on the human body. A WBAN is composed of a number of these miniature devices sampling signals of the body and sending them to a coordinator node for real-time monitoring or other medical purposes.

Energy is the scarcest resource in WBANs and it is therefore highly desirable to minimize energy dissipation in WBAN devices. One of the major sources of energy waste in WBANs originates from the interference between co-located WBANs working in the same frequency band. To mitigate this inter-network co-channel interference, transmission power control can be employed. A power controller adjusts the transmission power levels in order to maximize some utilities, such as throughput, with the least power. In this thesis, we aim to address the inter-network interference issue in WBANs by proposing practical power control mechanisms to reduce energy consumption and increase throughput as much as possible in WBANs.

We design a fuzzy-logic-based power controller which makes decisions on the transmission power level based on the SINR and interference power level. The proposed fuzzy power controller is then optimized off-line using genetic algorithms to increase throughput and reduce power consumption. Simulation results reveal that the proposed fuzzy power controller strongly outperforms a well-known power controller in the literature, called ADP[1], in terms of energy consumption per bit and also convergence.

We also propose a power controller based on the game theory where players of a non-cooperative game struggle to maximize their throughput with as low power for transmission as possible. We show that a pure unique Nash equilibrium exists in the game. Having found the best response of the players, we evaluate the performance of the proposed approach using simulation and compare its performance with the fuzzy power controller proposed earlier. Simulation results indicate that although the proposed power control game is outperformed by the fuzzy power controller in terms of

[1] Asynchronous Distributed Pricing Power Controller

energy consumption per bit, it is superior in terms of convergence. More importantly, the game power controller enables us to adjust the tradeoff between power and throughput easily and even adaptively, whereas in the fuzzy power controller this adjustment has to be carried out off-line sat the design stage by time-consuming genetic algorithms. We also propose adaptive methods to adjust pricing mechanism, taking into account the power budget and channel conditions of WBANs, which allows them to make the best use of their good conditions to achieve a higher throughput.

In an effort to enhance the adaptability and flexibility of the power controller, we employ learning algorithms and put forward a power controller which learns from experience to improve its performance. The proposed controller relies on the reinforcement learning to explore the environment and exploit the knowledge acquired from experience. We use approximation methods to tackle the curse of dimensionality issue and investigate a broad range of reinforcement learning algorithms. We scrutinize the performance of all the proposed approaches by extensive simulations and compare their performances in terms of throughput, power levels, energy consumption per bit and convergence. Simulation results illustrate that although the reinforcement-learning-based power controller suffers from a slower convergence compared to the fuzzy and game power controllers, it provides a better performance in terms of energy consumption per bit. Moreover, the reinforcement learning based power controller enjoys simplicity in design and high level of adaption to environment.

Moreover, for applications where meeting QoS requirements is more important than saving energy, we formulate the power control problem as an optimization problem which minimizes the total power consumption and meets the individual target rate of each WBAN. Having attained the optimal solution by using the Lagrange multipliers method, we present a distributed approach based on the Jacobi method for fixed-point calculations, which approximates the optimal solution and is suitable for practical WBANs without the need of any central arbiter. The simulation results indicate that the distributed approach is able to provide good performance which is reasonably close to that of the global optimum solution. Additionally, for cases where the optimization problem is not feasible, the proposed distributed approach provides a better QoS provisioning.

Contents

List of Figures

List of Tables

"Nothing can bring you peace but yourself."

Ralph Waldo Emerson

1 Introduction

1.1 Wireless Body Area Networks

Nowadays, the quality of medical services and health-care systems is regarded as a major benchmark of social welfare in the world. Over the last decade, however, health-care spending has shown a growing tendency to rise each year. According to a recent report [3], health expenditures increase 5.7 percent on average each year, jumping up from 17.9 percent of the GDP[1] of the USA in 2010 to 19.6 percent in 2021. In order to cope with this crisis, a new technology is needed to provide medical centres with the ability to remotely monitor the condition of patients particularly the elderly, children and chronically ill ones while they are at home or hospital. Hence in November 2007, IEEE established a new task group (TG6) under the IEEE 802.15 standard to develop a technology for short-range ultra-low-power wireless communication in/on and around the human body. The new-born technology was called Wireless Body Area Network (WBAN). Prior to this standard, wireless medical data collection systems were using standards such as ZigBee or Bluetooth that did not comply with the medical standard due to their size, power consumption and strong interference from other devices. Considering that potentially hundreds of sensors can be attached to a patient's body, such systems would become quite bulky and inefficient to be carried by patients.

WBAN technology benefits from the recent advancements in electronics and telecommunications which have made it possible to embed a micro-controller, sensors and radio interface for data transmission and reception in a single tiny chip that can be integrated into other wearable objects such as belts, wrist watches, glasses and head sets, or can

[1] Gross Domestic Product

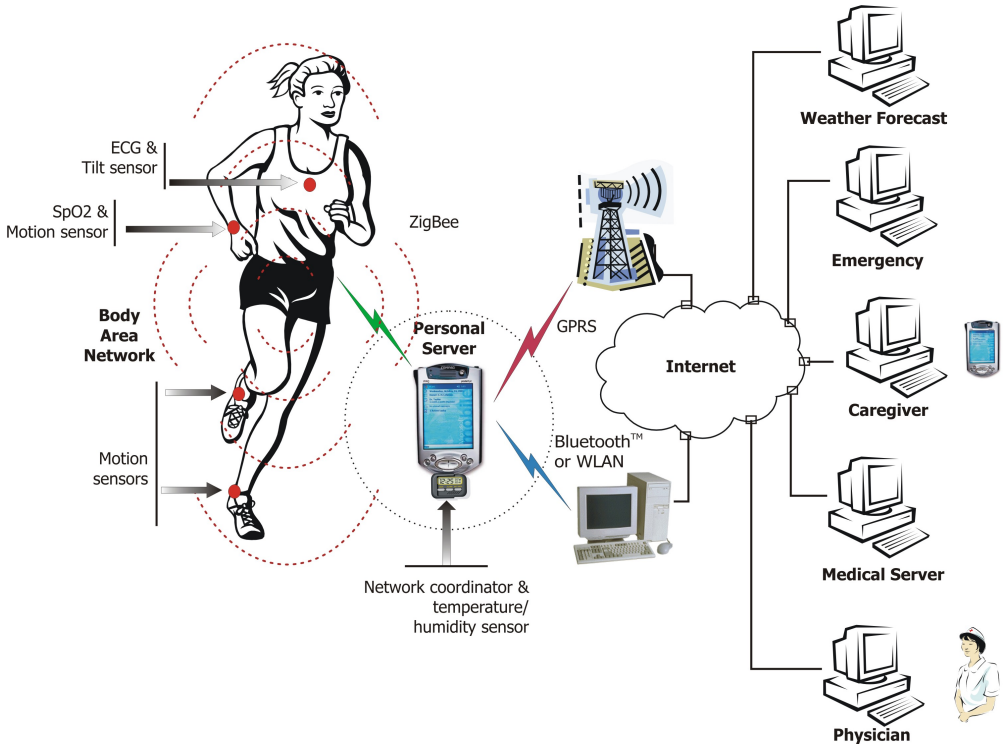

FIGURE 1.1: A Wireless Body Area Network along with other networking technologies; photo from The Journal of NeuroEngineering and Rehabilitation (JNER)

be implanted in/on the human body. A WBAN comprises a number of such miniature devices, known as BAN[1] Nodes (BNs)[2] forming a star topology[3] with a BAN Node Controller (BNC) node which serves as a coordinator, data collector, MAC controller and gateway for the WBAN. The sensor nodes detect an abnormality in the body or monitor physiological or physical signals of the body such as ECG[4], EEG[5], oxygen saturation level (SpO2), accelerometer and gyrometer signals, and report to the coordinator node.

1.1.1 Applications

With successful deployment of WBANs, patients can be examined, monitored and followed up remotely when they are at home, even asleep. WBANs enable elderly or after surgery patients to remain independently living in their own homes as long as possible, saving costs and time for both doctors and patients. Moreover, this will avoid

[1]Body Area Network

[2]Note that we use BN and sensor node interchangeably throughout the thesis.

[3]Relay communication between sensor nodes incorporating two-hop or multi-hop links in an extended star topology has been also studied in WBANs [4].

[4]Electrocardiography

[5]Electroencephalography

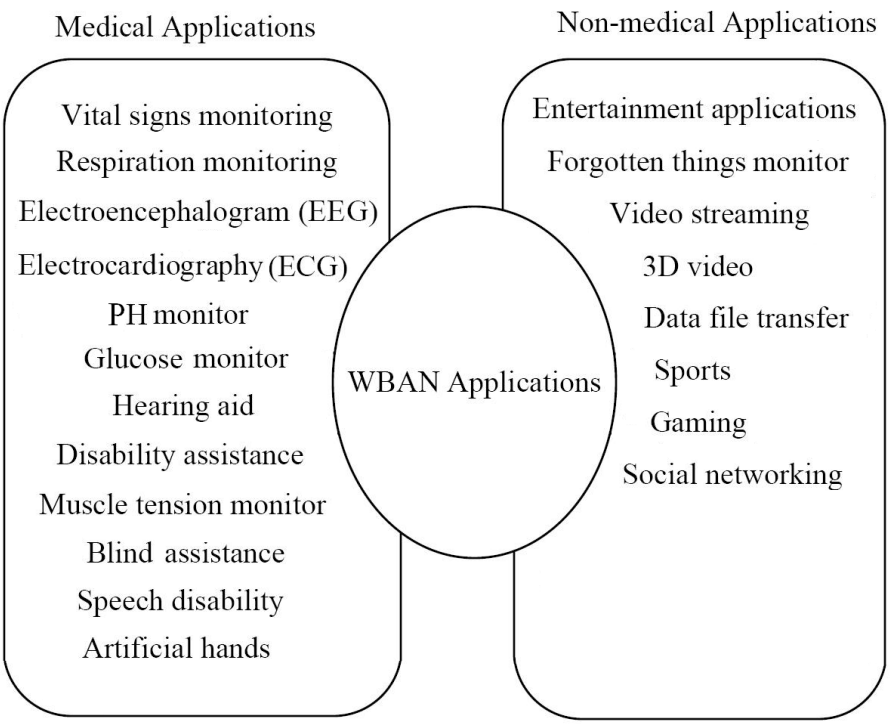

Figure 1.2: Applications of WBANs include medical and non-medical domains

patients' unnecessary presence at hospitals which further contributes to reduction of the health-care costs related to the cases where patients catch other infections known as HAI[1], when they visit or stay in hospitals [5].

WBANs may create interesting applications when combined with other technologies and devices which have become a necessary part of our daily life such as smart phones, PDAs, music players, etc. Checking the performance of our organs such as heart, liver and stomach, or finding any inefficiency in our body by just using our mobile phones will not be fictitious anymore and will be realized using WBANs in future.

In addition to the health-care and medical applications, WBAN is also a promising technology for ubiquitous and pervasive computations and has potential to serve applications in a wide range of domains including military services, battlefield management systems, sports and athletic training, surveillance, workplace safety, secure authentication, shopping, and entertainment and gaming. Figure 1.2 shows some applications of WBANs in the medical and non-medical domains.

[1]Hospital-Acquired Infection

1.2 Thesis Motivation

The lifetime of the sensor nodes in WBANs depend on the applications where they are being used which can vary from a couple of hours e.g. endoscopy capsules to a few years e.g. cardiac pacemakers or defibrillators. The sensor nodes all run on built-in batteries and in most cases, it is difficult or impractical to recharge or replace exhausted batteries, particularly when they are implanted inside the body. As a result, when the battery runs out, the senor node simply becomes inoperative. Although some energy harvesting techniques including motion [6] and body heat [7] scavenging have been considered, they still do not suffice to cope with the energy bottleneck in WBANs. Unlike other types of wireless communication technologies where bandwidth is the most valuable resource, energy is therefore considered to be the scarcest resource in WBANs. It is important to minimize energy waste and save power in WBANs.

One of the major sources of energy waste in wireless communications originates from interference between the signals of nearby nodes working in the same frequency band. The co-channel interference causes the SINR[1] to drop and thereby throughput degrades or bit error rate increases. In order to compensate for this SINR drop, an affected node has to raise its transmission power which will again produce more interference to the neighboring nodes and encourage them to transmit with a higher power in response. Such positive feedback behavior leads to a significant increase in power consumption in the system. Moreover, the packet loss caused by interference may result in more power consumption when collided packets have to be retransmitted.

Due to their structure, applications and mobility, a WBAN is quite likely to be affected by co-channel interference from other WBANs because their transmission ranges can easily overlap each other when for example patients stand or sit next to each other in the area where they the WBANs are deployed such as a hospital ward. The issue can be even more devastating in emergency medical applications where the reliability of the system is a key factor because the increased bit error rate or delay due to interference can jeopardize patients' lives.

The IEEE 802.15.6 standard has proposed two techniques for interference mitigation between neighboring WBANs, which are *beacon shifting* and *channel hopping*. It has also proposed two techniques for energy conservation which are *hibernation* and *sleeping*, known as *power management* techniques. However, the proposed interference mitigation techniques may not suffice to tackle the problem in practice. For example, due to the limited number of available frequency channels, hopping to another free channel may not be possible. On the other hand, the proposed power management techniques completely switch on/off the sensor nodes meaning that the radio is either on (full power) or off (zero power). In other words, the MAC of WBAN is *non-power-aware* and does not benefit from adjusting the transmission power levels to mitigate interference.

[1] Signal-to-Interference-and-Noise-Ratio

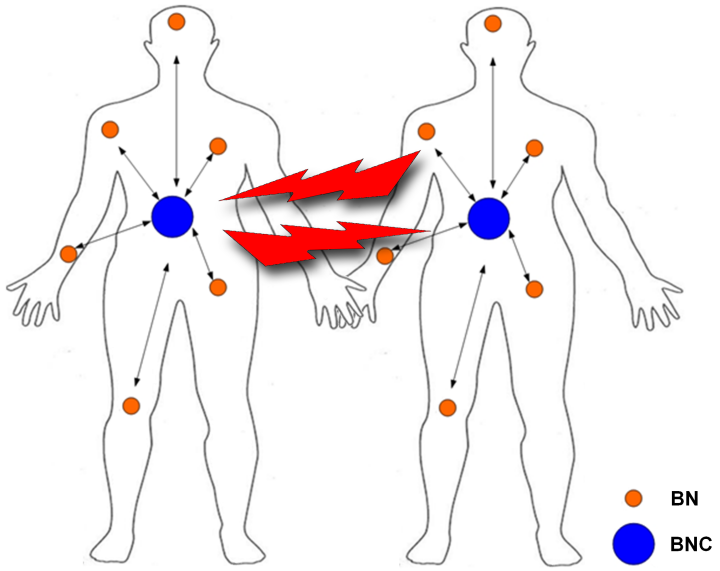

Figure 1.3: Inter-network interference between neighboring WBANs operating in the same frequency band

The co-channel interference between different WBANs, namely inter-network interference, can be managed by using Transmission Power Control (TPC). A power controller adjusts the transmission power level to achieve a specific goal in the system. The merits of using TPC include providing QoS[1], decreasing power consumption, enhancing throughput, controlling connectivity and network topology, increasing capacity, and so forth. The power control problem is often formulated as an optimization problem where the goal is to maximize a utility such as the total throughput in the system, with as little power for transmission as possible, or to minimize the total power consumption in the system while achieving the utility of interest such as a certain level of QoS.

In this thesis, we investigate techniques to mitigate inter-network interference in WBANs by proposing power control schemes to reduce energy consumption, increase throughput and keep WBANs reliable as much as possible.

1.3 WBAN Requirements

Any proposed approach for power control in medical WBANs must meet certain criteria to be applicable and suitable in the reality due to WBANs' special characteristics. It should ideally have the following properties:

[1]Quality of Service

1. *Distributed*: A centralized power controller requires a common arbiter which possesses all the information regarding all the nodes in the system and channel gains, and it controls all the transmission power levels in the network. On the other hand, a distributed approach adjusts the transmission power of one single node and is running by each node in the network separately until a global power allocation is found, usually in an iterative manner. Since there is no such central arbiter between different WBANs, any proposed solution for WBANs must be fully distributed and all WBANs should be treated in the same manner from the system point of view.
2. *Asynchronous*: Synchronous power control algorithms require that before proceeding to the next iteration, a radio interface that has finished its execution interval must wait for neighbouring nodes to finish their current iteration (i.e., the execution interval is synchronized). Such approaches need to employ synchronization techniques which usually involves negotiation between nodes in the network. In contrast, in asynchronous power control algorithms, each radio interface performs its power adjustments independently of other radio interfaces on the neighbouring nodes. In other words, such a radio interface can proceed to its next iteration without waiting for other radio interfaces to finish their current iteration interval.
3. *No Inter-WBAN Cooperation*: In most of medical applications, WBANs are reluctant to participate in any cooperation with other WBANs because each WBAN has to save its energy to do its critical tasks such as monitoring the heart beat signal. As a result solutions with the least cooperation between WBANs are highly preferred.
4. *No Negotiation or Message Exchange*: This is mainly because in almost all medical applications WBANs are assumed to be independent and non-cooperative. Moreover, the idea of negotiation of the interfering nodes may be infeasible in practice. This for example happens when the signals of WBANs are highly interfering with each other and they need to negotiate to find a solution reducing this interference while the negotiation itself will put more interference on them or even may not be possible at all in such high interference conditions.
5. *Least Processing Load*: In short-range wireless networks like WBANs (less than three meters), the processing power consumption may be not negligible compared to the transmission power. Since the tiny sensor nodes in WBANs are very energy-constrained, algorithms to be run on them should be light in terms of computation load and also memory requirements.
6. *Fast Convergence*: WBANs should run the power control algorithm distributively to find a global solution. As this normally needs to be done in an iterative manner, the convergence to the final solution becomes a concern not only for saving power but also for the sake of being real-time.

1.4 Scope of the Thesis

In this thesis, we consider techniques to manage inter-network interference between neighboring WBANs operating in the same frequency band using transmission power control in order to reduce energy consumption, increase throughput and keep WBANs reliable as much as possible.

The problem of minimizing the power consumption summed over all users in a system subject to meeting individual target SINR of users is a classic optimization problem which has been considered first in [8]. Although this model may be also suitable for some real-time medical applications of WBANs such as a surgery operation, where QoS is the most important, other WBANs' applications may be tolerable to reduce throughput for energy conservation. For example, at bad channel or high interference conditions, where the cost (e.g. energy per bit) to maintain a target rate is higher, a WBAN may prefer to postpone the transmission, or transmit at a lower data rate. Another case for motivating such tradeoff between power and throughput is when the battery is low and sensor nodes are willing to be more frugal for spending their battery. Therefore, instead of meeting target (data) rates at any cost, i.e. non-tradeoff scenario, in this thesis (Chapters 3,4 and 5) we mainly aim to achieve a tradeoff between throughput and power. However, in Chapter 6, we will turn back to the non-tradeoff problem and propose a solution for the sake of QoS-sensitive medical applications.

Making the tradeoff, we define a utility function which employs a penalty mechanism to discourage WBANs with high power levels motivating them to achieve a higher throughput with less power. This way, WBANs can reach a tradeoff between throughput and power consumption by adjusting appropriate price factor. We also propose adaptive methods to adjust the price factor to allow WBANs to make the best use of their good conditions in terms of channel and power budget to achieve a higher rate.

We propose three approaches which are based on *genetic-fuzzy systems*, *game theory*, and *Reinforcement Learning* (RL). Fuzzy control combined with genetic algorithms, known as genetic-fuzzy systems, provides a powerful tool for decision making and controlling complicated systems without requiring an exact mathematical model while providing a high level of flexibility. However, genetic algorithms are very time-consuming and are used only in off-line optimizations, which limits the adaptability of the system to on-line dynamic changes of the surrounding environment. Game theory, on the other hand, is a mathematical framework which analyzes the conflict of interests between agents and is needless of such off-line optimization. Although, game theory can improve the adaptability of the system compared to the genetic-fuzzy approach, it normally requires finding a stable solution called Nash Equilibrium (NE) in advance, and calculating a strategy such as the best response to reach that solution. In order to further boost the adaptability, reinforcement learning can be employed, which is a framework for finding optimal solutions without the need of the environment model and only by interacting with the environment. Reinforcement learning allows agents

to learn from experience and improve their performance over time.

The proposed fuzzy power controller in Chapter 3 makes decisions on the transmission power level based on the SINR, interference power level and the current level of transmission power. We design a genetic algorithm and a learning strategy to optimize the power controller. Simulation results indicate that the proposed fuzzy power controller strongly outperforms a well-known power controller in the literature, called ADP[1], in terms of energy consumption per bit and also convergence.

The proposed game-theoric power controller in Chapter 4 is a non-cooperative game in which players struggle to maximize their throughput with as low power for transmission as possible. We show that a pure unique NE exists in the game. Having found the best response of the players, we evaluate the performance of the proposed approach using simulation and compare its performance with the fuzzy power controller proposed earlier. Simulation results reveal that although the proposed power control game is outperformed by the fuzzy power controller in terms of energy consumption per bit, it is superior in terms of convergence. More importantly, the game power controller enables us to adjust the tradeoff between power and throughput easily and even adaptively, whereas in the fuzzy power controller this adjustment has to be carried out off-line at the design stage by time-consuming genetic algorithms.

The proposed RL-based power controller in Chapter 5 improves its performance by exploring the environment while exploiting the knowledge acquired from experience. We use approximation methods to tackle the issue of large state-action space and investigate a broad range of reinforcement learning algorithms. We scrutinize the performance of all the proposed approaches by extensive simulations and compare their performances in terms of throughput, power levels, energy consumption per bit and convergence. Simulation results illustrate that although the RL-based power controller suffers from a slower convergence compared to the fuzzy and game power controllers, it provides a better performance in terms of energy consumption per bit. Moreover, the RL-based power controller enjoys simplicity in design and a high level of adaption to environment.

In Chapter 6, we also deal with the non-tradeoff problem in WBANs. Using the Lagrangian multiplier method, we obtain the optimal solution and then propose an iterative approach based on the Jacobi method for fixed-point calculations to attain the optimal solution in a distributed manner. It should be clarified that the utilization of fixed-point calculation for distributively solving optimization problems has been well studied in the literature before (see the book [9] for detailed study and [10] for a survey). However, we adopt and apply the Jacobi method to our problem because it allows WBAN to *asynchronously* solve the problem while other methods such as Gauss-Seidel do not hold this property.

[1] Asynchronous Distributed Pricing Power Controller

1.5 Thesis Contribution

The main contributions of this thesis can be attributed as follows.

1. We employ transmission power control to increase the energy efficiency of WBANs as well as to enhance their reliability for medical applications by mitigating inter-network interference between different neighbouring WBANs operating in the same frequency band.

2. We propose a fuzzy power controller in WBANs to provide a tradeoff between throughput and power, and we develop a genetic algorithm and a learning mechanism to optimally design the fuzzy power controller to maximize throughput by reducing inter-network interference between different WBANs and at the same time reducing power consumption as much as possible. (Chapter 3)

3. We develop a power controller based on game theory to provide a tradeoff between throughput and power consumption. The proposed power control game utilizes a penalty on increasing power to motivate WBANs to achieve the maximum throughput with as little power for transmission possible. We analyze the linear and quadratic forms of the considered penalty function and prove the existence and uniqueness of the NE in the game. A best response strategy will be proposed for WBANs to reach the NE. (Chapter 4)

4. In order to adapt to dynamic changes in terms of channel conditions and power budget of WBANs, we propose adaptive pricing mechanisms for the power control game which increases the price of power for WBANs with bad channel conditions or low power budgets. The interference from such WBANs will reduce, allowing WBANs with good channels condition and power budgets to raise their power, resulting in increased system capacity. (Chapter 4)

5. We develop a power controller for WBANs based on reinforcement learning (RL) which allows WBANs to learn from experience and improve their performance thereby being able to accommodate dynamic changes of the environment adaptively and reduce energy consumption as much as possible. (Chapter 5)

6. We use radial basis function approximators to manage the issue of large state-action space and reduce the complexity of the learning process in the proposed RL-based power controller. (Chapter 5)

7. We investigate the effects of reward function and other parameters, including discount factor, learning rate and eligibility trace parameter on the performance of the RL-based proposed power controller. (Chapter 5)

8. We evaluate the performance of the RL-based power controller while employing different reinforcement algorithms including Q-learning and sarsa from the *value*

iteration algorithm category, as well as OLSPI[1] [11] from the *policy iteration algorithm* category. Their performances in terms of optimality of the solution and convergence are compared to each other and also against a counterpart approach based on game-theory without learning. (Chapter 5)

9. We model our power control problem as an optimization problem which minimizes the total power consumption in the system while keeping individual target rates of WBANs satisfied and obtain the optimum (but centralized) solution of the optimization problem using the Lagrangian multipliers method. (Chapter 6)

10. We develop a fully distributed version of the centralized solution based on fixed-point calculations using Jacobi method. The distributed approach approximates the centralized optimum solution and allows WBANs to find a solution asynchronously. (Chapter 6)

11. We scrutinize the performance of all the proposed approaches using extensive simulations and compare them against each other in terms of throughput, transmission power, interference power, energy consumption per bit, network lifetime and convergence. (Chapters 3, 4, 5 and 6)

12. Unlike other works which use the channel model in the ISM band for WBANs, we consider the MICS frequency band.

13. All the proposed power control approaches in this thesis rely only on local information and allow WBANs to find a solution independently and asynchronously. Moreover, they do not need any cooperation or message exchange between WBANs, which is highly favorable in WBANs with medical applications.

1.6 Thesis Outline

Chapter 2 is devoted to a review of the 802.15.6 standard and also presents related work on power control existing in the literature. A fuzzy power controller optimized by a genetic algorithm called WFPC[2] is proposed in Chapter 3 and its performance is evaluated and compared to a well-cited power controller, namely ADP. The system model and simulation framework described in this chapter will be also used for the rest of the thesis. Chapter 4 introduces a power controller based on game theory called WPCG[3] and presents the simulation results comparing its performance to ADP and the fuzzy power controller proposed previously, namely WFPC. In Chapter 5, we employ Reinforcement Learning (RL) and propose a highly adaptive power controller called WRLPC[4] and compare its performance to the previously proposed power controllers, namely WFPC and WPCG. Different RL algorithms are used and compared and also

[1] Online Least-Squares Policy Iteration
[2] WBAN Fuzzy Power Control
[3] WBAN Power Control Game
[4] WBAN RL-based Power Control

some design guidelines are presented. In Chapter 6, we formulate the power control with the approach of satisfying the target rate of each WBAN with at least power as possible. The centralized optimum solution and a distributed solution will be proposed. Finally Chapter 7, concludes the thesis and presents some open problems and future work.

"An expert is a person who has made all the mistakes that can be made in a very narrow field."

Niels Bohr

2

Related Work and Literature Review

In this chapter, we first look over some aspects of the IEEE 802.15.6 standard which may potentially affect power control because of the factors such as channel model, MAC superframe structure and energy conservation. Afterward, an extensive review on power control schemes existing in the literature will be presented and their suitability for WBANs will be examined.

2.1 IEEE 802.15.6 Standard

In this section, the components of the IEEE 802.15.6 standard which may be related to power control will be presented. This includes an overview of the frequency bands used by WBANs, the physical layer, the MAC layer, power management and interference mitigation techniques.

2.1.1 Frequency Bands

The frequency bands which are available for WBANs include Human Body Communications (HBC), Medical Implant Communication Service (MICS), Wireless Medical Telemetry Service (WMTS) and Ultra Wide Band (UWB). The Industrial, Scientific and Medical (ISM) band can be also used by WBANs. However, there are high chances of co-channel interference from other devices which operate in the ISM band such as IEEE 802.11 and IEEE 802.15.4. Although the ISM band is adopted worldwide, as Figure 2.1 shows, different countries use different frequencies in the WMTS and UWB

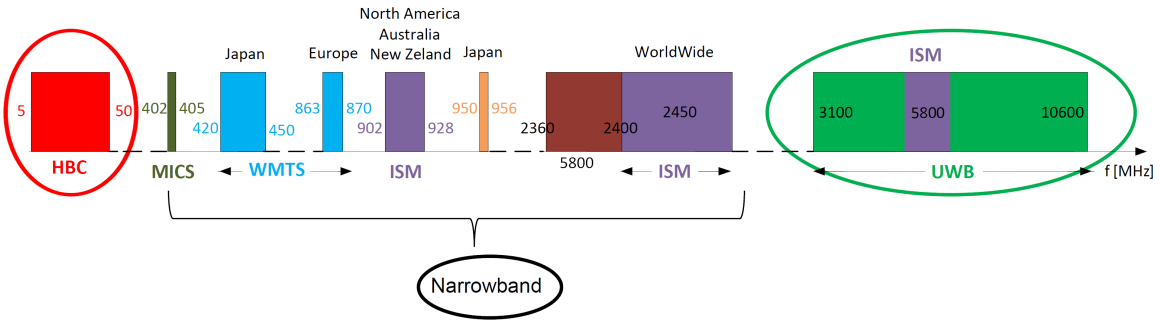

FIGURE 2.1: The frequency bands of WBANs in different countries

bands.

Human Body Communications (HBC)

HBC spans the frequencies between 5 MHz and 50 MHz and uses the human body as a medium to transmit signals. There are two methods for HBC which are electric field coupling and electromagnetic coupling. In the former method, there is one electrode placed on the human body, the signal return path is coupled by the near electric field. By means of the latter method, there are two electrodes attached on the human body, which is treated as a waveguide to propagate RF signals.

Medical Implant Communication Service (MICS)

The FCC[1] and ETSI[2] allocated the MICS band to enable WBAN to deliver a high level of comfort, mobility and better patient care. MICS is an ultra-low power, unlicensed, mobile radio service at 402-405 MHz with 300 kHz channels for short-range data transmission (up to 10 m) to support diagnostic or therapeutic functions associated with implanted medical devices (in-body communications). This frequency range and bandwidth allows for 10 non-overlapping channels. The MICS band permits individuals and medical practitioners to utilize ultra-low power medical implant devices, such as cardiac pacemakers and defibrillators, without causing interference to other users of the electromagnetic radio spectrum.

In addition, the 402-405 MHz frequencies have propagation characteristics conducive to the transmission of radio signals within human body and do not pose a significant risk of interference to other radio operations in that band [12]. The MICS band is located at an optimum frequency range that promises a high level of integration

[1]USA Federal Communications Commission
[2]European Telecommunications Standards Institute

Table 2.1: WMTS Channel Specifications

	Frequency (MHz)	Bandwidth (kHz)	Transmit Power (dBm)	Range
Band 1	608-614	25 or 50	> 1.8 and ≤ 10	100
Band 2	1395-1400	25	> 1.8 and ≤ 10	100
Band 3	1427-1432	25	> 1.8 and ≤ 10	100

with the use of advanced RFIC[1] technology. This results in miniaturization and low-power consumption. While higher frequency causes higher penetration loss, high-level integration becomes difficult at low frequencies. Moreover, there exists relatively insignificant penetration loss at these frequencies (10 dB with 10 mm tissue penetration) [12]. Additionally, a small antenna design is also difficult at lower frequencies such as HBC. Combining all these features with the availability of the 402-405 MHZ band internationally offers an attractive frequency choice for the targeted WBAN applications.

Wireless Medical Telemetry Service (WMTS)

WMTS defined by FCC is a service for data collection in medical applications and has a longer distance range than MICS which is up to 100m [13]. It is however used only for non-implantable devices (on-body communications). WMTS is split into three bands of 6MHz bandwidth with each band divided into 25KHz sub-channels. The first band can also support 50kHz subchannels. Table 2.1 presents some technical specifications of WMTS.

Ultra Wide Band (UWB)

UWB has been defined by FCC and ITU-R[2] in terms of a transmission from an antenna for which the emitted signal bandwidth exceeds the lesser of 500 MHz or 20% of the center frequency. The unlicensed use of UWB lies in the range of 3.1 GHz to 10.6 GHz. UWB enables WBANs to support real-time parameter measurement and can provide high data rate transfer (up to 10Mbps) for on-body communications. The advantages of UWB include low interference generation, resistance to multipath and low transceiver complexity as well as low transmission power levels in an order of those used in the MICS band, namely -16 dBm.

[1] Radio Frequency Integrated Circuits

[2] International Telecommunication Union (ITU) Radiocommunication Sector

2.1.2 Physical Layer

The IEEE 802.15.6 supports three different physical layers (PHYs) , which are Narrow-Band (NB), UWB, and Human Body Communications (HBC). In the following, the specifications of these PHYs are briefly described. For more detailed study, the reader is referred to [14].

Narrow-band PHY

The NB PHY is responsible for activation/deactivation of the radio transceiver, CCA[1] within the current channel and data transmission/reception. The Physical Protocol Data Unit (PPDU) frame of the NB PHY contains a Physical Layer Convergence Procedure (PLCP) preamble, a PLCP header, and a PHY Service Data Unit (PSDU) as seen in Figure 2.2.

The PLCP preamble helps the receiver in the timing synchronization and carrier-offset recovery. It is the first component being transmitted. The PLCP header conveys information necessary for a successful decoding of a packet to the receiver. The PLCP header is transmitted after the PLCP preamble using the given header data rate in the operating frequency band. The last component of a PPDU is a PSDU which consists of a MAC header, MAC frame body, FCS[2] and is transmitted after the PLCP header using any of the available data rates in the operating frequency band which can be either of MICS, WMTS or ISM. In the NB PHY, the standard uses DBPSK[3], DQPSK[4], and D8PSK[5] modulation techniques, except in the MICS band which uses only GMSK[6].

Human Body Communications PHY

The HBC PHY operates in two frequency bands centered at 16 MHz and 27 MHz with the bandwidth of 4 MHz. Both operating bands are valid for the United States, Japan, and Korea, and the operating band at 27 MHz is valid for Europe. HBC is the EFC[7] specification of PHY, which covers the entire protocol for WBAN such as packet structure, modulation, preamble/SFD, etc. Figure 2.3 describes the PPDU structure of EFC that is composed of a preamble, SFD, PHY header and PSDU. The preamble and SFD are fixed data patterns. They are pre-generated and sent ahead of the packet header and payload. The preamble sequence is transmitted four times in order to ensure packet synchronization while the SFD is transmitted only once. When

[1] Clear Channel Assessment
[2] Frame Check Sequence
[3] Differential Binary Phase-Shift Keying
[4] Differential Quadrature Phase-Shift Keying
[5] Differential 8-Phase-Shift Keying
[6] Gaussian minimum shift keying
[7] Electrostatic Field Communication

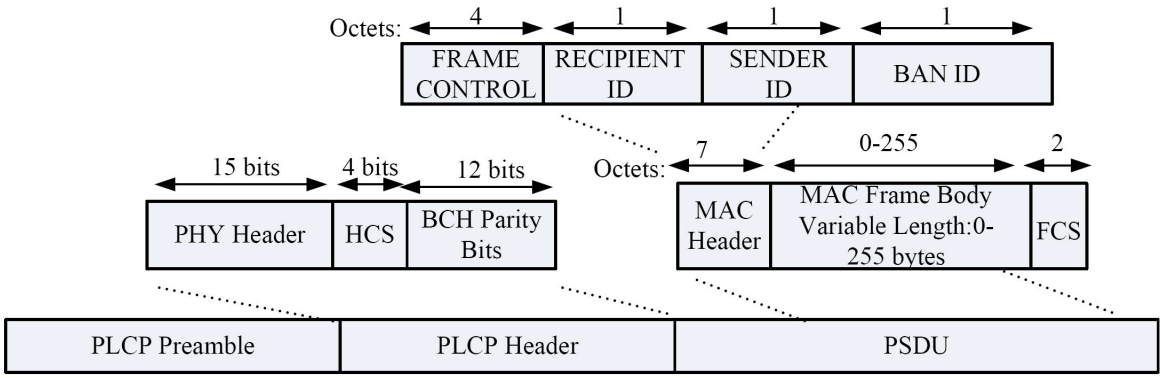

Figure 2.2: IEEE 802.15.6 NB PPDU structure [1]

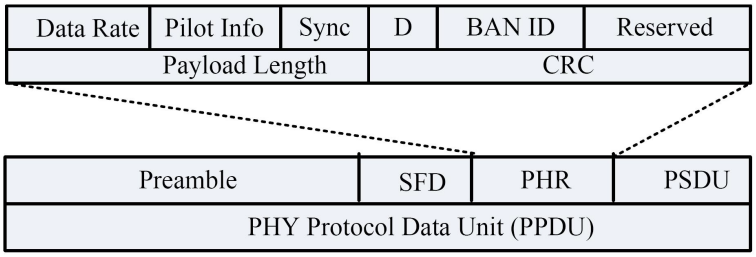

Figure 2.3: IEEE 802.15.6 EFC PPDU structure [1]

the packet is received by the receiver, it finds the start of the packet by detecting the preamble sequence, and then it finds the start of the frame by detecting the SFD.

Ultra Wide-band PHY

Figure 2.4 shows the UWB PPDU that contains a Synchronization Header (SHR), a PHY Header (PHR), and PSDU. The SHR is composed of a preamble and a Start Frame Delimiter (SFD). The PHR conveys information about the data rate of the PSDU, length of the payload and scrambler seed. The information in the PHR is used by the receiver in order to decode the PSDU. The SHR is formed of repetitions of Kasami code [15] of length 63. Typical data rates range from 0.5 Mbps up to 10 Mbps with 0.4882 Mbps as the mandatory one.

UWB PHY operates in two frequency bands: low band and high band. Each band is divided into channels with the bandwidth of 499.2 MHz. The low band consists of three channels (1-3) only. The channel 2 has a central frequency of 3993.6 MHz and is considered a mandatory channel. The high band consists of eight channels (4-11) where channel 7 with a central frequency 7987.2 MHz is considered a mandatory channel, while all other channels are optional. A typical UWB device should support at least one of the mandatory channels.

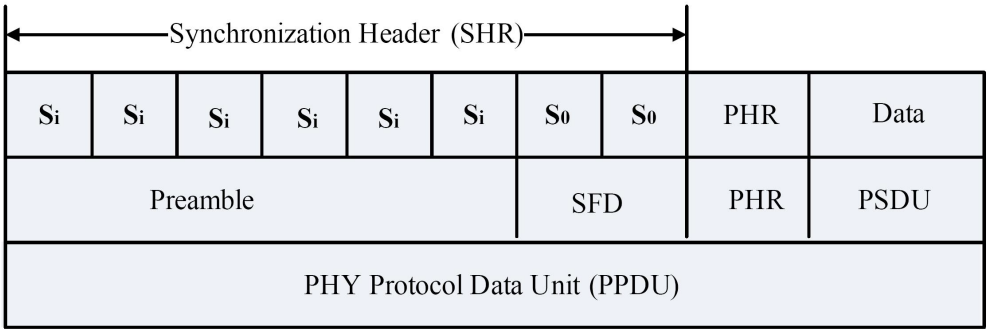

FIGURE 2.4: IEEE 802.15.6 UWB PPDU structure [1]

2.1.3 MAC Layer

In IEEE 802.15.6, the entire channel is divided into superframe structures. Each superframe is bounded by a beacon period of equal length. The BNC selects the boundaries of the beacon period and thereby selects the allocation slots. The BNC may also shift the offsets of the beacon period. Generally, the beacons are transmitted in each beacon period except in inactive superframes or unless prohibited by regulations such as in MICS band. A WBAN operates in one of the following modes.

1. *Beacon mode with beacon period superframe boundaries*: In this mode, the beacons are transmitted by the BNC in each beacon period except in inactive superframes or unless prohibited by regulations. Figure 2.5 shows the superframe structure of IEEE 802.15.6, which is divided into Exclusive Access Phase 1 (EAP1), Random Access Phase 1 (RAP1), Type I/II phase, Exclusive Access Phase 2 (EAP 2), Random Access Phase 2 (RAP 2), Type I/II phase, and a Contention Access Phase (CAP). In EAP, RAP and CAP periods, nodes contend for the resource allocation using either CSMA/CA or a slotted Aloha access procedure. The EAP1 and EAP2 are used for highest priority traffic such as reporting emergency events.

The RAP1, RAP2, and CAP are used for regular traffic only. The Type I/II phases are used for uplink allocation intervals, downlink allocation intervals, bilink allocation intervals, and delay bilink allocation intervals. In Type I/II phases, polling is used for resource allocation. Depending on the application requirements, the coordinator can disable any of these periods by setting the duration length to zero.

2. *Non-beacon mode with superframe boundaries*: In this mode, the entire superframe duration is covered either by a type I or a type II access phase but not by both phases.

3. *Non-beacon mode without superframe boundaries*: In this mode, the coordinator provides unscheduled Type II polled allocation only. The access mechanisms used in each period of the superframe are divided into three categories:

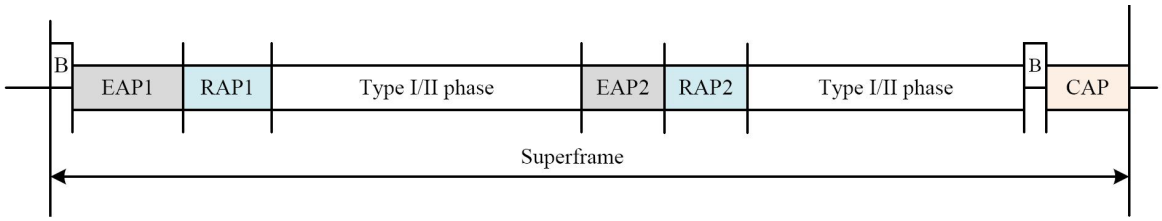

Figure 2.5: IEEE 802.15.6 superframe structure

- Random access mechanism, which uses either CSMA/CA or a slotted Aloha procedure for resource allocation.
- Improvised and unscheduled access (connectionless contention-free access), which uses unscheduled polling/posting for resource allocation.
- Scheduled access and variants (connection-oriented contention-free access), which schedules the allocation of slots in one or multiple upcoming superframes, also called 1-periodic or m-periodic allocations.

For a detailed study of the MAC layer, we refer the reader to the IEEE 802.15.6 standard [1].

2.1.4 Power Management

The power management techniques try to save energy by switching off sensor nodes when they are not supposed to send or receive. The sensor nodes in a WBAN are not always in the active state and may hibernate during a number of the entire beacon periods (superframes), and may also sleep over some time intervals even in its wake-up beacon periods.

The IEEE 802.15.6 standard has introduced two techniques for power management which are *hibernation* and *sleeping*. In the following we overview these techniques. For more detailed study, the reader is referred to the standard in [1].

Hibernation

Hibernation is referred to as a state for a sensor node without receiving or transmitting any traffic over one or more superframes. On the other hand, awake is referred to receiving or/and transmitting frames in every beacon period. To hibernate, the node shall set the *Wake-up Period* field in its last *Connection Request* frame to an integer larger than 1, while setting the *Wake-up Phase* field in the frame to a value specifying its intended next wake-up beacon period. To wake up, the node shall set the *Wake-up*

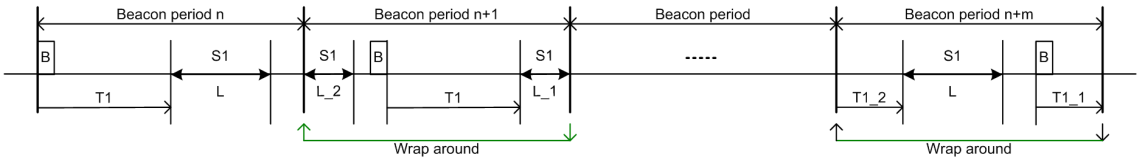

FIGURE 2.6: 1-periodic hibernation allocation [1]

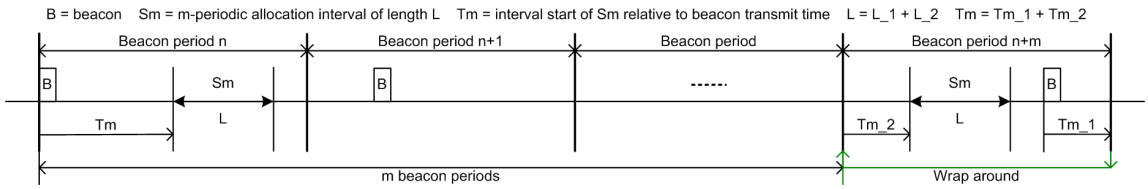

FIGURE 2.7: m-periodic hibernation allocation [1]

Period field in its last *Connection Request* frame to 1, while setting the Wake-up Phase field in the frame to a value identifying the next beacon period.

The intended recipient BNC of the *Connection Request* frame responds by a *Connection Assignment* frame. If the BNC sets the Wake-up Period field in its responding frame to an integer larger than 1, it may grant only m-periodic allocations to the node, with the allocation intervals being in the nodes wake-up beacon periods, in accordance with the nodes last *Connection Request* whenever possible, but shall not grant to the node any 1-periodic allocations. Likewise, if the BNC sets the Wake-up Period field to 1, it may grant only 1-periodic allocations to the node and shall not grant any m-periodic allocations. Figure 2.6 and 2.7 show the 1-periodic and m-periodic allocations respectively.

If the Wake-up Period value in the *Connection Assignment* frame last received from the BNC is larger than 1, the node shall wake up in each of its wake-up beacon periods based on the latest Wake-up Period and Wake-up Phase values provided in that frame by the BNC, to transmit or/and receive frames in the granted m-periodic allocation intervals, and to receive the beacon if needed. On the other hand, if the Wake-up Period value is 1, the node shall wake up in every beacon period, to transmit or/and receive frames in the granted 1-periodic allocation intervals, and to receive the beacon if appropriate. Figure 2.8 represents the hibernation mechanism of the macroscopic power management across beacon periods.

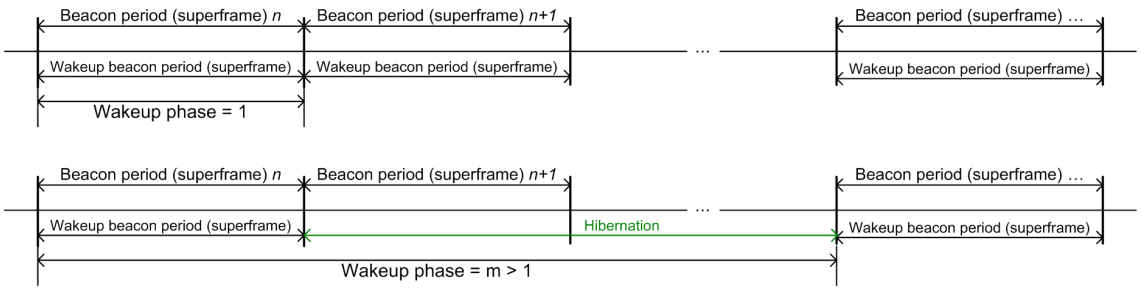

Figure 2.8: Hibernation mechanism [1]

Sleeping

Apart from the hibernation, a sensor node may also sleep —without receiving or transmitting any traffic— during a superframe, except over the following time intervals.

- The node shall wake up to receive a beacon from the BNC when it needs a beacon reception to synchronize with the BNC or to obtain certain information contained in a beacon.
- The node shall wake up to receive and transmit frames in its scheduled allocations in its wake-up beacon periods.
- The node shall stay active participating in frame transactions in its expected posted allocations. The BNC should arrange to have the posted allocations of a node to occur in the nodes wake-up beacon periods, if possible. If the node did not receive a frame at the announced time for a pending post, it should stay in receive mode until the BNC could have finished a frame transaction for the post and retransmitted a frame pSIFS later unless it needs to make a turnaround to transmit mode.
- If the node has indicated its support for polls through its MAC Capability field of its last Connection Request frame, it shall also stay active in such times as to receive announced polls and initiate frame transactions in its polled allocations. The BNC should arrange to have the polled allocations of a node to occur in the nodes wake-up beacon periods, if possible.

2.1.5 Interference Mitigation

The IEEE 802.15.6 standard has introduced two techniques for interference mitigation which are *beacon shifting* and *channel hopping*. In the following we overview these techniques. For more detailed study, the reader is referred to the standard in [1].

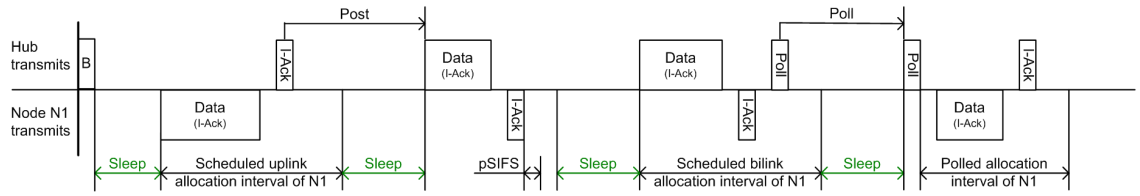

FIGURE 2.9: Sleeping mechanism [1]

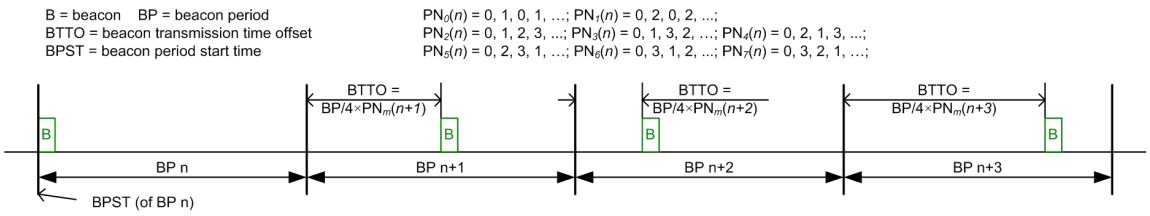

FIGURE 2.10: Beacon Shifting [1]

Beacon Shifting

A BNC may transmit its beacons at different time offsets relative to the start of the beacon periods by including a *Beacon Shifting Sequence* field in its beacons. A BNC should choose a beacon shifting sequence that is not being used by its neighbor BNC to mitigate potential repeated beacon collisions and scheduled allocation conflicts between overlapping or adjacent BANs operating on the same channel.

As shown in 2.10, the BNC shall transmit a beacon at a time $BTTO = \mathrm{PN}_m(n) \times \mathrm{BP}/4$ relative to the start of beacon period n. Here, PN_m is a pseudo-random beacon shifting sequence, m is the beacon shifting sequence index that the BNC has chosen for its WBAN, BP is the length of its beacon period, and n is the phase of the chosen sequence ($n = 0, 1, \cdots$) for this beacon period.

The allocation slots in a beacon period shift around with the beacon transmit time. The access phases (EAP1, RAP1, EAP2, RAP2, and CAP) are referenced to numbered allocation slots and shift around with the beacon in the beacon period accordingly. The RAP1 Length and RAP2 Length fields contained in the beacon of the current beacon period now refer to RAP1 and RAP2 in the next beacon period.

The BNC shall ensure in choosing these access phases and the beacon shifting sequence that beacon shift does not result in a split of any of the aforementioned access phases into two parts.

Scheduled allocation intervals are also referenced to numbered allocation slots and shift around with the beacon transmit time accordingly in the beacon period. A scheduled allocation interval in a beacon period may be split into two portions as a result of shifting around the beacon period, but the aggregate length remains the same.

Channel Hopping

A BNC may change its operating channel periodically by including the *Channel Hopping State* and *Next Channel Hop* fields in its beacons. The channel hopping sequence selected by the BNC must not be being used by its neighbor BNCs.

The BNC can not hop to a new channel in the middle of a beacon period and before hopping to another channel, it must dwell on the current channel for a fixed number of beacon periods as communicated to the nodes connected with the BNC through Connection Assignment frames.

If required for regulatory compliance, a BNC shall select a channel based on the applicable regulatory requirements, and may dwell on the channel for an indefinite period of time. To communicate this selection to nodes, the BNC shall set the channel hopping state to 0, which makes the connected sensor nodes to dwell on the current channel for an indefinite period of time as well.

The channel hopping sequences are generated by the BNC using the maximum-length Galois LFSR[1]. Given the current channel number, the next channel number will be chosen such that the difference between the two numbers is greater than a threshold parameter, *pChannelSeparation*, which is the minimum number of channels separated between two consecutive hops. The details of channel hopping sequences generation is beyond the scope of this thesis and the interested reader is referred to [1].

2.2 Transmission Power control

Transmission Power control (TPC) has been exploited for various objectives in various types of wireless networks. In the following, we examine power control schemes existing in the literature which have been proposed for non-WBANs and investigate their suitability for WBANs. This will be followed by summarizing the power control schemes currently found in the literature for WBANs. The network types we will explore consist WLANs[2], ad-hoc and WSNs[3] and cellular networks. Each of the selected networks has an analogy to WBANs in some way which will be pointed out in its own section. We also examine the existing power control schemes based on the methodology they utilized. The methodologies of interest comprises fuzzy control, game theory and

[1]Linear Feedback Shift Register
[2]Wireless Local Area Networks
[3]Wireless Sensor Networks

reinforcement learning.

2.2.1 Power Control in non-WBANs

Wireless Local Area Networks

In WLANs, power control is mainly employed to mitigate co-channel interference between adjacent APs[1] and thereby increasing spatial channel utilization.

In some works e.g. [16] [17] [18] [19], the authors suggested modifying the MAC layer to incorporate power control. The key idea is to send the RTS/CTS packets at the maximum power and use a minimum required transmission power for the DATA/ACK transmissions. However such approaches can not be applied to WBANs due to the lack of per data packet RTS/CTS control in WBANs' communication model and fundamental differences between the MAC layers of WLANs and WBANs.

In [20], the authors proposed a power control mechanism for WLANs taking into account factors including SINR, path loss and bit error rate as well as the dynamic changes in bit error rate by using a Markov chain model which includes good and bad states. They derived the minimum power needed to successfully transmit one bit at a certain bit error rate for various MAC packet lengths, and extended their scheme to support multiple power levels. Although they achieved good performance in terms of energy efficiency, their power control mechanism is a centralized approach and is specifically developed for WLAN channel model. A fully distributed power control mechanism for WLANs proposed by the authors in [21] where all APs cooperate to reduce their transmission power simultaneously until the point where it is impossible to improve the utility of at least one AP. Nevertheless their approach requires negotiation and cooperation between APs belonging to different WLANs.

There are also some works in WLAN which involve jointly optimizing power control and another parameter in the network such as channel assignment e.g. [22] [23], rate control e.g. [24] [25], scheduling e.g. [26], and session admission control e.g. [27] [28].

The reader is referred to [29] for further study on power control in WLANs.

Ad-hoc and Sensor Networks

In ad-hoc networks including MANETs[2] and WSNs, power control mostly deals with topology control [30] [31]. The multi-hop structure of packet delivery in ad-hoc networks obligates adjacent nodes to stay connected with each other. The dilemma is the tradeoff between network connectivity and energy efficiency: increasing power to avoid network from being partitioned or decreasing power to save energy.

[1] Access Points
[2] Mobile Ad-hoc NETworks

Some papers including [32] [33] [34] perform power control based on the number of neighboring nodes. In [32], the authors proposed a heuristics power control in WSNs where each node tries to keep a predefined number of neighbors. If the number of neighbors is less than the predefined threshold value, the transmitter will increase its power by a certain factor. They compared their algorithm to a centralized approach and showed a 50% improvement in network lifetime. However, they considered fixed nodes in their system model and ignored all spatial and temporal channel variations. Moreover their approach requires negotiation and cooperation between nodes in the network which does not suit WBANs.

El-Batt et. al. in [33] proposed a power control mechanism to adjust the tradeoff between reducing transmission power and enhancing throughput in ad-hoc networks. They investigated the effect which power reduction has on the increase in the number of intermediate hops. Each node broadcasts beacon messages at the maximum power level to discover its neighbors and builds up a connectivity table. The average received power level is used to pick up the nearest nodes having the highest average values. Each node then adapts its transmission power level for each of its direct neighbors. The problem with their approach is scalability as each node in the network has to store the global network topology information.

Taking advantage of the RTS/CTS or SYNC packets has been the idea of some power control schemes e.g. [35] and [36] in ad-hoc networks. Wu et al. in [35] proposed a power control scheme in mobile ad-hoc networks where the main idea is to use the RTS/CTS packets before transmitting data packets in an effort to determine the relative distance between two communicating nodes and then adjust transmission power level based on the estimated distance. Although their approach remarkably decreases co-channel interference, it needs an exact model of the channel and path loss for the distance estimation to be accurate. Moreover the lack of per packet RTS/CTS control in WBANs makes all such approaches inapplicable in WBANs.

There are also some works in ad-hoc networks e.g. [37] and [38] which employ node localization for computing link distance, which is subsequently used for adjusting power levels using a known channel model. However, in WBANs, performing localization for on-body sensor nodes is not feasible due to complexity and cost. The channel in WBAN greatly varies with body posture changes such as walking. Moreover, to perform localization, an additional transmission medium such as ultrasound is needed to perform TDOA[1] computation which is very costly in practice.

A well-cited power control algorithm, called Asynchronous Distributed Pricing Power Controller (ADP), is proposed by J. Huang in [39], where each user announces a price that reflects the compensation paid by other users for their interference. The authors present an asynchronous distributed algorithm for updating power levels and prices. Their approach is capable of finding the optimal power allocation which maximizes the utility $u_i(\gamma_i(\mathbf{p})) = \log(\gamma_i(\mathbf{p}))$ summed over all users, where $\gamma_i(\mathbf{p})$ is the SINR of user i. This problem is given by

[1] Time Delay of Arrival

$$\max_{\mathbf{p}} \sum_{i} u_i(p_i, p_{-i}) \tag{2.1}$$

$$\text{variable } \mathbf{p} : p_i \in [P_i^{\min}, P_i^{\max}] \tag{2.2}$$

The authors define a parameter as $\pi_i(p_i, p_{-i}) = -\partial u_i(p_i, p_{-i})/\partial I_i(p_{-i})$, which represents user i's marginal increase in utility per unit decrease in the total interference $I_i(p_{-i})$, given all other users' power levels p_{-i}. Each user i then maximizes the difference between its utility minus its payment to the other users affected by interference, i.e.

$$p_i = \arg\max_{p_i} \left\{ \log(\gamma_i(\mathbf{p})) - p_i \sum_{j \neq i} \pi_j h_{ji} \right\} \tag{2.3}$$

where h_{ji} is the channel gain from user i to user i.

At each iteration, the price π_i is updated according to the following equation and announced to other users in the network.

$$\pi_i = \frac{\partial u_i(\gamma_i(\mathbf{p}))}{\partial \gamma_i(\mathbf{p})} \frac{(\gamma_i(\mathbf{p}))^2}{B p_i h_{ii}} \tag{2.4}$$

where B is the bandwidth in Hz.

However the approach suffers from two main drawbacks. Firstly, users need to cooperative to announce their price updates to all other users in the network, which does not suit the systems in which users are reluctant to cooperate. Secondly, their approach does not rely only on local information and for example needs the adjacent channel gains, namely h_{ji}, to be known, which implies message exchange between users. However the approach suffers from two main drawbacks. Firstly, users need to cooperative to announce their price updates to all other users in the network, which does not suit the systems in which users are reluctant to cooperate. Secondly, their approach does not rely only on local information and for example needs the adjacent channel gains, namely h_{ji}, to be known, which implies message exchange between users.

The dynamic adjustment of transmission power in ad-hoc networks also affects routing link selection and has been well investigated in the literature bringing about numerous power-aware routing protocols [40] [41]. Although, there have been some studies on cooperative and relay communications using multi-hop links within a WBAN e.g. in [42] [43] or between different WBANs [44], network connectivity and routing are not pertinent in medical applications, where WBANs are not allowed to communicate with each other and are supposed to perform their vital health-care tasks. This leads to non-cooperative WBANs with single-hop links within each WBAN. Nevertheless, in some gaming and military applications WBANs may share their interests and need

to inter-communicate or even cooperate. Such applications are not however of our interest.

There are some good survey papers including [45] [46] and [47] on power control in ad-hoc and sensor networks, to which we refer the reader for further study.

Cellular Networks

Transmission power control has been studied widely in the context of cellular networks where it is employed for both uplink, from Mobile Station (MS) to Base Station (BS), and downlink, from BS to MS, although it is far more important and challenging in the uplink due to the mobility and energy limitation of MSs. The uplink power control problem is often attributed as maximizing a utility function of throughput or minimizing power consumption at MSs subject to a constraint on SINR.

Some works e.g. [48] [49] proposed closed-loop transmission power control schemes, where a separate feedback channel with universal frequency for control data is used. However, the lack of a separate feedback channel in a WBAN makes such approaches unsuitable. Additionally, as it is mentioned in [50], the time constants for power control in the cellular networks are much larger than what are needed in WBANs causing those closed-loop mechanisms to be too slow in the presence of high postural mobility in WBANs.

There exist numerous papers in the literature including [51] [52] [53] which try to solve the uplink power control problem centralizedly at BS. The optimum power allocation is calculated at BS and constantly instructed to the MSs in the cell. However, due to the lack of a central arbiter in WBANs, such approaches are not applicable.

Addressing the problem distributively at MSs, however, has been the goal of multiple works in the literature amongst which Distributed Power Control (DPC) is a key power control algorithm proposed by Foschini and Miljanic in [54] and has been further studied afterward in several papers including [55] [56] [57]. DPC is a heuristic power control in which each link attempts to continuously maintain its target SINR by overcoming the interference imposed by all the other links using as low power for transmission as possible. Although DPC was proposed heuristically, it is very efficient and is proved by Mitra et al. in [58] to be asynchronously convergent with geometric rate to the Pareto optimal power allocation. DPC is later extended by Bambos et al. in [59] and [60] to allow the links to trade off delay toleration for power conservation when the interference is high, which can lead to improvement in total throughput and power consumption.

A considerable amount of literature has been published on power control in cellular networks. For further study, we refer the reader to [61] [62] and references therein.

Due to the fundamental differences between WBANs and non-WBANs in terms of structure and requirements, the existing power control approaches in non-WBANs can

not be directly applied to WBANs.

2.2.2 Power Control in WBANs

Since WBAN is a very new-born technology, the literature on power control in WBANs is quite sparse. In the following, we summarize the existing power control schemes proposed in WBANs for energy conservation.

In [63], S. Xiao et al. presented an optimal transmission power control that minimizes energy usage subject to lower-bounds on the link quality, namely RSSI. Their optimal transmit power scheme is based on off-line calculations and impractical assumptions which require the sender to have a-priori knowledge of the link quality at the receiver. For practical scenarios, however, they proposed two simple online power control schemes called conservative and aggressive power control schemes which trade off reliability for energy savings by changing transmit power based on feedback information from the receiver. Their empirical results show that conservative scheme preserves reliability and yet reduces energy consumption by 9% on average when compared to using maximum transmit power, while the aggressive scheme saves 25% more energy on average, at the expense of slightly increased loss. However, the proposed approaches are very trivial and suffer from being too far from the optimal transmit power control. On the other hand, the optimal power control proposed which relies on the brute-force search is quite vague and it is unknown that how it converges to a stable solution when it is done by multiple neighboring WBANs simultaneously, and even if it converges, whether the final solution is still optimal or just sub-optimal.

M. Quwaider et al. in [64] modeled human body movements as a stochastic linear system and utilized a LQGI[1] to predict RF signal strength which then was used for regulating the RSSI of the receiver node at a fixed reference level for an on-body link. It was shown that power assignment with quantized LQGI model and small weight factor can provide lower error and energy performance compared with the search based strategies. Their approach is however very costly in terms of processing and memory usage which does not suit the tiny sensor nodes. Moreover, power levels can be very sensitive to prediction errors which may hinder its use in practice.

Smith et al. in [65] presented a power control scheme for WBANs based on channel predictions. They proposed a long-term predictor for WBAN channel which is accurate for up to 2 seconds, even with a nominal channel coherence time of 500ms. The predictor utilizes the partial-periodicity of measured WBAN channels and weights an alternate least-squares estimate for the desired prediction interval using the last 4s of received signal. Their approach shows concurrent improvements in reliability and power consumption in comparison to some typical WBAN transmission strategies with no channel prediction. However, they did not consider co-channel interference from nearby WBANs into account and their proposed power control scheme is only based

[1] Linear Quadratic Gaussian control with an Integrator

on body movements which cause partial periodicity in WBAN channel.

B. Moulton et al. in [66] extended Xiao's work and proposed a power control protocol which adaptively adjusts the period between each feedback transmission to accommodate run-time variation in the quality of channel. Their simulation results show that the adaptive approach improves the power savings compared to the full power with no feedback by 21%.

2.2.3 Power Control by Methodology

Fuzzy Control

Fuzzy control utilizes fuzzy logic, which incorporates qualitative linguistic variables, to control a system. The system is modeled by a number of fuzzy sets and fuzzy rules which are used by an inference engine to make decisions. The input of a fuzzy controller is a non-fuzzy value (called a *crisp* value) which is first fuzzified to produce a fuzzy input and then is used by the inference engine to produce a fuzzy output, which is finally defuzzified into a crisp value. This involves the utilization of membership functions and fuzzy operators to determine the degree of membership for the fuzzy input, consequences of fuzzy rules and the fuzzy output. For an introduction to fuzzy control, please refer to chapter 4.

Since linguistic variables are used to model a system in fuzzy control, an exact mathematical model of the system is not necessary which makes the design of the controller quite simple. In addition to having simplicity and flexibility in the controller design, fuzzy systems have shown a great ability to control complicated systems and have also been broadly employed in the literature for the purpose of power control. In the following we summarize some of the key related work which employ fuzzy logic for power control.

Sabitha et al. [67] replaced the currently existing common-range maximum transmission power at the MAC layer of the IEEE 802.15.4 with the concept of dynamic and adaptive transmission power control at link level. Various parameters like Link Quality Indicator (LQI), Received Signal Strength Indicator (RSSI) and MAC collisions are considered and fuzzified, then optimal transmission power levels are chosen based on the following algorithm:

- Step 1: Find RSSI from the physical layer. Store it in 'RSSI' variable.
- Step 2: Packetise and send it to MAC layer
- Step 3: In the MAC layer, find LQI, Source MAC address and the status of the frame (normal, corrupted or collided). Store them in the variables LQI, MACSRC and ERR respectively.
- Step 4: Calculate the average values of RSSI and LQI over a time period, i.e., 5 seconds and store it in variables ARSSI and ALQI respectively. Also store the total number of error frames in TOTERR.

- Step 5: Check if the packet type is DATA or ACK.
- Step 6: Based on the accumulated values of TOTRSSI and TOTLQI, calculate Average RSSI and Average LQI.
- Step 7: Decide the transmit power using the following fuzzy rule base. In case of Frame Errors > 2, increase the power level one step higher.
- Step 8: Repeat the above steps for every 5 seconds.

Zhang et al. in [68] proposed a transmission power adjustment scheme using the fuzzy control in WSNs for topology control by dynamically controlling the degree (number of neighboring nodes). All nodes start with the same initial power level. Each node acts as a controller and periodically broadcasts a message (Msg) including its unique identity. All other nodes, which receive such a Msg, reply with a feedback acknowledge message (FBMsg) including the identity of the Msg sender. Before the node issues the next Msg, it counts the number of FBMsgs received in current period, namely T_d in current period. If the error between T_d and E_d, the expected degree, is within a bound on error, e, the node converged and does not change its transmission power any more. Otherwise, the node runs the fuzzy control law to adjust its power again, and continues to broadcast its Msg. Although their approach improves network lifetime, it is a closed-loop power controller which need a feedback control channel to be existing between the nodes in the network.

Xia et al. in [69] considered the tradeoff between power consumption and packet delay in WSNs. At high interference conditions, deferring the transmission of packets which leads to experiencing longer delay. They proposed a fuzzy power controller to determine a threshold SINR used by sensor nodes to decide whether to send a packet or not. Average delay and distance of a node to the source node are the inputs for the controller. The output of the controller provides adjusting factor for the SINR threshold. Their simulation results indicate the proposed fuzzy controller can reduce the average delay by up to 28%.

Lakshmi et al. in [70] proposed a power controller to reduce interference in WSNs. The output of the fuzzy controller is the transmission power level and its inputs are end-to-end delay and RSSI. These parameters are fuzzified and optimal transmission power levels are calculated for each node by the fuzzy controller.

In [71], Jiang et al. presented a peer to peer fuzzy power controller in WSNs. The fuzzy controller adjusts transmission power adaptively based on diverse receiving QoS parameters. The inputs of the power controller are LQI, RSSI, SINR and and the output is a power adjust value which is fed back to the transmitter.

Some other power control works in wireless communications using fuzzy control include [72] [73] proposing power reduction algorithms to select cluster heads in WSNs using a fuzzy controller, [74] [75] presenting power efficient routing protocols in ad-hoc and sensor network using a fuzzy logic approach, [76] [77] [78] proposing fuzzy-based opportunistic spectrum access strategies in the Cognitive Radio Networks (CRN) which consider interference caused by CR links to the Primary Users (PUs) and enable the

CR links to select an optimal spectrum band and transmit at an optimal power.

Game Theory

Game theory is the formal study of decision-making agents known as players where their choices potentially affect the interests of the other players. A review on game theory can be found in chapter 5.

In wireless communications, game theory has been employed mostly to solve resource allocation problems in a competitive environment. Mobile nodes in a wireless system suffering from a limited transmission resource, i.e. energy and radio spectrum that imposes a conflict of interests. In an effort to resolve this conflict, they can make certain choices such as changing their transmission power level (power control), transmitting now or later (scheduling), choosing a beacon sequence (beacon shifting), changing their transmission channel (channel hopping), or adapting their transmission rate and modulation (AMC[1]).

Koskie et al. in [79] formulated the uplink power control problem in CDMA networks as a non-cooperative game in which users choose to trade off between SINR error and transmission power usage. That is, minimizing the SINR error at the cost of high transmission power usage. The cost function they have selected is:

$$J_i(p_i, \gamma_i) = b_i p_i + c_i(\gamma_i^{tar} - \gamma_i)^2 \tag{2.5}$$

where b_i and c_i are constant weighting factors, p_i is transmit power, γ_i and γ_i^{tar} are the SINR and the target SINR respectively. Using the best response method, the Nash equilibrium power p_i^* is given as:

$$p_i^* = \frac{2c_i}{b_i}\gamma_i^*(\gamma_i^{tar} - \gamma_i^*) \tag{2.6}$$

where γ_i^* is the Nash equilibrium SIR as:

$$\gamma_i^* = \begin{cases} \gamma_i^{tar} - \frac{b_i}{2c_i g_{ii}}\left(\frac{g_{ii}p_i^*}{\gamma_i^*}\right) & \text{if non-negative} \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

The authors have proposed distributed power control strategies based on the Newton iterations to accelerate the convergence of the static Nash power control algorithm. Their Newton iteration is of third-order rather than quadratic which appears to better eliminate the slight overshoot observed in early iterations. Their simulation results indicate that the use of Newton iterations notably improves convergence. A realistic CDMA cell model has been used to simulate the proposed algorithms. However, the CDMA cell model requires centralized arbiter and thus procedure on how independent users were assigned codes and with what power level was not discussed by the authors.

[1] Adaptive Modulation and Coding

In [80], Meshkati et al. proposed a game-theoretic approach for energy-efficient power control in multicarrier CDMA systems. The authors formulated power control problem as a non-cooperative game in which each user decides how much power to transmit over each carrier to maximize its own utility function being as follows:

$$u_k^{MC} = \frac{\sum_{l=1}^{D} T_{kl}}{\sum_{l=1}^{D} p_{kl}} \tag{2.8}$$

where T_{kl} is the throughput achieved by user k over the lth carrier, and is given by $T_{kl} = \frac{L}{MR_k f(\gamma_{kl})}$ with γ_{kl} denoting the received SINR for user k on carrier l; L and M are the number of information bits and the total number of bits in a packet, respectively; R_k is the transmission rate for the kth user; and $f(\gamma_{kl})$ is the efficiency function representing the packet success rate (PSR), i.e., the probability that a packet is received with no error.

The authors have showed that for all linear receivers including matched filter, decorrelator, and MMSE[1] detector, the utility function is maximized when the user transmits only on its *best carrier* which is the carrier that requires the least amount of power to achieve a particular target SINR at the output of the receiver. They derived the conditions on the channel gains for a Nash equilibrium to exist. The authors also characterized the distribution of users among the carriers at equilibrium and presented an iterative and distributed algorithm for reaching the equilibrium. Their approach results in significant improvements in the total utility achieved at equilibrium compared to a single-carrier system and also to a multicarrier system in which each user maximizes its utility over each carrier independently. However, the proposed technique trades-off complexity for optimality and thus efficient power consumption is hard to guarantee.

In [81], a game-theoretic power control in MIMO[2] ad-hoc networks has been proposed. The power allocation at each user is built into a non-cooperative game where the utility function is as follows:

$$u_l = C_l - \gamma_l p_l \tag{2.9}$$

where γ_l is a non-negative scaling factor, C_l is the channel capacity of link l and p_l is the transmission power of link l. Due to poor channel conditions, some users have very low data transmission rates even though their transmit powers are high. Therefore, a mechanism for shutting down such users is proposed in order to reduce co-channel interference and improve energy-efficiency. On the other hand, if the capacity of a particular link is more than enough to maintain a certain level of QoS, reducing the capacity by decreasing transmission power will mitigate the interference impose to

[1]Minimum Mean Square Error
[2]Multi Input Multi Output

other links. These two mechanisms are controlled via γ_l as follows:

$$\gamma_l = \begin{cases} \frac{\alpha_l}{p_0} & C_{l_0} \geqslant C_l^t \\ \infty & C_{l_0} < C_l^t \end{cases} \tag{2.10}$$

where α_l is a certain capacity value and p_0 is the initial transmit power; C_{l_0} is the initial multiuser water-filling capacity of link l and C_l^t is a capacity threshold assigned to link l by the external network controller.

The decision of whether to shut down a particular link depends on the minimum data rate C_l^t that is required by that link. This threshold is adaptively determined by the type of service in which the link is involved as well as the overall channel conditions which relate the QoS level to the threshold.

Compared to multiuser water-filling and gradient projection methods (e.g. [80]), the proposed game-theoretic approach with the link user shut-down mechanism allows the MIMO ad hoc network to achieve a higher energy saving and a higher system capacity.

Thomas et al. in [82] presented a cognitive network approach to achieve the objectives of power and spectrum management. The authors cast the problem as a two phased non-cooperative game and used the properties of potential game theory to ensure the existence of, and convergence to, a desirable Nash Equilibrium. The utility function they have used is:

$$u_i^{PC}(\mathbf{p}) = M f_i(\mathbf{p}) - p_i \tag{2.11}$$

where $\mathbf{p}$ is the transmission power vector; p_i is the transmission power of radio i; $f_i(\mathbf{p})$ is the number of the radios that can be reached (possibly over multiple hops) by radio i via bidirectional connections and paths. The scalar benefit multiplier M indicates the value each radio places on being connected to other radios; and it is assumed $M \geq \max_i \{p_{\max_i}\}$.

The authors proved that the game with this utility function is an OPG[1] with the global function as follows:

$$V^{PC}(\mathbf{p}) = M \sum_{i \in N} f_i(\mathbf{p}) - \sum_{i \in N} p_i \tag{2.12}$$

The authors showed that this selfish cognitive network constructs a topology that minimizes the maximum transmission power while simultaneously using, on average, less than 12% extra spectrum, as compared to the global optimum solution.

[1] Ordinal Potential Game

Closas et al. in [83] employed non-cooperative game theory to design a fully distributed network topology control algorithm in WSNs using optimal transmission adjustments. Their utility function is as follows:

$$u_i(p_i, p_{-i}) = \begin{cases} p_{\max_i} - p_i & \text{if network is connected} \\ -p_i & \text{otherwise} \end{cases} \tag{2.13}$$

The authors proved that the game is an EPG[1] with the following global function:

$$V(p_i, p_{-i}) = \begin{cases} p_{\max_i} - \sum_{i \in N} p_i & \text{if network is connected} \\ -\sum_{i \in N} p_i & \text{otherwise} \end{cases} \tag{2.14}$$

Their simulation results shows that for a relatively low node density, the proposed game leads to a connected network almost with probability one.

Huang et al. in [84] made use of game theory and proposed two auction mechanisms, SINR auction and power auction, that determined relay selection and relay power allocation respectively in a distributed fashion. For both single-relay networks and multiple-relay networks, the power auction achieves the efficient allocation by maximizing the total throughput, and the SNR auction is flexible in trading off fairness and efficiency. Users iteratively update and submit their bids based on the best response using the following update rule:

$$b(t+1) = F^s(\pi)b(t) + f^s(\pi)\beta \tag{2.15}$$

where $b(t+1)$ and $b(t)$ are the next and current bid vectors respectively, $F^s(\pi)$ is a $N \times N$ matrix with (i,j)th component being $f_i^s(\pi)$, and $f^s(\pi) = [f_1^s(\pi), f_2^s(\pi), \cdots, f_N^s(\pi)]'$; $\beta > 0$ is the reserve bid and $\pi > 0$ is the price.

Alpcan et al. in [85] proposed a non-cooperative power control game in CDMA networks with a utility function based on the outage probability, i.e. the probability that the SINR level of the mobile user is greater than a predefined individual threshold level. Having proved the uniqueness of the Nash equilibrium for a class of uniformly strictly convex pricing functions, the authors established the global convergence of continuous-time as well as discrete-time synchronous and asynchronous iterative power update algorithms to the unique NE of the game under some conditions. They also considered the uncertainty due to quantization and estimation errors and a proposed a stochastic version of the discrete-time synchronous update scheme which almost surely converged to the unique NE point.

Xing et al. in [86] put forward a stochastic learning solution for distributed discrete power control game in wireless data networks. They proposed two probabilistic power adaptation algorithms and analyzed their theoretical properties along with the numerical behavior. Their approach has been later formulated by Wang et al. in [87]

[1] Exact Potential Game

as a general-sum game in which each player evaluates a power strategy by computing a utility value. This evaluation was performed using a stochastic iterative procedure. The authors approximated the discrete power control iterations by an equivalent ODE[1] and proved that the proposed stochastic learning power control algorithm converges to a stable Nash equilibrium. The drawback is that the convergence times may be too long relative to the packet duration.

The literature is quite rich in game-theoric power control in wireless networks. For further study on this area, the reader is referred to [88], [89] and [90].

Reinforcement Learning

Reinforcement Learning (RL) is a form of machine intelligence which an agent can use to learn an optimal policy to achieve a given goal. The agent is ignorant of the environment model and performs trial-and-error interactions with the environment to find out the immediate reward resulted by the action taken at the current state of the environment, which will then be used to estimate the potential long-term reward for any state of the environment. A review on reinforcement learning can be found in chapter 6.

RL is a broad branch of machine learning and has attracted the attention of many scientists over the last few years. The utilization of RL for power control in wireless networks, however, is quite intact and the area has not been well investigated by the researchers in this area. In the following, we summarize the existing related power control works employing RL.

Pandana et al. in [91] employed reinforcement learning to maximize the average throughput per total consumed energy in WSNs by choosing the optimal modulation level and transmission power while adapting to the incoming traffic rate, buffer condition, and the channel condition. The reward function which they chose is the number of successfully transmitted packets per total consumed energy as follows

$$r = \frac{L_b}{L} \cdot \frac{R \cdot m \cdot S(\Gamma(\gamma, p_t), m)}{L \cdot p_t} \tag{2.16}$$

where L_b is the information carried by one packet in bits; L is the number of bits in the packet after adding error coding; R is the transmission rate in $bits/s$. γ is the current received channel gain fed back from receiver to the transmitter; m and p_t denote the modulation level and the transmission power respectively. $S(\Gamma(\gamma, p_t), m)$ is the probability of successful packet reception, where $\Gamma(\gamma, p_t)$ is the target SINR.

However, the authors kept the space-action space too small and did not consider the curse of dimensionality issue which simply occurs in large space-action spaces. Also, the computation of the optimal policy is vague and not addressed clearly. Moreover,

[1] Ordinary Differential Equation

they assumed an error-free feedback channel existing between transmitter and receiver which may not be feasible in practice.

In [92], the authors propose a distributed power control for SUs[1] to manage the interference at the receivers of the PUs[2] in a CRN[3]. They modeled it as a multiagent system where the multiple agents are the different secondary base stations in charge of controlling the secondary cells. Agents are independent learners which use Q-learning to find the optimal policy. The cost function considered is as follows:

$$c = (SINR_i^{(t)} - SINR^{Th})^2 \tag{2.17}$$

where $SINR_i^{(t)}$ is the instantaneous SINR in the control point of cell i, and $SINR^{Th}$ is the threshold SINR which should be reached.

The Q-learning employed aims to minimize this cost so that the SINR at the control points is $SINR^{Th}$, which guarantees that interference at the primary receivers is below the threshold. For the generalization of the state space, the authors employed neural networks to approximates the Q-functions. However, it needs off-line training. To take the error of the observation of the current state into account, the authors utilize POMDP[4] which makes use of a state estimator to compute the agents belief state as a function of the old belief state, the last action, and the current observation that the agent makes of the environment. This, however, adds complexity to the system thereby trading off optimality for the increased computation load.

In [93], Jiandong et al. presented a power control in CRNs to enhance the performance of secondary users in terms of the ratio of the spectrum efficiency to the power consumption level, as well as to improve the fairness among the SUs. The utility function considered is as follows:

$$c_i = \frac{\log(1 + SINR_i(t))}{p_i(t)} \tag{2.18}$$

The authors modeled the interaction among the agents and wireless environment as a Markovian game-theoretic and formulated the spectrum sharing issue among multiple secondary users as an expected utility maximization problem. Employing Q-learning, each SU can well obtain the fair and optimal power control strategy. However, their approach does not address the curse of dimensionality problem.

[1] Secondary Users
[2] Primary Users
[3] Cognitive Radio Network
[4] Partially Observable Markov Decision Process

"We are what we repeatedly do."

Aristotle

3

Rate-Power Tradeoff - Genetic-Fuzzy Approach

In this chapter, we begin with the system model description followed by the motivation for the tradeoff between rate and power in WBANs and also derive the utility function needed to achieve this. We then propose a fuzzy power controller, called WFPC[1], which makes a tradeoff between throughput and power by adjusting the transmission power level to mitigate inter-network interference between neighboring WBANs. The controller can be looked at as a decision-making system which makes decisions on the next transmission power level based on the current levels of the SINR, interference power and the current transmission power level which is fed back to the controller from the controller output. We utilize a genetic algorithm to attain the optimum design of the controller. This optimization takes place offline, i.e. at the design stage, and ensures the controller maximizes a utility function of throughput with a cost to penalize increasing power level. We compare the performance of WFPC to a literature well-cited power controller, namely ADP[2] [39]. Although, the offline genetic algorithm optimization required by WFPC can limit the controller adaptation to dynamic changes of the surrounding environment, it notably improves the performance of the controller as the simulation results illustrate that WFPC achieves a lower energy consumption per bit as well as a faster convergence compared to ADP.

[1] WBAN Fuzzy Power Controller

[2] Asynchronous Distributed Pricing Power Controller

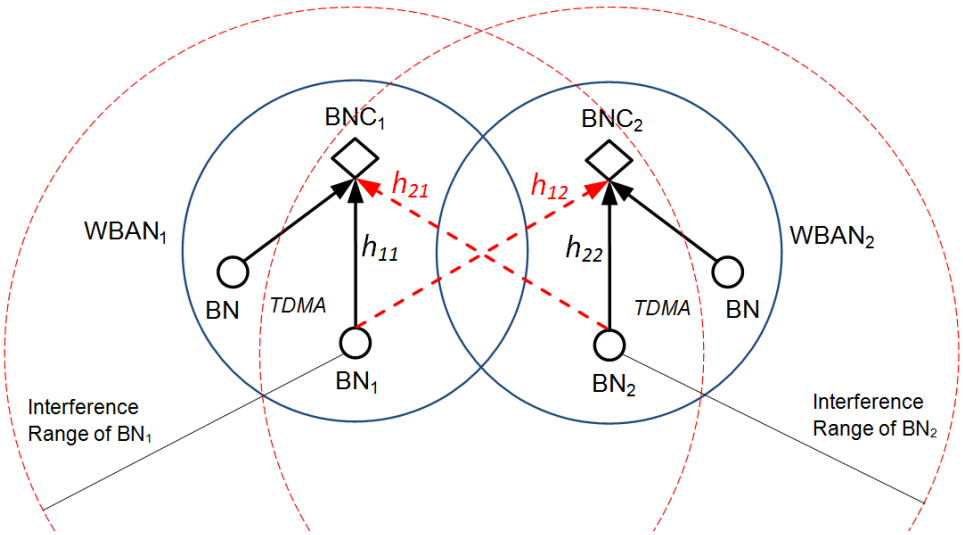

FIGURE 3.1: The System Model; as seen in the interference model of the system, while the signals of the BNs collide at the BNC of neighboring WBANs, there is no interference between BNs and their associated BNC within each WBAN, as they employ an orthogonal MAC communication scheme such as TDMA, as defined in the standard of IEEE 802.15.6

3.1 System Model

In this section, we elaborate the system model which will be used by all the proposed approaches in the thesis.

We consider a system of m WBANs denoted by set $M = \{1, ..., m\}$ operating in the same frequency channel where their transmission ranges overlap causing co-channel interference on each other. The transmissions considered within each WBAN are from BNs to BNCs, as the traffic flow is mostly of this direction in medical applications. We assume that WBANs rely on the MAC layer defined in the standard IEEE 802.15.6 [94], implying that there is no intra-network collisions between the BNs within a single WBAN. This interference model is shown in Figure 3.1.

We consider that each WBAN in our system has one instance of TPC algorithm running on its BNC node that determines the transmission power levels which the sensor nodes should use for the next MAC superframe. The sensor nodes are informed of their next transmission power levels by the BNC in the beacon period of each superframe. Running the power controller on BNCs —which are receivers in our system— rather than on the sensor nodes has two important advantages as follows: Firstly, bearing in mind that the sensor nodes are very resource-constrained in terms of battery, memory and processing power, keeping them away from running algorithms will save their resources remarkably. On the other hand, BNC nodes can be quite big like a PDA with powerful processors and enough memory to run complicated algorithms. Plus, their battery can be easily recharged or replaced. Secondly, since some power

controllers like the ones we will propose require some information at the receiver such as SINR, if the power controller was running on sensor nodes —which are transmitters in our system, we would need error-free feedback channels from BNC to each sensor node to carry back such information which can be quite costly in practice due to the extra power, delay and bandwidth needed for feedback channel coding.

From the system point of view, the task of the power controller is to allocate power levels to the sensor nodes across all WBANs to achieve a performance utility. For simplicity and without any loss of generality, we assume that there is only one sensor node within each WBAN and we use $\mathbf{p} = (p_i)_{i=1}^{m}$ to denote the power vector in the system, where $p_i \in \mathcal{P}_i = [0, P_{\max_i}]$ is the transmission power level of the sensor node in WBAN i.

The throughput of a WBAN is assumed to be the maximum data rate achievable by its sensor node in the channel from the sensor node to the BNC in the WBAN and is given by the Shannon channel capacity formula as follows

$$c_i = B \log_2(1 + \xi_i) \tag{3.1}$$

where c_i in bit per second (bps) is the throughput of WBAN i, B is the channel bandwidth in Hz and ξ_i is SINR at BNC in WBAN i given by:

$$\xi_i = \frac{h_{ii} p_i}{\sum_{j \neq i}^{m} h_{ji} p_j + n_i} \tag{3.2}$$

where p_i and p_j are the transmission power levels of BN i and BN j respectively; n_i is the thermal noise power over the entire channel bandwidth at BNC i. h_{ii} and h_{ji} are the elements of the power gain matrix $\mathbf{H}$ given by

$$\mathbf{H} = \begin{bmatrix} h_{11} & \dots & h_{1m} \\ \vdots & \ddots & \vdots \\ h_{m1} & \dots & h_{mm} \end{bmatrix} \tag{3.3}$$

where h_{ii} is the power gain between BN i and its corresponding coordinator, namely BNC i; h_{ji} is the power (interference) gain between BN j (transmitter) and BNC i (receiver).

It should be noted that we here abuse the notation $\mathbf{H}$, which is usually used for denoting channel matrix. It is, however, used here to denote the power gain matrix which can be thought of as the square of the norm ($\|\cdot\|^2$) of the channel matrix element-wisely. It therefore has all real-valued elements here.

3.2 Tradeoff Utility Motivation

In this thesis, we aim to make a tradeoff between throughput and power by using a utility function which encourages WBANs to achieve a higher throughput with less

power. To motivate such a utility function, consider the following optimization problem, which is the maximization of the total throughput, i.e. throughput summed over all WBANs in the system

$$\begin{aligned} &\max \sum_{i=1}^{m} c_i \\ &\text{subject to } \; 0 \leq p_i \leq P_{\max_i} \; , \forall i \\ &\text{variables } \; \mathbf{p} = (p_i)_{i=1}^{m} \end{aligned} \tag{3.4}$$

where c_i is the throughput of WBAN i given by Eq. 3.1.

This optimization problem is non-convex due to the non-convexity of throughput c_i, and the global optimum solution can not be attained analytically. As a baseline distributed approach, however, consider decomposing this problem into decoupled subproblems, where each WBAN chooses its transmission power level to maximize its own individual throughput. Since the throughput of each WBAN i is strictly increasing with its transmission power p_i (for fixed power levels of all other WBANs), provided that WBANs do not cooperate or exchange messages to solve their problems, the unique solution of the system is $\mathbf{p}^* = (P_{\max_i})_{i=1}^{m}$, i.e., each WBAN uses its maximum power to transmit, which is a very aggressive solution from the view point of internetwork interference and power consumption. One technique to prevent WBANs from aggressively raising their power is punishing them by setting a penalty for increasing transmission power. To this end, we define the following individual utility function

$$U_i(p_i, w_{p_i}) = \frac{c_i}{C_{\max_i}} - w_{p_i}(\frac{p_i}{P_{\max_i}})^{\alpha_i} \tag{3.5}$$

where p_i is the transmission power level of WBAN i; c_i is the throughput of WBAN i; w_{p_i} is a price factor which can be used to adjust the tradeoff between throughput and power; α is the price exponent which as we will see later in chapter 5, imposes sufficient conditions for existence of a stable solution as well as affecting the behavior of the power controller; $C_{\max_i}$ is the maximum channel capacity achievable by WBAN i at zero interference; and $P_{\max_i}$ is the maximum allowable transmission power of WBAN i. We normalize the values of throughput and power using $C_{\max_i}$ and $P_{\max_i}$ for the price factor to have meaningful values;

The considered utility function rewards WBANs for increasing their throughput and penalize them for increasing their power. The devised penalty mechanism prevents them from increasing their power uselessly and producing interference to others. This motivates them to achieve a higher throughput with less power and enables them to achieve a tradeoff between throughput and power, which can be controlled by using the price factor.

If the problem in (3.4) was convex, the optimum price factor vector $\mathbf{w}_p^* = (w_{p_i}^*)_{1=i}^{m}$

could be calculated as follows:

$$\begin{aligned}&\mathbf{w}_p^* = \min_{\mathbf{w}_p} U(\mathbf{p}^*, \mathbf{w}_p)\\&\text{subject to } w_{p_i} \geq 0\ , \forall i\\&\text{variables } \mathbf{w}_p\end{aligned}$$

where $\mathbf{p}^* = (p_i^*)_{1=i}^m$ and p_i^* is the optimum power calculated by each WBAN independently (and of course in a distributed manner).

Since $U_i(p_i, w_{p_i})$ is differentiable with respect to w_{p_i}, the optimum price factors could also be calculated distributedly using the gradient method iteration:

$$w_{p_i}(t+1) = [w_{p_i}(t) + \delta(p_i(t) - P_{\max_i})]^+ \tag{3.6}$$

where $0 < \delta < 1$ is a step size.

However, since the optimization problem is non-convex, there will be a dulaity gap and the price factor calculated as above will not be optimum which will cause the attained power allocation $\mathbf{p}^*$ not to be the global optimum of the primal problem. In fact due to the non-convexity, it is not possible to attain the global optimum solution in a distributed manner and the decoupled sub-problems will finally converge to a sub-optimal solution. In Chapter 4, we propose adaptive methodes to adjust the price factor.

3.3 Genetic Fuzzy Systems

A fuzzy control system [95] is a decision making machine which employs fuzzy logic [96] to make decisions for controlling a system. Fuzzy logic introduced by Zadeh [97] models the uncertainty expressed by the use of linguistic variables such as "High", "Medium", "Low", "Most", "Many", "Seldom", etc., similar to the way the human brain makes reasoning. While classical set theory requires an element to be either included by a set or not, fuzzy sets define intermediate values, known as *the degree of membership*, which allow an object to be a partial member of a set. For example, a person may be a member of the set short to a degree of 0.7, or the temperature of a room can be 70% freezing and 30% cold. Figure 3.2 shows sample fuzzy sets for describing temperature, which uses trapezoidal membership functions for "Freezing" and "Hot" fuzzy sets, and triangular membership functions for "Cool" and "Warm" fuzzy sets.

Fuzzy control is known as art of controlling using words. The dynamic behavior of the system to be controlled is attributed by a number of linguistic fuzzy rules based on the knowledge of a human expert. Fuzzy rules are of the general form: IF [*antecedents*] THEN [*consequents*], where antecedents and consequents are propositions containing linguistic variables. Antecedents of a fuzzy rule form a combination of fuzzy sets

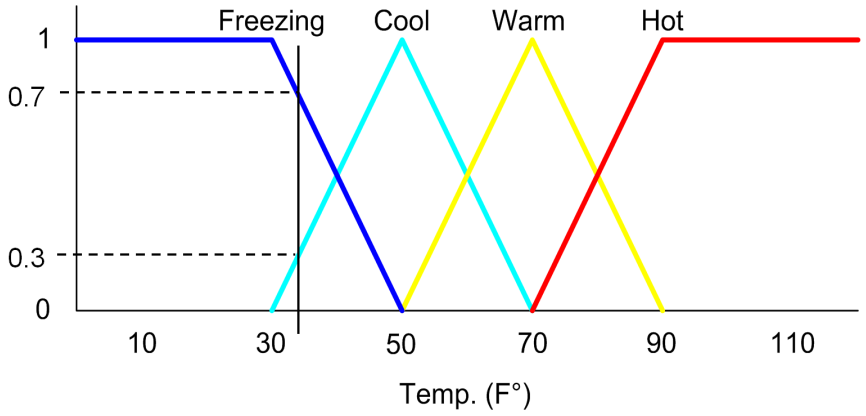

FIGURE 3.2: Fuzzy sets for describing the temperature of a room; the temperature can be a member of the "Freezing" set to a degree of 0.7 and a member of the "Cool" set to a degree of 0.3.

through the use of logic operations consisting AND, OR and NOT. Thus, fuzzy sets and fuzzy rules together form the *knowledge base* of a rule-based inference system as shown in Figure 3.3. Antecedents and consequents of a fuzzy rule form fuzzy input space and fuzzy output space respectively, which are defined by the combinations of fuzzy sets. Non-fuzzy (also known as *crisp*) inputs are scaled and mapped to their fuzzy representation in a process called *fuzzification*. This involves the utilization of membership functions such as Gaussian, triangular and trapezoidal. The inference engine maps the fuzzified inputs to the rule base to produce a fuzzy output which involves determining the consequent of rules and its membership to each output fuzzy set. Finally, the fuzzy output is *defuzzified* and scaled into a crisp value.

Since in fuzzy control, qualitative linguistic variables are used to define the system behavior, an exact mathematical model of the system is not necessary, and this makes the design of the controller quite simple and flexible. Fuzzy controllers enjoy the advantages of robustness, ease of design, simplicity, and flexibility. Besides, fuzzy controllers have shown a great ability to control nonlinear systems and gracefully map complicated relationships between input and output spaces over the last two decades [98]. In this chapter, we make use of the fuzzy logic and propose a fuzzy power controller, namely WFPC, which takes inter-network interference between neighbouring WBANs into consideration to maximize throughput aiming to use as little power for transmission as possible.

Although one of the most useful features of a fuzzy control system is to incorporate human expert knowledge for the controller design sake, designing and tuning a fuzzy controller using an automated learning process such as genetic algorithms is preferred and has received extensive attention by researchers in the literature [99] [100] [101].

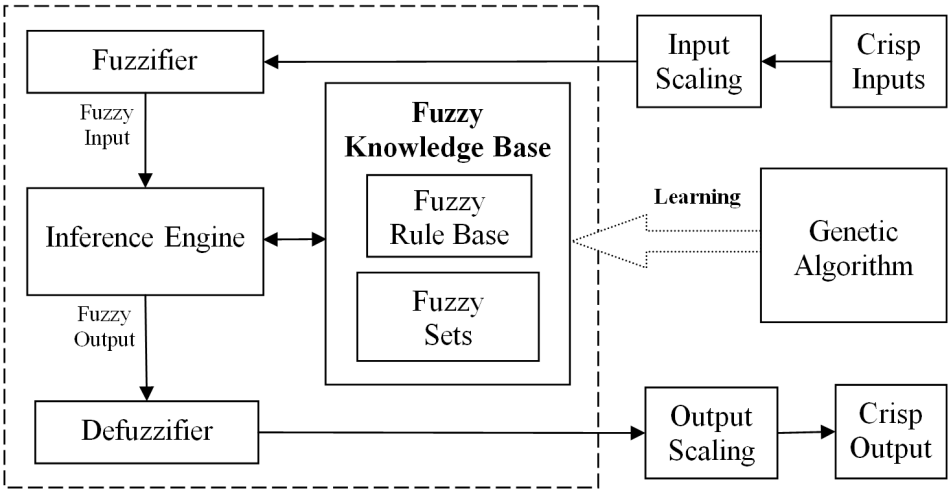

Figure 3.3: A Fuzzy control system; Genetic algorithms can be employed to design or tune the fuzzy controller by optimizing the fuzzy knowledge base

Genetic algorithms lie in the family of stochastic search algorithms based on Darwin's theory of evolution, where an objective function, called *fitness function*, is maximized. The inputs to the fitness function are the potential solutions to the problem coded as *chromosomes*. Each chromosome can be thought of as a point the search space. A genetic algorithm operates over a pool of such chromosomes, known as *population*, and randomly mates them using a *recombination* operator such as *crossover* to produce offspring called *individual*. The idea behind is that the new individual may be better than both of the parents if it takes the best genes from each of the parents.

The crossover operator randomly swaps some genes of the parents' chromosomes to make a new individual and can be carried out using different methods including N-point, uniform and arithmetic methods. In the N-point method, the crossover operator randomly chooses N points in the parents' chromosome which split them up into $N+1$ sections. Then it swaps the corresponding sections to make a new individual.

The individuals obtained constitute a *generation* which is then altered by a *mutation* operator where some genes in the chromosome are manipulated randomly. For example, with binary genes, this is done by flipping some bits, or with real-valued genes, by adding a small value.

The altered individuals are then *evaluated* by using the fitness function which quantifies the optimality of individuals. Afterward, some individuals from the current population are randomly chosen through a process called *selection* to form the mating pool for the next generation. The selection process can be done in various ways including roulette wheel selection, tournament selection and rank selection. Although the selection operator basically gives higher chances to the best individuals in the population to be chosen, there is always chances of losing the best individuals. To prevent this, a

few number of the best individuals in the current population are always selected first to be a part of the next generation. This process is called *elitism*.

The whole evolution process which involves recombination, mutation, evaluation, elitism and selection are repeated until a stop criterion, such as reaching a maximum number of generations, or finding a good enough individual, is fulfilled.

From the fuzzy system point of view, a genetic algorithm optimization is equivalent to parameterizing a fuzzy knowledge base, i.e. rules and membership functions, and to finding those parameter values that are optimal with respect to the design criteria (see Figure 3.3). The knowledge base parameters constitute the optimization space, which are transformed into a suitable genetic representation of chromosomes on which the genetic algorithm operates.

We employ a genetic algorithm to obtain the optimum design of WFPC. The optimization problem we get the genetic algorithm to solve is to design a fuzzy power controller which maximizes a utility function of throughput with a penalty on increasing power levels.

3.4 Proposed Approach

We will develop a fuzzy power controller, namely WFPC, to manage the effects of internetwork interference on throughput and power consumption. The output of WFPC is the next transmission power level $p(t+1)$ to be used by sensor nodes. WFPC makes decisions on transmission power based on the values of its inputs which are the current levels of the $SINR(t)$, interference power $p_I(t)$ [1], and transmission power $p(t)$, fed back from the output to the controller. As described earlier in the System Model section in chapter 1, in our system, the BNC nodes are responsible for running the power control algorithm within each WBAN. Both the SINR and interference power which are needed by WFPC to make decision on the transmission power level can be measured at a digital receiver and hence are available at BNC. The structure of WFPC is shown in Figure 3.4.

WFPC has three inputs indexed from 1 to 3 which represent $SINR(t)$, $p_I(t)$ and $p_T(t)$ respectively. We fuzzify each input by using K fuzzy sets corresponding to K linguistic terms. For example, for $K = 3$, these linguistic terms can be thought of as "Low", "Medium" and "High". For input i, the membership function (MF) corresponding to the fuzzy set j is denoted by $MF_{i,j}$ where $i \in [1, 3]$ and $j \in [1, K]$. The fuzzy output variable is however expressed by K_{out} fuzzy sets corresponding to K_{out} membership functions denoted by $MF_{o,j}$, where $j \in [1, K_{out}]$. A fuzzy rule then looks like as follows:

[1] This is actually interference power plus noise which can be measured at receivers, but to avoid cumbersome descriptions later on, it will be referred to as just interference power.

if $(SINR(t)$ is $MF_{1,i})$ and $(P_I(t)$ is $MF_{2,j})$ and $(P_T(t)$ is $MF_{3,k})$ then $\left(P_T(t+1)\right.$ is $\left.MF_{o,r_{i,j,k}}\right)$

where $r_{i,j,k} \in [1, K_{out}]$ is an integer valued number identifying the output fuzzy set for the fuzzy rule corresponding to fuzzy set i of input 1, fuzzy set j of input 2, and the fuzzy set k from input 3.

3.5 Genetic Algorithm Optimization

In this section, we obtain the optimum values for the parameters of membership functions as well as the fuzzy rules using GA.

3.5.1 Chromosome Structure

To codify the fuzzy controller as a chromosome, we consider that each chromosome is formed by two parts: *parametric genes* and *rule genes* that represent the fuzzy membership functions and the fuzzy rules, respectively. For the membership functions, we use the following Gaussian function

$$\mu_{ij}(x) = \exp\left(-\frac{(x - m_{ij})^2}{2\sigma_{ij}^2}\right) \tag{3.7}$$

where μ_{ij} is the membership function of the jth fuzzy set related to input i, and m_{ij} and σ_{ij} are the center position and spreading factor of the membership function. To be codified as genes, this function requires two real valued numbers, i.e. m_{ij} and σ_{ij}, bounded by the dynamic range of the corresponding input, as seen in Fig 3.5.

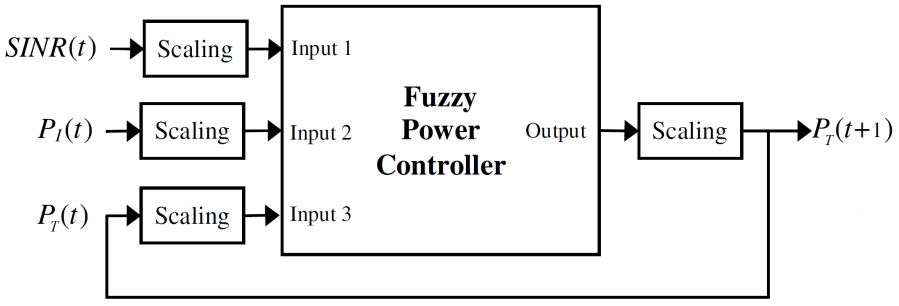

Figure 3.4: Structure of the WBAN Fuzzy Power Controller (WFPC); the SINR needed by WFPC can be measured at a digital receiver, i.e. the BNC nodes

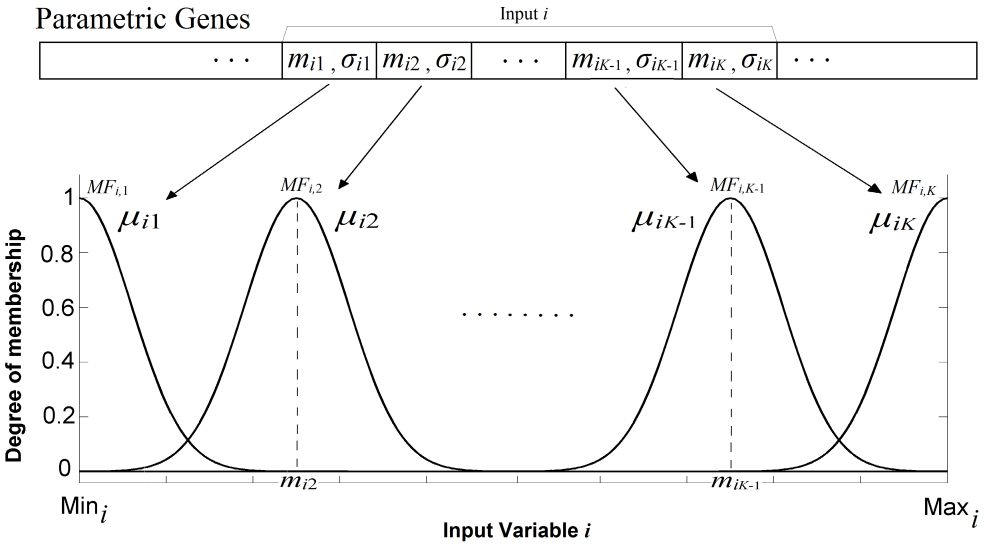

FIGURE 3.5: Membership functions and the corresponding genes; each membership function is codified using two real-valued parameters

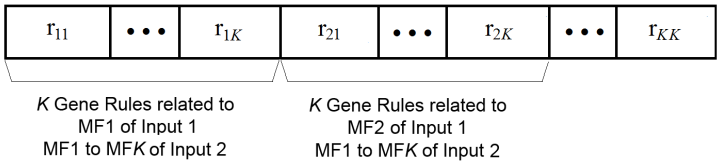

FIGURE 3.6: Rule genes in the chromosome structure

The chromosome part corresponding to the fuzzy rules is shown in Fig 3.6. There are K^2 rule genes each corresponds to a fuzzy rule which involves two fuzzy sets, one from each input.

3.5.2 Genetic Operators

Since we have real valued genes codifying the membership function parameters and integer valued genes codifying the fuzzy rules in the chromosome structure, we do not employ the usual binary code operators, i.e. N-point crossover and bit inversion for recombination and mutation respectively. Instead, for recombination, we use two crossover methods comprising arithmetic and heuristic crossovers being selected randomly with equal probabilities. In both methods, a linear combination of the parents'

corresponding genes is obtained based on the following equation

$$G_o = a.G_{P1} + (1 - a).G_{P2} \tag{3.8}$$

where G_o, G_{P1} and G_{P2} are the offspring's gene, the first parent's gene and the second parent's gene respectively; and a is a random number.

In arithmetic crossover, the offspring gene is an interpolation along the line formed by the parents' genes is performed ($0 < a < 1$), while in the heuristic crossover, it is an extrapolation outward in the direction of the better parent ($a < 0$ if P2 is better and $a > 1$ if P1 is better).

The mutation operator is uniformly selected from three methods which are Gaussian, uniform and non-uniform mutations. In the Gaussian mutation, the gene is changed with the probability of a normal Gaussian distribution. The other two methods change the value of the gene based on a uniform distribution and non-uniform distribution respectively in the specified range of the variable.

The selection strategy utilized to pick individuals into the mating pool to produce an offspring for the next generation is a ranking selection based on the geometric distribution. In the ranking selection, the individuals are first sorted according to their fitness values. Then the individuals are ranked based on their positions in the ordered list so that the rank 1 is assigned to the worst individual and N (the number of chromosomes in population) to the best individual. The ranking selection can lead to a more optimum controller design, although it may slows down the convergence to the optimum design. However, this convergence is not of importance in here, because this GA optimization is taking place offline.

3.5.3 Fitness Function

The genetic algorithm tries to find the individual maximizing a given fitness function which conveys the objectives of the problem at hand, which is maximizing throughput using as little power for transmission as possible. The fitness function we use to evaluate the optimality of individuals in our system is as follows.

$$f(p_i) = \frac{c_i}{C_{\max_i}} - w_{p_i} \frac{p_i}{P_{\max_i}} \tag{3.9}$$

3.5.4 Learning Strategy

We develop a new learning strategy for the system proposed by extending the existing learning process in GA. In addition to the elitism that takes place at each generation, namely *over-generation elitism*, we propose another elitism called *over-step elitism* in which after a certain number of generations, the WBANs move (according to a random

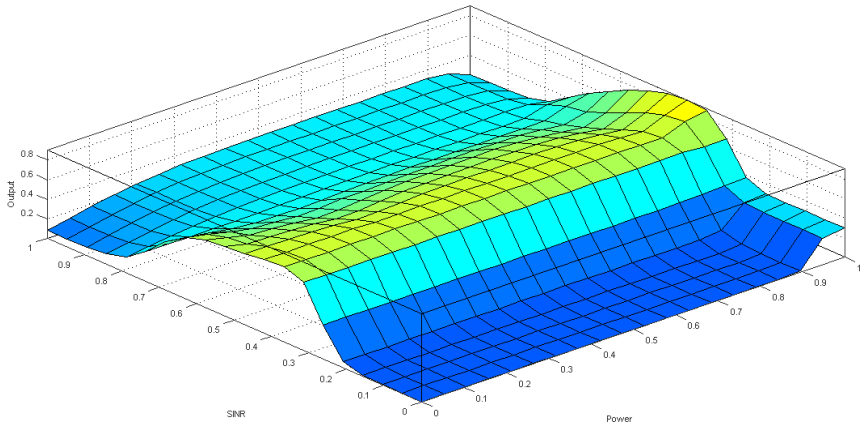

FIGURE 3.7: Fuzzy decision surface for power and the SINR after genetic algorithm optimization

walk model) and the best individuals are copied directly to the next newly initialized population. The number of best individuals copied during the over-step elitism is, however, not fixed but increases linearly with the number of random walk steps have been taken so far. Moreover, in order to get the controller adapted more, the number of WBANs in the system also changes after a certain number of steps, while the best individuals are again selected by the over-step elitism.

Figures 3.7, 3.8 and 3.9 demonstrate the fuzzy decision surfaces after performing the genetic algorithm optimization for each pair of inputs. Figure 3.7 shows how the fuzzy control output, i.e. the power level in the next tome slot changes with respect to the SINR and the current power level. The graph illustrates that the relation between the SINR and power is quite non-linear and complex. For example, for a certain level of the current power, the next power level will be low for both low and high SINR values, while it increases for mid values of SINR. On the other hand, for a certain SINR, the next power level will be almost independent of the current power level except for the cases where the current power is very high.

Figure 3.8 demonstrates how the next power level depends on interference power and the current power level. Again, a non-linear and complex behaviour is perceived. When the current power level is not very high, the next power level plateaus for a certain value of interference power. This behaviour, however, does not hold when interference is very high.

Figure 3.9 shows how the next power level varies with respect to interference and the SINR. As it can be seen, in very low SINR conditions, the next power level will be quite high in both low and high interference conditions, and it is very low for mid values of interference, while this behaviour is exactly the reverse for mid values of SINR. Besides, the controller decreases the next power level in very high SINR regime

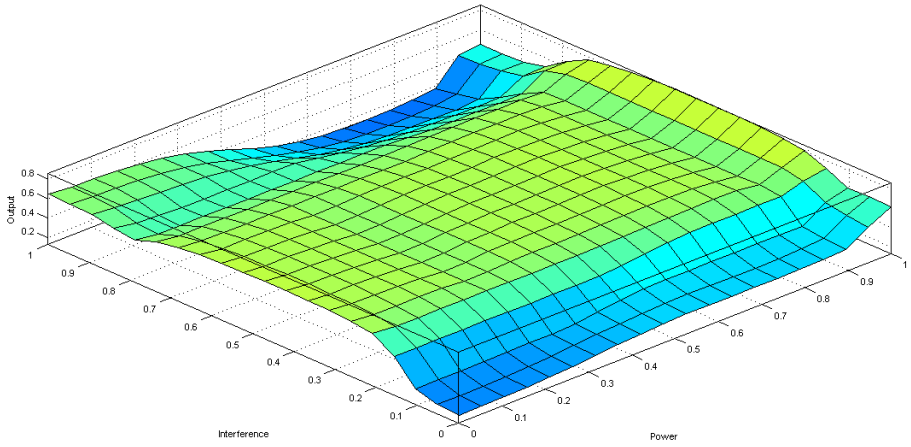

Figure 3.8: Fuzzy decision surface for power and the interference after genetic algorithm optimization

especially when interference is also very high.

3.6 Performance Evaluation

3.6.1 Simulation Framework

We use simulations to evaluate the performance of the proposed power control schemes in this research. In this section we elaborate the unified simulation framework used for the rest of the thesis. For a fair comparison, all the proposed approaches are evaluated using the same simulation framework and values for common parameters.

WBANs are confined inside a 10 m by 10 m room and walk around the room according to a random walk model as seen in Figure 3.10. For each WBAN, we consider one implanted sensor node inside the body at a depth of 50 mm from the body surface that communicates with its coordinator node that is located 1.1 m away from it on the body surface.

Channel gains are calculated based on the channel models (CM) as defined in [2]. We have different path types in our model as seen in Figure 3.11 which are as follows: an in-body path from a BN to the body surface, known as CM2; an on-body path from the body surface to the BNC in the same WBAN, referred to as CM3; and an off-body path from the body surface to a BNC in another WBAN, namely CM4, which can be with line of sight (LOS) or non-LOS. The transmission takes place in the MICS frequency band and is considered to be only from sensor nodes to BNC nodes.

We change the density of WBANs in the system and assess the performance of

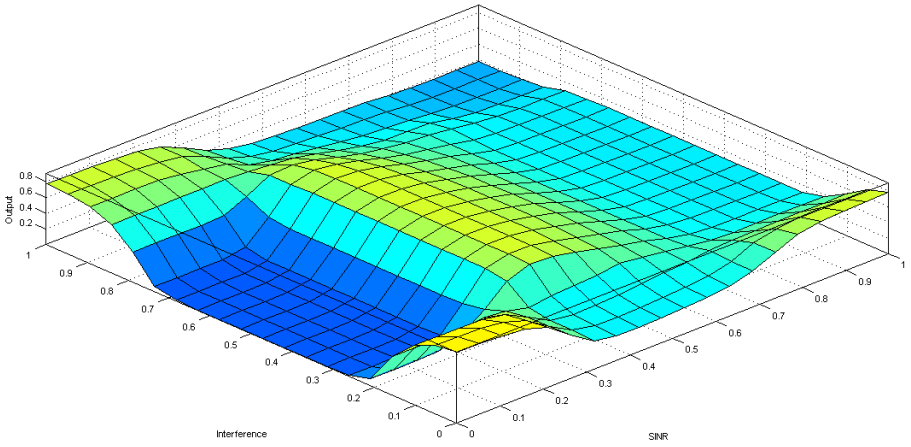

FIGURE 3.9: Fuzzy decision surface for interference and the SINR after genetic algorithm optimization

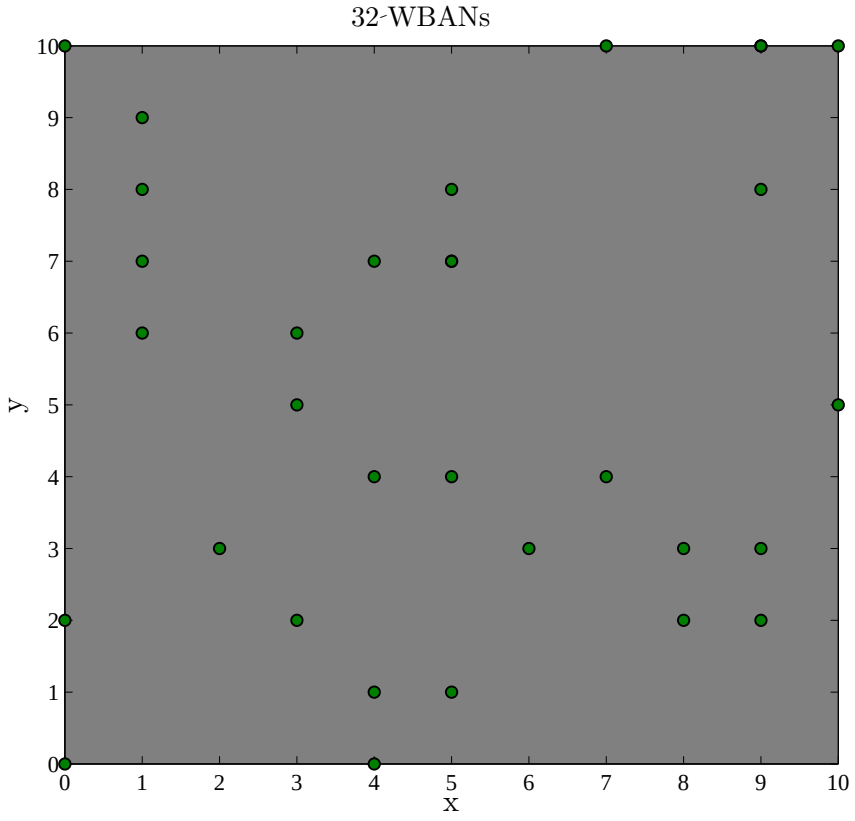

FIGURE 3.10: WBANs move around the simulation room according to the random walk model while transmitting and their signals interfering.

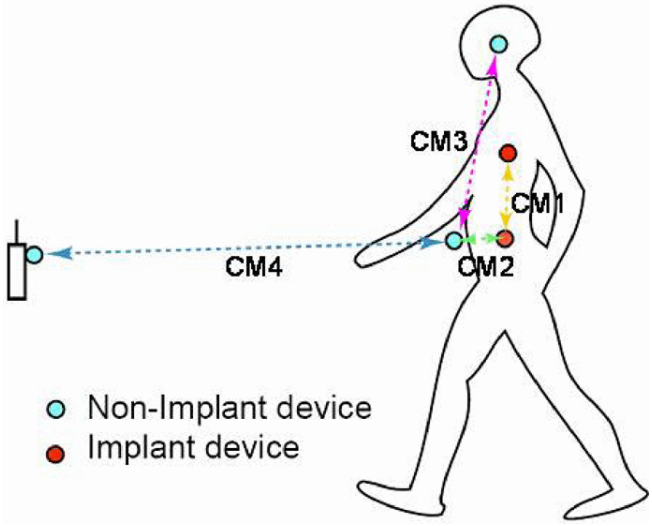

Figure 3.11: Different types of channel in WBANs [2]

algorithms in terms of power consumption, throughput, energy consumption per bit and convergence. We refer to a scenario as a setup of a certain number of WBANs in the system which move around the room and transmit simultaneously in the same frequency channel. Each scenario is run 1000 times and the confidence intervals of 95% will be provided.

3.6.2 Simulation Results

In this section, we evaluate the performance of WFPC by simulations and compare it to ADP. The simulation environment is as previously elaborated in the Simulation Framework section in chapter 3. Table 3.1 summarizes the parameters and their values used in this simulations. Each plot is the average of 1000 runs of the simulation.

Figure 3.12 represents the average transmission power level as a function of the number of WBANs in the system. As it can be clearly seen, WFPC strongly outperforms ADP and transmits at almost 9 μW less power than ADP for any number of WBANs in the system. This is equivalent to almost 40% and 50% improvement under sparse (4 WBANs) and dense conditions (32 WBANs) respectively.

Figure 3.13 shows the average throughput versus the number of WBANs in the system. The graph reveals that ADP slightly outperforms WFPC and delivers almost 20 kbps more throughput for any number of WBANs in the system. We can conclude that WFPC sacrifices a small portion of throughput, which is around 4%, for a great improvement in power being 9 μW (40%-50%). In order to find out how well this tradeoff works, we need to look at Figure 3.14 which shows the average energy consumption per bit versus the number of WBANs in the system for both WFPC and ADP. As it can be clearly seen, WFPC strongly outperforms ADP and consumes less

TABLE 3.1: Simulation Parameters and Values

Parameter Name	Symbol	Parameter Value
Bandwidth	B	300 kHz
Noise	n_i	-174 dBm/Hz
Minimum Transmission Power	P_i^{min}	0
Maximum Transmission Power	$P_{\max_i}$	25 μW ($\approx$ -16 dBm)
Minimum SINR	$SINR_{\min}$	-100 dB
Maximum SINR	$SINR_{\max}$	100 dB
Minimum Interference Power	$P_{I_{\min}}$	-100 dBm
Maximum Interference Power	$P_{I_{\max}}$	100 dBm
Price Factor	w_{p_i}	0.02
Number of Input Fuzzy Sets	K	3
Number of Output Fuzzy Sets	K_{out}	5
Genome Size		45
Number of Population		450
Mutation Probability		0.05
Crossover Probability		0.8

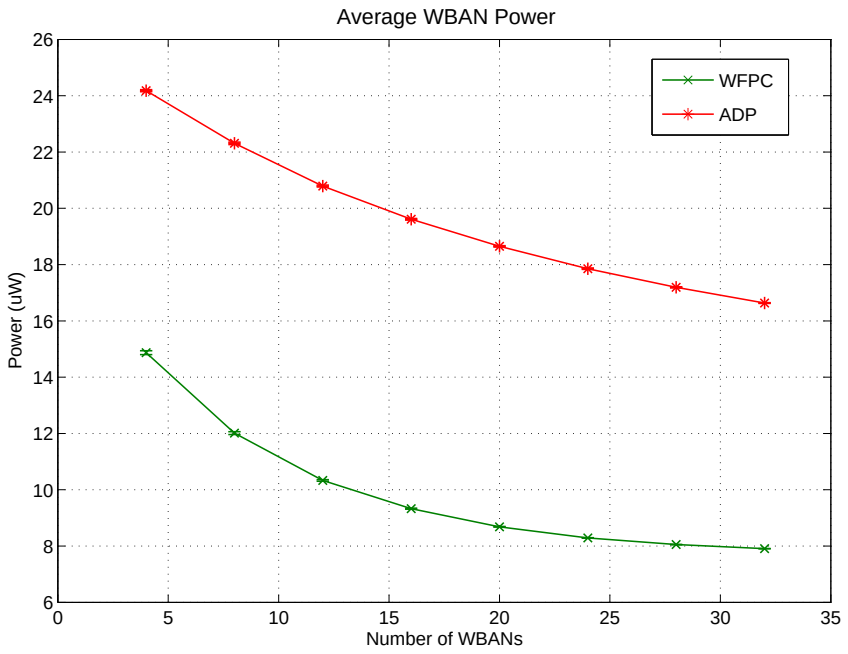

FIGURE 3.12: Average transmission power versus the number of WBANs

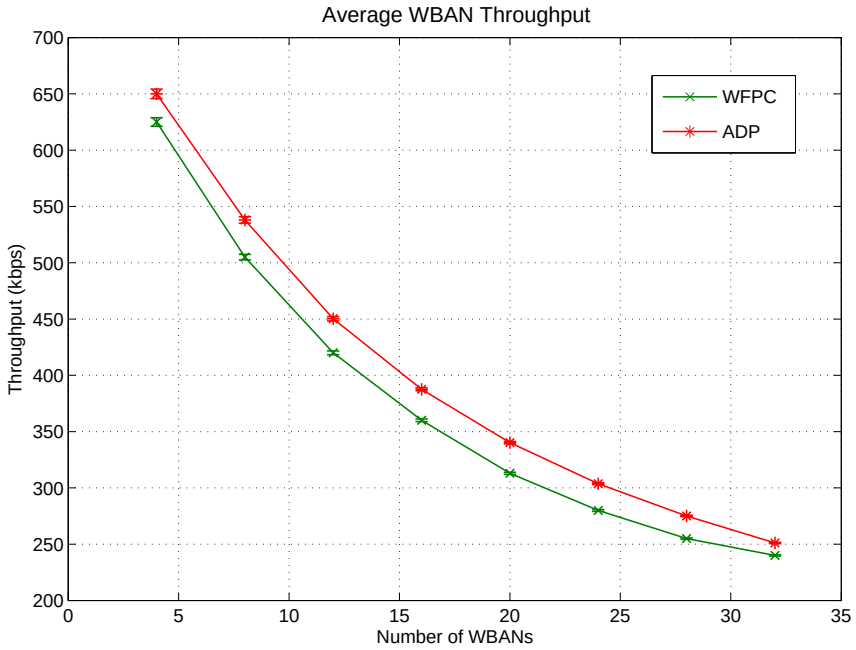

Figure 3.13: Average throughput versus the number of WBANs

energy than ADP for transmitting one bit particularly with large number of WBANs in the system where it shows over 45% improvement in energy conservation, while this figure amounts to around 27% under sparse condition.

Finally, Figure 3.15 shows the average number of iterations needed by each approach for converging to a stable solution. It is readily observed that WFPC strongly outperforms ADP in terms convergence by almost 70% under sparse condition and 60% under dense condition.

3.7 Conclusions

We proposed a power controller based on fuzzy logic, namely WFPC, to manage internetwork interference in WBANs. A genetic algorithm and a learning mechanism was developed to design and optimize WFPC. We evaluated and compared the performance of the proposed approach to the ADP algorithm, a well-cited power controller in the literature. Simulation results show that WFPC strongly outperforms ADP and improves both power consumption by 40%-50% and convergence by 60%-70% for different number of WBANs in the system, while sacrificing only 4% of throughput. Also, the average energy consumption per bit improves by 27%-45% for different number of WBANs in the system. This superiority basically originates from two factors which are the ability of genetic algorithms to find the best solution and the ability of fuzzy

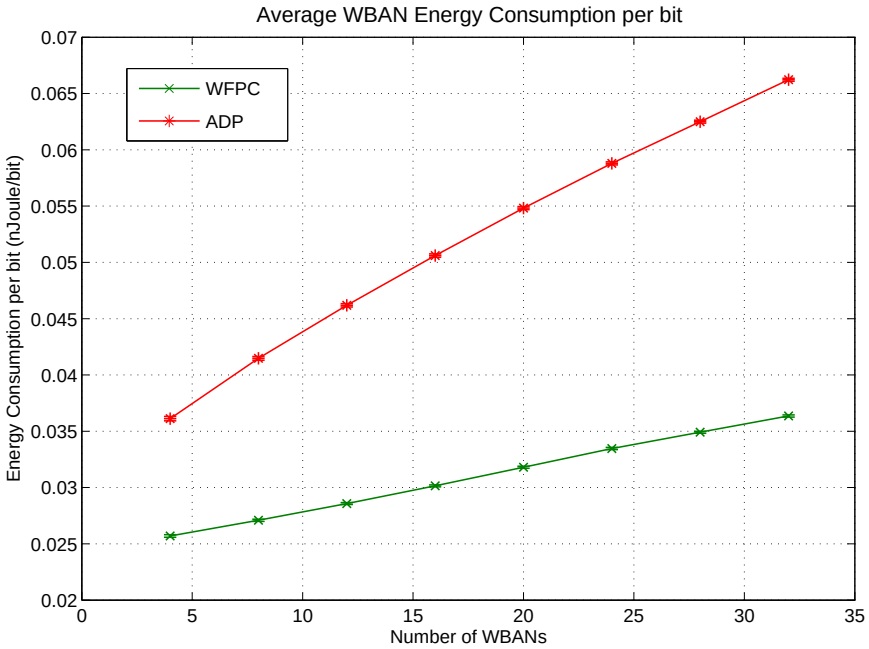

FIGURE 3.14: Average energy consumption per bit versus the number of WBANs

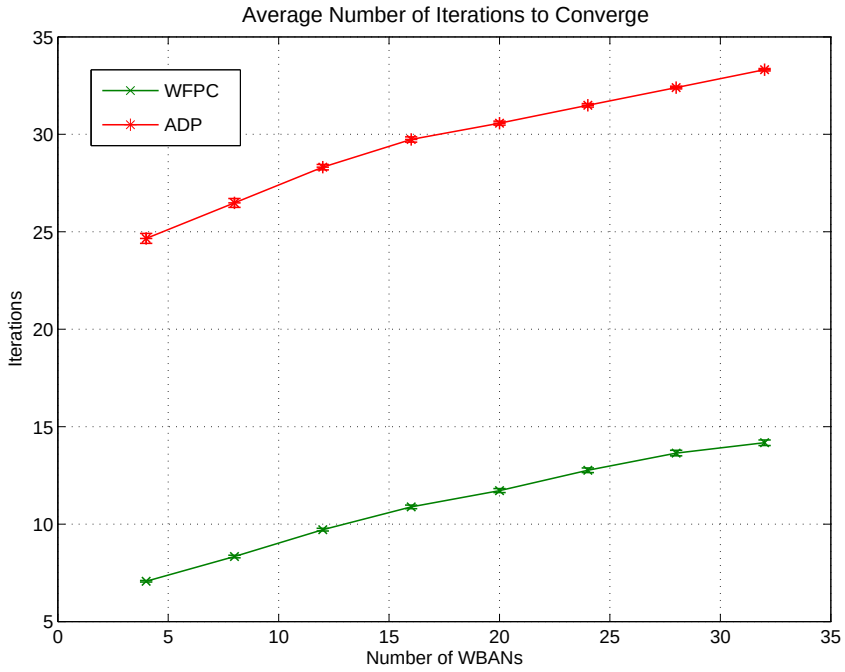

FIGURE 3.15: Average number of iterations versus the number of WBANs

controllers to cope with complicated non-linear systems. However, the genetic algorithm optimization required by WFPC decreases its flexibility to adjust the tradeoff between throughput and power adaptively and accommodate dynamic changes of the surrounding environment because it is performed offline at design stage. In the following chapters, we propose more adaptive approaches for power control in WBANs based on game-theory (Chapter 5) and based on reinforcement learning (Chapter 6). We also further evaluate and compare the performance of WFPC to the new power controllers which will be proposed.

"Gravitation cannot be held responsible for people falling in love."
Albert Einstein

4

Rate-Power Tradeoff - Game Theory Approach

Although the proposed genetic-fuzzy approach performs well, it requires off-line optimizations which means that the fuzzy controller needs to be optimized at the design stage using the time-consuming genetic algorithms. This is the main drawback of this approach because once the optimum design is achieved for a specific tradeoff between rate and power, it will not change anymore and the controller can not adapt to the dynamic changes of the surrounding environment. This makes the controller very sensitive to design parameters which is not favorable in practice. In order to remove the off-line optimization and design a more adaptive power controller, we employ the game theory. In this chapter, we put forward a non-cooperative game for power control, called WPCG[1], to mitigate inter-network interference in WBANs. The proposed power control game enables WBANs to coordinate transmission power levels so as to increase the system total throughput in the presence of interference from nearby WBANs using as little power for transmission as possible. We utilize both a first-order and a second-order pricing mechanism to penalize high power users and also introduce an adaptive pricing scheme to increase throughput in good channel conditions and high energy budgets. We investigate the Nash Equilibrium (NE) existence in the game and propose the best response strategy for players in the game to reach the NE. Finally we assess the performance of WPCG and compare it to the previously proposed fuzzy power controller, namely WFPC.

[1]WBAN Power Control Game

4.1 Game Theory

Game theory is a discipline aimed at modeling the interaction between decision makers with conflicting interests. Since 1944 when the first paper on game theory [102] was written by Von Neumann and followed by his textbook [103], many researchers have contributed to the field and just after a few years, game theory turned into a very active field of study. Although it was primarily used in economics to model competition between companies, game theory soon found its way to engineering fields to assist with solving optimization problems. Some references to game theory include [104], [105], [106] and [107].

Games are broadly categorized as *cooperative* and *non-cooperative*. A cooperative game deals with specifying what payoffs each potential group, or coalition, can obtain by the cooperation of its members. It however does not delineate the process by which the coalition forms. Cooperative game theory is mostly applied to situations emerging in political or economic relations, where concepts like power are the most important. For example, a cooperative game can describe which coalitions of parties can form a majority in a parliament, based upon the number of seats occupied by party members. It focuses solely on the outcome of such coalition formation, rather than specifying how this should be carried out. In contrast, non-cooperative game theory is concerned with the analysis of self-interested decision makers who make strategic choices based on their own interest. Unlike that of the cooperative games, the details of the ordering and timing of players choices are often crucial in determining the outcome of a non-cooperative game. In this thesis, we utilize non-cooperative games as they better suit to model WBANs with medical applications where cooperation between WBANs is not tolerable.

4.1.1 Non-Cooperative Games

Definition 6.1: A non-cooperative game (also known as a strategic game) is denoted by a tupple $\langle \{A_i\}, \{\pi_i\} \rangle_{\forall i \in M}$, where

- $M = \{1, 2, 3, \cdots, m\}$ is the set of a finite number, m, of players in the game.
- A_i is the set of pure strategies (or actions) available to player i. The Cartesian product of all players' individual strategy sets is known as the strategy profile set and is denoted by $A = \{\times A_i\}_{\forall i \in M}$. Each member of A is a pure strategy profile $a = (a_1, \cdots, a_m)$, where $a_i \in A_i$. The notation $a_{-i} = (a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_m)$ is used to delineate the pure strategies selected by all players except player i.
- $\pi_i : A \rightarrow \mathbb{R}$ is the payoff function of player i which is a real-valued function defined from the pure strategy profile set A to $\mathbb{R}$. Players are assumed to be *rational* which means that each player make choices to maximize his own payoff function.

In the following, we state some definitions and theorems which will be used in the rest of this chapter.

The Best Response

Definition 6.2: *A pure strategy $a_i = \mathrm{BR}_i(a_{-i}) \in A_i$ is called the best response of player i to the strategies of all other players a_{-i} if*

$$\pi_i(a_i, a_{-i}) \geq \pi_i(a_i^{'}, a_{-i}) \; \forall a_i^{'} \in A_i \tag{4.1}$$

This means that there is no alternative strategy to select for player i that gives him better profit, given all other players' strategies.

Pure Nash Equilibrium

Definition 6.3: *A pure strategy profile $a^* \in A$ is a pure Nash equilibrium if, for all players i*

$$\pi_i(a_i^*, a_{-i}^*) \geq \pi_i(a_i, a_{-i}^*), \forall a_i \in A_i \tag{4.2}$$

This means that no player can improve his payoff by unilaterally deviating from his strategy given that the other players stick to their selected strategies. In other words, the strategy profile a^* is the best response strategy profile where for each player i, a_i^* is the best response for a_{-i}^*.

Pure Nash Equilibrium Existence

Theorem 6.1: *In a strategic game, if each player's pure strategy set is a non-empty, compact and convex subset of an Euclidean space, and each player's payoff function is continuous and quasi-concave*[1] *over the player's strategies, then the game admits at least one pure Nash equilibrium.*

Proof: The proof can be found in Theorem 1.2 in [104].

Mixed Strategy

Definition 6.4: *A mixed strategy for player i denoted by $\Delta_i : A_i \to \mathbb{R}$ is a probability distribution over his pure strategy set, A_i. When A_i is a finite set, a mixed strategy*

[1] A function f defined on a convex set S is quasi-concave if for every value of a, the set $P_a = \{x \in S : f(x) \geq a\}$ is convex. Quasi-concavity generalizes the concavity, i.e. all concave functions are also quasi-concave, but not all quasi-concave functions are concave.

simply assigns a value to each pure strategy $a_i \in A_i$, which specifies the probability of choosing that pure strategy by player i, i.e. $\Delta_i(A_i) = \{\sigma_i(a_i) \mid \sum \sigma_i(a_i) = 1\}_{\forall a_i \in A_i}$, where $\sigma_i(a_i)$ is the probability of choosing the pure strategy i.

A mixed strategy profile denoted by $\sigma = (\sigma_1, \cdots, \sigma_m)$ comprises the mixed strategies chosen by all players. The expected payoff of each player i as a result of choosing the mixed strategy profile σ is as follows

$$u_i(\sigma) = \sum_{\forall a \in A} \prod_{j=1}^{m} \sigma_j(a_j) \cdot \pi_i(a_i, a_{-i}) \tag{4.3}$$

The notation $\sigma_{-i} = (\sigma_1, \cdots, \sigma_{i-1}, \sigma_{i+1}, \cdots, \sigma_m)$ is used to denote the mixed strategy profile chosen by all players except player i.

Mixed Nash Equilibrium

Definition 6.5: *A mixed strategy profile σ^* is a mixed Nash equilibrium if, for all players i*

$$u_i(\sigma_i^*, \sigma_{-i}^*) \geq u_i(\sigma_i^{'}, \sigma_{-i}^*), \forall \sigma_i^{'} \in \Delta_i \tag{4.4}$$

Strict Nash Equilibrium

Definition 6.6: *A strategy profile $a^* \in A$ is a strict Nash equilibrium if, for all players i*

$$\pi_i(a_i^*, a_{-i}^*) > \pi_i(a_i, a_{-i}^*), \forall a_i \in A_i \tag{4.5}$$

In other words, a Nash equilibrium is strict if each player has a unique best response to other players' strategies. A strict Nash equilibrium can not be mixed and is necessarily pure in strategies. There can be only one strict Nash equilibrium in a game.

Mixed Nash Equilibrium Existence

Theorem 6.2: *In a strategic game, if each player's pure strategy set is finite, the game admits at least one mixed Nash equilibrium.*

Proof: The proof can be found in Theorem 1.1 in [104].

4.2 Proposed Approaches

Because of the nature of their applications particularly in medicare and healthcare, WBANs do not cooperate in their power decision making implying that each WBAN should choose its transmission power independently based on its belief of other WBANs' choices. This suggests modeling WBANs as the rational players in a non-cooperative game where each player's goal is to maximize its own payoff. The payoff function can be intuitively inspired by decomposing the following non-convex optimization problem, which maximizes the total throughput, into distributed sub-problems and applying a pricing mechanism on the power levels so as to prevent WBANs from lavishly increasing their power levels.

$$\max \sum_{i=1}^{m} c_i \tag{4.6}$$

$$\begin{aligned} &\text{subject to } \ 0 \leq p_i \leq P_{\max_i} \ , \forall i \\ &\text{variables } \ \mathbf{p} = (p_i)_{i=1}^{m} \end{aligned} \tag{4.7}$$

where c_i is the throughput of WBAN i given by

$$c_i = B \log_2(1 + \frac{h_{ii} p_i}{\sum_{j \neq i}^{m} h_{ji} p_j + n_i}) \tag{4.8}$$

We call the resulted game, WPCG[1], and define it as follows.

Definition 6.6: The WBAN Power Control Game is a non-cooperative game denoted by a tupple $\langle \{A_i\}, \{\pi_i\} \rangle_{\forall i \in M}$, where

- Each player i models the link between the BN and BNC nodes within WBAN i at a certain time slot.
- The strategy set of player i is comprised of the transmission power levels which transmitter i can choose, i.e. $A_i = \{p_i \mid 0 \leq p_i \leq P_{\max_i}\}$.
- The payoff functions of each player i is defined as the difference between its normalized individual throughput and a pricing term on the normalized individual power as follows

$$\pi_i(p_i, p_{-i}) = \frac{c_i}{C_{\max_i}} - w_{p_i} (\frac{p_i}{P_{\max_i}})^{\alpha_i} \tag{4.9}$$

where w_{p_i} is the price factor; α is the price exponent which as we will see later, imposes sufficient conditions for existence of Nash equilibrium in the game as

[1] WBAN Power Control Game

well as affecting the behavior of the game; c_i is the throughput of WBAN i; $C_{\max_i}$ is the maximum channel capacity achievable at zero interference used for normalization; p_i is the transmission power level of WBAN i and $P_{\max_i}$ is the maximum allowable transmission power of WBAN i.

4.2.1 Nash Equilibrium

In this section we prove the existence of the Nash equilibrium in WPCG.

Proposition 6.1: *There exists a pure Nash equilibrium in WPCG if* $\alpha_i \geq 1$ *and* $w_{p_i} > 0$.

Proof: The payoff function $\pi_i(p_i, p_{-i})$ is continuous and twice differentiable with respect to p_i. Taking the first derivative, we get:

$$\frac{\partial \pi_i(p_i, p_{-i})}{\partial p_i} = \frac{B}{\log(2) C_{\max_i}} \cdot \frac{h_{ii}}{h_{ii} p_i + \sum_{j \neq i}^{m} h_{ji} p_j + n_i} - \alpha_i \frac{w_{p_i}}{P_{\max_i}^{\alpha_i}} p_i^{\alpha_i - 1} \tag{4.10}$$

Taking the second derivative gives:

$$\frac{\partial^2 \pi_i(p_i, p_{-i})}{\partial p_i^2} = -\frac{B}{\log(2) C_{\max_i}} \cdot \frac{h_{ii}^2}{\left(h_{ii} p_i + \sum_{j \neq i}^{m} h_{ji} p_j + n_i \right)^2} - \alpha_i (\alpha_i - 1) \frac{w_{p_i}}{P_{\max_i}^{\alpha_i}} p_i^{\alpha_i - 2} \tag{4.11}$$

It is straightforward to see that for $\alpha_i \geq 1$ and $w_{p_i} > 0$, the second derivative is less than zero for all p_i and a given p_{-i}, which means that $\pi_i(p_i, p_{-i})$ is strictly concave in p_i, for a given p_{-i} . On the other hand, the strategy set of player i, $A_i = [0, P_{\max_i}]$, is non-empty, convex and a compact subset of the Euclidean space $\mathbb{R}^n$. As a result according to Theorem 6.1, we can conclude that there exists a pure Nash equilibrium in the game.

Proposition 5.2: *The pure Nash equilibrium in WPCG when* $\alpha_i \geq 1$ *and* $w_{p_i} > 0$ *is unique.*

Proof: Let p_i^* denote the root of Eq. (4.10) which maximizes the payoff function of each player i, i.e.

$$\left. \frac{\partial \pi_i(p_i, p_{-i})}{\partial p_i} \right|_{p_i = p_i^*} = \frac{B}{\log(2) C_{\max_i}} \cdot \frac{h_{ii}}{h_{ii} p_i^* + \sum_{j \neq i}^{m} h_{ji} p_j + n_i} - \alpha_i \frac{w_{p_i}}{P_{\max_i}^{\alpha_i}} p_i^{*\alpha_i - 1} = 0 \tag{4.12}$$

Three cases may happen. Firstly, if p^* satisfies $p^* \in [0, P_{\max_i}]$, it can be easily shown that $\partial \pi_i(p_i, p_{-i}) / \partial p_i > 0$ for $p_i \in [0, p_i^*]$, meaning that $\pi_i(p_i, p_{-i})$ is strictly increasing,

and $\partial\pi_i(p_i, p_{-i})/\partial p_i < 0$ for $p_i \in [p_i^*, P_{\max_i}]$ when $\pi_i(p_i, p_{-i})$ is strictly decreasing. Secondly, if p^* hits the lower boundary of p_i, i.e. $p_i^* < 0$, then we still have $\pi_i(p_i, p_{-i})$ strictly decreasing in p_i. Thirdly, if p^* hits the upper boundary of p_i, i.e. $p_i^* > P_{\max_i}$, then $\pi_i(p_i, p_{-i})$ is strictly increasing in p_i. Hence, the payoff function of each player i is strictly concave in p_i implying that the best response of each player i to the strategies chosen by other players in unique. This concludes that the Nash equilibrium in the game is strict and thereby unique.

4.2.2 The Best Response

In this section, we provide a best response approach for WBANs to calculate their transmission power and reach the Nash equilibrium.

Finding the best response requires finding the solution of Eq. (4.12). The case of $\alpha_i = 2$ is easily solved and it is an example of a pricing function that penalizes high-power WBANs more severely than the linear cost case of $\alpha_i = 1$. The solution of Eq. (4.12) in the case of $\alpha_i = 2$ is given by the solution of the following quadratic equation

$$p_i^{*2} + \frac{p_i^*}{\eta_i} = \frac{2BP_{\max_i}^2}{\log(2)C_{\max_i}w_{p_i}} \tag{4.13}$$

where η_i is the sensitivity of SINR to power at WBAN i as defined in Definition 3.4.

Solving Eq. (4.13) gives

$$p_i^* = \frac{1}{2}\sqrt{\frac{1}{\eta_i^2} + \frac{2BP_{\max_i}^2}{\log(2)C_{\max_i}w_{p_i}}} - \frac{1}{2\eta_i} \tag{4.14}$$

We also need to take the power boundaries into account to obtain the best response. Since Eq. (4.14) is always greater than zero, we will have the following result.

Proposition 6.2: *The Nash equilibrium in WPCG with $\alpha_i = 2$ and $w_{p_i} > 0$ is the strategy profile $\{\mathrm{BR}_i(p_{-i})\}_{i\in M}$ where $\mathrm{BR}_i(p_{-i})$ is the best response of player i to the strategies of all other players in the game and is given by*

$$\mathrm{BR}_i(p_{-i}) = \min\left(P_{\max_i}, \frac{1}{2}\sqrt{\frac{1}{\eta_i^2} + \frac{2BP_{\max_i}^2}{\log(2)C_{\max_i}w_{p_i}}} - \frac{1}{2\eta_i}\right) \tag{4.15}$$

Using a similar approach, we can obtain the best response in the power control game for the linear cost case of $\alpha = 1$ as follows.

Proposition 5.3: *The Nash equilibrium in WPCG with $\alpha_i = 1$ and $w_{p_i} > 0$ is the strategy profile $\{\mathrm{BR}_i(p_{-i})\}_{i\in M}$ where $\mathrm{BR}_i(p_{-i})$ is the best response of player i to the*

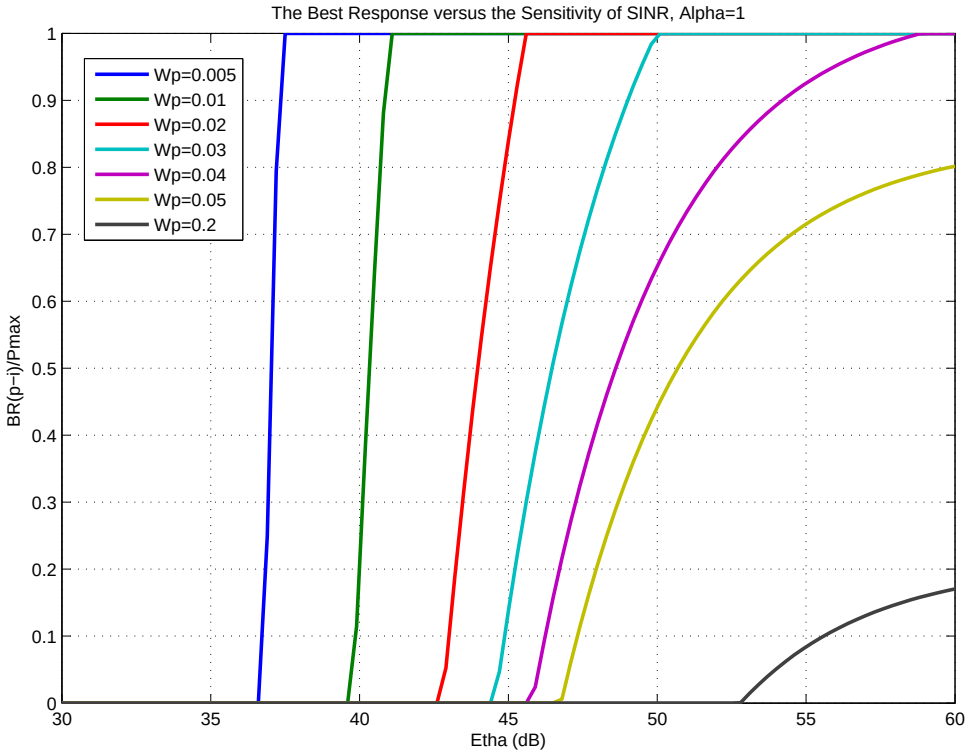

FIGURE 4.1: $\mathrm{BR}_i(p_{-i})/P_{\max_i}$ versus η_i for $\alpha_i = 1$

strategies of all other players in the game and is given by

$$\mathrm{BR}_i(p_{-i}) = \min\left(P_{\max_i}, \max\left(0, \frac{BP_{\max_i}}{\log(2)C_{\max_i}w_{p_i}} - \frac{1}{\eta_i}\right)\right) \tag{4.16}$$

The best response of each player gives him the highest payoff in response to the strategies chosen by all other players. If each player plays his best response, the game will finally settle at a Nash equilibrium. WBANs in our system independently and asynchronously update their power levels in an iterative manner using Eq. (4.15) or (4.16) until the game converges to the Nash equilibrium.

Figures 4.1 and 4.2 demonstrate the best response $\mathrm{BR}_i(p_{-i})$ versus η_i for different values of w_{p_i} with $\alpha_i = 1$ and $\alpha_i = 2$ respectively. As it can be clearly seen, as interference increases, which corresponds to smaller values of η_i, lower-clipping happens to the best response for $\alpha_i = 1$, which makes it more rigid in power decision making than $\alpha_i = 2$. For example, for $\alpha = 1$ with $w_{p_i} = 0.005$, if η_i is less than 37dBm, the best response will be 0 (i.e. fully shut down). If, however, η_i increases only 2dBm, the best response will be 1 (i.e. transmitting at full power). In other words, $\alpha = 1$ may lead to a binary power allocation, where the nodes are either transmitting at full power

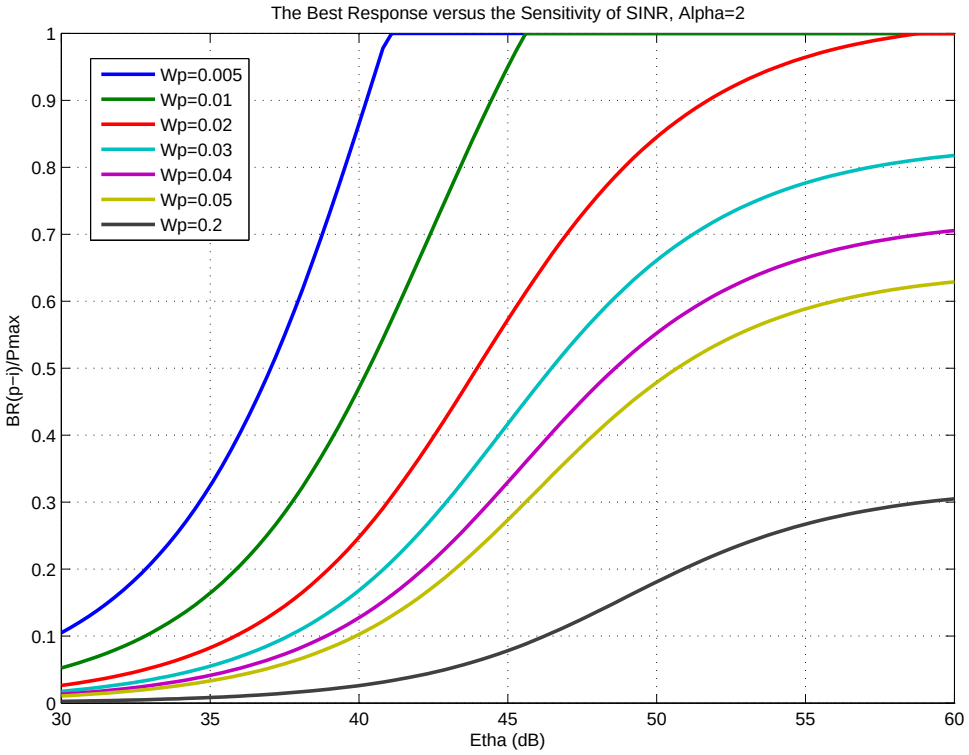

Figure 4.2: $\mathrm{BR}_i(p_{-i})/P_{\max_i}$ versus η_i for $\alpha_i = 2$

or switched off. In contrast, this behavior is much smoother with $\alpha = 2$.

4.2.3 Adapting to Dynamic Changes

The price parameters α_i and w_{p_i} can be utilized to adjust the tradeoff between throughput and power efficiency. This is of importance for WBANs where different WBANs have different throughput and power efficiency requirements because of the dynamics of the wireless channel states, different constraints on power and QoS requirements. Moreover, the price parameters can also be set by the WBAN's BNC according to some dynamic parameters such as the power budget, QoS metrics, and interference to and from nearby WBANs. We will employ w_{p_i} to dynamically adapt to channel gains and power budget so as to penalize WBANs with bad channel conditions or low power budgets more and to allow WBAN with good channels or high power budgets to take advantages of their good conditions. In order to employ w_{p_i} as an adaptive price factor, we should clarify its effect on the solution. Eq. (4.17) and (4.18) show the range of which leads to a best response within $[0, P_{\max_i}]$ without hitting the boundaries

for $\alpha_i = 1$ and $\alpha_i = 2$ respectively.

$$\frac{BP_{\max_i}\eta_i}{\log(2)C_{\max_i}(1+\eta_i P_{\max_i})} \leqslant w_{p_i} \leqslant \frac{BP_{\max_i}\eta_i}{\log(2)C_{\max_i}} \tag{4.17}$$

$$w_{p_i} \geqslant \frac{B\eta_i}{\log(2)C_{\max_i}\left(2P_{\max_i}^2\eta_i+1\right)} \tag{4.18}$$

Looking at Figures 4.1 and 4.2, we realize that for w_{p_i} smaller than the lower bounds in Eq. (4.17) and (4.18), also upper-clipping occurs on the best response due to the limitation on the maximum power for both $\alpha_i = 1$ and $\alpha_i = 2$. When interference decreases, which corresponds to greater values of η_i, the best response either is clipped by the maximum power (when w_{p_i} is quite small) or it is restricted by an upper bound which is smaller than the maximum power, even when the SINR is good. In other words, w_{p_i} plays the main role in limiting the power levels in high SINR conditions and can simply prevent WBANs to benefit form their good channel and high SINR conditions by limiting their power levels.

In order to dynamically adapt to channel changes and power budgets, the parameter w_{p_i} is no longer considered fixed but is a function of channel gains and power budget of WBANs as follows:

$$w_{p_i} = \frac{K_i}{h_{ii}}, \forall i \in M \tag{4.19}$$

where $K_i \geqslant K_i^{\min} > 0$ is a coefficient which reflects WBAN i's battery level. For full battery charge, it is set to $K_i^{\min}$ and as the WBAN's battery energy is consumed, increases. Different values of $K_i^{\min}$ distinguish different battery types and the importance given to battery capacity for different WBANs. Using Eq. (4.19), it will be possible to allow each user having a good channel state and power budget to take advantages of his good conditions by reducing his penalty for further increasing his power level, which will lead to increased system capacity. On the other hand, it will prevent the user from increasing his transmission power level uselessly when his channel gain is low or when he is running out of his battery power by increasing his penalty, and thereby avoiding interference to other WBANs. As a result, power consumption will reduce and network lifetime will increase.

4.2.4 SINR-based Adaptive Price Factor

In the previous section, we proposed an adaptive price parameter which adjusts the tradeoff between system capacity and power consumption according to the dynamic changes in the system including channel gains. Although parameters such as SINR are very simple to be measured at digital receivers, the availability of channel gains at receivers is still a rather serious problem and incurs a heavy cost and overload on

Table 4.1: Simulation Parameters and Values

Parameter Name	Symbol	Parameter Value
Bandwidth	B	300 kHz
Noise	n_i	-174 dBm/Hz
Maximum Transmission Power	$P_{\max_i}$	25 μW ($\approx$ -16 dBm)
Price Factor	w_{p_i}	0.02
Price Exponent	α	1

the system to be measured somehow as well as to be exchanged throughout the whole system. Consequently this makes all algorithms relying on channel gains at receivers less suitable and more expensive in practice. As a result, to overcome this shortcoming we suggest another adaptive price parameter which is not dependent on channel gains but relies only on SINR as:

$$w_{p_i} = \frac{K_i}{\xi_i}, \forall i \in M \tag{4.20}$$

where ξ_i is the SINR at the BNC node in WBAN i.

We evaluate the performance of different pricing schemes including the proposed adaptive pricing schemes in the following section by simulations.

4.3 Performance Evaluation

Table 4.1 summarizes the parameters and their values used in the simulations. Each plot is the average of 1000 runs of the simulation.

4.3.1 Pricing Mechanisms

Figures 4.3 and 4.4 show the average power and average throughput respectively versus the price factor w_{p_i} for $\alpha = 1$ and $\alpha = 2$. As it can be seen, both power and throughput drop when w_{p_i} increases. Although the performance of the system with $\alpha = 1$ and $\alpha = 2$ is similar for lower w_{p_i}, it shows some difference when w_{p_i} increases such that we see 2 μW less power and 130 kbps drop in throughput for $\alpha = 1$.

Figures 4.5 and 4.6 show the average power and average throughput respectively as functions of the number of WBANs in the system for different values of w_{p_i} and also the two adaptive schemes (1) and (2) based on Eq. (4.19) and (4.20) respectively. As it can be seen, for a given number of WBANs, decreasing the price factor w_{p_i} leads to higher power levels and also higher throughput for both $\alpha = 1$ and $\alpha = 2$. The graphs also reveal that the adaptive schemes (1) and (2) provide a moderate tradeoff

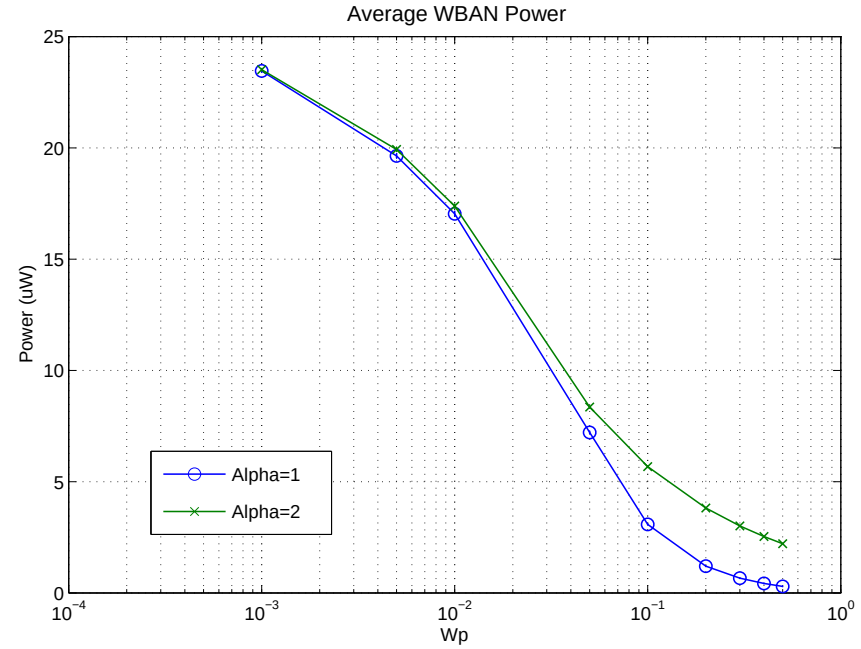

FIGURE 4.3: Average transmission power versus the price factor w_{p_i} with 16 WBANs

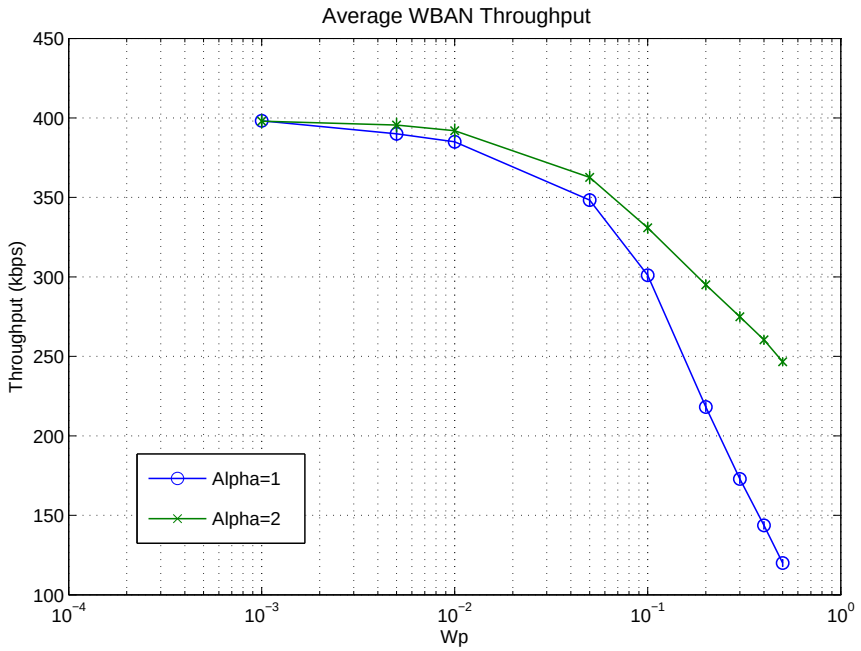

FIGURE 4.4: Average throughput versus the price factor w_{p_i} with 16 WBANs

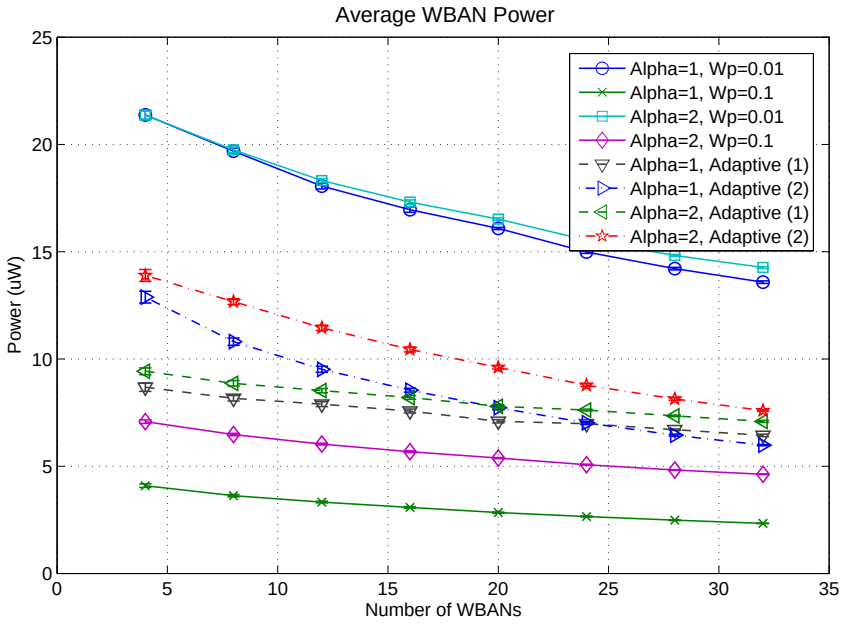

Figure 4.5: Average transmission power for different pricing schemes versus the number of WBANs

between throughput and power for both $\alpha = 1$ and $\alpha = 2$. Moreover, it is indicated that the adaptive scheme (2) which relies only on SINR performs almost similarly to the adaptive pricing scheme (1) which is based on channel gains.

4.3.2 WPCG versus WFPC and ADP

In the following, we compare the performance of WPCG to WFPC and ADP power controllers.

Figure 4.7 shows the average transmission power as a function of the number of WBANs in the system. As it can be clearly seen, all the approaches decrease the transmission power when number of WBANs in the system increases, leading to interference mitigation. WPCG surpasses ADP and uses almost 3 μW (i.e. about 12.5% under sparse and 17.5% under dense conditions) less power. However it is still outperformed by WFPC which transmits at almost 6 μW (i.e. about 27% under sparse and 40% under dense conditions) less power than WPCG.

Figure 4.8 represents the average throughput versus the number of WBANs in the system. The graphs illustrates that WPCG outperforms other approaches and provide almost 50 kbps (i.e. about 10% under sparse and 20% under dense conditions) more throughput than ADP and 100 kbps (i.e. about 20% under sparse and 40% under dense condition) more throughput than WFPC.

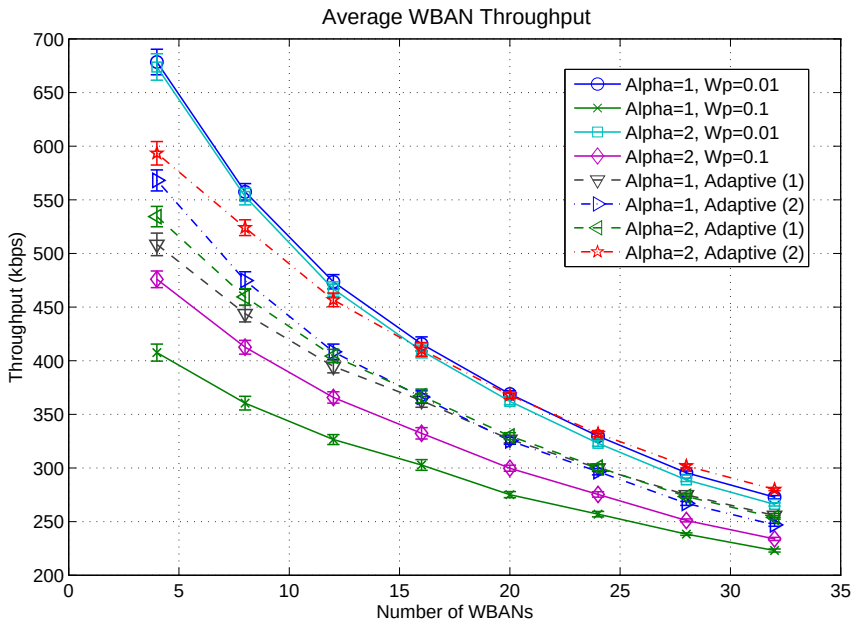

FIGURE 4.6: Average throughput for different pricing schemes versus the number of WBANs

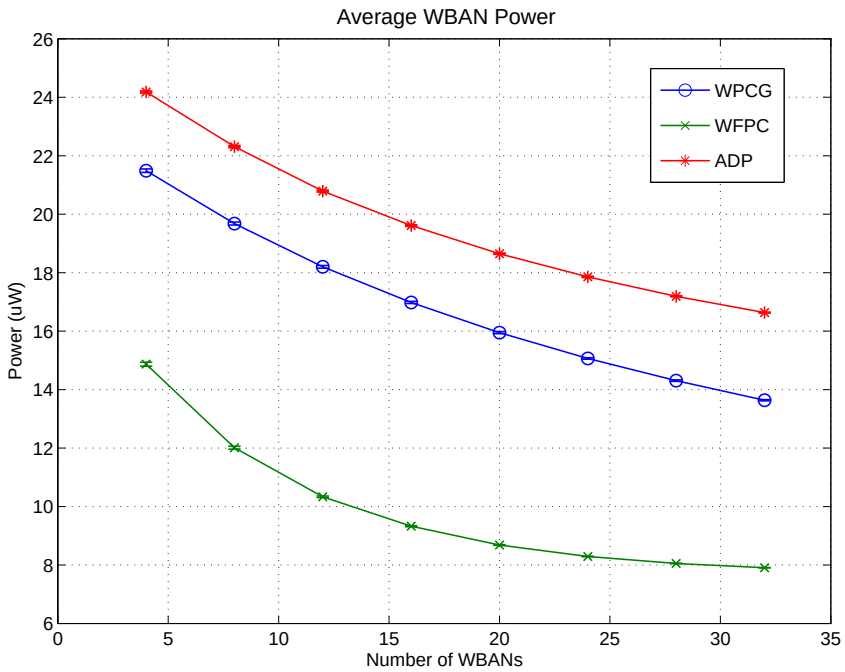

FIGURE 4.7: Average transmission power versus the number of WBANs

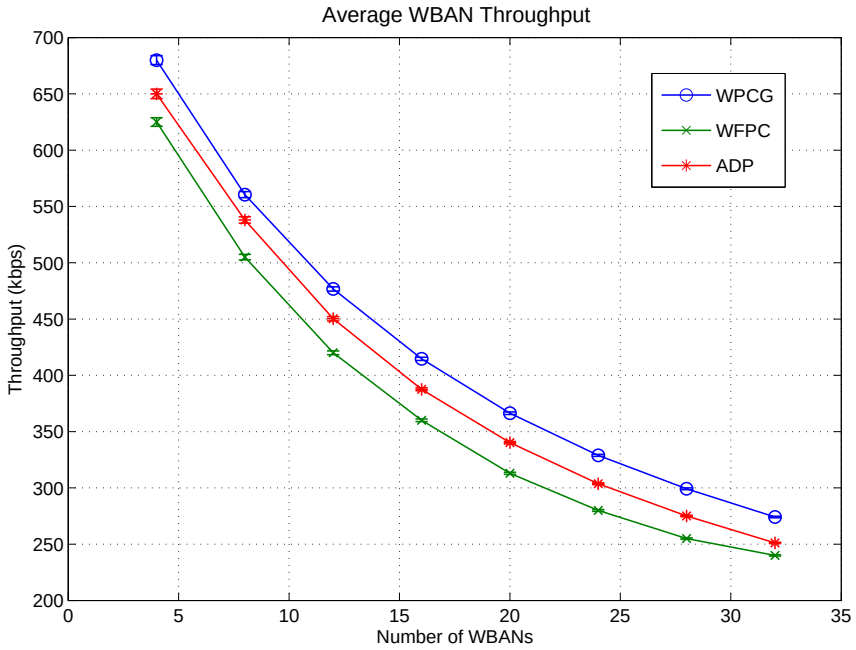

Figure 4.8: Average throughput versus the number of WBANs

Figure 4.9 demonstrates the average energy consumption per bit versus the number of WBANs in the system, and it reveals how the tradeoff between throughput and power is done. As it can be seen, WPCG ranks between WFPC and ADP. While it is outperformed by WFPC by almost 25%, it surpasses ADP by 25% under dense conditions

Finally, Figure 4.10 shows the number of iterations needed by each approach to reach the steady state versus the number of WBANs in the system. It is readily observed that WPCG strongly outperforms the other two approaches.

4.4 Conclusions

We proposed a non-cooperative power control game, namely WPCG, for inter-network interference mitigation in WBANs, and considered a broader family of pricing functions. We proved that there existed a Nash equilibrium in this game and proposed the dynamic rules of the game based on the best response. Moreover, we suggested an adaptive pricing mechanism to dynamically adjust the tradeoff between throughput and power based on the channel gains and WBANs' power budgets. This allows WBANs to benefit from good channel conditions and high energy budgets, leading to increased throughput, and on the other hand preventing them from increasing their transmission power levels at bad channel conditions or low power budgets, thereby

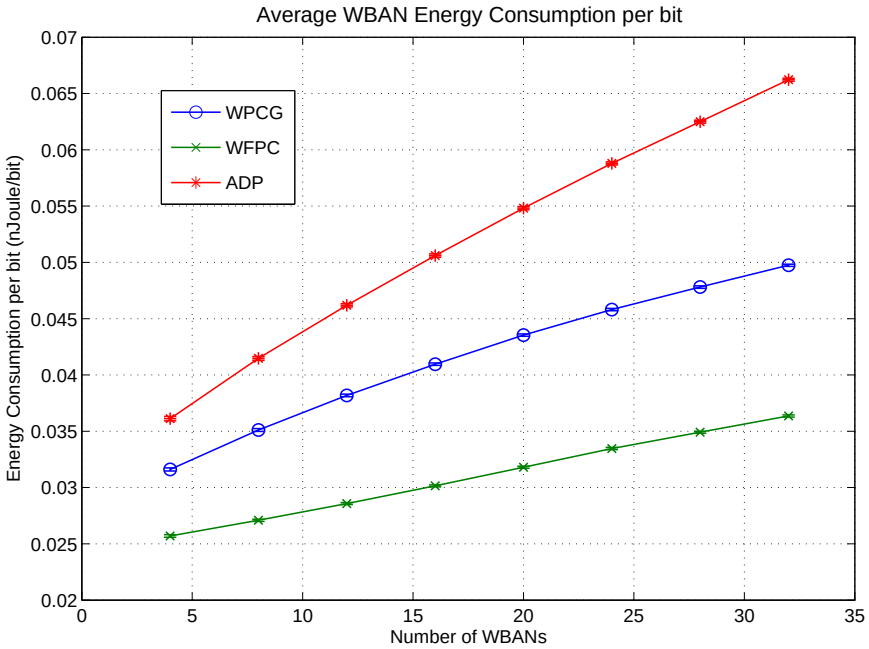

FIGURE 4.9: Average energy consumption per bit versus the number of WBANs

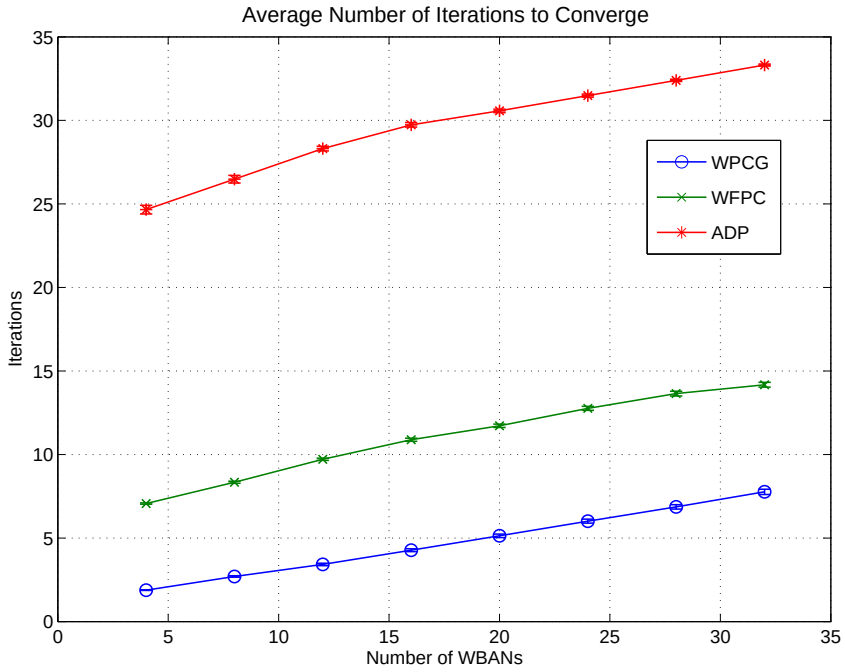

FIGURE 4.10: Average number of convergence iterations versus the number of WBANs

extending their battery lifetimes as well as decreasing interference to other WBANs. We also proposed another adaptive pricing method which did not rely on channel gains for calculations and required only SINR which was simply available at digital receivers at low cost, and it was showed by simulations to perform almost similarly to the first adaptive pricing scheme which was based on channel gains. We also compared the performance of WPCG to the previously proposed fuzzy power controller, namely WFPC, as well as to ADP. The simulation results indicate that WPCG strongly outperforms both WFPC and ADP in convergence while it is still overwhelmed by WFPC in terms of energy consumption per bit by 25%. WPCG consumes almost 6 μW more power than WFPC to produce 100 kbps more throughput.

"The less we deserve good fortune, the more we hope for it."
Moliere

5

Rate-Power Tradeoff - Reinforcement Learning Approach

The power control game proposed in Chapter 4 does not require any off-line optimization and using it, the tradeoff between rate and power can be adjusted adaptively. However, the solution of the game. i.e. the NE, needs to be devised in advance. In other words, the players of the game must implicitly know that they are playing the same game and are trying to reach the same point in the solution space, namely the NE. In an effort to make the power controller more adaptive and flexible, in this chapter, we propose a novel power controller called WRLPC[1], which employs Reinforcement Learning (RL) to learn from experience and improve its performance. We compare the performance of the proposed controller to the fuzzy and game-theoric power controllers proposed in the previous chapters. Simulation results indicate that WRLPC outperforms the fuzzy and game-theoric power controllers in terms of the solution optimality and provides a substantial saving in energy consumption per bit, while providing almost the same amount of throughput.

5.1 Reinforcement Learning

In this section, a review of reinforcement learning will be presented. For more detailed study, the reader is referred to the textbook [108] by Sutton and Barto.

Reinforcement learning is a branch of unsupervised learning algorithms which can

[1]WBAN RL-based Power Controller

solve control problems without using a model. Agents in RL learn to reach a goal by interacting with an environment in such a way that their long-term reward is maximized. For any state of environment, a given agent executes an action under a policy π, that, in general, is a probability distribution over all actions available to the agent at each state, and this action changes the current state of the environment to a new one. The environment responds to this change by giving an immediate reward to the agent. The learning agent tries to find an optimal policy π^*, which maps each state of the environment to the action(s) that the agent should take in that state so as to maximize its long-term rewards for any arbitrary starting state. The learned policy is deterministic or pure if the probability of choosing one action is 1 while the probability of choosing other actions is 0; otherwise it is called stochastic or mixed. The long-term reward which the agent tries to maximize is expressed by a discounted summation of immediate rewards r starting from time t over a time period T as:

$$R_t = \sum_{k=1}^{T} \gamma^{k-1} r_{t+k} \tag{5.1}$$

where γ is the discount factor and determines the level of far-sightedness of the agent. As γ approaches 1, the agent struggles more to achieve a possible high long-term reward at the expense of losing short-term rewards. For episodic tasks, i.e. the tasks having a terminal state, T is the time of reaching the terminal state and in this case R_t determines the episode reward. An episode is a sequence of $\langle$action-state-reward$\rangle$ triples the agent experiences starting from an arbitrary state and ending at the terminal state. However, for non-episodic tasks in which there is no terminal state, we have $T = \infty$ and γ must be less than one. In this study, we model our power control problem as an episodic task where each episode ends when the transmission power vector of WBANs converges to a stable solution, i.e. no WBAN changes its power anymore. In other words, the terminal state is the convergence point.

An RL problem can be regarded as determining which states are the most favorable according to the potential rewards of being in that state, obtained by evaluating the values of the states, and consequently choosing actions that are most likely to lead to the most valuable states. The value of each state is determined by the state value function $V^\pi(s)$ which is the expectation of the episode reward if the agent follows policy π, starting from state s:

$$V^\pi(s) = E\{R_t \mid s_t = s\} = \sum_a \pi(s,a) \sum_{s'} P(s,a,s') \left[R(s,a) + \gamma V^\pi(s')\right] \tag{5.2}$$

where $\pi(s,a)$ is the probability of choosing action a in state s under policy π; $P(s,a,s')$ is the transition probability for changing to state s' by choosing action a in state s and is given by:

$$P(s,a,s') : S \times A \times S \to \sigma = Pr\{s_{t+1} = s' \mid s_t = s, a_t = a\} \tag{5.3}$$

and $R(s,a)$ is the expectation of immediate reward received for choosing action a in state s:

$$R(s,a) : S \times A \to \mathbb{R} = E\left\{r_{t+1} \mid s_t = s, a_t = a\right\} \tag{5.4}$$

Alternatively to $V^\pi(s)$, the action value function $Q^\pi(s,a)$ explicitly denotes the value of taking action a in state s and following policy π afterward:

$$Q^\pi(s,a) = E\left\{R_t \mid s_t = s, a_t = a\right\} = \sum_{s'} P(s,a,s')\left[R(s,a) + \gamma V^\pi(s')\right] \tag{5.5}$$

The Q values for the optimal policy π^* are denoted by $Q^*(s,a)$ and according to the Bellman principle of optimality can be calculated by the following iteration:

$$Q^*_{t+1}(s_t,a_t) = R(s_t,a_t) + \gamma \sum_s P(s_t,a_t,s).\max_a Q^*_t(s,a)\ , t = 0,1,2,\cdots \tag{5.6}$$

where $Q^*_{t+1}(s_t,a_t)$ is the next update for the optimal Q value of the current state s_t and action a_t.

In this study, we will use deterministic rewards, i.e. $R(s,a) = r_{t+1}$, where the value of the next immediate rewards r_{t+1} is given by a reward function which will be defined later in the next section.

Once we have $Q^*(s,a)$, the greedy policy, which chooses the action with the greatest Q value amongst all actions available in state s, gives us the optimal policy:

$$\pi^*(s) = \operatorname*{argmax}_a Q^*(s,a) \tag{5.7}$$

Acting under the greedy policy is known as *exploitation* because the agent is exploiting its knowledge to take actions.

However the computation of $Q^*(s,a)$ from Eq. (5.6) requires the environment model parameters, i.e. $P(s,a,s')$ and $R(s,a)$, to be known. In reality, such perfect knowledge of the environment is rarely available. Reinforcement learning addresses this problem by estimating Q values without the need of the environment model. More precisely, a RL agent discovers the environment model using a technique called *exploration*, where the agent sometimes bypasses the exploitation and tries out a random action with the hope of achieving a higher long-term reward at the expense of possibly losing some short-term rewards. Interacting with the environment this way, the agent generates experience, i.e. a history of tupples $\langle a,s,r\rangle$, which will be used by RL to estimate Q values.

The policy used by an RL agent to generate experience is called the *behavior policy*, which normally makes a tradeoff between exploration and exploitation. Such tradeoff can be obtained by using an ϵ-greedy policy where with probability ϵ, the agent explores

and with probability $1-\epsilon$, it exploits, as seen in Eq. 5.8. The parameter ϵ is called the *exploration rate* and its value should be decreased gradually over time to make the ϵ-greedy policy hold the Greedy in the Limit with Infinite Exploration (GLIE) property. This allows agents to benefit more from their experience as they learn more and it is required for convergence to the greedy policy [109]. In our system, we decay the exploration rate from 1 by a factor of 0.9 after each episode.

$$\epsilon - greedy(Q,s) = \begin{cases} uniformly\ random\ action & \text{, with probability } \epsilon \\ \text{argmax}_a Q^*(s,a) & \text{, with probability } 1-\epsilon \end{cases} \tag{5.8}$$

where a is the action to be taken at state s.

Another popular behavior policy is Boltzmann exploration, where the probability of choosing an action is given by

$$\text{Prob}\left\{\text{choosing action } a\right\} = \frac{\exp(-Q(s,a))/T}{\sum_{a'} -\exp(Q(s,a')/T)} \tag{5.9}$$

where T is called temperature and initialized to a high value in order to do more exploration initially, and it is decayed over time to do more exploitation as the agent is learning more. At a very very low temperature, Boltzmann policy is equivalent to the greedy policy.

RL estimates Q values using the following equation, enabling the agent to learn the optimal policy directly through the estimated values without requiring the environment model to be known.

$$\hat{Q}_{t+1}(s_t,a_t) = \hat{Q}_t(s_t,a_t) + \alpha_t \delta_t \tag{5.10}$$

where $\hat{Q}_t(s_t,a_t)$ is the current estimation of Q^* at the current state-action pair, (s_t,a_t); $0 < \alpha \leq 1$ is the learning rate at time t; and δ_t is the temporal difference (TD) error at time t, given by:

$$\delta_t = r_{t+1} + \gamma \hat{Q}_t(s_{t+1}, \hat{\pi}(s_{t+1})) - \hat{Q}_t(s_t,a_t) \tag{5.11}$$

where $\hat{\pi}(s)$ is the policy being estimated and improved, known as the *estimation policy*, which generates the next action a_{t+1} for the estimations.

The learning rate in Eq. (5.11) should meet the following conditions for $\hat{Q}$ to converge to Q^*:

$$\sum_t \alpha_t = \infty \tag{5.12}$$
$$\sum_t \alpha_t^2 < \infty$$

1: **procedure** Q-learning
2: Initialize $Q(s, a)$ arbitrarily for all s,a
3: **for all** episodes **do**
4: $t \leftarrow 0$
5: Initialize s_t
6: **repeat**(for each step of the episode):
7: $a_t \leftarrow \epsilon$-greedy(Q_t, s_t) ▷ behavior policy
8: Take action a_t, observe r_{t+1}, s_{t+1}
9: $a_{t+1} \leftarrow \max_a Q_t(s_{t+1}, a)$ ▷ estimation policy
10: $Q_{t+1}(s_t, a_t) \leftarrow Q_t(s_t, a_t) + \alpha[r_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)]$
11: $a_t \leftarrow a_{t+1}$; $s_t \leftarrow s_{t+1}$
12: $t \leftarrow t + 1$
13: **until** s_t is terminal
14: **end for**
15: **end procedure**

Figure 5.1: Q-learning algorithm

In our system, we decrease the learning rate by a factor of 0.99 after each iteration within an episode and reset it to an initial value at the beginning of the next episode.

We should distinguish between the estimation policy, used for generating temporal difference, and the behavior policy, used for generating experience. As a matter of fact, they can be different and that is how different RL algorithms are obtained. If the greedy policy is used for the estimations, we obtain a so-called *off-policy* algorithm because the policy being used estimation is different from the behavior policy. The resulting algorithm called *Q-learning* is shown in Figure 5.1.

If we use the behavior policy also for estimation policy, we obtain an *on-policy* algorithm which is called *Sarsa*, as seen in Figure 5.3.

There are also some methods known as *indirect RL algorithms* where instead of estimating Q values directly, they try to build up the environment model, i.e. $P(s, a, s')$ and $R(s, a)$, from observations and then solve Eq. (5.6) to find the optimal policy. However, the applicability of these algorithms is very limited in practice.

An impacting factor in RL is credit assignment which deals with propagating rewards backward in time and space (across state-action pairs) to update the Q values of the visited state-action pairs while agents are learning forward in time. Q-learning and Sarsa are very hasty and use a one-step update or *backup* which is based on only the next immediate reward and update only the Q value of the last visited state-action pair. These algorithms with zero patience are referred to as $TD(0)$. On the contrary, algorithms like Monte-Carlo in which the agent waits until the end of episode to create a n-step backup (based on the immediate rewards of the previous n steps) to update the Q values of all visited state-action pairs during the episode, are called $TD(1)$. The drawback is that the agent does not benefit from its experience and acts blindly until

procedure SARSA
  Initialize $Q(s,a)$ arbitrarily for all s,a
  for all episodes **do**
    $t \leftarrow 0$
    Initialize s_t
    repeat(for each step of the episode):
      $a_t \leftarrow \epsilon$-greedy(Q_t, s_t) ▷ behavior policy
      Take action a_t, observe r_{t+1}, s_{t+1}
      $a_{t+1} \leftarrow \epsilon$-greedy$(Q_t, s_{t+1})$ ▷ estimation policy
      $Q_{t+1}(s_t,a_t) \leftarrow Q_t(s_t,a_t) + \alpha[r_{t+1} + \gamma Q_t(s_{t+1},a_{t+1}) - Q_t(s_t,a_t)]$
      $a_t \leftarrow a_{t+1}$; $s_t \leftarrow s_{t+1}$
      $t \leftarrow t+1$
    until s_t is terminal
  end for
end procedure

FIGURE 5.2: Sarsa algorithm

the end of episodes. The general case can be considered as $TD(\lambda)$, with λ being the *eligibility trace parameter* which allows a tradeoff between $TD(0)$ and $TD(1)$. In order to practically utilize this general credit assignment, an eligibility trace, a record of the occurrence of past visits, is used as:

$$e_{t+1}(s,a) = \begin{cases} 1 & \text{if } s_t = s \text{ and } a_t = a \\ \gamma\lambda e_t(s,a) & \text{otherwise} \end{cases} \tag{5.13}$$

The eligibility trace for each state-action pair is initialized to 0 at the beginning of each episode. Whenever a state-action pair is visited, its eligibility trace is set to 1, and at each iteration, it is decayed by a factor of $\gamma\lambda$. If a state-action pair is never visited, its eligibility trace will remain zero which means that its Q value will not change. When using an eligibility trace, all state-action pairs must be updated at each iteration where the amount of change for each state-action is proportional to its eligibility trace. The update rule in Eq. (5.10) will change as follows when eligibility trace is used:

$$\forall s,a: \ \hat{Q}_{t+1}(s,a) = \hat{Q}_t(s,a) + \alpha_t \delta_t e_t(s,a) \tag{5.14}$$

In our system, with the Q-learning algorithm, we do not use the eligibility trace whenever the next action is determined by exploration and also that all the eligibility trace values $e_t(s,a)$ should be reset to zero. This is because after exploration, the next backups will no longer have any necessary relationship to the estimation policy.

Calculating Q values by using either Eq. (5.10) or (5.14), however, requires keeping them in tables with size $|S| \times |A|$. The curse of dimensionality problem can arise

procedure Q-learning (λ)
  Initialize $Q(s, a)$ arbitrarily for all s,a
  for all episodes **do**
    $e(s, a) \leftarrow 0$ for all s,a
    $t \leftarrow 0$
    Initialize s_t, a_t
    repeat(for each step of the episode):
      Take action a_t, observe r_{t+1}, s_{t+1}
      $a_{t+1} \leftarrow \max_a Q_t(s_{t+1}, a)$
      $a^* \leftarrow \epsilon$-greedy(Q_t, s_{t+1})
      $e_t(s_t, a_t) \leftarrow 1$
      for all s, a **do**
        $Q_{t+1}(s, a) \leftarrow Q_t(s, a) + \alpha[r_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)]e_t(s, a)$
        if $a_{t+1} = a^*$ **then** ▷ exploiting?
          $e_t(s, a) \leftarrow \gamma\lambda e_t(s, a)$
        else ▷ exploring
          $e_t(s, a) \leftarrow 0$
        end if
      end for
      $a_t \leftarrow a_{t+1}$; $s_t \leftarrow s_{t+1}$
      $t \leftarrow t + 1$
    until s_t is terminal
  end for
end procedure

Figure 5.3: Q-learning algorithm with eligibility trace

when we are faced with problems with a large state-action space, which slows down the agent's learning markedly. This can simply happen when the state variables can take a very large or infinite —when they are continuous— number of possible values. In this case even an exact representation of the Q-function in a tabular format is not possible and we need to estimate them by using an approximation. An approximator can be denoted by an n-dimensional mapping

$$F(\theta) : \mathbb{R}^n \rightarrow \psi \tag{5.15}$$

where θ is the vector of approximator parameters, $\mathbb{R}^n$ is the approximator parameter space, and ψ is the space of Q-functions.

Instead of learning Q-functions directly, agents now try to learn the approximator parameter vector θ which provides a compact representation of the corresponding approximate Q-function as follows:

$$\hat{Q}(s, a) = [F(\theta)](s, a) \tag{5.16}$$

1: **procedure** SARSA(λ)
2: Initialize $Q(s, a)$ arbitrarily for all s,a
3: **for all** episodes **do**
4: $e(s, a) \leftarrow 0$ for all s,a
5: $t \leftarrow 0$
6: Initialize s_t, a_t
7: **repeat**(for each step of the episode):
8: Take action a_t, observe r_{t+1}, s_{t+1}
9: $a_{t+1} \leftarrow \epsilon$-greedy$(Q_t, s_{t+1})$
10: $e_t(s_t, a_t) \leftarrow 1$
11: **for all** s, a **do**
12: $Q_{t+1}(s, a) \leftarrow Q_t(s, a) + \alpha[r_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)]e_t(s, a)$
13: $e_t(s, a) \leftarrow \gamma\lambda e_t(s, a)$
14: **end for**
15: $a_t \leftarrow a_{t+1}$; $s_t \leftarrow s_{t+1}$
16: $t \leftarrow t + 1$
17: **until** s_t is terminal
18: **end for**
19: **end procedure**

FIGURE 5.4: Sarsa algorithm with eligibility trace

This remarkably reduces the problem complexity because normally the parameter space has many fewer dimensions than the space of Q-functions, i.e. $n << |S|$. Another advantage of using the approximation is the notion of *generalization*, which enables the agent to make reasonable decisions in the states which have not been encountered so far, only by approximating from the nearby states experienced before. This can improve decision making and also convergence speed. For further study of using approximations in RL we refer the reader to [11].

5.2 Design Considerations

5.2.1 Reward Function

A reward function is used by RL agents to calculate the immediate reward of taking an action. It reflects the goal of the task being optimized either implicitly or explicitly and can greatly impact the performance of the system in terms of convergence and optimality of the solution. Designing a good reward function which well targets the goal can be sometimes very hard and it is still an open problem. Recently, some researchers studied (e.g. in [110]) the approach of automatically constructing the reward function by the agent itself either from an explicit given fitness function induced directly by the goal with the aid of evolutionary search techniques such as genetic algorithms, or from a

set of expert demonstrations in the absence of an *a-priori* given reward function, known as *apprenticeship learning*. Although these approaches appear to be very promising, their application in the context of WBANs is currently infeasible because they are too resource-consuming which does not suit the very resource-limited WBAN nodes.

In order to motivate how the reward function can affect convergence, let us assume the WBANs in our system are intended to achieve the minimum energy consumption per bit. One may suggest the following reward function which is the negative of energy consumption per bit in Joules/bit as a result of an explicit reflection of the goal assumed:

$$r_i^{(t)} = -\frac{p_i^{(t)}}{c_i^{(t)}} \tag{5.17}$$

where $p_i^{(t)}$ is the transmission power level of WBAN i at iteration t and $c_i^{(t)}$ is the throughput of WBAN i at iteration t. Each agent tries to maximize its episode reward which is:

$$R_i^{(t)} = \sum_{k=t+1}^{T} \gamma_i^{k-t-1} r_i^{(t)} \tag{5.18}$$

At each state, this system can be thought of as a non-cooperative game with the set of RL agents $M = \{1, \cdots, m\}$ as players of the game, the transmission power levels $p_i^{(t)}$ as strategies and Q values as payoffs, bearing in mind that Q values are technically the expected episode reward $R_i^{(t)}$. The Nash Equilibrium (NE) is given by a strategy profile of the best response $p_i^{(t)}$ of each player i to all the other players strategies $p_{-i}^{(t)}$. The best response of player i in the game defined is the root of the following equation:

$$\left.\frac{\partial R_i^{(t)}(p_i, p_{-i})}{\partial p_i^{(t)}}\right|_{p_i^{(t)} = p_i^{*(t)}} = 0 \tag{5.19}$$

Using Eq. (5.18) and (5.17), we get:

$$\sum_{k=t+1}^{T} \gamma_i^{k-t-1} \left[\frac{\log(1+\xi_i^{(k)}) - p_i^{*(k)} \frac{\partial \xi_i^{(k)}}{\partial p_i^{(k)}} \left(\frac{1}{1+\xi_i^{(k)}} \right)}{\log^2(1+\xi_i^{(k)})} \right] = 0 \tag{5.20}$$

which gives:

$$\sum_{k=t+1}^{T} \gamma_i^{k-t-1} \left[\frac{\log(1+\xi_i^{(t)}) - \frac{\xi_i^{(t)}}{1+\xi_i^{(t)}}}{\log^2(1+\xi_i^{(t)})} \right] = 0 \tag{5.21}$$

because we have:

$$p_i^{*(k)} \frac{\partial \xi_i^{(k)}}{\partial p_i^{(k)}} = \xi_i^{(k)} \tag{5.22}$$

Defining $y_i^{(k)} = 1 + \xi_i^{(k)}$ simplifies Eq. (5.21) to:

$$\sum_{k=t+1}^{T} \gamma_i^{k-t-1} \left[\frac{\log(y_i^{(t)}) - \frac{1}{y_i^{(t)}} - 1}{\log^2(y_i^{(t)})} \right] = 0 \tag{5.23}$$

Since the discount factor γ_i and $y_i^{(k)}$ are always non-negative, the above summation can become zero only when the nominator is zero. This gives:

$$y_i^{(t)} \cdot \log(y_i^{(t)}) = y_i^{(t)} - 1 \tag{5.24}$$

The root of this equation is given by the Lambert W function as $y_i^{(k)} = -W(-e^{-1})$ which is equal to one and this gives us the best response of player i as: $p_i^{*(k)} = 0, k \in [t+1, T]$. This simply means to keep the sensor nodes switched off which does not seem a practical solution. In particular, with the non-cooperative agents which are interested to maximize their own rewards selfishly, this reward function does not admit a NE at which power levels stabilize. However, a stable solution may be attained by some cooperation between the agents which is not of our interest.

Now we revise the reward function as follows:

$$r_i^{(t)} = \frac{c_i^{(t)}}{C_{\max_i}} - w_p \frac{p_i^{(t)}}{P_{\max_i}} - 1 \tag{5.25}$$

where $c_i^{(t)}$ is is the throughput of WBAN i at iteration t, $C_{\max_i}$ is the maximum channel capacity calculated at zero interference used for normalization; $p_i^{(t)}$ is the transmission power level of WBAN i at time t, $P_{\max_i}$ is the maximum allowable transmission power of WBAN i; and, w_{p_i} is the price factor.

Using the same approach stated earlier, one can calculate the best response of player i as:

$$p_i^{*(t)} = \frac{P_{\max_i} B \log(2)}{w_p C_{\max_i}} - \frac{1}{\eta_i^{(t)}} \tag{5.26}$$

where $\eta_i^{(t)}$ is the sensitivity of the SINR at WBAN i.

A change in $\eta_i^{(t)}$ is a consequence of a change in the arrangement of WBANs, the transmission power levels of other WBANs and channel gains. Therefore, as WBANs

move around, NE in the system changes and the RL agents are faced with learning a non-stationary task with a moving terminal state. However we show through simulation that once they learn to reach the NE in an arbitrary arrangement, they are able to find the NE for the next given arrangements.

5.2.2 Impact of Immediate Rewards

Another important factor which can affect convergence and the optimality of solution is the range of immediate reward values. Considering $r_i^{\min} < r_i < r_i^{\max}$, three cases can happen:

1. $r_i^{\max} > r_i^{\min} \geq 0$: In this case the terminal state should be rewarded explicitly or a penalty should be considered for not being at the terminal state at each iteration otherwise the convergence speed may deteriorate. This is due to the fact that the episode reward is a discounted summation of these non-negative immediate rewards over an episode, and in this case, the longer the episode is, the greater the episode reward will be, which may be preferred by agents rather than reaching a solution with better quality in less number of iterations.

2. $r_i^{\max} \geq 0 > r_i^{\min}$: In this case also a special reward for reaching the terminal state or a penalty for not being at the terminal state at each iteration should be considered otherwise agents may compromise the optimality of the solution by longer episodes and not have any incentive to find a solution as fast as they can, particularly when the discount factor is large.

3. $r_i^{\min} < r_i^{\max} < 0$: In this case convergence is implicitly considered by the reward function and the terminal state is not necessarily needed to be rewarded. This case is also suitable for problems with moving terminal states such as our system where the terminal state is not known *a-priori* to be rewarded by the system designer, or it should not even be rewarded at all because it may change each time and does not remain the same state. The subtracting term (-1) in the reward function defined in Eq. (5.25) ensures we have negative immediate rewards in our system. In all three cases, the immediate rewards should be bounded for the RL algorithm to converge to the optimal policy.

5.2.3 Impact of Initial Q Values

Another factor which may affect the convergence and the optimality of the solution is the initial values of the Q-table. It can be easily shown that the episode reward lies in the following range:

$$\frac{r_i^{\min}}{1-\gamma_i} < R_i^{(t)} < \frac{r_i^{\max}}{1-\gamma_i} \tag{5.27}$$

If Q values are initially set to a value greater than the upper bound, this will lead to *optimistic initialization* [111], which promotes the exploration early in the learning phase because untried actions are always indicated to be better than the already tried ones. This can improve the optimality of the solution due to providing a broad exploration at early stages of learning, but can lead to a poor convergence because it may take quite a lot of time for the algorithm to get rid of optimism, propagate actual reward of actions, and converge to a stationary policy. In this study, we employ the optimistic initialization by setting the Q values to zero initially.

5.2.4 State Representation

Agents in RL make decisions based on only the current observation of the reinforcement signal (the next state and immediate reward) regardless of the sequence of past observation triples $\langle a_k, s_k, r_k \rangle, k = 1, 2, \cdots, t$. This means that RL is potentially able to solve only tasks where the state signal has the property to be a representative of the history of agents past interactions with environment, namely the Markov property. As a result, the state definition in RL problems is very important because the current state should summarize the previous observations in a compact and informative way to provide the agent with a good basis for predicting subsequent states and future rewards as well as selecting actions. An improper state representation can introduce non-Markov states and dramatically degrade the efficiency of the learning or even not leading to a solution. In our power control problem, as Eq. (5.26) shows, the NE in the system is a function of the sensitivities of the SINR to transmission power level:

$$\mathbf{P}^*(t) = f\left(\eta(t)\right) \tag{5.28}$$

$$\eta(t) = (\eta_i^{(t)})_{i=1}^m \tag{5.29}$$

This intuitively suggests to use the sensitivity of the SINR to the transmission power level as a Markov state variable because the power vector $\mathbf{P}^*(t)$ is dependent on only the current $\eta(t)$ regardless of the history of WBANs' movements, power levels and channel gains. We define the tupple $(p_i^{(t)}, \eta_i^{(t)})$, i.e. the current transmission power of BN i and the value of the sensitivity of the SINR at BNC i as the state of the environment from the view point of WBAN i at time t.

5.2.5 Approximation

Since in our system the state variables, i.e. the transmission power and the sensitivity of SINR, are continuous, we should use approximation to cope with the curse of dimensionality issue. To this end, we use Radial Basis Functions (RBF) [112] to approximate

Q-functions. RBF approximators lie in the category of linear approximators, as seen in the following:

$$\hat{y}(t) = \sum_{i=1}^{N} \theta_i \phi_i \left(\|x - x_i\| \right) \tag{5.30}$$

where $\hat{y}(t)$ is the function to be approximated and is represented as a sum of N radial basis functions $\phi(.)$, each associated with a different center x_i, and weighted by an appropriate coefficient θ_i. In this study, we use Gaussian shape basis functions as:

$$\phi(r) = \exp(-r/b)^2, r = \|x - x_c\| \tag{5.31}$$

where x_c and b are the center and the width of the Gaussian basis function respectively; and r is the Euclidean distance from the center point.

5.3 Performance Evaluation

Each run of the simulation has two phases which are a learning phase and a testing phase. For each scenario of a certain number of WBANs in the system, the learning phase starts with a random initial arrangement of the WBANs followed by running a number of episodes until the agents learn to find the optimal solution in the least number of iterations. At each arrangement, an episode starts with a random initial power vector and finishes when the power vector converges to a stable point, i.e. no WBAN changes its power anymore. In the testing phase, however, WBANs move around the room according to a random walk model and they just employ their knowledge to reach the solution and at the same time improve their policy. During the learning phase, the exploration rate is decayed from 1 by a factor of 0.9 after each episode. However, during the testing phase in which the environment (arrangement, channel, power levels) changes, in order to let agents adapt to the environment changes and be able to find the optimal solution, the exploration rate is kept at a fixed but small value, namely 0.05, that is 5% exploration. The learning rate is set to an initial value at the beginning of each episode and is decreased by a factor of 0.99 after each iteration to satisfy Eq. (5.12) for the sake of convergence.

Table 5.1 summarizes the parameters and their values used in the simulations.

5.3.1 Effects of RL Parameters

Figures 5.5 and 5.6 show the performance of RLPC with respect to the initial values of the learning rate (alpha) for three scenarios of 4, 16 and 32 WBANs in the system. Figure 5.5 represents the optimality of solution in terms of the average energy consumption per bit. As it illustrates, the energy consumption per bit decreases dramatically

TABLE 5.1: Simulation Parameters and Values

Parameter Name	Symbol	Parameter Value
Bandwidth	B	300 kHz
Noise	n_i	-174 dBm/Hz
Maximum Power	$P_{\max_i}$	-16 dBm
Initial Exploration Rate	ϵ_0	1
Exploration Rate Decay		0.9
Initial Learning Rate	α_0	0.7
Learning Rate Decay		0.99
Discount Factor	γ	0.1
Eligibility Trace Parameter	λ	1
Number of RBF approximators (per each state variable)		5
Number of action (power) levels		10

when alpha increases from 0 and almost remains unchanged for alpha greater than 0.5. Figure 5.6 demonstrates the number of iterations needed to converge to a stable solution. It indicates that the convergence improves as alpha increases and WBANs find the solution in less number of iterations. We conclude that both the optimality and convergence improve by using greater values for the initial learning rate.

The reason why the initial value of the learning rate affects the performance of the system is explained by the fact that at the begining of each episode in the training phase, the learning rate is reset to the initial value. Therefore it determines to what extent RL agents learn from early actions taken at the beginning of each episode. The figures reveal that the early actions of each episode are very important and learning more from them, considerably improves the performance of the system.

Figures 5.7 and 5.8 represents the performance of the system with respect to the eligibility trace parameter (lambda) for three scenarios of 4, 16 and 32 WBANs in the system. As seen in Figure 5.7, the quality of the solution does not show any dependency to lambda because energy consumption per bit does not change with lambda and is almost constant for the three scenarios. Figure 5.8 illustrates that convergence improves as lambda increases. Increasing the eligibility trace parameter, however, increases the amount of calculations which has not been considered in our system model because we are assuming that BNC nodes run the power control algorithms. In the systems where BN nodes are responsible for running the power control algorithm, this calculation overload should be also taken into account.

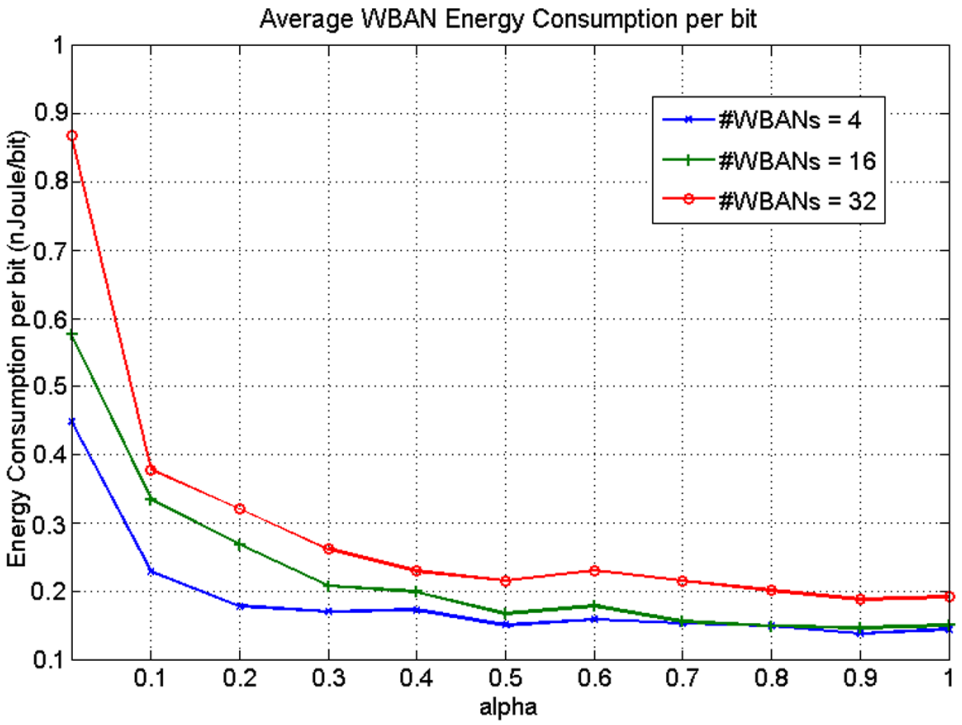

Figure 5.5: The average energy consumption per bit versus the initial learning rate

Figures 5.9 and 5.10 represents the performance of the system with respect to the discount factor (gamma) for three scenarios of 4, 16 and 32 WBANs in the system. As seen in Figure 5.9) while energy consumption per bit remains almost unchanged for small values of gamma up to 0.5, it gradually starts to increase after that and it increases substantially once gamma becomes greater than 0.8. Figure 5.10 also shows that convergence also deteriorates and WBANs need more iterations to converge to the solution when gamma increases. This suggests that the best value for the discount factor in our system is 0 which makes the agent completely near-sighted in terms of maximizing rewards. However this will also disable the eligibility trace (see Eq. (5.13)).

5.3.2 WRLPC versus WPCG and WFPC

Figure 5.11 represents the average power as a function of the number of WBANs in the system for different controllers. As it can be clearly seen, WRLPC transmits at the least transmission power compared to the other two approaches. While the fuzzy power controller, WFPC, transmits at almost 6 μW less power than WPCG, the RL-based power controller, WRLPC, consumes 10 μW less power than WPCG and almost 3 μW less power than WFPC for transmissions.

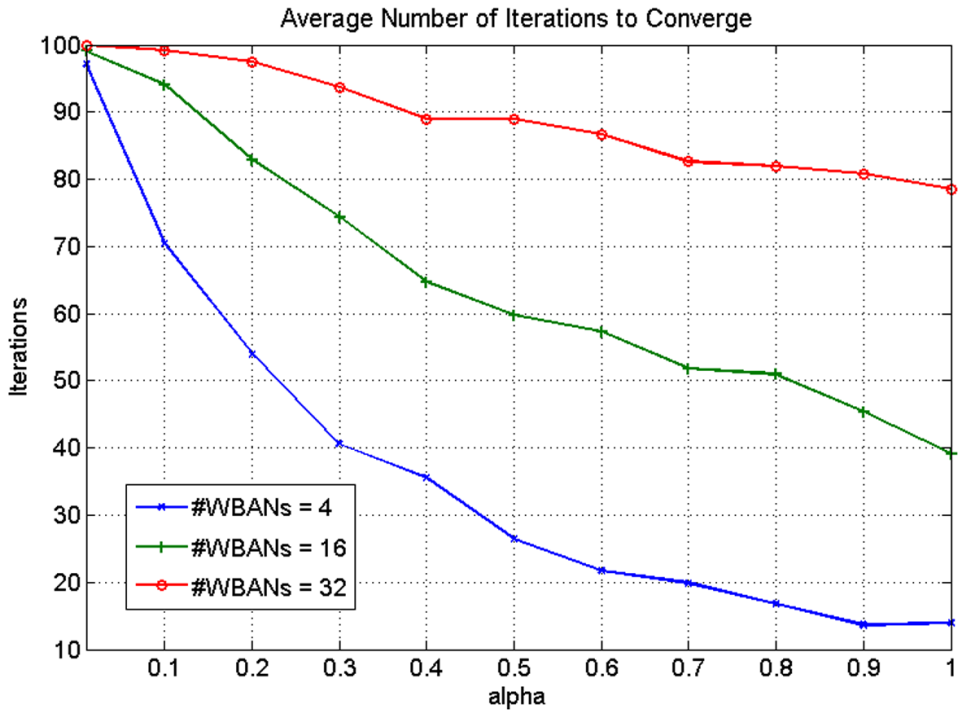

FIGURE 5.6: The average number of iteration versus the initial learning rate

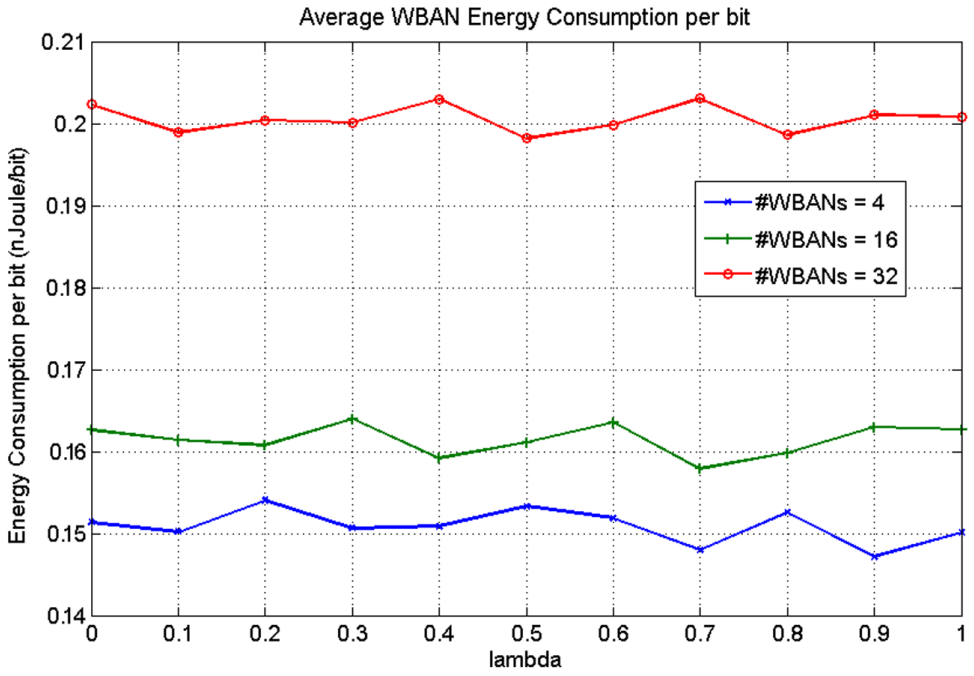

FIGURE 5.7: The average energy consumption per bit versus the eligibility trace parameter

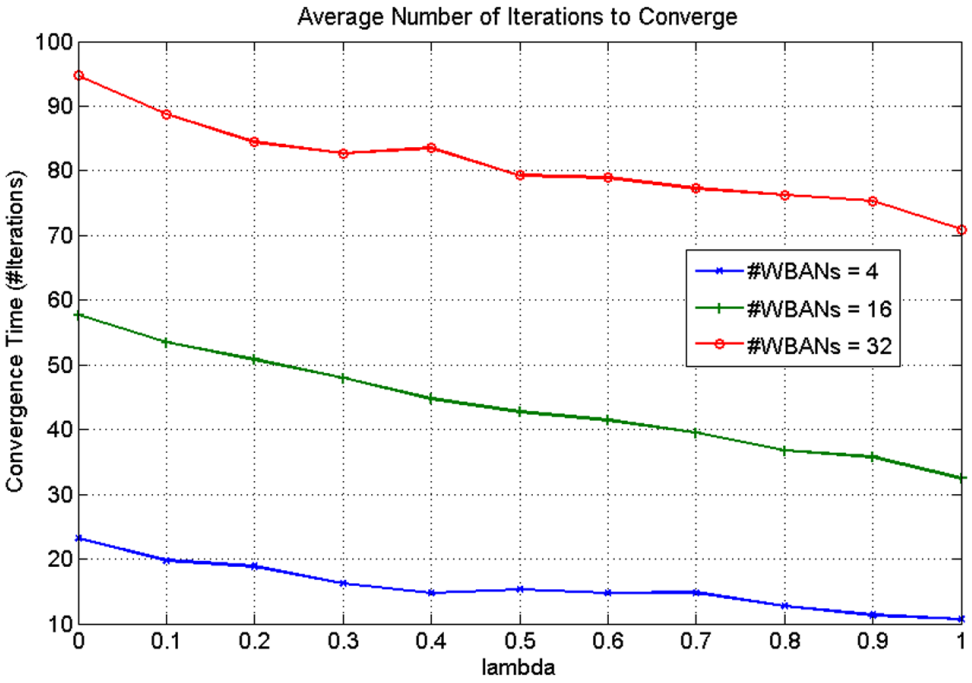

Figure 5.8: The average number of iteration versus the eligibility trace parameter

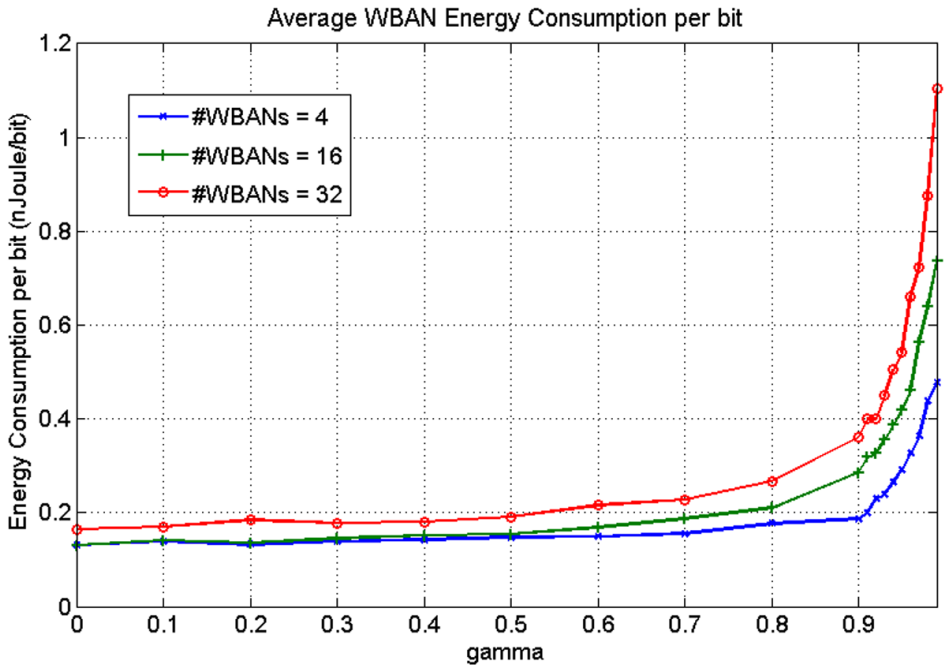

Figure 5.9: The average energy consumption per bit versus the discount factor

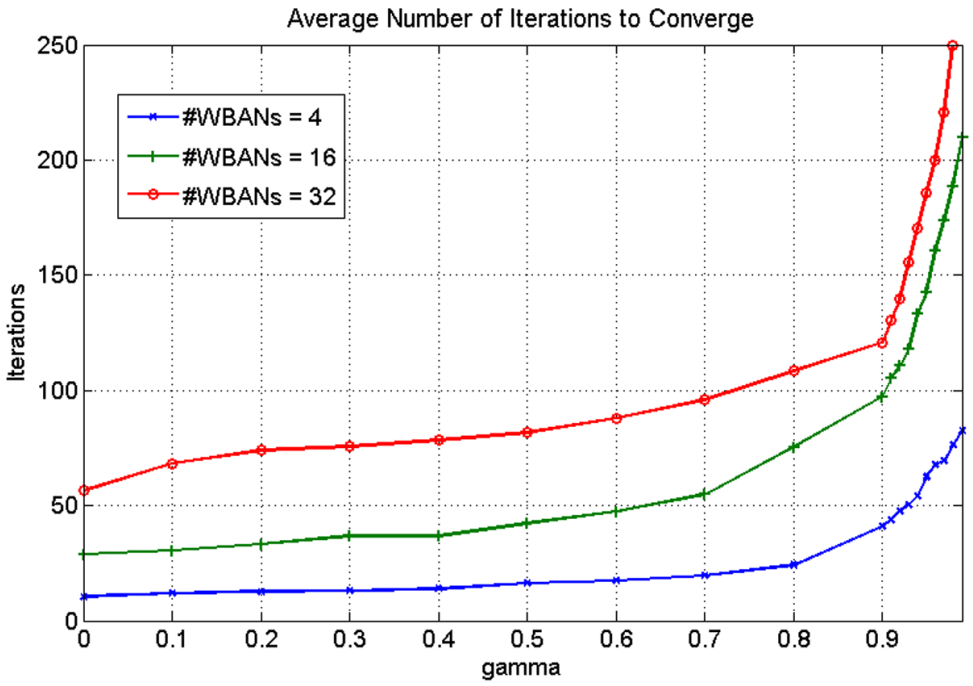

FIGURE 5.10: The average number of iteration versus the the discount factor

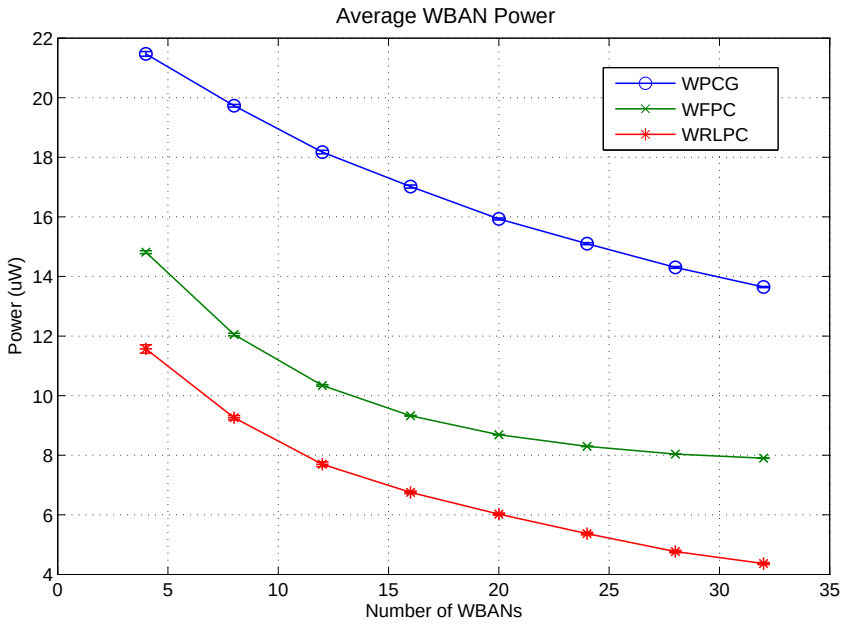

FIGURE 5.11: The average transmission power level versus the number of WBANs

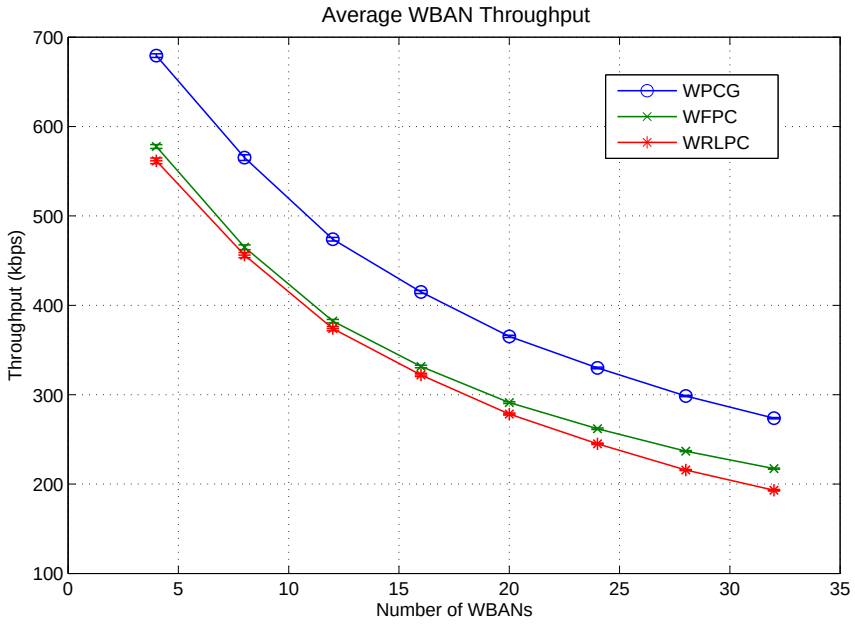

Figure 5.12: The average throughput versus the number of WBANs

Figure 5.12 shows the average throughput versus the number of WBANs in the system. The graph illustrates that WRLPC is slightly outperformed by WFPC, while both are beaten by WPCG. WRLPC delivers almost 10 kbps (i.e. around 2% under sparse and 5% under dense conditions) less throughput than WFPC and 110 kbps (i.e. around 15% under sparse and 30% under dense conditions) less throughput than WPCG. However, WRLPC saves a notable amount of power at the expense of sacrificing such throughput. In order to find out how each controller adjusts the tradeoff between power and throughput, we should look at the energy consumption per bit.

Figure 5.13 shows the average energy consumption per bit in nJoul/bit versus the number of WBANs in the system. As it can be clearly seen, WRLPC consumes the least energy per bit and strongly surpasses WPCG and WFPC. For all the controllers, the energy consumption per bit increases as the number of WBANs in the system goes up which can be explained by the increased inter-network interference leading to a decreased SINR. The rate of this rise is almost 1.2 nJoul/bit per WBAN for WPCG, 0.3 nJoul/bit per WBAN for WFPC and 0.07 nJoul/bit per WBANs for WRLPC. The energy consumption of WRLPC is only 60% of that of WPCG and 75% of that of WFPC under sparse condition, and these figures even drop and reach 40% and 50% respectively under dense condition. This concludes that WRLPC reduces energy consumption per bit by 40%-60% compared to WPCG, and by 25%-50% compared to WFPC.

Figure 5.14 shows the number of iterations needed by each approach to reach the steady state versus the number of WBANs in the system. It is noticed that WRLPC

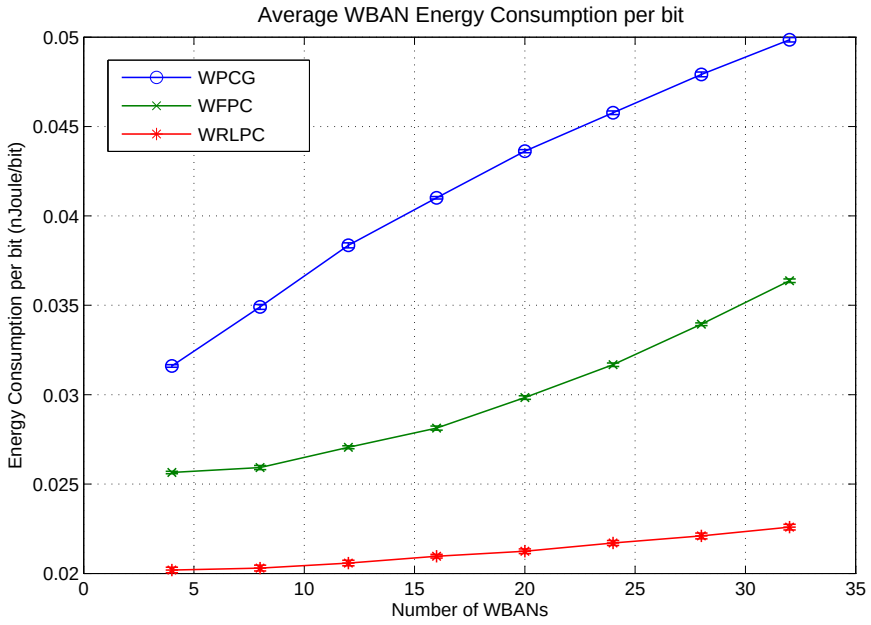

FIGURE 5.13: The average energy consumption per bit versus the number of WBANs

does not converge as fast as WPCG and WFPC and is outperformed strongly by them. This is the only drawback of the reinforcement learning approach compared to the fuzzy and game approaches.

Finally, the average lifetime of WBANs in days versus the number of WBANs in the system is shown in Figure 5.15, given that they transmit continuously. Note that we assume each BN is equipped with a 90 mWh Lithium coin battery and only transmission power is considered for energy consumption. We can vividly see that the WBANs featured with WRLPC survive longer than those with WPCG and WFPC. The average WBAN's lifetime with WRLPC is roughly 4 times longer than that of with WPCG and 1.3 times longer than that of with WFPC.

Discussion on the Results

The improvement of WRLPC over WPCG and WFPC is basically due to two factors. Firstly, WRLPC maximizes an agents rewards by taking the states of the environment into consideration. Secondly its dynamic behavior helps agents learn from their past experience and increase their knowledge of the environment to make better decisions. Game theory, genetic fuzzy systems and reinforcement learning are three different tools which have been extensively used in the machine learning area. The common concept between these techniques is having agents take the best actions to reach a goal with the most reward. Although all these techniques contribute to better decision making, they look at the problem from different points of view and try to solve the problem

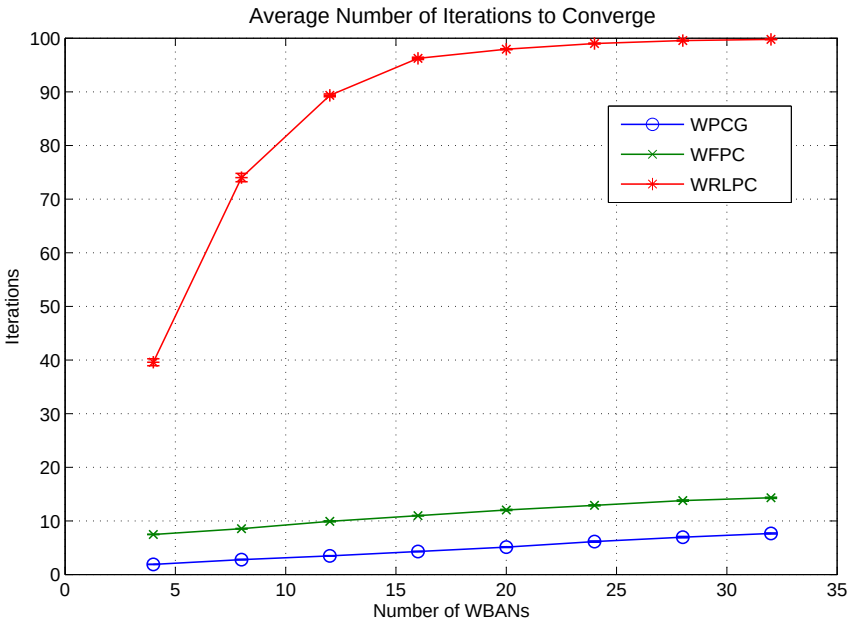

Figure 5.14: The average number of convergence iterations versus the number of WBANs

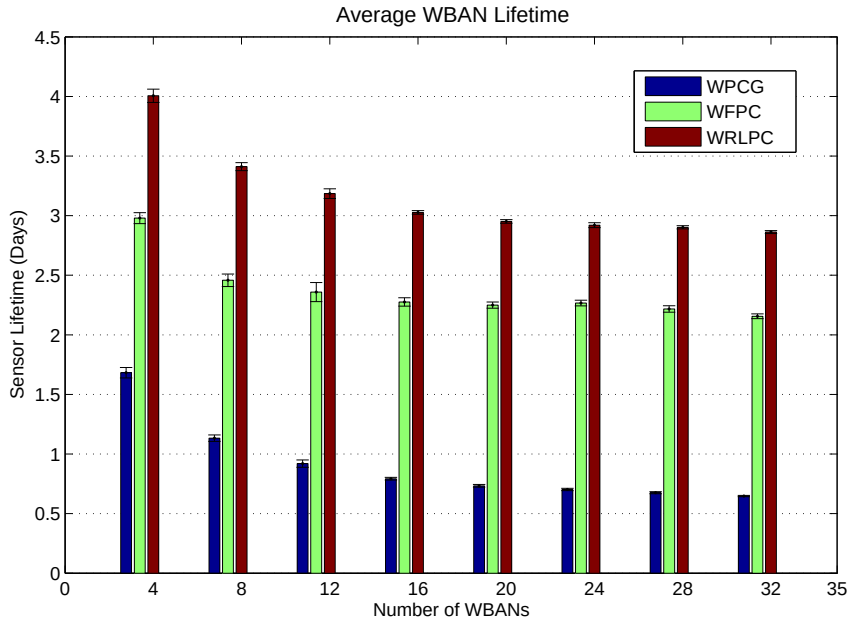

Figure 5.15: The average network lifetime versus the number of WBANs

using different approaches.

WPCG is a power controller based on the game theory where agents take the actions according to the best response strategy, which will maximize their reward function and eventually settle the game at the Nash equilibrium point. However, the best response is calculated directly from the reward function. As a result, the reward function implicitly has all the information agents need to know and it is the only thing which is directly taken into account by them to make decisions and reach the NE. In other words, agents in WPCG control their transmission power level regardless of the direct use of the current environment state. Although the states of the environment can have some effects on the value of the reward function, agents do not make use of them directly. Moreover, the NE itself is not necessarily the optimal solution but just a stable solution.

On the other hand, agents in WFPC, the fuzzy power controller, make decisions regarding only the current state of environment described by SINR, interference power level and transmission power level which are the controllers inputs. Although WFPC has been optimized by genetic algorithms to maximize a reward function during the design stage, agents do not consider the reward function afterward and make decisions only based on the current state of environment.

In contrast to WPCG and WFPC, agents in WRLPC, not only try to maximize their long-term rewards, but also take into account the current state of the environment directly to choose actions. Furthermore, WRLPC is quite dynamic in the sense that agents learn constantly from their experience and improve their policy to make better decisions, while WPCG and WFPC suffer from the lack of such a dynamic learning mechanism and everything needs to be considered accurately in advance during the design stage. Another advantage of WRLPC over WPCG and WFPC is its adaptability to unforeseen situations which may happen in practice and have not been modeled during the design stage. Due to its dynamic behavior, WRLPC is capable of learning from experience and adapting to any changes gracefully.

5.3.3 Various RL Algorithms

Figure 5.16 shows the average transmission power level of WBANs with respect to the number of WBANs in the system. As can be seen, all the RL-based power controllers achieve lower power levels than the counterpart game. The difference between power levels in Sarsa, which shows the lowest power level among the controllers, and the counterpart game is roughly 6 dBm for all the system densities. However, all the power controllers behave similarly and decrease power levels when the system density increases in order to control the interference.

Figure 5.17 depicts the average throughput of WBANs with respect to the number of WBANs in the system. As can be seen, the counterpart game outperforms the RL-based power controllers and delivers almost 100 kbps more throughput. However, the RL-based approaches perform almost the same and still deliver a reasonable

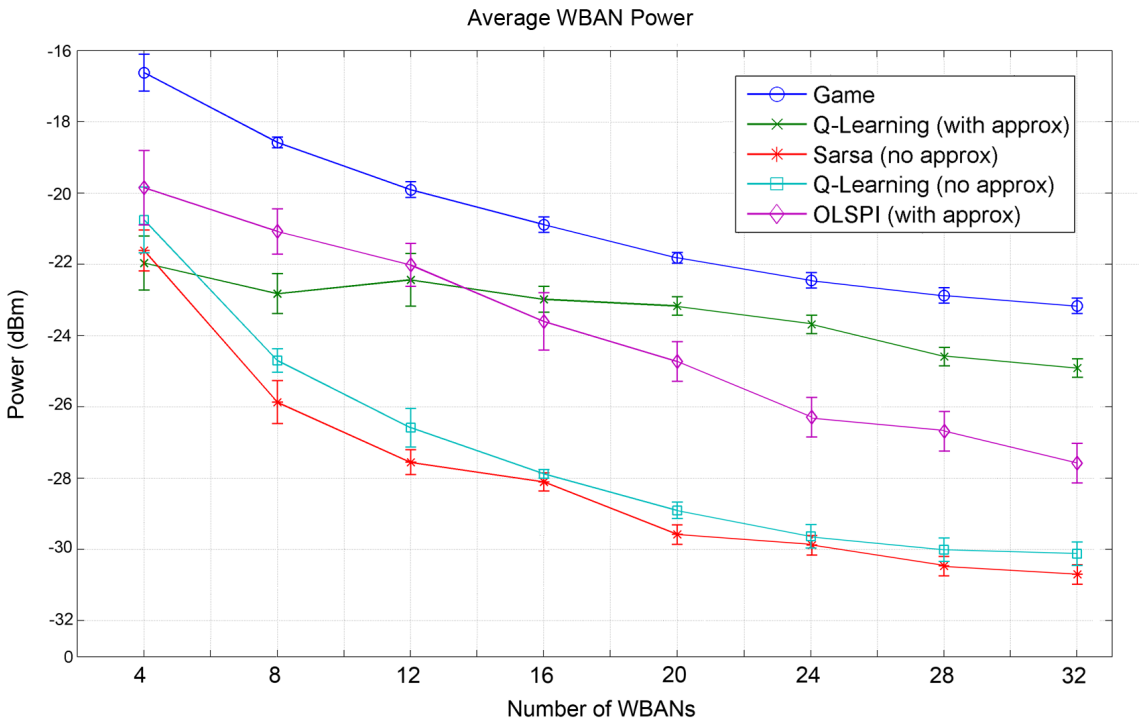

Figure 5.16: Average transmission power versus the number of WBANs

throughput. Figures 5.16 and 5.17 reveal that the RL-based power controllers sacrifice throughput to some extent for less power consumption. In order to find out how good this tradeoff is, we should have a look at Figure 5.18.

Figure 5.18 represents the average energy consumption of WBANs for transmitting one bit in nJouls/bit with respect to the number of WBANs in the system. The graphs show that the RL-based approaches perform almost the same and all outperform the counterpart game by achieving a less energy consumption per bit for any system density. However, for all the approaches, the energy consumption per bit increases when the number of interfering WBANs increases in the system.

Figure 5.19 compares the convergence time (the number of iterations) to reach NE with respect to the number of WBANs in the system. As expected, the counterpart game outperforms the RL-based approaches. This is due to the fact that the agents in the counterpart game are aware of the NE and are designed to reach it. However, the ignorance of the agents about a pre-designed NE in the RL-based power controllers conveys an important advantage and that is their ability to adapt to dynamic changes of the environment because they are not biased to any pre-designed NE but designed to find an optimal solution by interacting with the environment.

Finally, Figure 5.20 demonstrates the average lifetime of WBANs versus the number

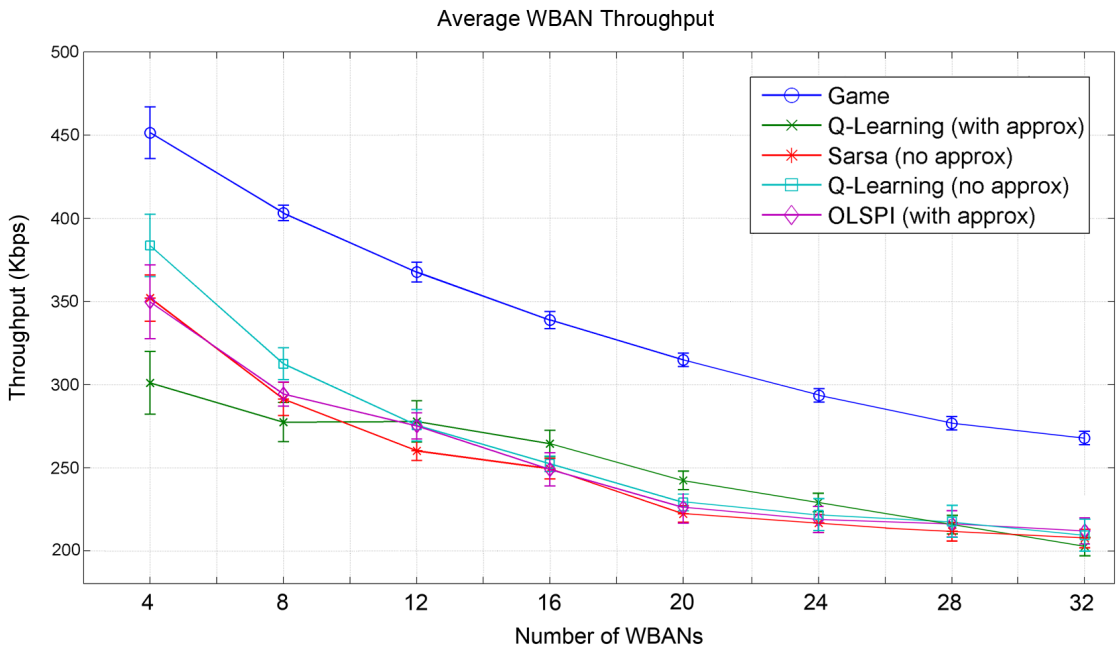

FIGURE 5.17: Average throughput versus the number of WBANs

of WBANs in the system. As it can be seen, WBANs with the RL-based power controllers live longer than those with the counterpart game power controller. This means that the proposed RL-based power controllers save energy more than the counterpart game.

5.4 Conclusions

We proposed a lightweight power controller based on reinforcement learning, namely WRLPC, to mitigate inter-network interference in WBANs. WRLPC learns from experience and improves its performance.

We showed through simulation that the proposed RL-based power controller provided a better tradeoff between throughput and power leading to 3 μW less power consumption for sacrificing 2%-5% of throughput compared to WFPC, and saving 6 μW of power for sacrificing 15%-30% of throughput compared to WPCG. Moreover, WRLPC was also able to improve energy consumption per bit by 40%-60% compared to WPCG, and by 25%-50% compared to WFPC. However, it was outperformed by WFLPC and WPCG in terms of convergence.

We also investigated the impact of reinforcement learning key factors including the

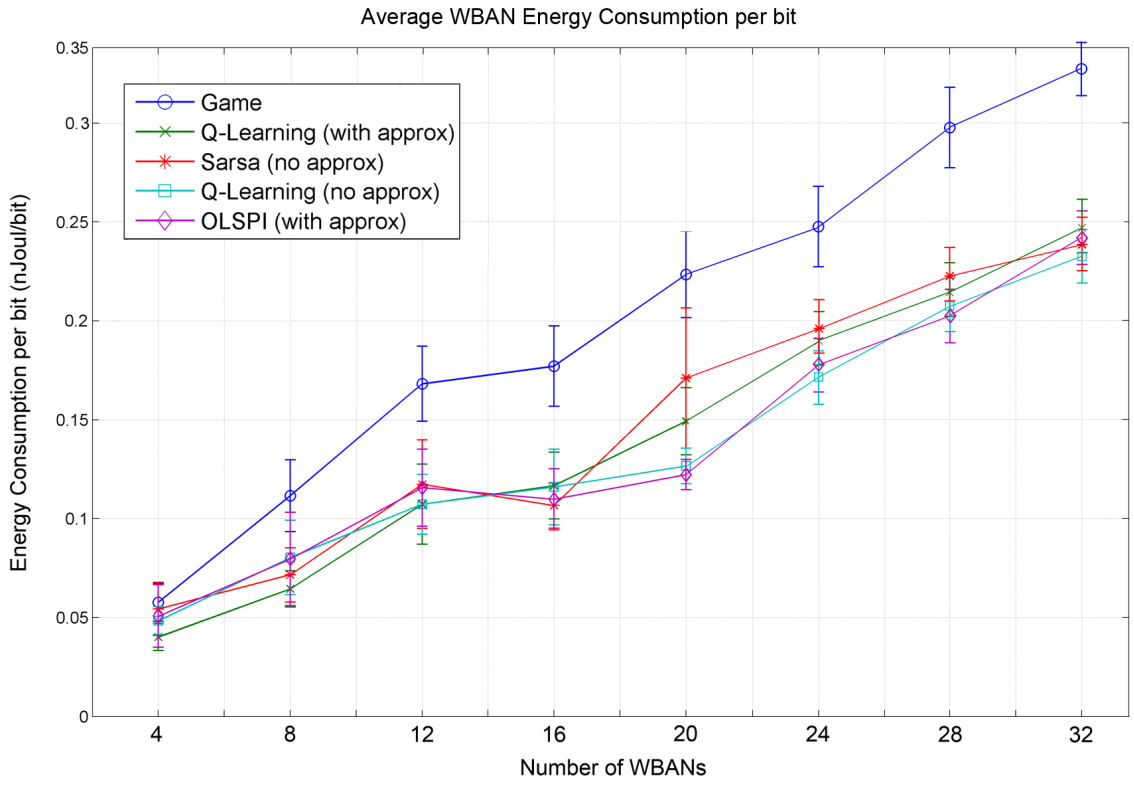

Figure 5.18: Average energy consumption per bit versus the number of WBANs

reward function, discount factor, learning rate and eligibility trace parameter on the performance of the system in terms of convergence and optimality. It was showed that increasing the learning rate could improve both convergence and energy consumption per bit. While increasing the eligibility trace parameter does not affect energy consumption per bit, it improves convergence, and finally, increasing the discount factor aggravates both energy consumption per bit and convergence.

Moreover, we evaluated and compared the performance of the different RL algorithms including Q-learning, sarsa and OLSPI to that of a counterpart non-cooperative game. Although RL-based approaches were outperformed by the counterpart game in terms of convergence, they were able to save more energy. More importantly, WBANs in the RL-based approaches do not need to be aware of a pre-designed Nah equilibrium as opposed to the game. This increases their adaptability to the dynamic changes of the environment.

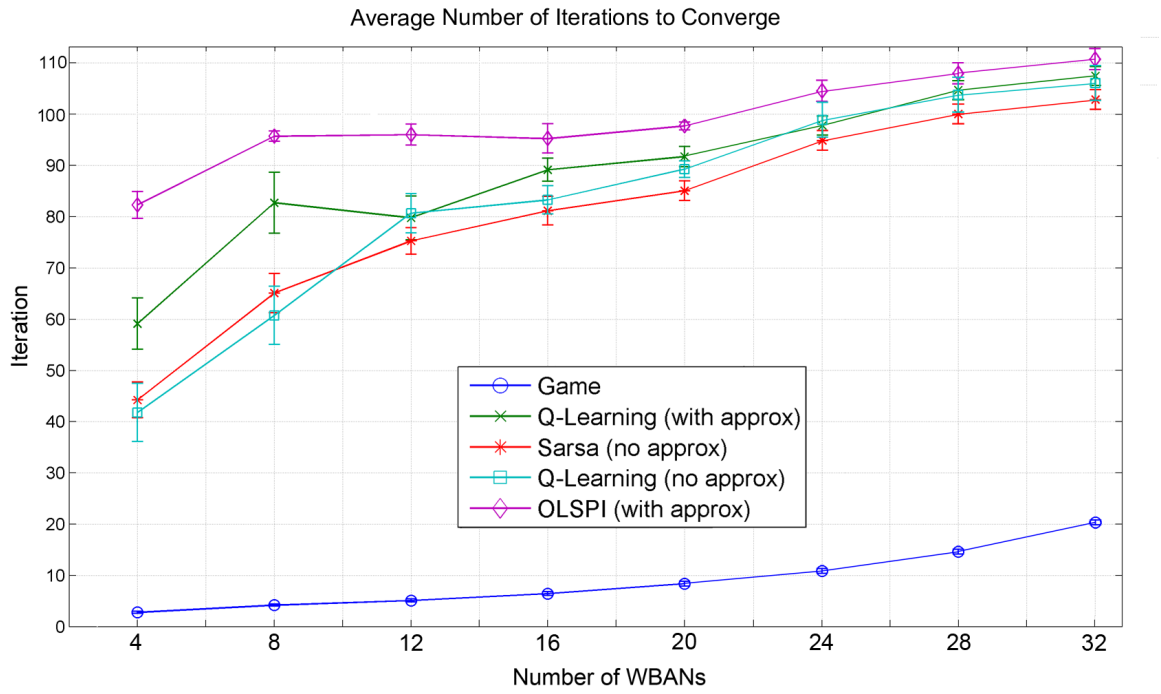

FIGURE 5.19: Average convergence iterations versus the number of WBANs

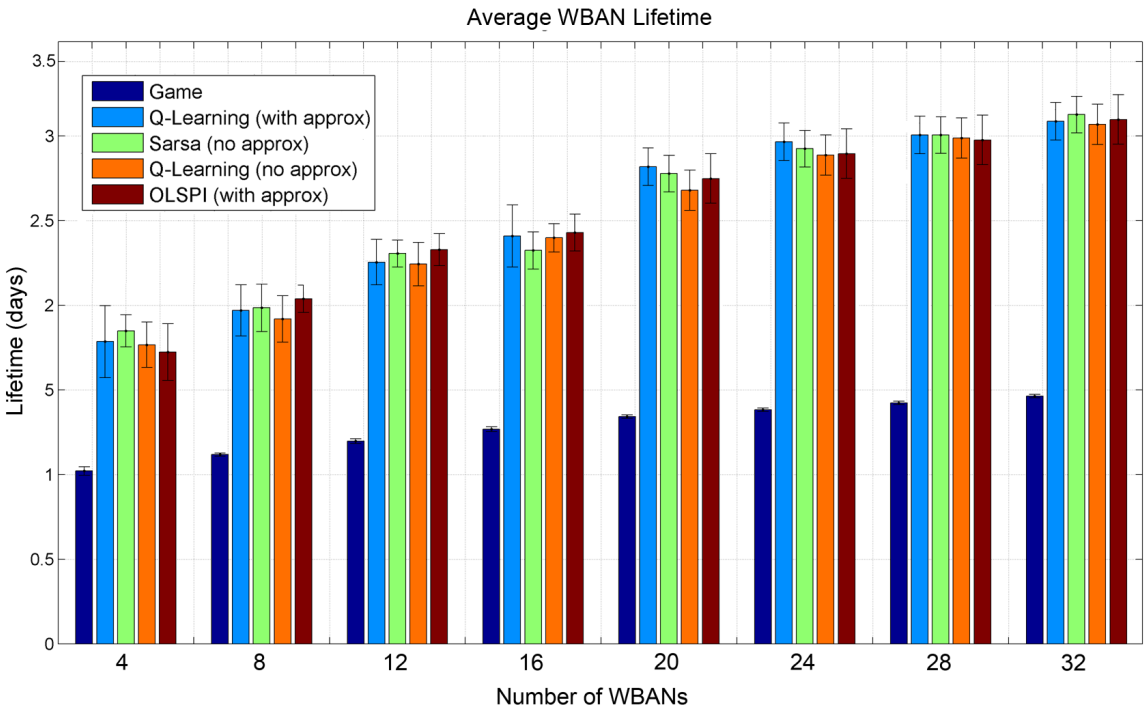

FIGURE 5.20: Average network lifetime versus the number of WBANs

"Make things as simple as possible, but not simpler."

Albert Einstein

6 Minimizing Power for Target Rate

In the previous chapters, we proposed three approaches which made a tradeoff between data rate and power. However, there are some medical applications with emergency traffic such as a surgery operation or ICU/CCU treatments, which require a certain level of QoS to be guaranteed. In this chapter, we take the non-tradeoff approach and aim to meet QoS requirements of WBANs at any cost. Although, the optimization problem we consider for this scenario is a classic optimization problem which has been considered before (e.g. in [8]) and also a distributed solution for it has been proposed by Foschini in [54], we present a different approach and formulation for solving this problem and employ the Jacobi method for fixed-point calculations, which leads to an asynchronous distributed power control algorithm.

6.1 Problem Formulation

We assume that each WBAN in our system demands to achieve a target (data) rate to meet its QoS requirements. For the sake of notational simplicity and without any loss of generality, we assume that the target data rate is the same for all WBANs and is denoted by r_{min}.

We formulate the power control problem as an optimization problem where the goal is to minimize the total power consumption in the system subject to the target data

rate constraint. The optimization problem is defined as follows:

$$\begin{aligned} &\min_{\mathbf{p}\in\mathcal{P}} f(\mathbf{p}) = \sum_{i=1}^{m} p_i \\ &\text{subject to } c_i \geq r_{min} \ , \forall i \\ &\text{variables } \mathbf{p} \end{aligned} \tag{6.1}$$

where $\mathcal{P} = \{\times\mathcal{P}_i\}_{\forall i\in M}$, $\mathcal{P}_i = [0, P_{\max_i}]$; and c_i is the throughput of WBAN i given by Eq. (3.1).

Proposition 4.1: *The optimization problem defined by (6.1) is convex.*

Proof: Since the objective function $f(\mathbf{p})$ is linear and $\mathcal{P}$ in the regional constraint is a convex subset on $\mathbb{R}^m$, for the optimization problem to be convex, just the inequality constraint should be convex. We can rearrange the data rate constraints as follows:

$$\begin{aligned} &c_i \geq r_{min} \\ &\Rightarrow B\log(1+\frac{h_{ii}p_i}{\sum_{j\neq i}^{m} h_{ji}p_j + n_i}) \geq \ r_{min} \\ &\Rightarrow 1+\frac{h_{ii}p_i}{\sum_{j\neq i}^{m} h_{ji}p_j + n_i} \geq 2^{r_{min}/B} \\ &\Rightarrow h_{ii}p_i - (2^{r_{min}/B}-1)\sum_{j\neq i}^{m} h_{ji}p_j - (2^{r_{min}/B}-1)n_i \geq 0 \\ &\Rightarrow g_i(\mathbf{p}) \geq 0 \end{aligned} \tag{6.2}$$

The new constraint function obtained, $g_i(\mathbf{p})$, is an affine function with respect to $\mathbf{p}$ and the constraints in Eq. (6.2) constitute a system of m inequalities which are linear in $\mathbf{p}$. This concludes that the optimization problem defined by Eq. (6.1) is convex.

■

The optimization problem with the new constraints obtained is as follows:

$$\begin{aligned} &\min_{\mathbf{p}\in\mathcal{P}} f(\mathbf{p}) = \sum_{i=1}^{m} p_i \\ &\text{subject to } g_i(\mathbf{p}) \geq 0 \ , \forall i \\ &\text{variables } \mathbf{p} \end{aligned} \tag{6.3}$$

6.2 Problem Solution

6.2.1 Centralized Solution

As the problem at hand is a convex constrained optimization problem, we can employ the method of Lagrange multipliers to obtain a solution which can be the global optimum solution under some conditions which will be stated later. The Lagrangian function for the optimization problem is given by:

$$\begin{aligned} L(\mathbf{p},\mathbf{u}) &= f(\mathbf{p}) - \mathbf{u}^{\mathrm{T}} \cdot \mathbf{g}(\mathbf{p}) \\ &= \sum_{i=1}^{m} p_i - \sum_{i=1}^{m} u_i [h_{ii} p_i - (2^{\frac{r_{min}}{B}} - 1) \sum_{j \neq i}^{m} h_{ji} p_j - (2^{\frac{r_{min}}{B}} - 1) n_i] \end{aligned} \tag{6.4}$$

where $\mathbf{g} = (g_1, \ldots, g_m)^{\mathrm{T}}$ is the vector of the new affine constraint functions and $\mathbf{u} = (u_1, \ldots, u_m)^{\mathrm{T}} \in \mathcal{U}$ is the vector of the Lagrange multipliers from the convex subset $\mathcal{U} = \{\mathbf{u} \in \mathbb{R}^m \mid u_i \geq 0, \forall i \in M\}$.

The Lagrangian function $L(\mathbf{p},\mathbf{u})$ can be thought of as the payoff function of a two player zero-sum game where $\mathbf{p} \in \mathcal{P}$ is the strategy of the first player $P1$ and $\mathbf{u} \in \mathcal{U}$ is the strategy of the second player $P2$, being chosen independently. P1 pays an amount of $L(\mathbf{p},\mathbf{u})$ to P2 (or equivalently P1 gains that from P2 if $L(\mathbf{p},\mathbf{u}) < 0$). P1 is concerned with making a large payoff to P2, so trying to minimize it, and on the other hand, P2 is worried about receiving a small payoff from P1, so willing to maximize it.

In the worst case, P1 expects that his choice $\mathbf{p}$ would lead to a payoff of at most $L^*(\mathbf{p}) = \sup_{\mathbf{u} \in \mathcal{U}} L(\mathbf{p},\mathbf{u})$ to P2. Minimizing this, P1 is faced with a so-called *min-max* optimization problem as follows, known as the *Lagrange primal problem*:

$$\min_{\mathbf{p} \in \mathcal{P}} \sup_{\mathbf{u} \in \mathcal{U}} \{L(\mathbf{p},\mathbf{u})\} \tag{6.5}$$

On the other hand, if P2 chooses $\mathbf{u}$, then in the worst case the payoff he receives is at least $L_*(\mathbf{u}) = \inf_{\mathbf{p} \in \mathcal{P}} L(\mathbf{p},\mathbf{u})$ from P1. Maximizing this, P2 deals with solving a so-called *max-min* optimization problem, referred to as the *Lagrange dual problem*:

$$\max_{\mathbf{u} \in \mathcal{U}} \inf_{\mathbf{p} \in \mathcal{P}} \{L(\mathbf{p},\mathbf{u})\} \tag{6.6}$$

The dual problem is often easier to solve than the primal one because $L_*(\mathbf{u})$ is always a convex function of $\mathbf{u}$. The solution of the dual problem is given by $\nabla_{\mathbf{p}} L = 0$ and $\nabla_{\mathbf{u}} L = 0$, which gives the following two systems, each with m equations:

$$\begin{cases} \partial L/\partial p_1 = 0 \\ \vdots \\ \partial L/\partial p_m = 0 \end{cases} \qquad \begin{cases} \partial L/\partial u_1 = 0 \\ \vdots \\ \partial L/\partial u_m = 0 \end{cases}$$

Taking partial derivations from Eq. (6.4) with respect to p_i yields the following system:

$$\begin{cases} h_{11}u_1 - (2^{r_{min}/B} - 1)\sum_{j\neq 1}^{m} h_{1j}u_j = 1 \\ \vdots \\ h_{mm}u_m - (2^{r_{min}/B} - 1)\sum_{j\neq m}^{m} h_{mj}u_j = 1 \end{cases} \tag{6.7}$$

Writing the linear system of equations given by Eq. (6.7) in the matrix form, we get:

$$\mathbf{H_d} \cdot \mathbf{u} - (2^{r_{min}/B} - 1)\mathbf{H_{d0}} \cdot \mathbf{u} = \mathbf{1} \tag{6.8}$$

where $\mathbf{1} = (1, ..., 1)^{\mathrm{T}}$; $\mathbf{H_d} = [h_{d_{ij}}]$ is a diagonal matrix with diagonal elements of the channel matrix, and $\mathbf{H_{d0}} = [h_{d0_{ij}}]$ is the channel matrix having the diagonal elements set to zero as follows:

$$h_{d_{ij}} = \begin{cases} h_{ij} & \text{if } i = j \\ 0 & else \end{cases} \qquad h_{d0_{ij}} = \begin{cases} 0 & \text{if } i = j \\ h_{ij} & else \end{cases}$$

This gives us the optimal Lagrange multipliers $\mathbf{u}^*$ as:

$$\mathbf{u}^* = \left[\mathbf{H_d} - (2^{r_{min}/B} - 1)\mathbf{H_{d0}}\right]^{-1} \cdot \mathbf{1} \tag{6.9}$$

Similarly, taking partial derivations from Eq. (6.4) with respect to u_i gives:

$$\begin{cases} h_{11}p_1 - (2^{r_{min}/B} - 1)\sum_{j\neq 1}^{m} h_{j1}p_j = (2^{r_{min}/B} - 1)n_1 \\ \vdots \\ h_{mm}p_m - (2^{r_{min}/B} - 1)\sum_{j\neq m}^{m} h_{jm}p_j = (2^{r_{min}/B} - 1)n_m \end{cases} \tag{6.10}$$

Writing this in the matrix form, we get:

$$\mathbf{H_d} \cdot \mathbf{p} - (2^{r_{min}/B} - 1)\mathbf{H_{d0}}' \cdot \mathbf{p} = (2^{r_{min}/B} - 1)\mathbf{n} \tag{6.11}$$

where $\mathbf{n} = (n_1, \ldots, n_m)^{\mathrm{T}}$ is the noise vector. Finally, we can get the optimal solution of the dual problem, $\mathbf{p}^* = (p_i^*)_{\forall i \in M}^{\mathrm{T}}$, as:

$$\mathbf{p}^* = \left[\mathbf{H_d} - (2^{r_{min}/B} - 1)\mathbf{H_{d0}}'\right]^{-1} \cdot (2^{r_{min}/B} - 1)\mathbf{n} \tag{6.12}$$

■

Duality Gap

Let x^* and y^* denote the solutions of the Lagrange primal and dual problems respectively for a given optimization problem. According to the weak duality theorem [113], the solution of the dual problem provides a lower bound on the solution of the primal problem, i.e. $x^* \geq y^*$. The difference between the solutions $x^* - y^*$ is called duality gap. The case $x^* = y^*$ is known as strong duality and happens only if the optimization problem meets some criteria called sufficient conditions for the strong duality. If the strong duality holds for an optimization problem, there is no duality gap between the primal and dual solutions which means that once the solution of the dual problem is found, provided that it is finite, the solution of the primal problem is attained.

Proposition 4.2: *The duality gap in the optimization problem defined by (6.3) is zero.*

Proof: According to the Slater theorem [113], if there exists a feasible solution for a convex optimization problem which normally satisfies affine inequality constraints and strictly satisfies nonlinear inequality constraints, then the strong duality holds. The optimization problem at hand is convex and the solution of the dual problem is given by Eq. (6.12). Assuming that $\mathbf{p}^*$ exists and is finite, it is a feasible solution for the optimization problem, i.e. satisfies the inequality constrains $\mathbf{g}(\mathbf{p})$. Since the inequality constraints are all affine, the strong duality is concluded.

■

As a result, the solution given by Eq. (6.9) is also the global optimum solution of the primal problem and hence $\mathbf{p}^*$ is the optimum power allocation for WBANs in the system which minimizes the total power consumption while satisfying the QoS constraints across the whole system.

Having calculated $\mathbf{p}^*$, we need to apply the regional constraint as the power allocation vector should lie in $\mathcal{P}$. The final power allocation would therefore be element-wise as:

$$p_i^* \leftarrow \min\left(P_{\max_i}, \max\left(0, p_i^*\right)\right) \tag{6.13}$$

6.2.2 Distributed Solution

The optimization problem at hand is inherently a centralized problem in the sense that all the WBANs aim to minimize the total power consumption in the system as opposed to minimizing individual power consumption. The solution given by Eq. (6.12) is also a centralized approach because firstly the optimum power allocation for all WBANs in the system is to be calculated in one place as the vector $\mathbf{p}^*$, and secondly the global information of the channel matrix and the noise vector is needed for this calculation.

However, multiple WBANs usually operate with no central arbiter existing and operating between different WBANs on which the power control algorithm can run. Therefore, the centralized solution proposed is not directly applicable to WBANs —it is very useful as a benchmark though, and the problem needs to be addressed distributively.

In the first effort to solve the problem distributively, let us decompose and decouple the dual problem. From the Lagrangian function in Eq. (6.4) we have

$$L(\mathbf{p},\mathbf{u}) = \sum_{i=1}^{m}\left[p_i(h_{ii}u_i - 1) - (2^{r_{min}/B} - 1)u_i n_{0_i}\right] - (2^{r_{min}/B} - 1)\sum_{i=1}^{m} u_i \sum_{j\neq i}^{m} h_{ji}p_j \tag{6.14}$$

It can be shown that the following equation holds:

$$\sum_{i=1}^{m} u_i \sum_{j\neq i}^{m} h_{ji}p_j = \sum_{i=1}^{m} p_i \sum_{j\neq i}^{m} h_{ij}u_j \tag{6.15}$$

Substituting (6.15) into (6.14) and factoring out p_i, we get:

$$L(\mathbf{p},\mathbf{u}) = \sum_{i=1}^{m} f_i(p_i,\mathbf{u}) \tag{6.16}$$

where $f_i(p_i,\mathbf{u})$ is given by:

$$f_i(p_i,\mathbf{u}) = p_i\left(h_{ii}u_i - 1 - (2^{r_{min}/B} - 1)\sum_{j\neq i}^{m} h_{ij}u_j\right) - (2^{r_{min}/B} - 1)u_i n_i \tag{6.17}$$

Provided that the optimal Lagrange multiplier vector $\mathbf{u}^*$ (the maximizer of the Lagrangian function) given by (6.9) is known by each WBAN, we can write:

$$L(\mathbf{p},\mathbf{u})\big|_{\mathbf{u}=\mathbf{u}^*} = \sum_{i=1}^{m} f_i(p_i,\mathbf{u}^*) = L^*(\mathbf{p}) \tag{6.18}$$

Considering the primal problem in (6.5), we have:

$$\begin{aligned}\min_{\mathbf{p}\in\mathcal{P}} \sup_{\mathbf{u}\in\mathcal{U}} \{L(\mathbf{p},\mathbf{u})\} & \\ &= \min_{\mathbf{p}\in\mathcal{P}} L^*(\mathbf{p}) \\ &= \min_{\mathbf{p}\in\mathcal{P}} \sum_{i=1}^{m} f_i(p_i,\mathbf{u}^*) \\ &\geq \sum_{i=1}^{m} \min_{p_i} f_i(p_i,\mathbf{u}^*) \end{aligned} \tag{6.19}$$

Thinking of $f_i(p_i,\mathbf{u}^*)$ as the individual objective function of WBAN i, this inequality implies that if each WBAN independently minimizes its own objective function, the reached solution bounds above the primal solution $\mathbf{p}^*$.

The approach proposed provides a partially-distributed solution to the problem because in order to calculate $f_i(p_i,\mathbf{u}^*)$, the problem in Eq. (6.9) should be solved centralizedly[1] before each iteration. Although WBANs can employ stochastic approximation methods to estimate $\mathbf{u}^*$, this leads to a more complicated solution. Moreover the approach proposed requires some negotiation taking place between WBANs in order to share partial information of channel gains which makes it less useful in practice. Each WBAN i needs power gain, h_{ii}, and also power gains of the interference it imposes on other WBANs, namely h_{ij}, to calculate $f_i(p_i,\mathbf{u}^*)$.

In the following, we present a fully distributed solution to the problem which does not need any centralized computation and can be executed distributively and asynchronously by independent WBANs. The concept underlying this approach is fixed-point calculations. From Eq. (6.11), the optimum power allocation for the system is obtained by solving the following equation:

$$\mathbf{A}.\mathbf{p} = \mathbf{b} \tag{6.20}$$

where

$$\mathbf{A} = \mathbf{H_d} - (2^{r_{min}/B} - 1)\mathbf{H_{d0}}' \tag{6.21}$$

$$\mathbf{b} = (2^{r_{min}/B} - 1)\mathbf{n} \tag{6.22}$$

Eq. (6.20) can be reformulated as a fixed-point $\mathbf{p} = F(\mathbf{p})$ by writing matrix $\mathbf{A}$ as $\mathbf{A} = \mathbf{A} + \mathbf{B} - \mathbf{B}$ with $\mathbf{B}$ an arbitrary non-singular matrix and multiplying the equation by $\mathbf{B}^{-1}$ as follows:

$$\mathbf{p} = \mathbf{C}\mathbf{p} + \mathbf{c} =: F(\mathbf{p}) \tag{6.23}$$

[1]The centralized problem which adjusts the Lagrangian multipliers is known as the master problem in decomposition theory [114].

where $\mathbf{C} = \mathbf{B}^{-1} \cdot (\mathbf{B} - \mathbf{A})$ is called the iteration matrix, $\mathbf{c} = \mathbf{B}^{-1} \cdot \mathbf{b}$ and $\mathbf{p}$ is a fixed point of the mapping $F : \mathbb{R}^m \to \mathbb{R}^m$. To attain this fixed point, one can use the following iteration, starting from an arbitrary point $\mathbf{p}_0$:

$$\mathbf{p}^{(k+1)} = F(\mathbf{p}^{(k)}), k = 0, 1, 2, 3, ... \tag{6.24}$$

According to Banach's theorem [115], if F is a contraction mapping, i.e. there is a constant $0 < q < 1$ such that $\|F(\mathbf{x}) - F(\mathbf{y})\| \leqslant q \|\mathbf{x} - \mathbf{y}\| \ , \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^m$, then this sequence converges to the fixed point. With the problem at hand where $F(\mathbf{p}) = \mathbf{C}\mathbf{p} + \mathbf{c}$, this would be determined by the iteration matrix $\mathbf{C}$. Whether the sequence converges or not to the fixed point depends only on the selection of $\mathbf{C}$.

One selection for the iteration matrix is obtained by the Jacobi method. By decomposing the matrix $\mathbf{A}$ to its lower-left sub-diagonal part $\mathbf{L}$, its diagonal part $\mathbf{D}$ and its upper-right sup-diagonal part $\mathbf{R}$, i.e. $\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{R}$, and then picking its diagonal part $\mathbf{D}$ for the matrix $\mathbf{B}$, we have the iteration matrix as:

$$\mathbf{C} = \mathbf{I} - \mathbf{B}^{-1}\mathbf{A} = \mathbf{I} - \mathbf{D}^{-1}(\mathbf{L} + \mathbf{D} + \mathbf{R}) = -\mathbf{D}^{-1}(\mathbf{L} + \mathbf{R})$$

with entries

$$c_{ij} = \begin{cases} -a_{ij}/a_{ii} & , \text{if } i \neq j \\ 0 & otherwise \end{cases} \tag{6.25}$$

Therefore, we can write the iteration as:

$$\mathbf{p}^{(k+1)} = \mathbf{C}\mathbf{p}^{(k)} + \mathbf{c} = -\mathbf{D}^{-1}(\mathbf{L} + \mathbf{R})\mathbf{p}^{(k)} + \mathbf{D}^{-1}\mathbf{b} \tag{6.26}$$

Writing component-wise, we get:

$$p_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j \neq i}^{m} a_{ij} p_j^{(k)} \right) \tag{6.27}$$

An important property of this iteration is that the elements of the vector $\mathbf{p}$ can be calculated independently and asynchronously. By inspecting Eq. (6.27), it is noticed that the computation of $p_i^{(k+1)}$ is independent of any other $p_j^{(k+1)}$. This is useful for parallel and distributed computations. Being regarded as the best response, it also fits well in the context of non-cooperative games where players play their best response independently and asynchronously to reach a Nash equilibrium.

By substituting the values of a_{ii}, a_{ij} and b_i from Eq. (6.21) and (6.22) into Eq. (6.27), we get the iteration for our system as:

$$p_i^{(k+1)} = \frac{2^{r_{min}/B} - 1}{h_{ii}} \left(n_i + \sum_{j \neq i}^{m} h_{ji} p_j^{(k)} \right) \tag{6.28}$$

We need to apply the regional constraint as the power allocation vector should lie in $\mathcal{P}$:

$$p_i^{(k+1)} \leftarrow \min\left(P_{\max_i}, \max\left(0, p_i^{(k+1)}\right)\right) \tag{6.29}$$

It is worth noting here that the regional constraint is applied to the power allocation at each iteration, whilst with the centralized approach, it should be applied only once and that to the final solution given by Eq. (6.12). This may cause the two approaches to end up with different solutions whenever a feasible power allocation does not exist in $\mathcal{P}$. Moreover, a WBAN that hits its maximum power bound at some iteration can not meet its QoS requirement and will get a rate lower than its target rate, even though it is transmitting at the maximum power. In this case, it can be shown that the power of those WBANs in the system that satisfy the rate constraint will converge to a feasible solution, whereas the other WBANs that cannot achieve their target rate will continue to do their best by transmitting at the maximum power.

Proposition 4.3: *The iteration in Eq. (6.28) converges for every starting point if the following condition holds*

$$\sum_{j\neq i}^{m} \left|2^{r_{min}/B} - 1\right| h_{ji} < h_{ii}\ , \forall i \in M \tag{6.30}$$

Proof:

$$\begin{aligned} &\|F(\mathbf{p_1}) - F(\mathbf{p_2})\| \\ &\qquad = \|C\mathbf{p_1} + c - (C\mathbf{p_2} + c)\| \\ &\qquad = \|C(\mathbf{p_1} - \mathbf{p_2})\| \\ &\qquad \leqslant \|C\|\,\|\mathbf{p_1} - \mathbf{p_2}\| \end{aligned}$$

For F to be a contraction mapping, we should have $\|C\| < 1$. Using the maximum absolute row sum norm to calculate the norm of the iteration matrix $\mathbf{C}$, we have:

$$\begin{aligned} &\|\mathbf{C}\|_\infty \\ &\qquad = \max_{0\leq i\leq m} \sum_{j=1}^{m} |c_{ij}| \\ &\qquad = \max_{0\leq i\leq m} \sum_{j=1}^{m} \frac{|a_{ij}|}{|a_{ii}|} \\ &\qquad = \max_{0\leq i\leq m} \sum_{j=1}^{m} \frac{\left|2^{r_{min}/B} - 1\right| h_{ji}}{h_{ii}} < 1 \end{aligned}$$

hence, we have the convergence.

■

It is highly desirable that WBANs coordinate their power levels without any message exchanges. However, for the iteration in Eq. (6.28) to be calculated at each WBAN i, the interference gains imposed by other WBANs, namely h_{ji}, as well as the power levels those WBANs have chosen currently, namely $p_j^{(k)}$, are needed which requires some messages to be exchanged between WBANs. In the following, we show that it is possible to use the iteration without the need of any message exchange between WBANs which is highly favorable in practice. First we define a parameter called *the sensitivity of SINR to power* as follows.

Definition 4.1: The sensitivity of SINR to power, denoted by η, is the rate of change in SINR at the receiver with respect to transmission power at the transmitter. For WBAN i in our system, it is given by

$$\eta_i = \frac{\partial \xi_i}{\partial p_i} = \frac{h_{ii}}{n_i + \sum_{j \neq i}^{m} h_{ji} p_j} \tag{6.31}$$

where ξ_i is the SINR at BNC in WBAN i.

The sensitivity of SINR in an average sense expresses the change in the SINR that the receiver experiences as a result of a change of one Watt in the transmission power at the transmitter. It can be thought of as a measure for the density of interfering WBANs in the system. For a fixed power level, η_i depends on the number of interfering WBANs and their arrangement around WBAN i because the interference gains are proportional to an order $n \geq 2$ of the inverted distance as $h_{ji} = f(d_{ji}^{-n})$. In a dense system with many interfering WBANs around, the sensitivity of SINR is low, hence a WBAN in order to increase its channel capacity has to increase its transmission power much more than that of a high sensitivity of SINR condition, where the interference is low.

Rewriting the iteration in Eq. (6.28) using the sensitivity of SINR yields:

$$p_i^{(k+1)} = \frac{2^{r_{min}/B} - 1}{\eta_i^{(k)}} \tag{6.32}$$

In each WBAN i, the BNC node can measure the SINR ξ_i and use it to approximate the sensitivity of SINR as follows:

$$\eta_i^{(k)} \approx \frac{\Delta^{(k)} \xi_i}{\Delta^{(k)} p_i} = \frac{\xi_i^{(k)} - \xi_i^{(k-1)}}{p_i^{(k)} - p_i^{(k-1)}} \tag{6.33}$$

This way, the distributed solution is attained using only local information which can be simply calculated at each BNC and there will be no need for any message exchange between WBANs.

Table 6.1: Simulation Parameters and Values

Parameter Name	Symbol	Parameter Value
Bandwidth	B	300 kHz
Noise	n_i	-174 dBm/Hz
Minimum Power	P_{min_i}	0
Maximum Power	$P_{\max_i}$	25 μW ($\approx$ -16 dBm)

6.3 Performance Evaluation

Table 6.1 shows the parameters and their values used in this simulations.

We assess the performance of the proposed centralized and distributed solutions using extensive simulations. The simulation environment is as described earlier in the previous section. We consider two cases for performance evaluation, one by changing the density of WBANs in the system and the other one by varying the target data rate required by WBANs to maintain their QoS. In the first case, the target data rate, r_{min}, is set to 100 kbps and the number of WBANs in the system is increased from 4 to 32, while in the second case, the number of WBANs is fixed at 16 and r_{min} is increased from 25 kbps to 200 kbps.

Figure 6.1 shows the average transmission power of sensor nodes in μW versus the number of WBANs in the system. As it can be seen, the power level has to rise when the number of WBANs increases to keep the minimum data rate constraint satisfied. Although in low density conditions, the difference between power levels in the centralized and distributed solutions is negligible, it goes up when the number of WBANs in the system increases. With only 4 WBANs in the system, sensor nodes transmit at just below 2 μW, while in a dense condition with 32 WBANs, this figure goes up to almost 3.2 μW in the centralized approach and to 5.8 μW in the distributed approach.

Figure 6.2 shows the average interference power in dBm versus the number of WBANs in the system. As it can be seen, the two approaches perform almost similarly under low density conditions, while the centralized solution outperforms the distributed solution under dense conditions. The interference power rises from just above -125 dBm and reaches around -122 dBm and -119 dBm for the centralized and distributed approaches respectively when the number of WBANs in the system increases from 4 to 32 nodes.

Since all the computations in our model take place at BNC nodes and not in sensor nodes, the major source of energy consumption in the sensor nodes will be dominantly related to transmissions. Figure 6.3 shows the average energy consumption at sensor nodes for transmission of one bit in nJoul/bit as a function of the number of WBANs in the system. We clearly notice that the energy needed to transmit one bit increases when the density of WBANs in the system goes up. It is due to the fact that the price

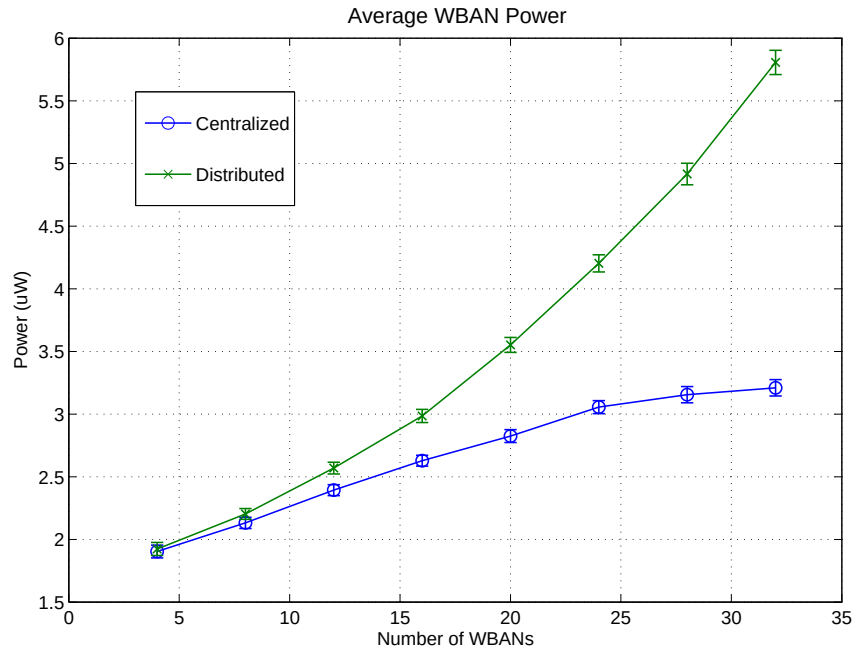

FIGURE 6.1: Average transmission power versus the number of WBANs with r_{min}= 100 kbps

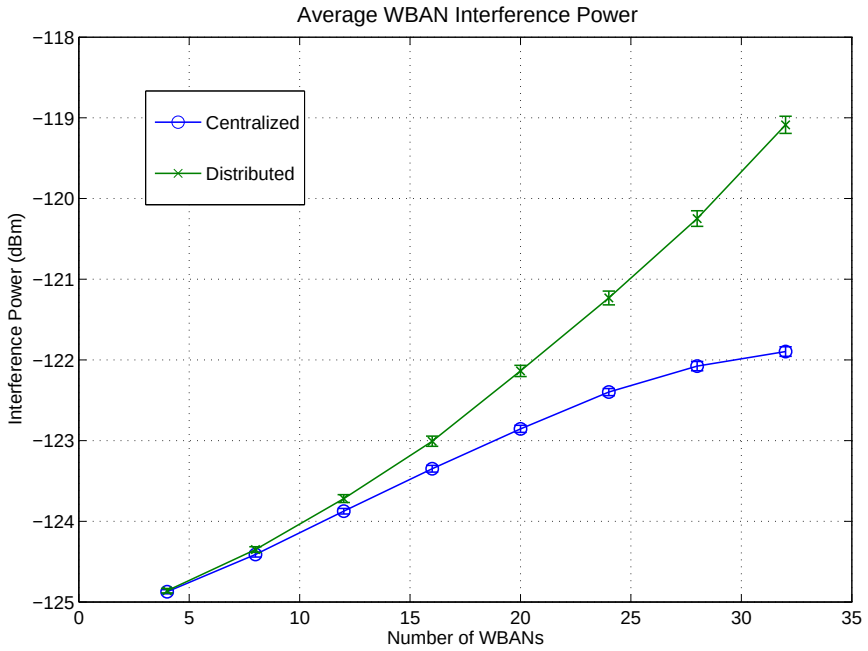

FIGURE 6.2: Average interference power versus the number of WBANs with r_{min}= 100 kbps

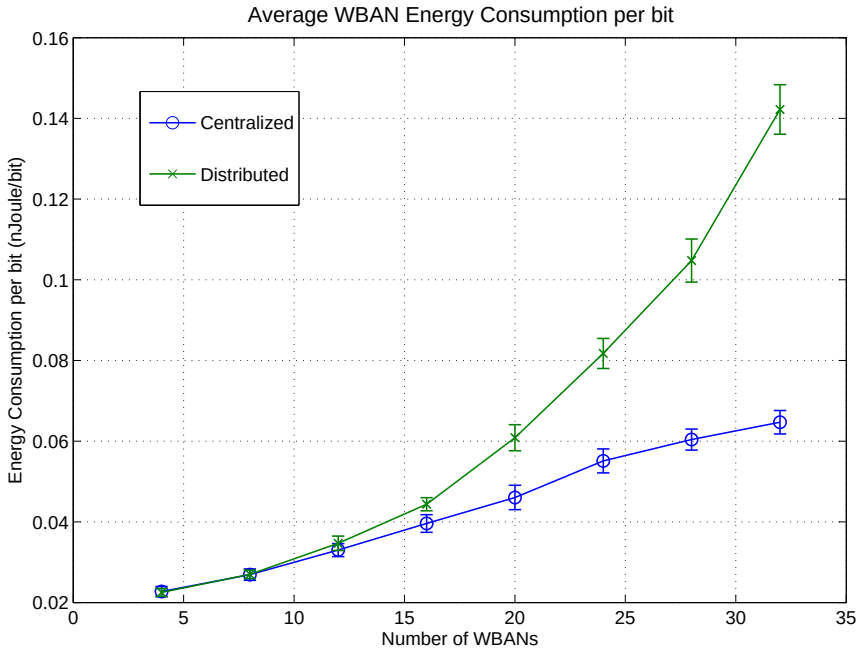

Figure 6.3: Average energy consumption per bit versus the number of WBANs with r_{min}= 100 kbps

to pay for keeping target rates increases when interference increases. With 4 WBANs in the system, the centralized and distributed solutions perform almost identically and consume just above 0.02 nJoul/bit, but as the network density goes up, the distributed solution does worse and consumes more energy than the centralized approach. With 32 WBANs in the system, the distributed solution consumes 0.14 nJoul per bit which is almost double that of the centralized solution.

Since WBANs in the distributed approach determine the solution in an iterative way, the convergence to a stable power allocation is of importance. The average number of iterations needed by the distributed approach to reach the solution at r_{min} = 100 kbps with respect to the number of WBANs in the system is depicted in Figure 6.4 . As it can be seen, the number of iterations increases almost linearly with the number of WBANs in the system. Each iteration takes almost 260 nS to run on a 2.67 GHz i5 Intel CPU. The convergence time, hence, under the most dense condition, i.e. 32 WBANs in the room, would be around 5.5 μS which is quite satisfactory in practice.

Finally, Figure 6.5 shows the average lifetime of the sensor nodes in months versus the number of WBANs in the system. As it can be noticed, sensors' lifetime decreases as the system density increases for both the approaches. The graph illustrates that the sensors with the centralized solution can live longer than those with the distributed solution for any number of WBANs in the system. The superiority of the centralized solution over the distributed version, however, decreases with the number of WBANs

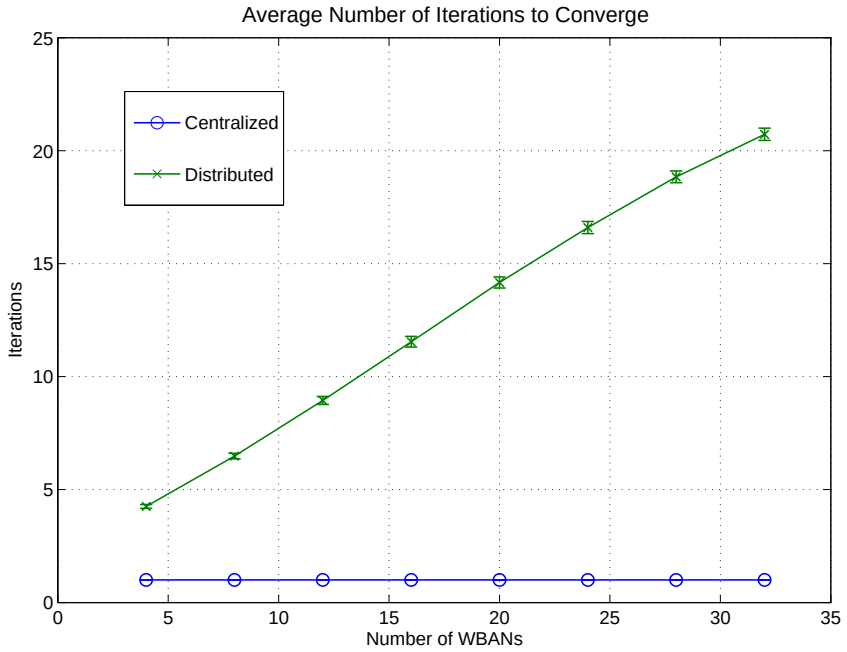

FIGURE 6.4: Average number of iterations versus the number of WBANs with r_{min}= 100 kbps

in the system such that with 32 WBANs in the room, the centralized solution allows WBANs to live double that of the distributed solution whereas with 4 WBANs, we see almost similar performances from both approaches allowing sensors to survive for almost 3.7 months respectively.

Now we evaluate the performance of the centralized and distributed approaches with respect to the target data rate, r_{min}. We keep the number of WBANs fixed at 16 nodes and increase r_{min} from 25 kbps to 200 kbps.

Figure 6.6 shows the average transmission power of sensor nodes versus the target data rate. As it can be seen, more power is needed to maintain a higher required data rate for both the centralized and distributed approaches. Transmission power in the centralized solution goes up with the target data rate and reaches from almost 0.5 μW at r_{min} = 25 kbps to 4.2 μW at r_{min} = 200 kbps. In the distributed solution, however, the power level markedly goes up from 0.1 μW at r_{min} = 25 kbps to roughly 7.8 μW at r_{min} = 200 kbps. Although the power consumption of the distributed solution is almost the same as the centralized solution at r_{min} =25 kbps, it is almost double that of the centralized solution at $r_{min} = 200kbps$.

Figure 6.7 shows the average interference power in dBm versus the target data rate in the system with 16 WBANs. As it can be seen, the interference power which WBANs experience increases as the QoS constraints tighten more. Although both approaches show similar performance at low target data rate, the centralized solution outperforms

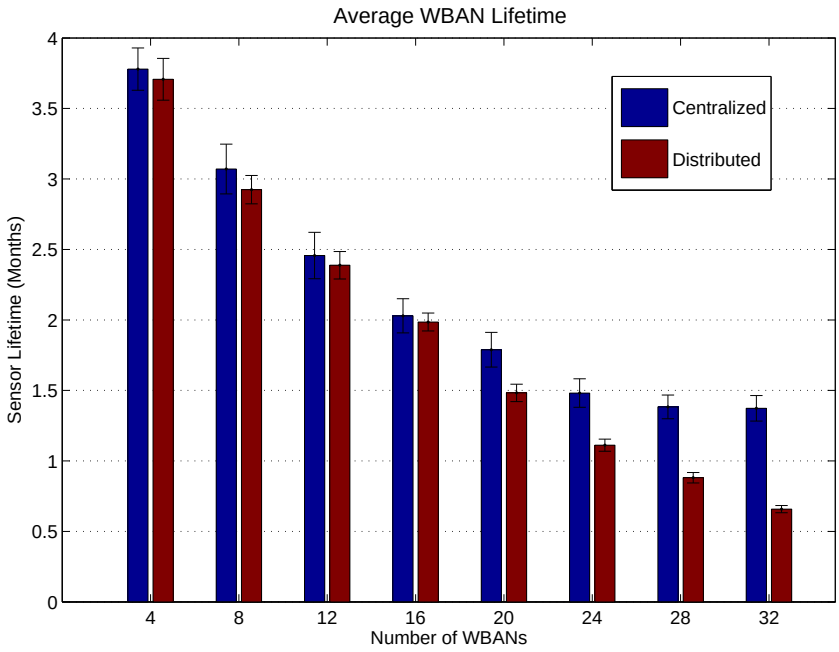

Figure 6.5: Average lifetime of sensor nodes versus the minimum required data rate constraint with 16 WBANs

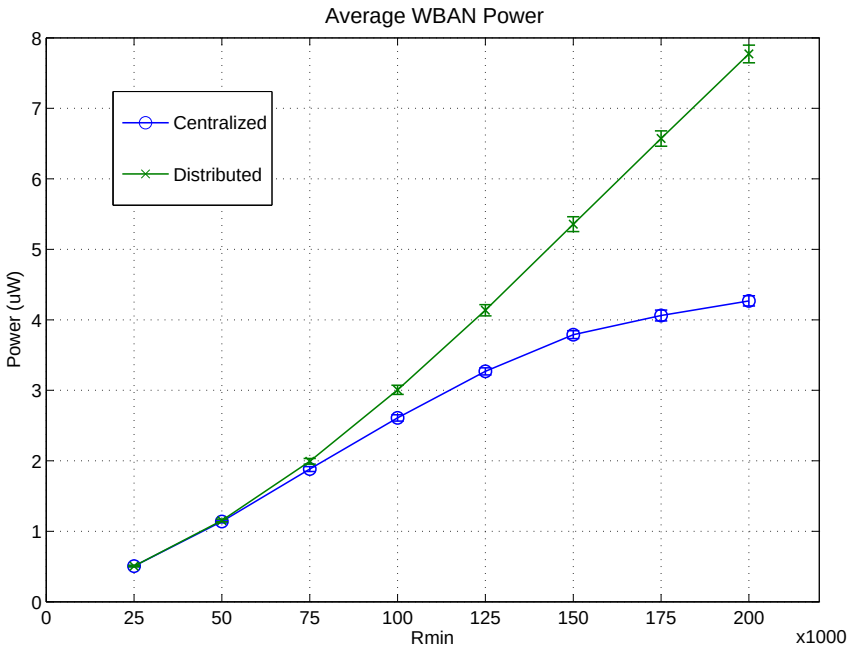

Figure 6.6: Average transmission power versus the minimum required data rate constraint with 16 WBANs

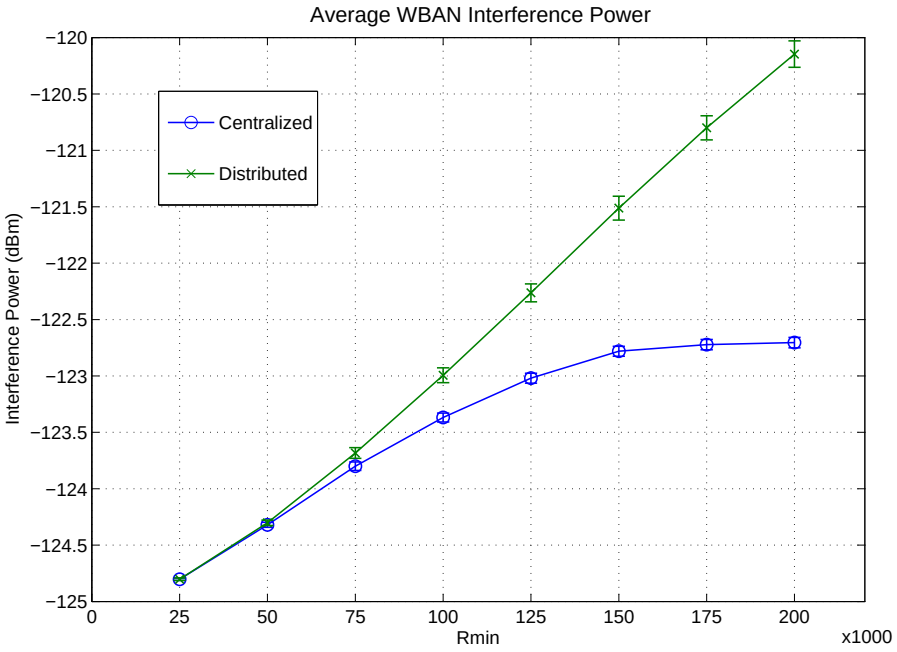

FIGURE 6.7: Average interference power versus the number of WBANs with 16 WBANs

the distributed solution at high data rate constraints. The interference power increase from just above -125 dBm at r_{min} =25 kbps and reaches almost -122.7 dBm and -120.2 dBm for the centralized and distributed solutions respectively at r_{min} = 200 kbps.

Figure 6.8 shows the average throughput of WBANs versus the target data rate both in kbps. Ideally, the graph should be a line with slope one for any r_{min}, implying providing an average throughput equal to the minimum required data rate. As it can be seen, both approaches are able to provide such QoS guarantee up to almost r_{min} =75 kbps. For greater data rates, however, a feasible solution in $\mathcal{P}$ may not exist and we see that the average throughput starts to drop slightly in both approaches, although it is more remarkable in the centralized solution. For r_{min} = 200 kbps, the distributed solution provides around 150 kbps averagely while the centralized solution delivers roughly 185 kbps.

The average energy consumed to transmit one bit is shown as a function of the target data rate in Figure 6.9. As it can be seen, more energy is needed as the minimum required data rate increases in both approaches. While they both consume almost 0.02 nJoul per bit at r_{min} = 25 kbps, the distributed solution needs almost 0.12 nJoul to transmit one bit at r_{min} = 200 kbps which is three times that of the centralized solution. In other words, WBANs in the distributed approach tend to transmit at a higher price which possibly leads to increased throughput to maintain the required QoS, whereas with the centralized solution, they tend to be more frugal and save power at the expense of decreased throughput.

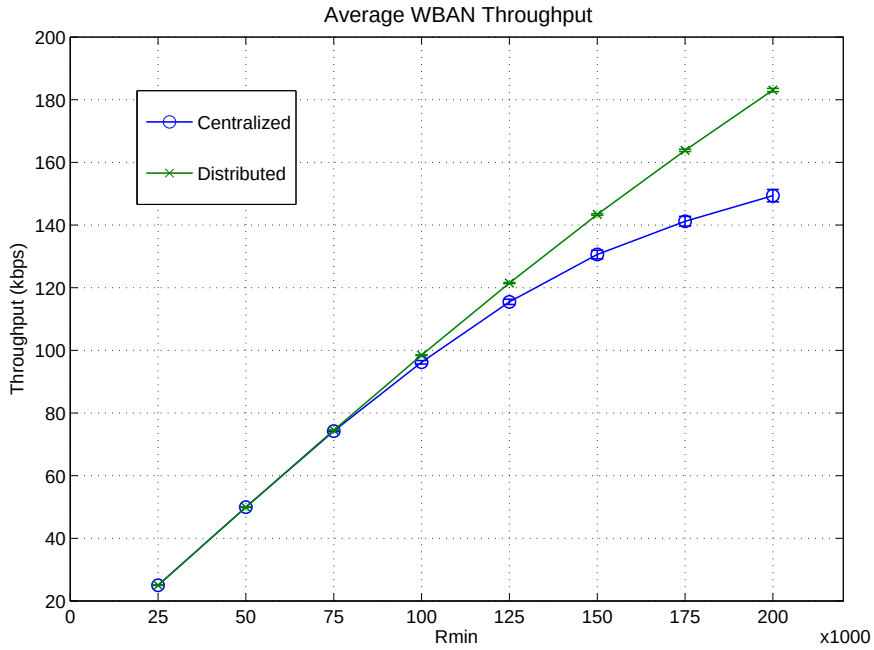

Figure 6.8: Average throughput versus the target rate with 16 WBANs

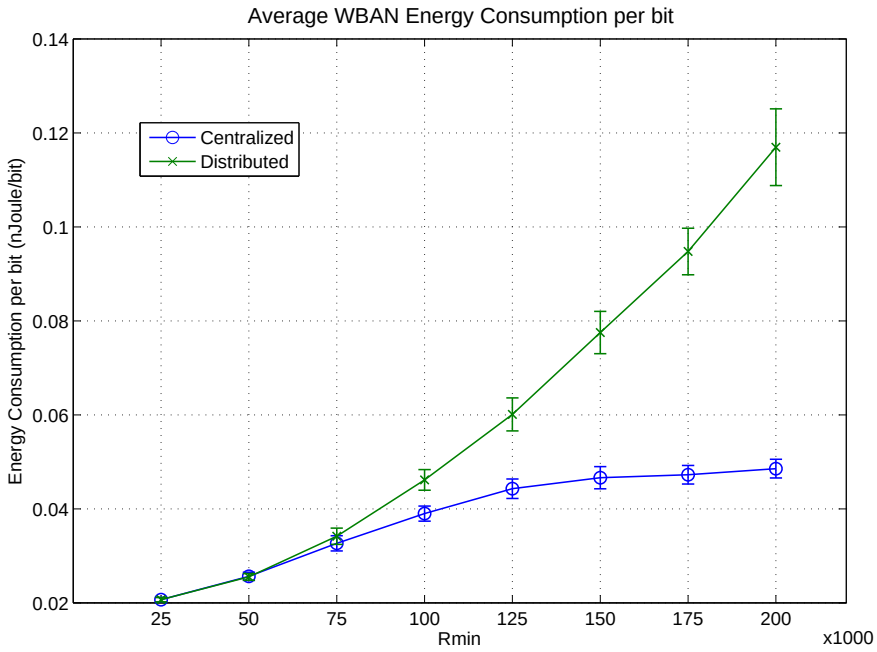

Figure 6.9: Average energy consumption per bit versus the target rate with 16 WBANs

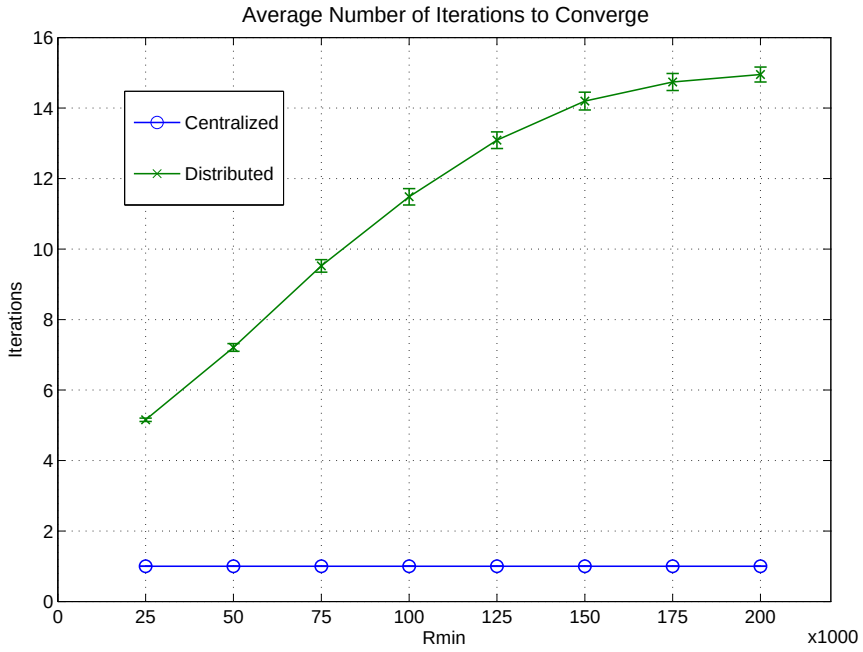

FIGURE 6.10: Average number of iterations versus the target rate with 16 WBANs

Figure 6.10 shows the average number of iterations needed by the distributed approach to reach a stable power allocation versus the target data rate in the system with 16 WBANs. As it can be seen, the number of iterations increases with the target data rate, i.e., a tighter QoS constraint, a slower convergence. Considering that each iteration takes almost 470 nS time of a 2.67 GHz i5 Intel CPU, the convergence time varies from almost 2.4 μS to 7.1 μS for a data rate range between 25 kbps and 200 kbps.

Finally figure 6.11 shows the average lifetime of the sensor nodes in months versus the target data rate. As it can be seen, the centralized solution slightly outperforms the distributed solution. The graph also illustrates that sensors' lifetime drops markedly as the QoS constraint becomes tighter. While the sensor nodes can run almost 16 months on their batteries at $r_{min} = 25$ kbps, they survive only around 1 months and half a month with the centralized and distributed approaches respectively at $r_{min} = 200$ kbps.

6.4 Conclusions

We formulated the power control problem as an optimization problem which minimized the total power consumption in the system while satisfying individual target (data) rates of WBANs. Having attained the centralized optimal solution using the

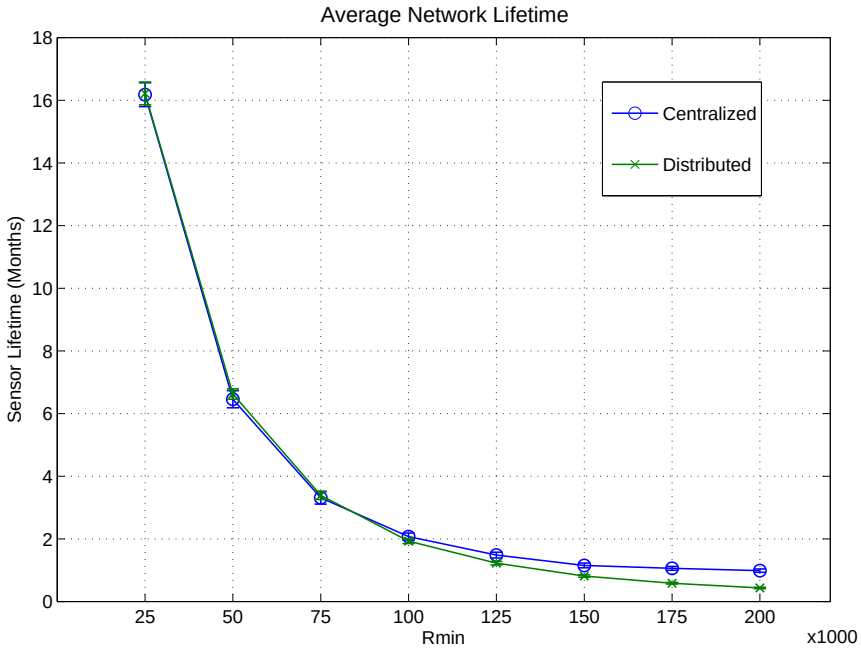

Figure 6.11: Average lifetime of sensor nodes versus the target rate with 16 WBANs

Lagrange multipliers, we employed fixed-point calculations and proposed a distributed solution based on Jacobi method which enabled WBANs to asynchronously find a solution, without the need of any central arbiter inter-operating between WBANs. Besides, the distributed approach did not require any cooperation or message exchange between WBANs to attain the solution which is highly favorable in medical applications. We evaluated the performance of the centralized and distributed approaches using extensive simulations by considering two cases. In the first case, we kept the target data rate fixed and increased the number of WBANs in the system, while in the second case we kept the number of WBANs fixed and increased the target data rate. This way we scrutinized the performance of the proposed approaches in terms of throughput, transmission and interference power levels, energy consumption per bit, network lifetime and convergence. The simulation results indicated that both the centralized and distributed approaches were able to manage the interference between WBANs when node density or target data rate increased in the system. Results also revealed that although the centralized solution allowed WBANs to live longer, it was outperformed by the distributed solution in terms of throughput under dense conditions, which leads to a worse QoS provisioning. The difference between the performance of the two solutions which mostly emerges under dense conditions is due to the fact that the regional constraint on power levels, which bounds them to $P_{\min_i}$ and $P_{\max_i}$, is a non-linear operator and is applied differently to the centralized and distributed solutions. While it is applied once to the final solution of the centralized approach, it is applied at each iteration to the solution in the distributed approach. This causes the two approach to

end up with different solutions whenever a feasible solution does not exist in the region of interest.

"The knowledge of yourself will preserve you from vanity."
Miguel de Cervantes

7 Conclusions and Future Work

7.1 Conclusions

The lifetime and reliability of WBANs can be enhanced by using transmission power control, which mitigates inter-network interference between neighboring WBANs operating in the same frequency band.

To this end, we considered two problems, where in the first one, the goal was to make a tradeoff between power and data rate for the sake of applications in which energy conservation is more important than QoS, and in the second problem, the aim was to maintain target data rates at any cost, for the sake of QoS-sensitive applications.

For the tradeoff scenario, we proposed three novel power controllers based on genetic-fuzzy control, game theory and reinforcement learning called WFPC, WPCG and WRLPC respectively.

We designed a genetic algorithm and a learning mechanism to optimize WFPC and compared its performance to a well-cited power controller in the literature, called ADP. Simulation results show that WFPC strongly outperforms ADP and improves both power consumption by almost 40%-50% and convergence by 60%-70% for different number of WBANs in the system, while sacrificing only 4% of throughput. Also, the average energy consumption per bit improves by 27%-45% for different number of WBANs in the system. This superiority basically originates from two factors which are the ability of genetic algorithms to find the best solution and the ability of fuzzy controllers to cope with complicated non-linear systems. However, the genetic algorithm optimization required by WFPC decreases its flexibility to adjust the tradeoff

TABLE 7.1: Comparing Power Controllers

	Solution Optimality	Convergence	Dynamic Adaptability	Message Exchange	Off-line Optimization
ADP	$\star$	$\star$	$\star\star$	YES	NO
WFPC	$\star\star$	$\star\star$	$\star$	NO	YES
WPCG	$\star$	$\star\star\star$	$\star\star$	NO	NO
WRLPC	$\star\star\star$	$\star$	$\star\star\star$	NO	NO

between throughput and power adaptively and accommodate dynamic changes of the surrounding environment because it is performed offline at design stage.

For the game theory-based power controller, WPCG, we considered a broader family of pricing functions and proved the existence of a Nash equilibrium in the game. We proposed the dynamic rules of the game based on the best response and suggested an adaptive pricing mechanism to dynamically adjust the tradeoff between throughput and power based on the channel gains and WBANs' power budgets. This allows WBANs to benefit from good channel conditions and high energy budgets, leading to increased throughput, and on the other hand preventing WBANs from increasing their transmission power levels at bad channel conditions or low power budgets, thereby extending their battery lifetimes as well as decreasing interference to other WBANs. We also proposed another adaptive pricing method which did not rely on channel gains for calculations and required only SINR which was simply available at digital receivers at low cost, and it was showed by simulations to perform almost similarly to the first adaptive pricing scheme which is based on channel gains. We also compared the performance of WPCG to WFPC and ADP. The simulation results indicate that WPCG strongly outperforms both WFPC and ADP in convergence while it is overwhelmed by WFPC in terms of energy consumption per bit by 25%. WPCG consumes almost 6 μW more power than WFPC to produce 100 kbps more throughput.

The proposed RL-based power controller, WRLPC, is a lightweight power controller based which improves its performance by learning from experience. We showed through simulation that the proposed RL-based power controller provided a better tradeoff between throughput and power leading to 3 μW less power consumption for sacrificing 2%-5% of throughput compared to WFPC, and saving 6 μW of power for sacrificing 15%-30% of throughput compared to WPCG. Moreover, WRLPC was also able to improve energy consumption per bit by 40%-60% compared to WPCG, and by 25%-50% compared to WFPC. However, it was outperformed by WFLPC and WPCG in terms of convergence.

Table 7.1 compares the power control approaches, where the measure of the solution optimality is considered to be the average energy consumption per bit.

Figure 7.1 demonstrates the most representative performance indicator, being the average energy consumption per bit in nJoul/bit, versus the number of WBANs in the system for all the power controllers. As it can be clearly seen, WRLPC consumes the

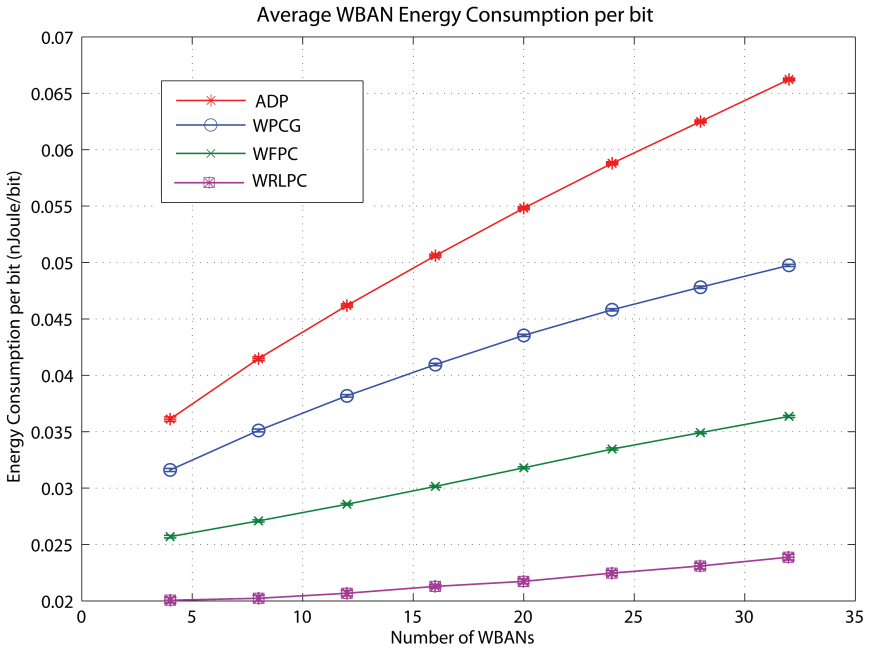

Figure 7.1: The average energy consumption per bit versus the number of WBANs

least energy per bit and strongly surpasses all other power controllers. WFPC ranks second and is followed by WPCG and ADP as the third and forth places respectively. ADP is outperformed by all the proposed power controllers.

We also investigated the impact of reinforcement learning key factors including the reward function, discount factor, learning rate and eligibility trace parameter on the performance of the system in terms of convergence and optimality. It was showed that increasing the learning rate could improve both convergence and energy consumption per bit. While increasing the eligibility trace parameter does not affect energy consumption per bit, it improves convergence, and finally, increasing the discount factor aggravates both energy consumption per bit and convergence.

Moreover, we evaluated and compared the performance of the different RL algorithms including Q-learning, sarsa and OLSPI to that of a counterpart non-cooperative game. Although RL-based approaches were outperformed by the counterpart game in terms of convergence, they were able to save more energy. More importantly, WBANs in the RL-based approaches do not need to be aware of a pre-designed Nah equilibrium as opposed to the game. This increases their adaptability to the dynamic changes of the environment.

For the non-tradeoff scenario, we modeled the problem as a convex optimization problem and utilized Lagrangian multipliers method to obtain the optimum solution. We then employed fixed-point calculations and proposed a distributed solution based

on Jacobi method which enabled WBANs to asynchronously find a solution. We evaluated the performance of the centralized and distributed approaches using extensive simulations and considered two cases where in the first case, we kept the target data rate fixed and increased the number of WBANs in the system, while in the second case we kept the number of WBANs fixed and increased the target data rate. This way we scrutinized the performance of the proposed approaches in terms of throughput, transmission and interference power levels, energy consumption per bit, network lifetime and convergence. The simulation results indicated that both the centralized and distributed approaches were able to manage the interference between WBANs when node density or target data rate increased in the system. Results also revealed that although the centralized solution allowed WBANs to live longer, it was outperformed by the distributed solution in terms of throughput under dense conditions, which leads to a better QoS provisioning. The difference between the performance of the two solutions which mostly emerges under dense conditions is due to the fact that the regional constraint on power levels boundaries is a non-linear operator and is applied differently to the centralized and distributed solutions. While it is applied once to the final solution of the centralized approach, it is applied at each iteration to the solution in the distributed approach. This causes the two approach to end up with different solutions whenever a feasible solution does not exist in the region of interest.

All the proposed power controllers in this thesis rely only on physically measurable local information, such as the SINR and interference power, and they do not need any negotiation or cooperation between WBANs to find a solution. Additionally, they are fully distributed and asynchronous.

7.2 Future Work and Open Problems

There exist a number of future work and open problems to be solved to potentially improve the performance of power control algorithms in WBANs. In the following, we list some of them.

- The proposed fuzzy power controller showed good performance in both convergence and the optimality of solution, while suffering from the lack of high adaptability, i.e. offline tuning. On the other hand, the RL-based power controller suffered from slow convergence, while enjoying a high level of adaptability and solution optimality. Combining these two techniques may potentially result in greater performance for the power control in WBANs. This can be basically achieved in two ways which are reinforcement learning power controllers being tuned by fuzzy logic and fuzzy power controllers being tuned by reinforcement learning.

  When fuzzy logic is employed to tune a RL-based power controller, it acts as a function approximator for the state-action space of reinforcement learning. Fuzzy controllers can gracefully approximate the Q functions and enable RL agents to

better generalize the state-action space and make better decisions. In other words, the power controller itself is based on reinforcement learning and it is only tuned by the fuzzy controller (as in [116], [117] and [118]).

On the other hand, reinforcement learning can also be employed to tune a fuzzy power controller. With the aid of reinforcement learning, a fuzzy power controller will be able to tune its knowledge-base parameters over time in an online manner by learning from experience (see for example [119] and [120]). This makes the fuzzy controller highly dynamic and adaptive to the changing environment.

Although the first approach, i.e. using fuzzy approximators in reinforcement learning, namely a *fuzzy RL-based power controller*, has been investigated by some researchers in the literature, the second approach, i.e. tuning fuzzy parameters using reinforcement learning, namely a *RL-based fuzzy power controller*, still deserves more attention and can be studied more extensively in the future.

It will be even more interesting if fuzzy RL-based power controllers and RL-based fuzzy power controllers incorporate POMDP[1] to model the error that may happen when RL agents observe the current state. In this study, however, we disregarded such errors and considered error-free state observations. Taking such errors into account will also remain for future work. We believe even better performance in practice can be achieved in WBANs by using specifically a POMDP RL-based fuzzy power controller.

- In this study, we considered that BNC nodes are responsible for running power control algorithms. However, if the power controller is embedded in BN nodes, the power consumption related to computations may not be negligible compared to the transmission power and should be considered in calculating the power efficiency of the power controller.

- In this study, we considered only single-action learner agents for the reinforcement learning approach. Although considering joint-action learner agents, which are aware of the actions of each other and maximize their reward over the joint actions, leads to more computation overhead, system complexity and the need for negotiation between WBANs, it may improve the quality of the solution and also convergence. The tradeoff between such complexity and the resulting improvements would be interesting to investigate.

- In this study we considered non-cooperative agents. However, in applications allowing cooperation between WBANs, considering cooperating agents may lead to a better solution. In the game theory approach, this can be done by using cooperative games (state-less) or stochastic games (state-full), and in the reinforcement learning approach, by employing Multi-Agent Reinforcement Learning (MARL), which can also help to solve stochastic games. However, the overhead imposed by such cooperation between agents in terms of computation, message exchange and power consumption should be modeled and investigated.

[1] Partially Observed Markov Decision Process

- In this thesis, we only considered the ϵ-greedy policy for the exploration-exploitation tradeoff. The performance of learning can be improved by using the Boltzmann policy. However, the application of the Boltzmann policy is limited to discrete space-action spaces. Developing such policy for continuous state-action spaces can be a possible future work.

List of Acronyms

ADP	Asynchronous Distributed Pricing Power Controller
AP	Access Point
BAN	Body Area Network
BN	BAN Node
BNC	BAN Node Controller
CAP	Contention Access Phase
CCA	Clear Channel Assessment
CTS	Clear To Send
CRN	Cognitive Radio Network
EAP	Exclusive Access Phase
ECG	Electrocardiography
EEG	Electroencephalography
EPG	Exact Potential Game
GA	Genetic Algorithm
GDP	Gross Domestic Product
HAI	Hospital-Acquired Infection
HBC	Human Body Communications
ISM	Industrial, Scientific and Medical
ITU	International Telecommunication Union
LFSR	Linear Feedback Shift Register
LQI	Link Quality Indicator

MANET	Mobile Ad-hoc NETwork
MARL	Multi Agent Reinforcement Learning
MICS	Medical Implant Communication Service
NE	Nash Equilibrium
ODE	Ordinary Differential Equation
OLSPI	Online Least-Squares Policy Iteration
OPG	Ordinal Potential Game
POMDP	Partially Observable Markov Decision Process
PPDU	Physical Protocol Data Unit
PU	Primary User
QoS	Quality of Service
RAP	Random Access Phase
RFID	Radio Frequency Integrated Circuits
RL	Reinforcement Learning
RSSI	Received Signal Strength Indicator
RTS	Ready To Send
SINR	Signal-to-Interference-and-Noise-Ratio
SU	Secondary User
TDMA	Time Division Multiple Access
UWB	Ultra Wide Band
WBAN	Wireless Body Area Network
WFPC	WBAN Fuzzy Power Controller
WLAN	Wireless Local Area Network
WMTS	Wireless Medical Telemetry Service
WPCG	WBAN Power Control Game
WRLPC	WBAN Reinforcement Learning-based Power Controller
WSN	Wireless Sensor Network

List of Symbols

c_i	Throughput of WBAN i
$c_i^{(t)}$	Throughput of WBAN i at time t
p_i	Transmission power of WBAN i
$p_i^{(t)}$	Transmission power of WBAN i at time t
B	Bandwidth
ξ_i	SINR of WBAN i
η_i	Sensitivity of SINR to the transmission power of WBAN i
a_t	RL action at time t
s_t	RL state at time t
α	Learning rate
γ	Discount factor
λ	Eligibility trace parameter
$Q(s, a)$	Action-value function at state s and action a
$V(s)$	Value function at state s
r_{min}	Target data rate (minimum required data rate)

References

[1] "IEEE standard for local and metropolitan area networks - part 15.6: Wireless body area networks," *IEEE Std 802.15.6-2012*, pp. 1–271, 2012. xv, 17, 18, 19, 20, 21, 22, 23

[2] K. Y. Yazdandoost, K. Sayrafian-Pour, *et al.*, "Channel model for body area network (ban)," *IEEE P802*, vol. 15, 2009. xvi, 49, 51

[3] S. P. Keehan, G. A. Cuckler, A. M. Sisko, A. J. Madison, S. D. Smith, J. M. Lizonitz, J. A. Poisal, and C. J. Wolfe, "National health expenditure projections: modest annual growth until coverage expands and economic growth accelerates," *Health Affairs*, vol. 31, no. 7, pp. 1600–1612, 2012. 1

[4] B. Braem, B. Latre, I. Moerman, C. Blondia, E. Reusens, W. Joseph, L. Martens, and P. Demeester, "The need for cooperation and relaying in short-range high path loss sensor networks," in *Sensor Technologies and Applications, 2007. SensorComm 2007. International Conference on*, pp. 566–571, IEEE, 2007. 2

[5] A. Pollack, "Rising threat of infections unfazed by antibiotics," *New York Times*, vol. 26, 2010. 3

[6] M. Renaud, K. Karakaya, T. Sterken, P. Fiorini, C. Van Hoof, and R. Puers, "Fabrication, modelling and characterization of mems piezoelectric vibration harvesters," *SENSORS AND ACTUATORS A-PHYSICAL*, vol. 145, pp. 380–386, 2008. 4

[7] V. Leonov, P. Fiorini, S. Sedky, T. Torfs, and C. Van Hoof, "Thermoelectric mems generators as a power supply for a body area network," in *Solid-State Sensors, Actuators and Microsystems, 2005. Digest of Technical Papers. TRANSDUCERS '05. The 13th International Conference on*, vol. 1, pp. 291–294 Vol. 1, 2005. 4

[8] A. Sampath, P. Sarath Kumar, and J. M. Holtzman, "Power control and resource management for a multimedia cdma wireless system," in *Personal, Indoor and Mobile Radio Communications, 1995. PIMRC'95.'Wireless: Merging onto the Information Superhighway'., Sixth IEEE International Symposium on*, vol. 1, pp. 21–25, IEEE, 1995. 7, 101

[9] D. P. Bertsekas and J. N. Tsitsiklis, "Parallel and distributed computation," 1989. 8

[10] D. P. Bertsekas and J. N. Tsitsiklis, "Some aspects of parallel and distributed iterative algorithmsa survey," *Automatica*, vol. 27, no. 1, pp. 3–21, 1991. 8

[11] L. Busoniu, R. Babuska, B. De Schutter, and D. Ernst, *Reinforcement learning and dynamic programming using function approximators*, vol. 39. CRC Press, 2010. 10, 82

[12] W. G. Scanlon, B. Burns, and N. E. Evans, "Radiowave propagation from a tissue-implanted source at 418 mhz and 916.5 mhz," *Biomedical Engineering, IEEE Transactions on*, vol. 47, no. 4, pp. 527–534, 2000. 14, 15

[13] H. S. Savci, A. Sula, Z. Wang, N. S. Dogan, and E. Arvas, "Mics transceivers: regulatory standards and applications [medical implant communications service]," in *SoutheastCon, 2005. Proceedings. IEEE*, pp. 179–182, IEEE, 2005. 15

[14] K. S. Kwak, S. Ullah, and N. Ullah, "An overview of IEEE 802.15. 6 standard," in *Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010 3rd International Symposium on*, pp. 1–6, IEEE, 2010. 16

[15] T. Kasami, "Weight distribution formula for some class of cyclic codes," 1966. 17

[16] S. Agarwal, R. Katz, S. Krishnamurthy, and S. Dao, "Distributed power control in ad-hoc wireless networks," in *Personal, Indoor and Mobile Radio Communications, 2001 12th IEEE International Symposium on*, vol. 2, pp. 59–66, 2001. 24

[17] J. Gomez, A. Campbell, M. Naghshineh, and C. Bisdikian, "Conserving transmission power in wireless ad hoc networks," in *Network Protocols, 2001. Ninth International Conference on*, pp. 24–34, 2001. 24

[18] P. Karn, "Maca-a new channel access method for packet radio," in *ARRL/CRRL Amateur radio 9th computer networking conference*, vol. 140, pp. 134–140, 1990. 24

[19] M. Pursley, H. Russell, and J. Wysocarski, "Energy-efficient transmission and routing protocols for wireless multiple-hop networks and spread-spectrum radios," in *EUROCOMM 2000. Information Systems for Enhanced Public Safety and Security. IEEE/AFCEA*, pp. 1–5, 2000. 24

[20] J.-P. Ebert, B. Stremmel, E. Wiederhold, and A. Wolisz, "An energy-efficient power control approach for wlans," *Journal of Communications and Networks*, vol. 2, no. 3, pp. 197–206, 2000. 24

[21] V. G. Douros, P. Frangoudis, K. Katsaros, and G. Polyzos, "Power control in wlans for optimization of social fairness," in *Informatics, 2008. PCI '08. Panhellenic Conference on*, pp. 239–243, 2008. 24

[22] R. Zhu and J. Wang, "Power-efficient spatial reusable channel assignment scheme in wlan mesh networks," *Mobile Networks and Applications*, vol. 17, no. 1, pp. 53–63, 2012. 24

[23] J. Chen, S. Olafsson, Y. Yang, and X. Gu, "Joint distributed transmit power control and dynamic channel allocation for scalable wlans," in *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, pp. 1–6, 2009. 24

[24] C. Yang, M. Sheng, J. Li, H. Li, and J. Li, "Energy-aware joint power and rate control in overlay cognitive radio networks: A nash bargaining perspective," in *Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference on*, pp. 520–524, 2012. 24

[25] A. Ibrahim, L. Muscariello, and J. Roberts, "A single channel signalling mechanism for power/rate control in wlans," in *Next Generation Internet Networks, 2009. NGI '09*, pp. 1–8, 2009. 24

[26] S. Oh, M. Gruteser, D. Jiang, and Q. Chen, "Joint power control and scheduling algorithm for wi-fi ad-hoc networks," in *Wireless Internet Conference (WICON), 2010 The 5th Annual ICST*, pp. 1–9, 2010. 24

[27] D. R. CS Rao, KCK Reddy, "Power control technique for efficient call admission control in advanced wirless networks," *International Journal on Computer Science and Engineering*, vol. 4, no. 6, 2012. 24

[28] H.-T. Lim, Y. Kim, S. Pack, and C.-H. Kang, "Joint uplink/downlink connection admission control in wlan/cellular integrated systems," in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2011 Fifth International Conference on*, pp. 470–474, 2011. 24

[29] V. G. Douros and G. C. Polyzos, "Review of some fundamental approaches for power control in wireless networks," *Computer Communications*, vol. 34, no. 13, pp. 1580–1592, 2011. 24

[30] F. Ingelrest, D. Simplot-Ryl, and I. Stojmenovic, "Optimal transmission radius for energy efficient broadcasting protocols in ad hoc and sensor networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 17, no. 6, pp. 536–547, 2006. 24

[31] M. Cardei, J. Wu, and S. Yang, "Topology control in ad hoc wireless networks with hitch-hiking," 2004. 24

[32] M. Kubisch, H. Karl, A. Wolisz, L. Zhong, and J. Rabaey, "Distributed algorithms for transmission power control in wireless sensor networks," in *Wireless Communications and Networking, WCNC 2003. IEEE*, vol. 1, pp. 558–563 vol.1, 2003. 25

[33] T. A. ElBatt, S. V. Krishnamurthy, D. Connors, and S. Dao, "Power management for throughput enhancement in wireless ad-hoc networks," in *Communications, ICC 2000. IEEE International Conference on*, vol. 3, pp. 1506–1513, IEEE, 2000. 25

[34] N. Hao and S.-J. Yoo, "Ancpc: Adaptive neighbor coordinated interference avoidance power control for cognitive radio ad hoc networks," in *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*, pp. 1–6, 2010. 25

[35] S. lin Wu, Y.-C. Tseng, and J. ping Sheu, "Intelligent medium access for mobile ad hoc networks with busy tones and power control," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1647–1657, 2000. 25

[36] Q. Hu and Z. Tang, "An adaptive transmit power scheme for wireless sensor networks," in *Ubi-media Computing (U-Media), 2010 3rd IEEE International Conference on*, pp. 12–16, 2010. 25

[37] T. L. Lim and G. Mohan, "Energy aware geographical routing and topology control to improve network lifetime in wireless sensor networks," in *Broadband Networks, 2005. BroadNets 2005. 2nd International Conference on*, pp. 771–773 Vol. 2, 2005. 25

[38] G. Xing, C. Lu, X. Jia, and R. Pless, "Localized and configurable topology control in lossy wireless sensor networks," *Ad Hoc Networks*, vol. 11, no. 4, pp. 1345 – 1358, 2013. 25

[39] J. Huang, R. Berry, and M. Honig, "Distributed interference compensation for wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 5, pp. 1074–1084, 2006. 25, 37

[40] R. Shah and J. Rabaey, "Energy aware routing for low energy ad hoc sensor networks," in *Wireless Communications and Networking Conference, 2002. WCNC 2002. IEEE*, vol. 1, pp. 350–355, 2002. 26

[41] S. M. Senouci and G. Pujolle, "Energy efficient routing in wireless ad hoc networks," in *Communications, 2004 IEEE International Conference on*, vol. 7, pp. 4057–4061, 2004. 26

[42] B. Braem, B. Latré, C. Blondia, I. Moerman, and P. Demeester, "Analyzing and improving reliability in multi-hop body sensor networks," *International Journal On Advances in Internet Technology*, vol. 2, no. 1, pp. 151–161, 2009. 26

[43] S.-H. Seo, S. Gopalan, S.-M. Chun, K.-J. Seok, J.-W. Nah, and J.-T. Park, "An energy-efficient configuration management for multi-hop wireless body area networks," in *Broadband Network and Multimedia Technology (IC-BNMT), 2010 3rd IEEE International Conference on*, pp. 1235–1239, IEEE, 2010. 26

[44] J. Dong and D. Smith, "Cooperative body-area-communications: Enhancing coexistence without coordination between networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, pp. 2269–2274, 2012. 26

[45] T. O. Olwal, B. J. Van Wyk, N. Ntlatlapa, K. Djouani, P. Siarry, and Y. Hamam, "Dynamic power control for wireless backbone mesh networks: a survey," *Network protocols and algorithms*, vol. 2, no. 1, pp. 1–44, 2010. 27

[46] N. A. Pantazis and D. D. Vergados, "A survey on power control issues in wireless sensor networks," *Communications Surveys & Tutorials, IEEE*, vol. 9, no. 4, pp. 86–107, 2007. 27

[47] L. H. Correia, D. F. Macedo, A. L. dos Santos, A. A. Loureiro, and J. M. S. Nogueira, "Transmission power control techniques for wireless sensor networks," *Computer Networks*, vol. 51, no. 17, pp. 4765–4779, 2007. 27

[48] H. Saghaei and A. Neyestanak, "Variable step closed-loop power control in cellular wireless cdma systems under multipath fading," in *Communications, Computers and Signal Processing, 2007. PacRim 2007. IEEE Pacific Rim Conference on*, pp. 157–160, 2007. 27

[49] M. Rintamaki, H. Koivo, and I. Hartimo, "Adaptive closed-loop power control algorithms for cdma cellular communication systems," *Vehicular Technology, IEEE Transactions on*, vol. 53, no. 6, pp. 1756–1768, 2004. 27

[50] M. Quwaider, J. Rao, and S. Biswas, "Transmission power assignment with postural position inference for on-body wireless communication links," *ACM Trans. Embed. Comput. Syst.*, vol. 10, pp. 14:1–14:27, Aug. 2010. 27

[51] M. Rintamaki, H. Koivo, and I. Hartimo, "Adaptive closed-loop power control algorithms for cdma cellular communication systems," *Vehicular Technology, IEEE Transactions on*, vol. 53, no. 6, pp. 1756–1768, 2004. 27

[52] C.-Y. Yang and B.-S. Chen, "Robust power control of cdma cellular radio systems with time-varying delays," *Signal Processing*, vol. 90, no. 1, pp. 363–372, 2010. 27

[53] M. Almgren, B. Engstrm, and M. Ericson, "Power control in a cdma mobile communication system," Oct. 19 2004. US Patent 6,807,164. 27

[54] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *Vehicular Technology, IEEE Transactions on*, vol. 42, no. 4, pp. 641–646, 1993. 27, 101

[55] R. D. Yates, "A framework for uplink power control in cellular radio systems," *Selected Areas in Communications, IEEE Journal on*, vol. 13, no. 7, pp. 1341–1347, 1995. 27

[56] A. Zappavigna, T. Charalambous, and F. Knorn, "Unconditional stability of the foschini-miljanic algorithm," *Automatica*, vol. 48, no. 1, pp. 219–224, 2012. 27

[57] I. Lestas, "Power control in wireless networks: Stability and delay independence for a general class of distributed algorithms," *Automatic Control, IEEE Transactions on*, vol. 57, no. 5, pp. 1253–1258, 2012. 27

[58] D. Mitra, "An asynchronous distributed algorithm for power control in cellular radio systems," in *Wireless and Mobile Communications*, pp. 177–186, Springer, 1994. 27

[59] N. Bambos and S. Kandukuri, "Power controlled multiple access (pcma) in wireless communication networks," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, pp. 386–395, IEEE, 2000. 27

[60] N. Bambos and S. Kandukuri, "Power-controlled matiple access schemes for next-generation wireless packet networks," *Wireless Communications, IEEE*, vol. 9, no. 3, pp. 58–, 2002. 27

[61] M. Chiang, P. Hande, T. Lan, and C. W. Tan, "Power control in wireless cellular networks," *Foundations and Trends® in Networking*, vol. 2, no. 4, pp. 381–533, 2008. 27

[62] S. Koskie and Z. Gajic, "Signal-to-interference-based power control for wireless networks: a survey, 1992-2005," *Dynamics of Continuous Discrete and Impulsive Systems Series B*, vol. 13, no. 2, p. 187, 2006. 27

[63] S. Xiao, V. Sivaraman, and A. Burdett, "Adapting radio transmit power in wireless body area sensor networks," in *Proceedings of the ICST 3rd international conference on Body area networks*, p. 14, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008. 28

[64] M. Quwaider, A. Muhammad, J. Choi, and S. Biswas, "Posture-predictive power control in body sensor networks using linear-quadratic gaussian control," in *Networks and Communications, 2009. NETCOM '09. First International Conference on*, pp. 52–59, 2009. 28

[65] D. Smith, L. Hanlen, and D. Miniutti, "Transmit power control for wireless body area networks using novel channel prediction," in *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, pp. 684–688, 2012. 28

[66] B. Moulton, L. Hanlen, J. Chen, G. Croucher, L. Mahendran, and A. Varis, "Body-area-network transmission power control using variable adaptive feedback periodicity," in *Communications Theory Workshop (AusCTW), 2010 Australian*, pp. 139–144, 2010. 29

[67] S. Ramakrishnan and T. Thyagarajan, "Energy efficient medium access control for wireless sensor networks," *IJCSNS International Journal of Computer Science and Network Security*, vol. 9, no. 6, pp. 273–279, 2009. 29

[68] J. Zhang, J. Chen, and Y. Sun, "Transmission power adjustment of wireless sensor networks using fuzzy control algorithm," *Wireless Communications and Mobile Computing*, vol. 9, no. 6, pp. 805–818, 2009. 30

[69] X. Xia and Q. Liang, "Packets transmission in wireless sensor networks: Interference, energy and delay-aware approach," in *Wireless Communications and Networking Conference, WCNC 2007. IEEE*, pp. 2501–2505, IEEE, 2007. 30

[70] A. Lakshmi, S. Manisekaran, and D. Venkatesan, "Fuzzified dynamic power control algorithm for wireless sensor networks," *International Journal of Science and Technology (IIJEST)*, vol. 3, 2011. 30

[71] T. Jiang, P. Wu, B. Shen, and K. Kwak, "A novel fuzzy algorithm for power control of wireless sensor nodes," in *Communications and Information Technology, 2009. ISCIT, 9th International Symposium on*, pp. 64–68, 2009. 30

[72] J. Anno, L. Barolli, A. Durresi, F. Xhafa, and A. Koyama, "A cluster head decision system for sensor networks using fuzzy logic and number of neighbor nodes," in *Ubi-Media Computing, 2008 First IEEE International Conference on*, pp. 50–56, 2008. 30

[73] I. Gupta, D. Riordan, and S. Sampalli, "Cluster-head election using fuzzy logic for wireless sensor networks," in *Communication Networks and Services Research Conference, 2005. Proceedings of the 3rd Annual*, pp. 255–260, IEEE, 2005. 30

[74] T. Srinivasan, R. Chandrasekar, and V. Vijaykumar, "A fuzzy, energy-efficient scheme for data centric multipath routing in wireless sensor networks," in *Wireless and Optical Communications Networks, 2006 IFIP International Conference on*, p. 5, 2006. 30

[75] Q. Bing, J. Lu, and W. Lili, "An improved multicast routing protocol based on fuzzy clustering," in *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on*, pp. 1–4, 2008. 30

[76] W. Mustafa, J. S. Yu, E. Rakus-Andersson, A. Mohammed, and W. J. Kulesza, "Fuzzy-based opportunistic power control strategy in cognitive radio networks," in *Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010 3rd International Symposium on*, pp. 1–5, IEEE, 2010. 30

[77] H.-S. T. Le, H. D. Ly, and Q. Liang, "Opportunistic spectrum access using fuzzy logic for cognitive radio networks," *International Journal of Wireless Information Networks*, vol. 18, no. 3, pp. 171–178, 2011. 30

[78] H.-S. Le and H. Ly, "Opportunistic spectrum access using fuzzy logic for cognitive radio networks," in *Communications and Electronics, 2008. ICCE 2008. Second International Conference on*, pp. 240–245, 2008. 30

[79] S. Koskie and J. Zapf, "Acceleration of static nash power control algorithm using newton iterations," *Dynamics of Continuous, Discrete and Impulse Systems B: Applications and Algorithms*, vol. 12, pp. 685–690, 2005. 31

[80] F. Meshkati, M. Chiang, H. Poor, and S. Schwartz, "A game-theoretic approach to energy-efficient power control in multicarrier cdma systems," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 6, pp. 1115–1129, 2006. 32, 33

[81] C. Liang and K. R. Dandekar, "Power management in mimo ad hoc networks: a game-theoretic approach," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 4, pp. 1164–1170, 2007. 32

[82] R. W. Thomas, R. S. Komali, A. B. MacKenzie, and L. A. DaSilva, "Joint power and channel minimization in topology control: A cognitive network approach," in *Communications, 2007. ICC'07. IEEE International Conference on*, pp. 6538–6543, IEEE, 2007. 33

[83] P. Closas, A. Pages-Zamora, and J. Fernandez-Rubio, "A game theoretical algorithm for joint power and topology control in distributed wsn," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 2765–2768, 2009. 34

[84] J. Huang, Z. Han, M. Chiang, and H. V. Poor, "Auction-based resource allocation for cooperative communications," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 7, pp. 1226–1237, 2008. 34

[85] T. Alpcan, X. Fan, T. Basar, M. Arcak, and J. Wen, "Power control for multicell cdma wireless networks: a team optimization approach," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2005. WIOPT 2005. Third International Symposium on*, pp. 379–388, 2005. 34

[86] Y. Xing and R. Chandramouli, "Stochastic learning solution for distributed discrete power control game in wireless data networks," *Networking, IEEE/ACM Transactions on*, vol. 16, no. 4, pp. 932–944, 2008. 34

[87] Y. Wang and X.-Y. Li, "Minimum power assignment in wireless ad hoc networks with spanner property," *Journal of combinatorial optimization*, vol. 11, no. 1, pp. 99–112, 2006. 34

[88] E. Altman, T. Boulogne, R. El-Azouzi, T. Jiménez, and L. Wynter, "A survey on networking games in telecommunications," *Computers & Operations Research*, vol. 33, no. 2, pp. 286–311, 2006. 35

[89] D. E. Charilas and A. D. Panagopoulos, "A survey on game theory applications in wireless networks," *Computer Networks*, vol. 54, no. 18, pp. 3421 – 3430, 2010. 35

[90] S. Sengupta, M. Chatterjee, and K. A. Kwiat, "A game theoretic framework for power control in wireless sensor networks," *Computers, IEEE Transactions on*, vol. 59, no. 2, pp. 231–242, 2010. 35

[91] C. Pandana and K. R. Liu, "Near-optimal reinforcement learning framework for energy-aware sensor communications," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 4, pp. 788–797, 2005. 35

[92] A. Galindo-Serrano and L. Giupponi, "Distributed q-learning for aggregated interference control in cognitive radio networks," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 4, pp. 1823–1834, 2010. 36

[93] J. Li and C. Yang, "A markovian game-theoretical power control approach in cognitive radio networks: A multi-agent learning perspective," in *Wireless Communications and Signal Processing (WCSP), 2010 International Conference on*, pp. 1–5, IEEE, 2010. 36

[94] N. Bradai, S. Belhaj, L. Chaari, and L. Kamoun, "Study of medium access mechanisms under ieee 802.15.6 standard," in *Wireless and Mobile Networking Conference (WMNC), 2011 4th Joint IFIP*, pp. 1–6, 2011. 38

[95] K. M. Passino and S. Yurkovich, *Fuzzy control*. Addison-Wesley, 1998. 41

[96] G. J. Klir and B. Yuan, *Fuzzy sets and fuzzy logic*. Prentice Hall New Jersey, 1995. 41

[97] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965. 41

[98] R.-E. Precup and H. Hellendoorn, "A survey on industrial applications of fuzzy control," *Computers in Industry*, vol. 62, no. 3, pp. 213–226, 2011. 42

[99] O. Cordon, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, "Ten years of genetic fuzzy systems: current framework and new trends," *Fuzzy sets and systems*, vol. 141, no. 1, pp. 5–31, 2004. 42

[100] O. Cordón, *Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases*, vol. 19. World Scientific Publishing Company, 2001. 42

[101] J. Schwarz, "Fuzzy genetic algorithm - a brief survey," in *Proceedings of the Colloquium Advanced Simulation of Systems*, pp. 353–358, 2000. 42

[102] J. Von Neumann and O. Morgenstern, "Theory of games and economic behavior," *Bull. Amer. Math. Soc*, vol. 51, pp. 498–504, 1945. 58

[103] J. Von Neumann and O. Morgenstern, *Theory of games and economic behavior (commemorative edition)*. Princeton University Press, 2007. 58

[104] J. T. Drew Funderburg, *Game Theory*. The MIT Press, Cambridge, MA, 1991. 58, 59, 60

[105] T. Basar, G. J. Olsder, G. Clsder, T. Basar, T. Baser, and G. J. Olsder, *Dynamic noncooperative game theory*, vol. 200. SIAM, 1995. 58

[106] R. B. Myerson, *Game theory: analysis of conflict*. Harvard University Press, 1997. 58

[107] R. Gibbons, *Game theory for applied economists*. Princeton University Press, 1992. 58

[108] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, vol. 1. Cambridge Univ Press, 1998. 75

[109] S. Singh, T. Jaakkola, M. L. Littman, and C. Szepesvári, "Convergence results for single-step on-policy reinforcement-learning algorithms," *Machine Learning*, vol. 38, no. 3, pp. 287–308, 2000. 78

[110] S. Niekum, L. Spector, and A. Barto, "Evolution of reward functions for reinforcement learning," in *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation*, pp. 177–178, ACM, 2011. 82

[111] I. Szita and A. Lőrincz, "Optimistic initialization and greediness lead to polynomial time learning in factored mdps," in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 1001–1008, ACM, 2009. 86

[112] P. V. Yee and S. S. Haykin, *Regularized radial basis function networks: Theory and applications*. John Wiley, 2001. 86

[113] J. M. Borwein and A. S. Lewis, *Convex analysis and nonlinear optimization: theory and examples*, vol. 3. Springer, 2006. 105

[114] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 8, pp. 1439–1451, 2006. 107

[115] C. Bessaga, "On the converse of banach," in *Colloquium Mathematicae*, vol. 7, pp. 41–43, Institute of Mathematics Polish Academy of Sciences, 1959. 108

[116] H. Shah and M. Gopal, "Fuzzy decision tree function approximation in reinforcement learning," *International Journal of Artificial Intelligence and Soft Computing*, vol. 2, no. 1, pp. 26–45, 2010. 125

[117] L. Buşoniu, D. Ernst, B. De Schutter, and R. Babuška, "Approximate dynamic programming with a fuzzy parameterization," *Automatica*, vol. 46, no. 5, pp. 804–814, 2010. 125

[118] L. Busoniu, D. Ernst, B. De Schutter, and R. Babuska, "Approximate reinforcement learning: An overview," in *Adaptive Dynamic Programming And Reinforcement Learning (ADPRL), 2011 IEEE Symposium on*, pp. 1–8, IEEE, 2011. 125

[119] H. Boubertakh, M. Tadjine, P.-Y. Glorennec, and S. Labiod, "Tuning fuzzy pd and pi controllers using reinforcement learning," *ISA transactions*, vol. 49, no. 4, pp. 543–551, 2010. 125

[120] R. Sharma and M. Spaan, "Fuzzy reinforcement learning control for decentralized partially observable markov decision processes," in *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*, pp. 1422–1429, 2011. 125