Power Control in Energy-Harvesting Small Cell Networks: Application of Stochastic Game

by

Thuc Kien Tran

A Thesis submitted to The Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg

May 2015

Abstract

Energy harvesting in cellular networks is an emerging technique to enhance the sustainability of power-constrained wireless devices. In this thesis, I consider the co-channel deployment of a macrocell overlaid with several small cells. In our model, the small cell base stations (SBSs) harvest their energy from environment sources (e.g., solar, wind, thermal) whereas the macrocell base station (MBS) uses conventional power supply. Given a stochastic energy arrival process, a power control policy for the downlink transmission of both MBS and SBSs is derived such that they can obtain their own objectives on a long-term basis (e.g., maintain the target signal-to-interference-plus-noise ratio [SINR] on a given transmission channel). To this end, I propose to use two different forms of stochastic game for the cases when the number of SBSs is small and when it becomes very large i.e. a very dense network. Numerical results demonstrate the significance of the developed optimal power control policy in both cases over the conventional methods.

Acknowledgements

There are a number of people I wish to thank for making my experience as a master student an interesting and rewarding one. First of all, I would like to thank my advisor Prof. Ekram Hossain for his extraordinary support and guidance. His energy, patience, and encouragement have been truly priceless. Without him, this thesis may not be completed. I also wish to acknowledge the financial support from University of Manitoba and the Natural Sciences and Engineering Research Council of Canada (NSERC) during my graduate work here.

Secondly, I would like to thank all the professors in our departments. Their courses were interesting and helpful, which contributed a lot to this thesis work. I would like to thank the staffs at the Department of Electrical and Computer Engineering, especially Amy Dario for her kind help. My sincere gratitude goes to all my colleagues in the lab whose support and friendship was invaluable.

Finally, I would like to thank my parents for their unbounded care, love, and support. You are always the persons I love the most.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 What is Small Cells and Why Energy Harvesting ?

A report made by Cisco has shown that the number of devices connected to Internet will reach 25 billions in 2015 [1], nearly four times the number of people on Earth. Also, the evolving fifth-generation (5G) cellular wireless networks are expected to support a super fast and stable download rate of 1Gb/s. That means a tremendous amount of data rate that the network must provide during peak time. The most viable way to alleviate this challenge is to make the cells smaller, smarter and denser. Small cells including micro, pico, and femto cells, can greatly increase the spatial reuse of radio resources and therefore improve the spectral efficiency (and hence the date rate). Moreover, small cell base stations (SBSs) can be easily deployed, avoiding the expensive cell site acquisition and operating cost. A typical small cell network will include:

- Macrocells: In a macrocell, the base station (BS) covers a large area (radius $\geq$ 1 km) and support many users. Total power consumption of a macrocell BS (MBS) is on the order of few hundred of Watts. Its installation requires careful planning and high set-up cost. The operating cost is high due to high power transmission.

- Small cells (e.g., microcells, picocells, femtocells): "Small cell" is an umbrella term for low-power radio access nodes that operate in both licensed and unlicensed spectrum and have a range of 10 meter to several hundred meters. They consume much less power compared to macro cell BS. SBSs can be installed without much planning.

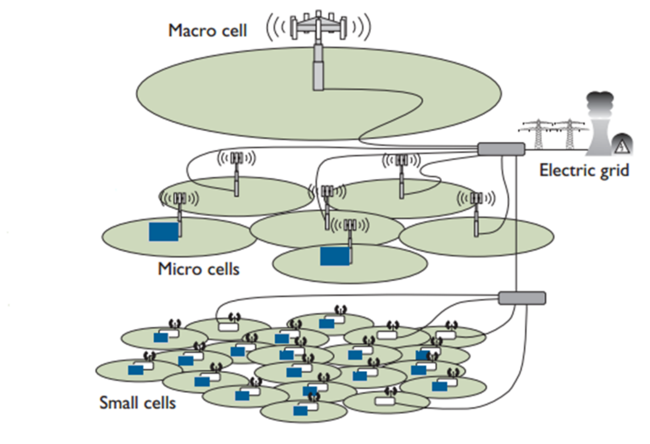

Figure 1.1: Architecture of a small cell network [2].

Obviously, increasing the number of small cells will increase coverage and spectral reuse. However, more and more small cells added into the network means higher power consumption. Also, SBSs cannot operate in some areas where accessibility to the power grid is not available. Therefore, using energy harvesting technologies to collect ambient renewable energy and power up the SBSs is a promising solution to reduce both power bill and the pollution while increasing the network coverage in difficult terrains.

1.2 Overview of Energy Harvesting Technologies for Wireless Communication Networks

With the popularity of IP-connected devices and the concerns about CO_2 pollution, energy harvesting techniques in wireless communication networks have recently gained a considerable attention. Energy-limited systems e.g., wireless sensor networks, are equipped with fixed energy supply devices such as batteries which possess limited operation time and energy. For applications where replacing the energy source is costly and unreliable e.g., in toxic environments or unreachable terrains, energy harvesting (EH) appears as a reasonable solution for safe and unlimited energy supply to communication networks. Moreover, Internet technology is growing so fast and soon it will be literally embedded into every aspects of our daily existence. From big things like fridge, car, laptop to small items like shoes and shirt, all of them will become smarter and interconnected. This densely interconnected system is referred to as the "Internet of Things" (IoT). However, to create such system, we must figure out how to design methods such that these IoT nodes can power itself and operate without the need of human intervention.

A traditional model requires an AC power line or a conventional battery. But such installment is costly and infeasible in many cases, also the maintenance cost will easily exceed the benefits. Using large battery to power them is also not sustainable since it can create dangerous source of both heat and chemicals when exposed. Therefore, small batteries with built-in energy harvesting device is the most promising method. ZigBee, which only requires a power of 1mW to transmit at rate 250kb/s, is a candidate that can be used along with energy harvesting to power a system of IoT nodes. The benefits of energy harvesting is obvious: self-sustainability, easiness for operation, less maintenance and set-up cost, and virtually no power bill. These features make energy harvesting wireless technology the ideal method for IoT to easily and reliably interconnect thousands of individual devices in a dense system.

1.2.1 Sources of Energy Harvesting

There are many harvesting approaches that have been successfully demonstrated including wind, solar, vibrational, biochemical, and motion based etc. The amount of energy captured from the environment is highly dependent on the source. Power densities of different harvesting technologies are shown in Table 1.1.

Table 1.1: Characterizations of energy sources [3]

Energy sources	Characteristics	Amount of available energy	Conversion efficiency	Amount of harvested energy
Solar	Ambient,uncontrollable, predictable	100mW/cm^2	15%	15mW/cm^2
Wind	Ambient,uncontrollable, predictable	-	-	1200mWh/day
Finger motion	Piezoelectric, Fully controllable	19mW	11%	2.1mW
Exhalation	Passive human power, uncontrollable, unpredictable	1W	40%	0.4W
Breathing	Passive human power, uncontrollable, unpredictable	0.83W	50%	0.42W
Blood pressure	Passive human power, uncontrollable, unpredictable	0.93W	40%	0.37W

Certainly, solar power has the highest harvesting rate and reliability compared to other methods. A solar panel converting light photon to electricity and the amount of energy harvested directly proportional to the size of the panel and the intensity of the light. Recently, researchers in Fraunhofer Institute for Solar Energy Systems have developed new technologies that can convert 46% of solar light into energy. That means by deploying 1 cm^2 solar panel during a normal sunny day we can obtain an average energy rate at 40 mW/cm^2 or 4 $W/100cm^2$. Solar power is uncontrollable but its pattern can be correctly estimated. The fairly cheap installation makes solar energy the best candidate to power SBSs, especially for developing countries in Asia and Africa where it is sunny most of the daytime.

Another source of energy is from wind turbine. The wind simply rotates the rotors and generate power. Fortunately, wind and solar power are good complement to each other. During daytime, solar panel is clearly favorable while at night wind turbine is used. Ambient radio frequency (RF) is also a promising technique to power energy harvesting devices. An AC voltage is generated when a time-varying electromagnetic RF field passes through an antenna coil. Experiments have shown that an RF device can harvest up to 189μW from a 3W source power at 5 meters away. This is quite small but if the number of sources increases then the harvested energy is also higher. It may not be possible for a device to run continuously with RF energy, however we can recharge the battery and use them to transmit data later. Clearly the design of an RF device is less bulky than wind/solar harvester. Therefore, RF-harvesting is suitable for in-door applications.

Piezoelectric technology can be also used for RF-harvesting. When a mechanical force deforms a piezoelectric material, a current will be produced. This kind of technology is applied for medical devices where blood pressure or body movement can power up their operations.

1.2.2 *Energy Harvesting Architecture*

The architectures of energy harvesting devices can be divided into two categories: Harvest-Use and Harvest-Storage-Use architectures.

- *Harvest-Use*: In this architecture, energy is harvested just in-time for use. There is no storage capacity for the device, so the device will only have two states "On" and "Off". If the energy harvested is enough the device is on, otherwise this amount will be wasted. This is usually the case with devices harvesting from piezoelectric sources. For example, a key/button when pushed will produce a small amount of energy to trigger the system. The mathematical model for this architecture is as follow: Let $P_h(t)$ is the power output from an energy source and $P_c(t)$ is the consumed power of the device. Then, at any time t we must have $P_h(t) \geq P_c(t)$.

- *Harvest-Storage-Use*: Devices have built-in rechargeable battery. For this case, energy can be stored for future usage and thus increase the operation time of the devices. This is usually deployed with wind/solar device where the harvested energy is large. Also, there are several ways to model the battery energy storage: i) An ideal case where it has infinite capacity, thus no harvested energy is wasted. ii) A practical model where battery has finite battery and also suffers from leakage.

  Again, if we use $P_h(t)$ and $P_c(t)$ as defined in previous case and denote E_0 as the initial battery, one must have:

$$B_0 + \sum_{t=0}^{T} [P_h(t) - P_c(t)] \geq 0$$

or in continuous form as:

$$B_0 + \int_{t=0}^{T} [P_h(t) - P_c(t)]dt \geq 0$$

Notice that in this case $P_h(t)$ can be smaller than $P_c(t)$. This means the device can adapt its strategy, transmit more or less depending on the conditions.

1.2.3 Feasibility of Energy Harvesting for SBSs

From [4], a traditional MBS requires large stable power supply thus energy harvesting is not a practical solution for macrocell base stations (MBSs) due to their high power consumption and stochastic nature of energy harvesting sources. On the other hand, it is appealing for small cell BSs (SBSs) (such as picocells or femtocells) that typically consume less power [5]. Following [4], a pico and femto MS only requires 7.3W and 5.2 W, respectively, in total. This amount can be easily provided with a 20×20 cm^2 solar panel which is cheap and available in the market. Other benefits are:

- The communication distance between SBSs and users is much smaller than that between an MBS and users when the density is increased. And since the transmit power increases with the distance, the SBSs require far less power to operate than an MBS.

- The baseband processing in an SBS is much simpler. Also, it serves less users and thus consumes less power than an MBS.

- Since an MBS transmits with a large power, it requires special cooling system. The SBSs on the other hand can use natural air circulation, which can save up to 10% of total power [6].

- Since no fixed power cables are required, it provides increased portability and thus makes the planning easier and more flexible.
- Lastly, it can reduce carbon emission thanks to renewable technologies.

1.3 Challenges and Current Trends of Research

The main reason that make energy management policy using harvesting technologies different and more difficult than other power control problem is its stochastic nature of energy arrival. For a system where power supply is fixed and unlimited, an operator only need to solve an optimization problem to obtain the optimal power. An action in time t will not depend on actions that happen before t and after t. It is not true with energy harvesting. We must optimize the transmit power in the long run while satisfying the causality constraint at any time instant. Here we need to find a power policy that specifies how the devices should transmit and with how much power constrained by a given available energy. Therefore, designing efficient power control policies with different objectives (e.g., maximizing system throughput) is a very challenging problem in energy-harvesting networks. Usually, we follow two directions as discussed below.

1.3.1 Offline Optimization

In offline method, we assume exact knowledge about the harvested energy $\varphi(t)$ at any time slot t. This method is usually applicable for solar source where the amount of energy $\varphi(t)$ is usually large and predictable at a specific time of the year. Offline throughput maximization for energy harvesting systems has recently received considerable interest. In the simplest case, given a duration T during which no energy is harvested, by using Jensen's inequality, it can be proven that transmitting using constant power during this period is the best solution to maximize the throughput. That means transmit power should be changed only when there is energy to arrive. In [7], with a harvest-store-transmit energy harvesting point-to-point systems using an unlimited capacity battery that operates over a static channel, the authors show that to achieve the minimum transmission completion time of a given amount of data, the transmit power should be an increasing function and also it is unchanged during the arrivals of two energy packets. In [8] and [9], the authors extend the work with finite capacity battery and fading channel. The power control policy is obtained using a directional waterfilling algorithm. The result is intuitive: when the channel gain is bad, the device transmits less and save this energy for future use. In [10], the authors show that: if we draw a graph of the cumulated energy harvested and the minimum amount energy that must be consumed over time t as $M(t)$, the optimal energy consumed over time will be the shortest path between the two end point of $H(t)$. Fig. 1.2 shows an example of the optimal energy consumption strategy. The dashed lines are the total harvested energy and minimum energy consumed, respectively. The pink line is the optimal energy consumption policy.

Maintaining the quality-of-services by minimizing the outage probability constraints in energy harvesting has also received much attention. In [11], the authors study a point-to-point energy harvesting system using Weibull fading model. The outage probability is shown to have concave-convex characteristics. From that, an algorithm based on one-dimensional search is proposed to find the optimal power policy such that the outage probability is minimized. In [12], the authors generalized the point-to-point model to the case where many sources supply energy to the destinations using a single relay. A water filling algorithm was proposed to minimize the

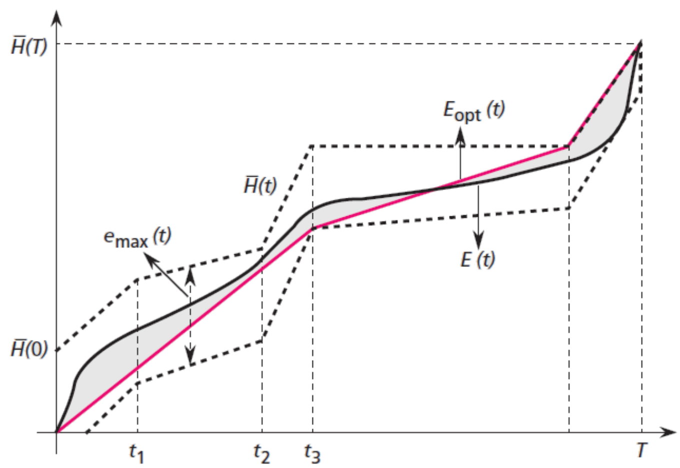

Figure 1.2: Optimal transmit power for offline model [10].

probability of outage.

1.3.2 Online Optimization

Although the offline power control policies provide an upper bound for online algorithms, centralized knowledge of energy/data arrivals is required which may not be feasible in practice. Therefore, recently a lot of works have focused on system models where the energy arrivals are stochastic. Markov Decision Process (MDP) and stochastic dynamic programming are very useful tools since they specialize in dealing with uncertainty in the future energy/data arrivals. Examples can be found in [11] and [13]. In [14], the authors proposed a two-state Markov Decision Process (MDP) model for a single energy-harvesting device considering random rate of energy arrival and different priority levels for the data packets. The authors proposed a low-cost balance policy to maximize the system throughput by adapting the energy harvesting state, such that, on average, the harvested and consumed energy remain balanced.

In [15], the outage performance analysis was conducted for a multi-tier cellular network in which all BSs are powered by the harvested energy and each BS has only two states: "On" when the battery is above threshold and "Off" otherwise. To reduce the effect of stochastic energy arrival, a hybrid system can be designed. In such a system, beside the energy harvesting functionality, the transmitter can receive energy from a conventional power source. Online resource optimization with hybrid power supply is studied in [16] and [17]. The objective here is to minimize the total power drawn from the conventional source while maintaining the QoS.

1.3.3 Research Challenges

- *Randomness in energy harvesting*: This is the main challenging aspect that makes energy harvesting difficult to apply into practice. Any change in the environment can affect the energy harvesting rate which in turn can degrade performance of the base station. Usually, a hybrid model where the base station can have both renewable and conventional power source should be used to maintain stable performance. Moreover, the transmit power of the base station will be constrained by its available energy which, in turn, depends on the transmit power and the harvested energy in previous time slot. That means the proposed mathematical model should optimize the long term average payoff rather than a one-shot game/optimization problem as usual.

- *Imperfect information*: Normally, to derive the the optimal power control policy, we need to assume that certain amount of information is available. For offline case, both energy and data arrivals for the whole duration are assumed to be known. This case is too extreme and almost impossible. This assumption is relaxed for online optimization where only the statistical information of the arrivals is available. However, in practice, these information also vary over time or in some cases cannot be predicted beforehand. Thus a learning-based algorithm such as Q-learning should be developed.

- *Interference in an energy harvesting (EH) system*: The randomness of energy arrival make it difficult to study a system with many energy harvesting devices. In this case, the battery of each device is a random variable and dependent on others. Therefore, most of the research focus on point-to-point communication which simply ignores the effect of interference. Because the transmit power of each device will directly affect the performance of others, the complexity will increase with the number of devices in the network. Also, since each device is not able to obtain the channel gain and battery levels of all devices in the network, solutions based on local information should be sought instead.

- *Accurate modeling of energy harvesting process and storage imperfection*: Many research assume Poisson distribution to model the arrivals of energy and data. In many cases these information is not accurate, energy arrivals are different at different time and places. Moreover, leaks and inefficiency of power usage may occur. This may not be a significant issue with MBSs, but for transmitters with small battery such as SBSs, this effect is not negligible.

1.4 Contribution and Outline of the Thesis

Different from the existing literature on energy-harvesting systems, this thesis considers the power control problem for downlink transmission in two-tier macrocell-small cell networks under co-channel deployment. Note that the power-control policies and their resulting interference levels directly affect the overall system performance. Thus, the impact of interference, which was ignored in most of the previous studies, is also considered along with the Gaussian noise. The problem is similar to that addressed in [18, 19], where the authors studied the power control problem for two-tier cellular networks using game models. However, in [18, 19], the stochastic nature of the energy arrival process, which dynamically constrains the transmission power of the SBSs, was not taken into account. Therefore, the problem addressed in this work is substantially more difficult since it optimizes the long-term payoff of both SBSs and macro base station under uncertainty.

For energy harvesting systems, MDP is a natural method to deal with randomness of the energy/data arrivals. When no information is available, a learning based method such as Q-learning can be applied so that the transmitter can adapt its transmission power based on the previous experiment [20]. However, most of the exisiting work in this context consider only point-to-point systems where there are only one transmitter and one receiver. To the best of my knowledge, there is no comprehensive work which studies the control policy for a large system where the transmitters in an energy harvesting system can interfere with each other. In this context, in Chapter 3 of this thesis, I propose the application of single-controller stochastic game into system of energy harvesting transmitters. The power control policy of the macrocell base station, which is often ignored in literature, is also taken into consideration. Moreover, using a centralized policy, the discounted payoff at each transmitter is optimized. Instead of each SBS harvesting its own energy, a central system is introduced that can harvest energy and then redistribute them to each SBSs. With this design, the problem can be modeled as a single-controller stochastic game and is able to capture the randomness of energy arrival. The Nash equilibrium power control policy is then obtained as the solution of a quadratic programming problem. Numerical results demonstrate that the proposed single-controller discrete stochastic power control policy offers reduced outage probability for the users served by the SBSs when compared to the greedy power control policies wherein each SBS tries to obtain the target SINR for the users without considering the strategies of other SBSs.

In Chapter 4, the work is extended to the case when the number the SBSs is very large such that using a central system to redistribute the harvested energy is no longer feasible. Instead, the system is remodeled where each SBS now is an autonomous harvest-store-transmit device. The stochastic game is then approximated by a mean field game (MFG). In the MFG model, the transmit power of each SBS is decided by its current battery level and the distribution of energy in the area. By solving a set of forward and backward partial differential equations, a distributed power control policy is derived for each SBS. Moreover, this model can also deal with energy leaks which is rarely mentioned in literature. Numerical results show that the proposed power control policy provides a better SINR compared to the MDP method where each SBS maximizes its own objective without consisdering the actions of other SBSs.

In summary, the contributions of the work can be summarized as follows.

1. For a two-tier macrocell-small cell network, a new centralized model, where energy is harvested and then distributed to the SBSs, is proposed. The power control problem for the MBS and SBSs is formulated as a discrete single-controller stochastic game with two players.

2. The existence of the Nash equilibrium and pure stationary strategies for this single-controller stochastic game is proven. The power control policy is derived as the solution of a quadratic-constrained quadratic programming problem.

3. When the network becomes very dense, a stochastic MFG model is used to obtain the power control policy as a solution of the forward and backward differential equations.

4. An algorithm using finite difference method is proposed to solve these forward-backward differential equations for the MFG model.

The rest of this thesis is organized as follows. Chapter 2 introduces some fundamentals of single-controller stochastic game and mean field game which will be used in following chapters. A centralized system model is introduced in Chapter 3 and an algorithm for downlink power control at the SBSs based on stochastic game is proposed. In Chapter 4, a very dense network using energy harvesting is studied using mean field game technique. Finally, Chapter 5 concludes the thesis and presents several potential extensions of the work presented in this thesis.

Chapter 2

Fundamentals of Stochastic Game

Stochastic game, introduced by Shapley in 1950, is a dynamic multiple stage game. The game changes its form at each stage with some probability. The total payoff to a player is often the discounted sum of the stage payoffs or the limit inferior of the averages of the stage payoffs. The transition of the game at each time instant follows Markovian property, i.e., the current stage only depends on the previous one. Therefore, a stochastic game can be viewed as a generalization of both matrix game and MDP. The single-controller game is a special form of stochastic game where the state of the game is directly decided by one player, namely, the controller. To understand stochastic game, we first start with MDP and some fundamentals of game theory.

2.1 Markov Decision Process

A **stochastic process** is simply a collection of random variables $\{X_t : t \in T\}$ where T is an index set that we usually think of as representing time. In this thesis, I only work with the discrete cases where index set T is a countable set usually denoted as the integer set $\mathcal{Z}$.

Definition 2.1.1. *A discrete **Markov Chain** $\{X_n; n \geq 0\}$ with values belonging to set Ω is a stochastic process if for every $n \geq 0$ and every set of $x_0, ..., x_n \in \Omega$ we have*

$$\mathbb{P}(X_{n+1} = x | X_0 = x_0, X_1 = x_1, ..., X_n = x_n) = \mathbb{P}(X_{n+1} = x | X_n = x_n).$$

The chain X is called homogeneous if

$$\mathbb{P}(X_{n+1} = i | X_n = j) = \mathbb{P}(X_1 = i | X_0 = j), \quad \forall n \in \mathbb{Z} \quad i, j \in \Omega.$$

Clearly in a homogeneous Markov Chain the transition probability that the system moves from one state to another state only does not depend on time.

Definition 2.1.2. *A (one-step) transition matrix $\mathbf{P}$ of a homogeneous Markov chain X is a $|\Omega| \times |\Omega|$ stochastic matrix where*

$$\mathbf{P}(i, j) = \mathbb{P}(X_1 = i | X_0 = j) \quad and \quad \sum_j \mathbf{P}(i, j) = 1.$$

By extending this definition, we can have an n-step transition matrix of the homogeneous Markov chain X as a stochastic matrix $\mathbf{P}^{(n)}$, where

$$\mathbf{P}^{(n)}(i, j) = \mathbb{P}(X_n = i | X_0 = j).$$

Theorem 2.1.1. ***(Chapman-Kolmogorov Equations)*** *Assume that X is a time-homogeneous Markov Chain with n-step transition probabilities matrix $\mathbf{P}^{(n)}$. Then,* *for any non-negative integer $r < n$, we have:*

$$\mathbf{P}^{(n)}(i,j) = \sum_{k\in\Omega} \mathbf{P}^{(r)}(i,k)\mathbf{P}^{(n-r)}(k,j), \quad \forall i,j \in \Omega.$$

One of the most important features of the Chapman-Kolmogorov equations is that they can be succinctly expressed in terms of matrix multiplication. It is easy to see that the n-step transition probability matrix can be calculated as a power of n of the one-step matrix, i.e., $\mathbf{P}^{(n)} = \mathbf{P}^n$.

The **Markov Decision Process** (MDP) is a discrete time stochastic control process where the only player decides which action should he choose at a specific state. I only consider a Stationary Discounted MDP $\mathbf{\Gamma}$ in this thesis. This type of MDP consists of:

- A decision maker or controller.
- The state s of $\mathbf{\Gamma}$ is a random variable which belongs to a finite set of state S.
- At each state s, denote by $A(s)$ the set of actions available to the controller.
- A probability distribution $p(s'|s,a)$ that shows the probability from state s to state s' following action $a \in A(s)$.
- A payoff function $u(s,a)$ which shows the profit the controller receive at state s if he chooses action a.
- A discount factor β that reduces the value of the payoff over time.

Without loss of generality, we assume the set of state as $\mathbf{S} = \{1, 2, .., S\}$ and the number of available actions for controller at state s is $m(s)$. Then a strategy is a concatenated row vector $\mathbf{f} = \{\mathbf{f}(1), ..., \mathbf{f}(s), ..., \mathbf{f}(S)\}$ where each s-th block is a non-negative row vector $\mathbf{f}(s) = (\mathbf{f}(s,1), \mathbf{f}(s,2), ..., \mathbf{f}(s, m(s)))$. Each entry $\mathbf{f}(s,a)$ is the probability that the controller chooses action $a \in \mathbf{A}(s)$ in state $s \in \mathbf{S}$ whenever s is visited. Obviously, one must have $\sum_{a=1}^{m(s)} \mathbf{f}(s,a) = 1$

A strategy is called *pure* if $\mathbf{f}(s,a) \in \{0,1\}$ for all $a \in A(s)$. This means every time the state s is visited, the controller should always choose action a_s to execute. This greatly simplifies the implementation and is preferred over randomly choosing an action. From $\mathbf{f}$, we can define an $S \times S$ transition matrix which shows the probability for the process to change from state s to s' as:

$$\mathbf{P}(\mathbf{f}) = (p(s'|s,\mathbf{f}))_{s,s'=1}^{S\times S},$$

where the entries given by $p(s'|s,\mathbf{f}) = \sum_{a=1}^{m(s)} p(s'|s,a)f(s,a)$

Let $\{U_t\}_{t=0}^{\infty}$ denote the sequence of payoff for the controller. The payoff at time t, U_t is a random variable whose distribution decided by the strategy $\mathbf{f}$ and the starting state s. We can denote the expected value of U_t as:

$$\mathbb{E}_{s\mathbf{f}}[U_t] = \mathbb{E}_{\mathbf{f}}(U_t|S_0 = s).$$

Denote by $u(s,\mathbf{f})$ the immediate expected payoff at state s if the controller uses strategy $\mathbf{f}$ at this stage. Different from U_t, this expected payoff is calculated regardless of the previous states and actions. That means $u(s,\mathbf{f}) = \sum_{a\in A(s)} u(s,a)f(s,a)$. The vector $\mathbf{u}$ of immediate expected payoff will be the concatenation of u over state s:

$$\mathbf{u}(\mathbf{f}) = (u(1,\mathbf{f}), u(2,\mathbf{f}), .., u(S,\mathbf{f}))^{\mathrm{T}}.$$

Denote by $[\mathbf{u}]_s$ the s-th element of the vectir $\mathbf{u}$. From Chapman-Kolmogorov Equations, we have

$$\mathbb{E}_{s\mathbf{f}}[U_0] = [\mathbf{u}(\mathbf{f})]_s,$$
$$\mathbb{E}_{s\mathbf{f}}[U_1] = [\mathbf{P}(\mathbf{f})\mathbf{u}(\mathbf{f})]_s,$$
$$\ldots$$
$$\mathbb{E}_{s\mathbf{f}}[U_n] = [\mathbf{P}^n(\mathbf{f})\mathbf{u}(\mathbf{f})]_s.$$

The discounted value of strategy $\mathbf{f}$ from initial state s will be defined as:

$$\phi(s, \mathbf{f}) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{s\mathbf{f}}[U_t] = \sum_{t=0}^{\infty} \beta^t [\mathbf{P}^t(\mathbf{f})\mathbf{u}(\mathbf{f})]_s.$$

Lemma 2.1.1. *Denote $\phi(\mathbf{f})$ as the vector that contains $\phi(s, \mathbf{f})$, then we have the vector form of the above equation as:*

$$\phi(\mathbf{f}) = \sum_{t=0}^{\infty} \beta^t \mathbf{P}^t(\mathbf{f})\mathbf{u}(\mathbf{f}) = [\mathbf{I} - \beta\mathbf{P}(\mathbf{f})]^{-1}\mathbf{u}(\mathbf{f}),$$

or equivalently, $\phi(\mathbf{f}) = \mathbf{u}(\mathbf{f}) + \beta\mathbf{P}(\mathbf{f})\phi(\mathbf{f})$

Proof. It can easily be seen that:

$$[\mathbf{I} - \beta\mathbf{P}][\mathbf{I} + \beta\mathbf{P} + ... + \beta^n\mathbf{P}^n] = \mathbf{I} - \beta^{n+1}\mathbf{P}^{n+1}.$$

Since $1 > \beta > 0$, letting $n \to \infty$, $[\mathbf{I} - \beta\mathbf{P}][\mathbf{I} + \beta\mathbf{P} + ...] = \mathbf{I}$
This means the determinant of matrix $[\mathbf{I} - \beta\mathbf{P}]$ is non-zero. Therefore, $\mathbf{I} + \beta\mathbf{P} + ... = [\mathbf{I} - \beta\mathbf{P}]^{-1}$. This completes the proof. □

Definition 2.1.3. ***Principle of Optimality****: An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.*

We want to find the strategy $\mathbf{f}$ such that it maximizes the discount value $\phi(s, \mathbf{f})$ where the starting state is s. We can apply this principle to derive the linear programming for finding optimal policy $\mathbf{f}$ as follows. Suppose at state s the controller knows how to optimally control the process from the next time period onward, then given that strategy, at this current time period he must choose the action such that it maximizes the payoff including that period, i.e.,

$$\phi(s) = \max_{a \in A(s)} \left\{ u(s,a) + \beta \sum_{s'=1}^{S} p(s'|s,a)\phi(s') \right\},$$

or equivalently,

$$\phi(s) \geq u(s,a) + \beta \sum_{s'=1}^{S} p(s'|s,a)\phi(s') \quad \forall a \in A(s),$$

For an arbitrary strategy $\mathbf{f}$, we multiply each of the above inequalities by the corresponding $\mathbf{f}(s,a)$ and sum over all $a \in \mathbf{A}(s)$, we obtain:

$$\phi(s) \geq u(s,\mathbf{f}) + \beta \sum_{s'=1}^{S} p(s'|s,\mathbf{f})\phi(s') \quad \forall s \in \mathbf{S},$$

or in matrix form

$$\phi \geq \mathbf{u}(\mathbf{f}) + \beta \mathbf{P}(\mathbf{f})\phi. \tag{2.1}$$

Substitute this inequality infinitely to itself we have $\phi \geq [I - \beta \mathbf{P}(\mathbf{f})]^{-1}\mathbf{u}(\mathbf{f}) = \phi(\mathbf{f})$. This suggests that an arbitrary vector ϕ satisfies the inequality above is an upper bound on the discounted value vector due to any stationary strategy $\mathbf{f}$. Denote by $\pi = (\pi(1), ..., \pi(S))$ as the vector of probabilities that system starts from a specific state, based on the previous remark, the optimal ϕ will be the solution of this following optimization problem:

$$\begin{aligned} &\min_{\phi} \quad \sum_{s=1}^{S} \pi(s)\phi(s) \\ &\text{s.t.} \quad \phi(s) \geq u(s,a)+\beta\sum_{s'=1}^{S} p(s'|s,a)\phi(s'), \quad a \in A(s), \quad s \in \mathbf{S} \end{aligned} \tag{2.2}$$

and its dual problem

$$\begin{aligned} &\max_{x_{s,a}} \quad \sum_{s=1}^{S}\sum_{a=1}^{|A(s)|} u(s,a)x_{s,a} \\ &\text{s.t.} \quad \sum_{s=1}^{S}\sum_{a=1}^{|A(s)|}[\delta(s,s')-\beta p(s'|s,a)]x_{s,a} = \pi(s') \quad s' \in \mathbf{S} \\ &\qquad x_{s,a} \geq 0; \quad a \in A(s), \quad s \in \mathbf{S}. \end{aligned} \tag{2.3}$$

Theorem 2.1.2. *[27](Validity of the Optimality Equation)*

- *There exists a unique solution u for the linear programming (2.2) and (2.3).*
- *For each $s \in \mathbf{S}$, select action $a_s \in A(s)$ such that*

$$a_s = \arg\max_{a \in A(s)} \left\{ u(s,a) + \beta \sum_{s'=1}^{S} p(s'|s,a)\phi(s',\mathbf{f}) \right\},$$

where ϕ is the solution of (2.2). Define $\mathbf{f}^*$ as

$$\mathbf{f}^*(s,a) = \begin{cases} 1, & \textit{if} \;\; a = a_s, \\ 0, & \textit{otherwise}, \end{cases}$$

for each $s \in \mathbf{S}$. Then $\mathbf{f}^*$ is the optimal deterministic strategy of the MDP.

This theorem shows that an MDP always has a deterministic strategy which tells the controller exactly which action it should follow at a specific state. Also it provides us the method to find such strategy using linear programming.

2.2 Two Player Single-Controller Stochastic Game

2.2.1 Game Theory Overview

Game theory is a branch of mathematics that deals with decision making. It studies mathematical models where there are conflicts of interest or mutual benefits between "rational" decision-makers. It was first developed to explain economic behaviors but later its applications have been seen in many other areas especially in wireless communications. Some examples are listed in Table 2.1.

Table 2.1: Applications of game theory in networking [35]

OSI layer	Applications	Specific application
Physical layer	Power control, Spectrum allocation, MIMO systems, Cooperative communications	Power control in CDMA, OFDMA networks, Spectrum sharing, spectrum bidding, Power managment in MIMO, Decode-and-forward cooperation
Data link	Medium access control	Slotted Aloha, Random access to the interference channel
Network	Routing	Routing and forwarding
Transport	Call admission control	Request distribution among providers

In this part, we only mention about non-cooperative game and some of its basic definitions.

Definition 2.2.1. *A strategic game is a tuple* $\langle \mathcal{I}, (A_i)_{i\in\mathcal{I}}, (u_i)_{i\in\mathcal{I}} \rangle$, *where*

- *There is a set of players* $\mathcal{I} = \{1, ..., I\}$.

- *For each player $i \in \mathcal{I}$, a set of available actions (actions profile) A_i; $a_i \in A_i$ is denoted as the action of player i.*
- *For each player i, a payoff function $u_i : A \to \mathbb{R}$, where $A = \prod_i A_i$ is the set of all action profiles.*

Also, we denote by $a_{-i} = \{a_j\}_{j \neq i}$ the set of actions for all players except i and by $A_{-i} = \prod_{j \neq i} A_j$ the set of all actions profiles of all players except player i. Then the pair $(a_i, a_{-i}) \in A$ defines a strategy profile of the game. The **strategy** is a complete description of how to play the game. In particular, a **pure strategy** determines the specific move a player will make for any situation it could face.

One of the basic assumptions of game theory is the rationality of players. That means a player always chooses the strategies that maximizes its payoff. This common knowledge is very important because it helps each player to predict what others will do given a situation. In other words, given the actions set of other player as a_{-i}, player i will choose a **best response** action $a_i^* \in A_i$ such that his payoff is optimal, i.e.,

$$u_i(a_i, a_{-i}) \leq u(a_i^*, a_{-i}), \quad \forall a_i \in A_i.$$

Since each player selfishly and rationally maximizes its own payoff, it is desirable to find a set of actions for every player such that an equilibrium is achieved, i.e., no player has any incentive to move away. This leads to the concept of Nash equilibrium as defined below.

Definition 2.2.2. *(pure strategy Nash equilibrium) A (pure strategy) Nash Equilibrium of a strategic game $\langle \mathcal{I}, (A_i)_{i \in \mathcal{I}}, (u_i)_{i \in \mathcal{I}} \rangle$ is a strategy profile $a^* = (a_i^*, a_{-i}^*) \in A$ such that for all $i \in \mathcal{I}$*

$$u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*), \qquad \forall a_i \in A_i.$$

However, in many cases, the pure strategy Nash equilibrium does not exist. To deal with this problem, we can assign a probability to each pure strategy of a player. That means a player will randomly choose its action based on some probability distribution.

Definition 2.2.3. *(Mixed strategy of a game) Given the strategic game $\mathcal{G}$ defined as above, for each player i, we denote the set of probability distribution over its set of strategies A_i as Σ_i. Then a mixed strategy $\sigma_i : \Sigma_i \to \mathbb{R}$ is a function of A_i such that $\sigma_i(a_i) \geq 0 \quad \forall a_i \in A_i$ and $\sum_{a_i \in A_i} \sigma_i(a_i) = 1$. Denote by $\sigma \in \Sigma = \prod_{i \in \mathcal{I}} \Sigma_i$ the mixed strategy profile for all players, then the payoff for mixed strategy of player i is the expected value of utility function over the set A_i of available actions as*

$$u_i(\sigma) = \sum_{a \in A} \left(\prod_{j=1}^{I} \sigma_j(a_j) \right) u_i(a_i, a_{-i}).$$

The Nash equilibrium can be extended to include the mixed-strategy case.

Definition 2.2.4. *(Mixed strategy Nash equilibrium) A (mixed strategy) σ^* of a strategic game $\langle \mathcal{I}, (A_i)_{i \in \mathcal{I}}, (u_i)_{i \in \mathcal{I}} \rangle$ is a mixed Nash Equilibrium if for each player i*

$$u_i(\sigma_i^*, \sigma_{-i}^*) \geq u_i(\sigma_i, \sigma_{-i}^*) \qquad \forall \sigma_i \in \Sigma_i.$$

In this case, it is guaranteed that a mixed strategy Nash equilibrium always exists. In fact, if the action set and state space are finite and discrete, the mixed strategy of each player is similar to the vector of probabilities $\mathbf{f}$ as in the MDP process.

2.2.2 Single-Controller Stochastic Game

Stochastic game can be considered as a combination of Markov Decision Process and game theory. In this thesis, I only consider stochastic games with only two players. Each player has different states and action spaces. The state of the game is then the combination of the states of two players and the transition and reward matrices will depend on actions of both players.

Assume that $\mathbf{S}$ is the set of states for the game. If the game is in state s at time t, player 1 chooses action $a_1 \in A_1(s)$ and player 2 chooses action $a_2 \in A_2(s)$, then the rewards will be $r_1(s, a_1, a_2)$ and $r_2(s, a_1, a_2)$, respectively. Furthermore the transition probability will be $p(s'|s, a_1, a_2) = p(S_{t+1} = s'|S_t = s, a_1, a_2)$. Denote by $\mathbf{f}$ and $\mathbf{g}$ the strategies of player 1 and player 2, then the discounted reward for player $i \in 1, 2$ will become $\phi_i(\mathbf{f}, \mathbf{g}) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{s\mathbf{fg}}(R_t^i)$, where $\mathbb{E}_{s\mathbf{fg}}(R_t^i)$ is the expected value of the reward of player k at state s at time t if player 1 and player 2 follow strategy $\mathbf{f}$ and $\mathbf{g}$, respectively.

Denote by $\mathbf{F_S}$ and $\mathbf{G_S}$ the sets of strategies for player 1 and player 2, respectively. Here each player will try to maximize its own payoff, i.e., given strategy $\mathbf{g}$ of player 2. The **best response** strategy $\mathbf{f}$ of player 1 is the solution of the following MDP:

$$\begin{aligned} \max_{\mathbf{f}} \quad & \phi_1(\mathbf{f}, \mathbf{g}) \\ \text{s.t.} \quad & \mathbf{f} \in \mathbf{F_S}, \end{aligned}$$

and similarly given $\mathbf{f}$, the best response of player 2 is:

$$\begin{aligned} \max_{\mathbf{g}} \quad & \phi_1(\mathbf{f}, \mathbf{g}) \\ \text{s.t.} \quad & \mathbf{g} \in \mathbf{G_S}. \end{aligned}$$

Definition 2.2.5. *(Nash equilibrium for discounted stochastic game) A Nash equilibrium point in a stochastic game is a pair of strategies* $(\mathbf{f}, \mathbf{g})$ *for player 1 and player*

2, respectively, such that the following conditions are satisfied:

$$\phi_2(\mathbf{f}, \mathbf{g}_0) \leq \phi_2(\mathbf{f}, \mathbf{g}), \quad \forall\ \mathbf{g}_0 \in \mathbf{G_S}$$

$$\phi_1(\mathbf{f}_0, \mathbf{g}) \leq \phi_1(\mathbf{f}, \mathbf{g}), \quad \forall\ \mathbf{f}_0 \in \mathbf{F_S}.$$

That means, with Nash equilibrium, no player has incentive to change its strategy because its payoff will become less.

Theorem 2.2.1. *(Existence of Nash Equilibrium) For every general sum, discounted stochastic game, there exists a Nash equilibrium in stationary strategies.*

The Nash equilibriums of a stochastic game are the solutions of a quadratic constraint quadratic (QCQP) problem and is usually very difficult to solve [27, Chapter 3, Theorem 3.8.2]. However, in some special cases where the transition probabilities only depend on actions of one player, usually chosen as player 2 i.e $p(s'|s, a_1, a_2) = p(s'|s, a_2)$, the QCQP can be reduced to a quadratic programming and can be solved more effectively. The structure of a single controller game can be detailed as follows:

A two-player single-controller stochastic game is a tuple $(\mathbb{S}, \mathcal{N}, A^s, P, R^s)$, where

- $\mathbf{S}$ is a finite set of states
- $\mathcal{N}$ is the set of players , $\mathcal{N} = \{1, 2\}$ in this case.
- $A^s = \{A_1^s, A_2^s\}$ where A_i^s is the set of actions available at state s to player i.
- $\mathbf{P} = \mathbf{S} \times A \times \mathbf{S} \longmapsto [0, 1]$ is the transition probability function; $p(s'|s, a_2)$ is the probability that system transits to state s' from state s after action a_2 of player 2.

- $R^s = \{R_1^s, R_2^s\}$, where $R_i^s : S \times A^s \longmapsto R$ is the $m_s \times n_s$ real-valued payoff matrix for player i at state s (with m_s and n_s denoting, respectively, the number of available actions for the first and second player):

$$R_1^s = (u_1(s,i,j))_{i,j=1}^{m_s,n_s}, \qquad R_2^s = (u_2(s,i,j))_{i,j=1}^{m_s,n_s},$$

  where $u_k(s,i,j)$ is the expected payoff for player k if he plays action i while the other plays action j.

First, notice that if we somehow know the optimal strategy $\mathbf{f}$ of the first player, the problem of finding the best response strategy $\mathbf{g}$ of the second player is exactly a discounted MDP problem introduced previously:

$$\begin{aligned} &\min_{\phi_2} \sum_{s=0}^{S} \pi_s \phi_2(s,\mathbf{f},\mathbf{g}), \\ \text{s.t.} \quad &\phi_2(s,\mathbf{f},\mathbf{g}) \geq u_2(s,\mathbf{f},j) + \beta \sum_{s'=0}^{S} p(s'|s,j)\phi_2(s',\mathbf{f},\mathbf{g}), \\ &\forall s,j, \quad 0 \leq j \leq s \ \text{ and } \ 0 \leq s \leq S. \end{aligned} \tag{2.4}$$

The expected payoff u_2 can be calculated as $u_2(s,\mathbf{f},j) = \sum_{i \in A_1(s)} u_2(s,i,j)\mathbf{f}(s,i)$. This linear programming has the dual problem as:

$$\begin{aligned} &\max_{x} \sum_{s=0}^{S} \sum_{j=0}^{s} u_2(s,\mathbf{m}_0,j)x_{s,j}, \\ \text{s.t.} \quad &\sum_{s=0}^{S} \sum_{j=0}^{s} [\delta(s-s') - \beta p(s'|s,j)]x_{s,j} = \pi_{s'}, \quad 0 \leq s' \leq S, \\ &x_{s,j} \geq 0 \ \ \forall s,j, \quad 0 \leq j \leq s \ \text{ and } \ 0 \leq s \leq S, \end{aligned} \tag{2.5}$$

where $\delta(s) = 1$ if $s = 0$ and $\delta(s) = 0$ otherwise.

Using some algebraic formulation [28], these two primal and dual problems above can be expressed in matrix form as:

$$\begin{aligned} &\min_{\phi_2} \pi^{\mathrm{T}} \phi_2 \\ \text{s.t.} \quad & H\phi_2 \geq R_2^{\mathrm{T}} f \end{aligned}$$

and the dual

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{f}^{\mathrm{T}} R_2 \mathbf{x}, \\ \text{s.t.} \quad & \mathbf{x}^{\mathrm{T}} H = \pi^{\mathrm{T}}, \\ & \mathbf{x} \geq 0. \end{aligned} \tag{D}$$

The matrix R_2 can be constructed by placing the sequence of matrices $R_2^1, R_2^2, ..., R_2^S$ along the diagonal and set all other cells to zero. We can use a similar method to build reward matrix R_2.

Theorem 2.2.2. *(Nash equilibrium of single-controller stochastic game) If the state space and the action space are finite and discrete, and the transition probabilities are controlled only by player 2, then a pair* $(\mathbf{f}, \mathbf{g})$ *is a Nash equilibrium point of a general-sum single-controller discounted stochastic game if and only if it is an optimal solution* *of a (bilinear) quadratic program given by*

$$\begin{aligned}
\max_{\mathbf{f},\mathbf{x},\phi,\xi} \quad & [\mathbf{f}(R_1 + R_2)\mathbf{x} - \pi^T\phi_2 - \mathbf{1}^T\xi], \\
s.\ t. \quad & H\phi_2 \geq R_2^T\mathbf{f}, \\
& \mathbf{x}^T H = \pi^T, \\
& R_1^s\mathbf{x}(s) \leq \xi_s\mathbf{1}, \quad \forall s = 0, ..., S, \\
& \mathbf{f}(s)^T\mathbf{1} = 1, \quad \forall s = 0, ..., S, \\
& \mathbf{f}, \mathbf{x} \geq \mathbf{0},
\end{aligned} \tag{2.6}$$

where ξ_s *is the maximum average payoff of the MBS at state* s*. The sub-vector strategy* $\mathbf{g}(s)$ *of the second player at state* s *is calculated from* $\mathbf{x}$ *as:*

$$\mathbf{g}(s) = \frac{\mathbf{x}(s)}{\mathbf{x}(s)^T\mathbf{1}}. \tag{2.7}$$

This theorem is very useful since it provides us the necessary and sufficient condition to find the Nash equilibrium strategies for both players.

2.3 Stochastic Mean Field Game

When there are many players, it is very difficult to find a Nash equilibrium. Because everyone can change the state of the system so the complexity of is proportional to the number of players. Moreover the discrete set of actions and states spaces of each player make it nearly impossible to derive a closed form for the equilibrium strategies. To deal with these problems, mean field game (MFG) considers a very large game where the number of players is approximately infinite and the influence from one player to another is negligible. This is called **indistinguishability** property. All players are similar except their "state" which is assumed to be a vector of continuous values. The player is assumed to respond only to the distribution of the states in the system which is also called the **mean field** [31]. Since the number of players is infinite we assume that **this distribution is smooth** (so that it can be differented and integrated). Every player now has the same objective function which greatly simplifies the mathematical model and in many cases can give us the optimal strategies in closed-forms.

2.3.1 Formulation of the MFG

Denote by $X(t)$ the state of one player at time t and by $m(t, X)$ the probability distribution of the state X of an agent at time t. Thanks to the "indistinguishability" property, this distribution is also the density of the states of all players in the game, i.e., the "mean field" m. Assuming that m is known, the best response strategy $p(t, X)$ for each player is given by the solution of the following optimal control problem:

$$\min_{p} U = \mathbb{E}\left[u(T, x(T)) + \int_{0}^{T} C(p(t, X), X(t), m(t, X))dt\right], \tag{2.8}$$

where $C(p, X, m)$ is the cost of a player at time instant t. The cost only depends on the current state/action of the players and the mean field m. The state of the player is modeled by using the following differential equation:

$$\begin{aligned} dX &= p(t, X)dt + \sigma(t, X)dW_t, \\ X(0) &= x_0, \end{aligned} \tag{2.9}$$

where W_t is a Wiener process. W_t has independent increments: $W_t - W_s \approx N(0, t - s)$ $\forall 0 \leq s \leq t$, where N denotes the normal distribution and T is the stopping time of the process.

With MFG, everyone will need to solve the same optimal control problem. The only difference is the starting state x_0. Thus, the problem is much simpler since we only need to solve one optimal control problem for everyone.

The Hamilton-Jacobi-Bellman (HJB) equation for the optimal control problem above can be derived as:

$$\partial_t U(t,X) + \min_p \left\{ C(p,X,m) + p(t,X)\partial_X U(t,X) + \frac{\sigma^2(t,X)}{2}\partial^2_{XX} U(t,X) \right\} = 0. \tag{2.10}$$

Notice that, the function $p(t,X)$ obtained from solving the sub-optimal problem defined in the HJB gives us the optimal control policy for each player.

By applying Fokker-Planck theorem for differential equation (2.9), we have the backward equation:

$$\partial_t m(t,X) = -\partial_X(p(t,X)m) + \frac{\sigma^2(t,X)}{2}\partial^2_{XX}(m). \tag{2.11}$$

Together, we have the system of Forward-Backward equations that can give us the solutions of MFG:

$$\begin{cases} \partial_t U(t,X) + \min_p \left\{ C(p,X,m) + p\partial_X U(x,t) + \frac{\sigma^2(t,X)}{2}\partial^2_{XX} U(t,X) \right\} = 0, \\ \partial_t m(t,X) = -\partial_X(p(t,X)m) + \frac{\sigma^2(t,X)}{2}\partial^2_{XX} m, \\ \int_{X\in\Omega} m(t,X) = 1 \quad \text{and} \quad m(t,X) \geq 0. \end{cases}$$

The existence and uniqueness of the solutions of these equations are proven, and some special cases [31] and the numerical solutions can be obtained using Finite Difference method [34]. The details of the steps depend on the applications and will be explained in Chapter 4.

Chapter 3

Power Control in Small Cell Network With Centralized Energy Harvesting Queue

3.1 Introduction

In this chapter, I consider a centralized network where one macro cell is laid out with M small-cells. Each small cell is equipped with a harvest-transmit energy harvesting device. The small-cells will receive power from a central energy queue (CEQ) which can harvest energy from environment. The CEQ and the Macro Base Station (MBS) will try adapt their power respectively such that their long term equilibrium can be achieved. A stochastic game model between two players CEQ and MBS will be derived and the equilibrium policies are obtained by solving a Quadratic Constraint Quadratic Programming problem. This concept of CEQ is somewhat similar to the concept of dedicated power beacons that are responsible for wireless energy transfer to users in cellular networks [21, 22]. Moreover, in the cloud-RAN architecture [23], where along with data processing resources, a centralized cloud can also act as an energy farm that distributes energy to the remote radio heads each of which acts as an SBS. The experimental result also shows that if the battery size of CEQ is large enough, the usage of central energy queue outperforms greedy method where each base stations selfishly try to achieve it SINR regardless of others activities.

3.2 System Model and Assumptions

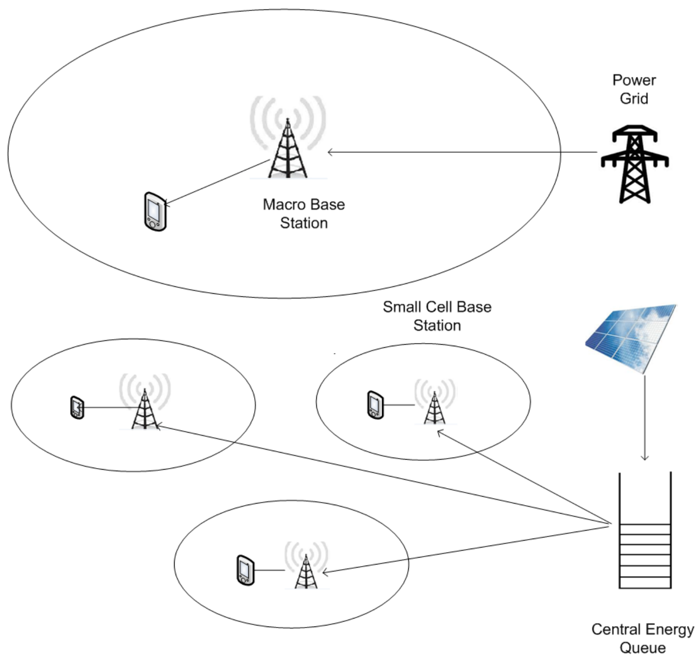

Figure 3.1: Network model with one macrocell overlaid with multiple small cells and a central energy queue (CEQ).

3.2.1 Energy Harvesting Model

I consider a single macrocell overlaid with M small cells. The downlink co-channel transmission of the MBS and SBSs is considered and it is assumed that each BS can serve only a single user on a given transmission channel during a transmission interval (e.g., time slot). The MBS uses a conventional power source and its transmit power level is quantized into a discrete set of power levels $\mathcal{P} = \{p_0^{min}, ..., p_0^{max}\}$, where the subscript 0 denotes the MBS. This discrete model of transmit power can also be found in [24]. On the other hand, the SBSs receive energy from a centralized energy queue (CEQ) which harvests renewable energies from the environment. I assume that only the CEQ can store energy for future use and each SBS must consume all the energy they receive from the CEQ at every time slot. The energy arrives at the CEQ in the form of packets (one energy packet corresponds to one energy level in CEQ). The number of energy packet arrivals $\varphi(t)$ during any time interval t is discrete and follows an arbitrary distribution, i.e., $\Pr(\varphi(t) = X)$. I assume that the battery at the CEQ has a finite storage S. Therefore, the number of energy packet arrivals is constrained by this limit and all the exceeding energy packets will be lost, i.e., $\Pr(\varphi(t) = S) = \Pr(\varphi(t) \geq S)$. Moreover, the statistics of energy arrival is known a priori at both the MBS and the CEQ. At time t, given the battery level $E(t)$, the number of energy packet arrivals $\varphi(t)$, and the energy packets $Q(t)$ that the CEQ distributes to the M SBSs, the battery level $E(t+1)$ at the next time slot can be calculated as follows:

$$E(t+1) = E(t) - Q(t) + \varphi(t). \tag{3.1}$$

Given $Q(t)$ energy packets to distribute, the CEQ will choose the best allocation method for the M SBSs according to their desired objectives. Denote slot duration as ΔT and the volume of one energy packet as K, I have the energies distributed to the M SBSs at time t as $(p_1(t)\Delta T, p_2(t)\Delta T, \cdots, p_M(t)\Delta T)$ where $p_i(t)$ is the transmit power of SBS i at time t. Clearly I must have:

$$\sum_{i=1}^{M} p_i(t) = \frac{K}{\Delta T} Q(t). \tag{3.2}$$

From the causality constraint, $E(t) \geq Q(t) \geq 0$, i.e., the CEQ cannot send more energy than that it currently possesses. Note that $E(t)$ is the current battery level which is an integer and has its maximum size limited by S. Since the battery level of CEQ and the number of packet arrivals are integer values, it follows from (3.1) that $Q(t)$ is also an integer.

Without a centralized CEQ-based architecture, each SBS can have different harvested energy and in turn battery levels at each time slot, which will make this problem a multi-agent stochastic game [25]. Although this kind of game can be heuristically solved by using Q-learning [26], the conditions for convergence to a Nash equilibrium are often very strict and in many cases impractical. By introducing CEQ, the state of the game is simplified into the battery size of CEQ, and the multi-player game is converted into a two-player game. Another benefit of the centralized CEQ-based architecture is that the energy can be distributed based on the channel conditions of the users in SBSs so that the total payoff will be higher than the case where each SBS individually stores and consumes the energy.

All the symbols that are used in the system model and section III are listed in **Table I**.

3.2.2 Channel Model

The received SINR at the user served by SBS i at time slot t is defined as follows:

$$\gamma_i(t) = \frac{p_i(t) g_{i,i}}{I_i(t)}, \tag{3.3}$$

Table 3.1: List of symbols used for the single-controller stochastic game model

M	Number of SBSs
$\bar{g}_i$	Average channel gain between BS i and its associated user
$\bar{g}_{i,j}$	Average channel gain between BS j and user of BS i
λ_0 (λ_1)	Target SINR for MBS (SBS)
$E(t)$	(Discrete) Battery level of CEQ at time t
ΔT	Duration of one time slot in seconds
$Q(t)$	Number of quanta distributed by the CEQ at time t
$\bar{I}_0(t)$	Average interference at the user served by the MBS at time t
$\bar{I}_i(t)$	Average interference at the user served by SBS i at time t
S	Maximum battery level of the CEQ
$\mathcal{P}$	Finite set of transmit power of the MBS
$\mathbf{m}$, $\mathbf{n}$	Concatenated mixed-strategy vector for the MBS and the CEQ, respectively
$\mathbf{m}(s)$, $\mathbf{n}(s)$	Probability mass function for actions of the MBS and the CEQ, respectively, when $E(t) = s$
$\mathbf{m}(s,p)$	Probability that the MBS chooses power level $p \in \mathcal{P}$ when $E(t) = s$
$\mathbf{n}(s,i)$	Probability that the CEQ sends i quanta when $E(t) = s$
$\varphi(t)$	Energy harvested at time t
β	Discount factor of the stochastic game
U_0, U_1	Utility function of the MBS and the CEQ, respectively
R_0, R_1	Payoff matrix for the MBS and the CEQ, respectively
ϕ_0, ϕ_1	Discounted sum of the value function of the MBS and the CEQ, respectively
π_s	Probability that the CEQ starts with battery level s

where $I_i(t) = \sum_{j \neq i}^{M} p_j g_{i,j} + p_0 g_{i,0}$ is the interference caused by other BSs. $g_{i,0}$ is the channel gain between MBS and the user served by SBS i, $g_{i,i}$ represents the channel gain between SBS i and the user it serves, and $g_{i,j}$ is the channel gain between SBS j and the user served by SBS i. Finally, $p_i(t)$ represents the transmit power of SBS i at time t. The transmit power of MBS $p_0(t)$ belongs to the set $\{p_0^{min}, ..., p_0^{max}\}$. The thermal noise is ignored assuming that it is very small compared to the cross-tier interference. Later, in **Remark 2**, I will show that the interference and transmit power of each SBS can be derived from Q and p_0.

Similarly, the SINR at a macrocell user can be calculated as follows:

$$\gamma_0(t) = \frac{p_0(t) g_{0,0}}{I_0(t) + N_0}, \tag{3.4}$$

where $I_0(t) = \sum_{i=1}^{M} p_i g_{0,i}$ is the interference from other SBSs to the macrocell user, $g_{0,0}$ denotes the channel gain between the MBS and its user it serves, $g_{0,i}$ represents the channel gain between SBS i and macrocell user, and N_0 is the constant thermal noise.

The channel gain $g_{i,j}$ is calculated based on path-loss and fading gain as follows:

$$g_{i,j} = |h|^2 r_{i,j}^{-\alpha}, \tag{3.5}$$

where $r_{i,j}$ is the distance from BS j to user served by BS i, h follows a Rayleigh distribution, and α is the path-loss exponent. It is assumed that the M SBSs are randomly located around the MBS and the users are uniformly distributed within their coverage areas.

3.3 Formulation and Analysis of the Single-Controller Stochastic Game

For the system with one MBS and one CEQ, the source of randomness is the arrivals of energy at CEQ. This means the battery size of CEQ is a random process which implies that I can formulate a two-player non-cooperative stochastic power control game for MBS and CEQ where the state is the battery level at CEQ. The MBS and the SBSs try to maintain the average SINR at their users. Following [18], I define the utility function of the MBS at time t as:

$$U_0(p_0(t), Q(t), t) = -(p_0(t)\bar{g}_0 - \lambda_0(\bar{I}_0(t) + N_0))^2, \tag{3.6}$$

where $\bar{I}_0(t) = \sum_{i=1}^{M} p_i(t)\bar{g}_{0,i}$ is the average interference at the macrocell user at time t, and λ_0 is the target SINR. The MBS wants to achieve the target SINR for its user and thus reduce its outage probability . Note that, the target SINR λ_0 should be chosen such that it is higher than the MBS' SINR outage threshold $\lambda_{outage,0}$. $\bar{g}_{i,j}$ is the average channel gain for the link from BS j to the user served by BS i and $Q(t)$ is the number of energy packets distributed by the CEQ. Similarly, the utility function of the CEQ is defined as follows:

$$U_1(p_0, Q, t) = -\frac{1}{M}\sum_{i=1}^{M}(p_i(t)\bar{g}_i - \lambda_1 \bar{I}_i(t))^2, \tag{3.7}$$

where $\bar{I}_i(t) = \sum_{j \neq i}^{M} p_j(t)\bar{g}_{i,j} + p_0(t)\bar{g}_{i,0}$ is the average interference at the user served by SBS i at time t and $(p_1, p_2, .., p_M)$ satisfy (3.2). By maximizing U_1, each SBS tries to obtain an average SINR at its user close to the target SINR λ_1.

The arguments of both the utility functions demonstrate that the action at time t for the MBS is its *transmit power* $p_0(t)$ while the action of the CEQ is the *number of energy packets* $Q(t)$ *that is used to transmit data from the SBSs.* The conflict in the payoffs of both the players arises from their transmit powers that directly impact the cross-tier interference among transmissions from the BSs.

3.3.1 Formulation of the Game Model

Unlike a traditional power control problem the action space of the CEQ changes at each time and is limited by its battery size. Given the distribution of energy arrival and the discount factor β, the power control problem can be modeled by using a single-controller discounted stochastic game as follows:

- There are two players: one MBS and one CEQ.
- The state of the game is the battery level of the CEQ, which is $\{0, ..., S\}$.
- At time t and state s, the action $p_0(t)$ of the MBS is its transmission power and belongs to the finite set $\mathcal{P} = \{p_0^{min}, ..., p_0^{max}\}$. On the other hand, the action of the CEQ is $Q(t)$, which is the number of energy packets distributed to M SBSs.

$Q(t)$ belongs to the set $\{0, ..., s\}$.

- Let $\mathbf{m}$ and $\mathbf{n}$ denote the concatenated mixed-stationary-strategy vectors of the MBS and the CEQ, respectively. The vector $\mathbf{m}$ is constructed by concatenating $S+1$ sub-vectors into one big vector as $\mathbf{m} = [\mathbf{m}(0), \mathbf{m}(1), ..., \mathbf{m}(S)]$, in which each $\mathbf{m}(s)$ is a vector of probability mass function for the actions of the MBS at state s. For example, if the game is in state s, $\mathbf{m}(s, p)$ gives the probability that the MBS transmits with power p. Therefore, the full form of $\mathbf{m}$ will include the state s and power p. However, to make the formulas simple, in the later parts of the paper, I will use $\mathbf{m}$ or $\mathbf{m}(s)$ to denote, respectively, the whole vector or a sub-vector at state s, respectively.

- Similarly, for the CEQ, $\mathbf{n}(s, i)$ gives the probability that the CEQ distributes i energy packets. Note that the available actions of the CEQ dynamically vary at each state whereas the available actions for the MBS remain unchanged at every state.

- **Pay-offs:** At state s, if the MBS transmits with power p_0 and the CEQ distributes Q energy packets, the payoff function for the MBS is $U_0(p_0, Q)$ while the payoff function for the CEQ is $U_1(p_0, Q)$. I omit t since t does not directly appear in U_1 and U_0.

- **Discounted Pay-offs:** Denote by β the discount factor ($\beta < 1$), then the discounted sum of payoffs of the MBS is given as:

$$\phi_0(s, \mathbf{m}, \mathbf{n}) = \lim_{T \to \infty} \sum_{t=1}^{T} \beta^t \mathbb{E}[U_0(\mathbf{m}, \mathbf{n}, t)], \tag{3.8}$$

where $\mathbb{E}[U_0(\mathbf{m}, \mathbf{n}, t)]$ is the average utility of macrouser at time t if the MBS and the CEQ are using strategy $\mathbf{m}$ and $\mathbf{n}$, respectively. Similarly, I define the discounted sum of payoffs ϕ_1 at the CEQ. In Chapter 2, it was proven that the limit of ϕ_0 and ϕ_1 always exist when $T \to \infty$.

- **Objective:** To find a pair of strategies $(\mathbf{m}^*, \mathbf{n}^*)$ such that ϕ_0 and ϕ_1 become a Nash equilibrium, i.e., $\phi_0(s, \mathbf{m}^*, \mathbf{n}^*) \geq \phi_0(s, \mathbf{m}^*, \mathbf{n}) \quad \forall \mathbf{n} \in \mathcal{N}$ and $\phi_1(s, \mathbf{m}^*, \mathbf{n}^*) \geq \phi_1(s, \mathbf{m}, \mathbf{n}^*) \quad \forall \mathbf{m} \in \mathcal{M}$ where $\mathcal{M}$ and $\mathcal{N}$ are the sets of strategies of MBS and CEQ respectively.

Given the distribution of energy arrival at the CEQ, the transition probability of the system from state s to state s' under action Q $(0 \leq Q \leq s)$ of the CEQ is given as follows:

$$q(s'|s, Q) = \begin{cases} \Pr(\varphi = s' - (s - Q)), & \text{if } s' < S \\ 1 - \sum_{X=0}^{S-s} \Pr(\varphi = X), & \text{otherwise.} \end{cases} \tag{3.9}$$

The states of the game can be described by a Markov chain for which the transition probabilities are defined by (3.9). Clearly, the CEQ controls the state of the game while MBS has no direct influence. Therefore the single-controller stochastic game can be applied to derive the Nash equilibrium strategies for both the MBS and the CEQ. The two main steps to find the Nash equilibrium strategies are:

- First, the payoff matrices for the MBS are built and the CEQ for every state s, where $S \geq s \geq 0$. Denote them by R_0 and R_1, respectively.

- Second, using these matrices, a quadratic programming problem is solved to obtain the Nash equilibrium strategies for both the MBS and the CEQ.

3.3.2 Calculation of the Payoff Matrices

To build R_0 and R_1, U_0 and U_1 are calculated for every possible pair (p_0, Q), where $p_0 \in \mathcal{P}$ and $0 \leq s \leq S$. In this regard, the average channel gain $\bar{g}_{i,j}$ is derived first. Second, given the energy packets Q and transmission power p_0 of the CEQ and the MBS, respectively, the CEQ decides how to distribute this amount of energy Q among the SBSs. Then, I calculate the transmit power at each SBS and obtain U_1 and U_2. The next two remarks provide me with the methods to calculate U_1 and U_2.

Remark 1. *Given two BSs A and B, assume that a user D, who is associated with B, is uniformly located within the circle centred at B with radius r (Fig. 3.2). Assume that A does not lie on the circumference of the circle centred at B and $\alpha = 4$. Denote $AB = R$ and $AD = d$, then the expected value of d^{-4}, i.e., $\mathbb{E}[d^{-4}]$ is $\frac{1}{(R^2-r^2)^2}$. If $A \equiv B$, then $\mathbb{E}[d^{-4}] = \frac{1-r^{-2}}{r^2}$ given that $r \geq BD \geq 1$.*

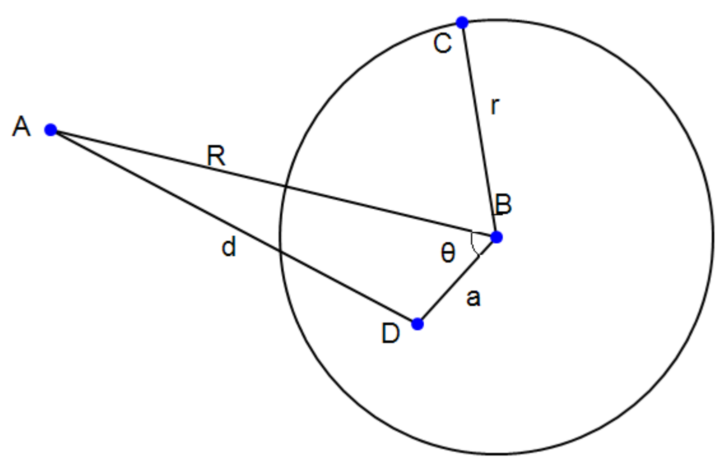

Figure 3.2: Graphical illustration of the two BSs A and B and the user D located within the disk centred at B.

Proof. Denote the distance between the BS B to its user D as $BD = a$. If D is uniformly located inside the disk centred at B, the probability density function of

BD is $f_D(BD = a) = \frac{2a}{r^2}$. Moreover, denote by θ the value of the angle $\angle ABD$. Clearly, θ is uniformly distributed between $(0, 2\pi)$. Now, using the cosine formula, $d^2 = R^2 + a^2 - 2aR\cos(\theta)$,

$$\mathbb{E}[d^{-4}] = \int_0^{2\pi} \int_0^r (R^2 + a^2 - 2aR\cos\theta)^{-2} \frac{1}{2\pi} \frac{2a}{r^2} \, da \, d\theta. \tag{3.10}$$

First, the indefinite integral over θ is solved as $\int (R^2 + a^2 - 2aR\cos\theta)^{-2} d\theta = f_1(a, \theta) + f_2(a, \theta) + L$, where L is a constant and

$$\begin{aligned} f_1(a,\theta) =& \frac{2(R^2+a^2)}{(R^2-a^2)^3} \arctan \frac{(R+a)\tan\frac{\theta}{2}}{R-a}, \quad \text{and} \\ f_2(a,\theta) =& \frac{2aR\sin\theta(R^2+a^2-2aR\cos\theta)}{(R^2-a^2)^2}. \end{aligned}$$

These closed-form equations can be verified by manually taking the derivative of the left hand side. Since $\sin 0 = \sin 2\pi = 0$, after integrating f_2 over $[0, 2\pi]$, I can ignore it. Also, notice that arctan is a multi-valued function so I have to keep track of which branch I am on. The result is

$$\int_0^{2\pi} (R^2 + a^2 - 2aR\cos\theta)^{-2} d\theta = \pi \frac{2(R^2+a^2)}{(R^2-a^2)^3}.$$

The indefinite integral of the above result is:

$$\frac{1}{r^2} \int a \frac{2(R^2+a^2)}{(R^2-a^2)^3} da = \frac{1}{r^2} \frac{a^2}{(R^2-a^2)^2} + L. \tag{3.11}$$

Applying the upper and lower limits of a, I complete the proof. $\square$

Recalling that $g_{i,j} = |h|^2 r_{ij}^{-4}$ and that the fading and path-loss are independent, I have $\bar{g}_{i,j} = \mathbb{E}[h^2]\mathbb{E}[r_{ij}^{-4}]$. Assuming that the fading gain follows Rayleigh distribution with scale parameter λ, then h^2 is an exponential random variable with mean λ. The expected value $\mathbb{E}[r_{ij}^{-4}]$ can then be calculated using remark above. For other values of α, $\mathbb{E}[r_{ij}^{-\alpha}]$ can be computed numerically using tools such as `MATHEMATICA`.

Next I need to find how the CEQ distribute its energy to each SBS such that U_1 is maximized.

Remark 2. *(Optimality in energy distribution at the CEQ) If at time t, the CEQ distributes Q energy packets to M SBSs and the MBS transmits with power p_0, then the transmit powers $(p_1, p_2, ..., p_M)$ at the M SBSs are the solutions of the following optimization problem (t is omitted for brevity):*

$$\begin{aligned} \max_{p_1,p_2,...,p_M} \; & -\frac{1}{M}\sum_{i=1}^{M}\left(p_i\bar{g}_i - \lambda_1(\sum_{j\neq i}^{M} p_j\bar{g}_{ij} + p_0\bar{g}_{i,0})\right)^2, \\ s.t. \; & \sum_{i=1}^{M} p_i = \frac{K}{\Delta T}Q, \\ & P_{max} \geq p_i \geq 0, \quad \forall i = 1, 2, ..., M, \end{aligned} \tag{3.12}$$

where P_{max} is the maximum transmit power of each SBS. Since this problem is strictly concave, the solution $(p_1, ..., p_M)$ always exists and is unique for each pair (Q, p_0). Thus, for each pair (Q, p_0), where $Q \in \{0, ..., S\}$ and $p_0 \in \{p_0^{min}, ..., p_0^{max}\}$, I have unique values for $U_0(p_0, Q)$ and $U_1(p_0, Q)$.

Based on the remarks above, for each combination of Q and p_0, I can find the unique payoff U_0 and U_1 of MBS and CEQ. Since Q and p_0 belongs to discrete sets I can find the payoff for all of the possible combinations between them. Thus, I can build the payoff matrix R_0 for the MBS and R_1 for the CEQ. The matrix R_0 has the form of a block-diagonal matrix $\text{diag}(R_0^0, ..., R_0^S)$, where each sub-matrix

$R_0^s = (U_0(p_0, j))^{\mathcal{P}\times\{0,...,s\}}$, with $p_0 \in \mathcal{P}$ and $j \in \{0, ..., s\}$ is the matrix of all possible payoffs for the MBS at state s. Similarly, I can build R_1, which is the payoff matrix for the CEQ. A detailed explanation on how I use them will be given in the next subsection.

3.3.3 Derivation of the Nash Equilibrium

If the strategy $\mathbf{m}_0$ of the MBS is known, the discount factor β, and the probability π_s that the CEQ starts with s energy packets in the battery, then the stochastic game is reduced to a simple MDP problem with only one player, the CEQ. For this case, denote the CEQ's best response strategy to $\mathbf{m}_0$ by $\mathbf{n}$. From Chapter 2, the CEQ's value function $\phi_1(s, \mathbf{m}_0, \mathbf{n})$, where $s = 0, ..., S$, is the solution of the following MDP problem:

$$\begin{aligned} &\min_{\phi_1} \sum_{s=0}^{S} \pi_s \phi_1(s, \mathbf{m}_0, \mathbf{n}), \\ \text{s.t.} \quad &\phi_1(s, \mathbf{m}_0,\mathbf{n}) \geq r_1(s, \mathbf{m}_0, j) + \beta \sum_{s'=0}^{S} q(s'|s, j)\phi_1(s', \mathbf{m}_0, \mathbf{n}), \\ &\forall s, j, \quad 0 \leq j \leq s \text{ and } 0 \leq s \leq S \end{aligned} \tag{3.13}$$

with $r_1(s, \mathbf{m}_0, j) = \sum_{p_0\in\mathcal{P}} U_1(p_0, j)\mathbf{m}_0(s, p_0)$, where $\mathcal{P}$ is the set of transmit power levels of the MBS. This $r_1(s, \mathbf{m}_0, j)$ is the average payoff for the CEQ at state s when it consumes j quanta of energy. Using the Dirac function δ, the dual problem can be expressed as

$$
\begin{aligned}
&\max_x \sum_{s=0}^{S}\sum_{j=0}^{s} r_1(s,\mathbf{m}_0,j)x_{s,j},\\
\text{s.t.}\quad &\sum_{s=0}^{S}\sum_{j=0}^{s}[\delta(s-s')-\beta q(s'|s,j)]x_{s,j}=\pi_{s'},\quad \forall 0\le s'\le S,\\
&x_{s,j}\ge 0\ \ \forall s,j,\quad 0\le j\le s \ \text{ and } \ 0\le s\le S
\end{aligned}
\tag{3.14}
$$

where $\delta(s)=1$ if $s=0$ and $\delta(s)=0$ otherwise.

By solving the pair of linear programs above, the probability that the SBS chooses action j at state s can be found as $\mathbf{n}(s,j)=\frac{x_{s,j}}{\sum_{j=0}^{s}x_{s,j}}$. Using some algebraic manipulations, the optimization problem in (3.13) can be converted into a matrix form as:

$$
\begin{aligned}
\min_{\phi_1}\ \ &\pi^{\mathrm{T}}\phi_1,\\
\text{s.t.}\ \ &H\phi_1\ge R_1^{\mathrm{T}}\mathbf{m}_0,
\end{aligned}
\tag{P}
$$

and its dual as

$$
\begin{aligned}
\max_{\mathbf{x}}\ \ &\mathbf{m}_0^{\mathrm{T}}R_1\mathbf{x},\\
\text{s.t.}\ \ &\mathbf{x}^{\mathrm{T}}H=\pi^{\mathrm{T}},\\
&\mathbf{x}\ge 0,
\end{aligned}
\tag{D}
$$

where R_1 is the payoff matrices of the CEQ. Combining the primal and dual linear programs (i.e., (P) and (D) above) and using the same notations, I have the following theorems.

Theorem 3.3.1 (Nash equilibrium strategies [28])**.** *If the state space and the action* *space are finite and discrete, and the transition probabilities are controlled only by player 2 (i.e., the CEQ), then there always exists a Nash equilibrium point* $(\mathbf{m}, \mathbf{n})$ *for this stochastic game. Moreover, a pair* $(\mathbf{m}, \mathbf{n})$ *is a Nash equilibrium point of a general-sum single-controller discounted stochastic game if and only if it is an optimal solution of a (bilinear) quadratic program given by*

$$\begin{aligned}
\max_{\mathbf{m},\mathbf{x},\phi,\xi} \quad & [\mathbf{m}(R_0 + R_1)\mathbf{x} - \pi^T\phi_1 - \mathbf{1}^T\xi], \\
s.\ t. \quad & H\phi_1 \geq R_1^T\mathbf{m}, \\
& \mathbf{x}^T H = \pi^T, \\
& R_0^s\mathbf{x}(s) \leq \xi_s\mathbf{1}, \quad \forall s = 0, ..., S, \\
& \mathbf{m}(s)^T\mathbf{1} = 1, \quad \forall s = 0, ..., S, \\
& \mathbf{m}, \mathbf{x} \geq \mathbf{0},
\end{aligned} \tag{3.15}$$

where ξ_s *is the maximum average payoff of the MBS at state* s*. The sub-vector strategy* $\mathbf{n}(s)$ *of the CEQ at state* s *is calculated from* $\mathbf{x}$ *as:*

$$\mathbf{n}(s) = \frac{\mathbf{x}(s)}{\mathbf{x}(s)^T\mathbf{1}}. \tag{3.16}$$

I can define different utility functions for the MBS and the SBS and apply the same method to achieve the Nash equilibrium. As long as the number of states is finite and the transition and the payoff matrices remain unchanged over time, a Nash equilibrium point always exists.

Theorem 3.3.2 (Best response strategy for the MBS). *For any given stationary strategy* $\mathbf{n}$ *of the CEQ, there exists a pure stationary strategy* $\mathbf{m}$ *as the best response for the MBS. Similarly, for any stationary strategy* $\mathbf{m}$ *of the MBS, there exists a pure stationary best response* $\mathbf{n}$ *of the CEQ.*

Proof. First it is proved that given $\mathbf{n}$, there exists a pure stationary strategy $\mathbf{m}$ which is the best response of the MBS against $\mathbf{n}$. Since the action set of the MBS is fixed, at each state s, given strategy $\mathbf{n}(s)$ of the CEQ, the MBS just needs to choose a mixed stationary strategy $\mathbf{m}(s)$ such that its average payoff is maximized.

Denote by $\mathcal{P}$ the set of power levels available at the MBS. At state s, the expected value of the utility function for the MBS is

$$\mathbb{E}[U_1] = \sum_{p_0^s \in \mathcal{P}} \sum_{j=0}^{s} -(p_0^s \bar{g}_0 - \lambda_0 \bar{I}(p_0^s, j))^2 \mathbf{m}(s, p_0^s)\mathbf{n}(s, j), \tag{3.17}$$

where p_0^s and $j \in \{0, ..., s\}$ are the transmit power of the MBS and the number of energy packets distributed at the CEQ at state s, respectively. $\bar{I}(p_0^s, j)$ is the average interference from other SBSs to the MBS if the CEQ distributes j energy packets and the MBS transmits with power p_0^s. I have

$$\bar{I}(p_0^s, j) = \sum_{i=1}^{M} p_i \bar{g}_{0,i} + N,$$

where $(p_1, p_2, ..., p_M)$ is the solution of **Remark 2**, with Q and p_0 replaced by j and p_0^s, respectively. Since $\sum_{p_0^s \in \mathcal{P}} \mathbf{m}(s, p_0^s) = 1$ and $\mathbf{m}_s$ is a non-negative vector,

$$\mathbb{E}[U_1] \leq \max_{p_0^s \in \mathcal{P}} \left\{ -\sum_{j=0}^{s} (p_0^s \bar{g}_0 - \lambda_0 \bar{I}(p_0^s, j))^2 \mathbf{n}(s, j) \right\}. \tag{3.18}$$

Since the set $\mathcal{P}$ is fixed and finite, there always exists at least one value of p_0^s that achieves the maximum for the right hand side. That means when the game in state s, the MBS can choose this power level with probability of 1.

It is difficult to obtain p_0^s in closed-form, because first I need to find $(p_1, ..., p_M)$ in closed-form by solving (3.12). However, if the average channel gains from each

SBS to the macrocell user are the same, I can obtain p_0^s in closed form. In this case, denote by $\bar{g}_{0,SBS}$ the channel gain from an SBS to the macrocell user, then

$$\begin{aligned}\bar{I}(p_0^s, j) &= \sum_{i=1}^{M} p_i \bar{g}_{0,SBS} + N_0 = \bar{g}_{0,SBS} \sum_{i=1}^{M} p_i + N_0 \\ &= \frac{K}{\Delta T} j \bar{g}_{0,SBS} + N_0.\end{aligned}$$

The final equality is from (3.2). Replacing this result back into (3.18),

$$\mathbb{E}[U_1] \leq \max_{p_0^s \in \mathcal{P}} - \sum_{j=0}^{s} (p_0^s \bar{g}_0 - \lambda_0 (\frac{K}{\Delta T} j \bar{g}_{0,SBS} + N_0))^2 \mathbf{n}(s, j). \tag{3.19}$$

The right hand side of this inequality is a strictly concave function (specifically it is a downward parabola) with respect to (w.r.t.) p_0^s. Note that $\sum_{j=0}^{s} \mathbf{n}(s, j) = 1$. The parabola will achieve the maximum value at its vertex given by

$$p_0^{s*} = \frac{\lambda_0 \sum_{j=0}^{s} \left(\frac{K}{\Delta T} \bar{g}_{0,SBS} j + N_0 \right) \mathbf{n}(s, j)}{\bar{g}_0}. \tag{3.20}$$

If p_0^{s*} is not available in $\mathcal{P}$, since the right hand side of the inequality above is a parabola w.r.t. p_0^s, the best response p_0^s to $\mathbf{n}(s)$ is the one nearest to the vertex p_0^{s*}.

On the other hand, given strategy $\mathbf{m}$ of the MBS, the problem of finding the best response strategy $\mathbf{n}$ for the CEQ is simplified into a simple MDP in (3.13). Then, there always exists a pure stationary strategy $\mathbf{n}$ [27, Chapter 2]. This completes the proof. $\square$

Because for any mixed strategy of CEQ, MBS can find a pure stationary strategy as a best response, I only need to find Nash equilibrium where the strategy of MBS is deterministic. This problem can be converted to a mixed-integer program with $\mathbf{m}$ as a vector of 0 and 1. I can use a brute-force search to obtain an equilibrium point. For each feasible integer value of $\mathbf{m}$ I insert it into (3.15) to obtain $\mathbf{n}$. If the objective is zero, then $(\mathbf{m}, \mathbf{n})$ is the equilibrium point. This theorem implies that the optimization problem in (3.15) can be solved in a finite amount of time. .

Although there exists a Nash equilibrium point, the uniqueness is not guaranteed. To make the Nash equilibrium point more meaningful, the following lemma from [28] is used.

Lemma 3.3.1. *(Necessary and sufficient conditions for the Nash equilibrium)* $\mathbf{m}$ *and* $\mathbf{n}$ *constitute a pair of Nash equilibrium policies for the MBS and the CEQ if and only if*

$$\mathbf{m}(R_0 + R_1)\mathbf{x} - \pi^T \phi_1 - \mathbf{1}^T \xi = 0. \tag{3.21}$$

Since π_s is the probability that the CEQ starts with s energy level in the battery at starting time, from (3.13), $\pi^{\mathrm{T}} \phi_1$ is the average value function of the CEQ with respect to the energy arrival rate and the strategies $\mathbf{m}, \mathbf{n}$. Using the lemma above, the problem in (3.15) is changed to a quadratic-constrained quadratic programming (QCQP) as stated below.

Proposition 1. *(Nash equilibriums that favor the SBSs) The Nash equilibrium* $(\mathbf{m}, \mathbf{n})$ *that has the best payoff for the CEQ is a solution of the following QCQP problem:*

$$\begin{aligned} &\max_{\mathbf{m},\mathbf{x},\phi,\xi} \pi^T \phi_1, \\ s.t. \quad &\mathbf{m}(R_0 + R_1)\mathbf{x} - \pi^T \phi_1 - \mathbf{1}^T \xi = 0, \\ &\textit{all constraints from (3.15).} \end{aligned} \tag{3.22}$$

By solving this QCQP, I obtain a Nash equilibrium that returns the best average payoff for the CEQ. This bias toward the SBSs is crucial as the available energy of the CEQ is limited by the randomness of the environment and thus the SBSs are more likely to suffer when compared to the MBS. Again, I can still have multiple Nash equilibrium in this case, but all of them must return the same payoff for CEQ. Because there may be multiple solutions, CEQ and MBS need to exchange information so that they agree on the same Nash equilibrium.

Algorithm 1 Nash equilibrium for the stochastic game

1: The MBS and the CEQ build their reward matrices R_0 and R_1. For each possible pair of energy level and transmit power (Q, p_0), the CEQ solves (3.2) to obtain a unique tuple $(p_1, p_2, ..., p_M)$ and record these results.
2: The MBS and the CEQ calculate their strategy $\mathbf{m}$ and $\mathbf{n}$, respectively, by solving (3.22).
3: At time t, the CEQ sends its current battery level s to the MBS. It also randomly chooses an action Q using the probability vector $\mathbf{n}(s)$.
4: The MBS then randomly picks power p_0 using distribution $\mathbf{m}(s)$ and sends it back to the CEQ. Based on p_0 and Q, the CEQ searches its records and retrieves the corresponding tuple $(p_1, ..., p_M)$.
5: The CEQ distributes energy $(p_1\Delta T, p_2\Delta T, ..., p_M\Delta T)$, respectively, to the M SBSs.

From Theorem 2, I know that there exist an equilibrium with pure stationary strategies for both MBS and CEQ. Recall that with pure strategy, the action of each player is a function of the state. Thus, if I can obtain this equilibrium, CEQ can predict which transmit power p_0, the MBS will use based on the current state without exchanging information with MBS and vice versa.

3.3.4 Implementation of the Discrete Stochastic Game

For discrete stochastic control game with CEQ, each SBS first needs to send its location and average fading channel information $\mathbb{E}[h^2]$ of its user to the CEQ and the MBS so that complete channele state information is known at both the MBS and CEQ. Since I only use average channel gain, the CEQ and the MBS only need to re-calculate the Nash equilibrium strategies when either the locations of SBSs change, e.g., some SBSs go off and some are turned on, or when the average of channel fading gain h changes, or when distribution of energy arrival $\varphi(t)$ at the CEQ is changes. Thanks to the central design, SBSs and MBS only need to send information about their channel gains to the CEQ thus create less communication overhead compared to a fully distributed system.

3.4 Simulation Results and Discussions

3.4.1 Single-Controller Stochastic Game

In this section, the efficacy of the developed stochastic policy is quantified in comparison to the greedy power control policy. The stochastic policy is obtained from the QCQP problem. On the other hand, in the greedy policy, I follow a hierarchical method. First, the MBS chooses its transmit power, then each SBS tries to transmit with the power such that the SINR at its user is nearest to its target value, ignoring the co-interference from other SBSs. Next, the MBS records its current interference and then chooses a new transmit power to achieve the target SINR at its user and so on.

To solve the QCQP in (3.22), the *fmincon* function from Matlab is used. In the simulations, it is assumed that the CEQ has a maximum battery size of $S = 21$ and the volume of one energy packet is $K = 25\mu$J. The duration of one time interval is $\Delta T = 5$ ms and the thermal noise is $N_0 = 10^{-8}$ W. The MBS has two levels of transmission power [10; 20] W and the SINR outage threshold is set to 5. It is assumed that the energy arrival at each SBS follows a Poisson distribution with unit rate. Also, the volume of each energy packet arriving at the CEQ is assumed to be C times larger than that for one energy packet collected by each SBS. Thus, the amount of energy in each packet at the CEQ will be CK μJ. This shows that the CEQ should have a more efficient method to harvest energy than each SBS (in the case of greedy method). However, as the maximum battery size of the CEQ is S, its total available energy is always limited by the product SCK μJ regardless of M. For both the cases, each SBS can receive up to 150 μJ of energy from either the CEQ or the environment. In the simulations, I investigate the impact of choosing different target SINRs, number of SBSs in the macrocell, and the parameter C of the CEQ against the outage probability of small cell and macrocell users. Also, I set the minimum threshold for outage for small cell and macrocell users to $\lambda_{outage,1} = 0.001$ and $\lambda_{outage,0} = 5$, respectively. It is assumed that at the beginning, the CEQ has full battery.

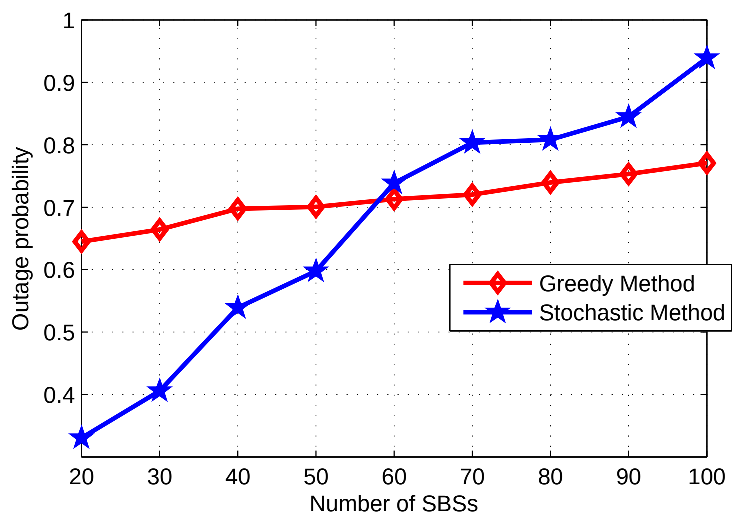

Figure 3.3: Outage probability of a small cell user with different number of SBSs when $S = 21$ states, $C = 40$, $\lambda_1 = 0.002$, $\lambda_0 = 10$.

From Fig. 3.3, it can be seen that when the number of SBSs is smaller than some value, the stochastic method achieves better results. That is because, the CEQ can share energy among the SBSs, and also the QCQP in (3.22) gives a Nash equilibrium that favors the CEQ. However, at some point, the greedy method will provide better

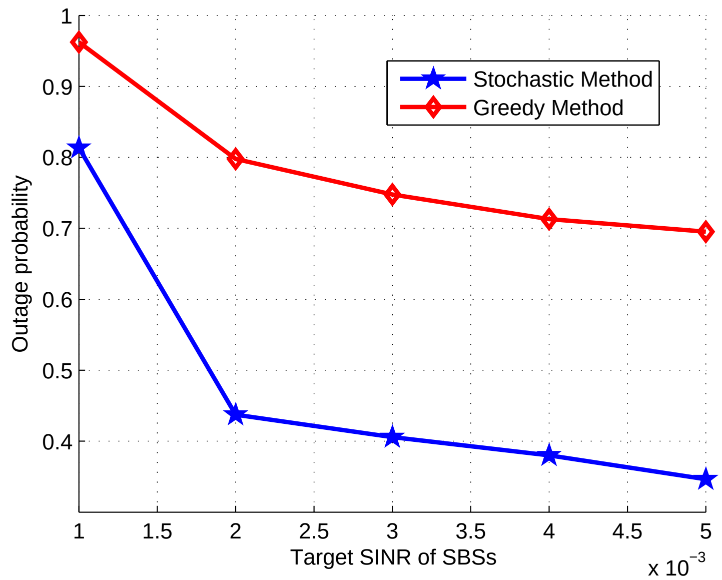

Figure 3.4: Outage probability of a small cell user with different target SINR when $S = 21$ states, $C = 40$, $M = 30$ SBSs, $\lambda_0 = 10$.

results. This is not surprising since the CEQ can only store at most $S \times K \times C$ μJ of energy. Therefore, when the number of SBSs increases, the allocated power per SBS by the CEQ reduces to zero while the greedy method allows each SBS to harvest up to 150 μJ no matter how large M is. This means the greedy method will provide a better performance compared to using the CEQ when M is large.

Following Fig. 3.4, by increasing the threshold SINR target λ_1, the outage probability of a user served by an SBS can be reduced. This is understandable since the average SINR will increase to approach the higher target and thus reduce the outage probability. However, for both the greedy and stochastic methods, the slope is nearly flat when the target SINR is larger than some value. This is because, to increase the average SINR, the SBSs need to transmit with higher power to mitigate cross-tier interference. However, since the battery capacity is limited for the CEQ and each of the SBSs, a higher transmit power means a higher consumption of harvested energy, which can cause a shortage of energy later. Thus at some point increasing the target SINR does not bring any benefits.

Fig. 3.5 shows the outage probability when increasing the quanta volume by choosing a higher multiplier C for the CEQ. It is easy to see that, with a higher C, i.e., choosing a more effective method to harvest energy at the CEQ, I can achieve a better performance. The greedy method does not use the CEQ, so the outage probability remains unchanged. Note that, since the battery size of each SBS is limited to 150 μJ, at some point, a higher C does not improve the outage probability.

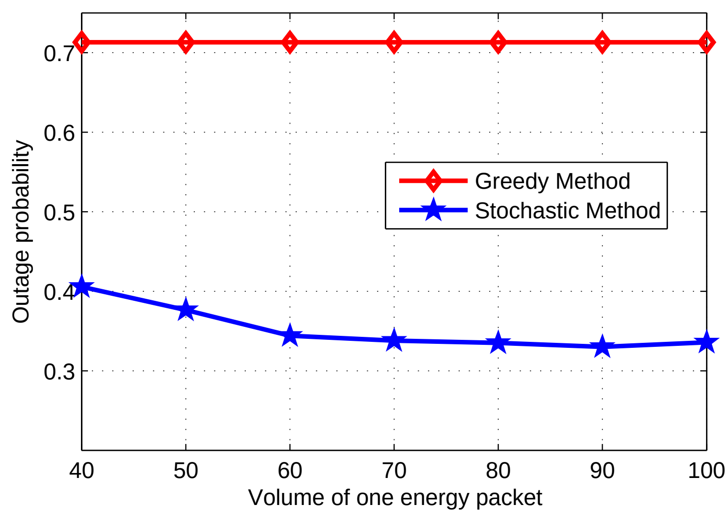

Figure 3.5: Outage probability of a small cell user with different quanta volume when $S = 21$ states, $M = 30$ SBSs, $\lambda_1 = 0.002$, $\lambda_0 = 10$.

Fig. 3.6 shows the outage probability of the macrocell user when $M = 80$ and $C = 50$. The stochastic method gives better results in this case since the SBSs are more “rational” in choosing their transmit powers. Also, unlike the greedy method, the CEQ has a fixed-energy battery, so when M is large, the average amount of energy distributed to an SBS will be small, which in turn limits the cross-interference to the MBS. With the greedy method, the SBSs use higher transmit power to compete against the MBS; therefore, it creates a larger cross-interference and in turn increases

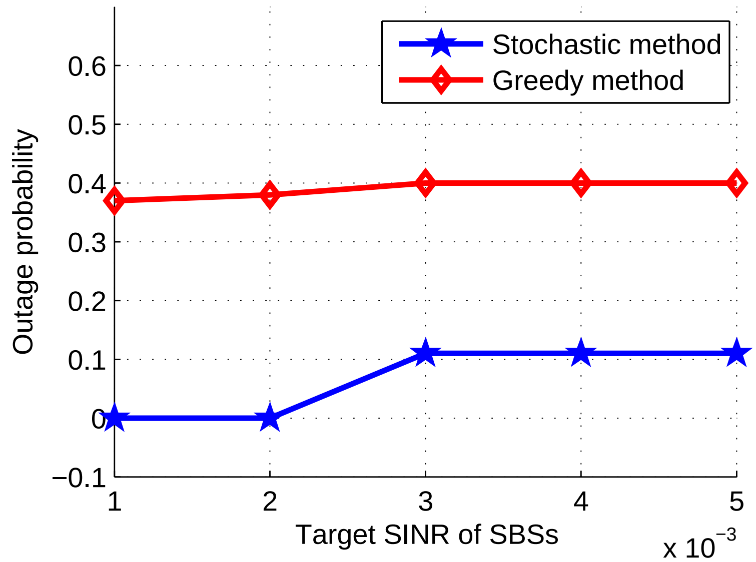

Figure 3.6: Outage probability of a macrocell user when $S = 21$ states, $M = 80$ SBSs, $C = 50$, $\lambda_0 = 10$.

the outage probability of MBS.

In summary, the centralized method using a CEQ can provide a better performance in terms of outage probability for both the MBS and the SBSs. However, since the CEQ has a fixed battery size, the centralized method performs poorer when it needs to support a large number of SBSs. To improve this inflexibility, other parameters can be adjusted as follows: change target SINR, increase the multiplier C, or increase the battery size of each SBS.

Chapter 4

Power Control in Energy-Harvesting Small Cells Using Mean Field Game

4.1 Introduction

The main problem of the two-player single-controller stochastic game is the "curse of dimensions". The time complexity of **Algorithm 1** increases exponentially with the number of states S or the maximum battery size. Note that R_0 and R_1 have dimensions of $|\mathcal{P}| \times S(S+1)/2$, so the complexity increases proportionally to S. Moreover, unlike other optimization problems, I am unable to relax the QCQP in (3.22), because the Nash equilibrium can only be obtained at the global optimal. Last but not the least, the centralized method requires information and coordination from all the SBSs within the macrocell area, which can be challenging when M is large. To tackle these problems, the stochastic game model is extended to an MFG model for very large number of players.

As discussed in Chapter 2, the main idea of an MFG is the assumption of similarity, i.e., all players are identical and follow the same strategy. They can only be differentiated by their "state" vectors. If the number of players is very large, I can assume that the effect of a specific player to other players is nearly negligible. In my energy harvesting game, the state is the battery E and the mean field $m(t, E)$ is the probability distribution of energy in the area I am considering. When the number of players M is very large, I can assume that $m(t, E)$ is a smooth continuous distribution function. As usually assumed in a MFG, the SBS does not care about others' states but only act according to a "mean field" $m(t, \mathbf{s})$. Since SBSs affect each other directly through its transmit power, I will show that the average interference received at a SBS as a function of the mean field m.

All the symbols that are used in this section are listed in **Table II**.

Table 4.1: List of symbols used for the MFG model

$\bar{g}$	Average channel gain from a generic SBS to another user
E	(Continuous) Battery level of an SBS
R	Energy coefficient $e^R = E$
W_t	Wiener process at time t
$p(t, E)$	Transmit power at a generic SBS as a function of E
$p(t, R)$	Transmit power at a generic SBS as a function of R
$m(t, E)$	Probability distribution of energy E at time t
$m(t, R)$	Probability distribution of energy coefficient R at time t
$\bar{p}(t)$	Average transmit power of a SBS at time t
σ	Intensity of energy arrival or loss.

4.2 Formulation of the MFG

Denote by $E(v)$ the available energy in the battery of an SBS at time v. Given the transmission strategies of other SBSs, each SBS will try to maximize its long run generic utility value function by solving the following optimal control problem:

$$\min_{p} U(0, E(0)) = \mathbb{E}\left[\int_0^T (p(v, E(v))g - \lambda(I(v) + N_0))^2 dv\right], \tag{4.1}$$

$$\text{s.t.} \quad dE(v) = -p(v, E(v))dv + \sigma dW_v, \tag{4.2}$$

$$E(v) \geq 0, \quad p(v) \geq 0, \tag{4.3}$$

where $I(v)$ is the generic interference at a user served by an SBS at time v and g is the channel gain between a generic SBS and its user. The mean field $m(v, E)$ is the probability distribution of energy E in the area at time v. Using M as the number of SBSs in a macrocell and assuming that the other SBSs have the same average channel gain $\bar{g}$ to the user of the current generic SBS, then, the average interference $I(v)$ at the user served by a generic SBS can be expressed as $I(v) = M\bar{g}\bar{p}(v)$, where $\bar{p}(v) = \int_0^\infty p(v, E)m(v, E)dE$ can be understood as the average transmit power of "another" generic SBS. Since the MFG assumes similarity, $\bar{p}(v)$ can be considered as the average transmit power of a generic SBS at time v. To make the notation simpler, I denote $\bar{\lambda} = \lambda\bar{g}M$.

Thanks to similarity, all the SBSs have the same set of equations and constraints, so the optimal control problem for the M SBSs reduces to finding the optimal policy for only one generic SBS. Mathematically, if an SBS has infinite available energy, i.e., $E(0) = \infty$, it will act as an MBS. However, for simplicity, I will assume that only the SBSs are involved in the game and the interference from the MBS is constant, which is included in the noise N_0 as in [32]. Except that, the system model and the optimization problem here are similar to the case with the discrete stochastic game model.

Assuming that the SBSs are uniformly distributed within the macrocell with radius

r centered at the MBS, the average interference from the MBS to a generic user served by an SBS can be easily derived by using a method similar to that described in **Remark 1**. For the MFG model, the energy level E is a continuous non-negative variable. The equality in (4.2) shows the evolution of the battery, where σ is a constant which is proportional to the maximum energy arrival during a time interval. W_v is a Wiener process, thus $dW_v = \epsilon_v dv$, where ϵ_v is a Gaussian random variable with mean zero and variance 1. This model of evolution for battery energy was mentioned in [29]. The inflexibility of the energy arrival is the main disadvantage of using the MFG model compared to the discrete stochastic game model. The random arrival of energy is configured as "noise", so this can be either positive or negative. I consider the negative part as the battery leakage and internal energy consumption. The final inequalities are the causality constraints: The battery state $E(v)$ and transmit power must always be non-negative. To guarantee this positivity I follow [30] and change the energy variable $E(v)$ to $E(v) = e^{R(v)}$. This conversion is a bijection map from $E(v)$ to $R(v)$, thus I can write $m(v, E) = m(v, R)$ and $p(v, E) = p(v, R)$, where $\infty > R > -\infty$. The new optimal control problem can be rewritten as

$$\min_{p(.)} U(0, R(0)) = \mathbb{E}\left[\int_0^T (p(v, R(v))g - \bar{\lambda}\bar{p}(v) - \lambda N_0)^2 dv\right], \tag{4.4}$$

$$\text{s.t.} \quad dR(v) = -p(v, R(v))e^{-R(v)}dv + \sigma e^{-R(v)}dW_v, \tag{4.5}$$

$$p(v) \geq 0. \tag{4.6}$$

To obtain the power policy p, the following steps are followed:

- First, the Forward-Backward differential equations are derived from the above optimal control problem.
- Second, Finite Difference method is applied to numerically solve these equations.

4.2.1 Forward-Backward Equations of MFG

Assuming that the optimal control above starts at time t with $T \geq t \geq 0$, the Bellman function $U(t, R)$ is obtained as

$$U(t, R(t)) = \mathbb{E}\left[\int_t^T (p(v, R(v))g - \bar{\lambda}\bar{p}(v) - \lambda N_0)^2 dv\right]. \tag{4.7}$$

From this function, at time t, I obtain the following Hamilton-Jacobi-Bellman (HJB) [30] equation as:

$$\partial_t U + \min_{p \geq 0}\left\{\left(p(t, R)g - \bar{\lambda}\bar{p}(t) - \lambda N_0\right)^2 - p(t, R)e^{-R}\partial_R U(t, R)\right\} + \frac{\sigma^2}{2}e^{-2R}\partial^2_{RR}U = 0 \tag{4.8}$$

The Hamiltonian $\min_{p \geq 0}\left\{\left(p(t, R)g - \bar{\lambda}\bar{p}(t) - \lambda N_0\right)^2 - p(t, R)e^{-R}\partial_R U(t, R)\right\}$ is given by the Bellman's principle of optimality. By applying the first order necessary condition, I obtain the optimal power control as follows:

$$p^*(t, R) = \left[\frac{\bar{\lambda}\bar{p}(t) + \lambda N_0}{g} + \frac{e^{-R}\partial_R U}{2g^2}\right]^+. \tag{4.9}$$

Remark 3. *The Bellman U, if exists, is a non-increasing function of time and energy. Therefore, I have $\partial_R U \leq 0$ and $\partial_t U \leq 0$.*

From equation (4.9), given the current interference $\bar{\lambda}\bar{p}(t) + \lambda N_0$ at a user, the corresponding SBS will transmit less power based on the future prospect $\frac{e^{-R}\partial_R U}{2g^2}$. If the future prospect is too small, i.e., $\frac{e^{-R}\partial_R U}{2g^2} < -\frac{\lambda\bar{p}(t)+\lambda N_0}{g}$, it stops transmission to save energy.

Replacing p^* back to the HJB equation,

$$\partial_t U + \frac{\sigma^2}{2}e^{-2R}\partial^2_{RR}U + (\bar{\lambda}\bar{p}(t) + \lambda N)^2 - \left(\left[\bar{\lambda}\bar{p}(t) + \lambda N + \frac{e^{-R}\partial_R U}{2g}\right]^+\right)^2 = 0, \tag{4.10}$$

which has a simpler form as follows:

$$\partial_t U + \frac{\sigma^2}{2} e^{-2R} \partial^2_{RR} U = (pg)^2 - (\bar{\lambda}\bar{p}(t) + \lambda N)^2. \tag{4.11}$$

Also, from (4.5), at time t, I have the Fokker-Planck equation [31] as:

$$\partial_t m(t, R) = \partial_R(pe^{-R}m) + \frac{\sigma^2}{2}\partial_{RR}(me^{-2R}), \tag{4.12}$$

where $m(t, R)$ is the probability density function of R at time t.

Combining all these information, I have the following proposition.

Proposition 2. *The value function and the mean field (U, m) of the MFG defined in (4.4) is the solution of the following partial differential equations:*

$$\partial_t U + \frac{\sigma^2}{2} e^{-2R} \partial^2_{RR} U = (pg)^2 - (\bar{\lambda}\bar{p}(t) + \lambda N_0)^2, \tag{4.13}$$

$$p(t, R) = \left[\frac{\bar{\lambda}\bar{p}(t) + \lambda N_0}{g} + \frac{e^{-R}\partial_R U}{2g^2}\right]^+, \tag{4.14}$$

$$\partial_t m(t, R) = \partial_R(pe^{-R}m) + \frac{\sigma^2}{2}\partial^2_{RR}(e^{-2R}m), \tag{4.15}$$

$$\bar{p}(t) = \int_{-\infty}^{\infty} e^R p(t, R) m(t, R) dR, \tag{4.16}$$

$$\int_{-\infty}^{\infty} m(t, R) dR = 1, \quad \textit{where} \quad m(t, R) \geq 0. \tag{4.17}$$

Lemma 4.2.1. *The average transmit power $\bar{p}(t)$ of a generic SBS is a derivative with respect to time of the average energy available and can be calculated as*

$$\bar{p}(t) = -\frac{d}{dt}\int_{-\infty}^{\infty} e^{2R} m(t, R) dR. \tag{4.18}$$

Proof. Using from the stochastic differential equation in (4.2) at time t, $dE(t) =$

$-p(t, E(t))dt + \sigma dW_t$, I get the integral form as follows:

$$\begin{aligned} E(t+t') - E(t) &= -\int_t^{t+t'} p(v, E(v))dv + \int_t^{t+t'} \sigma dW_v \\ &= -p(\bar{t}, E(\bar{t}))t' + \sigma(W_{t+t'} - W_t), \end{aligned} \tag{4.19}$$

where $\bar{t} \in (t, t+t')$. I obtain the second equality using the mean value theorem for integrals: If $G(x)$ is a continuous function and $f(x)$ is integrable function that does not change sign on the interval $[a, b]$, then there exists $x \in [a, b]$ such that $\int_a^b G(t)f(t)dt = G(x)\int_a^b f(t)dt$. Since equation (4.19) is true for all SBSs, taking expectation of this equality above for all SBSs (or all possible values of E), I have

$$\mathbb{E}[E(t+t')] - \mathbb{E}[E(t)] = -\mathbb{E}[p(\bar{t}, E(\bar{t}))]t' + \sigma\mathbb{E}[W_{t+t'} - W_t] \tag{4.20}$$

$$\int_0^\infty Em(t+t', E)dE - \int_0^\infty Em(t, E)dE = -t'\mathbb{E}[p(\bar{t}, E(\bar{t}))], \tag{4.21}$$

where $m(t, E)$ is the distribution of E in the system at time instant t. Using the fact that W is a Wiener process, $W_{t+t'} - W_t$ follows a normal distribution with mean zero, I have $\mathbb{E}[W_{t+t'} - W_t] = 0$.

By dividing both sides by t' and letting t' to be very small (or $t' \to dt$), I have $\bar{t} \to t$ and

$$\frac{d\int_0^\infty Em(t, E)dE}{dt} = -\left(\int_0^\infty p(t, E)m(t, E)dE\right) = -\bar{p}(t). \tag{4.22}$$

Using $m(t, R) = m(t, E)$, $dE = e^R dR$, and changing the variable E to R I complete the proof. □

Since $\bar{p}$ is always non-negative, the average energy in an SBS's battery is a decreasing function of time. That is the distribution m should shift toward left when t

increases. This is because I use the Wiener process in (4.2). Since dW_t has a normal distribution with mean zero, the energy harvested will be equal to the energy leakage. So for the entire system, the total energy reduces when time increases.

Lemma 4.2.2. *If (U_1, m_1) and (U_2, m_2) are two solutions of* ***Proposition 2*** *and $m_1 = m_2$, then I have $U_1 = U_2$.*

Proof. First, from (4.5), the Fokker-Planck equation is derived as follows:

$$\partial_t m_1(t, R) = \partial_R(p_1 e^{-R} m_1) + \frac{\sigma^2}{2}\partial^2_{RR}(e^{-2R} m_1),$$
$$\partial_t m_2(t, R) = \partial_R(p_2 e^{-R} m_2) + \frac{\sigma^2}{2}\partial^2_{RR}(e^{-2R} m_2).$$

Since $m_1 = m_2 = m$, I subtract the first equation from the second one to obtain $\partial_R((p_1 - p_2)e^{-R} m) = 0$. This means $(p_1 - p_2)e^{-R} m$ is a function of t. Let's denote $f(t) = (p_1 - p_2)e^{-R} m$, I have $(p_1 - p_2)m = f(t)e^R$. From **Lemma 4.2.1**, $\bar{p}$ is a function of m, thus $\bar{p}_1(t) = \bar{p}_2(t)$. Since $\bar{p}(t) = \int e^R pm dR$, I have

$$\int_{-\infty}^{\infty} e^R p_1 m dR = \int_{-\infty}^{\infty} e^R p_2 m dR \;\Rightarrow\; \int_{-\infty}^{\infty} e^R (p_1 - p_2) m dR = 0, \quad \forall t.$$

Now, substituting $(p_1 - p_2)m = f(t)e^R$ results in $\int_{-\infty}^{\infty} f(t)dR = 0 \ \ \forall t$. This means $f(t) = 0$ or $p_1 = p_2$. From the definition, U is a function of $\bar{p}$ and p. Since $\bar{p}_1 = \bar{p}_2$ and $p_1 = p_2$, it follows that $U_1 = U_2$. This lemma confirms that an SBS will act only against the mean field m. Thus m is the one that determines the evolution of the system. Two systems with the same mean field will behave similarly. □

4.3 Solving MFG Using Finite Difference Method (FDM)

To obtain U and m, the finite difference method (FDM) [30, 33] is used. The time and energy coefficient R is discretized into large intervals as $[0, ..., T_{max}\Delta t]$ and $[-R_{max}\Delta R, ..., R_{max}\Delta R]$ with Δt and ΔR as the step sizes, respectively. Then U, m, p become matrices with size $T_{max} \times (2R_{max} + 1)$. To keep the notations simple, t and R are used as the index for time and energy coefficient in these matrices with $t \in \{0, ..., T_{max}\}$ and $R \in \{-R_{max}, ..., R_{max}\}$. For example, $m(t, R)$ is the probability distribution of energy $e^{R\Delta R}$ at time $t\Delta t$. Using the FDM, $\partial_R U$, $\partial_t U$, and $\partial^2_{RR} U$ are replaced with the discrete formula as follows [34]:

$$\partial_t U(t, R) = \frac{U(t+1, R) - U(t, R)}{\Delta t}, \tag{4.23}$$

$$\partial_R U(t, R) = \frac{U(t, R+1) - U(t, R-1)}{2\Delta R}, \tag{4.24}$$

$$\partial^2_{RR} U(t, R) = \frac{U(t, R+1) - 2U(t, R) + U(t, R-1)}{(\Delta R)^2}. \tag{4.25}$$

By using them in (4.13) and after some simple algebraic steps,

$$U(t-1, R) = U(t, R) + e^{-2R} \frac{\sigma^2 \Delta t}{2(\Delta R)^2} A_1 - \Delta t B_1, \tag{4.26}$$

where

$$A_1 = U(t, R+1) - 2U(t, R) + U(t, R-1),$$
$$B_1 = (p(t, R)g)^2 - \left(\bar{\lambda}\bar{p}(t) + \lambda N\right)^2.$$

Similarly, discretizing (4.15),

$$m(t, R) = \frac{\Delta t}{2\Delta R} A_2 + \frac{\sigma^2 \Delta t}{2(\Delta R)^2} B_2 + m(t-1, R), \tag{4.27}$$

where

$$
\begin{aligned}
A_2 =& e^{-(R+1)\Delta R} p(t-1, R+1) m(t-1, R+1) - e^{-(R-1)\Delta R} p(t-1, R-1) m(t-1, R-1), \\
B_2 =& e^{-2(R+1)\Delta R} m(t-1, R+1) - 2e^{-2R\Delta R} m(t-1, R) + e^{-2(R-1)\Delta R} m(t-1, R-1).
\end{aligned}
$$

To obtain $U, m, p,$ and $\bar{p}$ using **Proposition 2**, I need to have some boundary conditions. First, to find m, it is assumed that there is no SBS that has the battery level equal to or larger than $e^{R_{max}\Delta R}$ so that $m(t, R_{max}) = 0, \quad \forall t$. This is true if it is assumed that $e^{(R_{max}-1)\Delta R}$ is the largest battery size of an SBS. Also, when $R = -R_{max}$, from the basic property of probability distribution

$$
\begin{aligned}
&\sum_{R=-R_{max}}^{R_{max}} m(t, R)\Delta R = 1 \\
&\Rightarrow m(t, -R_{max}) = \frac{1}{\Delta R} - \sum_{R=-R_{max}+1}^{R_{max}} m(t, R).
\end{aligned}
\tag{4.28}
$$

Next, to find U, again some boundary conditions need to be set. Notice that $U(T_{max}, R) = 0$ for all R. I further assume the following:

- Intutitively, if the battery level of a SBS is full, i.e., when $R = R_{max}$, this SBS should transmit something (because the thermal noise $N > 0$), or equivalently, $p(t, R_{max}) > 0$. That means

$$
p(t, R_{max}) = \frac{\bar{\lambda}\bar{p}(t) + \lambda N}{g} + \frac{e^{-R_{max}\Delta R}\partial_R U(t, R_{max})}{2g^2}. \tag{4.29}
$$

  Therefore, if $U(t, R_{max} - 1)$ and p are known, $U(t, R_{max})$ can be calculated.

- Similarly, it must be true that when available energy is 0, i.e., $R = -R_{max}$, a

SBS will stop transmission. Therefore, it can be assumed that

$$\frac{\bar{\lambda}\bar{p}(t) + \lambda N}{g} + \frac{e^{-R_{max}\Delta R}\partial_R U(t, -R_{max})}{2g^2} = 0. \tag{4.30}$$

Again, if $U(t, -R_{max}+1)$ is known, $U(t, -R_{max})$ can be calculated.

- During simulations, in some cases when the density is very high, very large (unrealistic) values of transmit power are obtained. Therefore, an extra constraint must be put for the upper limit. I use $E(t) > p(t, E(t))\Delta T$, or $e^{R(t)\Delta R} \geq p(t, R(t))\Delta T$, where ΔT is the duration of one time slot. This means, the transmit power has to be limited during one time step Δt to be smaller than the maximum power that can be transmitted during one time interval ΔT.

Based on the above assumptions, an iterative algorithm (**Algorithm 2**) is developed, similar to the one in [30]. First, a transmit power p is assumed. Next, p and $m(0, R)$ are used to calculate m by using the Fokker-Planck equation. Then, from m and p, U is calculated by using the HJB equation. Finally, p is updated by using equation (4.14). The process is iterated until the algorithm converges.

4.4 Implementation of MFG

For the MFG, the location information for each SBS is not required. However, the following information are required: the average channel gain g, $\bar{g}$, the number of SBSs M in one macrocell, and the initial distribution m_0 of the energy of SBSs in the macrocell. Therefore, some central system should measure these information, solve the differential equations, and then broadcast the power policy p to all the SBSs. It is more efficient than broadcasting all the information to all SBSs and let them solve the differential equations by themselves. Again, the central system only needs

Algorithm 2 Iterative algorithm for FDM

Initialize input

Set up $T_{max} \times (2R_{max} + 1)$ matrices U, m, p, and $T \times 1$ vector $\bar{p}$.

Guess arbitrarily initial values for power p, i.e., $p(t, R) = e^{R\Delta R}$.

Initialize $i = 1$, $U(T_{max}, .) = 0$, $m(0, .) = m_0(.)$, $m(t, R_{max}) = 0$, and $p(t, 0) = 0$.

Initialize ΔR and Δt as the step size of energy and time with $(\Delta R)^2 > \Delta t$.

Set MAX as the number of iteration

Solve PDEs with FDM

while $i <$ MAX **do**:

Solve the Fokker-Planck equation to get m using (4.27) and (4.28) with given p, m_0.

Update $\bar{p}(t)$ for $T_{max} \geq t \geq 0$ using discrete form of equation (4.15).

Calculate U for all $t < T_{max}$ by using (4.26), (4.29) and (4.30) with p, $\bar{p}$.

Calculate new transmission power p_{new} using (4.14)

Regressively update $p = ap + bp_{new}$ with $a + b = 1$.

for $R \in \{-R_{max}, ..., R_{max}\}$

if $p(t, R) > \frac{e^{R\Delta R}}{\Delta T}$ **then** $p(t, R) = \frac{e^{R\Delta R}}{\Delta T}$.

end

$i \leftarrow i + 1$.

end

Loop in T_{max} time slots

At time slot t, SBS with energy battery $e^{R\Delta R}$ transmits with power $p(t, R)$.

to re-calculate and broadcast to all SBSs a new power policy if there are changes in g, $\bar{g}$, or M.

4.5 Simulation Results

In this section, the efficacy of the developed mean field game policy is quantified in comparison to the power control policy using Markov Decision Process. The transmit power at the MBS is fixed at 10W and it results in a constant noise at the user served by a generic SBS. The radius of the macrocell is $r = 1000$ meter, so the constant cross-interference is $N_0 = 10^{-5}$ W. The target SINR is $\lambda = 0.002$ and assume that $g = \bar{g} = 0.001$. I discretize the energy coefficient R into 80 intervals, i.e., $R_{max} = 40$ and $T_{max} = 1000$ intervals. Similar to the discrete stochastic case, each SBS can hold up to 150 μJ in the battery, so the maximum transmit power is 30 mW. The threshold is chosen such that an SBS will not transmit at $R = -R_{max} = -40$ or $E = 0.6$ μJ. The intensity of energy loss/energy harvesting, σ is 1.

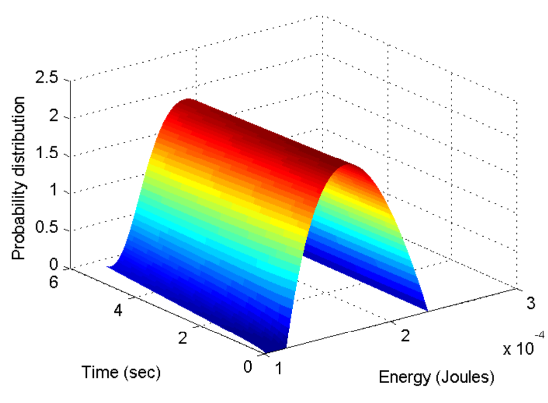

Figure 4.1: Energy distribution over time when $M = 400$ SBSs/cell.

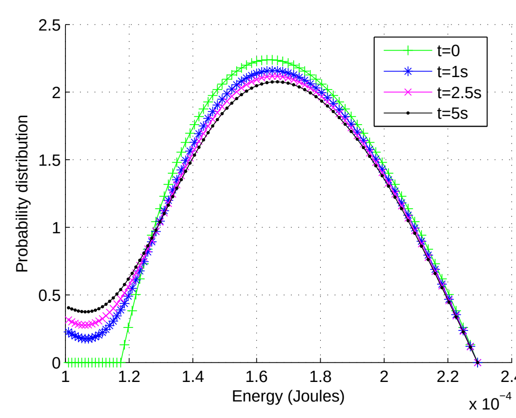

Figure 4.2: Energy distribution over time when $M = 400$ SBSs/cell.

For $M = 400$ SBSs/cell, I have $g = 0.001 > \bar{\lambda} = \lambda \bar{g} M = 0.0008$, so a generic SBS does not need to use a large amount of power in order to obtain the target SINR.

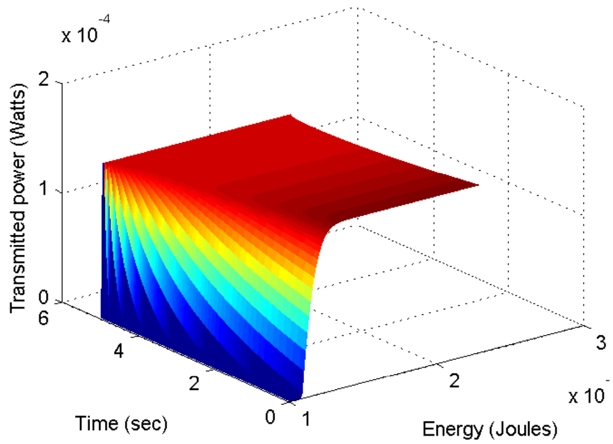

Figure 4.3: Transmit power to serve a generic user using MFG when $M = 400$ SBSs/cell.

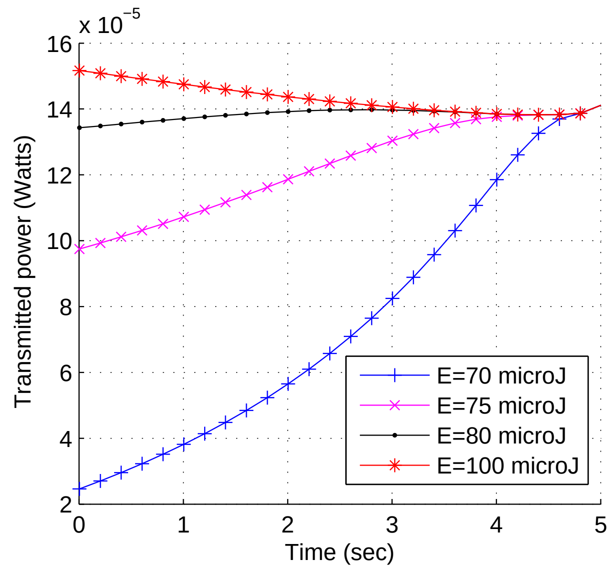

Figure 4.4: Transmit power for different energy levels when $M = 400$ SBSs/cell.

Notice that $\bar{p}$ is the average transmit power of a generic SBS. Therefore, if a generic SBS reduces its transmit power, the cost term $\bar{\lambda}\bar{p}$ also reduces. Thus the difference between the cost and the received power pg will be smaller, which is desirable. It makes sense that a generic SBS will try to reduce its power as much as possible in this case. The power cannot be zero though, because $N_0 > 0$. Moreover, from Fig. 4.3 and Fig. 4.4, it can be seen that, at the beginning, the SBS with higher energy (i.e., 100 μJ) will transmit with a high power and will gradually reduce to some value. The SBSs with smaller battery will increase their power gradually. Since the transmit power is small,as shown in Fig. 4.1 and Fig. 4.2, the energy distribution shifts to the left slowly.

On the other hand, when $M = 500$ SBSs/cell, I have $g = \bar{\lambda} = 0.001$. In this case, the effect is more complicated because reducing the transmit power may not reduce the gap between the received power pg and the cost term $\bar{\lambda}\bar{p} + \lambda N$. Again, as can

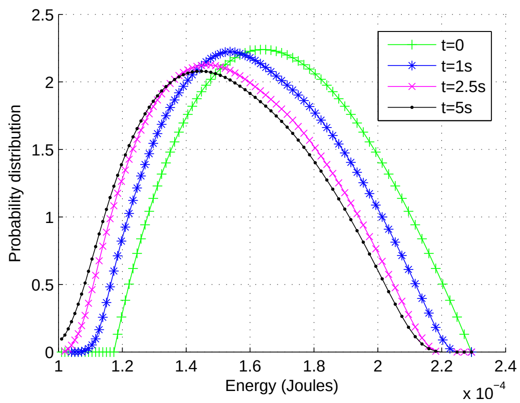

Figure 4.5: Energy distribution over time when $M = 500$ SBSs/cell.

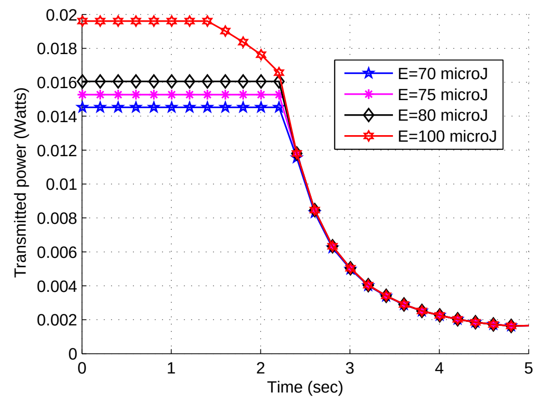

Figure 4.6: Transmit power for different energy levels when $M = 500$ SBSs/cell.

be seen from Fig. 4.6, the SBSs with larger available energy will transmit with large power first and after sometime when there is less energy available in the system, all of them start to use less power. Therefore, as can be seen in Fig. 4.5, the energy distribution shifts toward the left with a faster speed than the previous case.

For $M = 600$ SBSs/cell, I have $g < \bar{\lambda}$. This means each SBS needs to transmit with a power larger than the average $\bar{p}$ to achieve the target SINR. In Fig. 4.7,the behavior of each SBS is the same as in the previous case. That is, the SBSs with higher energy transmit with larger power first and then reduce it, while the poorer SBSs increase their transmit power over time.

The MFG model is compared against the stochastic discrete model for different values of M. For simplicity, I assume that each SBS has the same link gain to its user as $g = 0.001$. Also, I assume that the channel gain from each SBS to the user of another SBS is $\bar{g} = 0.001$. Using **Remark 2**, it can easily be proven that in this case, each SBS will transmit with the same power, i.e., if the CEQ sends QCK μJ of energy to M SBSs, then each SBS receives QCK/M μJ. Then, the interference at each SBS will be calculated as $\frac{g}{(M-1)\bar{g}+N_0 M\Delta T/(QCK)}$ with multiplier $C = 20$ and the

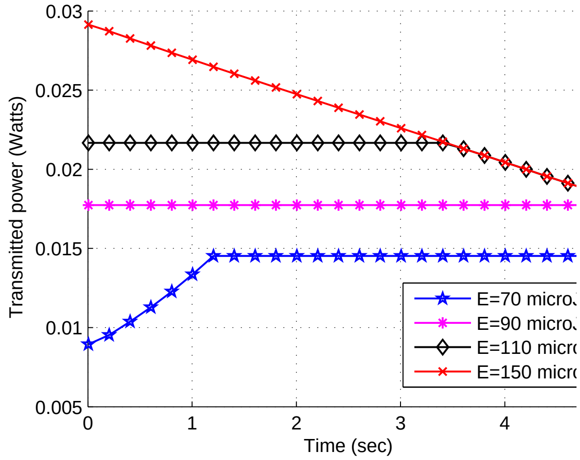

Figure 4.7: Transmission power over time when $M = 600$ SBSs/cell.

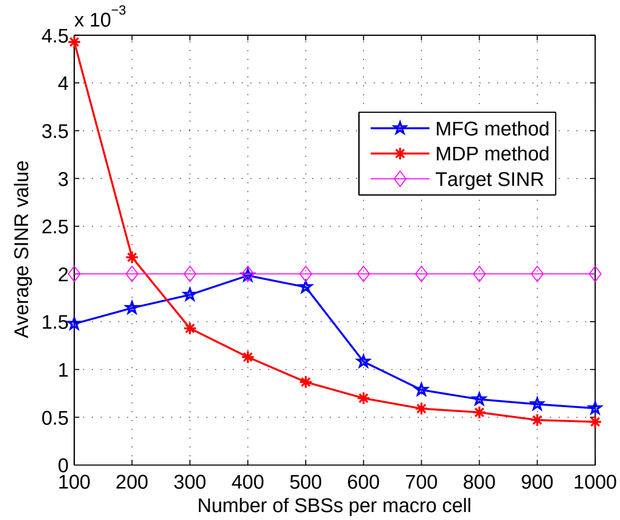

Figure 4.8: Average SINR at a generic SBS.

maximum battery size of the CEQ as $S = 101$. Because the MBS is not a player of the game, the simulation step becomes simpler, and only a linear program for the MDP needs to be solved instead of a QCQP. Therefore, it is referred to as the MDP method to accurately reflect the difference.

For the discrete stochastic case, the Gaussian distribution is discretized to model the energy arrivals at the CEQ. The battery size of each SBS is still 150 μJ. The average SINR of a generic small cell user using both MFG and MDP models with different density is plotted in Fig. 4.8. It can be seen that using the MFG model, the average SINR increases at the beginning and then it starts falling at some point. This is because, when the density is low, the interference from the MBS is noticeable (i.e., 10^{-5} W in my simulation). From the previous figures, it can be seen that an SBS will increase its power when the density is higher. Therefore, after some point the co-tier interference becomes dominant and the average SINR will begin to drop. It means at some value of the density, e.g., $M = 400$ SBSs/cell in Fig. 4.8, the optimal average SINR is obtained. It can be noticed that the MFG model performs better than the

MDP model with the CEQ. One of the reasons is that the CEQ must consume an integer number of quanta, which limits its flexibility compared to the MFG model. Also, the size of the battery at the CEQ is a limitation. Lastly, one of the reasons I use the CEQ is to redistribute the energy based on channel gains. However, in a dense network scenario, these gains are the same for every SBS and the CEQ loses this advantage.

In summary, two important remarks can be made for the MFG model. First, if the density of small cells is high, the SBSs will transmit with higher power. Second, from Fig. 4.8, it can be seen that by choosing a suitable density of SBSs I can obtain the highest average SINR. Notice that from (4.14), it can be easily proven that the average SINR at a user served by an SBS will always be smaller than the target SINR (because $\partial_R U < 0$). Therefore, the highest average SINR is also the closest to the target SINR, which is my objective in the first place.

Chapter 5

Summary and Future Work

5.1 Summary

In this thesis, I have proposed a two-tier system model with many SBSs and one MBS share the same communication channel. The SBSs harvest energy from environment for transmission.

In Chapter 3, a discrete single-controller discounted two-player stochastic game has been applied to address the power control problem for energy harvesting small cell networks. A Central Energy Queue (CEQ), which can harvest and distribute energy to each SBS, is deployed to simplify the game model and optimize the average payoff for all SBSs. For this discrete case, the strategies for both the macrocell BS and the small cell BSs have been derived by solving a quadratic optimization problem. The numerical results have shown that using CEQ can provide a better performance in terms of outage probability for both the MBS and the SBSs. It also shows that when the number of SBSs increases, I can improve the outage probability by adjusting other parameters such as: changing the target SINR, increasing the multiplier C, or the battery size of each SBS.

In Chapter 4, the system model has been extended to the extreme case where there are a large number of SBSs in the system. Since CEQ and discrete stochastic game are inapplicable with large number of SBSs, I have applied a mean field game model to obtain the optimal power for this case. Subsequently, the forward-backward equations are derived from this model. I have applied the finite difference method to numerically solve this problem. At the end, I have also discussed the implementation aspects of these models in a practical network. The numerical results show that if the density of SBSs is high, the average transmit power also increases. Moreover, over time, the total energy will be reduced and the distribution of energy will shift to the left. Finally, experimental results also show that MFG method performs better in term of average SINR compared to policies using Markov Decision Process when the density is high.

5.2 Future Work

In this thesis, I have assumed that CEQ distributes energy to each SBS over a perfect medium, i.e., no energy loss. It can be considered that the CEQ wirelessly transfers energy to each SBS instead, to increase the portability of the system. In this case, there will be some uncertainty over the channel where the energy is wirelessly transfered. Obviously I do not want to spend the harvested energy carelessly over high-uncertainty channel so a robust solution should be found. To make the problem more practical, conventional energy source can be added to the CEQ, i.e., the CEQ can consume more than it harvests. The objective of the sub-problem in Remark 2 can then be modified to minimize the power consumed from conventional source such that all SBSs obtain their target SINRs. The new sub-problem will be a linear optimization problem. If the uncertainty of the channel is mdeled using ellipsoidal model, then a robust counterpart can be derived.

Lastly, the central method with CEQ is infeasible if the number of SBSs increases. In the system model considered here, the CEQ does everything from finding the Nash equilibrium between MBS and itself to distributing energy. One possible way to remove this bottleneck is to make the SBSs smarter. In this approach CEQ will only need to harvest energy and distribute it. SBSs now will need to bid for the energy it requires to transmit. To make problem simpler, it can be assumed that the MBS maintains a constant throughput. Using the SINR formula, the transmit power at the MBS can be easily derived as a function of its interference. The CEQ will distribute energy based on the offers it receives from the SBSs. To help the SBSs with poor channel conditions, a conventional energy source can be added. Obviously, the price per energy unit in conventional source must be fixed and more expensive than the bidding price from the CEQ. This model is somewhat similar to a dynamic auction game model proposed in [41].

Bibliography

[1] Cisco's White Paper ,"The Internet of Things How the Next Evolution of the Internet Is Changing Everything", April 2011.

[2] G. Piro et al, "HetNets powered by renewable energy sources sustainable next-generation cellular networks", *IEEE Internet Computing*, vol. 17, no. 1, pp. 32–39, Jan. 2013.

[3] S. Sudevalayam, P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Communications Surveys & Tutorials,* vol. 13, no. 3, Sep. 2011, pp. 443–461.

[4] G. Auer *et al*, " How much energy is needed to run a wireless network," *IEEE Wireless Communications*, vol. 18, no. 5, Oct. 2011, pp. 40–49.

[5] M. Deruyck, D. D. Vulter, W. Joseph, and L. Martens, "Modelling the power consumption in femtocell networks," in *IEEE WCNC 2012 Workshop on Future Green Communications.*

[6] Y. Mao, Y. Luo, J. Zhang and K. Letaief, "Energy Harvesting Small Cell Networks: Feasibility, Deployment and Operation", http://arxiv.org/pdf/1501.02620v1.pdf, 12 Jan. 2015.

[7] J. Yang and S. Ulukus. "Optimal packet scheduling in an energy harvesting communication system," *IEEE Transactions on Communications*, vol. 60, no. 1, January 2012, pp. 220–230.

[8] K. Tutuncuoglu and A. Yener, " Optimum transmission policies for battery limited energy harvesting nodes." *IEEE Transactions on Wireless Communications,* vol. 11, no. 3, March 2012, pp. 1180–1189.

[9] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener. " Transmission with energy harvesting nodes in fading wireless channels: Optimal policies." *IEEE Journal on Selected Areas in Communications,* vol. 29, no. 8, September 2011, pp. 1732–1743.

[10] D. Gunduz, K. Stamatiou, M. Zorzi, "Designing intelligent energy harvesting communication systems," *IEEE Communications Magazine*, vol. 52, no. 1, Jan. 2014, pp. 210–216

[11] C. Huang, R. Zhanga, and S. Cui, "Optimal power allocation for outage probability minimization in fading channels with energy harvesting constraint," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, Feb. 2014, pp. 1074–1087.

[12] Z. Ding, S. Perlaza, I. Esnaola, and H. Poor, "Power allocation strategies in energy harvesting wireless cooperative networks," *IEEE Transactions on Wireless Communications*, Jan. 2014, pp. 846–860.

[13] C. K. Ho and R. Zhang, " Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, Sep. 2012, pp. 4808–4818.

[14] N. Michelusi, K. Stamatiou, and M. Zorzi, "Transmission policies for energy harvesting sensors with time-correlated energy supply," *IEEE Transactions on Communications*, vol. 61, no. 7, Jul. 2013, pp. 2988–3001.

[15] H. Dhillon, Y. Li, P. Nuggehalli, Z. Pi, and J. Andrews, "Fundamentals of heterogeneous cellular networks with energy harvesting," www.arxiv.org/pdf/1307.1524, Jan. 2014.

[16] D. W. K. Ng, E. S. Lo and R. Schober, " Energy-efficient resource allocation in OFDMA systems with hybrid energy harvesting base station," *IEEE Transactions on Wireless Communications,* vol.12, no. 7, July 2013, pp. 3412–3427.

[17] X. Kang, Y. K. Chia, C. K. Ho and S. Sun, "Cost minimization for fading channels with energy harvesting and conventional energy," *IEEE Transactions on Wireless Communications,* vol. 13, no. 8, Aug. 2014, pp. 4586–4598.

[18] V. Chandrasekhar and J. Andrews, "Power control in two-tier femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, Aug. 2009, pp. 4316–4328.

[19] S. Guruacharya, D. Niyato, D. I. Kim, and E. Hossain, "Hierarchical competition for downlink power allocation in OFDMA femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, Apr. 2013, pp. 1543–1553.

[20] P. Blasco, D. Gndz, and M. Dohler, "A learning theoretic approach to energy harvesting communication system optimization," *IEEE Transactions on Wireless Communications,* vol. 12, no. 4, Apr. 2013, pp. 1872–1882.

[21] K. Huang and V. K. Lau, "Enabling wireless power transfer in cellular networks: Architecture, modeling and deployment," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, Feb. 2014, pp. 902–912.

[22] H. Tabassum, E. Hossain, A. Ogundipe, and D.I. Kim, Wireless-Powered Cellular Networks: Key Challenges and Solution Techniques," *IEEE Communications Magazine*, to appear.

[23] "C-RAN – the road towards green RAN," White Paper, China Mobile Research Institure, October 2011.

[24] P. Semasinghe, E. Hossain, and K. Zhu, "An evolutionary game for distributed resource allocation in self-organizing small cells," *IEEE Transactions on Mobile Computing*, DOI 10.1109/TMC.2014.2318700.

[25] M. Bowling and M. Veloso, "An analysis of stochastic game theory for multiagent reinforcement learning," *Technical Report*, Carnegie Mellon University.

[26] J. Hu and M. P. Wellman, "Nash Q-learning for general-sum stochastic games," *Journal of Machine Learning Research*, vol. 4, Nov. 2003, pp. 1039–1069.

[27] J. Filar and K. Vrieze, *Competitive Markov Decision Process*. Springer 1997.

[28] J. A. Filar, "Quadratic programming and the single-controller stochastic games," *Journal of Mathematical Analysis and Applications*, vol. 113, no. 1, Jan. 1986, pp. 136–147.

[29] H. Tembine, R. Tempone, and P. Vilanova, "Mean field games for cognitive radio networks," *in Proc. of American Control Conference*, June 2012.

[30] MFG Labs,"A mean field game approach to oil production," http://mfglabs.com/wp-content/uploads/2012/12/cfe.pdf accessed in 14-Nov-2014.

[31] O. Gueant, J. M. Lasry, and P. L. Lions, "Mean field games and applications," *Paris-Princeton Lectures on Mathematical Finance*, Springer, 2011, pp. 205-266.

[32] A. Y. Al-Zahrani, R. Yu, and M. Huang, "A joint cross-layer and co-layer interference management scheme in hyper-dense heterogeneous networks using mean-field game theory," *IEEE Transactions of Vehicular Technology*, to appear, Oct. 2014.

[33] D. Bauso, H. Tembine, and T. Basar "Robust mean field games with application to production of an exhaustible resource," *Proc. of the 7th IFAC Symposium on Robust Control Design*, Aalborg, 2012.

[34] J. Li and Yi-Tung Chen, *Computational Partial Differential Equations Using MATLAB*. Chapman and Hall/CRC, 2008.

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

[36] E. Altman et al., "Constrained stochastic games in wireless networks," *IEEE Globecom General Symposium*, Washington D.C., 2007.

[37] E. Altman et al., "Dynamic discrete power control in cellular networks," *IEEE Transactions on Automatic Control*, vol. 54, no. 10, Oct. 2009, pp. 2328–2340.

[38] M. Huang, Member, P. E. Caines, and R. P. Malham, “Uplink power adjustment in wireless communication systems: A stochastic control analysis,” *IEEE Transactions on Automatic Control*, vol. 49, no. 10, Oct. 2004, pp. 1693–1708.

[39] Y. Achdou, I. C. Dolcetta, “Mean field games: Numerical methods” *SIAM Journal on Numerical Analysis*, vo. 48, no. 3, 2010, pp. 1136–1162.

[40] W. A. Strauss, *Partial Differential Equations: An Introduction*. 2nd edition, Wiley, 2007.

[41] M. J. Bonet, and M. Pesendorfer, “Estimation of a dynamic auction game,” *Econometrica*, vol. 71, 2003, 1443–1489.

[42] G. N. Iyengar, “Robust dynamic programming,” *Mathematics of Operations Research,* vol. 30, no. 2 May 2005, pp. 257–280.

Publications

1. K. T. Tran, H. Tabassum, and E. Hossain, "A stochastic power control game for two-tier cellular networks with energy harvesting small cells," *Proc. IEEE Globecom'14*, Austin, TX, USA, 8-12 December, 2014.

2. K. T. Tran, E. Hossain, and H. Tabassum, "Downlink Power Control in Two-Tier Cellular Networks with Energy-Harvesting Small Cells as Stochastic Games", submitted to the *IEEE Transaction of Communications*