Green Wireless Transmissions for Advanced Internet of Things: Session-Specific Analysis and Design with Deep Reinforcement Learning

by

Fang Xu

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electrical and Computer Engineering

Green Wireless Transmissions for Advanced Internet of Things: Session-Specific Analysis and Design with Deep Reinforcement Learning

by

Fang Xu

Supervisory Committee

Dr. Hong-chuan Yang, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Xiaodai Dong, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Yang Shi, Outside Member
(Department of Mechanical Engineering)

Supervisory Committee

Dr. Hong-chuan Yang, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Xiaodai Dong, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Yang Shi, Outside Member
(Department of Mechanical Engineering)

ABSTRACT

Reliable and energy-efficient wireless data transmission is of great importance to the success of future Internet of Things (IoT). In this thesis, we analyze and minimize the energy consumption of wireless data transmissions from an individual transmission perspective. In particular, we investigate three fundamental transmission scenarios, namely point-to-point transmission, data collection from wirelessly-powered sensor, and cooperative relaying transmission with wireless power transfer.

For point-to-point transmission, we analyze and optimize ideal continuous rate adaptation and continuous power adaptation transmission schemes, where the closed-form expressions for all optimal parameters are derived. These results establish an energy consumption lower limit for wireless data transmissions over fading channels. In the case of data collection from wirelessly-powered sensor scenario, we derive closed-form expressions for optimal transmission parameters for ideal rate adaptive transmission with linear energy harvesting setting. Under more practical assumptions of finite block-length transmission with nonlinear energy harvesting, we propose a deep reinforcement learning (DRL) based approach to arrive at a deep policy network for determining the near-optimal transmission

parameters in real time. An online tuning method is also proposed to adjust the policy network using online experience to cater for model inaccuracy and environment variation. Finally, for the case of cooperative relaying transmission with wireless power transfer, under ideal rate adaptive transmission with piecewise linear energy harvesting, we derive the closed-form expressions for all optimal transmission parameters. Then, the optimal design problem is again generalized to the finite block-length transmission with nonlinear energy harvesting setting, where we again apply the DRL-based method to train and tune a deep policy network for determining the near-optimal transmission parameters in real time. For all cases, we illustrate various design tradeoffs through selected numerical examples. Besides improving the energy efficiency of wireless transmissions for future IoT applications, our proposed DRL-based method can serve as a general solution for real-time optimal design problems in wireless communications.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

My dearest families, for their encouragement, patience, love and support throughout my life. Without them, I can not achieve anything.

Pro. Hong-chuan Yang, for his instruction, support, help, patience and tolerance.

Pro. Xiaodai Dong, for serving as a committee member for me.

Pro. Yang Shi, for serving as a committee member for me.

My lab members, Tingnan Bao, Shuang Li, Yuan Liu, Huimin Hu, Hyemin Yu, Ramy Zaghloul, Wenjing Wang, for their advices and help.

Chapter 1

Introduction

1.1 Background and motivation

Internet of Things (IoT) can effectively extend the connectivity of Internet to numerous terminal devices, e.g. various sensors, cell phones, RF tags, etc, and enable future smart world. It is expected to revolutionize the world similar to what the Internet itself did [1] at the beginning of this century. As such, IoT has attracted a lot of attention from both academia and industry over the past decade. More specifically, through facilitating the interconnection of numerous devices anywhere and anytime, IoT will enable many important emerging applications, e.g. smart transportation, automatic industry, intelligent agriculture, smart city and so on, as shown in Fig. 1.1. However, there are still many issues needed to be addressed, such as low energy efficiency, massive random access, privacy and security concerns, interference-free connectivity [2], etc. The total number of connected IoT devices may be up to 100 billion by 2030 [3] and so, total energy consumption of IoT devices will explosively grow. This will lead to much more carbon emissions into the atmosphere and constitute a very big challenge for the goal of energy saving. Note that the amount of carbon dioxide (CO2) emissions from the communication networks may already exceed 345 million tons in 2020 and it is still increasing rapidly [4], which will further accelerate the global warming. Accordingly, a novel green technology is more and more desirable for suppressing the emission of greenhouse gases. Meanwhile, there is a pressing need to utilize the resource more efficiently and reduce the energy consumption for the operation of all connected devices to achieve green IoT.

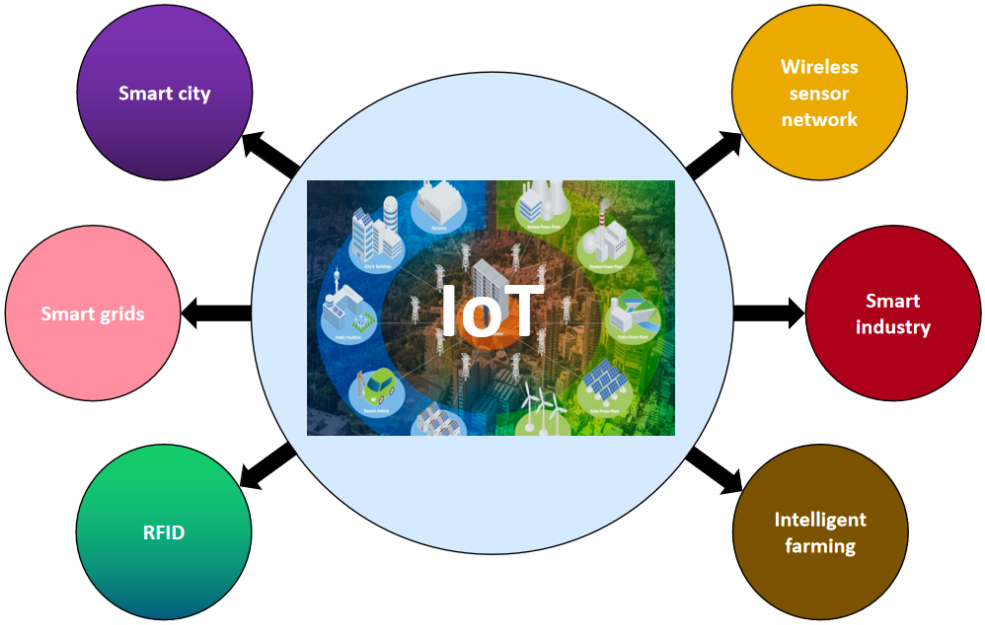

Figure 1.1: The diagram of IoT applications.

There are three key technologies for the implementation of green IoT [3], i.e. green RFID tags, green sensor network, and green Internet. The energy consumption of RFID and sensor networks is mainly from data collection and transmission and hence, improving the energy efficiency of wireless transmissions is extremely valuable. The data transmission sessions for different IoT applications usually have dramatically different quality of service (QoS) requirements. For mission critical IoT applications, high reliability and low latency are essential. For example, with factory automation applications, packet loss rate and latency are demanded to be lower than 10^{-9} and less than 1 ms, respectively. Meanwhile, IoT system involves a large amount of wireless devices with strictly constrained energy supply. These devices, usually powered by nonchargeable and nonreplaceable batteries, are expected to operate for many years. However, most existing transmission technologies were mainly designed for mobile broadband (MBB) services, which are very different from IoT services [5]. The diverse requirements for IoT applications, e.g. high reliability, low latency, good sustainability and high energy efficiency, etc, are still very hard to be satisfied. Accordingly, novel transmission technologies need to be investigated for effectively and efficiently supporting advanced IoT applications.

1.2 Conventional approaches for energy efficient transmission

In the past, researchers already performed much work to enhance the energy efficiency of wireless transmissions [6-27]. Generally, proposed techniques mainly include cross-layer design, power adaptation, resource allocation, adaptive modulation scheme, adaptive switching strategy, relay based transmission, and so on.

Cross-layer design

Wireless transmission systems have conventionally been designed independent of upper layers, including link layer, network layer, transport layer, and application layer. Through performing optimal joint design across these layers, the overall system performance may be considerably enhanced. The work of [6–8] investigated the energy-efficient transmission schemes on the basis of cross-layer design. The work of [6] presented a comprehensive overview of recent advances in cross-layer design for green wireless communications, where link-level transmission schemes and network/mac layer resource management strategies were discussed in detail. To achieve low energy consumption in ad hoc wireless networks, The work of [7] investigated the requirements of cross-layer design across the link, medium access, network and application layers. The interactions across different protocol layers were discussed as well. The work of [8] considered a potential 5G spectrum-sharing transmission system, where a modified energy per good bit (MEPG) metric was proposed and minimized by cross-layer optimization for two scenarios. In addition, the optimal scheme of associated resource allocation was obtained accordingly.

Transmit power adaptation

Transmit power directly affects the quality of received signal in wireless transmissions. So, power adaptation is a promising method to effectively ameliorate the transmission performance. The work of [9–12] investigated the energy-efficient wireless transmissions by utilizing transmit power adaptation. The work of [9] investigated energy-efficient design in multi-cell scenarios with inter-cell interference, where transmit power was optimized to boost energy efficiency and refine the trade-off of energy efficiency and spectral efficiency (SE). The work of [10] considered a multi-cell massive multiple-input–multiple-output (MIMO) system, where transmit power adaptation was utilized to minimize the energy consumption of a base station. The work of [11] considered a point-to-point trans-

mission, where energy efficiency was maximized by adapting both overall transmit power and its allocation. The work of [12] investigated energy efficient power adaptation in spectrum sharing cognitive radio systems, where the secondary users first obtained channel state information through sensing and then initiated data transmission with two power levels.

Resource allocation

Note that available resources, e.g. transmit power, available channel bandwidth, etc, are strictly constrained in realistic wireless transmissions, and they usually need to be allocated to multiple users in many systems. Hence, proper resource allocation strategies are critical to realize good performance. Resource allocation schemes were optimized to improve the energy efficiency of wireless transmissions in [13–18]. The work of [13] investigated the energy-efficient OFDMA system with consideration of circuit power consumption within frequency-selective fading channels, where the overall energy efficiency was maximized by optimizing transmit power allocation among subcarriers. The work of [14] considered a network with two users, where the power was optimally allocated to maximize the energy efficiency of each user in a distributed way. The work of [15] investigated the energy-efficient resource allocation, based on physical-layer network coding, in OFDM bidirectional transmission, where overall transmit power consumption, with constrained throughput, was minimized by jointly optimizing subcarriers and power allocation. The work of [16] investigated an energy-efficient UAV-enabled communication network, where user communication scheduling, UAV trajectory, transmit power and bandwidth allocation were jointly optimized to maximize energy efficiency with constrained QoS. The work of [17] considered a cooperative relaying system with non-orthogonal multiple access, where the global energy efficiency was maximized by optimizing power allocation. The work of [18] proposed an energy-efficient power allocation for a multicarrier link, over a frequency-selective fading channel, with a delay outage probability constraint, where the energy efficiency of system was maximized accordingly.

Adaptive modulation and switching strategy

Since the wireless channels are generally time-varying, the transmission strategy also should be time-varying. As such, utilizing an adaptive transmission strategy can lead to much better transmission performance. Adaptive modulation schemes and switching strategies were utilized to realize highly energy-efficient wireless transmissions [19–23]. The work of [19] investigated the energy-efficient design for general OFDMA system under flat fad-

ing channels, where the energy efficiency of uplink transmission was enhanced by employing adaptive modulation. The work of [20] investigated the alamouti diversity based MIMO system, where optimal modulation and transmission strategies were investigated to minimize the energy consumption of transmission. The work of [21] pointed out that with MIMO system, switching off some radio frequency (RF) amplifier units at night can save energy significantly while maintaining QoS of active users. The work of [22] utilized adaptive switching between MIMO mode and SIMO mode to save energy at mobile terminals. The work of [23] utilized adaptive switching strategy among different MIMO modes in the transmission, where it showed a better energy efficiency than non-adaptive systems.

Relay selection

With poor channel quality or long transmission distance, it is difficult for direct transmission to guarantee desirable performance. Hence, relay transmission was introduced into the wireless transmissions to enhance reliability, energy efficiency, and effective data rate. In terms of system with multiple relays, selecting proper relays, for forwarding data, is very critical to achieve good performance. Relay selection was utilized to improve the energy efficiency in transmissions in [24–27]. The work in [24] and [25] investigated the effect of relay selection schemes on the overall energy efficiency of transmission system, where they showed that more relay nodes could not lead to a higher energy efficiency. The work of [26] considered the wireless transmission with cooperative hybrid automatic retransmission request, where the total consumed energy was reduced by utilizing an optimal relay selection scheme. The work of [27] considered a discrete modulation based nonregenerative multi-relay system, where energy efficiency, subject to data rate constraint, was effectively improved by employing relay selection and resource allocation.

1.3 Session-specific transmission design

IoT applications usually entail massive connectivity and sporadic short transmission sessions [111]. Most existing works characterize the energy efficiency of wireless transmission in terms of average transmission rate with respect to energy utilization [29]. Specifically, achieved data rate per unit power consumption, with unit of bits/s/Watt or equivalently bits/Joule, is a widely used energy efficiency metric [30]. The work of [31] calculated the average energy consumption for each transmitted bit with consideration of possible retransmission. For fading channels, the average energy efficiency is evaluated with the

application of channel ergodic capacity [32, 33]. Such average energy efficiency characterization, however, is not sufficient for IoT applications that typically involve sporadic short transmission sessions. Also note that the energy consumption of each IoT transmission session varies dramatically with prevailing fading channel condition. Recently, a data-oriented approach was proposed for wireless transmission system analysis and design to best satisfy stringent QoS requirement of future IoT applications [34]. The basic idea is to design transmission schemes from an individual transmission session perspective instead of targeting average channel quality. A statistical energy efficiency characterization, namely energy outage rate (EOR), was developed to capture the randomly varying nature of fading wireless channels. The work of [35] analyzed the energy consumption of individual data transmission session, where big data transmission with practical adaptive modulation and coding was investigated. However, the results are not applicable to small data sporadic transmission in IoT. Moreover, the energy efficiency optimization and the effect of latency constraint were not investigated there.

In general, the achievable energy efficiency of wireless transmission is limited by the QoS requirement. The effect of QoS requirement on average energy efficiency was previously studied using the concept of effective capacity [32, 33]. However, the data to be transmitted over different sessions may have considerably different QoS requirements. The transmission strategy that is optimal in the average sense may not be the best choice for an individual data transmission session [34]. Furthermore, the average-based characterization can only take into account the statistical QoS requirement, using the concept of effective capacity. For example, the work in [35–39] investigated the average energy efficiency over fading channels under statistical QoS constraint for different transmission scenarios. It is of great interest to study the effect of instantaneous QoS requirement on the energy consumption of an individual data transmission session.

Advanced IoT applications usually have very strict QoS constraints and hence, the energy consumption for them is very high. Accordingly, the investigation on the reduction of energy consumption for wireless transmissions is extremely important. Typically, such energy consumption minimization subject to QoS constraint can be formulated as an optimization problem. Since most optimization problems are very complicated in IoT, closed-form solutions can not be derived in most cases. Usually, iterative algorithms are employed to handle these complicated problems, under which some time has to be consumed for iterative calculation during each optimization. Note that IoT applications typically involve sporadic short transmission sessions with strict latency constraints, where the channel condition dramatically varies from one channel coherence time to another. So, the solutions

obtained from iterative algorithms may be based on outdated channel knowledge, and corresponding time consumption may violate the constraints of latency. Combining the above descriptions, conventional iterative algorithms are not suitable for IoT applications. As such, more advanced optimization algorithms need to be developed to effectively support the machine-type communications with high energy efficiency and diverse QoS requirements.

1.4 Finite block-length transmission

Note that as compared to Shannon capacity, in terms of small data transmissions, finite block-length formula is a more accurate accessible upper bound for real channel coding rate [40]. As such, researchers performed some work on the investigation of finite block-length transmission, e.g. [40–43]. The work of [40] derived the finite block-length expression to approximate the upper bound of achievable channel coding rate under given block-length and error probability. The work of [41] investigated the validity of Shannon's separation theorem within finite block-length regime, where joint source-channel coding scheme was shown to have better performance than separate coding scheme. The work of [42] proposed a self-adaptive ordered statistics decoder (S-OSD) scheme for finite block-length nonbinary raptor code (NBRC), where some strategies were proposed to minimize the complexity of decoding. The work of [43] investigated the interference constrained performance of spectrum sharing networks using adaptive power allocation, where finite block-Length analysis was performed to show the effect of codeword length on the spectrum sharing networks.

1.5 Machine learning techniques for wireless communications

Machine learning (ML) is usually used to train a model for approximating a complex mapping relationship, which can't be represented by an exact mathematical expression. It has already shown very excellent performance for handling some tasks, such as regression, classification, and the learning of action policy, etc. Recently, a lot of ML based methods were applied in wireless communications for addressing some very complicated design problems, e.g. [44-78]. In general, they can be roughly divided into four categories, namely

supervised/unsupervised learning, reinforcement learning (RL), DRL, and federated learning (FL).

1.5.1 Supervised/unsupervised learning

Supervised learning is a ML technique to train a model that maps the input to the output based on training data with deterministic labels. Unsupervised learning is another ML technique to train a mapping model from unlabelled or partial labelled training data. Supervised learning and unsupervised learning were used to handle the complex problems of wireless transmissions [44-53]. The work of [44] considered a realistic noncoherent wireless body area network (WBAN) model, where transmissions and receptions were performed without any channel knowledge due to the fast-varying channels in the human body. Here, a supervised learning approach was proposed to overcome the noncoherent issue for detecting received data symbol. The work of [45] considered an adaptive MIMO-OFDM system, where a supervised learning framework was proposed to predict the best modulation order and coding rate for increasing the accuracy of link adaptation. The work of [46] considered a MIMO system with low-resolution analog-to-digital converters (ADCs), where a supervised learning aided successive-interference-cancellation (SIC) method was proposed to reduce the training overhead as well as detection complexity. The work of [47] proposed an unsupervised learning method to fully utilize the decision factors with different types for optimal network selection in heterogeneous wireless networks. The work of [48] considered a reconfigurable intelligent surface (RIS) assisted system, where a unsupervised learning approach was proposed to reduce the complexity of passive beamforming design. The work of [49] considered the massive access problem with both machine-to-machine and human-to-human traffic, where a novel machine learning scheme was proposed to achieve better control for machine-to-machine traffic under the interference of human-to-human traffic. The work of [50] presented an overview of the solutions toward ameliorating random access network (RAN) congestion, where potential advantages and challenges for emerging ML techniques were discussed. The work of [51] considered the coordinated multiple point (CoMP) transmission within a ultra-dense small cell network (SCN), where a supervised learning method was proposed to create data model from input data, for reducing feedback load, to estimate optimal step size in iterative calculation. The work of [52] proposed a big data and supervised learning enabled wireless channel estimation framework, where channel statistical properties could be estimated based on carrier frequency and the coordinates of transmitter and receiver. The work of [53] considered the orthogo-

nal frequency division multiplexing (OFDM) based wireless communication system, where sparse learning algorithms were proposed to eliminate the impulsive noise. However, both supervised learning and unsupervised learning can not be used for addressing optimization issues.

1.5.2 Reinforcement learning

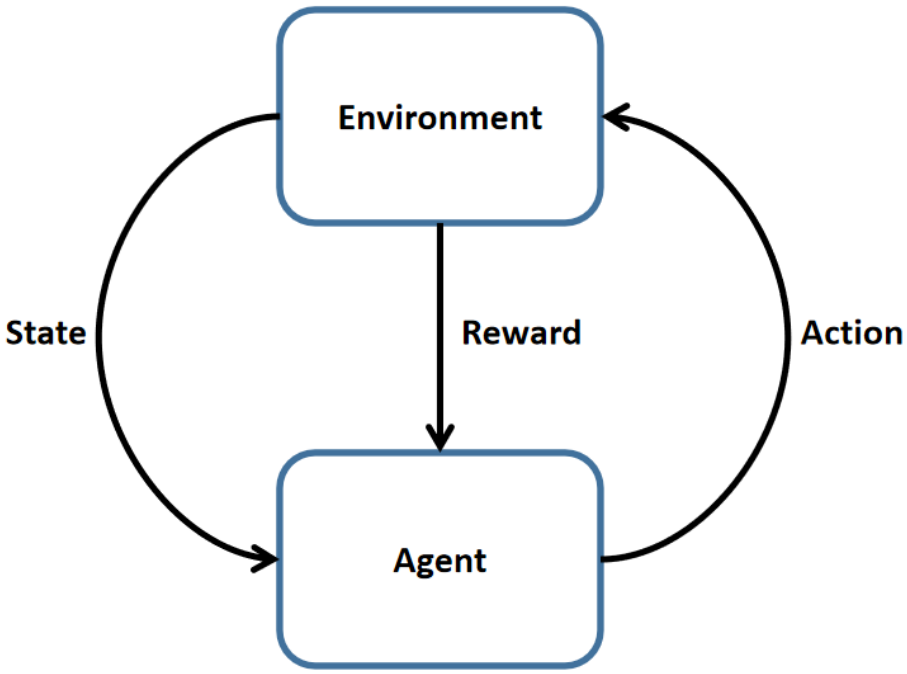

Figure 1.2: The interaction between agent and environment within the reinforcement learning.

RL is a very popular ML technique to train an optimal action policy within discrete state space and action space. In particular, as shown in Fig. 1.2, during each iteration, an intelligent agent can update the action policy through maximizing the expected value function, for its current action, obtained from the environment. After a certain amount of training iterations, the agent usually can obtain a good action policy. The expected value function is calculated by Bellman equation, shown as follows

$$V_\pi(s,a) = \mathbb{E}[r_{t+1} + \gamma V_\pi(s_{t+1}, a_{t+1}) | s_t = s, a_t = a]. \tag{1.1}$$

Here, $V_\pi(.)$ denotes the expected value function under the action policy π, γ denotes the discounting factor, r_{t+1} denotes the instantaneous reward at time instant $t+1$, s_j and a_j denote the state and action at time instant j ($j = t,\ t+1$). The work in [54-59] applied RL method to improve the performance of wireless transmissions. The work of [54] considered the packetized sensor communications, where a RL based strategy was proposed to choose the optimal modulation level and transmit power for maximizing the throughput per unit energy consumption. The work of [55] considered a cellular internet of UAVs, where RL was applied to solve key problems on trajectory control and resource management. The work of [56] considered a mobility-aware mobile edge computing (MEC) network, where a RL approach was proposed to design the optimal joint scheme for task offloading and migration. The work of [57] proposed an improved sarsa RL algorithm to improve the model's convergence property and learning efficiency, which will significantly improve the efficiency for the design of artificial intelligent communication networks. The work of [58] considered the wireless sensor networks, where a multi-agent RL based algorithm was proposed to achieve intelligent anti-jamming communication. The work of [59] considered a multi-agent communication system over noisy channels, where a RL based framework was proposed to achieve better coordination and cooperation for all agents. Although RL showed very good performance for some complicated problems, it can not be applicable for continuous action space and state space.

1.5.3 Deep reinforcement learning

DRL can be used to train an optimal action policy within a continuous state space, which is implemented by combining deep learning and RL. Particularly, a deep neural network is employed to approximate the Q value function through deep Q learning, and the optimal action can be determined through comparing the Q values of discrete actions. Meanwhile, for implementing RL within the high-dimensional or continuous action space, policy gradient method was proposed to directly learn the parameterized policy function, e.g. Actor-Critic [60], where the agent's policy can be optimized by performing gradient ascent on the parameters of deep neural network. Note that deep Q learning and Actor-Critic were further combined in the algorithm of deep deterministic policy gradient (DDPG) [61] for extending DRL to continuous action space, where a deep Q network is set as the critic agent and another deep network is set as the actor agent. Worthy of mentioning is that the actor model, trained from DRL based method, will not involve any iterative calculation within the process of online optimization, which is very suitable for IoT applications under strict

latency constraints. However, as mentioned above, the energy consumption and QoS constraint of transmission session varies from one session to another in IoT. Optimal strategy for future expected reward, obtained from conventional DRL, is not necessarily the best option for a specific transmission session.

Recently, DRL based method was widely applied into wireless communications for improving the performance of systems over a period of time, such as [58-69]. The work of [62] considered the self-organizing UAV networks, where a DRL based method was proposed to construct the backhaul framework for enhancing the backhaul rate with limited information exchange. The work of [63] considered the cooperative transmissions in the wireless sensor networks (WSNs), where a DRL based relay selection scheme was proposed to improve the performance of system on outage probability, system capacity, and energy consumption. The work of [64] considered the wireless networks with massive access requests from a hugh number of devices, where a DRL based distributed algorithm was proposed to meet both reliability and latency constraints on ultra-reliable low latency communications (URLLC) services within the massive access scenario. The work of [65] considered an intelligent reflecting surface (IRS) aided wireless secure communication system, where a DRL based secure beamforming approach was proposed to dynamically achieve the optimal beamforming policy for improving the secrecy rate. The work of [66] considered a MIMO full duplex (FD) system, where a hybrid DRL based method was proposed to jointly clusters the MIMO antennas between energy harvesting and information transmission. Also, optimal precoding matrix could be found in real time for both transmitter and receiver. On the basis of DRL, The work of [67] proposed a non-cooperative and real-time approach to achieve energy-efficient power allocation under the QoS constraints of D2D communication. The work of [68] investigated the problem of dynamic spectrum sensing and aggregation in a wireless network with N correlated channels, where a DRL based algorithm was used to maximize the number of successful transmissions without interrupting the primary users (PUs). The work of [69] proposed a novel energy management strategy, based on the DRL, for the sensors in IoT with renewable energy supply. As compared to other state-of-the-art algorithms, this algorithm showed better performance in terms of long-term average net bit rate. The work of [70] considered a UAV assisted wireless sensor network, where the algorithm of deep deterministic policy gradient (DDPG) was used to train the UAV to understand the environment and provide effective scheduling to accomplish its data collection mission. The work of [71] considered the data collection in mobile crowd sensing, where various unmanned mobile terminals (MTs) are equipped with different sensors that aid to collect data. A DDPG based algorithm was proposed to maximize

the total amount of collected data with the limited energy reserve, in which geographical fairness among all point-of-interests (PoIs) was also guaranteed. The work of [72] considered the next generation radio access networks (RANs) with intelligent controllers, where an DRL based user access control scheme was proposed to maximize the longterm throughput and avoid frequent handovers. The work of [73] considered the relay transmissions in WSNs, where a DRL based relay selection scheme was proposed to achieve low energy consumption and outage probability.

1.5.4 Federated learning

FL is a ML technique to train a model across multiple decentralized edge devices with local data samples, in which all devices only need to exchange their local models rather than local data. As such, this method can effectively protect the privacy of all devices. The work of [74-78] applied the technique of FL into wireless transmissions for enhancing the privacy protection during the process of learning. The work of [74] considered a reconfigurable intelligent surface (RIS) assisted mmWave communication system, where a FL based privacy-preserving design paradigm was proposed to optimize the RIS' configuration matrix for maximizing the throughput. The work of [75] proposed a FL based framework to harness wireless channel perturbations and interference for improving privacy, bandwidth-efficiency, and scalability within the process of model training. The work of [76] investigated the problem of energy efficient transmission and computation resource allocation for federated learning over wireless communication networks, where overall energy consumption was minimized under a latency constraint through optimizing the resource allocation. The work of [77] considered the FL process over a realistic wireless network, where FL loss function was minimized by jointly optimizing learning parameters, wireless resource allocation, and user selection. The work of [78] considered the cooperative perception within the vehicular communication network, where a federated RL approach was proposed to maximize vehicles' satisfaction in terms of the received sensory information as well as speed up the overall training process.

1.5.5 Research Gap

Conventional DRL only targets at the maximization of future expected reward. Note that IoT applications typically involve sporadic short transmission sessions, where QoS requirement and channel condition may dramatically vary from one transmission session to another. As such, expected reward is not sufficient for accurately evaluating the performance

of data transmissions in IoT. On the other hand, conventional DRL is mainly based on online learning, where all learning experience is collected from real operation. This is computationally intensive and highly complex for wireless communication systems. Accordingly, there is a pressing need to investigate some novel learning methods to perform real-time optimal design for data transmissions in advanced IoT.

1.6 Contributions and thesis outline

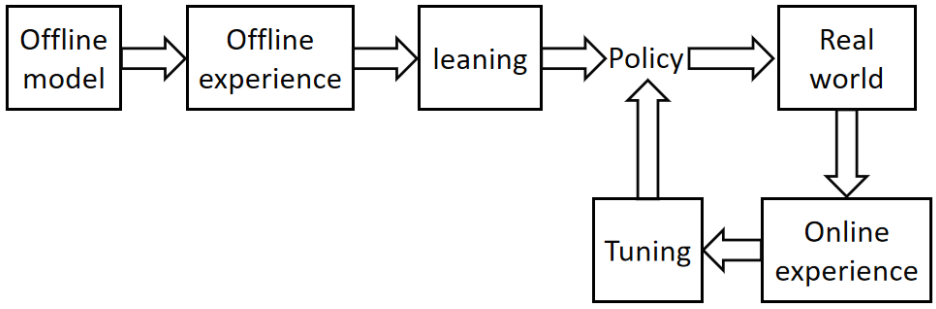

Figure 1.3: The configuration for our proposed DRL-based approach.

In this thesis, we analyze and minimize the energy consumption of wireless transmissions from an individual transmission perspective for three fundamental transmission scenarios, namely point-to-point transmission, data collection from wirelessly-powered sensor and cooperative relaying transmission with wireless power transfer. In particular, under ideal rate adaptive transmission and/or linear energy harvesting, we derive the closed-form expressions for all optimal transmission parameters with a latency constraint in terms of all scenarios. These results establish the lower bound of energy consumption for wireless transmissions over fading channels. Under practical finite block-length transmission with nonlinear energy harvesting, we propose a model-based DRL solution to arrive at a deep policy network for determining near-optimal transmission parameters in real time for sensor data collection and cooperative relaying transmission. In addition, online experience is also used for further adjusting the well-trained offline policy network to cater for model inaccuracy and environment variation. The configuration of our proposed DRL solution is shown as Fig. 1.3. Note that our DRL-based approach does not involve any online iterative calculation and can be generalized to any complicated problem. As such, we provide a

general solution for effectively handling the optimal design problems in wireless communications. The remaining parts of this thesis are organized as follows.

In chapter 2, we investigate the optimal design of a point-to-point transmission system from a session specific perspective, where a continuous rate adaptation scheme and a continuous power adaptation scheme are both discussed. The closed-form expressions of optimal transmission parameters, with and without latency constraint, are derived. By calculating the resulting energy outage rate, we establish a statistical energy efficient limit for point-to-point transmissions over fading channels.

In chapter 3, we investigate the optimal design of a sensor data collection system from a session-specific perspective. Specifically, the agent firstly charges the sensor and then collects data from it. With consideration of ideal rate adaptive transmission with linear energy harvesting, we derive the closed-form expressions for all optimal transmission parameters. With consideration of finite block-length transmission with nonlinear energy harvesting, we propose a DRL-based approach to train a deep policy network for determining near-optimal transmission parameters in real time for the online operation of the agent. Meanwhile, we also propose an online tuning method to adjust the well-trained policy network for better adapting to the variation of environment. Through comparing with the results from exhaustive search, we verify the validity of our analysis and proposed approaches.

In chapter 4, we investigate the optimal design of a cooperative relaying transmission system with wireless power transfer. With consideration of ideal rate adaptive transmission with piecewise linear energy harvesting, for time switching mode and power splitting mode at the relay, we minimize the energy consumption of the system when transmitting a fixed amount of data. Closed-form expressions for all optimal transmission parameters are derived with and without latency constraint. With consideration of finite block-length transmission with nonlinear energy harvesting, we again apply the DRL-based approach from chapter 3 to train and adjust a deep policy network for determining near-optimal transmission parameters in real time.

In chapter 5, we conclude the whole thesis and present the future research direction.

Chapter 2

Green Point-to-Point Wireless Transmission over Fading Channels

Future IoT applications demand highly energy-efficient wireless transmission solutions. In this chapter, we characterize the energy efficiency of point to point wireless transmissions over fading channels from an individual transmission perspective. Specifically, we apply a statistical energy efficiency metric to analyze and optimize ideal continuous rate adaptation (CRA) and continuous power adaptation (CPA) transmission schemes. The resulting statistical characterization establishes an energy efficiency performance limit for data transmissions over fading channels, and facilitates a direct tradeoff analysis between maximal achievable energy efficiency and instantaneous latency requirement.

2.1 Introduction

The minimum energy consumption of wireless point-to-point transmission is typically a constant value for Gaussian noise channels [80, 81], but become random in fading environment. To develop a more thorough energy efficiency characterization for wireless transmission particularly targeting IoT applications, the minimum energy consumption, for reliably transmitting a given amount of data over fading channels, needs to be investigated. Therefore, we follow a statistical approach and analyze the probability that the energy consumption for delivering a given amount of data exceeds a certain threshold level. The resulting energy efficiency metric, termed as energy outage rate (EOR), was first introduced in [34], where a data-oriented approach for wireless transmission system design was proposed. However, little analytical results were reported in [34] due to its magazine paper

nature.

Here, we analyze the EOR performance of two ideal adaptive transmission strategies, namely CRA and CPA, to establish statistical energy efficiency performance limit for fading environment. With CRA, the transmission rate is adaptively selected according to the channel condition while fixing the transmit power, whereas with CPA, the transmit power is varying to support a target transmission rate, subject to a peak transmit power constraint [82]. For the CRA scheme, we determine the optimal transmit power in terms of minimizing the energy consumption and study the resulting EOR performance with and without a latency constraint, which allow us to directly analyze the effect of instantaneous latency constraint on achievable energy efficiency. Additionally, we qualitatively demonstrate that the CPA scheme achieves better energy efficiency performance at the cost of larger data transmission latency. To the best knowledge of me, the only other previous work that characterized energy efficiency statistically is [35], where the energy consumption for big data transmission with practical adaptive transmission scheme was analyzed. However, the results are not applicable to small data sporadic transmission in IoT. Moreover, the statistical energy efficiency optimization and the effect of latency constraint were not investigated there.

2.2 Problem formulation

We consider a generic data transmission session over a point-to-point flat fading channel. We assume that the amount of data to be transmitted, denoted by H, is small and slow fading is introduced in the channel, in which data transmission can be completed within one channel coherence time. Note that the transmitter can be either a data center or a sensor. The received signal during data transmission is given by

$$y(t) = h \cdot s(t) + n(t), \tag{2.1}$$

where h is the complex channel gain, $s(t)$ is the transmitted signal, and $n(t)$ is the additive white Gaussian noise signal, with noise spectral density N_0. Therefore, the instantaneous received SNR is equal to $P_t g/N_0 B$, where P_t is the transmit power, B is the channel bandwidth, and $g = |h|^2$ is the instantaneous channel power gain.

Due to the randomly varying nature of fading wireless channels, the energy consumption for successfully transmitting H bits of information, denoted by $E_{\mathrm{c}}(H)$, varies with the channel realization. To analyze the energy efficiency of an individual data transmis-

sion session, we adopt the statistical metric EOR [34], which is mathematically defined as EOR $= \Pr[E_{\mathrm{c}}(H) > E_{\mathrm{th}}]$, where E_{th} denotes an energy consumption threshold. In the following sections, we will analyze and optimize the EOR of two ideal adaptive transmission schemes, namely CRA and CPA. We assume that the transmitter has the knowledge of channel power gain g to perform the rate and power adaptation. This can be achieved through either feedback or exploring channel reciprocity in time-division duplexing (TDD) mode.

2.3 Continuous rate adaptation

With ideal CRA transmission, the transmitter adapts the transmission rate with the channel condition while using a fixed transmit power P_t [82]. We assume that with proper modulation and coding schemes, achieved rate can follow a similar trend as Shannon capacity. Accoridngly, applying the Shannon capacity formula, the maximum instantaneous data rate for reliable transmission is equal to $C(g) = B \cdot \log_2(1 + P_t g / N_0 B)$. As such, the minimum time duration to finish data transmission is determined as $H/C(g)$.[1] Denoting the circuit power of the transmitter by P_c, we can calculate the energy consumption with CRA as the product of the total power and the minimum transmission time, given by

$$E_{\mathrm{c}}(H) = (P_t + P_c)\frac{H}{B \cdot \log_2(1 + P_t g / N_0 B)}, \tag{2.2}$$

which varies with the instantaneous channel power gain g. It follows that the EOR of data transmission with CRA can be calculated as

$$\mathrm{EOR}_{\mathrm{CRA}} = \Pr[E_{\mathrm{c}}(H) \geq E_{\mathrm{th}}] = F_g\left(\frac{N_0 B}{P_t}\left(2^{\frac{(P_t+P_c)H}{E_{\mathrm{th}}B}} - 1\right)\right), \tag{2.3}$$

where $F_g(\cdot)$ denotes the CDF of channel power gain g. For the Rayleigh fading scenario, we arrived at the following closed-form expression for EOR

$$\mathrm{EOR}_{\mathrm{CRA}} = 1 - \exp\left(-\frac{N_0 B}{P_t \bar{g}}\left(2^{\frac{(P_t+P_c)H}{BE_{\mathrm{th}}}} - 1\right)\right), \tag{2.4}$$

where $\bar{g}$ is the average channel power gain.

[1] Note that the achievable rate may be smaller than $C(g)$ due to finite data amount [84]. In this work, we apply the Shannon capacity formula to establish energy efficiency performance limits.

2.3.1 Transmit power optimization

We can minimize the energy consumption of CRA transmission by optimally setting the transmit power.

Theorem 1. *The transmit power P_t that minimizes the energy consumption for transmitting a fixed amount of data with CRA is given by*

$$P_t^*(g) = \frac{P_c - N_0 B/g}{W_0[\frac{g}{N_0 Be}(P_c - N_0 B/g)]} - N_0 B/g, \tag{2.5}$$

where $W_0[\cdot]$ is the principal branch of Lambert W function [85]. Also note that the effect of latency constraint on the optimal power level will be discussed in section 2.3.2*.*

Proof. It is straightforward to verify that the second derivative of (2.2) with respect of P_t is positive for $P_t \geq 0$, and as such, (2.2) is a strictly convex function of P_t. Taking the derivative of (2.2) with respect to P_t and setting the result to zero, the optimal P_t value should satisfy

$$(P_t^* + P_c)\frac{g}{N_0 B} = (1 + \frac{P_t^* g}{N_0 B})\ln(1 + \frac{P_t^* g}{N_0 B}). \tag{2.6}$$

After some manipulation and applying the definition of Lambert W function, we obtain the optimal transmit power given in (2.5). □

Fig. 2.1 presents the behavior of optimal transmit power versus instantaneous channel power gain for different circuit power consumption levels. We can observe that the optimal transmit power decreases as channel power gain increases and/or circuit power consumption decreases. It can be verified from (2.5) that when P_c or g becomes very large, the optimal transmit power P_t^* approaches 0.

After substituting (2.5) into (2.2) and some manipulations, the energy consumption with CRA while using optimal transmit power can be determined as

$$E_c^*(H) = \frac{(P_c - \frac{N_0 B}{g})\frac{H \ln 2}{B}}{W_0[\frac{g}{N_0 Be}(P_c - \frac{N_0 B}{g})]}. \tag{2.7}$$

Following the analysis in the Appendix, the EOR of CRA with optimal transmit power is

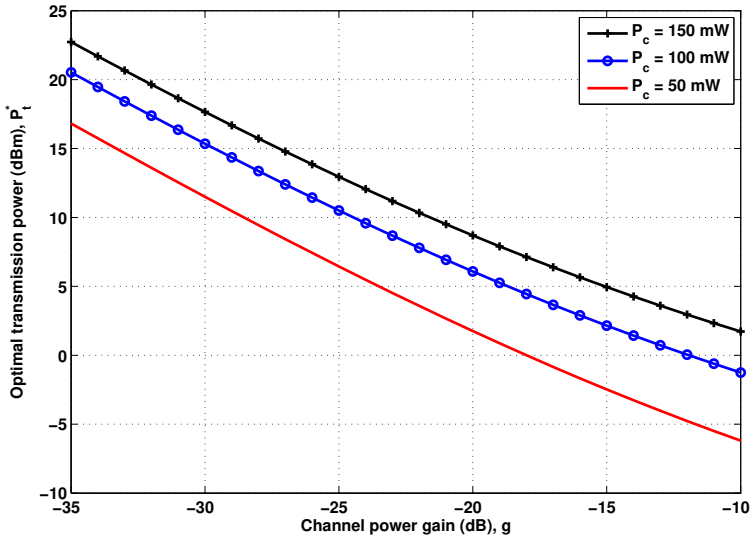

Figure 2.1: Optimal P_t values as function of channel power gain g for different circuit power consumption levels P_c.

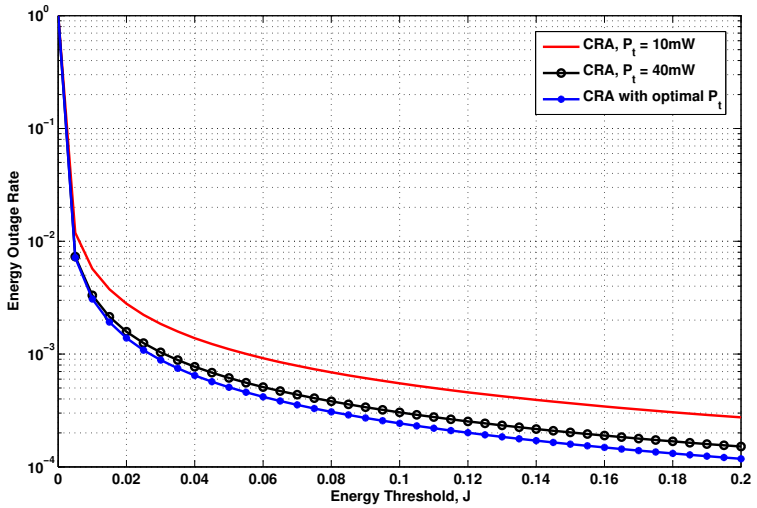

Figure 2.2: Energy outage rate of CRA with constant and optimal transmit power over slow Rayleigh fading channels ($H = 10$ kB, $B = 200$ KHz, $P_c = 15$ mW, and $\overline{g}$ = -5 dB).

given by

$$\begin{aligned}
&\mathrm{EOR}_{\mathrm{CRAw}P_t^*} = \\
&F_g\Big(\frac{-HN_0\ln 2}{E_{\mathrm{th}}W_0(-\exp(-\frac{P_cH\ln 2}{BE_{\mathrm{th}}})/e)}\Big) - F_g\Big(\frac{-HN_0\ln 2}{E_{\mathrm{th}}W_{-1}(-\exp(-\frac{P_cH\ln 2}{BE_{\mathrm{th}}})/e)}\Big) \\
&+ F_g\Big(\min\Big\{\frac{N_0B}{P_c+BE_{\mathrm{th}}/(H\ln 2)}, \frac{-HN_0\ln 2}{E_{\mathrm{th}}W_{-1}(-\exp(-\frac{P_cH\ln 2}{BE_{\mathrm{th}}})/e)}\Big\}\Big),
\end{aligned} \tag{2.8}$$

where $W_{-1}[\cdot]$ denotes the negative branch of Lambert W function [85].

Fig. 2.2 presents the EOR performance of CRA over slow Rayleigh fading channels as the function of the energy threshold E_{th} for different transmit power settings. We can see that higher transmit power leads to larger EOR. Typically, larger transmit power help improve the received SNR for the same channel realization, which allows for higher transmission rate with CRA and in turn reduces the data transmission duration. The transmission time reduction is, however, in logarithm with respect to P_t. As such, the energy consumption for the same amount of data increases with P_t, which leads to higher EOR. From Fig. 2.2, we can also clearly see that EOR of CRA with optimal transmit power are considerably lower than those with constant transmit power.

2.3.2 Effect of latency constraint

The optimal transmit power value in terms of minimizing the energy consumption tends to be small, especially when the circuit power level is low, which may lead to very low transmission rate. In certain IoT applications, the data needs to be delivered with a hard delay constraint, denoted by $T_{\rm th}$. To satisfy such a latency constraint, the transmit power with CRA transmission should be lower-bounded by $P_t(t) \geq (2^{\frac{H/T_{\rm th}}{B}}-1)\frac{N_0 B}{g} \triangleq \widetilde{P}_t$. The corresponding energy consumption is given by

$$E_{\rm c}(H) = \begin{cases} (\widetilde{P}_t + P_c)T_{\rm th}, & \widetilde{P}_t \geq P_t^*; \\ \frac{(P_c - \frac{N_0 B}{g})\frac{H\ln 2}{B}}{W_0[\frac{g}{N_0 B e}(P_c - \frac{N_0 B}{g})]}, & \widetilde{P}_t < P_t^*. \end{cases} \tag{2.9}$$

The EOR under the latency constraint can be calculated by considering the $\widetilde{P}_t \geq P_t^*$ and $\widetilde{P}_t < P_t^*$ cases separately. In particular, when $\widetilde{P}_t \geq P_t^*$, which can be shown to be equivalent to $g < \frac{N_0 B(\frac{H}{T_{\rm th}B}\ln 2 - 1)}{P_c}\exp(\frac{H}{T_{\rm th}B}\ln 2) + \frac{N_0 B}{P_c} \triangleq g_c$, energy outage occurs when $g < \frac{N_0 B(2^{\frac{H}{T_{\rm th}B}}-1)}{HE_{\rm th}/T_{\rm th} - P_c H}$. Note that we assume that $E_{\rm th} > P_c T_{\rm th}$ here. If $\widetilde{P}_t < P_t^*$, the result of the Appendix can applies.

Finally, the EOR of CRA transmission with optimal power under a latency constraint can be shown, by considering the different value ranges of g_c, to be given by

$$\begin{aligned} {\rm EOR}_{{\rm CRA}|T_{\rm th}} &= F_g(\min\{\frac{N_0 B(2^{\frac{H}{T_{\rm th}B}}-1)}{HE_{\rm th}/T_{\rm th} - P_c H},\, g_c\}) + \\ &\begin{cases} F_g(\frac{-HN_0\ln 2}{E_{\rm th}W_0(-\exp(-\frac{P_c H\ln 2}{BE_{\rm th}})/e)}) - F_g(\frac{-HN_0\ln 2}{E_{\rm th}W_{-1}(-\exp(-\frac{P_c H\ln 2}{BE_{\rm th}})/e)}) \\ -F_g(g_c) + F_g(\min\{\frac{N_0 B}{P_c + BE_{\rm th}/(H\ln 2)},\, \frac{-HN_0\ln 2}{E_{\rm th}W_{-1}(-\exp(-\frac{P_c H\ln 2}{BE_{\rm th}})/e)}\}), \\ g_c \leq \min\{\frac{N_0 B}{P_c + BE_{\rm th}/(H\ln 2)},\, \frac{-HN_0\ln 2}{E_{\rm th}W_{-1}[-\exp(-\frac{HP_c\ln 2}{BE_{\rm th}})/e]}\}; \\ F_g(\frac{-HN_0\ln 2}{E_{\rm th}W_0[-\exp(-\frac{HP_c\ln 2}{BE_{\rm th}})/e]}) - F_g(\frac{-HN_0\ln 2}{E_{\rm th}W_{-1}[-\exp(-\frac{HP_c\ln 2}{BE_{\rm th}})/e]}), \\ \min\{\frac{N_0 B}{P_c + BE_{\rm th}/(H\ln 2)},\, \frac{-HN_0\ln 2}{E_{\rm th}W_{-1}[-\exp(-\frac{HP_c\ln 2}{BE_{\rm th}})/e]}\} < g_c \leq \\ \frac{-HN_0\ln 2}{E_{\rm th}W_{-1}[-\exp(-\frac{HP_c\ln 2}{BE_{\rm th}})/e]}; \\ F_g(\frac{-HN_0\ln 2}{E_{\rm th}W_0[-\exp(-\frac{P_c H\ln 2}{BE_{\rm th}})/e]}) - F_g(g_c), \\ \frac{-HN_0\ln 2}{E_{\rm th}W_{-1}[-\exp(-\frac{HP_c\ln 2}{BE_{\rm th}})/e]} < g_c \leq \frac{-HN_0\ln 2}{E_{\rm th}W_0[-\exp(-\frac{HP_c\ln 2}{BE_{\rm th}})/e]}. \end{cases} \end{aligned} \tag{2.10}$$

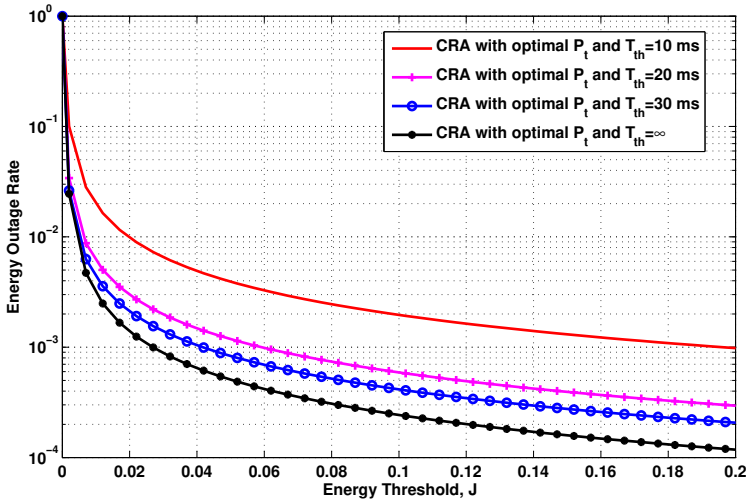

Figure 2.3: Effect of latency constraint on EOR performance of CRA with optimal transmit power ($H = 10$ kB, $B = 200$ KHz, $P_c = 15$ mW and $\overline{g} =$ -5 dB).

We study the effect of latency constraint on the EOR performance of CRA in Fig. 2.3. In particular, the EOR of CRA with optimal transmit power under different delay threshold T_{th} are plotted as function of energy threshold E_{th}. We can see that more stringent latency constraint leads to poorer EOR performance.

In Fig. 2.4, we compare the minimum energy consumption under average energy efficiency metric and that under session specific energy efficiency metric for a specific transmission session. Here, average energy efficiency denotes average energy consumption over a period of time for a specific session. Session specific energy efficiency denotes instantaneous energy consumption for a specific session. Note that for the case of average energy efficiency metric, we use average latency constraint over a period of time instead of instantaneous latency constraint. As expected, we see that session specific metric based design can reduce the energy consumption considerably.

In Fig. 2.5, we compare the outage probability of QoS constraint under average energy efficiency metric and that under session specific energy efficiency metric for an individual transmission session. Here, QoS constraint denotes that instantaneous latency needs to be less than a threshold T_{th}. We see that as compared to average energy efficiency metric, session specific energy efficiency metric can arrive at a much lower outage probability for QoS constraint. Combining Fig. 2.4 and Fig. 2.5, we can show that in terms of optimal design for a specific transmission session, session specific metric is more suitable than average metric.

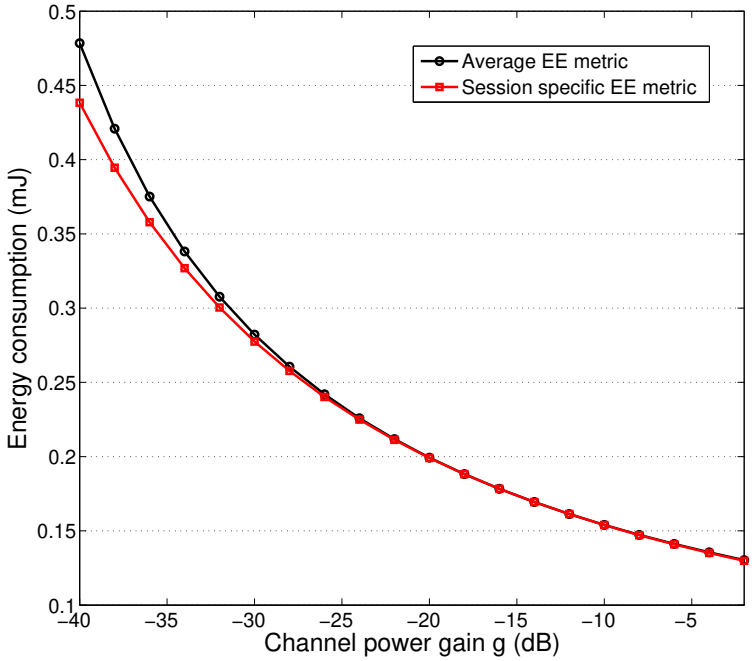

Figure 2.4: The comparison of minimum energy consumption for average and session-specific design ($H = 10$ kB, $B = 100$ KHz, $T_{\text{th}} = 20$ ms, $P_c = 20$ mW and $P_{\max} = 300$ mW).

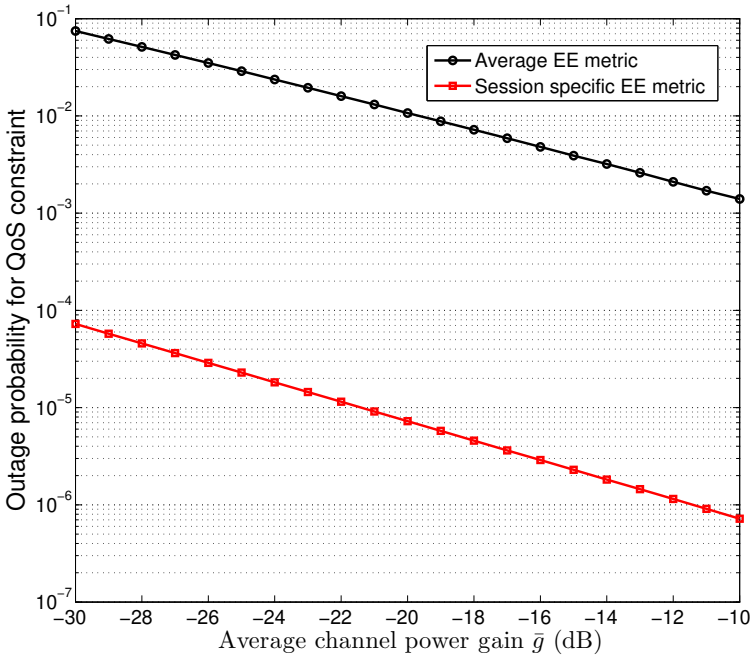

Figure 2.5: The comparison of the outage probability of QoS constraint for average and session-specific design ($H = 10$ kB, $B = 100$ KHz, $T_{\text{th}} = 20$ ms, $P_c = 20$ mW and $P_{\max} = 300$ mW).

2.4 Continuous power adaptation

Rate adaptive transmission requires reconfigurable transceivers that can adjust coding and modulation schemes with prevailing channel condition. Many IoT devices cannot afford such complexity. Alternatively, these devices may simply adapt the transmit power with the channel condition to support a constant transmission rate R. According to the Shannon capacity result, the minimum received SNR required for error free transmission at rate R is given by $\gamma_c = 2^{R/B-1}$. Under a peak transmit power constraint $P_{\max}$, the transmit power P_t should be set to $\gamma_c N_0 B/g$ when $g \geq \gamma_c N_0 B/P_{\max}$, and 0 otherwise (i.e. truncated channel inversion [82]). The energy consumption for transmitting data amount H with such continuous power adaptation (CPA) strategy can be calculated as

$$E_{\mathrm{c}}(H) = (\frac{\gamma_c N_0 B}{g} + P_c)\frac{H}{R}, \tag{2.11}$$

when $g \geq \gamma_c N_0 B/P_{\max}$. The EOR of data transmission with CPA can be calculated as the probability that $E_{\mathrm{c}}(H)$ is greater than the energy threshold E_{th}, which leads to

$$\mathrm{EOR}_{\mathrm{CPA}} = \Pr[(E_{\mathrm{th}}R/H - P_c)g < \gamma_c N_0 B]. \tag{2.12}$$

Apparently, when E_{th} is less than the circuit energy consumption during transmission, i.e. $E_{\mathrm{th}} < \frac{H}{R}P_c$, $\mathrm{EOR}_{\mathrm{CPA}}$ is equal to 1. Here, circuit energy consumption is equal to the product of circuit power and transmission duration. When $E_{\mathrm{th}} \geq \frac{H}{R}P_c$, while noting that data transmission occurs only when $g \geq \gamma_c N_0 B/P_{\max}$, the EOR of CPA can be determined as

$$\begin{aligned}\mathrm{EOR}_{\mathrm{CPA}} &= \Pr[g < \frac{\gamma_c N_0 B}{E_{\mathrm{th}}R/H - P_c}] \\ &= \begin{cases} \frac{F_g(\frac{\gamma_c N_0 B}{E_{\mathrm{th}}R/H - P_c}) - F_g(\frac{\gamma_c N_0 B}{P_{\max}})}{1 - F_g(\frac{\gamma_c N_0 B}{P_{\max}})}, & E_{\mathrm{th}} \leq (P_{\max} + P_c)H/R; \\ 0, & E_{\mathrm{th}} > (P_{\max} + P_c)H/R. \end{cases}\end{aligned} \tag{2.13}$$

We examine the effect of peak transmit power and circuit power on the EOR performance of CPA in Fig. 2.6. We first note that larger circuit power consumption will degrade the EOR performance of CPA transmission, as expected by intuition. We also observe that larger peak transmit power value results in poorer EOR performance, as the transmitter may transmit over poor channel condition. Finally, we notice that CPA scheme achieves better EOR performance than the CRA scheme. This somewhat counter-intuitive result (as CRA

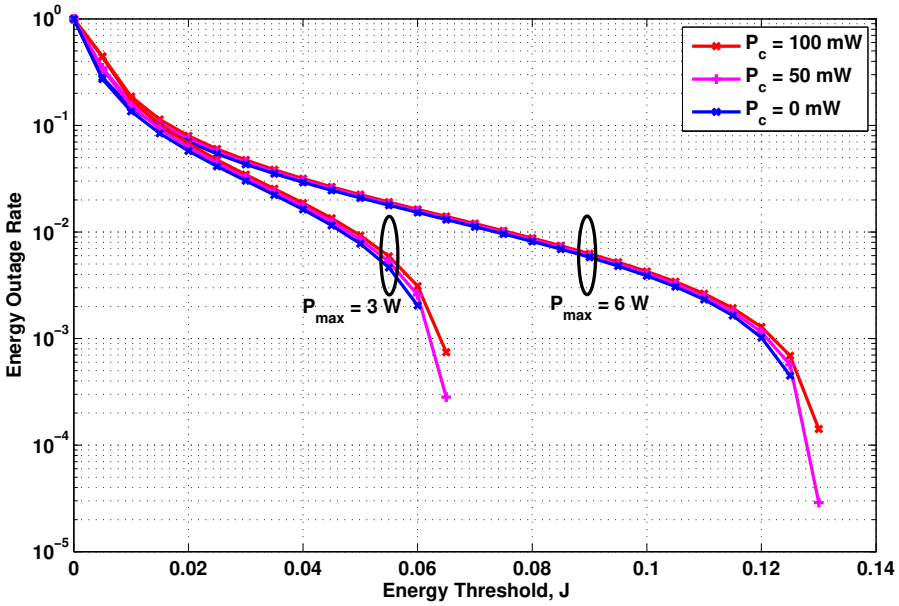

Figure 2.6: Energy outage rate of CPA over slow Rayleigh fading channel ($H = 10$ kB, $B = 200$ kHz, $\gamma_c = 16$ dB, and $\overline{g} =$ -5 dB).

has higher implementation complexity than CPA) can be explained from two perspectives. First, we can see that $E_{\mathrm{c}}(H)$ for CPA is inverse proportional to the channel gain g, whereas $E_{\mathrm{c}}(H)$ for CRA, given in (2.2), is approximately proportional to $1/\log_2(1+g)$. Secondary, the EOR performance advantage of CPA comes at the cost of occasional transmission outage. On the other hand, we can see the cut-off effect when the energy threshold increases. Note that with CPA, transmission occurs only if $g \geq \gamma_c N_0 B/P_{\max}$ due to peak transmit power constraint. Meanwhile, energy outage occurs when $g < \gamma_c N_0 B/(E_{\mathrm{th}}R/H - P_c)$. When E_{th} is large enough, $\gamma_c N_0 B/(E_{\mathrm{th}}R/H - P_c)$ will be less than $\gamma_c N_0 B/P_{\max}$. In this scenario, EOR will be equal to zero, as shown in Eq. (13).

Note that if $g < \gamma_c N_0 B/P_{\max}$ upon data arrival, CPA will hold data transmission for a certain duration until the channel condition improves. It can be shown with the application of outage duration result in [86] that the CDF of waiting duration over Rayleigh fading can be obtained as

$$F_{T_w}(\tau) = 1 - F_g(g_T)\frac{2\overline{\tau}}{\tau} I_1\left(\frac{2\overline{\tau}^2}{\pi(\tau)^2}\right)\exp\left(-\frac{2\overline{\tau}^2}{\pi(\tau)^2}\right), \tag{2.14}$$

where $I_1(\cdot)$ is the modified Bessel function and $\overline{\tau}$ is given by

$$\overline{\tau} = \frac{\exp(\rho^2) - 1}{\sqrt{2\pi}\rho f_m}, \tag{2.15}$$

with $\rho = \gamma_c N_0 B/(P_{\max}\bar{g})$ and f_m being the maximum Doppler frequency shift. Note that this CDF result considers the possible mobility [86]. The tradeoff between EOR performance and waiting duration for CPA transmission was already illustrated in [83].

2.5 Conclusion

In this work, we analyzed the statistical energy efficiency, namely EOR, of wireless data transmissions over fading channels from an individual transmission session perspective. We focused on ideal rate and power adaptive transmission schemes and established a statistical energy efficiency limit for wireless transmissions. For CRA scheme, the optimal transmit power in terms of maximizing energy efficiency was derived. Based on these analytical results, the transmitter will learn how to adapt transmission parameters in real time for minimizing the EOR of a specific transmission session. Also, we showed that session-specific metric is more suitable for achieving green transmissions in advanced IoT.

Chapter 3

Green Data Collection from Wirelessly-powered Sensors

Reliable and energy-efficient data collection from resource-limited sensors is essential to the success of future IoT. In this chapter, we study the energy consumption minimization problem during the data collection from a generic wirelessly-powered sensor. Specifically, we determine the optimal data collection parameters, in terms of charging duration and charging power as well as sensor transmission rate, in real time according to the instantaneous channel condition while satisfying a certain latency constraint. For the scenario of ideal rate adaptive transmission with linear energy harvesting, we derive closed-form expressions for optimal transmission parameters. We also establish the condition on channel quality for successful data collection within a latency constraint. For the more practical case of finite block-length transmission with non-linear energy harvesting, we develop a DRL solution for efficient online implementation. We also propose an online tuning scheme to cater for model inaccuracy and environment variation. The accuracy and effectiveness of our proposed approaches are verified by comparing with benchmark schemes. Our DRL-based approach has broad applicability and can solve other real-time optimal design problems in wireless communications.

3.1 Introduction

Many IoT applications, e.g. smart city and intelligent manufacture, rely on reliable and timely data collection from numerous wireless sensors [87]. These sensors are expected to operate, often in remote and/or hard-to-reach area, for several years without human inter-

vention. Wireless power transfer (WPT) is one of the most convenient solution to supply energy to these sensors and power their autonomous operation [88-92]. With WPT, the data collecting agent first transfer energy to the sensor by emitting RF signal. Then the sensor will transmit its collected data to the agent using harvested energy. Given the large amount of sensors in the field, it is of great interest to achieve reliable and highly energy efficient data collection from them. Note that the data-collecting agents may also be energy constrained, as they could be powered by compact battery mounted on unmanned aerial vehicles (UAV) [93]. In this paper, we design optimal strategy for data collection from individual wirelessly-powered sensors to minimize the total energy consumption while satisfying a latency requirement.

Over the past several years, some novel data collection schemes/protocols were proposed to improve the energy efficiency of wireless sensor networks (WSNs), e.g. [94-100]. The work in [94] proposed an energy efficient data collection scheme based on denoising autoencoder (DCDA) for WSNs, with which lower energy consumption was achieved as compared to conventional schemes. The work in [95] proposed an energy-aware path construction (EAPC) scheme for mobile data sink to collect data from a preselected set of sensors in WSNs, where network energy consumption was significantly reduced. The work in [96] proposed a novel protocol, based on mobile fog computing, for WSNs to enhance the energy efficiency and reliability of data collection. The work in [97, 98] optimized the clustering strategy to suppress the energy consumption of network. The work in [99, 100] utilized the method of adaptive data collection to suppress the energy consumption. However, although these methods above can effectively diminish the energy consumption of WSNs, it is still very hard to collect data from numerous sensors in some remote and/or hard-to-reach regions for ground objects, e.g. BS, moving data sink and so on. While noting that unmanned aerial vehicle (UAV) has low cost and great deployment flexibility, it is a promising solution to wireless data collection from IoT sensors at those remote locations mentioned above. Also, UAV can employ WPT to power the IoT sensors for transmitting their data. As such, UAV enabled wireless data collection has attracted a lot of attentions from both academia and industry. Most recent work on such scheme mainly focused on trajectory optimization and hover scheduling, such as [93, 101–104]. Specifically, The work in [93] optimized the trajectory and hovering locations of multiple UAVs to efficiently charge IoT devices through magnetic resonance-coupled power transfer. The work in [101] considered a UAV-assisted sequential sensor charging and data collection system and formulates a trajectory planning problem to minimize the average Age of Information (AoI) of collected data. The work in [102] solved for the optimal hovering location and duration

and then the flying route of a rotary-wing UAV to maximize the lifetime of the wireless sensor network. The work in [103, 104] considered the scenario that a UAV is serving multiple IoT sensors and minimize its energy consumption by trajectory optimization and time allocation for flying and hovering, respectively. To the best of our knowledge, there is little work on energy consumption minimization for individual sensor charging and data collection session.

On the other hand, IoT sensors typically transmit a small amount of data during each data collection session. Due to the short block-length, error free transmission at ergodic capacity may not be feasible [105]. In addition, most practical energy harvesting circuits typically exhibit non-linear characteristics [106]. We then generalize the energy consumption minimization problem to the practical scenario of finite block-length transmission with non-linear energy harvesting. The resulting optimization problem becomes non-convex, which prohibits analytical solution. While iterative algorithms can be developed to gradually approach sub-optimal/optimal solutions [107, 108], such solutions are not suitable for real-time implementation, as the algorithms may not converge even to a sub-optimal solution within the latency constraint. In this work, we develop a deep reinforcement learning (DRL) based solution that can determine near-optimal parameter configuration in real time during the online operation.

Machine learning technique has been applied into wireless system design [65, 66, 69, 73]. The key advantage of machine learning solutions is the capability of extracting and exploring hidden relationship from experience data, when accurate system model is not available [109]. For physical wireless transmissions, however, rather accurate models of the transmission medium and transceiver structure have already been developed, whereas the amount of real-world experience is limited. In this work, we propose a novel approach to apply data-driven machine learning algorithms to solve real-time optimal design for data collection from wirelessly powered sensors. Specifically, we reformulate the energy consumption minimization problem into a one-step Markov decision process (MDP) and train a deep policy network that determines the optimal transmission strategy for a given system state. Considering the continuous state/action spaces for the MDP formulation, we apply policy gradient algorithm to train the policy network offline with simulated data collection sessions.

During the online operation, the trained policy network can quickly generate near optimal transmission parameters without any iterative calculation. Meanwhile, due to modelling inaccuracy and/or environment variation, the model used for offline training may not accurately reflect real world operating environment. With limited online experience avail-

able, it is impractical to develop new systems model or to adjust specific model parameters. To maintain near-optimal online operation, we propose an online tuning algorithm for the policy network through applying model-free DRL algorithms, such as actor-critic and DDPG [60, 61]. In particular, we introduce a critic network to approximate the energy harvesting/data transmission model and track its variation with sample gradient calculated using online experience. The policy network is then fine-tuned using the gradient calculated with the updated critic network. The resulting solution can overcome slight model inaccuracy and track gradual environment variation to achieve near optimal online performance.

3.2 Ideal adaptive rate transmission with linear energy harvesting

In this section, we mainly analyze and minimize the energy consumption of data collection system under ideal adaptive rate transmission with linear energy harvesting.

3.2.1 System model and problem formulation

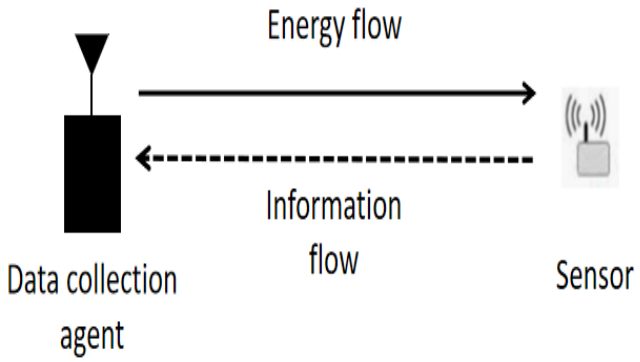

Figure 3.1: Data collection from wirelessly-powered sensor.

We consider a sensor data collection system as shown in Fig. 3.1. Particularly, a data collection agent first transfers energy to an individual sensor wirelessly with its emitted

RF signal for a certain duration τ_C. We assume a slow frequency flat fading channel between the agent and the sensor, with effective channel power gain (after considering path loss/shadowing/fading effects and potential beamforming/combining operations) denoted by g. Note that in terms of multi-antenna agent, analog beamforming is needed for enhancing the performance of energy harvesting and maximum ratio combining is needed for enhancing the performance of information reception. Assuming linear energy harvesting model, the collected energy at the sensor is given by $P_S g \eta \tau_C$ J, where P_S is the transmitted power of the agent and η denotes the DC energy conversion efficiency of harvesting circuit at the sensor. Such linear model can well approximate real harvested energy when charging power is in the linear region of practical nonlinear energy harvesting model [106].

After energy harvesting, the sensor will send its collected data of H bits to the agent using all of its harvested energy. Let τ_I denote the information transmission duration. The transmit power of the sensor is given by $\frac{P_S g \eta \tau_C}{\tau_I}$. Assuming ideal rate adaptive transmission based on the prevailing channel condition, While noting that channel power gain is the same for charging and data transmission, the information transmission duration τ_I should satisfy

$$H = \tau_I B \log_2(1 + \frac{P_S g \eta \tau_C}{\tau_I} \frac{g}{\sigma^2}), \tag{3.1}$$

where we denote the channel bandwidth and average noise power by B and σ^2, respectively. Here we assume that sensor charging and data transmission occur over the same frequency band, and as such, channel reciprocity applies. While error free transmission with Shannon capacity is very difficult to achieve in practice due to short transmission duration, we adopt it here to establish the energy consumption lower limit for individual data collection sessions. To avoid the outdatedness of channel gain information, the total duration of sensor charging and information transmission should be less than a channel coherence time T_C, i.e. $\tau_C + \tau_I \leq T_C$.

Our goal is to minimize the energy consumption of the agent during the data collection at the sensor through optimal design. Neglecting the circuit power consumption of the agent, the total energy consumption is equal to $E = P_S \tau_C$. The energy consumption

minimization problem can be formulated, with the application of Eq. (3.1), as

$$\begin{aligned} \min_{P_S,\ \tau_I} \quad & E = P_S \tau_C = \frac{\tau_I \sigma^2}{g^2 \eta}(\exp(\frac{H \ln 2}{B \tau_I}) - 1), \\ s.t. \quad & \frac{\tau_I \sigma^2}{P_S g^2 \eta}(\exp(\frac{H \ln 2}{B \tau_I}) - 1) + \tau_I \leq T_C; \\ & 0 < P_S \leq P_{\max},\ 0 < \tau_I. \end{aligned}$$

Since the latency constraint is not jointly convex with respect to P_S and τ_I, we can not apply KKT condition to derive global optimal solution for this problem. While iterative algorithms for numerically searching optimal parameter values can be devised, they may not be suitable for online implementation at the agent. This is because that the time consumption for iterative calculation may violate the latency constraint. In the following section, we derive the analytical expressions of the optimal P_S and τ_I in terms of minimizing agent energy consumption, which can be used to calculate optimal τ_C.

3.2.2 Analytical solution

Noting that the optimization problem may have no feasible solution due to the latency constraint, we first derive the condition for solution existence, from which we can analyze the probability of data collection failure during a particular channel coherence time. After that, we will derive the analytical expressions for optimal P_S, τ_I, and τ_C. We will conclude the section with some discussions on how these results can apply during online operation.

Condition for solution existence

We first transform the latency constraint equivalently to

$$(\frac{P_S g^2 \eta}{\sigma^2}(\frac{T_C}{\tau_I} - 1) + 1) \exp(-\frac{\tilde{H}}{\tau_I}) \geq 1, \tag{3.2}$$

where we define the constant $\tilde{H} = H \ln 2 / B$. Then, we convert the left hand side to the form of $x \exp(x)$ by multiplying both sides with appropriate terms as

$$(-\frac{\tilde{H}}{\tau_I} + \frac{\tilde{H}}{T_C}(1 - \frac{\sigma^2}{P_S g^2 \eta})) \exp(-\frac{\tilde{H}}{\tau_I} + \frac{\tilde{H}}{T_C}(1 - \frac{\sigma^2}{P_S g^2 \eta})) \leq -\frac{\tilde{H} \sigma^2}{P_S g^2 \eta T_C} \exp(\frac{\tilde{H}}{T_C}(1 - \frac{\sigma^2}{P_S g^2 \eta})). \tag{3.3}$$

Note that the minimum value of $x\exp(x)$ for any real variable x is $-1/e$. As such, if we treat $-\frac{\tilde{H}}{\tau_I}+\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2\eta})$ as x, the right hand side of Eq. (3.3) should be greater than $-1/e$ for the latency constraint to be satisfiable, which leads to

$$-\frac{\tilde{H}\sigma^2}{P_S g^2\eta T_C}\exp(-\frac{\tilde{H}\sigma^2}{P_S g^2\eta T_C}) > -\exp(-1-\frac{\tilde{H}}{T_C}). \tag{3.4}$$

Applying the definition of Lambert W function [85], while noting that it has two function values when the independent variable is negative, the above inequality can be shown to be equivalent to

$$-\frac{\tilde{H}\sigma^2}{P_S g^2\eta T_C} > W_0[-\exp(-1-\frac{\tilde{H}}{T_C})] \text{ or } -\frac{\tilde{H}\sigma^2}{P_S g^2\eta T_C} < W_{-1}[-\exp(-1-\frac{\tilde{H}}{T_C})], \tag{3.5}$$

where $W_0[\cdot]$ denotes the principle branch and $W_{-1}[\cdot]$ the negative brach of Lambert W function. After some manipulation, we arrive at the following necessary condition for the existence of feasible solution in terms of channel power gain g as

$$\begin{aligned} g &> \sqrt{\frac{-\tilde{H}\sigma^2}{P_S\eta T_C W_0[-\exp(-1-\tilde{H}/T_C)]}} \quad \text{or} \\ g &< \sqrt{\frac{-\tilde{H}\sigma^2}{P_S\eta T_C W_{-1}[-\exp(-1-\tilde{H}/T_C)]}}. \end{aligned} \tag{3.6}$$

Under the necessary condition, the latency constraint can be rewritten, after applying the Lambert W function to Eq. (3.3), into

$$\begin{aligned} -W_0[-\frac{\tilde{H}\sigma^2}{P_S g^2\eta T_C}\exp(\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2\eta}))] &\le \frac{\tilde{H}}{\tau_I}-\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2\eta}) \\ &\le -W_{-1}[-\frac{\tilde{H}\sigma^2}{P_S g^2\eta T_C}\exp(\frac{\tilde{H}\ln 2}{T_C}(1-\frac{\sigma^2}{P_S g^2\eta}))]. \end{aligned} \tag{3.7}$$

Noting that τ_I has to be positive, we arrive at the following sufficient condition for the existence of feasible solution

$$W_{-1}[-\frac{\tilde{H}\sigma^2}{P_S g^2\eta T_C}\exp(\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2\eta}))] < \frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2\eta}). \tag{3.8}$$

Proposition 1. *The sufficient condition in Eq. (3.8) is satisfied if and only if* $\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2\eta}) > -1$

Proof. Let us consider the cases of $\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta})$ greater and less than or equal to -1 separately. Since $W_{-1}[\cdot]$ is less than -1, Eq. (3.8) always holds when $\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}) > -1$. When $\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}) \leq -1$, noting that $W_{-1}(x\exp(x)) = x$ for $x < -1$ by definition, we can write $\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta})$ as $W_{-1}[\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta})\exp(\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}))]$, which will always be less than $W_{-1}[-\frac{\tilde{H}\sigma^2}{P_S g^2 \eta T_C}\exp(\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}))]$ as $W_{-1}[\cdot]$ is a monotonically decreasing function. This completes the proof. □

The sufficient condition in Eq. (3.8) can be equivalently rewritten as

$$g > \sqrt{\frac{\tilde{H}\sigma^2}{P_S \eta T_C (1+\tilde{H}/T_C)}}. \tag{3.9}$$

Considering together with the necessary condition given in Eq. (3.6) and noting that

$$-W_0[-\exp(-1-\frac{\tilde{H}}{T_C})] < 1+\frac{\tilde{H}}{T_C} < -W_{-1}[-\exp(-1-\frac{\tilde{H}}{T_C})], \tag{3.10}$$

we arrive at the condition for feasible solution existence as

$$g > \sqrt{\frac{-\tilde{H}\sigma^2}{P_{\max} \eta T_C W_0[-\exp(-1-\tilde{H}/T_C)]}}, \tag{3.11}$$

where the peak transmission power of the agent is used. With applicable distribution function of the channel power gain, we can evaluate the probability that the charging and data collection cannot finished within one channel coherence time and therefore lead to data collection failure.

Optimal solution

Assuming that the channel power gain is sufficiently large to guarantee feasible solution, we now derive the analytical expressions of the optimal transmission parameters. We first note that the objective function E is monotonically decreasing with τ_I, since its derivative with respect to τ_I is always negative. As such, τ_I should be designed as large as possible, while satisfying the latency constraint. Let us now consider again the equivalent latency constraint in Eq. (3.7). In particular, we first prove the following proposition.

Proposition 2. *If* $\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}) > -1$*, then we have*

$$W_0[-\frac{\tilde{H}\sigma^2}{P_S g^2 \eta T_C}\exp(\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}))] < \frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}). \tag{3.12}$$

Proof. Note that based on definition, $W_0(x\exp(x)) = x$ for $x > -1$. We can write $\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta})$ equivalently as $W_0[\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta})\exp(\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}))]$. Note the independent variable of $W_0[\cdot]$, i.e. $\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta})\exp(\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}))$, must be greater than that of $W_0[\cdot]$ on the left hand side of Eq. (3.12), i.e. $-\frac{\tilde{H}\sigma^2}{P_S g^2 \eta T_C}\exp(\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}))$. The proof is completed while noting that $W_0[\cdot]$ is a monotonically increasing function. □

Based on this proposition, we can show that the upper bound of τ_I while satisfying the latency constraint is given by

$$\tau_I \leq \hat{\tau}_I = (-\frac{1}{\tilde{H}}W_0[-\frac{\tilde{H}\sigma^2}{P_S g^2 \eta T_C}\exp(\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}))] + \frac{1}{T_C}(1-\frac{\sigma^2}{P_S g^2 \eta}))^{-1}. \tag{3.13}$$

It can be shown, by taking derivative of $\hat{\tau}_I$ with respect of P_S and checking the sign of the result, that $\hat{\tau}_I$ is an increasing function of transmit power P_S. Note that since $\hat{\tau}_I$ must be greater than zero when P_S ranges from 0 to $P_{\max}$, equivalent latency constraint in Eq. (3.13) can be always satisfied. As such, we can determine the optimal parameter values to minimize the energy consumption of the agent during the sensor charging and data collection as

$$P_S^* = P_{\max}, \tag{3.14}$$

and

$$\tau_I^* = (-\frac{1}{\tilde{H}}W_0[-\frac{\tilde{H}\sigma^2}{P_{\max} g^2 \eta T_C}\exp(\frac{\tilde{H}}{T_C}(1-\frac{\sigma^2}{P_{\max} g^2 \eta}))] + \frac{1}{T_C}(1-\frac{\sigma^2}{P_{\max} g^2 \eta}))^{-1}. \tag{3.15}$$

Correspondingly, the optimal charging duration and minimum energy consumption can be determined as

$$\tau_C^* = \frac{\exp(\tilde{H}/\tau_I^* - 1)\tau_I^* \sigma^2}{P_{\max} g^2 \eta}, \tag{3.16}$$

and

$$E_{\min} = \frac{\exp(\tilde{H}/\tau_I{}^* - 1)\tau_I{}^* \sigma^2}{g^2 \eta}, \tag{3.17}$$

respectively

Implementation consideration

With the analytical results derived in this section, the operation of the agent during remote data collection with wireless energy transfer can be implemented as follows. First of all, the agent will estimate channel power gain to the sensor during the current channel coherent time. If the channel power gain is large enough to guarantee a feasible solution, the agent will start charging the sensor with its peak power for a duration of $\tau_C{}^*$ given in Eq. (3.16). The sensor will then transmit its data to the agent with all harvested energy for a duration of $\tau_I{}^*$ with ideal rate adaptation. If the channel power gain is too low to guarantee a feasible solution, the agent will wait for a channel coherence time duration and re-estimate the channel quality. This process can repeat for multiple coherence time intervals permitted by the application scenario until the data is successfully collected. With such implementation strategy, we can achieve the highest possible energy utilization efficiency during data collection at individual sensors.

3.2.3 Effect of circuit power consumption

In this section, we consider the effect of circuit power consumption of the agent during sensor charging and data reception. Let P_C denote the circuit power consumption of the agent. The total energy consumption becomes $E = (P_S + P_C)\tau_C + P_C\tau_I$. The energy consumption minimization problem is updated to

$$\begin{aligned} \min_{P_S,\, \tau_I} \quad & E = (1+\frac{P_C}{P_S})\frac{\tau_I\sigma^2}{g^2\eta}(\exp(\frac{H\ln 2}{B\tau_I})-1)+P_C\tau_I, \\ s.t. \quad & \frac{\tau_I\sigma^2}{P_S g^2\eta}(\exp(\frac{H\ln 2}{B\tau_I})-1)+\tau_I \le T_C; \\ & 0 < P_S \le P_{\max},\ 0 < \tau_I. \end{aligned}$$

We first note that both total energy consumption and charging duration decrease as P_S increases. Therefore, the optimal P_S should should set to its peak value $P_{\max}$. We can also show that objective function is strictly convex with respect to τ_I and that the τ_I value that minimizes the objecting function is given by

$$\tilde{\tau}_I = \frac{\tilde{H}}{W_0[e^{-1}(P_C P_{\max} g^2 \eta/\sigma^2/(P_C + P_{\max}) - 1)] + 1}. \tag{3.18}$$

From the latency constraint, we arrive at the same existence condition of feasible solution, given in Eq. (3.11). Under this condition, we can show τ_I is bounded by $\tau_I^\dagger \leq \tau_I \leq \tau_I^*$, where $\tau_I^\dagger$ is defined as

$$\tau_I^\dagger = \tilde{H}(\frac{\tilde{H}}{T_C}(1 - \frac{\sigma^2}{P_{\max} g^2 \eta}) - W_{-1}[-\frac{\tilde{H}\sigma^2}{P_{\max} g^2 \eta T_C} \exp(\frac{\tilde{H}}{T_C}(1 - \frac{\sigma^2}{P_{\max} g^2 \eta}))])^{-1}, \tag{3.19}$$

and τ_I^* was the optimal result without considering circuit power consumption, given in Eq. (3.15).

Finally, the optimal transmission time with the consideration of circuit power consumption, while satisfying the latency constraint, is determined as

$$\widetilde{\tau_I}^* = \max\{\tau_I^\dagger,\ \min\{\tilde{\tau}_I,\ \tau_I^*\}\}. \tag{3.20}$$

Corresponding optimal charging duration $\widetilde{\tau_C}^*$ and the resulting minimum energy consumption can be calculated accordingly.

3.2.4 Numerical results

Table 3.1: System parameters of sensor data collection.

Notation	Parameter names	Values
B	Channel bandwidth	200 KHz
η	Energy conversion efficiency	0.8
σ^2	Noise power	10^{-5} mWatt
$P_{\max}$	Peak charging power	1 Watt
P_C	Circuit power consumption	10 mWatt
a	Non-linear energy harvesting constant	1
b	Non-linear energy harvesting constant	1
P_H	Peak harvesting power	1 Watt

In this section, we present some selected numerical examples to illustrate the analytical results. Unless otherwise indicated, we use system parameter values listed in Table I. Note that these parameters will be used in section 3.2.4 and section 3.3.

We first study the likelihood that the sensor charging and data collection can not successfully finish within the latency requirement T_C due to poor channel quality. This proba-

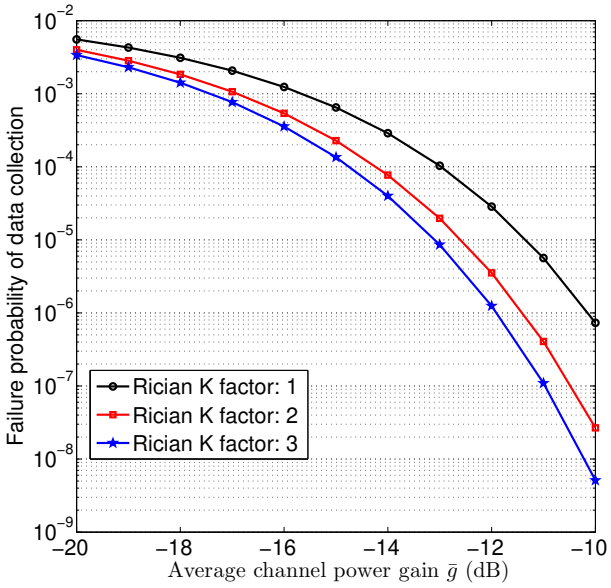

Figure 3.2: Probability of data collection failure under a latency constraint $T_C = 0.02$ s.

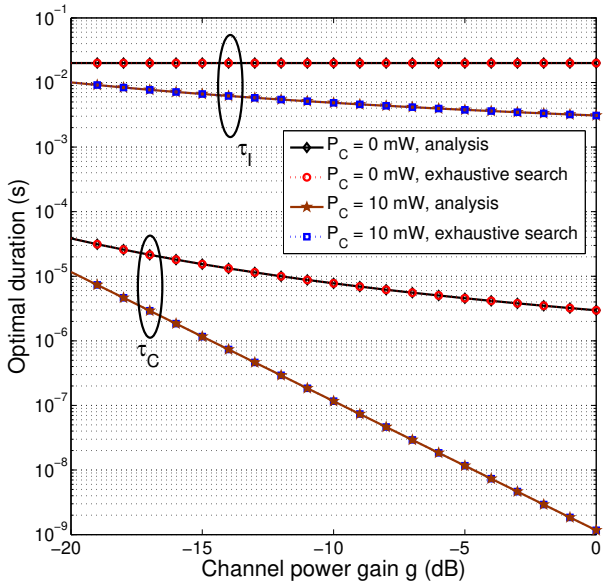

Figure 3.3: Optimal sensor charging and data transmission durations, $T_C = 0.02$ s, $H = 10000$ bits.

bility can be obtained by evaluating the distribution function of channel power gain g at the threshold level given in Eq. (3.11). Fig. 3.2 plots the probability of data collection failure as a function of average channel power gain over Rician fading environment [82]. We can see that, as expected, the failure probability decreases as the average channel quality improves. In addition, the larger the Rician factor K, the lower the failure probability, as the wireless power transfer enjoys the channel with stronger line-of-sight component.

Fig. 3.3 presents the optimal durations for sensor charging and data transmission, calculated using the analytical results, as the functions of channel power gain g for different circuit power levels. The corresponding results obtained from exhaustive search are also plotted. The perfect match validates our analytical approach. We can see that when the circuit power consumption is negligible ($P_C = 0$ mW), the optimal duration of data transmission τ_I slightly increases with increasing g while the optimal charging duration τ_C decreases dramatically. Note that in this case, the energy consumption is only dependent on τ_C. The energy consumption minimization will make τ_C as small as possible while ensuring the data transmission completes within a channel coherence time. The behaviour becomes different when circuit power consumption is not negligible ($P_C = 10$ mW), where the duration of sensor charging and data transmission decreases when the channel condition improves.

As an additional numerical example, Fig. 3.4 plots the minimum energy consumption as

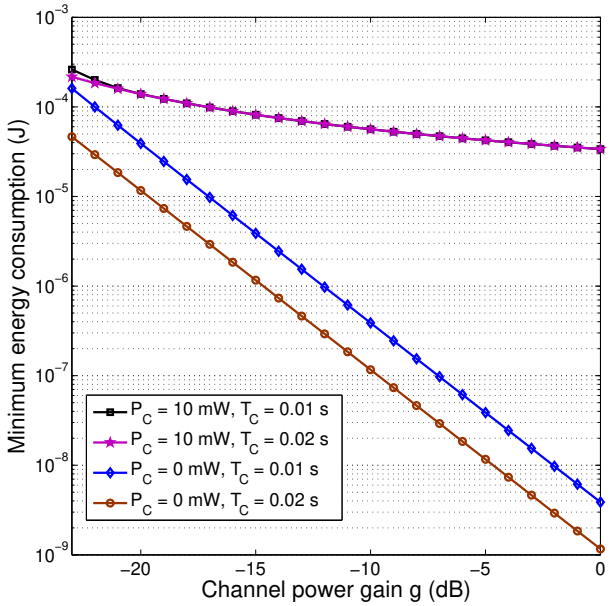

Figure 3.4: Minimum energy consumption under different latency constraints with and without considering circuit power consumption, $H = 10000$ bits.

a function of channel power gain g for different circuit power levels and latency constraints. We can see that the energy consumption decreases when channel power gain g increases, as expected. We also note that a higher circuit power level leads to higher energy consumption even with optimized parameters. Finally, the energy consumption generally decreases as the latency requirement T_C increases. Meanwhile, when the circuit power is not negligible, T_C has limited effect on the energy consumption unless the channel quality is very poor.

3.3 Generalization with finite block-length transmission with nonlinear energy harvesting

In this section, we consider the more practical scenario of finite block-length transmission with nonlinear energy harvesting. Specifically, we propose a DRL-based solution to arrive at a policy network for minimizing the effective energy consumption of such generic system. Also, well-trained policy network can be further tuned using online experience to better cater for small model inaccuracy and environment variation.

3.3.1 Generalized formulation

The data collecting agent again transfers energy to the sensor by emitting RF signal with power P_S for a duration of τ_C. Here, to better match the real application scenarios, we adopt a nonlinear energy harvesting model [106]. Such nonlinear model can well approximate practical energy harvesting process when charging power ranges from zero to infinity. In particular, the harvested energy at the sensor over the duration of τ_C is given by

$$E_H = P_H\left(\frac{1+\exp(-ab)}{1+\exp(-a(P_S g - b))} - \exp(-ab)\right)\tau_C, \tag{3.21}$$

where P_H is the peak harvesting power, a, and b are constants depending upon the EH circuit. The sensor again uses all of its harvested energy for data transmission. We assume finite block-length transmission from the sensor, as the data amount is typically small. Note that τ_I will depend upon channel use number in terms of this assumed condition. If the sensor uses n channel use, while noting that the duration of each channel use is equal to the inverse of the channel bandwidth, i.e. $1/B$ [110], the duration of data transmission is approximately equal to $\tau_I = n/B$. This is different from the ideal model in section 3.2. The

received SNR at the agent during data transmission is given by

$$\gamma = \frac{E_H}{n/B}\frac{g}{\sigma^2}. \tag{3.22}$$

The corresponding block error rate for transmitting H bits of sensor data is given by [105]

$$\varepsilon = Q\left(\frac{\log_2(1+\gamma) - H/n}{\sqrt{\frac{\gamma(\gamma+2)}{2n(1+\gamma)^2}}\log_2 e}\right). \tag{3.23}$$

The energy consumed, during a data collection session with the consideration of circuit power P_C, is equal to $E_C = P_S\tau_C + P_C(\tau_C + n/B)$. Due to the possible block error, the data collection may be unsuccessful, rendering the energy wasted. To account for such situation, we introduce the concept of *effective energy consumption*, which is defined as the ratio of the energy consumption of a data collection session over the probability of successful collection, i.e. $E_C/(1-\varepsilon)$. For each data collection session with specific instantaneous channel realization, we optimally select the transmission parameters, including transmission power P_S, charging duration τ_C, and the number of channel use n for sensor data transmission, to minimize effective energy consumption. To avoid the outdatedness of channel knowledge and/or to satisfy a latency constraint, we impose the latency constraint of T_C. As such, while assuming that achieved effective data rate is less than Shannon capacity, we arrive at the following optimization problem

$$\begin{aligned} \min_{P_S,\ \tau_C,\ n} \quad & \frac{P_S\tau_C + P_C(\tau_C + n/B)}{1-\varepsilon}, \\ s.t. \quad & 0 < P_S \le P_{\max},\ 0 < \tau_C,\ 0 < n, \\ & 0 < \tau_C + n/B \le T_C. \end{aligned} \tag{3.24}$$

Note that the objective function is a complex function of the optimization variables, as ε depends on n, P_S, and τ_C either directly or through the received SNR γ. The mixed-integer nonlinear nature of this problem make closed-form solution unfeasible, especially when objective function is not jointly convex with respect to all continuous variables. While an iterative algorithm can be developed to approach the optimal solution, executing the iterative algorithm for every data collection session during online operation will consume the valuable computing/power resource of the agent. Furthermore, the iterative algorithm may not converge before the latency constraint expires. In what follows, we develop a deep reinforcement learning solution to determine the near optimal transmission parameters for

each transmission session in real time with minimal computation.

3.3.2 Deep reinforcement learning formulation

Our goal is to build an intelligent agent that can determine near-optimal transmission parameters to minimize the effective energy consumption for each data collection session. Considering the very complex relationship between these parameters, we resort to deep neural networks (DNN) for its universal function approximation capability and develop a deep reinforcement learning algorithm to effectively train such network. In particular, we reformulate the parameter optimization problem into the one-step Markov decision process (MDP) with continuous state and action spaces. The state space $\mathscr{S}$ is defined by the vector $\vec{s} = [H, g, T_C]^T$, which may vary from one session to another. We assume that the collecting agent obtains accurate state information through channel estimation and previous experience. The action space $\mathscr{A}$ is defined by the vector $\vec{a} = [P_S, \tau_C,\ n]^T$. Note that while n is an integer, both P_S and τ_C are continuous. Since our goal is to minimize the effective energy consumption for current data collection session, we define the immediate reward as its inverse, i.e.

$$R = \frac{1-\varepsilon}{P_S\tau_C + P_C(\tau_C + n/B)}. \tag{3.25}$$

Note that data collection sessions are independent with each other and hence, the MDP will terminate after each state-action-reward tuple $\{\vec{s}, \vec{a}, R\}$, which leads to a one step MDP.

We will train a DNN to approximate an optimal policy in terms of maximizing the reward for each data collection session. Let π denote the policy network, parameterized by θ, as shown in Fig. 3.5. With state $\vec{s} = [H, g, T_C]^T$ given, the output action $\vec{a} = [P_S,\ \tau_C,\ n]^T$ from policy network is given by

$$\vec{a} = \pi(\vec{s}|\theta). \tag{3.26}$$

Considering the continuous state/action spaces, we apply policy gradient method to reach a deterministic policy. For that purpose, we need to determine the gradient of the action policy, or equivalently the gradient for immediate reward R with respect to network parameter θ, i.e. $\bigtriangledown_\theta R(\vec{s},\vec{a})|_{\vec{a}=\pi(\vec{s}|\theta)}$. Applying the chain rule, such gradient can be calculated as

$$\bigtriangledown_\theta R(\vec{s},\vec{a})|_{\vec{a}=\pi(\vec{s}|\theta)} = \bigtriangledown_{\vec{a}} R(\vec{s},\vec{a})\, \bigtriangledown_\theta\, \pi(\vec{s}|\theta). \tag{3.27}$$

With the energy harvesting/transmission models introduced earlier, we can analytically calculate $\bigtriangledown_{\vec{a}} R(\vec{s},\vec{a})$. The gradient $\bigtriangledown_\theta \pi(\vec{s}|\theta)$ can be obtained using back-propagation tech-

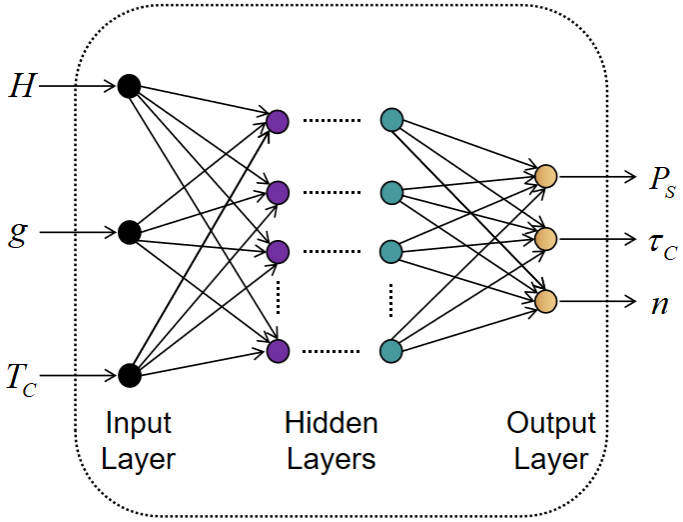

Figure 3.5: Deep policy network for data collection

nique. Then, we can update the network parameters along the gradient direction. Additionally, such policy network can be split into three small networks during training process, where each small network corresponds to an action output and θ can denote combined parameter set for three networks. Corresponding parameters for small networks still can be updated along the direction of joint gradient operation for them. After training is completed, three well-trained networks can be combined into a policy network again.

Offline training

We now explain the offline training process in further details. First of all, to improve sample efficiency, we adopt off-policy learning with replay buffer [141, 142]. Specifically, we simulate various data collection scenarios, with randomly-generated channel power gains, typical data amounts, and latency requirements. Then we feed the state vector $\vec{s}_i$ into the current policy network, which generates corresponding action vector $\vec{a}_i$. Finally, the resulting reward R_i can be calculated using the formula presented earlier. The resulting experience tuple of $\{\vec{s}_i,\ \vec{a}_i,\ R_i\}$ is saved into a replay buffer. When the replay buffer reaches its capacity, a randomly selected existing experience tuple will be replaced. The size of the replay buffer should be sufficiently large to guarantee experience diversity, but not extremely large to ensure reasonably probability of selecting the recent experience tuples.

In addition, we apply mini-batch training with soft update to mitigate training divergence [142]. In particular, a mini-batch of N experience tuples will be randomly extracted from the replay buffer and used to calculate the gradient for parameter update. The resulting sample gradient is given by

$$\bigtriangledown_\theta R(\vec{s},\vec{a})|_{\vec{a}=\pi(\vec{s}|\theta)} \approx \frac{1}{N}\sum_{n=1}^{N} \bigtriangledown_{\vec{a}} R(\vec{s},\vec{a})|_{\vec{s}=\vec{s}_n,\ \vec{a}=\pi(\vec{s}_n|\theta)} \bigtriangledown_\theta \pi(\vec{s}_n|\theta). \tag{3.28}$$

The parameters of the policy network will be updated to

$$\theta \leftarrow \theta + \xi \bigtriangledown_\theta R(\vec{s},\vec{a})|_{\vec{a}=\pi(\vec{s}|\theta)}. \tag{3.29}$$

where $\xi \in (0,1)$ denotes the learning rate.

We introduce random exploration to reduce the probability of converging to suboptimal solutions, which may occur in policy gradient method. Specifically, a zero mean Gaussian random variable is added into each entry of the action vector generated by current policy network [109]. Finally, to ensure the feasibility and validity for the output of policy network, in terms of satisfying peak power and latency constraints, we apply hard thresholds

to corresponding actions. For example, the charging power P_S during training is calculated using the first entry of policy network output, $\pi(\vec{s}|\theta)[1]$, as

$$P_S = \max\{0,\ \min\{\pi(\vec{s}|\theta)[1] + v_P,\ P_{\max}\}\}, \tag{3.30}$$

where v_P is a zero-mean Gaussian random variable with variance σ_P^2. At the end of each training iteration, the variance σ_P^2 is updated to $\beta\sigma_P^2$, where β is a constant slightly smaller than 1 to ensure sufficient exploration. Similarly, the charging duration and the number of channel use n during training is calculated as

$$\tau_C = \max\{0,\ \min\{\pi(\vec{s}|\theta)[2] + v_\tau,\ T_C\}\}, \tag{3.31}$$

and

$$n = \lfloor \max\{1,\ \min\{\pi(\vec{s}|\theta)[3] + v_n,\ (T_C - \tau_C)/B\}\}\rfloor, \tag{3.32}$$

respectively. Here v_τ and v_n are zero-mean Gaussian random variables with σ_τ^2 and σ_n^2, respectively. Again, σ_τ^2 and σ_n^2 will be reduced by a factor of β after each iteration.

The pseudo code of the proposed offline training algorithm is shown in Algorithm 4. When the number of iterations T is sufficient large, the exploration variances will eventually approach zero. The policy network will converge to a deterministic policy. The most computational intensive step of Algorithm 4 is the calculation of sample gradient, which can be efficiently implemented using introduced models earlier and back-propagation technique. As such, the proposed offline training solution has low computational complexity. The network may still converge to a suboptimal policy even with the random exploration. We can solve this problem by performing several training sessions and pick the network with the highest average reward after convergence. As shown in the numerical example section, at most three training sessions will lead to a policy network with performance similar to that of exhaustive search.

Online tuning

During each online data collection session, the collecting agent will observe environment state information and feed them to the offline-trained policy network, which will output a near-optimal action vector. Note that no iterative calculation is involved and so, the action vector can be quickly obtained. If the energy harvesting/data transmission models used during offline training match real-world environment perfectly, the data collection with action vector determined by well-trained policy network will achieve minimum effective

Algorithm 1 The pseudo code of offline training

Initialize policy network with random parameter θ.
Initialize exploration variances σ_P^2, σ_τ^2 and σ_n^2.
Set up an empty replay buffer that can hold M state-action-reward tuples.
for t $\in [1,\ 2,\ \cdots,\ T]$ **do**
  Generate state vector $\vec{s}_t$.
  Determine action vector $\vec{a}_t$ by feeding $\vec{s}_t$ to current policy network and apply random exploration/thresholding to the output as Eq. (3.30), Eq. (3.31) and Eq. (3.32).
  Calculate reward R_t using Eq. (3.25) together with Eq. (3.23), Eq. (3.22), and Eq. (3.21).
  Add new state-action-reward tuple $\{\vec{s}_t, \vec{a}_t, R_t\}$ to the replay buffer, and remove randomly an existing tuple if necessary.
  Extract N state-action-reward tuples randomly from replay buffer and calculate sample gradient using Eq. (3.28).
  Update the policy network parameter θ as $\theta \leftarrow \theta + \xi \bigtriangledown_\theta R(\vec{s},\vec{a})|_{\vec{a}=\pi(\vec{s}|\theta)}$.
  Update exploration variances by multiplying β.
end for

energy consumption. If there exists certain modelling inaccuracy and/or environment variation over time, the offline-trained policy network needs to be further adjusted using online experience.

With online experience, We can adjust the policy network π by again performing policy gradient. Since the energy harvesting/data transmission models may be inaccurate, we can on longer use them to calculate the gradient. Building new models with online experience will be inefficient, as the number of experience is typically limited. Inspired by the idea of model-free DRL algorithms, such as actor-critic and DDPG [60, 61], we establish a DNN that predicts the instantaneous reward for a given state vector and action vector pair. We denote the resulting critic network by $Q(\vec{s}_t, \vec{a}_t|\mu)$ with parameter set μ, as shown in Fig. 3.6. Note that since the gradient operation of offline reward function may be NAN in terms of some training data, the critic network is also employed to approximate the reward function during offline training. Through properly supervised learning using offline experience, the resulting critic network will closely approximate the relationship of Eq. (3.25) together with Eq. (3.23), Eq. (3.22), and Eq. (3.21), and as such, capture offline training models.

During online operation, we will adjust the parameters of the critic network using online experience to improve its accuracy. Particularly, we propose to perform online tuning using a mini-batch of N' online experience tuples with soft update. In particular, after a mini-batch of N' online experience tuples $\{\vec{s}_i, \vec{a}_i, R_i\}$, $i = 1,2,3,\cdots,N'$, are available, the agent will adjust the parameter of the critic network μ by performing gradient descent for the

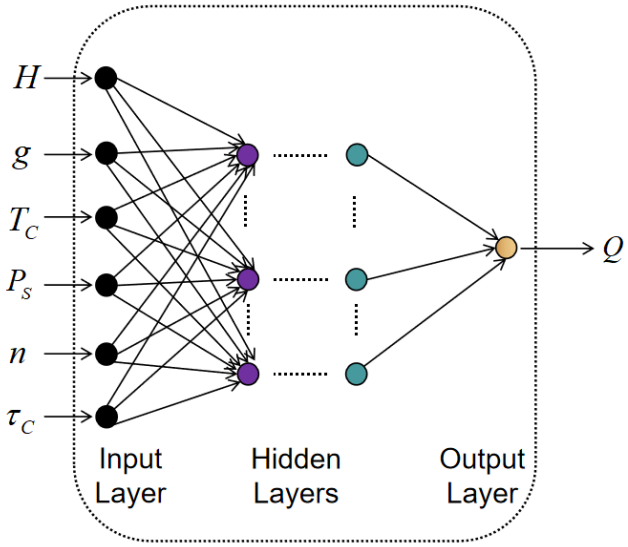

Figure 3.6: Deep critic network for approximating the reward during online operation

sample loss function, defined as

$$J(\mu) = \frac{1}{N'} \sum_{i=1}^{N'} (R_i - Q(\vec{s}_i, \vec{a}_i|\mu))^2. \tag{3.33}$$

Note that the instantaneous reward R_i for online experience will be equal to either $\frac{1}{P_S \tau_C + P_C(\tau_C + n/B)}$ for successful data collection or 0 for data collection failure. Correspondingly, the parameters of the critic network can be updated to

$$\mu \leftarrow \mu + \frac{\rho}{N'} \sum_{i=1}^{N'} (R_i - Q(\vec{s}_i, \vec{a}_i|\mu)) \bigtriangledown_\mu Q(\vec{s}_i, \vec{a}_i|\mu), \tag{3.34}$$

where $\rho \in (0,1)$ controls the learning rate of the critic network. We can then adjust the policy network using the sample gradient calculated with critic network $Q(\vec{s}, \vec{a}|\mu)$, given by

$$\bigtriangledown_\theta Q(\vec{s}, \vec{a}|\mu)|_{\vec{a}=\pi(\vec{s}|\theta)} \approx \frac{1}{N'} \sum_{i=1}^{N'} \bigtriangledown_{\vec{a}} Q(\vec{s}, \vec{a}|\mu)|_{\vec{s}=\vec{s}_i,\ \vec{a}=\pi(\vec{s}_i|\theta)} \bigtriangledown_\theta \pi(\vec{s}_i|\theta). \tag{3.35}$$

The pseudo code of the proposed online tuning algorithm is shown in Algorithm 2. The most computational intensive step of Algorithm 2 are the calculation of sample gradient, which can be efficiently implemented using back-propagation technique. With this tuning algorithm, the policy network can effectively track environment change with low computational complexity.

Algorithm 2 The pseudo code of online tuning

Input: offline trained policy network $\pi(\vec{s}|\theta)$ and critic network $Q(\vec{s}, \vec{a}|\mu)$.

for every N' data collection sessions **do**

Update critic network parameters μ as $\mu \leftarrow \mu + \frac{\rho}{N'} \sum_{i=1}^{N'} (R_i - Q(\vec{s}_i, \vec{a}_i|\mu)) \bigtriangledown_\mu Q(\vec{s}_i, \vec{a}_i|\mu)$.

Calculate sample gradient using Eq. (3.35).

Update policy network parameters θ as $\theta \leftarrow \theta + \xi \bigtriangledown_\theta Q(\vec{s}, \vec{a}|\mu)|_{\vec{a}=\pi(\vec{s}|\theta)}$.

end for

Table 3.2: Hyper parameters for policy and critic networks training.

Notation	Parameter names	Values
M	Replay buffer size	5000
N	Offline mini-batch size	60
ξ	Learning rate for offline training	10^{-6}
σ_P^2	Initial exploration variance for P_S	16.2
σ_τ^2	Initial exploration variance for τ_C	0.27
σ_n^2	Initial exploration variance for n	540
β	Variance reduction factor	0.995
N'	Online mini-batch size	100
ρ	Learning rate of online critic network tuning	0.005

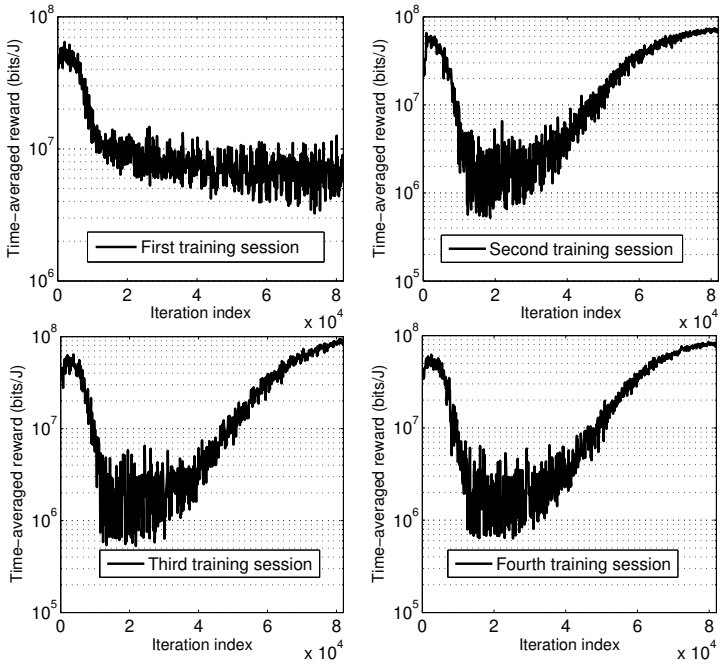

Figure 3.7: Time-averaged reward of different training sessions, $T_C = 25$ ms.

3.3.3 Numerical results

In this section, we present selected numerical examples to illustrate our proposed DRL solution. The system parameter values are shown as Table I, unless otherwise indicated. The hyper parameters for training the policy and critic networks are summarized in Table II for convenience.

In Fig. 3.7, we illustrate the offline training process of the deep policy network. We implement three hidden layers for the policy network with 128, 128 and 300 hidden nodes respectively. The network parameters are randomly generated. We approximately use the same number of sigmoid, tanh, and relu functions as the activation functions in each hidden layer and leave it to the training process to determine their preference. After training, we found that those nodes contributing more significantly to P_S use sigmoid function and those to τ_C use tanh function, and the remaining use relu function. Fig. 3.7 plots the time-averaged reward of 100 consecutive training steps as the function of the index of training steps. We notice that the time-averaged reward typically exhibits a large amount of variation initially and eventually converge to a stable value, as the exploration variances gradually reduce to zero. We can also see that even with random exploration, the offline training may converge to a local optimal solution, as shown in the first training session.

To address this problem, we can repeat the training of policy network and pick the

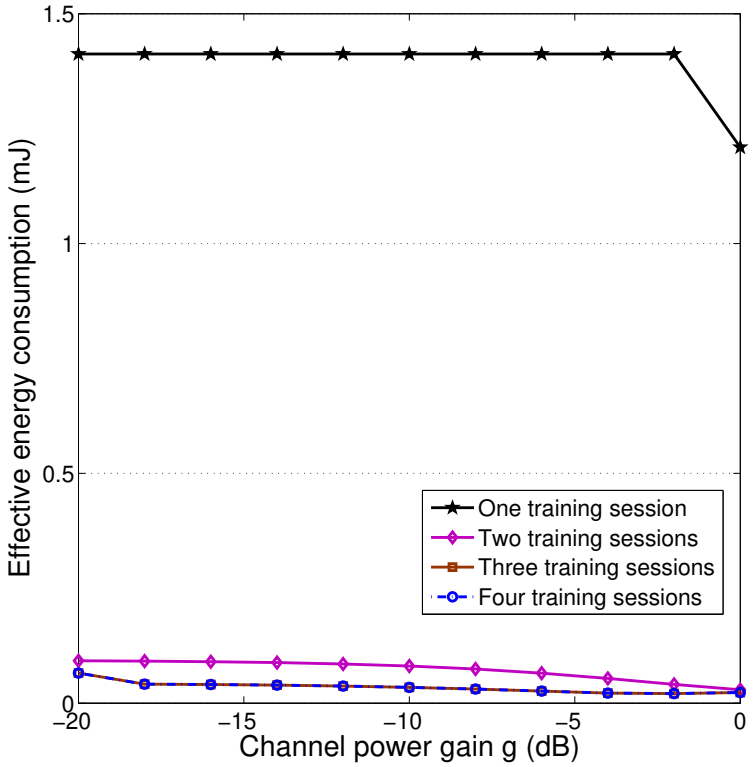

Figure 3.8: Effective energy consumption of the best network selected from multiple training sessions, $T_C = 25$ ms, $H = 2000$ bits.

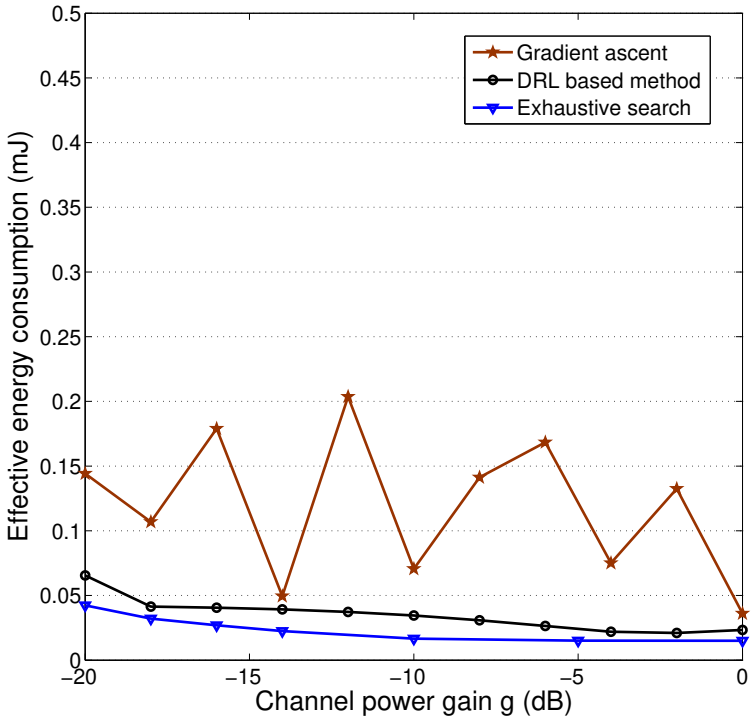

Figure 3.9: Effective energy consumption of different algorithms, $T_C = 25$ ms, $H = 2000$ bits.

resulting network with the highest reward upon convergence. Fig. 3.8 presents the effective energy consumption as the function of channel power gain g for the best network picked from increasing number of training sessions. We see that resulting energy consumption becomes smaller with increasing number of training sessions. Typically, we can achieve near-optimal performance after two or three sessions.

Fig. 3.9 compares the performance of our offline trained policy network with benchmark algorithms. In particular, we plot effective energy consumption as the function of channel power gain g for gradient ascent, proposed DRL, and exhaustive search algorithms. The results of gradient ascent are obtained through performing several gradient operations, where all initial points are randomly generated. Note that due to the latency constraint, the gradient ascent may not converge after a limited number of iterations, leading to the suboptimal performance. We can see that the result from our offline-trained network is very close to that from exhaustive search. Note that both gradient ascent and exhaustive search require much higher computational complexity and are not suitable for real-time data collection strategy design. Our offline trained policy network can generate near-optimal transmission parameters without any iteration, and as such, it is more suitable for the online operation.

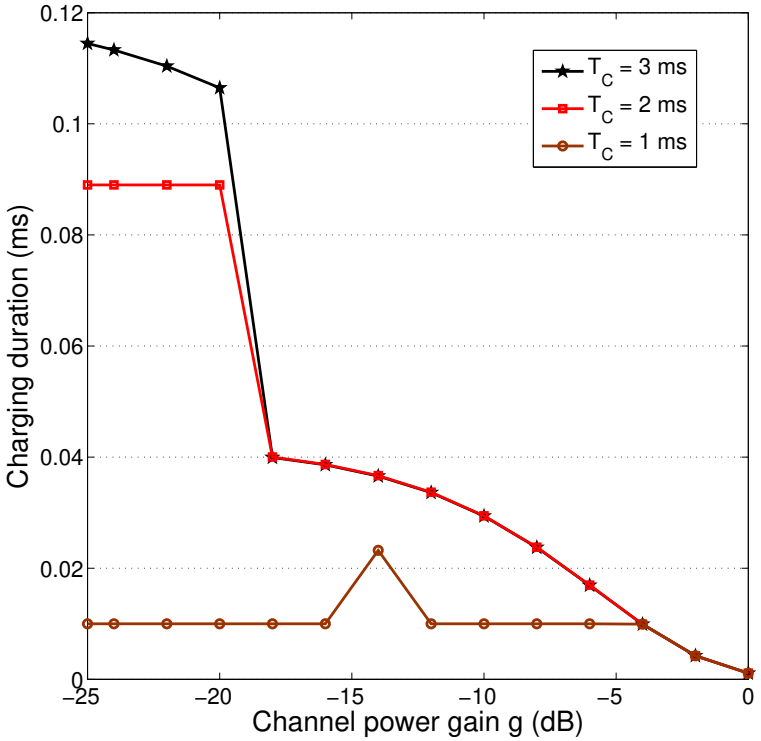

Figure 3.10: Charging duration from the offline-training policy network for different latency requirement, $H = 2000$ bits.

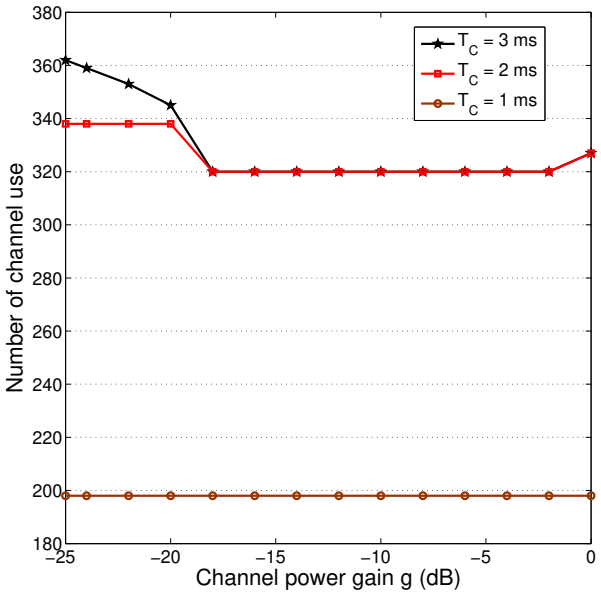

Figure 3.11: Channel use from the offline-training policy network for different latency requirement, $H = 2000$ bits.

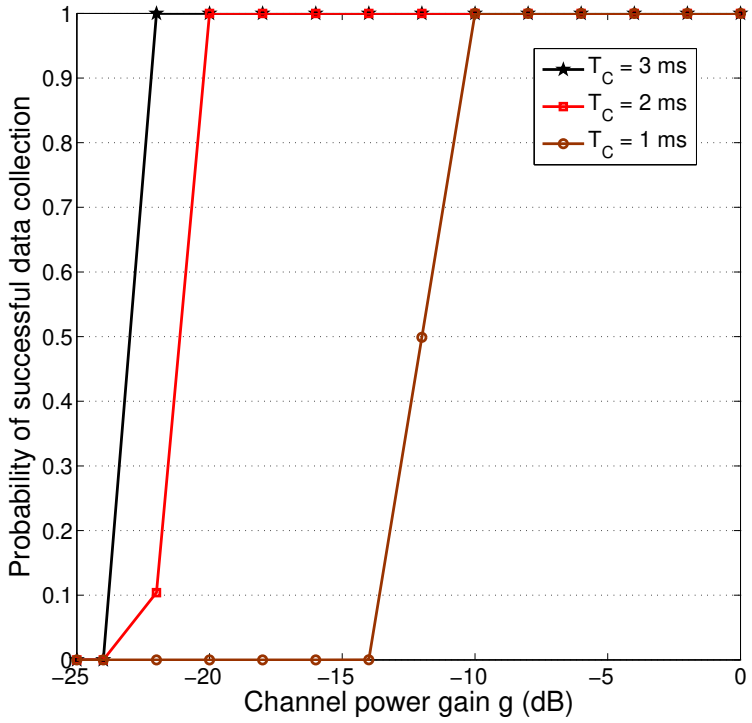

Figure 3.12: Probability of successful data collection for different latency requirement, $H = 2000$ bits.

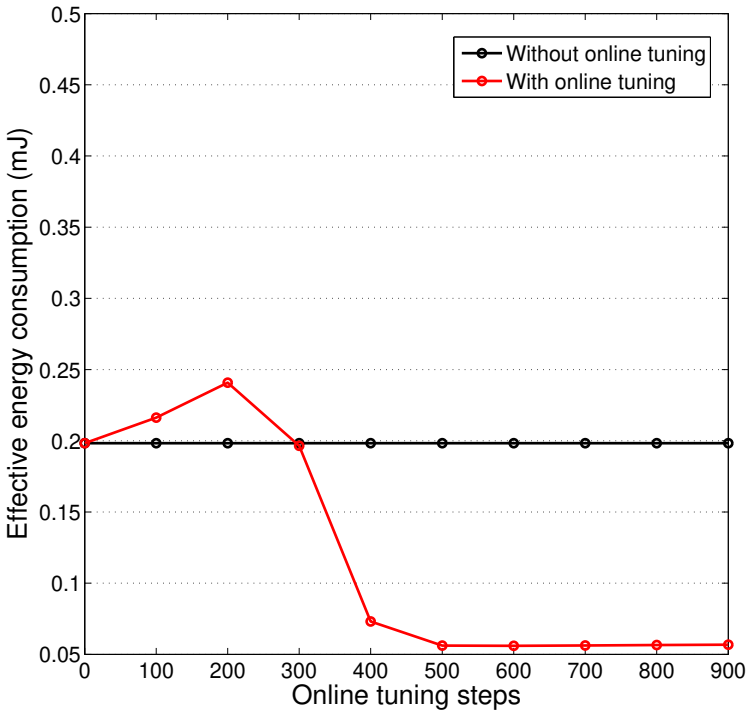

Figure 3.13: Effectively energy consumption with and without online tuning, $T_C = 25$ ms.

Fig. 3.10 and Fig. 3.11 illustrate the behavior of our policy network for different channel gains and latency requirement. We can see from Fig. 3.10 that the charging duration is generally increasing as the channel power quality degrades, unless limited by latency constraint. For the chosen system parameters, the charging duration is much less than 1 ms, leaving more time for data transmission as permitted by the latency requirement, as shown in Fig. 3.11. When the latency requirement is too stringent, i.e. $T_C = 1$ ms, the charging duration remains nearly constant and the data transmission duration is reduced dramatically to satisfy the latency constraint, at the cost of smaller probability of successful data collection. We omit the plot of the charging power P_S as the network output is always very closed to $P_{\max}$. Such behavior of policy network is consistent with our intuition and partially verifies the near-optimality of the trained policy network. The corresponding probability of successful data collection is plotted in Fig. 3.12. As we can see, the probability of successful data collection quickly reduces to zero when the channel power gain falls below a certain level, i.e. less than -10 dB for $T_C = 1$ ms case. Note that the goal of the policy network is to maximize the effective energy consumption, not to maximize the probability of successful data collection, under the given channel realization.

Fig. 3.13 illustrates the effect of online tuning, while setting the channel power gain

to -10 dB. We assume due to model inaccuracy or environment variation that the system parameters during online operation change to: $a = 2$, $b = 1.5$, $P_c = 30$ mW, and $\sigma^2 = 2 \times 10^{-5}$ mW. As we can see, if the offline trained policy is used without online tuning, the effective energy consumption remains at 0.2, which is much higher than the case when the model is accurate as shown in Fig. 3.9. To apply the proposed online tuning algorithm, we build a critic network with one hidden layer of 64 neuron and relu activation function. The critic network will be initially trained using offline experience tuples in the replay buffer and then gradually updated using online experience. We can see from Fig. 3.13 that with online tuning, the effective energy consumption, after some initial variation, is dramatically reduced to a low level after a certain number of tuning operations. As such, the policy network after online tuning can achieve near optimal performance during practical online operation.

3.4 Conclusion

In this chapter, we studied the energy consumption minimization problem during data collection at wireless-powered IoT sensors. For the ideal scenario of linear energy harvesting with ideal rate adaptive transmission, we derive the closed-form analytical expressions of the optimal transmission parameters for energy efficient data collection under given channel realization. For the practical scenario of nonlinear energy harvesting with finite block-length transmission, we develop an efficient solution based on DRL algorithms. Using an offline trained deep neural network, near-optimal transmission parameters can be determined in real time. We also proposed a solution for online tuning to mitigate the effect of model inaccuracy or/and environment variation. We presented some selected numerical examples to illustrate the effectiveness of our proposed solutions. With our analysis and DRL-based approach, optimal data collection from energy-constrained sensor, in terms of minimizing energy consumption, can be realized through adapting time-dependent optimal parameter configurations. On the other hand, following a similar process with our approach, other complicated optimal-design problems in real communication systems can be effectively handled as well.

Chapter 4

Green Relaying Transmission with Wireless Power Transfer

Reliable and energy-efficient wireless transmission is of critical importance to the success of future IoT. Due to the sporadic nature of IoT transmissions, the energy consumption of a specific data session varies dramatically with the instantaneous operating environment as well as the QoS requirement. In this chapter, we analyze and design the energy-efficient relaying transmission system from a session-specific perspective. Specifically, we consider a dual-hop transmission system with a decode-and-forward (DF) relay that is solely powered by wireless power transfer from source node. With consideration of ideal rate adaptive transmission and piecewise linear energy harvesting, for both time switching and power splitting modes of operation at the relay, we analyze and minimize the total energy consumption of the system when transmitting a fixed amount of data. Closed-form expressions for all optimal transmission parameters are derived with and without latency constraint. With consideration of finite block-length transmission and nonlinear energy harvesting, we follow the DRL-based algorithm proposed in chapter 3 to arrive at a policy network for determining near-optimal transmission parameters in real time during online operation. By comparing with the results from exhaustive search, the validity of our proposed approach and analysis can be effectively verified. In addition, we illustrate various design tradeoffs for such system through presenting some selected numerical examples.

4.1 Introduction

Internet connectivity can be extended to numerous terminal devices with the help of IoT technology, where many important applications will be enabled accordingly, such as remote sensing, intelligent manufacture, and smart farming, etc [111, 112]. Highly energy-efficient wireless transmission is very critical to the success of these applications. Also note that within some scenarios, high reliability is very difficult to be guaranteed for direct transmission, e.g. data transmission over long distance and poor channel. Hence, cooperative relaying transmission, as a promising solution, attracted a lot of attentions from academia. Meanwhile, since the energy storage of relay was strictly constrained in usual, some researchers investigated the wireless transmission system with energy harvesting relay [114–127], where the energy of relay was harvested from either environment or source node. On the other hand, to more properly characterize the energy efficiency for IoT applications, the work in [34] proposed a data-oriented approach to analyze the energy consumption of an individual transmission session. The work in [128] further applied this approach to the analysis and minimization of energy consumption for transmitting a fixed amount of data over a point-to-point wireless fading channel. The work in [35] investigated the wireless big data transmission with adaptive modulation and coding, where the statistical energy consumption of a specific data session was analyzed.

There has been ongoing interest for improving the energy efficiency of wireless transmission systems. Most previous efforts target at traditional broadband communication services. The energy efficiency of these services was characterized by the ratio of average data rate over corresponding power consumption level, with unit of bps/W [129]. For example, the work in [31] calculated the average energy consumption for each transmitted bit with the consideration of possible retransmissions. For fading wireless channel, the channel ergodic capacity is typically used to evaluate the energy efficiency [32, 33]. However, most IoT transmission sessions are very short and sporadic. For example, many IoT devices only send a short measurement update periodically, or upon the reception of a control command. In addition, to support advanced IoT applications, some IoT transmission sessions, e.g. for control commands, may require much higher reliability and lower latency than others, e.g. measurement updates. The resulting energy consumption of IoT transmission sessions will vary dramatically from one session to another. The transmission strategy that is optimal, in the average sense, may perform poorly for a specific transmission session [34]. As such, the average energy efficiency characterization is not suitable for optimizing wireless transmissions to/from IoT devices.

In this chapter, we extend previous work by considering relaying transmission system with wireless power transfer from a session-specific perspective. More specifically, we consider the general transmission scenario from a source to a destination with the help of an intermediate DF relay. The relay can harvest RF energy from the transmitted signal of the source and use it to power its relaying operation. With consideration of ideal rate adaptive transmission and piecewise linear energy harvesting, for both power-splitting and time-switching modes of operation, we analyze and minimize the overall energy consumption for transmitting a fixed amount of data with and without instantaneous latency constraint. The closed-form expressions for all optimal parameters are derived. With consideration of finite block-length transmission and nonlinear energy harvesting, we apply the algorithm of one-step DRL, proposed in chapter 3, to train a policy network to determine the near-optimal transmission parameters in real time for the online operation of such system. On the basis of these results, we can investigate whether the extra energy consumption to power the relay can enhance the energy efficiency of wireless transmission. Selected numerical examples are also discussed to illustrate the effect of our proposed approach and analysis.

4.1.1 Previous work

There have been continuing interest in wireless transmission systems with energy harvesting (EH) relays [114–118]. The work in [114, 115] applied power and packet size adaptation to optimize the energy efficiency of wireless transmission systems with EH relays. The work in [116] formulated and solved an energy efficiency optimization problem considering relays' energy harvesting constraints and average throughput constraint. The work in [117] considered a multi-hop transmission system with EH relays, where the energy consumption of source node is minimized. The work in [118] studied the joint transmit power minimization and EH relay selection problem to save the energy of multiple non-energy-harvesting transmitters. While freely available, the energy harvested from the environment is typically unstable and unpredictable, which makes the system design and planning very challenging.

Alternatively, relays can collect energy from the RF signal transmitted by the source node [79, 119–127]. In [79], the RF energy harvested from source transmission powers a full-duplex amplify-and-forwarding (AF) relay node. The work in [119] considered cooperative sensor network, where energy-constrained relay nodes harvest energy from the transmitter. The work in [120] considered a RF-powered AF relaying system, where the relay can also harvest energy from its own transmitted signal. The relay node in [121]

optimally decides to use its own battery energy or harvest RF energy from the source to enhance system energy efficiency. The work in [122,123] optimally design precoding scheme to maximize the energy efficiency of a MIMO two-way DF relay system with simultaneous wireless information and power transfer (SWIPT). Generally, relay node can adopt either power splitting or time switching operation mode for SWIPT operation. The work in [124,125] considered time-switching operation for single-carrier and multi-carrier transmission systems, respectively, whereas the work in [126, 127] employed power splitting mode at the relay. These works apply power allocation, precoding optimization, and relay selection to improve the average energy efficiency of relay transmission.

4.1.2 Contribution

This work is the first work to analyze the energy consumption of relaying transmission system with wireless power transfer from a session-specific perspective. By examining the minimum energy consumed for transmitting a fixed amount of data with and without latency constraint, we illustrate various design tradeoffs for such system. Additionally, we investigate the ideal rate adaptive transmission with piecewise linear energy harvesting and finite block-length transmission with nonlinear energy harvesting, respectively. These results will greatly facilitate the design of energy-efficient relaying transmission system for advanced IoT applications.

4.2 Ideal rate adaptive transmission with piecewise linear energy harvesting

In this section, we analyze and minimize the energy consumption for the system under ideal rate adaptive transmission with piecewise linear energy harvesting.

4.2.1 System and channel model

We consider a specific IoT transmission session where the source node transmits H bits of data to the destination node. Due to the poor channel quality of the direct link, the transmission is carried out with the help of an intermediate DF relay node. Note that we adopt DF relay as it can achieve higher reliability and capacity than AF relay [130]. To encourage relay cooperation, we assume that the relaying operation will be solely powered by the RF energy harvested from the transmitted signal of the source. Under the energy

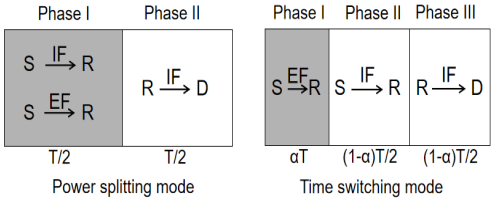

Figure 4.1: Operating modes of relay transmission with wireless power transfer.

and information causality constraints, the relay will operate in a half-duplexing fashion. In particular, the relay node first performs data reception and energy harvesting over the first hop and then forward the decoded data to the destination using its harvested energy. Our goal is to minimize the total energy consumption by designing the transmission parameters optimally considering the instantaneous channel condition, reliability requirement, and latency constraint, if applicable.

For most IoT transmission scenarios, the data amount H is typically small, at most several kbits. As such, such data transmission session will typically complete within one channel coherence time. It follows that the channel gains will remain constant for the whole data transmission duration. Assuming frequency flat fading environment, we denote the complex channel gains of the first hop and the second hop by h_{SR} and h_{RD}, respectively. The received signal at the relay can be written as

$$y_R(t) = h_{SR}s(t) + n_R(t), \tag{4.1}$$

where $s(t)$ is the signal transmitted by the source with power P_S and $n_R(t)$ is the noise at the relay with power σ^2. The relay performs simultaneous information detection and energy harvesting on $y_R(t)$, following either power splitting or time switching mode of operation. After that, it forwards a copy of its decoded signal, denoted by $r(t)$, to the destination using harvested RF energy. The received signal at the destination given by

$$y_D(t) = h_{RD}r(t) + n_D(t), \tag{4.2}$$

where $n_D(t)$ is the noise at the destination also with the average power of σ^2. We will consider the optimal design for power splitting and time switching modes separately in the following two sections, assuming that the instantaneous channel gains have been accurately estimated before the transmission session starts.

4.2.2 Power splitting mode

Energy consumption analysis

With power splitting (PS) operation mode, the total duration of the transmission session, T, is divided into two equal-length slots, one for source transmission and the other for relay transmission, as shown in Fig. 4.1(a). The relay will divide its received signal $y_R(t)$ into energy signal and information signal, with a power splitting factor β $(0 < \beta < 1)$,

for energy harvesting and information decoding, respectively. As most practical energy harvesting circuit is characterized by a linear region and a saturation region, we adopt a piecewise linear EH model [131]. As compared to the linear EH model in section 3, this piecewise linear model can better match the real EH process. In particular, the harvested energy at the relay over the first slot of duration $T/2$ is given by

$$E_H^{PS} = \begin{cases} \eta\beta P_S g_{SR} \frac{T}{2}, & P_S g_{SR}\beta < P_{sat}; \\ \eta P_{sat} \frac{T}{2}, & P_S g_{SR}\beta \geq P_{sat}, \end{cases} \tag{4.3}$$

where $g_{SR} = |h_{SR}|^2$ is the channel power gain from source to relay, $0 < \eta < 1$ denotes the energy conversion efficiency, and P_{sat} is the power saturation threshold of energy harvesting circuit. Meanwhile, the received signal-to-noise ratio (SNR) at the relay for information detection is given by

$$\gamma_R^{PS} = \frac{(1-\beta)P_S g_{SR}}{\sigma^2}. \tag{4.4}$$

In general, the energy consumption for information decoding at the relay is negligible compared to that for data forwarding. Accordingly, we assume that the relay uses all of its harvested energy for data forwarding over the second slot. The transmission power of relay is then given by

$$P_R^{PS} = \frac{E_H^{PS}}{T/2} = \eta \min\{P_S g_{SR}\beta, P_{sat}\}. \tag{4.5}$$

To establish the performance limit of energy consumption, we assume ideal rate adaptive transmission. Therefore, the effective data rate of the transmission system with power splitting mode can be calculated as

$$R^{PS} = \frac{B}{2}\log_2(1 + \frac{1}{\sigma^2}\min\{P_S(1-\beta)g_{SR},\ P_R^{PS} g_{RD}\}), \tag{4.6}$$

where B is the channel bandwidth and $g_{RD} = |h_{RD}|^2$ is the channel power gain from relay to destination. It follows that the total transmission duration is equal to $T = H/R^{PS}$.

Finally, the total energy consumption for transmitting H bits of data with the power splitting mode can be calculated as the product of transmission duration and power consumption for the source, which is given by

$$E_C^{PS} = (P_S + P_c)T/2 = \frac{H(P_S + P_c)/B}{\log_2(1 + \min\{P_S(1-\beta)g_{SR},\ \eta \min\{P_S g_{SR}\beta, P_{sat}\} g_{RD}\}/\sigma^2)}, \tag{4.7}$$

where P_c denotes the circuit power consumption of the source node.

Optimal design

We now determine the optimal source transmit power P_S and power splitting factor β, based on the instantaneous channel realization, to minimize the total energy consumption, given in Eq. (4.7). Note that this objective function is non-convex. As such, we can not directly apply KKT condition to obtain the optimal solution. Note also that we need to determine the optimal parameter values for each data transmission session based on the instantaneous channel gain. Iterative algorithms are not suitable as they need time to converge, even to suboptimal solution. In what follows, we derive the closed-form expressions of optimal transmission parameters through mathematical analysis.

We consider the case of $P_S g_{SR}\beta < P_{sat}$ and $P_S g_{SR}\beta \geq P_{sat}$ separately. When $P_S g_{SR}\beta < P_{sat}$, the energy consumption minimization problem is formulated as

$$\begin{aligned} \min_{P_S,\,\beta} \quad & \frac{H(P_S+P_c)}{B\log_2(1+P_S g_{SR}(\min\{(1-\beta),\ \eta\beta g_{RD}\})/\sigma^2)}, \\ s.t. \quad & 0<P_S<\min\{P_{sat}/(g_{SR}\beta), P_{\max}\},\ 0<\beta<1, \end{aligned}$$

where $P_{\max}$ is the peak transmit power of the source node. We first observe that objective function is monotonically decreasing with $\min\{1-\beta,\ \beta\eta g_{RD}\}$, which will achieve its maximum value when $1-\beta=\beta\eta g_{RD}$. Thus, the optimal value of power splitting factor β is given by

$$\beta^* = \frac{1}{1+g_{RD}\eta}. \tag{4.8}$$

Applying optimal power splitting factor, the optimization problem simplifies to

$$\begin{aligned} \min_{P_S} \quad & \frac{(P_S+P_c)H}{B\log_2(1+P_S(g_{SR}/\widetilde{g}_{RD})/\sigma^2)}, \\ s.t. \quad & 0<P_S<\min\{P_{sat}(1+g_{RD}\eta)/g_{SR}, P_{\max}\}, \end{aligned}$$

where we define $\widetilde{g}_{RD}=1+\frac{1}{g_{RD}\eta}$ for notation conciseness. It can be verified that, by examining its second derivative with respect to P_S, this objective function is strictly convex with respect to P_S. After taking derivative with respect to P_S and setting the result to zero, the optimal source transmission power, under the the constraint of $0<P_S<P_{sat}/(g_{SR}\beta^*)$, is

determined as

$$P_S^{*PS} = \min\{\frac{P_c - \widetilde{g}_{RD}\sigma^2/g_{SR}}{W_0[e^{-1}(\frac{P_c g_{SR}}{\widetilde{g}_{RD}\sigma^2} - 1)]} - \frac{\widetilde{g}_{RD}\sigma^2}{g_{SR}},\ \frac{P_{sat}(1+g_{RD}\eta)}{g_{SR}},\ P_{\max}\}, \tag{4.9}$$

where $W_0[\cdot]$ denotes the principle branch of the Lambert W function [85].

For the case of $P_S g_{SR}\beta \geq P_{sat}$, the energy consumption minimization problem becomes, assuming that $P_{sat}/(g_{SR}\beta)$ is less than $P_{\max}$ [1]

$$\begin{aligned} \min_{P_S,\,\beta} \quad & \frac{H(P_S + P_c)}{B\log_2(1 + \min\{P_S(1-\beta)g_{SR},\ \eta P_{sat} g_{RD}\}/\sigma^2)}, \\ s.t. \quad & P_{sat}/(g_{SR}\beta) \leq P_S \leq P_{\max},\ 0 < \beta < 1, \end{aligned}$$

The objective function will be minimized when P_S and β satisfy $P_S(1-\beta)g_{SR} = \eta P_{sat} g_{RD}$. Under this condition, the optimization problem can be simplified to

$$\begin{aligned} \min_{P_S} \quad & \frac{H(P_S + P_c)}{B\log_2(1 + \eta P_{sat} g_{RD}/\sigma^2)}, \\ s.t. \quad & P_{sat}/(g_{SR}\beta) \leq P_S \leq P_{\max}. \end{aligned}$$

Since the objective function is monotonically increasing with P_S, the optimal P_S should be equal to $P_{sat}/(g_{SR}\beta)$. Applying this optimal P_S, we arrive at the same optimal β^* as given in Eq. (4.8).

Combining the above two cases, the optimal transmission parameters for power splitting mode in terms of minimizing total energy consumption is given in Eq. (4.8) and Eq. (4.9). The resulting transmission rate of both hops is equal to

$$R^{*PS} = B\log_2(1 + \frac{P_S^{*PS}\eta g_{SR} g_{RD}}{(1+\eta g_{RD})\sigma^2}). \tag{4.10}$$

Correspondingly, to minimize the energy consumption of the IoT transmission session under consideration, the source will transmit at rate R^{*PS} with power P_S^{*PS} for a duration of $H/R^{*PS}/2$. The relay will apply power splitting operation with β^* and then forward its decoded data with the same rate for the same duration, but with power level $\eta P_S^{*PS} g_{SR}\beta^*$.

[1] Otherwise, no solution exists for this case.

The resulting minimum energy consumption can be determined as

$$E_{\min}^{PS} = \begin{cases} \frac{H(P_c - \widetilde{g}_{RD}\sigma^2/g_{SR})\ln 2}{BW_0[e^{-1}(\frac{P_c g_{SR}}{\widetilde{g}_{RD}\sigma^2}-1)]}, \\ \frac{(P_c g_{SR} - \widetilde{g}_{RD}\sigma^2)\beta^*}{W_0[e^{-1}(\frac{P_c g_{SR}}{\widetilde{g}_{RD}\sigma^2}-1)]} - \widetilde{g}_{RD}\sigma^2\beta^* < P_{sat}; \\ \frac{H(P_{sat}/g_{SR}+P_c)}{B\log_2(1+\frac{\eta P_{sat} g_{RD}}{\sigma^2})}, \\ \frac{P_c g_{SR} - \widetilde{g}_{RD}\sigma^2\beta^*}{W_0[e^{-1}(\frac{P_c g_{SR}}{\widetilde{g}_{RD}\sigma^2}-1)]} - \widetilde{g}_{RD}\sigma^2\beta^* \geq P_{sat}. \end{cases} \tag{4.11}$$

4.2.3 Time switching mode

Energy consumption analysis

With time switching (TS) operation mode, the total transmission duration T is divided into three slots, as shown in Fig. 4.1(b). Specifically, the source first transfers energy to the relay with power P_{SC} over the first slot of duration αT, where $\alpha \in (0,1)$ is the time switching factor. Then, the source transmits information to the relay over the second slot of duration $(1-\alpha)T/2$. Finally, the relay forwards its decoded information to the destination over the last slot. As such, applying a piecewise linear EH model, the harvested energy at the relay is given by

$$E_H^{TS} = \begin{cases} \alpha T \eta P_{SC} g_{SR}, & P_{SC} g_{SR} < P_{sat}; \\ \alpha T \eta P_{sat}, & P_{SC} g_{SR} \geq P_{sat}. \end{cases} \tag{4.12}$$

The received SNR at the relay over the second slot is equal to

$$\gamma_R^{TS} = \frac{P_S g_{SR}}{\sigma^2}. \tag{4.13}$$

Since the relay uses all of its harvested energy for data forwarding over the last slot, the transmit power of the relay is determined as

$$P_R^{TS} = \frac{E_H^{TS}}{(1-\alpha)T/2} = \frac{2\alpha\eta}{1-\alpha}\min\{P_{SC} g_{SR}, P_{sat}\}. \tag{4.14}$$

The effective data rate with time switching mode can be calculated, while noting that only $1-\alpha$ portion of the total duration is used for information transmission, as

$$R^{TS} = \frac{(1-\alpha)B}{2}\log_2(1 + \frac{1}{\sigma^2}\min\{P_S g_{SR}, P_R^{TS} g_{RD}\}). \tag{4.15}$$

It follows that the total duration of the transmission session T is equal to H/R^{TS}.

With time switching mode, the source consumes energy at power level $P_{SC}+P_c$ in the first slot for a duration of $\alpha H/R^{TS}$ and at power level P_S+P_c in the second slot for $(1-\alpha)H/(2R^{TS})$ duration. As such, we can calculate the total energy consumption as

$$E_C^{TS}=\frac{H(P_S+P_c+\widetilde{\alpha}(P_c+P_{SC}))/B}{\log_2(1+\min\{P_S g_{SR},\ \widetilde{\alpha}\eta\min\{P_{SC}g_{SR},P_{sat}\}g_{RD}\}/\sigma^2)}, \tag{4.16}$$

where we define $\widetilde{\alpha}$ as $2\alpha/(1-\alpha)$ for conciseness.

Optimal design

We now optimize the charging power P_{SC}, information transmission power P_S, and time switching factor α to minimize the total energy consumption of a specific transmission session, given in Eq. (4.16). Note that the above objective function is not convex. Again, iterative algorithms are not applicable as the channel state information will soon become outdated. In the follows, we still perform some mathematical analysis to derive the closed-form expressions of all optimal transmission parameters.

First of all, we determine the optimal value of charging power P_{SC} with the following proposition.

Proposition 3. *For time switching mode, the optimal charging power in terms of minimizing the total energy consumption of a data transmission session is equal to* $P_{SC}^*=\min\{P_{sat}/g_{SR},P_{\max}\}$.

Proof. When $P_{SC}g_{SR}\le P_{sat}$, we can rewrite the total energy consumption as

$$E_C^{TS}=\frac{(P_S+(1+m/P_{SC})P_c+m)H}{B\log_2(1+g_{SR}\min\{P_S,\ m\eta g_{RD}\}/\sigma^2)}, \tag{4.17}$$

where we define an auxiliary variable $m=\widetilde{\alpha}P_{SC}$. Note that m can be adjusted independent of P_{SC} since $\widetilde{\alpha}$ ranges from 0 to $+\infty$. As such, total energy consumption is a monotonically decreasing function of P_{SC}. So, the optimal P_{SC} is equal to P_{sat}/g_{SR}.

When $P_{SC}g_{SR}\ge P_{sat}$, the total energy consumption becomes

$$E_C^{TS}=\frac{(P_S+P_c+\widetilde{\alpha}(P_c+P_{SC}))H}{B\log_2(1+\min\{P_S g_{SR},\ \widetilde{\alpha}\eta P_{sat}g_{RD}\}/\sigma^2)}, \tag{4.18}$$

which is monotonically increasing with P_{SC}. Therefore, the optimal P_{SC} in this case is also equal to P_{sat}/g_{SR}. The proof is completed while noting the peak transmit power of the

source $P_{\max}$. □

Applying the optimal value of charging power, i.e. P_{SC}^*, the energy consumption minimization problem can be formulated as

$$\begin{aligned} \min_{P_S,\ \alpha} \quad & \frac{(P_S + P_c + \widetilde{\alpha}(P_c + P_{SC}^*)H}{B\log_2(1 + g_{SR}\min\{P_S,\ \widetilde{\alpha}\eta P_{SC}^* g_{RD}\}/\sigma^2)}, \\ s.t. \quad & 0 < P_S \leq P_{\max},\ 0 < \alpha < 1. \end{aligned}$$

Note that the objective function is minimized when α and P_S satisfy

$$P_S = \widetilde{\alpha}\eta P_{SC}^* g_{RD}. \tag{4.19}$$

Under this condition, the optimization problem simplifies, after setting $\widetilde{\alpha} = P_S/(P_{SC}^*\eta g_{RD})$, to

$$\begin{aligned} \min_{P_S} \quad & \frac{(P_S\hat{G} + P_c)H}{B\log_2(1 + P_S g_{SR}/\sigma^2)}, \\ s.t. \quad & 0 < P_S \leq P_{\max}, \end{aligned}$$

where we define $\hat{G} = 1 + \frac{P_c + P_{SC}^*}{\eta P_{SC}^* g_{RD}}$ for notation conciseness. Since the objective function is strictly convex with respect to P_S, we can determine the optimal information transmission power for time switching mode as

$$P_S^{*TS} = \min\{\frac{P_c/\hat{G} - \sigma^2/g_{SR}}{W_0[e^{-1}(\frac{P_c g_{SR}}{\hat{G}\sigma^2} - 1)]} - \frac{\sigma^2}{g_{SR}},\ P_{\max}\}. \tag{4.20}$$

The optimal time switching factor can be obtained, by setting P_S to P_S^{*TS}, as

$$\alpha^* = 1 - \frac{2g_{RD}\eta P_{SC}^*}{P_S^{*TS} + 2g_{RD}\eta P_{SC}^*}. \tag{4.21}$$

With these results, we can calculate the effective information transmission rate as

$$R^{*TS} = B\log_2(1 + \frac{P_S^{*TS} g_{SR}}{\sigma^2}). \tag{4.22}$$

Accordingly, to minimize the energy consumption of the transmission session under consideration, the source first transfers energy to the relay with power $P_{SC}^* = \min\{P_{sat}/g_{SR}, P_{\max}\}$

for a duration of $\alpha^* H/R^{*TS}$. Then, it will transmit data at rate R^{*TS} for duration of $(1-\alpha^*)H/(2R^{*TS})$ with power P_S^{*TS}. For the remaining $(1-\alpha^*)H/(2R^{*TS})$, the relay will forward its decoded data to the destination at the same rate with power $\frac{2\alpha^*\eta}{1-\alpha^*}P_{SC}^* g_{SR}$. Combining the above descriptions, the resulting minimum energy consumption with time switching mode can be determined as

$$E_{\min}^{TS} = \begin{cases} \frac{H(P_c-\hat{G}\sigma^2/g_{SR})\ln 2}{BW_0[e^{-1}(\frac{P_c g_{SR}}{\hat{G}\sigma^2}-1)]}, \\ \frac{P_c/\hat{G}-\sigma^2/g_{SR}}{W_0[e^{-1}(\frac{P_c g_{SR}}{\hat{G}\sigma^2}-1)]} - \frac{\sigma^2}{g_{SR}} < P_{\max}; \\ \frac{(P_{\max}+P_c+\frac{2\alpha^*\eta}{1-\alpha^*}(P_c+P_{SC}^*))H}{B\log_2(1+g_{SR}\min\{P_{\max}, \frac{2\alpha^*\eta}{1-\alpha^*}P_{SC}^* g_{RD}\}/\sigma^2)}, \\ \frac{P_c/\hat{G}-\sigma^2/g_{SR}}{W_0[e^{-1}(\frac{P_c g_{SR}}{\hat{G}\sigma^2}-1)]} - \frac{\sigma^2}{g_{SR}} \geq P_{\max}. \end{cases} \tag{4.23}$$

4.2.4 Numerical results

In this section, we present selected numerical examples to illustrate the analytical results in previous subsections. Table I summarizes the common parameter values used when generating these results.

Table 4.1: Common parameter values of relaying transmission system with ideal transmission and EH models.

Parameters	Values
noise power, σ^2	1×10^{-5} mWatt
channel bandwidth, B	100 kHz
circuit power, P_c	20 mWatt
energy conversion efficiency, η	0.9
maximum transmit power, $P_{\max}$	300 mWatt
power saturation threshold, P_{sat}	20 mWatt

In Fig. 4.2, we plot the energy consumption of relay transmission system with power splitting mode when transmitting $H = 5$ kbits of data. In particular, we plot the energy consumption as the function of g_{SR} with different values of g_{RD} for the cases with and without parameter optimizations. We can see that the energy consumption generally decreases as the channel gains g_{RD} and g_{SR} increase, as intuitively expected. We can also see that parameter optimization can dramatically reduce the energy consumption of the session. The perfect match between the analytical results and the results from exhaustive search verifies the validity of our analytical results. Similar behavior can be observed for the energy consumption for time switching mode.

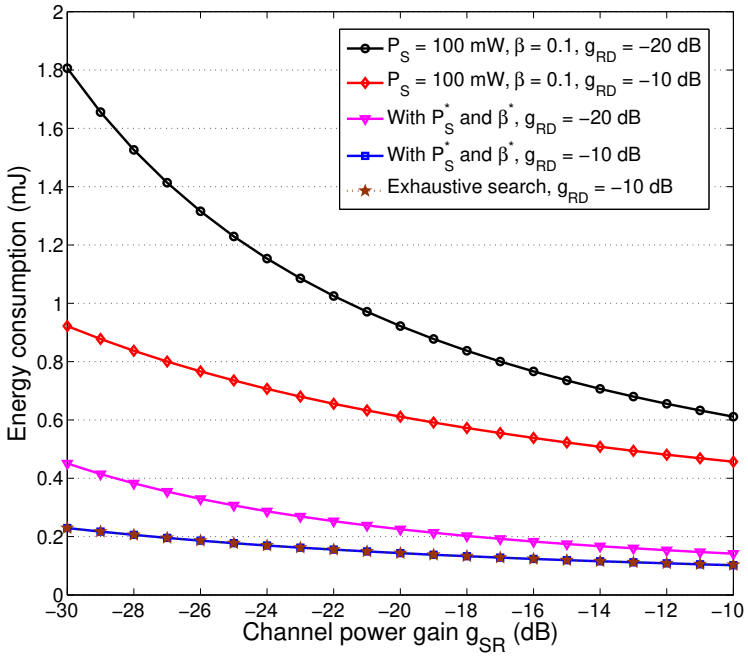

Figure 4.2: Energy consumption of a transmission session with power splitting energy transfer mode ($H = 5$ kbits).

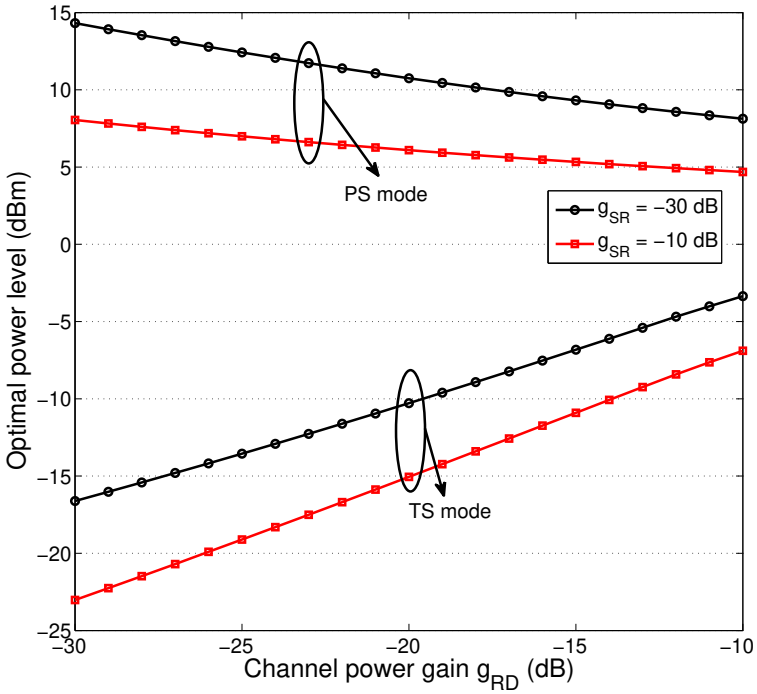

Figure 4.3: Optimal transmit power levels of source node ($H = 5$ kbits).

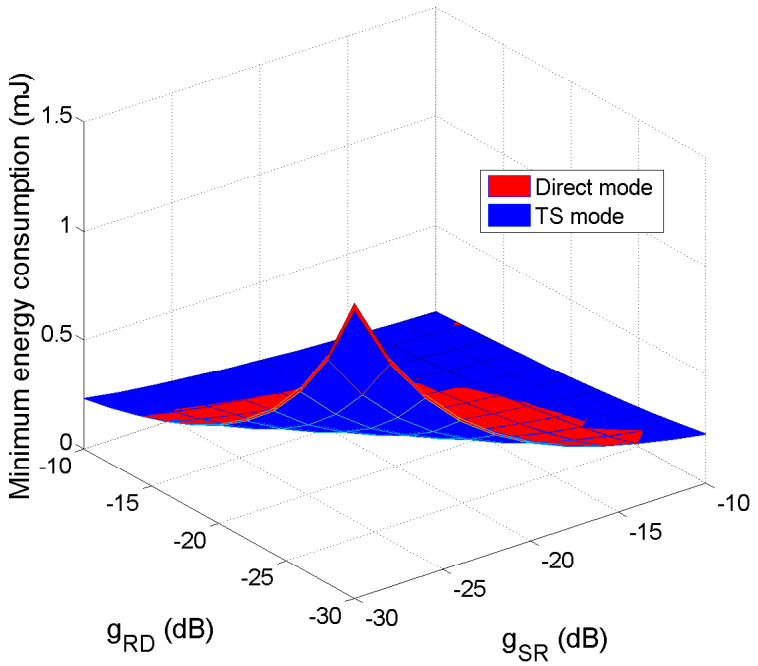

Figure 4.4: Energy consumption comparison between relay transmission with time switching mode and direct transmission with optimal parameters ($H = 5$ kbits).

In Fig. 4.3, we plot the optimal information transmission power of the source P_S^* in terms of minimizing the energy consumption for different channel realizations. We can see that when g_{SR} increases, the optimal power for both modes decreases as intuitively expected. On the other hand, when g_{RD} increases, while that for power splitting mode decreases, the optimal transmit power for time switching mode increases. This interesting behavior of time switching mode can be explained as follows. When g_{RD} is relatively small, the second hop will likely be the bottleneck link. The best strategy for the source is to transmit at low rate with low transmit power. When g_{RD} increases and relay link can support higher rate, the source will use a higher transmit power level to match the transmission rate of the second hop, which will reduce the overall transmission duration and hence increase the power consumption.

Fig. 4.4 compares the energy consumption of relay transmission with time switching mode and direct transmission [128] using optimal transmission parameters. In particular, the minimum energy consumption is plotted as function of both g_{SR} and g_{RD}. For fair comparison, we assume that the channel attenuation of direct transmission g_{SD} is approximately equal to the sum of the attenuation of both hops, which corresponds to the case that relay is situated on the line connecting the source and the destination. We can see that when the

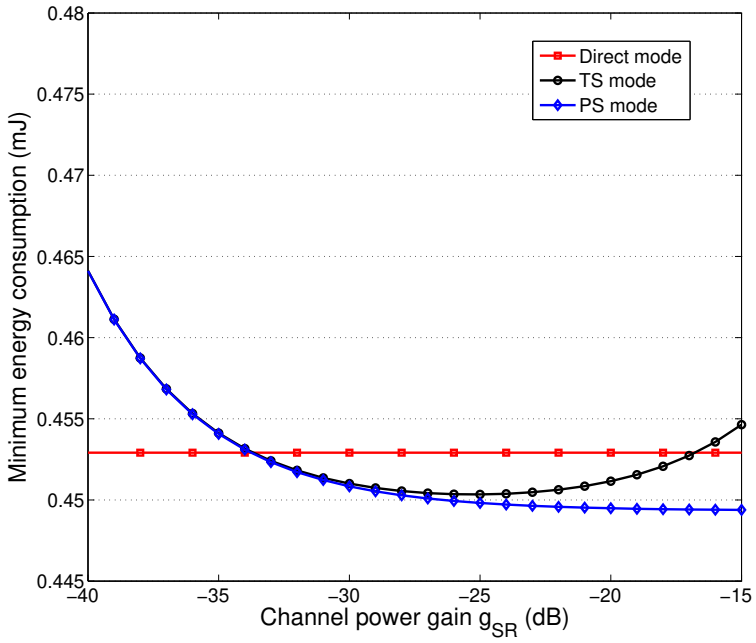

Figure 4.5: Energy consumption comparison between power splitting, time switching, and direct transmission modes with optimal parameters ($g_{SD} = -50$ dB and $H = 5$ kbits).

channel quality is poor, i.e. $g_{SD} < -30$ dB, relay transmission with energy transfer can save considerable amount of energy compared to direct transmission. On the other hand, when $g_{SD} > -25$ dB, direct transmission leads to lower energy consumption. As such, we can conclude that relay transmission with wireless power transfer is more energy efficient than direct transmission when the direct link experience poor channel quality.

Fig. 4.5 compares the energy consumption of power splitting, time switching, and direct transmission while fixing the direct link channel power gain g_{SD} to around -50 dB and varying g_{SR} from -40 dB to -15 dB. Correspondingly, g_{SR} decreases from -10 dB to -35 dB. The energy consumption of direct transmission is constant as g_{SD} remains the same. The time switching mode leads to lower energy consumption than direct transmission as long as neither g_{SR} nor g_{RD} is too small. On the other hand, the power splitting mode enjoys the lowest energy consumption as long as g_{SR} is not very small. With the exact analytical expressions for the minimum energy consumption determined above, the source can determine which transmission mode leads to the minimum energy consumption for a given channel realization. The resulting transmission system will achieve the highest possible energy utilization efficiency for each transmission session and save the valuable energy resource of source node.

Fig. 4.6 plots the time duration required to transmit a fixed amount of data using the

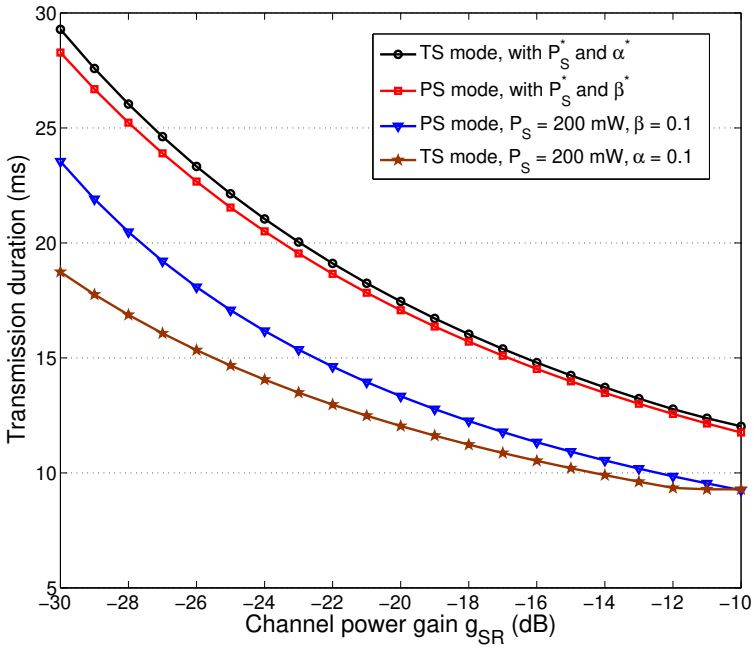

Figure 4.6: Transmission duration comparison of power splitting and time switching modes with and without energy consumption minimization ($g_{RD} = -20$ dB, $H = 5$ kbits).

relay transmission system under consideration. The results of both power splitting and time switching modes, with and without energy consumption minimization, are presented. We can see that the transmission duration for all cases decreases with increasing g_{SR}. We also observe that the transmission duration increases for both operating modes after parameter optimization, demonstrating a tradeoff between energy consumption and latency performance. Finally, the optimized power splitting mode leads to smaller transmission duration than the optimized time switching mode. Considering together with the energy consumption comparison in Fig. 4.5, we can conclude that power splitting is the preferred operating mode as it leads to lower minimum energy consumption and shorter transmission duration, at the cost of slightly higher hardware complexity due to simultaneous energy harvesting and data reception.

4.2.5 Effect of latency constraint

The results in previous section show that the optimal transmission parameters that minimizes energy consumption will lead to large transmission duration for transmitting a fixed amount of data. Meanwhile, the QoS requirement of certain IoT applications may mandate a hard latency constraint. In this section, we study the effect of such latency constraint on minimum energy consumption, again from an individual data transmission session perspective. We assume that the source needs to transmit H bits data to the destination under a hard latency constraint, denoted by T_{th}.

With hard latency constraint, the energy consumption minimization problem for power splitting mode becomes

$$\begin{aligned} \min_{P_S,\,\beta} \quad & \frac{H(P_S+P_c)/B}{\log_2(1+\frac{\min\{P_S(1-\beta)g_{SR},\ \eta\min\{P_S g_{SR}\beta, P_{sat}\}g_{RD}\}}{\sigma^2})}, \\ s.t. \quad & 0<P_S\le P_{\max},\ 0<\beta<1, \\ & \frac{2H}{B\log_2(1+\min\{P_S(1-\beta)g_{SR},\ \eta\min\{P_S g_{SR}\beta, P_{sat}\}g_{RD}\}/\sigma^2)} < T_{\text{th}}. \end{aligned}$$

Following a similar process in previous section, while considering the cases of $P_S g_{SR}\beta \le P_{sat}$ and $P_S g_{SR}\beta > P_{sat}$ separately, we can show that the power splitting factor β obtained in Eq. (4.8) remains optimal under the latency constraint. Accordingly, the optimal transmit power is given by

$$\widetilde{P}_S^{*PS} = \max\{P_S^{*PS},\ (2^{2H/(BT_{\text{th}})}-1)\widetilde{g}_{RD}\sigma^2/g_{SR}\}, \tag{4.24}$$

where P_S^{*PS} was given in Eq. (4.9). Essentially, the latency constraint translates to a minimum transmission power requirement. Note that when the minimum power $(2^{2H/(BT_{\text{th}})}-1)\widetilde{g}_{RD}\sigma^2/g_{SR}$ is greater than $P_{\max}$, the above optimization problem will have no feasible solutions, i.e. the data transmission can not complete within the latency constraint under the given channel realization. In this case, the source node may decide not to transmit to save its energy.

For time switching mode, the energy consumption minimization problem is updated to

$$\begin{aligned}\min_{P_S,\,\alpha} \quad & \frac{(P_S+P_c+\widetilde{\alpha}(P_c+P_{SC}))H}{B\log_2(1+\min\{P_S g_{SR},\ \widetilde{\alpha}\eta\min\{P_{SC}g_{SR},P_{sat}\}g_{RD}\}/\sigma^2)},\\ s.t. \quad & 0<P_S\le P_{\max},\ 0<\alpha<1,\\ & \frac{2H}{B(1-\alpha)\log_2(1+\min\{P_S g_{SR},\ \widetilde{\alpha}\eta\min\{P_{SC}g_{SR},P_{sat}\}g_{RD}\}/\sigma^2)}<T_{\text{th}}.\end{aligned}$$

Note that increasing effective transmission rate will reduce the energy consumption as well as the transmission duration. Therefore, the optimal charging power is still equal to $\min\{P_{sat}/g_{SR},P_{\max}\}$ under the latency constraint. And the optimal α and P_S still need to satisfy Eq. (4.19). After eliminating α using Eq. (4.19), the optimization problem simplifies to

$$\begin{aligned}\min_{P_S} \quad & \frac{(P_S\hat{G}+P_c)H}{B\log_2(1+P_S g_{SR}/\sigma^2)},\\ s.t. \quad & 0<P_S\le P_{\max},\\ & \frac{g_{RD}\eta P_{SC}^*}{P_S+2g_{RD}\eta P_{SC}^*}B\log_2(1+P_S g_{SR}/\sigma^2)\ge H/T_{\text{th}}.\end{aligned}$$

The optimal power determined in Eq. (4.20) should also satisfy the latency contraint. After some manipulations, we can equivalently transform the latency constraint to

$$\begin{aligned}(\frac{U}{g_{SR}g_{RD}}-\frac{U}{\sigma^2 g_{RD}}P_S)\exp(\frac{U}{g_{SR}g_{RD}}-\frac{U}{\sigma^2 g_{RD}}P_S)\le\\ -\frac{U}{g_{SR}g_{RD}}\exp(-\frac{U}{g_{SR}g_{RD}}+\frac{2H\ln 2}{BT_{\text{th}}}).\end{aligned}\tag{4.25}$$

where U denotes $\frac{H\sigma^2\ln 2}{\eta P_{SC}^* BT_{\text{th}}}$. Applying the definition of Lambert W function, the value range of transmit power that satisfies the latency constraint is determined as

$$P_0^{TS}\le P_S\le P_{-1}^{TS},\tag{4.26}$$

where

$$P_0^{TS} = -\sigma^2\left(\frac{1}{g_{SR}} + \frac{g_{RD}}{U} W_0\left[-\frac{U}{g_{SR}g_{RD}} \exp\left(-\frac{U}{g_{SR}g_{RD}} + \frac{2H\ln 2}{BT_{\text{th}}}\right)\right]\right), \tag{4.27}$$

and

$$P_{-1}^{TS} = -\sigma^2\left(\frac{1}{g_{SR}} + \frac{g_{RD}}{U} W_{-1}\left[-\frac{U}{g_{SR}g_{RD}} \exp\left(-\frac{U}{g_{SR}g_{RD}} + \frac{2H\ln 2}{BT_{\text{th}}}\right)\right]\right). \tag{4.28}$$

Here, $W_{-1}[\cdot]$ denotes the negative branch of Lambert W function. It can be verified using the properties of the W function that P_0^{TS} is greater than zero. As such, the optimal information transmission power for the time switching mode under the latency constraint is determined as

$$\widetilde{P_S}^{TS} = \max\{P_0^{TS},\ \min\{P_{-1}^{TS},\ P_S^{*TS}\}\}, \tag{4.29}$$

where P_S^{*TS} was given in Eq. (4.20). The corresponding optimal time switching factor for this case becomes

$$\widetilde{\alpha} = 1 - \frac{2g_{RD}\eta P_{SC}^*}{\widetilde{P_S}^{TS} + 2g_{RD}\eta P_{SC}^*}. \tag{4.30}$$

The energy consumption minimization problem for time switching mode may also have no solution due to the latency constraint. Specifically, for a particular channel realization, if the right hand side of Eq. (4.25), i.e. $-\frac{U}{g_{SR}g_{RD}} \exp\left(-\frac{U}{g_{SR}g_{RD}} + \frac{2H\ln 2}{BT_{\text{th}}}\right)$, is less than $-1/e$ or P_0^{TS} in Eq. (4.27) is greater than $P_{\max}$, the latency constraint can not be satisfied. In this case, the source node may again decide not to transmit to save its energy.

In Fig. 4.7, we plot the minimum energy consumption of the relay transmission system, with optimal transmission parameters, as the function of the latency constraint threshold T_{th}. As we can see, the minimum energy consumption of relay transmission system under different channel realization follows a similar decreasing trend as latency threshold increases. This confirms again the tradeoff between energy consumption and transmission latency. We also observe that the energy consumption of time switching mode is slightly higher than that with power splitting mode. Finally, better source-to-relay channel quality leads to lower energy consumption as less energy is consumed during wireless power transfer to the relay.

In Fig. 4.8, we compare the minimum energy consumption of relay transmission system with direct transmission under a hard latency constraint of $T_{\text{th}} = 0.015$ s. For fair comparison, we again assume that the relay is situated on the line between the source and destination and fix the channel power gain of direct link to -50 dB. The minimum energy consumption is plotted as function of g_{SR}, while the corresponding g_{RD} is decreasing from -10 dB to -35 dB. We can see that under the given latency constraint, relay transmission

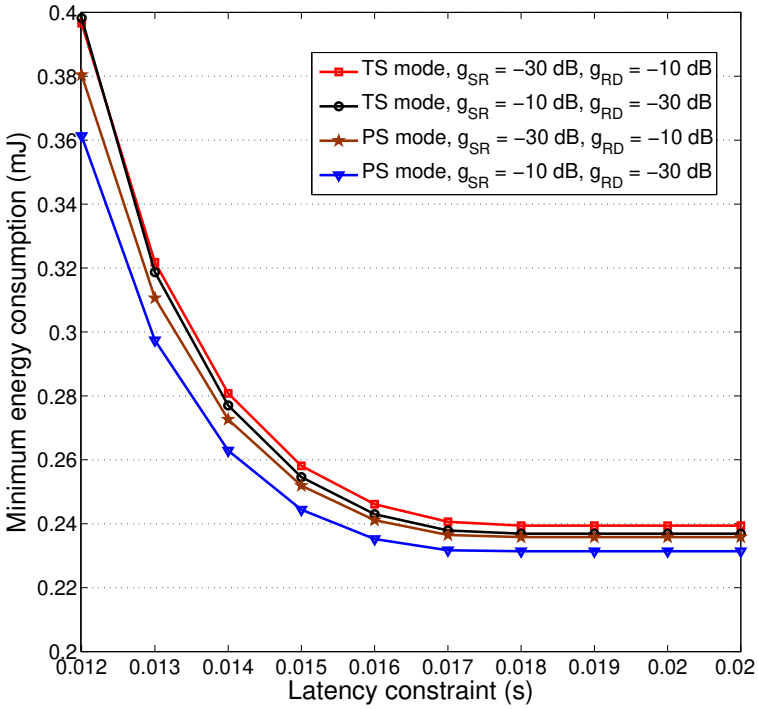

Figure 4.7: Effect of latency constraint on the energy consumption of the relay transmission system with optimal parameter values ($H = 5$ kbits).

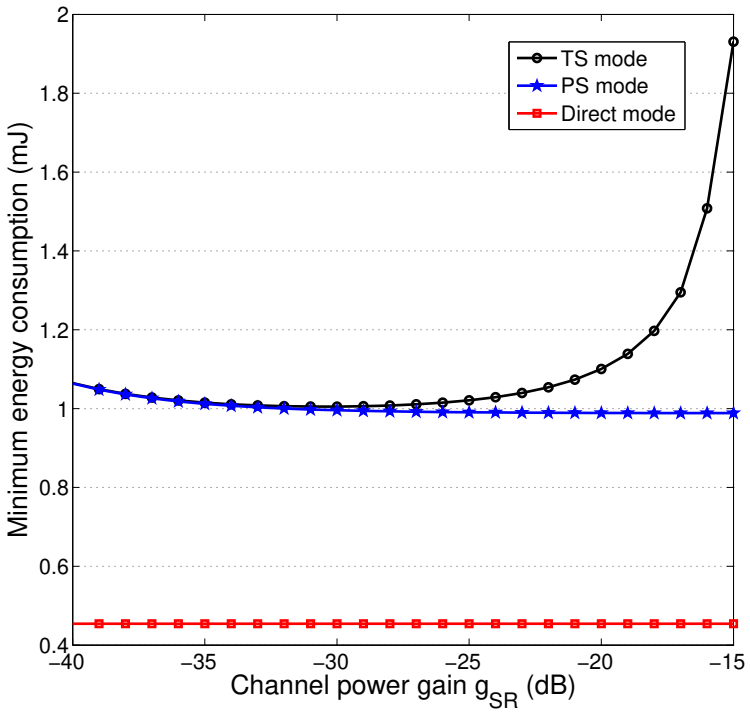

Figure 4.8: Minimum energy consumption comparison for power splitting, time switching, and direct modes under latency constraint ($g_{SD} = -50$ dB, $H = 5$ kbits, $T_{\text{th}} = 0.015$ s).

with power transfer requires much higher energy consumption than direct transmission. This is because that a much high transmission power is required for relay transmission system to satisfy the latency requirement. We also note that power splitting mode leads to lower energy consumption than time switching mode, especially when g_{SR} is large, which again justifies that power splitting mode is preferred in terms of energy consumption minimization.

4.3 Finite block-length transmission with nonlinear energy harvesting

In this section, we investigate the optimal design for the system of cooperative relaying transmission under finite blocklength data rate and nonlinear energy harvesting.

4.3.1 Generalized formulation

We consider a generic relaying transmission system with wireless power transfer, where time switching mode is used for the operation of relay. Specifically, a source node with M antennas transmits H bits data, by the help of a wireless powered DF relay, to a destination sensor, where direct link is ignored. We also assume that perfect CSI is available for the BS and the relay, and transmission task will fail if any error occurs. Total transmission duration is composed of three unequal time slots. In the first slot with the duration of T_{SC}, the source charges the relay with the power of P_{SC} using wireless power transfer. Applying a practical non-linear EH model from [106], while noting that optimal beamforming vector is $\vec{h}_{SR}^{*}/\left|\vec{h}_{SR}\right|$, the harvested energy of the relay is denoted by

$$E_H = P_H\left(\frac{1+\exp(-ab)}{1+\exp(-a(P_{SC}\left|\vec{h}_{SR}\right|^2-b))}-\exp(-ab)\right)T_{SC}, \tag{4.31}$$

where $\vec{h}_{SR}$ is the channel gain vector from the source to the relay, P_H denotes the maximum harvested power of EH circuit, a and b denote the constants related to EH circuit specifications.

In the second slot, the source transmits information to the relay with transmit power P_S. Applying optimal beamforming and finite block-length transmission, block transmission error rate from the source to the relay [132] is given by

$$\varepsilon_{SR} = Q\left((\log_2(1+\gamma_R)-H/n_{SR})/\left(\sqrt{\frac{\gamma_R(\gamma_R+2)}{2n_{SR}(1+\gamma_R)^2}}\log_2 e\right)\right), \tag{4.32}$$

where n_{SR} is the number of channel use from the source to the relay, γ_R is the received SNR at the relay, and $Q(\cdot)$ is the complementary gaussian CDF fuction. γ_R is denoted by

$$\gamma_R = P_S\left|\vec{h}_{SR}\right|^2/\sigma^2, \tag{4.33}$$

where σ^2 is average noise power.

In the last slot, the relay decodes received signal and forwards it to the destination using harvested energy. Following a similar process above, block transmission error rate from the relay to the destination is given by

$$\varepsilon_{RD} = Q\left((\log_2(1+\gamma_D)-H/n_{RD})/\left(\sqrt{\frac{\gamma_D(\gamma_D+2)}{2n_{RD}(1+\gamma_D)^2}}\log_2 e\right)\right). \tag{4.34}$$

Here, γ_D is the received SNR of the destination, and n_{RD} is the number of channel use from the relay to the destination. We assume that the duration of each channel use is equal to the inverse of the bandwidth [110], i.e. $1/B$. Since the energy for relaying operation is all harvested from the BS, the transmit power of the relay is equal to harvested energy over transmission duration, i.e. $E_H/(n_{RD}/B)$. Accordingly, γ_D can be denoted by

$$\gamma_D = E_H \frac{|h_{RD}|^2}{\sigma^2 n_{RD}/B}, \tag{4.35}$$

where h_{RD} denotes the channel gain from the relay to the destination.

Also note that block error will occur for DF relaying transmission in case that any one hop has transmission error. As such, the overall error probability of block transmission can be calculated by $\varepsilon = 1-(1-\varepsilon_{SR})(1-\varepsilon_{RD})$. The energy consumption of the source, during transmission process, is equal to $E_C = P_{SC}T_{SC} + P_S n_{SR}/B + P_C(T_{SC} + n_{SR}/B)$. With consideration of transmission errors, we use effective energy consumption to evaluate the performance of such system, defined as the ratio of energy consumption over successful transmission rate, i.e $E_C/(1-\varepsilon)$. For each transmission session, with instantaneous channel realization, all transmission parameters, i.e. T_{SC}, P_{SC}, P_S, n_{SR}, and n_{RD}, need to be optimally designed for minimizing the effective energy consumption. To avoid the outdatedness of channel knowledge and keep the effectiveness of information, we impose a latency requirement of T_C to this case. Accordingly, we formulate the optimization problem as follows

$$\begin{aligned} \min_{T_{SC},\, P_{SC},\, P_S,\, n_{SR},\, n_{RD}} \quad & E = \frac{P_{SC}T_{SC} + P_S n_{SR}/B + P_C(T_{SC} + n_{SR}/B)}{1-\varepsilon}, \\ s.t. \quad & 0 < P_{SC} \le P_{\max},\ 0 < P_S \le P_{\max},\ \tfrac{H}{n_{RD}} \le \log_2(1+\gamma_D), \\ & \tfrac{H}{n_{SR}} \le \log_2(1+\gamma_R),\ \tfrac{n_{SR}+n_{RD}}{B} + T_{SC} \le T_C. \end{aligned}$$

Note that this formulated problem is a mixed-integer nonlinear issue and its objective function is not jointly convex with respect to all continuous variables. Accordingly, the closed-form solutions are not possible to be derived. Although an iterative algorithm can be designed to arrive at near-optimal or sub-optimal solutions, it will consume valuable computing/power resource for iterative calculation during each online session. Additionally, the iterative algorithm may not converge under a strict latency constraint.

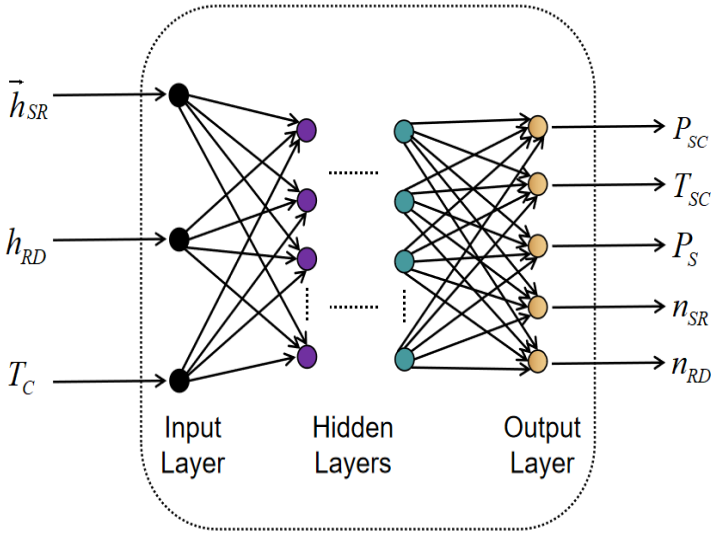

Figure 4.9: Deep policy network for relaying transmission.

4.3.2 Deep reinforcement learning formulation

Intuitively, supervised learning is a promising technique to train a deep policy network for determining near-optimal transmission parameters with minimum computation and latency. However, we need to generate numerous channel condition and find corresponding optimal parameter configurations as training data through exhaustive search. Note that numerical search, for these optimal configurations, will lead to low training efficiency.

As DRL does not involve any numerical search in offline training, it enjoys much higher efficiency than supervised learning. Accordingly, we apply DR-based approach, proposed in chapter 3, to train a near-optimal policy network. Particularly, under continuous state space and action space, we formulate an one-step markov decision process (MDP) problem. The state space $\mathbb{S}$ is defined as $\{T_C,\ \vec{h}_{SR},\ h_{RD}\}$ and the action space $\mathbb{R}$ is defined as $\{P_{SC},\ T_{SC},\ P_S,\ n_{SR},\ n_{RD}\}$, where the state information may vary from one transmission session to another. Our goal is to minimize the effective energy consumption and as such, the reward function is designed as the inverse of the above objective function, i.e. $R = \frac{H(1-\varepsilon)}{P_{SC}T_{SC}+P_S n_{SR}/B+P_C(T_{SC}+n_{SR}/B)}$. We assume that the policy network with the parameter of θ is denoted by π, as shown in Fig. 4.9. With input state vector $\vec{s} = [T_C,\ \vec{h}_{SR},\ h_{RD}]^T$ given, the output action vector $\vec{a} = [P_{SC},\ T_{SC},\ P_S,\ n_{SR},\ n_{RD}]^T$ can be also denoted by $\pi(\vec{s}|\theta)$.

Similar to chapter 3, we still employ policy gradient to gradually optimize the network parameter θ. In particular, we use the chain rule to calculate the gradient of reward function

with respect to θ, i.e. $\bigtriangledown_{\theta} R(\vec{s},\vec{a})|_{\vec{a}=\pi(\vec{s}|\theta)}$, shown as follows

$$\bigtriangledown_{\theta} R(\vec{s},\vec{a})|_{\vec{a}=\pi(\vec{s}|\theta)} = \bigtriangledown_{\vec{a}} R(\vec{s},\vec{a}) \bigtriangledown_{\theta} \pi(\vec{s}|\theta). \tag{4.36}$$

Here, $\bigtriangledown_{\vec{a}} R(\vec{s},\vec{a})$ can be calculated using an exact or approximate reward function and $\bigtriangledown_{\theta} \pi(\vec{s}|\theta)$ can be calculated using backpropagation method. Accordingly, within each gradient operation, the network parameter can be updated along the gradient direction of reward function.

Offline training

Generally, following a similar process shown in chapter 3, we can perform offline training for the above initialized policy network. Experience tuples, i.e. $\{\vec{s}_i,\ \vec{a}_i,\ R_i\}$ $(i = 1,\ 2, \cdots)$, can be generated to save into the memory buffer. Note that one existing tuple will be removed if the memory buffer reaches its capacity. Then, the gradient of reward function with respect to network parameter θ can be calculated still using mini-batch method, shown as below

$$\bigtriangledown_{\theta} R(\vec{s},\vec{a})|_{\vec{a}=\pi(\vec{s}|\theta)} \approx \quad \frac{1}{N} \sum_{i=1}^{N} \bigtriangledown_{\vec{a}} R(\vec{s},\vec{a})|_{\vec{s}=\vec{s}_i,\vec{a}=\pi(\vec{s}_i|\theta)} \bigtriangledown_{\theta} \pi(\vec{s}_i|\theta). \tag{4.37}$$

Accordingly, the parameter of the policy network can be updated to

$$\theta \leftarrow \theta + \xi \bigtriangledown_{\theta} R(\vec{s},\vec{a})|_{\vec{a}=\pi(\vec{s}|\theta)}, \tag{4.38}$$

where ξ denotes the learning rate of the policy network.

Similar to chapter 3, for performing random exploration, we add five Gaussian random variables into corresponding output actions [109], respectively, i.e. $v_{n_{SR}}$, $v_{T_{SC}}$, $v_{n_{RD}}$, v_{P_S} and $v_{P_{SC}}$. While noting that some constraints have to be strictly satisfied, the output actions of the policy network can be denoted by the following formulas, i.e. Eq. (4.39)- Eq. (4.43). In particular, n_{SR} is denoted by

$$n_{SR} = \lfloor \max\{1,\ \min\{\pi(\vec{s}|\theta)[1] + v_{n_{SR}},\ BT_C\}\} \rfloor, \tag{4.39}$$

where $v_{n_{SR}}$ is a Gaussian random variable with the mean value of zero and the variance value of $\sigma^2_{v_{n_{SR}}}$. T_{SC} is denoted by

$$T_{SC} = \max\{0,\ \min\{\pi(\vec{s}|\theta)[2] + v_{T_{SC}},\ T_C - n_{SR}/B\}\}, \tag{4.40}$$

where $v_{T_{SC}}$ is a Gaussian random variable with the mean value of zero and the variance value of $\sigma^2_{v_{T_{SC}}}$. n_{RD} is denoted by

$$n_{RD} = \lfloor \max\{1,\ \min\{\pi(\vec{s}|\theta)[3] + v_{n_{RD}},\ BT_C - n_{SR} - BT_{SC}\}\}\rfloor, \tag{4.41}$$

where $v_{n_{RD}}$ is a Gaussian random variable with the mean value of zero and the variance value of $\sigma^2_{v_{n_{RD}}}$. P_{SC} is denoted by

$$P_{SC} = \max\{P_{SCmin},\ \min\{\pi(\vec{s}|\theta)[4] + v_{P_{SC}},\ P_{\max}\}\}, \tag{4.42}$$

where $v_{P_{SC}}$ is a Gaussian random variable with the mean value of zero and the variance value of $\sigma^2_{v_{P_{SC}}}$. Here, P_{SCmin} can be derived from the constraint of $H/n_{RD} < \log_2(1+\gamma_D)$, equal to $\frac{ab}{\left|\vec{h}_{SR}\right|^2} - \frac{a}{\left|\vec{h}_{SR}\right|^2}\ln(\frac{P_H(1+\exp(ab))}{P_H+\exp(ab)\sigma^2(2^{\frac{H}{n_{RD}}}-1)n_{RD}/(B|h_{RD}|^2T_{SC})} - 1)$. P_S is denoted by

$$P_S = \max\{P_{Smin},\ \min\{\pi(\vec{s}|\theta)[5] + v_{P_S},\ P_{\max}\}\}, \tag{4.43}$$

where v_{P_S} is a Gaussian random variable with the mean value of zero and the variance value of $\sigma^2_{v_{P_S}}$. Here, P_{Smin} can be derived from the constraint of $H/n_{SR} < \log_2(1+\gamma_R)$, equal to $(2^{H/n_{SR}} - 1)\sigma^2 / \left|\vec{h_{SR}}\right|^2$.

Likewise, at the end of each iteration, the exploration variances of all output actions will be reduced by a factor of β ranging from 0 to 1.

The pseudo code of offline training is shown as Algorithm 3. After a certain amount of training iterations, the actor network can be well trained for the online operation. Note that as mentioned in chapter 3, the model is not necessarily convergent to optimal performance after one training session. Accordingly, we still need to repeat offline training around three times and pick the resulting model with highest average reward.

Online tuning

During the online operation, the source and the relay can use well-trained actor network to determine near-optimal transmission parameters in real time. However, as illustrated in chapter 3, since the models used for offline training may have some deviations from real models, well-trained actor network still needs to be further tuned for better adapting to the real environment. Generally, we can still follow a similar process, presented in chapter 3, to perform online tuning for offline trained policy. More specifically, the BS can collect online experience and use them to update the policy network. The critic network with the parameter of μ, used for approximating online reward function, is denoted by $Q(\vec{s},\vec{a}|\mu)$,

Algorithm 3 The pseudo code of offline training

Initialize the policy network and all parameters.
for t $\in [1,\ 2,\ \cdots,\ T]$ **do**
Generate a state information vector $\vec{s}_t$.
Feed the $\vec{s}_t$ to the policy network for obtaining an action vector $\vec{a}_t$ as Eq. (4.40)-Eq. (4.43).
Calculate corresponding reward value R_t based on given $\vec{s}_t$ and $\vec{a}_t$.
Save this state-action-reward tuple $\{\vec{s}_t,\ \vec{a}_t,\ R_t\}$ into the memory buffer, and randomly remove one existing tuple if necessary.
Randomly extract N state-action-reward tuples $\{\vec{s}_i,\ \vec{a}_i,\ R_i\}$ $(i = 1,\ 2, \cdots,\ N)$ from the memory buffer to update the parameter set of the policy network as Eq. (4.37) and Eq. (4.38).
Update exploration variances for all output actions by multiplying a factor of β.
end for

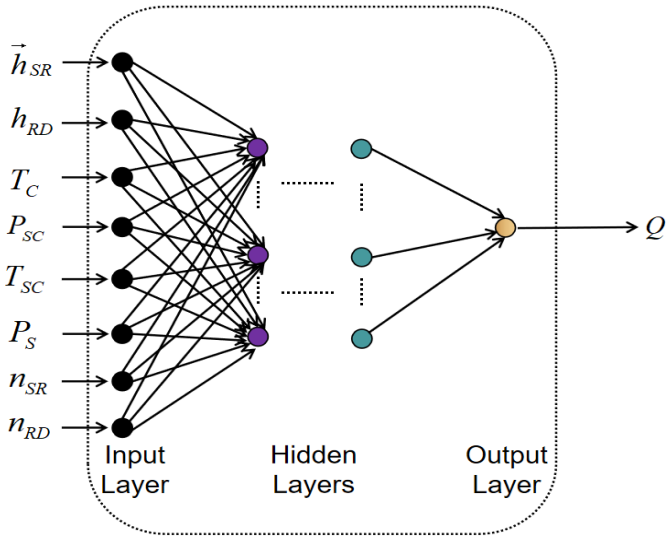

Figure 4.10: Deep critic network for approximating the reward during online operation.

as shown in Fig. 4.10. Such critic network can be constructed through supervised learning using offline experience, and then gradually updated using online experience. Note that since the gradient operation of offline reward function may be NAN in terms of some training data, the critic network is also employed to approximate the reward function during offline training.

According to the descriptions in chapter 3, in each tuning step, the source can collect N' online state-action-reward tuples $\{\vec{s}_i,\ \vec{a}_i,\ R'_i\}$ $(i = 1,\ 2,\ \cdots,\ N')$ from memory buffer and use them to update the critic network, shown as follows

$$\mu \leftarrow \mu + \frac{\rho}{N'} \sum_{i=1}^{N'} (R'_i - Q(\vec{s}_i, \vec{a}_i|\mu)) \bigtriangledown_\mu Q(\vec{s}_i, \vec{a}_i|\mu). \tag{4.44}$$

Here, ρ denotes the learning rate of the critic network. Then, we further update the parameter of the policy network to

$$\theta \leftarrow \theta + \frac{\xi}{N'} \sum_{i=1}^{N'} \bigtriangledown_{\vec{a}} Q(\vec{s}, \vec{a}|\mu)|_{\vec{s}=\vec{s}_i, \vec{a}=\pi(\vec{s}_i|\theta)} \bigtriangledown_\theta \pi(\vec{s}_i|\theta). \tag{4.45}$$

Here, ξ still denotes the learning rate of the policy network within the process of online tuning. On the other hand, as mentioned before, well-trained policy network should be already close to near-optimal performance. As such, after only a few tuning steps, offline trained policy model has a good chance to arrive at the near-optimal model of real environment.

The pseudo code of online tuning is presented as Algorithm 4.

Algorithm 4 The pseudo code of online tuning

Input the critic network and well-trained actor network
for collecting N' online experience tuples **do**
  Update the parameter of the critic network using Eq. (4.44).
  Update the parameter of the actor network using Eq. (4.45).
end for

4.3.3 Numerical results

In this section, we present some numerical examples to illustrate the effect of our proposed solution. The values of system parameters used in simulation are shown as Table I. The size of memory buffer is set to 10000, the size of online/offline mini-batch is set to 100, $\sigma^2_{n_{SR}}$ is set to 600, $\sigma^2_{n_{RD}}$ is set to 600, $\sigma^2_{I_{SC}}$ is set to 1.5, $\sigma^2_{P_S}$ is set to 2.7, $\sigma^2_{P_{SC}}$ is set to 2.7, β is set to 0.9995, ξ is set to 0.0000005, ρ is set to 0.05, P_H is set to 1 watt, a and b are

Table 4.2: System parameters of relaying transmission with generalized transmission and EH models.

Parameter	Meaning	Value
H	Transmitted data amount	2000 bits
σ^2	Average noise power	10^{-5} mWatt
B	Channel bandwidth	200 KHz
P_C	Circuit power consumption	10 mWatt
$P_{\max}$	Peak power	500 mWatt
η	DC energy conversion efficiency	0.8

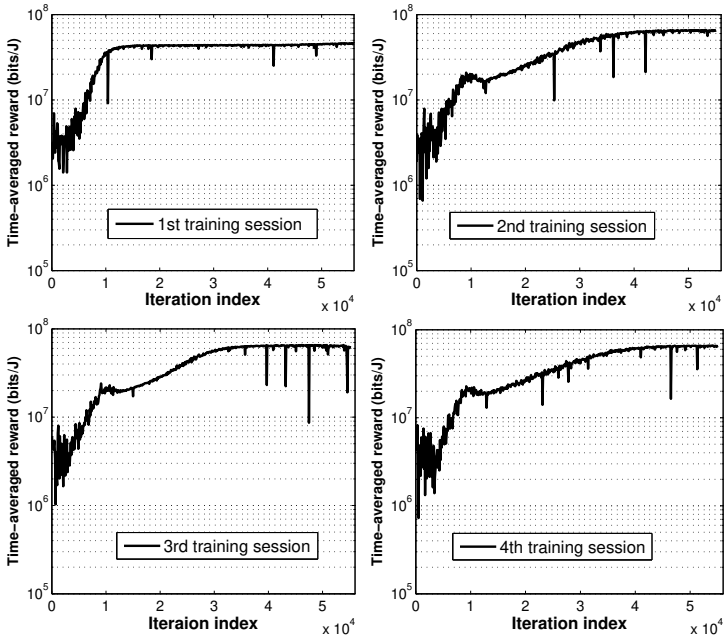

Figure 4.11: The time-averaged reward during the offline training.

both set to 1. Note that the initial parameter of deep networks are all randomly generated prior to offline training. In terms of the actor network, there are three hidden layers, where the number of neurons is 98, 80 and 78, respectively. Note that we approximately use the same number of tanh functions and relu functions as the activation functions in the hidden layers. As for the critic network, there is only one hidden layer with 128 neurons, where relu function is employed as the activation function.

Fig. 4.11 presents the time-average reward, over 100 training steps, as the function of the number of iteration step for different training sessions. We see that the reward converges to different values in different training sessions, where the initial fluctuation of curves is mainly caused by random exploration. We can also observe that even with random exploration, the offline training may converge to the local optimal solution, shown as the first training session. Note that we usually select the actor network with highest average reward as the final well-trained model.

Fig. 4.12 presents the effective energy consumption of the source as the function of channel power gain $g_{RD} = |h_{RD}|^2$ for exhaustive search, DRL based method, and gradient ascent, where $g_{SR} = \left|\vec{h}_{SR}\right|^2$ is set to -10 dB. The result from gradient ascent is obtained after several gradient operations, in which corresponding initial points are randomly generated. As expected, we observe that the effective energy consumption decreases when channel power gain g_{RD} increases, and the result of DRL based method is very close to

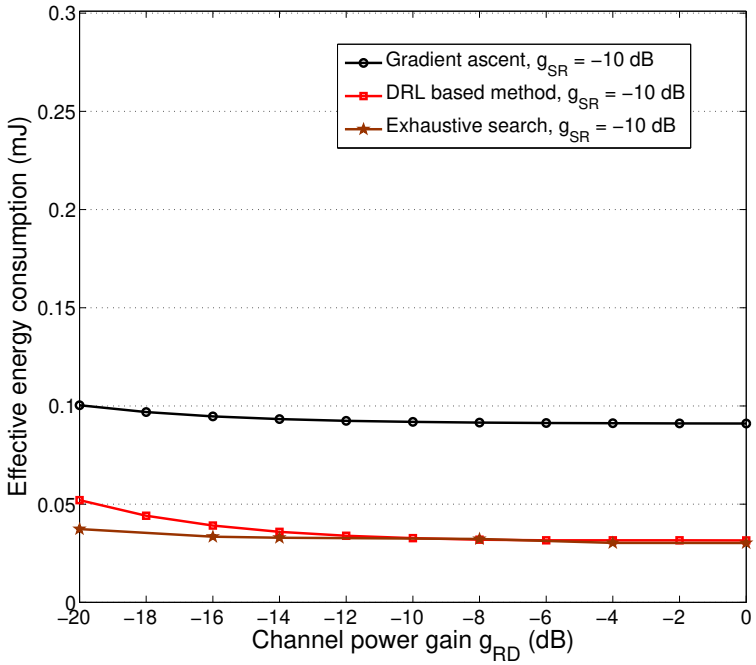

Figure 4.12: Effective energy consumption for different algorithms. ($T_C = 25$ ms)

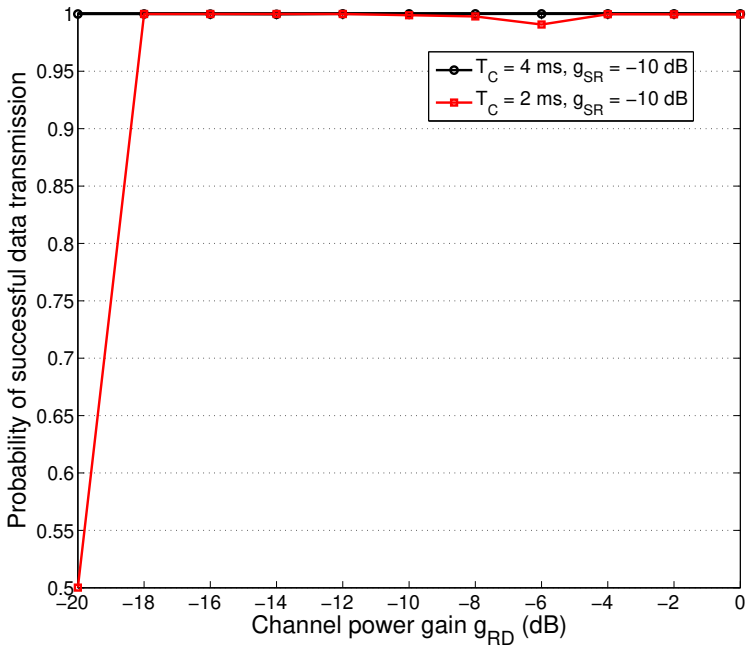

Figure 4.13: The successful block transmission rate.

that of exhaustive search. Accordingly, near optimal performance has been achieved by our proposed approach. Also, we see that the method of gradient ascent has the worst performance as compared to other two algorithms. While noting that the computational complexity is very high for gradient ascent and exhaustive search, they are not suitable for determining optimal transmission parameters in real time. Since there is not any online iterative calculation, we can show that DRL based method is a better option for IoT applications.

Fig. 4.13 presents the successful block transmission rate as the function of channel power gain g_{RD} for different latency requirements, where g_{SR} is set to -10 dB. As expected by our intuition, a more stringent latency constraint will considerably decrease the successful transmission rate, in particular for poor channel quality.

Fig. 4.14 presents the effective energy consumption as the function of the number of online tuning steps with and without online tuning. Average noise power σ^2, EH parameter a, and circuit power consumption are set to 1×10^{-5} mW, 1, and 10 mW within the offline training, while they are set to 2×10^{-5} mW, 1.5, and 30 mW within the online operation. We see that the effective energy consumption is considerably reduced after online tuning since accurate model parameters are used into the updation of the actor network. Additionally, a higher number of tuning steps can lead to better effect. Accordingly, we can show that online tuning can effectively improve the performance of the policy network for better

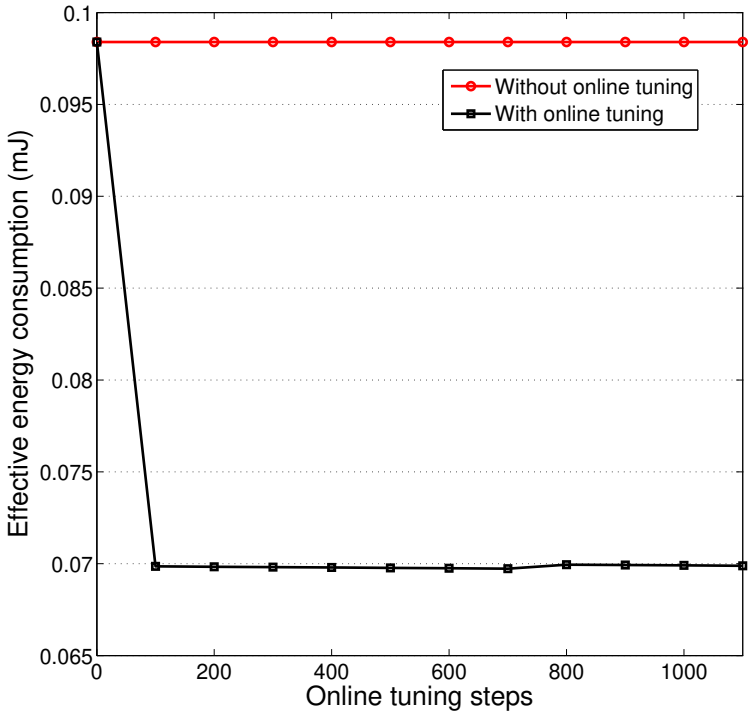

Figure 4.14: The expected energy consumption with/without online tuning. ($g_{SR} = g_{RD} = -10dB$, $T_C = 25$ ms)

adapting to the real environment.

4.4 Conclusion

In this chapter, we investigate the energy consumption of relaying transmission with wireless powered transfer from a session-specific perspective. Under ideal rate adaptive transmission with piecewise linear energy harvesting, in terms of both TS and PS modes, we analyzed and optimized the energy consumption of such system for transmitting a fixed amount of data. We derived the closed-form expressions of all optimal parameters with and without latency constraint. Under finite block-length transmission with nonlinear energy harvesting, we apply DRL-based method to obtain a policy model for determining the near-optimal transmission parameters in real time during the online operation. Selected numerical results are presented to illustrate various design tradeoffs for such system. Based on our analysis and proposed approach, transmitter can effectively determine near-optimal transmission parameters with low computation complexity and latency, for minimizing the energy consumption, during the process of practical wireless relaying transmissions. As such, they will greatly facilitate the design of green wireless relaying transmission system for advanced IoT.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

In this thesis, we analyzed and minimized the energy consumption of three IoT transmission scenarios from an individual transmission perspective, i.e. point-to-point transmission, data collection from wirelessly-powered sensor, and relaying transmission with wireless power transfer. For point-to-point transmission scenario, we followed a data oriented approach to optimize the energy consumption of a specific transmission session for ideal CPA and CRA schemes. The resulting EOR was also derived for both schemes. These results establish the limits of energy consumption for wireless transmissions over fading channels. For sensor data collection scenario, we optimized the transmission parameters to reduce the energy consumption of the agent. We derive the closed-form expressions of all optimal transmission parameters under ideal adaptive rate transmission with linear energy harvesting. Under practical finite block-length transmission with nonlinear energy harvesting, we propose a DRL-based method to train and adjust a deep policy network to determine the near-optimal transmission parameters in real time for the online operation of the agent. For relaying transmission with wireless power transfer scenario, we minimize the energy consumption of the source node through optimizing transmission parameters. Under ideal rate adaptive transmission with piecewise linear energy harvesting, we derive the closed form expressions for all optimal transmission parameters. Subsequently, the problem is again extended to finite block-length transmission with nonlinear energy harvesting, where we again apply DRL to train and tune a deep policy network to determine optimal transmission parameters for the online operation of source node.

With our proposed DRL-based approach, we can construct a complicated model for

a real communication system, and use it to train an optimal offline policy model. Then, obtained offline policy can be further tuned using real operation experience for arriving at higher accuracy. Accordingly, optimal parameter configurations for real-time wireless transmissions can be effectively designed to diminish consumed energy.

5.2 Future work

Over the past several years, I have performed much work on investigating the method for minimizing the energy consumption of link-level transmissions, i.e. point-to-point transmission and cooperative relaying transmission. However, optimal design for the network level transmissions has not be well investigated yet and as such, I plan to extend my research to learning based multi-tier heterogeneous communication network. The system configuration is shown as Fig. 5.1.

With the fast development of wireless communications, more and more energy will be consumed for numerous services, e.g. the exchange of large volume of data among a ever-increasing amount of users. Note that the cellular network, as an important architecture of communication technology, needs to consume a large amount of energy in real applications [133]. As such, the investigation, for enhancing the energy efficiency of communication networks, attracted a lot of attentions from both academia and industry over the past decade. The work in [134] proposed a statistical performance metric, i.e. generalized area spectral efficiency (GASE), to quantify the spectral utilization efficiency and energy efficiency for wireless transmissions. The work in [135] applied the metric of GASE to evaluate the energy efficiency performance of a wireless ad-hoc network with poisson distributed nodes, where co-channel interference was taken into account. However, energy efficiency was not optimized in the above two literatures. The work in [136] considered a heterogeneous cellular network with involving information decoding (ID) femto users (FUs) and energy harvesting FUs, where the scheme of transmit power allocation, under two beamformers, was optimized for maximizing the efficiency of energy harvesting and information transmission. The work in [137] studied and proposed energy and radio allocation mechanisms for enhancing the energy efficiency of cellular networks supplied with hybrid energy sources. The work in [138] considered device-to-device (D2D) communications underlying cellular networks, where an energy-efficient downlink resource reuse strategy was proposed to maximize the total energy efficiency of all D2D links under QoS constraints. Worthy of noting that fairness was not considered in the optimization of energy efficiency for the above [136–138]. To our best knowledge, in terms of multi-tier heteroge-

neous cellular network, fairness based analysis and maximization for the energy efficiency has not been investigated yet. I plan to solve such challenging research problem in the coming months.

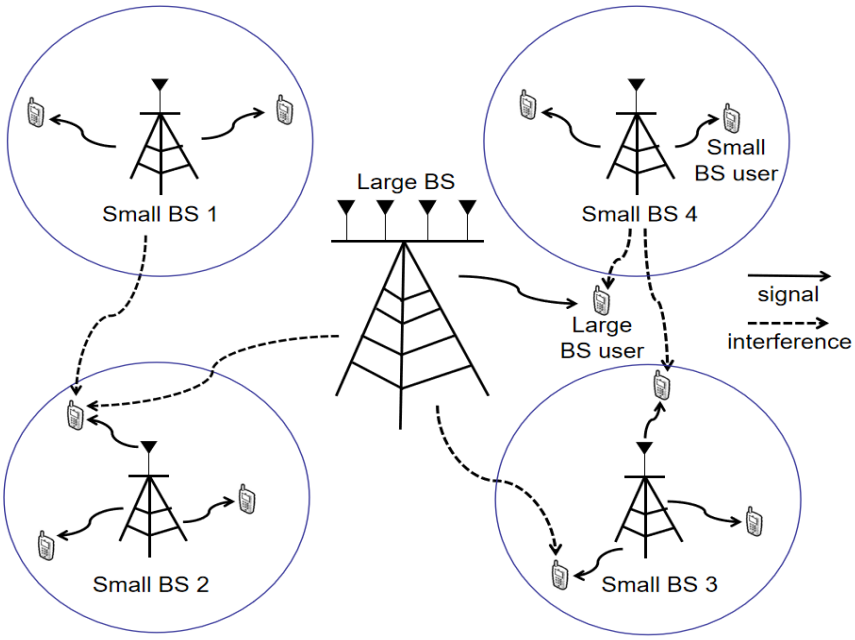

Figure 5.1: Two-tier heterogeneous cellular network.

Appendix A

The derivation process for Eq. (2.8)

The energy outage with CRA while using optimal transmit power occurs when

$$\frac{(P_c - \frac{N_0 B}{g})\frac{H \ln 2}{B}}{W_0[\frac{g}{N_0 B e}(P_c - \frac{N_0 B}{g})]} \geq E_{\text{th}}. \tag{A.1}$$

Noting that $W_0[\cdot]$ is always greater than -1, as illustrated in Fig. A.1, we can infer that energy outage will always occur if

$$(P_c - \frac{N_0 B}{g})\frac{H \ln 2}{B} \leq -E_{\text{th}}, \tag{A.2}$$

or equivalently,

$$g \leq \frac{N_0 B}{P_c + B E_{\text{th}}/(H \ln 2)}. \tag{A.3}$$

When $\frac{N_0 B}{P_c + B E_{\text{th}}/(H \ln 2)} < g \leq \frac{N_0 B}{P_c}$, energy outage occurs if and only if

$$(P_c - \frac{N_0 B}{g})\frac{H \ln 2}{B E_{\text{th}}} \geq W_0[\frac{g}{N_0 B e}(P_c - \frac{N_0 B}{g})]. \tag{A.4}$$

Note that $W_0[.]$ function is a monotonically increasing function. By some manipulations, we update Eq. (A.4) to

$$(-\frac{H N_0 \ln 2}{g E_{\text{th}}})\exp(-\frac{H N_0 \ln 2}{g E_{\text{th}}}) \leq -\exp(-\frac{H P_c \ln 2}{B E_{\text{th}}})/e. \tag{A.5}$$

Since $-1/e < -\exp(-\frac{H P_c \ln 2}{B E_{\text{th}}})/e < 0$, we can rewrite the above inequality, while referring

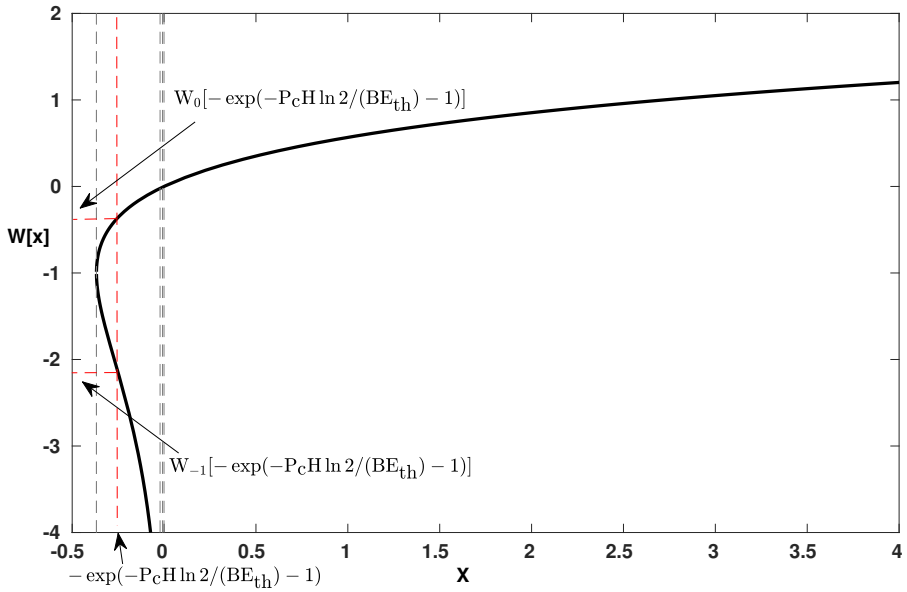

Figure A.1: The diagram for derivation of EOR of CRA with optimal power.

to Fig. A.1, to

$$W_{-1}[-\exp(-\frac{HP_c\ln 2}{BE_{\text{th}}})/e] \leq -\frac{HN_0\ln 2}{gE_{\text{th}}} \leq W_0[-\exp(-\frac{HP_c\ln 2}{BE_{\text{th}}})/e], \tag{A.6}$$

where $W_{-1}[\cdot]$ denotes the negative branch of Lambert W function [85].

When $g > \frac{N_0B}{P_c}$, we can also follow a similar process to show that energy outage occurs if and only if Eq. (A.6) holds.

Combining results above, the energy outage of CRA with optimal transmit power occurs when either

$$\frac{-HN_0\ln 2}{E_{\text{th}}W_{-1}[-\exp(-\frac{HP_c\ln 2}{BE_{\text{th}}})/e]} \leq g \leq \frac{-HN_0\ln 2}{E_{\text{th}}W_0[-\exp(-\frac{HP_c\ln 2}{BE_{\text{th}}})/e]}, \tag{A.7}$$

or

$$g \leq \min\{\frac{N_0B}{P_c+BE_{\text{th}}/(H\ln 2)},\ -\frac{HN_0\ln 2}{E_{\text{th}}W_{-1}[-\exp(-\frac{HP_c\ln 2}{BE_{\text{th}}})/e]}\}. \tag{A.8}$$

As such, the corresponding EOR can be calculated using the CDF of the channel power gain g as in (2.8).

Appendix B

Lambert W Function

In this appendix, we mainly illustrate the property of Lambert W function. It will be helpful for readers to better understand the theoretical derivations in this proposal. By assuming $x = ye^y$, we can write y as a function of x, shown as follows

$$y = W[x], \tag{B.1}$$

where $W[\cdot]$ denotes the Lambert W function. Fig. B.1 presents the curves of this function with respect to x. We can see the principal branch function $W_0[x]$ and negative branch function $W_{-1}[x]$, where $W_0[x]$ is a monotonically increasing function and $W_{-1}[x]$ is a monotonically decreasing function. Also, we observe that the domain of $W[\cdot]$ is from $-\exp(-1)$ to $+\infty$, $W_0[x]$ is always greater than -1, and $W_{-1}[x]$ is always less than -1. Note that with this function, there are two function values corresponding to one real variable for $x > -\exp(-1)$.

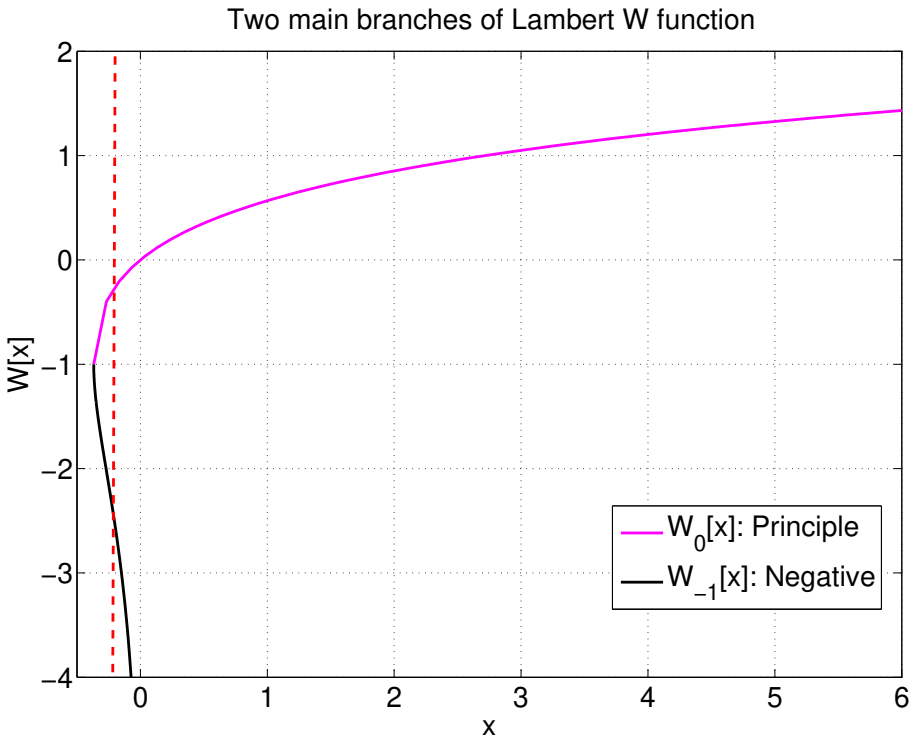

Figure B.1: The Lambert W function.

Appendix C

List of publications

[J1]. Hong-Chuan Yang, Fang Xu, Mohamed-Slim Alouini, "Statistical Energy Efficiency Characterization for Wireless Transmission Over Fading Channels," *IEEE Transactions on Vehicular Technology,* vol. 69, no. 11, pp. 13947 - 13951, nov. 2020. (Corresponding author)

[J2]. Fang Xu, Hong-Chuan Yang, Mohamed-Slim Alouini, "Energy-efficient Data Collection from Wirelessly-powered IoT Sensors: Session-Specific Optimal Design with DRL," *IEEE Sensors Journal*, Under review.

[J3]. Fang Xu, Hong-Chuan Yang, Mohamed-Slim Alouini, "Ultra-green Relay Transmission with Wireless Power Transfer for Advanced IoT: Session-Specific Analysis and Optimization," *IEEE Internet of Things Journal*, Under reivew.

[C1]. Fang Xu, and Hong-Chuan Yang, "Session-specific Energy Consumption Minimization for UAV Enabled Data Collection System," *IEEE Wireless Communications & Networking Conference (WCNC),* April. 2022, Accepted.

[C2]. Fang Xu, Hong-Chuan Yang, Mohamed-Slim Alouini, "Session-specific Energy-efficient Relaying Transmission With Wireless Power Transfer: Optimal Design With DRL," Submitted to IEEE Global Communications Conference (Globecom) 2022.

Bibliography

[1] J. Chase, “The evolution of the Internet of Things,” *Texas Instrum,* Dallas, TX, USA, White Paper, pp. 1-7, Feb. 2013.

[2] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, “Context Aware Computing for The Internet of Things:A Survey,” *IEEE Communications Surveys Tutorials,*vol. 16, no. 1, pp. 414-445, 2014.

[3] M. A. M. Albreem, A. A. E. Saleh, M. Isa, W. Salah, M. Jusoh, M. M. Azizan, A. Ali, “Green internet of things (IoT): An overview,” *2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA),* pp. 1-6, Mar. 2018.

[4] Green Power for Mobile, ”The Global Telecom Tower ESCO Market,” Technical Report, 2015.

[5] F. J. Velez, L. M. Correia, “Mobile Broadband Services: Classification, Characterization, and Deployment Scenarios,” *IEEE Communications Magazine ,* vol. 40, pp. 142-150, Aug. 2002.

[6] G. Miao et al., “Cross-Layer Optimization for Energy-Efficient Wireless Communications: A Survey,” *Wiley J. Wireless Commun. Mobile.,* vol. 9, no. 4, pp. 529-542, Jan. 2010.

[7] A. J. Goldsmith and S. B. Wicker, “Design challenges for energy constrained ad hoc wireless networks,” *IEEE Wireless Communications.,* vol. 9, no. 4, pp. 8-27, Aug. 2002.

[8] A. Alabbasi, and B. Shihada, “Optimal Cross-Layer Design for Energy Efficient D2D Sharing Systems,” *IEEE Trans. Wireless Commun.,* vol. 16, no. 2, pp. 839-855, Feb. 2017.

[9] G. Miao, N. Himayat, G. Y. Li, A. T. Koc, and S. Talwar, "Interference-Aware Energy-Efficient Power Optimization," *2009 IEEE International Conference on Communications.,* pp. 1-5, Aug. 2009.

[10] V. Khodamoradi, A. Sali, O. Messadi, A. A. Salah, M. M. A. Wani , B. M. Ali, and R. S. A. R. Abdullah, "Optimal Energy Efficiency Based Power Adaptation for Downlink Multi-Cell Massive MIMO Systems," *IEEE Access.,* vol. 8, pp. 203237-203251, Nov. 2020.

[11] G. Miao, N. Himayat and G. Y. Li, "Energy-Efficient Link Adaptation in Frequency-Selective Channels," *IEEE Trans. Commun.,* vol. 58, no. 2, pp. 545-554, Feb. 2010.

[12] G. Ozcan and M. C. Gursoy, "Energy-Efficient Power Adaptation for Cognitive Radio Systems under Imperfect Channel Sensing," *IEEE Conference on Computer Communications Workshops.,* pp. 706-711, Jul. 2014.

[13] M. Bohge, J. Gross, and A. Wolisz, "Energy-Efficient Design in Wireless OFDMA," *2008 IEEE Network.,* vol. 21, no. 1, pp. 53–59, Feb. 2007.

[14] T. C. Y. Ng and W. Yu, "Joint Optimization of relay Strategies and Resource Allocations in Cooperative Cellular Networks," *IEEE J. Select. Areas Commun.,* vol. 25, no. 2, pp. 328–339, Feb 2007.

[15] Z. Liu, X. Tao, W. u. Rehman, "Resource Allocation for Two-Way Amplify and Forward OFDM Relay Networks," *China Communications.,* vol. 14, no. 8, pp. 76-82, Aug 2017.

[16] F. Zeng, Z. Hu, Z. Xiao, H. Jiang, S. Zhou, W. Liu and D. Liu, "Resource Allocation and Trajectory Optimization for QoE Provisioning in Energy-Efficient UAV-Enabled Wireless Networks ," *China Communications.,* vol. 69, no. 7, pp. 7634-7647, Jul 2020.

[17] Q. Liu, T. Lv and Z. Lin, "Energy-Efficient Transmission Design in Cooperative Relaying Systems Using NOMA," *IEEE Commun. Lett.,* vol. 22, no. 3, pp. 594–597, Mar 2018.

[18] A. Helmy, L. Musavian, and T. L. Ngoc, "Energy-Efficient Power Adaptation over a Frequency-Selective Fading Channel with Delay and Power Constraints," *IEEE Trans. Wireless Commun.,* vol. 12, no. 9, pp. 4529-4541, Sep. 2013.

[19] G. Miao, N. Himayat, Y. Li, and D. Bormann, "Energy-Efficient Design in Wireless OFDMA," *2008 IEEE International Conference on Communications.,* pp. 3307-3312, May. 2008.

[20] S. Cui, A. J. Goldsmith and A. Bahai, "Energy-Efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks," *IEEE J. Select. Areas Commun.,* vol. 22, no. 6, pp. 1089–1098, Aug 2004.

[21] 3GPP, R1-101084, "Energy Saving Techniques to Support Low Load Scenarios," *www.3gpp.org.,* Huawei, Tech. Rep., 2010.

[22] H. Kim, C. B. Chae, G. d. Veciana and R. W. H. Jr, "A Cross-Layer Approach to Energy Efficiency for Adaptive MIMO Systems Exploiting Spare Capacity," *IEEE Trans. Wireless Commun.,* vol. 8, no. 8, pp. 4264-4275, Aug. 2009.

[23] B. Bougard, G. Lenoir, A. Dejonghe, L. V. d. Perre, F. Catthoor and W. Dehaene , "An Energy-Aware Adaptive MIMO-OFDM Radio Link Control for Next-Generation Wireless Local Area Networks," *EURASIP Journal on Wireless Communications and Networking.,* vol. 2007, no. 3, pp. 1–15, June 2007.

[24] Z. Zhou, S. Zhou, J. H. Cui and S. Cui, "Energy Efficient Cooperative Communication Based on Power Control and Selective Single-Relay in Wireless Sensor Networks," *IEEE Trans. Wireless Commun.,* vol. 7, no. 8, pp. 3066–3078, Aug 2008.

[25] R. Madan, N. B. Mehta, A. F. Molisch and J. Zhang, "Energy-Efficient Cooperative Relaying over Fading Channels with Simple Relay Selection," *IEEE Trans. Wireless Commun.,* vol. 7, no. 8, pp. 3013–3025, Aug 2008.

[26] J. Kim, K. Kim, and J. Lee, "Energy-Efficient Relay Selection of Cooperative HARQ Based on the Number of Transmissions Over Rayleigh Fading Channels," *IEEE Trans. Veh. Technol.,* vol. 66, no. 1, pp. 610–621, Jan 2017.

[27] Y. Chen, X. Fang, and B. Huang, "Energy-Efficient Relay Selection and Resource Allocation in Nonregenerative Relay OFDMA Systems," *IEEE Trans. Veh. Technol.,* vol. 63, no. 8, pp. 3689–3699, Oct 2014.

[28] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal,* vol. 1, no. 1, pp. 22-32, Feb. 2014.

[29] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Communications Magazine,*vol. 49, no. 6, pp. 30-37, June 2011.

[30] M. H. Alsharif, R. Nordin, and M. Ismail, "Survey of green radio communications networks: techniques and recent advances," *Journal of Computer Networks and Commun.,* vol. 2013, 2013.

[31] F. Rosas and C. Oberli, "Modulation and SNR optimization for achieving energy-efficient communications over short-range fading channels," *IEEE Transactions on Wireless Communications,* vol. 11, no. 12, pp. 4286-4295, December 2012.

[32] M.-S. Alouini and A. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *IEEE Trans. Vehicular Technol.,* vol. 48, no. 4, pp. 1047-1066, Jul 1999.

[33] L. Zhang, H.-C. Yang, and M. O. Hasna, "Generalized area spectral efficiency: An effective performance metric for green wireless communications," *IEEE Trans. on Commun.,* vol. 62, no. 2 pp. 747-757, February 2014.

[34] H.-C. Yang and M. S. Alouini, "Data-oriented transmission in future wireless systems: Toward trustworthy support of advanced Internet of Things" in *IEEE Veh. Technol. Mag.,* pp. 1- 1, Jul 2019.

[35] W. Wang, H.-C Yang, and M. S. Alouini, "Energy consumption analysis for adaptive transmission of big data over fading channels: A statistical characterization," *IEEE Trans. on Green Commun. and Networking,*vol. 4, no. 2, pp. 365-374, Jun. 2020.

[36] M. C. Gursoy, D. Qiao, and S. Velipasalar, "Analysis of energy efficiency in fading channels under QoS constraints," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 8, pp. 4252-4263, August 2009.

[37] A. Helmy, L. Musavian and T. Le-Ngoc, "Energy-efficient power adaptation over a frequency-selective fading channel with delay and power constraints," *IEEE Trans. on Wireless Commun.*, vol. 12, no. 9, pp. 4529-4541, September 2013.

[38] D. Qiao, "The impact of statistical delay constraints on the energy efficiency in fading channels," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 2, pp. 994-1007, February 2016.

[39] W. Cheng, X. Zhang, and H. Zhang, "Statistical-QoS driven energy-efficiency optimization over green 5G mobile wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3092-3107, December 2016.

[40] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel Coding Rate in the Finite Blocklength Regime," *IEEE Trans. Inf. Theory.*, vol. 56, no. 5, pp. 2307-2359, May 2010.

[41] J. Ho, J. Meng, and E. h. Yang, "On Separation of Source and Channel Coding in the Finite Block Length Regime," *13th Canadian Workshop on Information Theor.*, pp. 92-95, Jun 2013.

[42] F. Wang, J. Jiao, K. Zhang, S. Wu, Y. Li, and Q. Zhang, "Self-Adaptive Ordered Statistics Decoder for Finite Block Length Raptor Codes Toward URLLC," *IEEE Internet of Things Journal.*, vol. 9, no. 5, pp. 3282-3297, Mar 2022.

[43] B. Makki, T. Svensson, and M. Zorzi, "Finite Block-Length Analysis of Spectrum Sharing Networks: Interference-Constrained Scenario," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 433-436, Aug 2015.

[44] J. Cha, J. Choi, and D. J. Love "Practical Distributed Reception for Wireless Body Area Networks Using Supervised Learning," in *IEEE Trans. Wireless Commun.,* pp. 1-12, Dec. 2021 (Accepted for publication).

[45] R. C. Daniels, C. M. Caramanis, and R. W. Heath "Adaptation in Convolutionally Coded MIMO-OFDM Wireless Systems Through Supervised Learning and SNR Ordering," in *IEEE Trans. Veh. Technol.,* Vol. 59, No. 1, pp. 114-126, Jan. 2010.

[46] Y. S. Jeon, S. N. Hong, and N. Lee "Supervised-Learning-Aided Communication Framework for MIMO Systems With Low-Resolution ADCs," in *IEEE Trans. Veh. Technol.,* Vol. 67, No. 8, pp. 7299-7313, Aug. 2018.

[47] Y. Wang, and K. Zhang "Decision Tree Based Unsupervised Learning to Network Selection in Heterogeneous Wireless Networks," in *The 8th Annual IEEE Consumer Communications and Networking Conference.,* pp. 1108-1109, Aug. 2018.

[48] J. Gao, C. Zhong, X. Chen, H. Lin, and Z. Zhang "Unsupervised Learning for Passive Beamforming," in *The 8th Annual IEEE Consumer Communications and Networking Conference.,* Vol. 24, No. 5, pp. 1052-1056, May. 2020.

[49] C. Di, B. Zhang, Q. Liang, S. Li, and Y. Guo "Learning Automata-Based Access Class Barring Scheme for Massive Random Access in Machine-to-Machine Communications," in *IEEE Internet Of Things Journal.,* Vol. 6, No. 4, pp. 6007-6017, Aug. 2019.

[50] S. K. Sharma, and X. Wang "Toward Massive Machine Type Communications in Ultra-Dense Cellular IoT Networks: Current Issues and Machine Learning-Assisted Solutions," in *IEEE Communications Surveys and Tutorials.,* Vol. 22, No. 1, pp. 426-471, 2020.

[51] W. S. Liao, O. Zhao, M. G. Kibria, G. P. Villardt, K. Ishizu, and F. Kojima "Machine Learning-Based Signal Detection for CoMP Downlink in Ultra-Dense Small Cell Networks," in *IEEE Access.,* Vol. 8, pp. 17454-17463, 2020.

[52] J. Huang, C. X. Wang, L. Bai, J. Sun, Y. Yang, J. Li, O. Tirkkonen, and M. T. Zhou "A Big Data Enabled Channel Model for 5G Wireless Communication Systems," in *IEEE Transactions On Big Data.,* Vol. 6, No. 2, pp. 211-222, Jun. 2020.

[53] S. Liu, L. Xiao, L. Huang, and X. Wang "Impulsive Noise Recovery and Elimination: A Sparse Machine Learning Based Approach," in *IEEE Trans. Veh. Technol.,* Vol. 68, No. 3, pp. 2306-2315, Mar. 2019.

[54] C. Pandana and K. J. R. Liu "Near-Optimal Reinforcement Learning Framework for Energy-Aware Sensor Communications," in *IEEE Journal On Selected Areas In Communications.,* Vol. 23, No. 4, pp. 788-797, Apr. 2015.

[55] J. Hu, H. Zhang, L. Song, Z. Han, and H. V. Poor "Reinforcement Learning for a Cellular Internet of UAVs: Protocol Design, Trajectory Control, and Resource Management," in *IEEE Wireless Communications.,* Vol. 27, No. 1, pp. 116-123, Feb. 2020.

[56] D. Wang, X. Tian, H. Cui, Z. Liu "Reinforcement Learning-Based Joint Task Offloading and Migration Schemes Optimization in Mobility Aware MEC Network," in *China Communications.,* Vol. 17, No. 8, pp. 31-44, Aug. 2020.

[57] H. Jiang, R. Gui, Z. Chen, L. Wu, J. Dang, and J. Zhou "An Improved Sarsa() Reinforcement Learning Algorithm for Wireless Communication Systems," in *IEEE Access.,* Vol. 7, pp. 115418-115427, Aug. 2019.

[58] Q. Zhou, Y. Li, and Y. Niu "Intelligent Anti-Jamming Communication for Wireless Sensor Networks: A Multi-Agent Reinforcement Learning Approach," in *IEEE Access.,* Vol. 2, pp. 775-784, Feb. 2021.

[59] T. Y. Tung, S. Kobus, J. P. Roig, and D. Gunduz "Effective Communications: A Joint Learning and Communication Framework for Multi-Agent Reinforcement Learning Over Noisy Channels," in *IEEE J. Sel. Areas Commun.,* Vol. 39, No. 8, pp. 2590-2603, Aug. 2021.

[60] T. Degris, M. White, R. S. Sutton "Off-Policy Actor-Critic," in *Proceedings of the 29th International Conference on Machine Learning.,* pp. 179–186, Jun. 2012.

[61] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra "Continuous Control with Deep Reinforcement Learning," in *International Conference on Learning Representations.,* pp. 1-14, May. 2016.

[62] W. Xu, H. Lei, and J. Shang "Joint Topology Construction and Power Adjustment for UAV Networks: A Deep Reinforcement Learning Based Approach," in *China Communications.,* Vol. 18, pp. 265-283, Jul. 2021.

[63] Y. Su , X. Lu, Y. Zhao, L. Huang, and X. Du "Cooperative Communications With Relay Selection Based on Deep Reinforcement Learning in Wireless Sensor Networks," in *IEEE sensors journal.,* Vol. 19, No. 20, pp. 9561-9569, Oct. 2019.

[64] H. Yang, Z. Xiong, J. Zhao, D. Niyato, C. Yuen, and R. Deng "Deep Reinforcement Learning Based Massive Access Management for Ultra-Reliable Low-Latency Communications," in *IEEE Trans. Wireless Commun.,* Vol. 20, No. 5, pp. 2977-2990, may. 2021.

[65] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu "Deep Reinforcement Learning-Based Intelligent Reflecting Surface for Secure Wireless Communications," in *IEEE Trans. Wireless Commun.,* Vol. 20, No. 1, pp. 375-388, jan. 2021.

[66] Y. A. Eryani, M. Akrout, and E. Hossain "Antenna Clustering for Simultaneous Wireless Information and Power Transfer in a MIMO Full-Duplex System: A Deep Reinforcement Learning-Based Design," in *IEEE Trans. Wireless Commun.,* Vol. 69, No. 4, pp. 2331-2345, Apr. 2021.

[67] K. K. Nguyen, T. Q. Duong, N. A. Vien, N. A. L. Khac, and M. N. Nguyen "Non-Cooperative Energy Efficient Power Allocation Game in D2D Communication: A Multi-Agent Deep Reinforcement Learning Approach," in *IEEE Access.,* Vol. 7, pp. 100480-100490, Aug. 2019.

[68] Y. Li, W. Zhang, C. X. Wang, J. Sun, and Y. Liu "Deep Reinforcement Learning for Dynamic Spectrum Sensing and Aggregation in Multi-Channel Wireless Networks," in *IEEE Trans. Cogn. Commun. Netw.,* Vol. 6, no. 2, pp. 464-475, Jun. 2020.

[69] C. Qiu, Y. Hu, Y. Chen, and B. Zeng "Deep Deterministic Policy Gradient (DDPG)-Based Energy Harvesting Wireless Communications," in *IEEE Internet Of Things Journal.,* Vol. 6, No. 5, pp. 8577-8588, OCT. 2019.

[70] O. Bouhamed, H. Ghazzai, H. Besbes, and Y. Massoud "A UAV-Assisted Data Collection for Wireless Sensor Networks: Autonomous Navigation and Scheduling," in *IEEE Access.,* Vol. 8, pp. 110446-110460, Jun. 2020.

[71] C. H. Liu, Z. Chen, and Y. Zhan "Energy-Efficient Distributed Mobile Crowd Sensing: A Deep Learning Approach," in *IEEE Journal On Selected Areas In Communications.,* vol. 37, No. 6, pp. 1262-1276, Jun. 2019.

[72] Y. Cao, S. Y. Lien, Y. C. Liang, K. C. Chen, and X. Shen, "User Access Control in Open Radio Access Networks: A Federated Deep Reinforcement Learning Approach," in *IEEE Trans. Wireless Commun.,* pp. 1-1, Nov. 2021 (Accepted for publication).

[73] Y. Su , X. Lu, Y. Zhao, L. Huang, and X. Du "Cooperative Communications With Relay Selection Based on Deep Reinforcement Learning in Wireless Sensor Networks," in *IEEE Sensors Journal.,* Vol. 19, No. 20, pp. 9561-9569, Oct. 2019.

[74] L. Li, D. Ma, H. Ren, D. Wang, X. Tang, W. Liang, and T. Bai "Enhanced Reconfigurable Intelligent Surface Assisted mmWave Communication: A Federated Learning Approach," in *China Communications.,* Vol. 17, pp. 115-128, Oct. 2020.

[75] A. Elgabli, J. Park, C. B. Issaid, and M. Bennis "Harnessing Wireless Channels for Scalable and Privacy-Preserving Federated Learning," in *IEEE Trans. Commun.,* Vol. 69, No. 8, pp. 5194-5208, Aug. 2021.

[76] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. S. Bahaei "Energy Efficient Federated Learning Over Wireless Communication Networks," in *IEEE Trans. wireless Commun.,* Vol. 20, No. 3, pp. 1935-1949, Mar. 2021.

[77] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui "A Joint Learning and Communications Framework for Federated Learning Over Wireless Networks," in *IEEE Trans. wireless Commun.,* Vol. 20, No. 1, pp. 269-283, Jan. 2021.

[78] M. K. A. Aziz, C. Perfecto, S. Samarakoon, M. Bennis, and W. Saad "Vehicular Cooperative Perception Through Action Branching and Federated Reinforcement Learning," in *IEEE Trans. Commun.,* pp. 1-1, Nov. 2021 (Accepted for publication).

[79] Y. Zeng and R. Zhang, "Full-Duplex Wireless-Powered Relay With Self-Energy Recycling," in *IEEE Wireless Communications Lett.,* vol. 4, No. 2, pp. 136-146, Apr 2015.

[80] R. Landauer, "Minimal energy requirements in communication," *Science,* vol. 272, pp. 1914-1918, Jun 1996.

[81] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.,*vol. 4, no. 5, pp. 2349-2360, Sep. 2005.

[82] A. Goldsmith, *Wireless Communications*, Cambridge University Press., 2005.

[83] H.-C. Yang and M.-S. Alouini, "Data-oriented transmission in future wireless systems: Toward trustworthy support of advanced Internet of Things," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 78 - 83, 2019.

[84] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[85] F. Blondeau, A. Monir, "Evaluation of the Lambert W function and application to generation of generalized Gaussian noise with exponent 1/2," *IEEE Trans. Signal Process.,* vol. 50, No. 9, pp. 2160-2165, Sep 2002.

[86] S. O. Rice, "Distribution of the duration of fades in radio transmission: Gaussian noise model," *The Bell System Technical Journal.,* vol. 37, no. 3, pp. 581-635, May 1958.

[87] Y. Fathy and P. Barnaghi, "Quality-Based and Energy-Efficient Data Communication for the Internet of Things Networks," *IEEE Internet of Things Journal,* vol. 6, no. 6, pp. 10318- 10331, Dec. 2019.

[88] K. Li, W. Ni, L. Duan, M. Abolhasan, and J. Niu, "Wireless Power Transfer and Data Collection in Wireless Sensor Networks," *IEEE Transactions On Industrial Informatics,* vol. 67, no. 3, pp. 2686- 2697, Mar 2018.

[89] Y. Shu, H. Yousefi, P. Cheng, J. Chen, Yu. Gu, T. He, and K. G. Shin, "Near-Optimal Velocity Control for Mobile Charging in Wireless Rechargeable Sensor Networks," *IEEE Trans. Mobile Comput,* vol. 15, no. 7, pp. 1699- 1713, Jul 2016.

[90] Z. Wang, L. Duan, and R. Zhang,, "Adaptively Directional Wireless Power Transfer for Large-Scale Sensor Networks," *IEEE Trans. Mobile Comput,* vol. 34, no. 5, pp. 1785- 1800, May 2016.

[91] L. Fu, P. Cheng, Y. Gu, J. Chen, and T. He,, "Optimal Charging in Wireless Rechargeable Sensor Networks," *IEEE Trans. Veh. Technol,* vol. 65, no. 1, pp. 278- 291, Jan 2016.

[92] X. Zhang, X. Zhang, and L. Han, "An Energy Efficient Internet of Things Network Using Restart Artificial Bee Colony and Wireless Power Transfer," in *IEEE Access,* vol. 7, pp. 12686-12695, Jan. 2019.

[93] P. Wu, F. Xiao, H. Huang, C. Sha, and S. Yu "Adaptive and Extensible Energy Supply Mechanism for UAVs-aided Wireless Powered Internet of Things," in *IEEE Internet of Things Journal,* vol. 7, no. 9, pp. 9201-9213, Jun. 2020.

[94] G. Li, S. Peng, C. Wang, J. Niu, and Y. Yuan, "An Energy-Efficient Data Collection Scheme Using Denoising Autoencoder in Wireless Sensor Networks," *Tsinghua Science And Technology,* vol. 24, no. 1, pp. 86-96, Feb 2019.

[95] W. Wen, S. Zhao, C. Shang, and C. Y. Chang, "EAPC: Energy-Aware Path Construction for Data Collection Using Mobile Sink in Wireless Sensor Networks," *IEEE Sensors Journal,* vol. 18, no. 2, pp. 890-901, Jan 2018.

[96] T. Wang, L. Qiu, A. K. Sangaiah, G. Xu, and A. Liu, "Energy-Efficient and Trustworthy Data Collection Protocol Based on Mobile Fog Computing in Internet of Things," *IEEE Transactions On Industrial Informatics,* vol. 16, no. 5, pp. 3513-3539, May 2020.

[97] R. Velmani and B. Kaarthick, "An Efficient Cluster-Tree Based Data Collection Scheme for Large Mobile Wireless Sensor Networks," *IEEE Sensor Journal,* vol. 15, no. 4, pp. 2377-2390, Apr 2015.

[98] S. Chen, J. Zhou, X. Zheng, and X. Ruan, "Energy-Efficient Data Collection Scheme for Environmental Quality Management in Buildings," *IEEE Access,* vol. 6, pp. 57324-57333, 2018.

[99] H. Harb and A. Makhoul, "Energy-Efficient Sensor Data Collection Approach for Industrial Process Monitoring," *IEEE Transactions On Industrial Informatics,* vol. 14, no. 2, pp. 661-672, Feb 2018.

[100] G. A. Salaam, A. H. Abdullah, and M. H. Anisi, "Energy-Efficient Data Reporting for Navigation in Position-Free Hybrid Wireless Sensor Networks," *IEEE Sensors Journal,* vol. 17, no. 7, pp. 2289-2297, Apr 2017.

[101] H. Hu, K. Xiong, G. Qu, Q. Ni, P. Fan, and K. B. Letaief "AoI-Minimal Trajectory Planning and Data Collection in UAV-Assisted Wireless Powered IoT Networks," in *IEEE Internet of Things Journal,* vol. 8, No. 2, pp. 1211- 1223 Jan. 2021.

[102] J. Baek, S. I. Han, and Y. Han "Optimal UAV Route in Wireless Charging Sensor Networks," in *IEEE Internet of Things Journal,* vol. 7, No. 2, pp. 1327- 1335, Feb. 2020.

[103] Z. Wang, W. Xu, D. Yang, and J. Lin "Joint Trajectory Optimization and User Scheduling for Rotary-Wing UAV-Enabled Wireless Powered Communication Networks," in *IEEE Access,* vol. 7, pp. 181369- 181380, Dec. 2019.

[104] H. T. Yey, X. Kangy, J. Joungz, and Y. C. Liangy "Optimization for Full-Duplex Rotary-Wing UAV-Enabled Wireless-Powered IoT Networks," in *IEEE Trans. Wireless Commun.,* pp. 1- 1, Apr 2020.

[105] Y. Polyanskiy, H. V. Poor, and S. Verdú "Channel Coding Rate in the Finite Blocklength Regime," in *IEEE Trans. Inf. Theory.,* vol. 56, no. 5, pp. 2307-2359, May. 2010.

[106] E. Boshkovska, D. W. K. N. N. Zlatanov, and R. Schober "Practical Non-Linear Energy Harvesting Model and Resource Allocation for SWIPT Systems," in *IEEE Commun. Lett.,* vol. 19, no. 12, pp. 2082-2085, Dec. 2015.

[107] W. Cheng, X. Zhang, and H. Zhang, "Statistical-QoS driven energy-efficiency optimization over green 5G mobile wireless networks," *IEEE J. Sel. Areas Commun*, vol. 34, no. 12, pp. 3092-3107, Dec 2016.

[108] J. Baek, S. I. Han, and Y. Han "Energy-Efficient UAV Routing for Wireless Sensor Networks," in *IEEE Trans. Veh Technol.,* vol. 69, no. 2, pp. 1741- 1750, Feb 2020.

[109] T. Hester, M. Lopes and P. Stone "Learning Exploration Strategies in Model-Based Reinforcement Learning," in *The Twelfth International Conference on Autonomous Agents and Multiagent Systems.,* pp. 1069– 1076, May. 2013.

[110] R. J. Marks "Introduction to Shannon Samplingand Interpolation Theory," in *Springer-Verlag.,* New York, 1991.

[111] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal,* vol. 1, no. 1, pp. 22-32, Feb. 2014.

[112] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," in *IEEE Wireless Communications*, vol. 24, no. 1, pp. 120-128, Feb 2017.

[113] K. Liu and P. Lin, "Toward self-sustainable cooperative relays: state of the art and the future," in *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 56-62, June 2015.

[114] D. Liu, M. Zhao, W. Zhou, "Energy Efficiency Optimization in Energy Harvesting Incremental Relay System," in *10th International Conference on Wireless Communications and Signal Processing (WCSP).,* Dec 2018.

[115] J. Zhao, M. Zhao and W. Zhou, "Energy Efficiency Optimization in Energy Harvesting Cooperative Relay Systems," in *Sixth International Conference on Wireless Communications and Signal Processing.,* Dec 2014.

[116] M. Zhao, J. Zhao, W. Zhou, J. Zhu and S. Zhang, "Energy efficiency optimization in relay-assisted networks with energy harvesting relay constraints," in *China Communications.,* vol. 12, pp. 84-94, Feb 2015.

[117] M. Rezaee, M. Mirmohseni , V. Aggarwal and M. R. Aref, "Optimal Transmission Policies for Multi-Hop Energy Harvesting Systems," in *IEEE Transactions On Green Communications And Networking.,* vol. 2, No. 3, pp. 751-763, Sep 2018.

[118] Y. Luo, J. Zhang, and K. B. Letaief, "Transmit Power Minimization for Wireless Networks With Energy Harvesting Relays," in *IEEE Trans. Commun.,* vol. 64, No. 3, pp. 987-1000, Mar 2016.

[119] S. Guo, F. Wang, Y. Yang, and B. Xiao, "Energy-Efficient Cooperative Transmission for Simultaneous Wireless Information and Power Transfer in Clustered Wireless Sensor Networks," in *IEEE Trans. Commun.,* vol. 63, No. 11, pp. 4405- 4417, Nov 2015.

[120] C. Li ; P. Wu ; M. Xia, "Energy Efficiency Maximization of AF Relaying SWIPT Systems with Energy Recycling," in *2019 IEEE 89th Vehicular Technology Conference.,* Jun 2019.

[121] D. O. Akande, M. F. M. Salleh, "Energy-efficiency-based CMAC protocol with hybrid time-power splitting relaying for wireless ad-hoc networks," in *IET Communications.,* vol. 13, Iss. 17, pp. 2778- 2785, Sep 2019.

[122] J. Rostampoor, S. M. Razavizadeh, and I. Lee, "Energy efficient precoding design for swipt in MIMO two-way relay networks," in *IEEE Trans. Veh. Technol,* vol. 66, No. 9, pp. 7888-7896, Sep 2017.

[123] Q. Li and L. Yang, "Robust optimization for energy efficiency in MIMO two-way relay networks with swipt," in *IEEE systems journal,* vol. 14, No. 1, pp. 196-207, Mar 2020.

[124] A. Gupta, K. Singh and M. Sellathurai, "Time-Switching EH-Based Joint Relay Selection and Resource Allocation Algorithms for Multi-User Multi-Carrier AF Relay Networks," in *IEEE Transactions On Green Communications And Networking.,* vol. 3, No. 2, pp. 505- 522, Jun 2019.

[125] W. Chen, H. Ding, S. Wang, D. B. d. Costa and F. Gong, "Adaptive Discrete-Time-Switching-Based Energy-Harvesting Relaying Systems," in *IET Communications.,* vol. 68, No. 11, pp. 11064- 11079, Nov 2019.

[126] Y. Yan, S. Zhang, "Energy Efficiency Maximization of the SWIPT-based Full-Duplex Relay Cooperative System," in *2019 IEEE 19th International Conference on Communication Technology (ICCT).,* pp. 398- 403, Jan 2020.

[127] F. K. Ojo and M. F. M. Salleh, "Energy Efficiency Optimization for SWIPT-Enabled Cooperative Relay Networks in the Presence of Interfering Transmitter," in *IEEE Communications Lett.,* vol. 23, no. 10, pp. 1806- 1810, Oct 2019.

[128] H.-C. Yang, F. Xu, and M. S. Alouini, "Statistical energy efficiency characterization for wireless transmission over fading channels" in *IEEE Trans. Veh. Technol.,* 2020.

[129] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Communications Magazine,*vol. 49, no. 6, pp. 30-37, June 2011.

[130] T. P. Do, and T. V. T. Le, "Power allocation and performance comparison of full duplex dual hop relaying protocols," in *IEEE Communications Letters,* vol. 19, no. 5, pp. 791-794, May 2015.

[131] S. Pejoski, Z. H. Velkov, and R. Schober, "Optimal power and time allocation for WPCNs with piece-wise linear EH model," in *IEEE Communications Letters,* vol. 7, no. 3, pp. 364-367, Jun 2018.

[132] Y. Polyanskiy, H. V. Poor, and S. Verdú "Channel Coding Rate in the Finite Block-length Regime," in *IEEE trans. inf. theory.,* vol. 56, no. 5, pp. 2307-2359, May. 2010.

[133] J. Wu, Y. Zhang, M. Zukerman, and E. K. N. Yung, "Energy-Efficient Base-Stations Sleep-Mode Techniques in Green Cellular Networks: A Survey," in *IEEE Communication Surveys Tutorials.,* vol. 17, no. 2, pp. 803- 826, 2015.

[134] L. Zhang, H. C. Yang, and M. O. Hasna, "Generalized Area Spectral Efficiency: An Effective Performance Metric for Green Wireless Communications," in *IEEE Trans. Commun.,* vol. 62, no. 2, pp. 747- 757, Feb 2014.

[135] J. Zhu, L. Zhang, H. C. Yang, and M. O. Hasna "Generalized Area Spectral Efficiency of Wireless Ad-hoc Networks over Rayleigh Fading," in *Journal of communications and networks.,* vol. 22, no. 4, pp. 293- 302, Aug 2020.

[136] M. Sheng, L. Wang, X. Wang, Y. Zhang, C. Xu, and J. Li, "Energy Efficient Beamforming in MISO Heterogeneous Cellular Networks With Wireless Information and Power Transfer," in *Journal of communications and networks.,* vol. 34, no. 4, pp. 954-968, Apr 2016.

[137] A. E. Amine, H. A. H. Hassan, and L. Nuaymi, "Battery-Aware Green Cellular Networks Fed By Smart Grid and Renewable Energy," in *Journal of communications and networks.,* vol. 18, no. 2, pp. 2181- 2192, Jun 2021.

[138] J. Hu, W. Heng, X. Li, and J. Wu, "Energy-Efficient Resource Reuse Scheme for D2D Communications Underlaying Cellular Networks," in *IEEE Commun. Lett.,* vol. 21, no. 9, pp. 2097- 2100, Sep 2017.

[139] Yonghai Lin, Zhen Yang, Haiyan Guo, "Proportional Fairness-Based Energy-Efficient Power Allocation in Downlink MIMO-NOMA Systems with Statistical CSI," *China Communications.*, vol. 16, no. 12, pp. 47-55, Dec 2019.

[140] A. Zappone, P. Cao, and E. A. Jorswieck, "Energy Efficiency Optimization in Relay-Assisted MIMO Systems With Perfect and Statistical CSI," *IEEE Trans. Signal Process*, vol. 62, no. 2, Jan 2014.

[141] G. E. Hinton, S. Osindero, and Y. W. Teh "A fast learning algorithm for deep belief nets," in *Neural Computation.,* Vol. 18, pp. 1527-1554, 2006.

[142] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou "Playing Atari with Deep Reinforcement Learning," in *Proceedings of the 26th International Conference on Neural Information Processing Systems.,* pp. 1-9, Dec. 2013.

[143] S. S. Gupta, S. K. Pallapothu and N. B. Mehta "Ordered Transmissions for Energy-Efficient Detection in Energy Harvesting Wireless Sensor Networks," in *IEEE Trans. Commun.,* vol. 68, No. 4, pp. 2525- 2537, Apr 2020.

[144] E. Kartsakli, C. V. Verikoukis and L. Alonso, "Performance Analysis of the Distributed Queuing Collision Avoidance (DQCA) Protocol with Link Adaptation," in *IEEE Trans. Wireless Commun.,* vol. 8, No. 2, pp. 644- 647, Feb 2009.

[145] Y. Sim and D. H. Cho, "Performance Evaluation of Partially Clustered Access Scheme for Massive Machine Type Communications," in *IEEE Communications Letters.,* vol. 24, No. 3, pp. 626- 629, Mar 2020.

[146] S. Ray, "A Quick Review of Machine Learning Algorithms," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMIT-Con).,* pp. 35- 39, Feb 2019.