

ENERGY EFFICIENT RESOURCE ALLOCATION FOR FUTURE WIRELESS
COMMUNICATION SYSTEMS

by

MEHDI SALEHI HEYDAR ABAD

Submitted to the Graduate School of Engineering
and Natural Sciences in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Sabancı University

December 2019

Energy Efficient Resource Allocation for Future Wireless Communication Systems

by Mehdi Salehi Heydar Abad

APPROVED BY

Prof. Dr. Özgür Erçetin
(Thesis Advisor)

Prof. Dr. Özgür Gürbüz

Assoc. Prof. Dr. Deniz Gündüz

Prof. Dr. Alper Erdoğan

Assist. Prof. Dr. Sinan Yıldırım

DATE OF APPROVAL: December 6, 2019

To Deniz

©Mehdi Salehi Heydar Abad, 2019

All Rights Reserved

ENERGY EFFICIENT RESOURCE ALLOCATION FOR FUTURE WIRELESS COMMUNICATION SYSTEMS

Mehdi Salehi Heydar Abad

PhD Thesis, 2019

Thesis Advisor: Prof. Dr. Özgür Erçetin

Keywords: *Energy Harvesting, Resource Allocation, Machine Learning*

Next generation of wireless communication systems envisions a massive number of connected battery powered wireless devices. Replacing the battery of such devices is expensive, costly, or infeasible. To this end, energy harvesting (EH) is a promising technique to prolong the lifetime of such devices. Because of randomness in amount and availability of the harvested energy, existing communication techniques require revisions to address the issues specific to EH systems. In this thesis, we aim at revisiting fundamental wireless communication problems and addressing the future perspective on service based applications with the specific characteristics of the EH in mind.

In the first part of the thesis, we address three fundamental problems that exist in the wireless communication systems, namely; multiple access strategy, overcoming the wireless channel, and providing reliability. Since the wireless channel is a shared medium, concurrent transmissions of multiple devices cause interference which results in collision and eventual loss of the transmitted data. Multiple access protocols aim at providing a coordination mechanism between multiple transmissions so as to enable a collision free medium. We revisit the random access protocol for its distributed and low

energy characteristics while incorporating the statistical correlation of the EH processes across two transmitters. We design a simple threshold based policy which only allows transmission if the battery state is above a certain threshold. By optimizing the threshold values, we show that by carefully addressing the correlation information, the randomness can be turned into an opportunity in some cases providing optimal coordination between transmitters without any collisions.

Upon accessing the channel, a wireless transmitter is faced with a transmission medium that exhibits random and time varying properties. A transmitter can adapt its transmission strategy to the specific state of the channel for an efficient transmission of information. This requires a process known as channel sensing to acquire the channel state which is costly in terms of time and energy. The contribution of the channel sensing operation to the energy consumption in EH wireless transmitters is not negligible and requires proper optimization. We developed an intelligent channel sensing strategy for an EH transmitter communicating over a time-correlated wireless channel. Our results demonstrate that, despite the associated time and energy cost, sensing the channel intelligently to track the channel state improves the achievable long-term throughput significantly as compared to the performance of those protocols lacking this ability as well as the one that always senses the channel. Next, we study an EH receiver employing Hybrid Automatic Repeat reQuest (HARQ) to ensure reliable end-to-end communications. In inherently error-prone wireless communications systems, re-transmissions triggered by decoding errors have a major impact on the energy consumption of wireless devices. We take into account the energy consumption induced by HARQ to develop simple-to-implement optimal algorithms that minimizes the number of retransmissions required to successfully decode the packet.

The large number of connected edge devices envisioned in future wireless technologies enable a wide range of resources with significant sensing capabilities. The ability to collect various data from the sensors has enabled many exciting smart applications. Providing data at a certain quality greatly improves the performance of many of such applications. However, providing high quality is demanding for energy limited sensors. Thus, in the second part of the thesis, we optimize the sensing resolution of an EH wireless sensor in order to efficiently utilize the harvested energy to maximize an application dependent utility.

GELECEK NESİL KABLOSUZ HABERLEŐME SİSTEMLERİ İÇİN ENERJİ VERİMLİ KAYNAK TAHSİSİ

Mehdi Salehi Heydar Abad

Doktora Tezi, 2019

Tez Danıőmanı: Prof. Dr. ÖZGÜR ERÇETİN

Anahtar Kelimeler: *Enerji hasadı, kaynak tahsisi, makina öğrenme*

Gelecek nesil kablosuz iletişim sistemleri batarya ile çalışan çok sayıda kablosuz cihaz olmasını öngörmektedir. Bu tür cihazların bataryasını deęiőtirmek pahalı, maliyetli veya olanaksızdır. Enerji hasadı (EH) teknięi bu tür cihazların ömrünü uzatmak için umut veren bir yöntemdir. Hasat edilen enerjinin miktarı ve varolmasındaki rastgelelik, EH sistemlerine özgü sorunları ele almak suretiyle mevcut iletişim tekniklerinin guncellenmesini gerektirmektedir. Bu tezde EH'in kendine has özelliklerini ele alarak, temel kablosuz iletişim sorunlarını ve gelecekteki hizmete dayalı uygulamalarını incelemeyi hedefliyoruz.

Tezin birinci bölümünde, kablosuz iletişim sistemlerinde var olan üç temel sorunu ele alıyoruz; çoklu erişim strateji, kablosuz kanalın olumsuz etkilerini azaltmak ve güvenilirlik sağlamak. Kablosuz kanal paylaşılan bir ortam olduğundan, çoklu cihazların eşzamanlı yayınları girişime neden olur ve çarpışma sonucunda iletilen veri kaybolur. Çoklu erişim protokolleri çoklu veri aktarımı için kordinasyon sağlayarak çarpışmasız bir ortam sağlamayı hedefler. Daęıtık ve düşük enerji özelliklerine sahip olan rastgele erişim protokolü, EH süreçlerinin istatistiksel olarak ilişkili olan iki verici için tekrar gözden geçiriyoruz. Sadece

batarya seviyesi belli bir eşik üstüne çıktığında iletme izin veren basit bir eşik tabanlı erişim protkolu tasarlanmaktadır. Korelasyon bilgisini dikkatlice ele alarak eşik değerleri optimize edildiğinde rastgeleliğin fırsata dönebileceğini ve bazı durumlarda vericiler arasında en uygun koordinasyonu sağlayan çarpışmasız bir protokol olabileceği gösterilmektedir.

Kanala erişildiğinde, bir kablosuz verici rastgele ve zamanla değişen özellikler gösteren bir iletim ortam ile karşılaşmaktadır. Bir verici iletim stratejisini kanalın durumuna uyarlayarak daha etkin bir iletişim sağlayabilir. Kanal durumunu öğrenmek, kanal algılama sürecinin başlatılmasını gerektirir ki bu zaman ve enerji açısından maliyetlidir. Kanal durumu algılamanın enerjiye olan katkısını EH kablosuz cihazlarında olan etkisi ihmal edilemez, bu yüzden uygun optimizasyon gerekmektedir. Zamanda ilişkili bir kablosuz kanal üzerinden iletişim kuran bir EH vericisi için akıllı bir kanal algılama stratejisi geliştirilmektedir. Elde edilen sonuçlara göre zaman ve enerji maliyetine rağmen kanalın akıllı bir şekilde algılanması bu kabiliyetten yoksun ve ya her zaman kanalı algılayan protokollere göre önemli ölçüde veri aktarımını arttırmaktadır. Sonrasında, güvenilir uçtan uca iletişim sağlamak için hibrit yeniden gönderimli sistem (HYGS) kullanan bir EH alıcısı incelenmektedir. Doğası gereği hataya açık kablosuz iletişim sistemlerinde, hatalardan kaynaklanan yeniden gönderimin tetiklenmesi enerji tüketiminde büyük bir etkiye sahiptir. HYGS'ın neden olduğu enerji tüketimini ele alarak, gereken yeniden gönderim sayısını en aza düşürmek için basit bir şekile uygulanabilen algoritma geliştirilmektedir.

Gelecek nesil kablosuz teknolojilerde öngörülen çok sayıda bağlı cihazlar geniş bir algılama kabiliyetine sahip kaynak yelpazesi sağlar. Çeşitli verileri algıçlar tarafından toplayabilmek, pek çok heyecan verici akıllı uygulamayı mümkün kılmaktadır. Bu tür uygulamaların başarısı önemli bir ölçüde aktarılan verilerin kalitesine bağlıdır. Ancak, yüksek kaliteli veri sağlamak, enerjisi sınırlı olan algıçların yeteneklerinin sınırlarını aşmaktadır. Böylece tezin ikinci bölümünde, algıçın verisine bağlı olan uygulamanın başarısını enerji verimli bir şekilde en yüksek seviyeye çıkarabilmek için, algıçın algılama çözünürlüğü optimize edilmektedir.

Acknowledgments

There are many individuals who have played an important role in my success. First and foremost my sincere appreciation and thanks goes to my advisor Dr. Özgür Erçetin. I would not have been successful without his generous support and guidance. His patience, motivation, and immense knowledge had guided me to be an independent researcher today. I could not have imagined having a better advisor and mentor for my PhD study.

Additionally, I would also like to thank Dr. Deniz Gündüz whom I learned a lot through numerous collaborations and studies. I would also like to thank the rest of my thesis committee: Dr. Özgür Gürbüz, Dr. Sinan Yıldırım, and Dr. Alper Erdoğan, for their insightful comments and encouragement.

I am also thankful for the support of my family and friends. I would like to express my special gratitude towards my wife, Deniz, who had made this journey possible through significant amount of support and sacrifice. Finally, I am grateful for my good friend Emre who has contributed towards my academic accomplishment with his insights and numerous collaborations.

Contents

1	Introduction	2
1.1	Overview	2
1.2	Focus	4
1.2.1	Multiple Access strategy	4
1.2.2	Wireless Medium	6
1.2.3	Reliability	7
1.2.4	Service Based Optimization	8
1.3	Contributions	8
1.4	Publication Lists	10
1.4.1	Journal Papers	10
1.4.2	Conference Papers	11
2	Literature Review	12
3	Random Access Protocol with Correlated Energy Sources	17
3.1	Overview	17
3.2	System Model	19
3.3	Maximizing the Throughput	21
3.4	Special Cases	24
3.4.1	The Case of High Negative Correlation	24
3.4.2	The Case of High Positive Correlation	25
3.5	Numerical Results	29
3.6	Chapter Summary	30
4	Intelligent Channel Sensing Protocol over a Channel with Memory	33
4.1	Overview	34

4.1.1	Organization of the chapter	35
4.2	System Model	36
4.2.1	Channel and energy harvesting models	36
4.2.2	Transmission protocol	37
4.3	POMDP Formulation	38
4.4	The Structure of The Optimal Policy	43
4.4.1	General Case	43
4.4.2	Special Case: $R_1 = 0$	49
4.5	Numerical Results	52
4.5.1	Evaluating the optimal policy	52
4.5.2	Throughput performance	54
4.5.3	Optimal policy evaluation with two different transmission rates	59
4.6	Chapter Summary	60
4.7	Optimality of always transmitting in a GOOD state	61
5	Reliable Communication for a SWIPT enabled Receiver	65
5.1	Overview	66
5.1.1	Background and Motivation	66
5.1.2	Contributions	67
5.1.3	Related Work	68
5.2	System Model and Preliminaries	71
5.2.1	Channel Model and Receiver Architecture	71
5.2.2	Brief Overview of HARQ	72
5.2.3	Energy Harvesting and Consumption Model	73
5.3	The Minimum Expected Number of Re-transmissions For I.I.D. Channels	74
5.3.1	Markov Decision Process (MDP) Formulation	75
5.3.2	Absorbing Markov Chain Formulation	77
5.4	Optimal Class of Policies for i.i.d. Channels	86
5.5	Expected Number of Re-Transmissions for a Correlated Channel	91
5.6	Numerical Results	95
5.6.1	Simple ARQ	96
5.6.2	i.i.d. Channel States	96
5.6.3	Correlated Channel	98

5.7	Chapter Summary	99
6	Optimal Sensing Strategy for Wirelessly Powered Devices	104
6.1	Overview	105
6.1.1	Motivation	105
6.1.2	Contributions	108
6.1.3	Related Work	109
6.1.4	Outline	111
6.2	System Model	112
6.3	Problem Formulation	114
6.4	Finite Horizon Throughput Maximization	115
6.4.1	Optimal Offline Policy	116
6.4.2	Optimal Online Policy	120
6.5	Optimal Sensing	130
6.6	Numerical Results	133
6.6.1	Rate-Energy Trade-off	134
6.6.2	Performance Evaluation	134
6.6.3	MAB	136
6.7	Chapter Summary	139
7	Conclusions and Future Works	141

List of Figures

3.1	System Model	18
3.2	Associated DTMC with joint battery states	22
3.3	Transitions of joint battery states for high positive correlation case.	25
3.4	%AE and %RE versus γ_1 with $\gamma_2 = 9$ and $\delta' = 30$	29
3.5	Expected total throughput for high negative correlation with $\delta' = 5$	30
3.6	Expected total throughput for high positive correlation with $\delta' = 0.04$	31
3.7	Expected total throughput for high positive correlation with $\delta' = 30$	32
4.1	System model.	36
4.2	Optimal thresholds for taking the actions D (blue), O (green), H (yellow) for $B_{max} = 50$, $\mathcal{E}_T = 10$, $\tau = 0.2$, $\beta = 0.98$, $\lambda_1 = 0.9$, $\lambda_0 = 0.6$, $R = 3$ and $q = 0.1$	53
4.3	Value function associated with $B_{max} = 50$, $\mathcal{E}_T = 10$, $\tau = 0.2$, $\beta = 0.98$, $\lambda_1 = 0.9$, $\lambda_0 = 0.6$, $R = 3$ and $q = 0.1$	54
4.4	Optimal thresholds for taking the actions D (blue), O (green), H (yellow) for $B_{max} = 50$, $\mathcal{E}_T = 10$, $\tau = 0.1$, $\beta = 0.9$, $\lambda_1 = 0.8$, $\lambda_0 = 0.4$, and $R = 3$	55
4.5	Optimal thresholds for taking the actions D (blue), O (green), H (yellow) for $B_{max} = 50$, $\beta = 0.9$, $\mathcal{E}_T = 10$, $\lambda_1 = 0.8$, $\lambda_0 = 0.4$, $R = 3$ and $q = 0.8$	56
4.6	Throughputs by the optimal, greedy, single-threshold and opportunistic policies as a function of the EH rate, q	58
4.7	Optimal thresholds for taking the actions D (blue), L (red), OD (green), OT (black), H (yellow) for $B_{max} = 800$, $\mathcal{E}_T = 200$, $\tau = 0.035$, $\beta = 0.7$, $\lambda_1 = 0.98$, $\lambda_0 = 0.81$, $R_1 = 2.91$, $R_2 = 3$ and $q = 0.1$	60
5.1	A brief overview of the chapter.	78
5.2	State transition probabilities of the Markov chain associated with ρ	79

5.3	The effect of the encoding rate on the minimum expected number of re-transmissions for $R_1 = 10$, $e = 1$, $E_d = 5$ and $p = 0.1$	100
5.4	The effect of the channel quality and correlation on the minimum expected number of re-transmissions for $R_1 = 10$, $R_0 = 3$, $e = 1$, $E_d = 5$ and $p = 0.1$	101
5.5	The effect of the EH rate on the minimum expected number of re-transmissions for $R_1 = 10$, $R_0 = 5$, $E_d = 10$ and $p = 0.1$	102
6.1	System model.	112
6.2	The comparison of monomial and actual transmission rate and required signal-to-noise (SNR) ratio per symbol for $m = 3$ and $\lambda = 0.025$ as given in [1]. d represents the minimum distance between signal points.	114
6.3	An illustrative example of the battery evolution, $E(t)$, where $T = 10$. . .	120
6.4	The effect of channel discretization and deadline duration on the expected throughput.	135
6.5	Expected total throughput of the WPD with respect to the number of channel discretization levels in $T = 100$ time slots.	136
6.6	Expected total throughput of the WPD with $N = 20$ channel levels with respect to the frame length, T	137
6.7	Expected transmission rate of the WPD with $N = 20$ channel levels with respect to the frame length, T	138
6.8	Per-period regret comparison of TS and ϵ -greedy algorithms for $\epsilon = 0, 0.05, 0.1$	139

Chapter 1

Introduction

1.1 Overview

Wireless sensors has been utilized for decades to collect information from the environment. Prior use cases for the utilization of sensors were limited to simple applications such as environmental monitoring, animal tracking, monitoring catastrophic events such as volcano and etc. The recent scope envisions the utilization of huge number of wireless sensors in emerging services and applications such as the internet of things (IoT), entertainment, haptics, automation and many more. Regardless of the use case scenario of the sensors, an important issue related to employment of such sensors is the limited lifetime of their batteries. Consequently, many early research in this field proposed solutions for prolonging the lifetime of these devices. Some prominent proposed approaches include the energy aware MAC [2] protocols, duty cycle optimization [3], adaptive sensing [4] and etc. Although these solutions prolong the lifetime of sensors, it should be noted that eventually the lifetime remains finite. Note that many of such sensors are deployed in toxic, hostile or inaccessible environments where the replacement of batteries is often difficult, cost-prohibitive or impossible [5]. Another major difficulty is the sheer number of sensors in futuristic use cases which makes their tracking and hence battery replacement costly.

As a promising solution for the battery replacement problem, harvesting of energy from natural resources has become an important research area as a mean of achieving an ultimate perpetually available networks [6, 7]. Energy harvesting (EH) refers to scavenging energy from the environmental sources such as solar and wind, or other sources such as body heat and foot striking. The harvested energy is then converted to electric-

ity to be used by electrical devices. The various sources for EH include wind turbines, photovoltaic cells, thermoelectric generators and mechanical vibration devices such as piezoelectric devices, electromagnetic devices [8]. Unlike the sole dependence on the energy of a battery, EH provides periodical charging opportunities for EH devices which can extend their lifetimes significantly.

The fundamental challenge of EH paradigm is that the harvested energy is minuscule, comes at random amounts and times. This puts a heavy emphasis on communication schemes that specifically account for the randomness of the EH process. Based on the EH characteristic, the communication design can be categorized based on:

- energy storage structure: the sensor can either have a dedicated energy storage unit such as a battery and (super) capacitor, or without a storage unit such as passive RFID tags.
- energy arrival process: the energy arrival process can be either offline or online.

The energy storage unit enables storing energy to be used in the future. This correlates the resource management decisions over time making the resource management problem a dynamic one. Earlier research in design of energy management policies for EH systems aim at maximizing a given concave utility of consumed energy (e.g., transmitted bits, delay, etc) for the offline scenario in which the amount of harvested energy is known non-causally [9, 10]. Such a non-causal assumption on EH process enables an offline optimization framework that can be solved using well-known techniques such as Lagrangian optimization framework. On the contrary, when the EH arrival process is online, only causal information about the EH process is available and future realization of the EH process is unknown. The online EH process itself can be categorized in to two cases regarding the availability of EH statistics. The statistics governing the random processes can be available at the transmitter while their realizations are known only causally [11–13]. In this case, the EH communication system is usually modeled as a Markov decision process (MDP) [14], and dynamic programming (DP) [15] can be used to solve the MDP. There is also the possibility that in an online case even the statistics about the EH process is unknown. Such cases usually require tools such as machine learning to be able to learn the optimal resource allocation policy through interacting with the specific environment in which the energy is harvested [16].

Such challenges are induced by obvious uncontrollable nature of EH resources. The stochastic nature of the EH process dictates the amount and availability of the harvested energy that is beyond the control of system designers. To this end, radio frequency (RF) EH has been gaining popularity since it has the potential to provide network administrators a leverage for seamless charging opportunities. Radio signals with frequencies ranging from 300 GHz to 3 KHz can be used as a medium for transferring energies using electromagnetic propagation. Wireless power transfer (WPT) is a technology providing the network a way to replenish the batteries of the remote devices by utilizing RF transmissions. In wireless powered communication networks (WPCNs) [17–19], WPT occurs in the down-link (DL) to replenish the battery of WPDs which in turn is used for information transmission (IT) in the up-link (UL). Recently, the concurrent use of RF signals for both delivering energy and information has gained interest. In simultaneous wireless information and power transfer (SWIPT), the incoming RF signal is used for both energy harvesting and decoding of information bits. We emphasize that although WPT and SWIPT can be employed on demand to provide energy to the EH devices, the medium in which the RF signal is delivered has stochastic properties that may cause random variations in the received RF signal.

1.2 Focus

We divide the focus of this work into two parts. In the first part, we study three important challenging problems existing for communications systems; i) multiple access strategy for shared mediums, ii) channel state acquisition, and, iii) reliability. We develop policies that are tailored carefully for EH systems in addressing these problems by taking into account the challenges associated with EH systems. In the second part, we address the service based perspective of the future generations of the wireless communication technology. To this end, we take a general view of a service quality as an optimization metric and design a cross layer optimization framework specifically for this purpose.

1.2.1 Multiple Access strategy

At the physical layer, a wireless transmitter’s job is to convert the digital bits into electrical signals, modulate them into higher frequencies suitable for propagation and then feed

them into an antenna for propagation. The medium in which the electromagnetic wave is transmitted is shared with other electromagnetic waves giving an additive property to the electromagnetic waves. Hence, when multiple transmitters transmit a signal concurrently, their electromagnetic bearers interfere in the air and result in collision and eventual loss of data. Thus designing efficient multiple access policies for wireless channels is popular. The resources that a transmitter may use is time, frequency and space. The use of these resources can be orthogonalized to allow multiple transmitters to communicate with a common receiver. One way is to use a centralized entity to allow a given transmitter to communicate only on allowed resources, similar to a moderator in a debate. Such approaches require various signalling steps to enable orthogonalization which is costly in terms of energy.

Distributed algorithms do not rely on a centralized entity for mitigating collisions. As an example, carrier sense multiple access (CSMA) algorithm requires each transmitter to monitor the wireless channel for a certain time and allows them to transmit only when no other transmitter is transmitting. This technique relies on continuous sensing of the channel activity resulting in high energy consumption for low capability EH devices. Random channel access strategy is a frequently used technique preferred for its distributed and stateless implementation, which is particularly suitable for low power and low duty-cycle sensor networks. In random channel access no specific signaling is required to coordinate the transmissions and thus, enabling a low energy access scheme. However, this comes at the cost of occasional packet collisions. Specifically, transmitting with high probability increases the chances of collision events and accessing the channel with low probability decreases the resource utilization.

To this end, we adopt a random channel access strategy with the aim of introducing a form of coordination with the help of the statistics of the EH processes. More specifically, depending on the spatial distribution of EH devices, the amount of energy harvested by different devices is typically correlated. For example, consider EH devices harvesting energy from tidal motion [20]. The locations of two EH devices may be such that one is located at the tidal crest, while the other one is located in a tidal trough. In such a case, there may be a time delay equal to the speed of one wavelength between the generation of energy at each device. Such correlation information can be used to coordinate the transmissions of these devices without passing messages between them which is usually costly

in terms of energy consumption. We consider a network with two EH nodes transmitting data to a common base station over a random access channel. We develop and analyze a simple threshold-based transmission policy which grants access to an EH device only when its battery state exceeds a given threshold value. Threshold values are optimized based on the battery capacities and the correlation among EH processes of the devices to maximize the long-term throughput of the system.

1.2.2 Wireless Medium

The transmitted electromagnetic waves undergo various processes that attenuate signal power through absorption, reflection, scattering, and diffraction. When the attenuation is strong, the signal is blocked. Moreover, due to mobility, change of environment, or interference from other signals, the signal power may change randomly over time. Such variations in the wireless channel profile (amplitude and phase) is known as channel state information (CSI). A transmitter can adapt its transmission strategy to the specific state of the channel for an efficient transmission of information. One way to achieve this is to sense the channel¹ by consuming a fraction of available resources such as energy and time required for transmitting and receiving a pilot signal.

Often times, the wireless channel exhibits correlation in time in which the past history of the wireless channel can be used to predict the future channel state saving time and valuable energy for the EH transmitter. For an EH transmitter, we aim to utilize the memory of the channel to design intelligent channel sensing protocol so that harvested energy can be used efficiently by only expending energy when it is required. The correlation information can be mapped to a belief state which represents the conditional probability of the channel quality given its history. The EH transmitter, if believes that channel is in a good state can transmit without sensing the channel to save energy. Meanwhile if it believes that the channel quality is bad, it can opt to remain silent and save energy. The ultimate goal is to map the belief of the EH transmitter about the channel to the sensing, transmitting and deferring actions of an EH device to maximize the expected throughput over an infinite time horizon.

¹If the transmitter receives a known signal, known as pilot, it can calculate the channel state by investigating the received pilot's amplitude and phase.

1.2.3 Reliability

As discussed, wireless medium behaves randomly over time resulting in eventual loss of data due to unpredictable events such as random interfering RF signals. Thus, providing a mechanism for a reliable end-to-end transmission protocol is another important research topic for communication systems in recovering lost data. Automatic repeat and reQuest (ARQ) was the simplest form of reliable transmission protocols. The data stream is segmented into units of data known as packets and transmitted one by one. The receiver upon receiving the packets informs the transmitter whether a packet is corrupted and there is a need for retransmission. An obvious drawback of the ARQ protocol is that upon packet corruption, the whole corrupted packet is retransmitted which is inefficient. Hybrid ARQ (ARQ) protocols [21] provide a mechanism for forward error correcting (FEC) which is enabled by introducing redundancy to the packets. Specific HARQ protocols such as chase combining (CC) and incremental redundancy allow for the corrupted packets to be combined to potentially reducing the number of retransmissions. Nevertheless, this comes at the expense of extra processing time and energy associated with the enhanced error-correction decoders.

Recall that in simultaneous wireless information and power transfer (SWIPT), the incoming RF signal is used for both energy harvesting and decoding of information bits. More specifically, receiver architectures often adopted for a SWIPT receiver is the separated and co-located architectures. In separated architecture, both receivers have separate antennas, whereas in co-located architecture a single antenna is shared by both. In general, EH devices have small footprints necessitating a co-located architecture. This arises a resource allocation problem of sharing the RF signal among the two receivers. The incoming RF signal is fed to information decoding (ID) and energy harvesting (EH) circuitry by applying either time-switching (TS) or power splitting (PS) schemes. In TS, the RF signal is split over two different parts of the time slot, one for EH and the other for ID, whereas in PS the incoming RF signal is fed to both, proportional to a given factor. A receiver employing HARQ encounters two major energy consuming operations: (1) sampling or Analog-to-Digital Conversion (ADC), which includes all RF front-end processing, and (2) decoding. The energy consumption attributed to sampling, quantization and decoding plays a critical role in energy-constrained networks which makes their study a non-trivial problem.

We consider a point-to-point link where an energy-abundant transmitter employs HARQ to deliver a message reliably to an EH receiver. The receiver has no energy source, so it relies on harvesting energy from the information-bearing RF signal. The channel is time-varying where the amount of energy harvested and information collected varies depending on the quality of the channel. We minimize the expected number of re-transmissions needed to successfully deliver a message by optimally splitting the incoming RF signal between EH and ID receivers.

1.2.4 Service Based Optimization

The future generation of wireless communication technologies envision a service based approach where the wireless network should be tailored to realize the specific service based requirements. With the rapid development of hardware technologies for sensors, we are witnessing an increasing amount of data that can be collected to be used in various data driven machine learning applications. The performance of such applications and services greatly depends on the quality of the sensor generated data as measured by the resolution of the data points. On the other hand, generating high resolution data by wireless sensors induces a higher energy consumption and reduces the chance of successfully delivering the sensed data. We study a utility maximization problem in data driven applications for a wireless powered device (WPD) that is able to generate and transmit data at different resolution settings. We balance a trade-off between the utility gained by providing a high resolution data and the extra energy consumption associated with it.

1.3 Contributions

- In Chapter 3, we investigate the effects of the correlation between the EH processes at different EH devices in a wireless network. To this end, we consider a network with two EH nodes transmitting data to a common base station over a random access channel. We develop and analyze a simple threshold-based transmission policy which grants access to an EH node only when its battery state exceeds a given threshold value. Threshold values are selected based on the battery capacities and the correlation among EH processes of the nodes to maximize the long-term throughput of the system. We derive the average throughput of the network by

modeling the system as a discrete time Markov chain (DTMC) and obtaining its steady-state distribution. We then investigate two important special cases to obtain further insights into the selection of optimal transmission thresholds. In the first special case, only one node harvests energy at any time, while in the second case the nodes always harvest energy simultaneously. These two cases demonstrate completely different optimal threshold characteristics.

- In chapter 4, we utilize the information conveyed by a time correlated channel to design an intelligent channel sensing protocol to maximize the throughput of the EH transmitter. we take into account the energy cost of acquiring the CSI. We formulated the problem as a partially observable MDP (POMDP), which is then converted into an MDP with continuous state space by introducing a belief parameter for the channel state. We prove that the optimal transmission policy has a threshold structure with respect to the belief state, where the optimal threshold values depend on the battery state.
- In Chapter 5, we consider a class of wireless powered devices employing HARQ to ensure reliable end-to-end communications over a two-state time-varying channel. A receiver, with no power source, relies on the energy transferred by a SWIPT enabled transmitter to *receive* and *decode* information. We develop low complexity algorithms for the receiver to be able to decode the information with the minimum number of re-transmissions over independent and identically distributed (i.i.d.) as well as time correlated channel.
- We address the service based optimization perspective of next generation of wireless technologies in Chapter 6. To this end, we consider data driven applications in which the output quality depends on the resolution of the data generated by the sensors. We study a sensing resolution optimization problem for a WPD that is powered by wireless power transfer WPT from an access point (AP). We study a class of harvest-first-transmit-later type of WPT policy, where an AP first employs RF power to recharge the WPD in the down-link, and then, collects the data from the WPD in the up-link. The WPD optimizes the sensing resolution, WPT duration and dynamic power control in the up-link to maximize an application dependant utility at the AP. The utility of a transmitted packet is only achieved if the data

is delivered successfully within a finite time. Thus, we first study a finite horizon throughput maximization problem by jointly optimizing the WPT duration and power control. We prove that the optimal WPT duration obeys a time-dependent threshold form depending on the energy state of the WPD. In the subsequent data transmission stage, the optimal transmit power allocations for the WPD is shown to possess a channel-dependent fractional structure. Then, we optimize the sensing resolution of the WPD by using a Bayesian inference based multi armed bandit problem with fast convergence property to strike a balance between the quality of the sensed data and the probability of successfully delivering it.

1.4 Publication Lists

1.4.1 Journal Papers

- M. Salehi Heydar Abad and O. Ercetin, “Optimal Finite Horizon Sensing for Wirelessly Powered Devices,” in *IEEE Access*, vol. 7, pp. 131473-131487, 2019.
- Mehdi Salehi Heydar Abad, Ozgur Ercetin, Eylem Ekici, “Throughput optimal random medium access control for relay networks with time-varying channels,” *Computer Communications*, Volume 133, Pages 129-141, 2019.
- M. Salehi Heydar Abad, O. Ercetin and D. Gunduz, “Channel Sensing and Communication Over a Time-Correlated Channel With an Energy Harvesting Transmitter,” in *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 114-126, March 2018.
- M. Salehi Heydar Abad, O. Ercetin, T. ElBatt and M. Nafie, “SWIPT Using Hybrid ARQ Over Time Varying Channels,” in *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 4, pp. 1087-1100, Dec. 2018.
- M. S. Heydar Abad, E. Ozfatura, D. Gunduz and O. Ercetin, “Hierarchical Federated Learning Across Heterogeneous Cellular Networks”, arXiv, 2019.

1.4.2 Conference Papers

- M. S. Heydar Abad, E. Ozfatura, O. Ercetin and D. Gunduz, “Dynamic Content Updates in Heterogeneous Wireless Networks,” 2019 15th Annual Conference on Wireless On-demand Network Systems and Services (WONS), Wengen, Switzerland, 2019, pp. 107-110.
- M. S. H. Abad and O. Ercetin, “Finite Horizon Throughput Maximization for a Wirelessly Powered Device Over a Time Varying Channel,” 2018 IEEE Globecom Workshops (GC Wkshps), Abu Dhabi, United Arab Emirates, 2018, pp. 1-6.
- M. S. H. Abad, O. Ercetin, T. Elbatt and M. Nafie, “Wireless energy and information transfer in networks with hybrid ARQ,” 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, 2018, pp. 1-6.
- M. S. H. Abad, D. Gunduz and O. Ercetin, “Communication over a time correlated channel with an energy harvesting transmitter,” 2017 International Symposium on Wireless Communication Systems (ISWCS), Bologna, 2017, pp. 331-336.
- M. S. H. Abad, D. Gunduz and O. Ercetin, “Energy harvesting wireless networks with correlated energy sources,” 2016 IEEE Wireless Communications and Networking Conference, Doha, 2016, pp. 1-6.
- M. S. Heydar Abad, E. Ozfatura, D. Gunduz and O. Ercetin, “Hierarchical Federated Learning Across Heterogeneous Cellular Networks” ICASSP 2020, under review.

Chapter 2

Literature Review

Due to the tremendous increase in the number of battery-powered wireless communication devices over the past decade, harvesting of energy from natural resources has become an important research area as a mean of prolonging life time of such devices [6, 7]. The various sources for energy harvesting (EH) are wind turbines, photovoltaic cells, thermoelectric generators and mechanical vibration devices such as piezoelectric devices, electromagnetic devices [8]. EH technology is considered as a promising solution especially for large scale wireless sensor networks (WSNs), where the replacement of batteries is often difficult or cost-prohibitive [5]. However, due to the random nature of the harvested energy from ambient sources, the design of the system requires a careful analysis.

Early research in the design of optimal energy management policies for EH networks consider an offline optimization framework [9, 10, 22–24], in which non-causal information on the exact realization of the EH processes are assumed to be available. In the online optimization framework [11–13, 25], the statistics governing the random processes are assumed to be available at the transmitter, while their realizations are known only causally. In the learning optimization framework, knowledge about the system behavior is further relaxed and even the statistical knowledge about the random processes governing the system is not assumed, and the optimal policy scheduling is learned over time [16, 26].

The EH sensors communicate with a destination for reporting their data over wireless channels. Since the wireless channel is a shared medium, concurrent transmissions of sensors create interference. Thus, efficient multiple access protocols are needed to utilize the harvested energy which is usually of minuscule amount. In [22], for an offline

setting, the goal is to minimize the time in which all the data from both users are transmitted to the destination by optimizing the power allocations and departure rates. In [9], a transmitter with non-causal information schedules packets to be transmitted for two EH receivers and the objective is to minimize the transmission completion time which is the time both users have received their packets. [27] studies a resource allocation problem over a finite horizon to characterize the boundary of the maximum departure region for a multiple access channel in which the users can communicate with each other.

Concurrent transmissions of multiple devices over a shared wireless channel result in collision and eventual loss of data. Orthogonal schemes such as time division multiple access (TDMA) [17,28] and frequency division multiple access (FDMA) [29] allocate non overlapping resources to users to mitigate collision. Ensuring orthogonalization requires message passing to synchronize the transmission which comes at the cost of extra energy consumption for energy limited sensors. Random access protocols such as ALOHA [30] require no coordination at the cost of allowing occasional collisions among transmitters. In chapter 3, we aim at developing a random access policy for two energy harvesting sensors that transmit their data to a common destination. Different from literature, we take into account the possibility that the harvested energy by the sensors maybe correlated across them. We incorporate this information in designing a simple threshold based transmission policy that coordinate their transmissions for maximizing the sum throughput of the network. We show that the inherent randomness in the EH system can be turned into an opportunity by carefully addressing the correlation information in the random access policy.

Upon accessing the channel, the wireless sensor needs to overcome the challenges imposed by another source of randomness which is the state of the wireless channel that vary randomly over time. For an efficient utilization of energy, the transmission strategy should be properly adapted to the channel state. In [31], the authors develop an optimal transmission policy for maximizing the bit rate of a EH sensor by adapting the transmission parameters, allocated power and modulation type to the channel state. The optimality of a single-threshold policy is proven in [32] when an EH transmitter sends packets with varying importance. The allocation of energy for collecting and transmitting data in an EH communication system is studied in [33] and [34]. The scheduling of EH transmitters with time-correlated energy arrivals to optimize the long term sum throughput is inves-

tigated in [35]. Finite time horizon throughput optimization is addressed in [36], when either the current or future energy and channel states are known by the transmitter. In [37], power allocation to maximize the throughput is studied when the amount of harvested energy and channel states are modeled as Markov and static processes, respectively. In [38], an energy management scheme for sensor nodes with limited energy being replenished at a variable rate is developed to make the probability of complete depletion of the battery arbitrarily small, which at the same time asymptotically maximizes a utility function (e.g., Gaussian channel capacity) that depends on the energy consumption scheme. In [39] a simple online power allocation scheme is proposed for communication over a quasi-static fading channel with an i.i.d. energy arrival process, and it is shown to achieve the optimal long-term average throughput within a constant gap. However, [31–39] assume that the transmitter is aware of the wireless channel prior to transmissions. In practice, channel state is obtained through channel sensing which is realized by utilization of pilot signals. This procedure results in both energy and time overheads when the channel state is sensed at every transmission. We argue that the energy consumption for limited EH devices cannot be neglected and intelligent channel sensing algorithms is required to only sense the channel when it is needed. Thus, in Chapter 4, we show that when the channel state is correlated over time (e.g., when strong line of sight exists) it is possible to provide an intelligent frame work for the EH sensor to refrain from channel sensing and save its energy for future.

In inherently error-prone wireless communications systems, re-transmissions triggered by decoding errors have a major impact on the energy consumption of wireless devices. Hybrid automatic repeat request (HARQ) schemes are frequently used in order to reduce the number of re-transmissions by employing various channel coding techniques [21]. Nevertheless, this comes at the expense of extra processing time and energy associated with the enhanced error-correction decoders.

Note that EH devices harvest energy only in minuscule amounts (orders of μW s), so the energy consumption of the receiver circuitry to perform simple sampling and decoding can no longer be neglected. The authors in [40] addressed the energy consumption of sampling and decoding operations over a point-to-point link where the receiver harvests energy at a constant rate. In [41], a decision-theoretic approach is developed to optimally manage the transmit energy of an EH transmitter transmitting to an EH receiver, where

both the transmitter and the receiver harvests energy independently from a Bernoulli energy source. The receiver uses selective sampling (SS) and informs the transmitter about the SS information and its delayed battery state by feedback. Based on this feedback, the transmitter adjusts its transmission policy to minimize the packet error probability.

Meanwhile, in [42], the performance of different HARQ schemes for an EH receiver harvesting energy from a deterministic energy source with a constant energy rate was studied. In [43], the impact of the battery's internal resistance at the receiver was analyzed for an EH receiver with imperfect battery, with the aim of maximizing the amount of information decoded by the EH receiver. While ignoring the sampling energy cost at the receiver, [44] investigates the performance of TS policies to maximize the amount of information decoded at the receiver operating over a binary symmetric channel (BSC), by optimizing the fraction of time used for harvesting energy and for extracting information. For an EH transmitter and an EH receiver pair both harvesting ambient environmental energy with possible spatial correlation, [45] addresses the problem of outage minimization over a fading wireless channel with ACK-based re-transmission scheme by optimizing the power allocation at the transmitter. In [46], for a pair of EH transmitter-receiver employing ARQ and HARQ with binary EH process, packet drop probability over fading channels is minimized by optimally allocating power over different rounds of re-transmissions. In [47], an adaptive feedback mechanism for an EH receiver is proposed by taking into account the energy cost of sampling and decoding. The receiver is allowed to transmit a delayed feedback with the aim of efficiently utilizing the harvested energy in order to minimize the packet drop probability in the long run. In [48], the outage probability for an EH receiver powered by RF transmissions is minimized by implementing HARQ.

In Chapter 5, we consider a point-to-point link where an energy-abundant transmitter employs HARQ to deliver a message reliably to an EH receiver. The receiver has no energy source, so it relies on harvesting energy from the information-bearing RF signal. We develop optimal low complexity algorithms that can minimize the number of retransmissions required for successfully decoding information by an EH receiver, thus, addressing reliability issue in communication systems for EH receivers.

We revisited some the most fundamental problems of communication system in Chapter 3, 4 and 5 by specifically accounting for the characteristics of EH systems. Nowadays, the application scope of the sensors have evolved from simply reporting fixed size

packets about quantities such as temperature to advanced applications for hands-free interaction with the physical world, detection of unsafe behaviors, leveraging visual context for advertising, life logging and etc. For example, in [49], a camera sensor is trained to estimate the gaze location of person. Such an application heavily depends on the quality of the reported data by the sensors as measured by its resolution. In the example of [49], a high resolution image results in a lower gaze error while consuming more energy with the risk of not being able to deliver the message. Such a service based vision of sensors for future technologies motivated us to consider a general service based optimization in Chapter 6 where we optimize sensing resolution of a sensor to maximize the application dependant utility.

Chapter 3

Random Access Protocol with Correlated Energy Sources

This chapter considers a system with two energy harvesting (EH) nodes transmitting to a common destination over a random access channel. The amount of harvested energy is assumed to be random and independent over time, but correlated among the nodes possibly with respect to their relative position. A threshold-based transmission policy is developed for the maximization of the expected aggregate network throughput. Assuming that there is no a priori channel state or EH information available to the nodes, the aggregate network throughput is obtained. The optimal thresholds are determined for two practically important special cases: i) at any time only one of the sensors harvests energy due to, for example, physical separation of the nodes; ii) the nodes are spatially close, and at any time, either both nodes or none of them harvests energy.

3.1 Overview

Depending on the spatial distribution of EH devices, the amount of energy harvested by different devices is typically correlated. For example, consider EH devices harvesting energy from tidal motion [20]. The locations of two EH devices may be such that one is located at the tidal crest, while the other one is located in a tidal trough. In such a case, there may be a time delay equal to the speed of one wavelength between the generation of energy at each device.

In this chapter, we aim to investigate the effects of the correlation between the EH

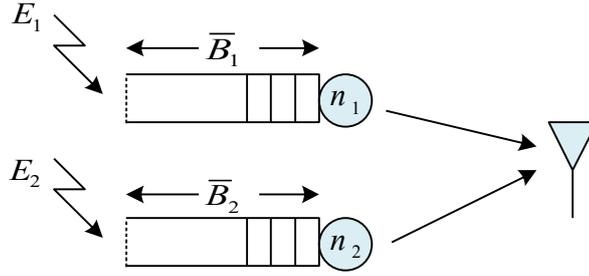


Figure 3.1: System Model

processes at different EH devices in a wireless network. To this end, we consider a network with two EH nodes transmitting data to a common base station over a random access channel as shown in Fig. 3.1. Random channel access is a frequently used technique preferred for its distributed and stateless implementation, which is particularly suitable for low power and low duty-cycle sensor networks. In random channel access, the nodes transmit probabilistically over time resulting in occasional packet collisions. However, packet collisions are especially harmful in EH networks due to scarce resources, and should be avoided as much as possible. In this chapter, we develop and analyze a simple threshold-based transmission policy which grants access to an EH node only when its battery state exceeds a given threshold value. Threshold values are selected based on the battery capacities and the correlation among EH processes of the nodes to maximize the long-term throughput of the system.

To illustrate the importance of choosing these threshold values intelligently, consider the following example. Let both EH nodes have a battery capacity of two energy units. Suppose that the EH nodes are spatially close, so they harvest energy simultaneously when energy is available. If the transmission thresholds are such that both nodes transmit a packet whenever they have one unit of energy, transmissions always result in a collision, and thus, the total network throughput is essentially zero. Meanwhile, if the thresholds are selected such that one EH node transmits a packet whenever it has one unit of energy, and the other node transmits a packet whenever it has two units of energy, there will be a collision once every two transmissions. Hence, with the latter choice of thresholds throughput increases to 0.5 packets.

We first derive the average throughput of the network by modeling the system as a discrete time Markov chain (DTMC) and obtaining its steady-state distribution. We then investigate two important special cases to obtain further insights into the selection of

optimal transmission thresholds. In the first special case, only one node harvests energy at any time, while in the second case the nodes always harvest energy simultaneously. These two cases demonstrate completely different optimal threshold characteristics.

We assume that EH nodes have no knowledge about the EH processes, and can only observe the amount of harvested energy in their own battery. Optimal threshold policies for an EH network is considered in [50] based on a game theoretic approach. In [51], authors optimize the throughput of a heterogeneous *ad hoc* EH network by formulating it as an optimal stopping problem. In [52] multiple energy harvesting sensor nodes are scheduled by an access point which does not know the energy harvesting process and battery states of the nodes. However, in these works the EH processes at different devices are assumed to be independent.

3.2 System Model

We adopt an interference model, where the simultaneous transmissions of two EH nodes result in a collision, and eventual loss of transmitted packets at the base station. Each node is capable of harvesting energy from an ambient resource (solar, wind, vibration, RF, etc.), and storing it in a finite capacity rechargeable battery. EH nodes have no additional power supplies. The nodes are data backlogged, and once they access the channel, they transmit until their battery is completely depleted. Note that assuming that the nodes are always backlogged allows us to obtain the saturated system throughput. In the following, we neglect the energy consumption due to generation of data to better illustrate the effects of correlated EH processes¹.

Time is slotted into intervals of unit length. In each time slot, the energy is harvested in units of δ joules. Let $E_n(t)$ be the energy harvested in time slot t by node $n = 1, 2$. We assume that $E_n(t)$ is an independent and identically distributed (i.i.d.) Bernoulli process with respect to time t . However, at a given time slot t , $E_1(t)$ and $E_2(t)$ may not be

¹For example, data may be generated by a sensor continuously monitoring the environment. Then, the energy consumption of a sensor may be included as a continuous drain in the energy process, but due to possible energy outages, the data queues may no longer be backlogged. We leave the analysis of this case as a future work.

independent. The EH rates are defined as follows:

$$\begin{aligned}
\Pr(E_1(t) = \delta, E_2(t) = \delta) &= p_{11}, \\
\Pr(E_1(t) = \delta, E_2(t) = 0) &= p_{10}, \\
\Pr(E_1(t) = 0, E_2(t) = \delta) &= p_{01}, \\
\Pr(E_1(t) = 0, E_2(t) = 0) &= p_{00},
\end{aligned} \tag{3.1}$$

where $p_{00} + p_{10} + p_{01} + p_{11} = 1^2$.

We assume that the transmission time ε is much shorter than the time needed to harvest a unit of energy, i.e., $\varepsilon \ll 1$, and the nodes cannot simultaneously transmit and harvest energy. Transmissions take place at the beginning of time slots, and the energy harvested during time slot t can be used for transmission in time slot $t + 1$. The channel is non-fading, and has unit gain. Given transmission power P , the transmission rate, $r_n(t)$, $n = 1, 2$ is given by the Shannon rate, i.e., $r_n(t) = \log(1 + P/N)$ (nats/sec/Hz), where N is the noise power.

We consider a deterministic transmission policy which only depends on the state of the battery of an EH node. Each EH node independently monitors its own battery level, and when it exceeds a pre-defined threshold, the node accesses the channel. If more than one node accesses the channel, a collision occurs and both packets are lost. Note that, by considering such an easy-to-implement and stateless policy, we aim to achieve low-computational power at EH devices.

The battery of each EH node has a finite capacity of \bar{B}_n , $n = 1, 2$. Let $B_n(t)$ be the state of the battery of EH node $n = 1, 2$ at time t . Node n transmits whenever its battery state reaches $\gamma_n \leq \bar{B}_n$ joules, $n = 1, 2$. When node n accesses the channel, it transmits at power $\frac{B_n(t)}{\varepsilon}$, i.e., the battery is completely depleted at every transmission. Hence, the time evolution of the battery states is governed by the following equation.

$$\begin{aligned}
B_n(t + 1) &= \min \{ \bar{B}_n, \\
&B_n(t) + E_n(t) \mathbb{1}_{\{B_n(t) < \gamma_i\}} - \mathbb{1}_{\{B_n(t) \geq \gamma_i\}} B_n(t) \},
\end{aligned} \tag{3.2}$$

²Note that if $p_{00} = p_{10} = p_{01} = p_{11} = 1/4$, then EH nodes generate energy independently from each other.

where $\mathbb{1}_{a < b} = \begin{cases} 1 & \text{if } a < b \\ 0 & \text{if } a \geq b \end{cases}$ is the indicator function.

Let $R_n(t)$ be the rate of *successful* transmissions, i.e.,

$$R_1(t) = \log \left(1 + \frac{B_1(t)/\varepsilon}{N} \right) \mathbb{1}_{\{B_1(t) \geq \gamma_1, B_2(t) < \gamma_2\}}, \quad (3.3)$$

$$R_2(t) = \log \left(1 + \frac{B_2(t)/\varepsilon}{N} \right) \mathbb{1}_{\{B_1(t) < \gamma_1, B_2(t) \geq \gamma_2\}}. \quad (3.4)$$

3.3 Maximizing the Throughput

We aim at maximizing the long-term average total throughput by choosing the transmission thresholds intelligently, taking into account the possible correlation between the EH processes. Let $\bar{R}_n(\gamma_1, \gamma_2)$ be the long-term average throughput of EH node n when the thresholds are selected as γ_1, γ_2 , i.e.,

$$\bar{R}_n(\gamma_1, \gamma_2) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T R_n(t), \quad n = 1, 2. \quad (3.5)$$

Then, the optimization problem of interest can be stated as

$$\max_{\gamma_1, \gamma_2} \sum_n \bar{R}_n(\gamma_1, \gamma_2), \quad (3.6)$$

$$\text{s.t. } 1 \leq \gamma_n \leq \bar{B}_n \quad n = 1, 2. \quad (3.7)$$

In order to solve the optimization problem (3.6)-(3.7), we first need to determine the long term average total throughput in terms of the thresholds. Note that for given γ_1, γ_2 , the battery states of EH nodes, i.e., $(B_1(t), B_2(t)) \in \{0, \dots, \gamma_1 - 1\} \times \{0, \dots, \gamma_2 - 1\}$ constitute a finite two dimensional discrete-time Markov chain (DTMC), depicted in Fig. 3.2. Let $\pi(i, j) = \Pr(B_1(t) = i, B_2(t) = j)$ be the steady-state distribution of the Markov chain for $i = 0, \dots, \gamma_1 - 1$ and $j = 0, \dots, \gamma_2 - 1$.

Theorem 3.1. *The steady state distribution of DTMC associated with the joint battery state of EH nodes is $\pi(i, j) = \frac{1}{\gamma_1 \gamma_2}$, for $i = 0, \dots, \gamma_1 - 1$ and $j = 0, \dots, \gamma_2 - 1$.*

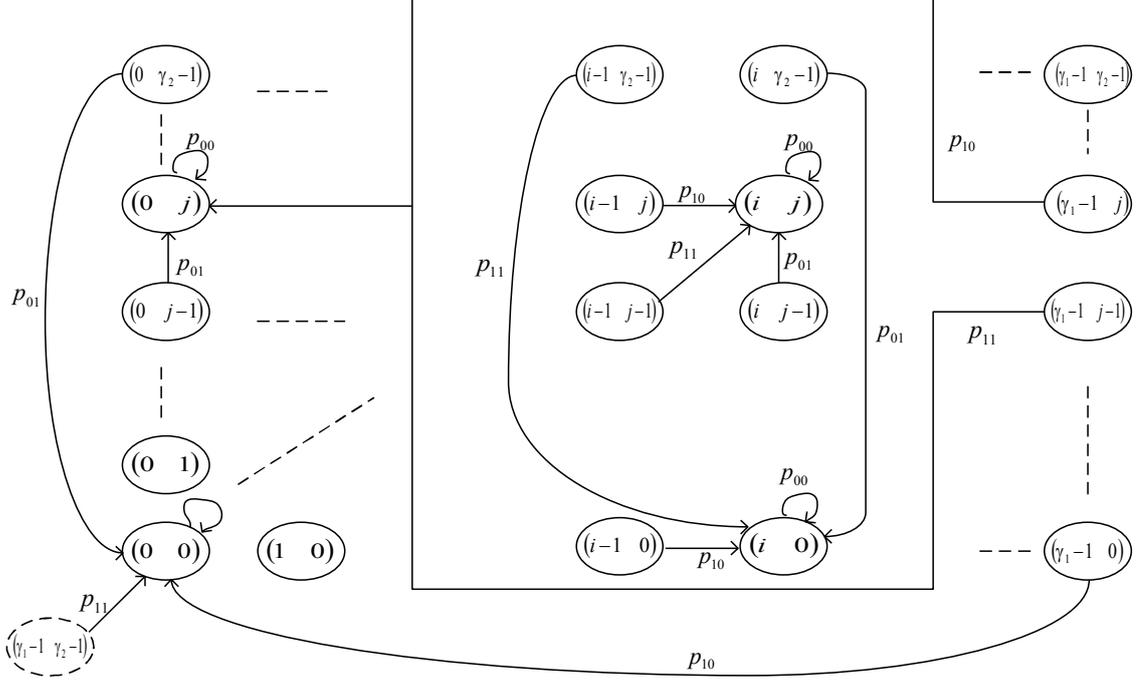


Figure 3.2: Associated DTMC with joint battery states

Proof. The detailed balance equations for $i = 1, \dots, \gamma_1 - 1$ and $j = 1, \dots, \gamma_2 - 1$ are:

$$\begin{aligned} \pi(i, j)(1 - p_{00}) &= \pi(i-1, j-1)p_{11} \\ &+ \pi(i-1, j)p_{10} + \pi(i, j-1)p_{01}. \end{aligned} \quad (3.8)$$

Whenever the battery state of node n reaches $\gamma_n - 1$, in the next state transition, given that it harvests energy, there is a transmission. Since the transmission time is much shorter than a time slot, i.e., $\varepsilon \ll 1$, after reaching state γ_n , node n immediately transmits and transitions back to state 0. Thus, the detailed balance equations for state 0 are given as:

$$\begin{aligned} \pi(i, 0)(1 - p_{00}) &= \pi(i-1, 0)p_{10} + \pi(i, \gamma_2 - 1)p_{01} \\ &+ \pi(i-1, \gamma_2 - 1)p_{11}, \quad 1 \leq i \leq \gamma_1 - 1, \end{aligned} \quad (3.9)$$

$$\begin{aligned} \pi(0, j)(1 - p_{00}) &= \pi(0, j-1)p_{01} + \pi(\gamma_1 - 1, j)p_{10} \\ &+ \pi(\gamma_1 - 1, j-1)p_{11}, \quad 1 \leq j \leq \gamma_2 - 1, \end{aligned} \quad (3.10)$$

$$\begin{aligned} \pi(0, 0)(1 - p_{00}) &= \pi(\gamma_1 - 1, \gamma_2 - 1)p_{11} \\ &+ \pi(\gamma_1 - 1, 0)p_{10} + \pi(0, \gamma_2 - 1)p_{01}. \end{aligned} \quad (3.11)$$

From (3.8), it is clear that if $p_{01}, p_{10} \neq 0$ then $\pi(i, j) \neq 0$ for all $i = 1, \dots, \gamma_1 - 1$ and $j = 1, \dots, \gamma_2 - 1$. Then, it can be verified that $\pi(i, j) = \pi(l, k)$ satisfies (3.8)-(3.11) for all i, j, k , and l . Hence, the theorem is proven since $\sum_{j=0}^{\gamma_2-1} \sum_{i=0}^{\gamma_1-1} \pi(i, j) = 1$. \square

Once the steady state distribution of DTMC is available, we can obtain the average throughput values. Let $\delta' = \frac{\delta/\varepsilon}{N}$.

Lemma 3.1. *The average throughput of EH nodes 1 and 2 for $p_{01}, p_{10} \neq 0$ are given as*

$$\begin{aligned} \bar{R}_1(\gamma_1, \gamma_2) &= \log(1 + \gamma_1 \delta') \\ &\times \left((p_{10} + p_{11}) \sum_{j=0}^{\gamma_2-2} \pi(\gamma_1 - 1, j) + p_{10} \pi(\gamma_1 - 1, \gamma_2 - 1) \right) \\ &= \frac{\log(1 + \gamma_1 \delta') [(\gamma_2 - 1)(p_{10} + p_{11}) + p_{10}]}{\gamma_1 \gamma_2}, \end{aligned} \quad (3.12)$$

$$\begin{aligned} \bar{R}_2(\gamma_1, \gamma_2) &= \log(1 + \gamma_2 \delta') \\ &\times \left((p_{01} + p_{11}) \sum_{i=0}^{\gamma_1-2} \pi(i, \gamma_2 - 1) + p_{01} \pi(\gamma_1 - 1, \gamma_2 - 1) \right) \\ &= \frac{\log(1 + \gamma_2 \delta') [(\gamma_1 - 1)(p_{01} + p_{11}) + p_{01}]}{\gamma_1 \gamma_2}. \end{aligned} \quad (3.13)$$

Proof. Consider node 1. Note that whenever the batteries are in one of the states $(\gamma_1 - 1, j)$ for $j = 0, \dots, \gamma_2 - 2$, a unit of energy (of δ joules) is harvested at node 1 with probability of $p_{10} + p_{11}$, and it transmits in the subsequent transition. Meanwhile, whenever the batteries are in state $(\gamma_1 - 1, \gamma_2 - 1)$, both nodes harvest a unit energy with probability p_{11} , and transmit in the subsequent transition resulting in a collision. Thus, in state $(\gamma_1 - 1, \gamma_2 - 1)$, EH node 1 successfully transmits with probability p_{10} . Similar arguments apply for node 2. \square

The following optimization problem is equivalent to (3.6)-(3.7).

$$\begin{aligned} \max_{\gamma_1, \gamma_2} z(\gamma_1, \gamma_2) &\triangleq \frac{\log(1 + \gamma_1 \delta') [(\gamma_2 - 1)(p_{10} + p_{11}) + p_{10}]}{\gamma_1 \gamma_2} \\ &+ \frac{\log(1 + \gamma_2 \delta') [(\gamma_1 - 1)(p_{01} + p_{11}) + p_{01}]}{\gamma_1 \gamma_2}, \end{aligned} \quad (3.14)$$

$$\text{s.t. } 1 \leq \gamma_n \leq \bar{B}_n, \quad n = 1, 2. \quad (3.15)$$

Note that (3.14)-(3.15) is an integer program. Since our main motivation is to investigate the effects of the correlated energy arrivals on the operation of EH networks,

rather than to obtain exact optimal thresholds, we may relax the optimization problem by omitting the integrality constraints. Nevertheless, the resulting relaxed optimization problem is still difficult to solve since the objective function is non-convex. Hence, in the following, we obtain the optimal solution for two important special cases.

3.4 Special Cases

Depending on the energy source and relative locations of the nodes, correlation among their EH processes may significantly vary. For example, if mechanical vibration is harvested, and the nodes are located far from each other, e.g., one EH device on one side of the road whereas the other one on the other side of a two-lane road, only the EH device on the side of the road where a car passes may generate energy from its vibration. This is a case of *high negative correlation*. Meanwhile, if solar cells are used as an energy source, EH processes at nearby nodes will have *high positive correlation*.

3.4.1 The Case of High Negative Correlation

We first analyze the case of high negative correlation. In particular, we have $p_{00} = p_{11} = 0$, $p_{10} = p$ and $p_{01} = 1 - p$ with $0 < p < 1$. Note that only one EH device generates energy at a given time. Let $z^{(-)}(\gamma_1, \gamma_2)$ be the total throughput of EH network when the thresholds are γ_1, γ_2 , obtained by inserting the values of $p_{00}, p_{11}, p_{10}, p_{01}$ in (3.14). We have

$$z^{(-)}(\gamma_1, \gamma_2) = \frac{\log(1 + \gamma_1 \delta') p}{\gamma_1} + \frac{\log(1 + \gamma_2 \delta') (1 - p)}{\gamma_2}. \quad (3.16)$$

The following lemma establishes that an EH device transmits whenever it harvests a single unit of energy. Interestingly, the optimal thresholds prevent any collisions between transmissions of EH devices, since at a particular time slot only one EH device has sufficient energy to transmit.

Lemma 3.2. *The optimal solution of (3.14)-(3.15) when $p_{00} = p_{11} = 0$, $p_{10} = p$ and $p_{01} = 1 - p$ with $0 < p < 1$, is $\gamma_1^* = 0$, $\gamma_2^* = 0$.*

Proof. Assume that γ_1 and γ_2 are non-negative continuous variables. Then, the gradient

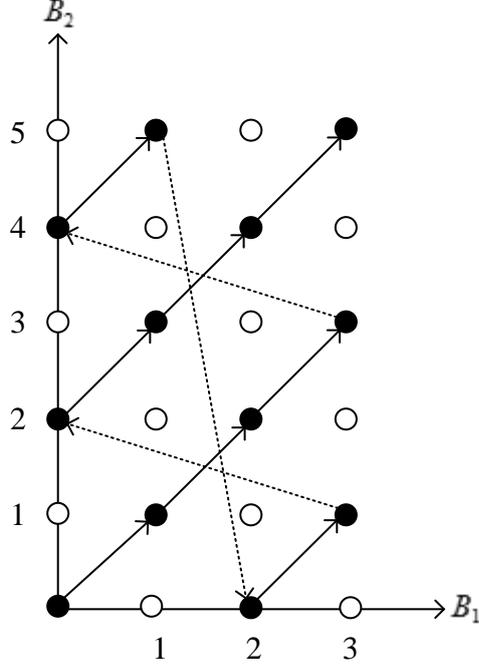


Figure 3.3: Transitions of joint battery states for high positive correlation case.

of $z^{(-)}(\gamma_1, \gamma_2)$ is:

$$\nabla z^{(-)}(\gamma_1, \gamma_2) = \left[\frac{p(\delta'\gamma_1 - (1 + \delta'\gamma_1) \log(1 + \gamma_1\delta'))}{\gamma_1^2(1 + \delta'\gamma_1)}, \frac{(1-p)(\delta'\gamma_2 - (1 + \delta'\gamma_2) \log(1 + \gamma_2\delta'))}{\gamma_2^2(1 + \delta'\gamma_2)} \right]. \quad (3.17)$$

Note that $\nabla z^{(-)}(\gamma_1, \gamma_2) < 0$ for all $\gamma_1 \geq 0$, $\gamma_2 \geq 0$ and p . Since $\nabla z^{(-)} < 0$, we have $z^{(-)}(\gamma_1, \gamma_2) > z^{(-)}(\hat{\gamma}_1, \hat{\gamma}_2)$ for every $\gamma_1 < \hat{\gamma}_1$ and $\gamma_2 < \hat{\gamma}_2$. Then, the lemma follows. \square

3.4.2 The Case of High Positive Correlation

Now, we consider the case of high positive correlation. In particular, we investigate the optimal solution when EH process parameters are $p_{01} = p_{10} = 0$, $p_{11} = p$ and $p_{00} = 1 - p$ with $0 < p < 1$; that is, either both EH devices generate energy or neither of them does. Note that in Theorem 3.1 the steady state distribution of DTMC is derived assuming that all of the states are visited. However, in the case of high positive correlation, only a part of the state space is visited.

In order to better illustrate this case, consider an EH network with thresholds $\gamma_1 = 4$ and $\gamma_2 = 6$. The state space of the corresponding DTMC is given in Fig. 3.3. Large solid and empty circles represent visited and unvisited battery states, respectively. The solid

lines represent the transitions of battery states when thresholds are not yet reached, and the dotted lines represent transitions when at least one of the nodes transmits. Also, arrows show the direction of transitions between the states. Since only a subset of the state space is visited infinitely often, the average throughputs given in Lemma 3.1 are no longer valid. We establish the average throughput of EH network with high positive correlation by the following lemma.

Lemma 3.3. *The average throughput $\bar{R}_n^{(+)}(\gamma_1, \gamma_2)$ of node $n = 1, 2$ for $p_{01} = p_{10} = 0$, $p_{11} = p$ and $p_{00} = 1 - p$ is given as*

$$\bar{R}_n^{(+)}(\gamma_1, \gamma_2) = p \cdot \frac{\left[\frac{LCM(\gamma_1, \gamma_2)}{\gamma_n} - 1 \right]}{LCM(\gamma_1, \gamma_2)} \cdot \log(1 + \gamma_n \delta'), \quad n = 1, 2 \quad (3.18)$$

where $LCM(\gamma_1, \gamma_2)$ is the least common multiple of γ_1 and γ_2 .

Proof. Due to our transmission policy, EH node n transmits whenever its battery level reaches γ_n , $n = 1, 2$. Note that both nodes reach their respective thresholds simultaneously every $LCM(\gamma_1, \gamma_2)$ instances of EH events. Since they transmit simultaneously, a collision occurs, and they both exhaust their batteries, i.e., the joint battery state transitions into state $(0, 0)$. The process repeats afterwards. Hence, the renewal period of this random process is $LCM(\gamma_1, \gamma_2)$. In every renewal period, EH node $n = 1, 2$ makes $\frac{LCM(\gamma_1, \gamma_2)}{\gamma_n} - 1$ number of successful transmissions. Hence, by using renewal reward theory, and noting that on the average a unit of energy is harvested in $p < 1$ proportion of time slots, we obtain (3.18). \square

Let $z^{(+)}(\gamma_1, \gamma_2) = \bar{R}_1^{(+)}(\gamma_1, \gamma_2) + \bar{R}_2^{(+)}(\gamma_1, \gamma_2)$ be the total throughput of a system with high positive correlation. Note that $z^{(+)}(\gamma_1, \gamma_2)$ is a non-convex function with respect to γ_1 , and γ_2 . Hence, in the following, we analyze the system in two limiting cases, i.e., when unit of energy harvested per slot, i.e., δ' , is either very small or very large.

Small Values of δ'

For small values of δ' , $\log(1 + \gamma_n \delta')$ can be approximated by $\gamma_n \delta'$. Let $GCD(\gamma_1, \gamma_2)$ be the greatest common divisor of γ_1 and γ_2 . By substituting $LCM(\gamma_1, \gamma_2) = \frac{\gamma_1 \gamma_2}{GCD(\gamma_1, \gamma_2)}$

we obtain

$$z^{(+)}(\gamma_1, \gamma_2) = 2\delta'p - GCD(\gamma_1, \gamma_2) \left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2} \right) \delta'p. \quad (3.19)$$

Note that maximizing (3.19) is equivalent to minimizing $GCD(\gamma_1, \gamma_2) \left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2} \right)$. Lemma 3.4 establishes that it is optimal to choose the thresholds as large as possible as long as the greatest common divisor of the two thresholds is equal to 1. This is due to the fact that the objective function in (3.19) is linear, and the optimum thresholds minimize the number of collisions.

Lemma 3.4. *The optimal thresholds for the case of high positive correlation for small values of δ' , and for $\bar{B}_2 > \bar{B}_1$ are $\gamma_1^* = \bar{B}_1$, $\gamma_2^* = \arg \max_j \bar{B}_2 - j$ for $j = 1, \dots, \bar{B}_2$, s.t., $GCD(\bar{B}_1, j) = 1$.*

Proof. Note that $0 < \frac{1}{\gamma_1} + \frac{1}{\gamma_2} \leq 2$, for $1 \leq \gamma_n \leq \bar{B}_n$, $n = 1, 2$. Let $\Gamma = \{(\gamma_1, \gamma_2) : GCD(\gamma_1, \gamma_2) = 1\}$. Note that if $(\gamma_1, \gamma_2) \notin \Gamma$, then $GCD(\gamma_1, \gamma_2) \geq 2$. Hence, it can be shown that $z^{(+)}(\gamma_1, \gamma_2) \geq z^{(+)}(\gamma'_1, \gamma'_2)$, for all $(\gamma_1, \gamma_2) \in \Gamma$, and $(\gamma'_1, \gamma'_2) \notin \Gamma$. Among $(\gamma_1, \gamma_2) \in \Gamma$, we choose the one that minimizes $\frac{1}{\gamma_1} + \frac{1}{\gamma_2}$, and thus, proving the lemma. \square

Large Values of δ'

For large values of δ' , $\log(1 + \gamma_n \delta')$ can be approximated by $\log(\gamma_n \delta')$. Also by substituting $LCM(\gamma_1, \gamma_2) = \frac{\gamma_1 \gamma_2}{GCD(\gamma_1, \gamma_2)}$ in $z^{(+)}(\gamma_1, \gamma_2)$ we have:

$$z^{(+)}(\gamma_1, \gamma_2) = \frac{(\gamma_2 - GCD(\gamma_1, \gamma_2)) \log(\gamma_1 \delta') p}{\gamma_1 \gamma_2} + \frac{(\gamma_1 - GCD(\gamma_1, \gamma_2)) \log(\gamma_2 \delta') p}{\gamma_1 \gamma_2}. \quad (3.20)$$

The optimal thresholds for this case is established in Lemma 3.5. Since the objective function in (3.20) has the property of *diminishing returns*, i.e., the rate of increase in the function decreases for higher values of its parameters, each device will choose transmitting more often, equivalently short messages, using less energy. However, transmissions are scheduled every time each node exceeds a threshold, which dictates small thresholds. When both EH devices transmit with small thresholds, there will be a large number of collisions, so the following lemma suggests that the aggregate throughput is maximized when one EH device transmits short messages, whereas the other transmits long messages.

Lemma 3.5. *The optimal thresholds for the case of high positive correlation for large values of δ' are $\gamma_1^* = B_1$, $\gamma_2^* = 1$ for $\bar{B}_1 > \bar{B}_2$, and they are $\gamma_1^* = 1$, $\gamma_2^* = B_2$ for $\bar{B}_2 > \bar{B}_1$.*

Proof. Let \hat{z} be an upper envelope function for $z^{(+)}$, obtained by substituting $GCD(\gamma_1, \gamma_2) = 1$ in (3.20):

$$\hat{z}(\gamma_1, \gamma_2) = \frac{(\gamma_2 - 1) \log(\gamma_1 \delta') p}{\gamma_1 \gamma_2} + \frac{(\gamma_1 - 1) \log(\gamma_2 \delta') p}{\gamma_1 \gamma_2}. \quad (3.21)$$

Note that since $GCD(\gamma_1, \gamma_2) \geq 1$, for every value of γ_1 and γ_2 , we have $\hat{z}(\gamma_1, \gamma_2) \geq z^{(+)}(\gamma_1, \gamma_2)$. First, we maximize \hat{z} for a given γ_2 by obtaining the corresponding optimal γ_1 . Taking the partial derivative of \hat{z} with respect to γ_1 , we obtain:

$$\frac{\partial \hat{z}}{\partial \gamma_1} = \frac{p}{\gamma_1^2 \gamma_2} [\log(\gamma_1 \delta) + \log(\gamma_2 \delta) - \gamma_2 (\log(\gamma_1 \delta) - 1) - 1]. \quad (3.22)$$

Note that $\gamma_2 \in \{1, \dots, \bar{B}_2\}$. If $\gamma_2 = 1$, (3.22) reduces to

$$\frac{\partial \hat{z}(\gamma_1, 1)}{\partial \gamma_1} = \frac{p}{\gamma_1^2 \gamma_2} \log \delta > 0. \quad (3.23)$$

Since $\frac{\partial \hat{z}(\gamma_1, 1)}{\partial \gamma_1} > 0$, the maximum value of \hat{z} is attained when $\gamma_1 = B_1$. For $\gamma_2 = 2$, (3.22) reduces to

$$\begin{aligned} \frac{\partial \hat{z}(\gamma_1, 2)}{\partial \gamma_1} &= \frac{p}{\gamma_1^2 \gamma_2} (-\log(\gamma_1 \delta) + \log(2\delta) + 1) \\ &= \begin{cases} < 0 & \text{if } \gamma_1 > 2e, \\ \geq 0 & \text{if } \gamma_1 \leq 2e, \end{cases} \end{aligned} \quad (3.24)$$

where e is the Euler's constant. Since $\frac{\partial^2 \hat{z}(2e, 2)}{\partial \gamma_1^2} = -\frac{1}{16e^3} < 0$, the maximum value of \hat{z} is attained when $\gamma_1 = 2e$. Finally, if $\gamma_2 \geq 3$, it can be shown that (3.22) is always negative as long as $\delta > 3e^2$. Hence, the maximum value of \hat{z} is attained for $\gamma_1 = 1$, if $\gamma_2 \geq 3$. By comparing the optimal values of \hat{z} for all $\gamma_2 \in \{1, \dots, \bar{B}_2\}$, one can show that \hat{z} is maximized for $(\gamma_1, \gamma_2) = (B_1, 1)$ when $B_1 > B_2$ and $(\gamma_1, \gamma_2) = (1, B_2)$ when $B_2 > B_1$. Since $GCD(1, B_2) = GCD(B_1, 1) = 1$, and $\hat{z} = z^{(+)}$ when $GCD(\gamma_1, \gamma_2) = 1$, it follows that optimal points for \hat{z} are also the optimal for $z^{(+)}$. \square

3.5 Numerical Results

We first verify (3.14) and (3.18) by Monte Carlo simulations. In the simulation, we model the battery states using equation (3.2). At each time slot t , we generate the joint EH process $(E_1(t), E_2(t))$ randomly. We run the simulation for 10^4 time slots and calculate the expected throughput by evaluating time average of the instantaneous rates as in (3.5).

Fig. 3.4 depicts the reliability of our analytical derivations. In particular, we measure both the percent relative error (%RE), which is defined as $\%RE = \frac{\text{Analytical value} - \text{Simulation value}}{\text{Analytical value}} \times 100$, and the absolute error (%AE), which is defined as $\%AE = (\text{Analytical value} - \text{Simulation value}) \times 100$, for $\gamma_2 = 9$ versus γ_1 . The results show a good match between the analytical and simulation results.

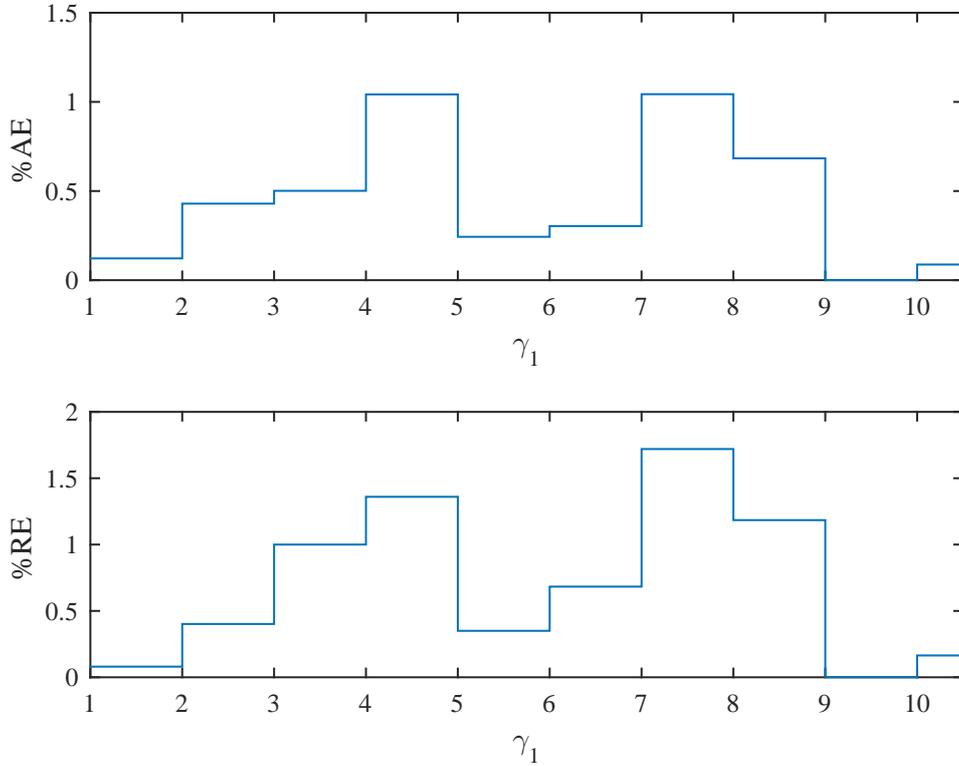


Figure 3.4: %AE and %RE versus γ_1 with $\gamma_2 = 9$ and $\delta' = 30$.

Next, we verify the optimal thresholds by numerically evaluating (3.14) and (3.18) for the cases of high negative and high positive correlation. We assume that $\bar{B}_1 = \bar{B}_1 = 10$ and $p = 0.5$. The aggregate throughput of the network with respect to the thresholds γ_1 and γ_2 for the case of high negative correlation is depicted in Fig. 3.5. It can be seen that the optimal thresholds are $\gamma_1^* = 1, \gamma_2^* = 1$, which is in accordance with Lemma 3.2.

Fig. 3.6 illustrates the aggregate throughput of the network for the case of high

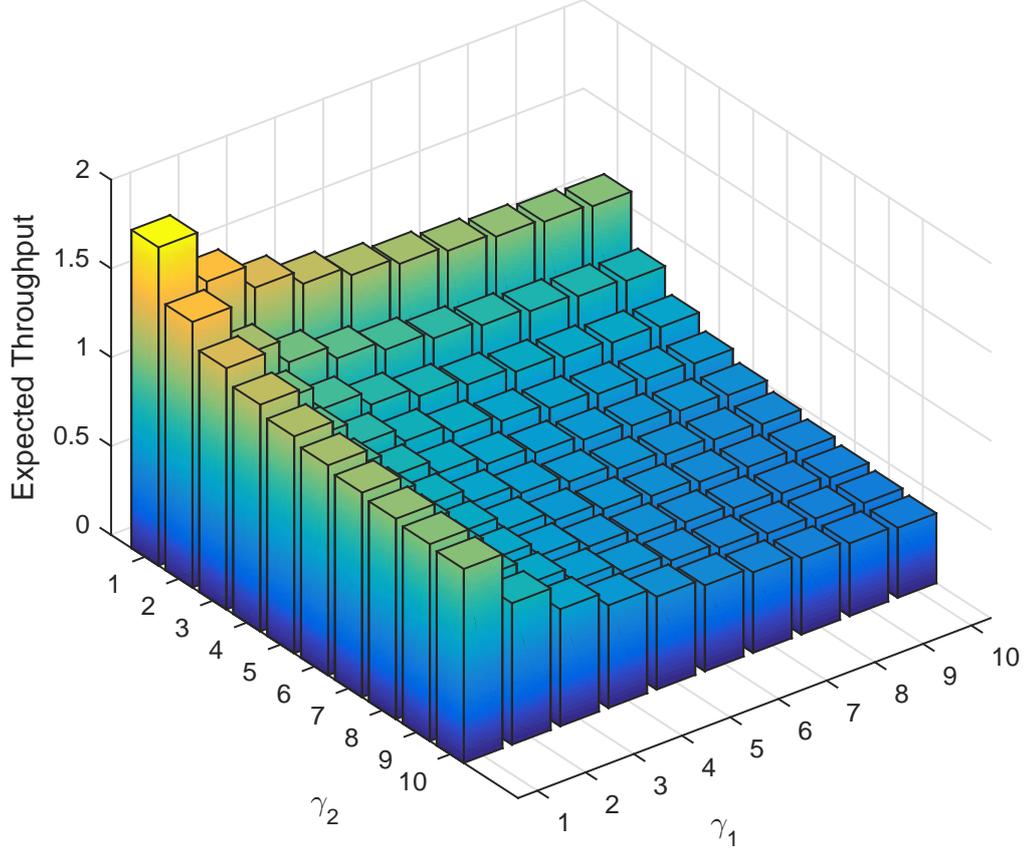


Figure 3.5: Expected total throughput for high negative correlation with $\delta' = 5$.

positive correlation with respect to γ_1 and γ_2 for $\delta' = 0.04$. The abrupt drops in the value of the aggregate throughput are due to the fact that $GCD(\gamma_1, \gamma_2)$ varies at least by a factor of two, which shows consistency with Lemma 3.4.

In Fig. 3.7, the aggregate throughput is depicted for the case of high positive correlation with respect to γ_1 and γ_2 for $\delta' = 30$. As expected from the results established in Lemma 3.5, the optimal thresholds are either $(\gamma_1^*, \gamma_2^*) = (1, 10)$ or $(\gamma_1^*, \gamma_2^*) = (10, 1)$.

3.6 Chapter Summary

We have investigated the effects of correlation among the EH processes of different EH nodes as encountered in many practical scenarios. We have developed a simple threshold based transmission policy to coordinate EH nodes' transmissions in such a way to maximize the long-term aggregate throughput of the network. In the threshold policy, nodes have no knowledge about each other, and at any given time they can only monitor their own battery levels. Considering various assumptions regarding the EH statistics and

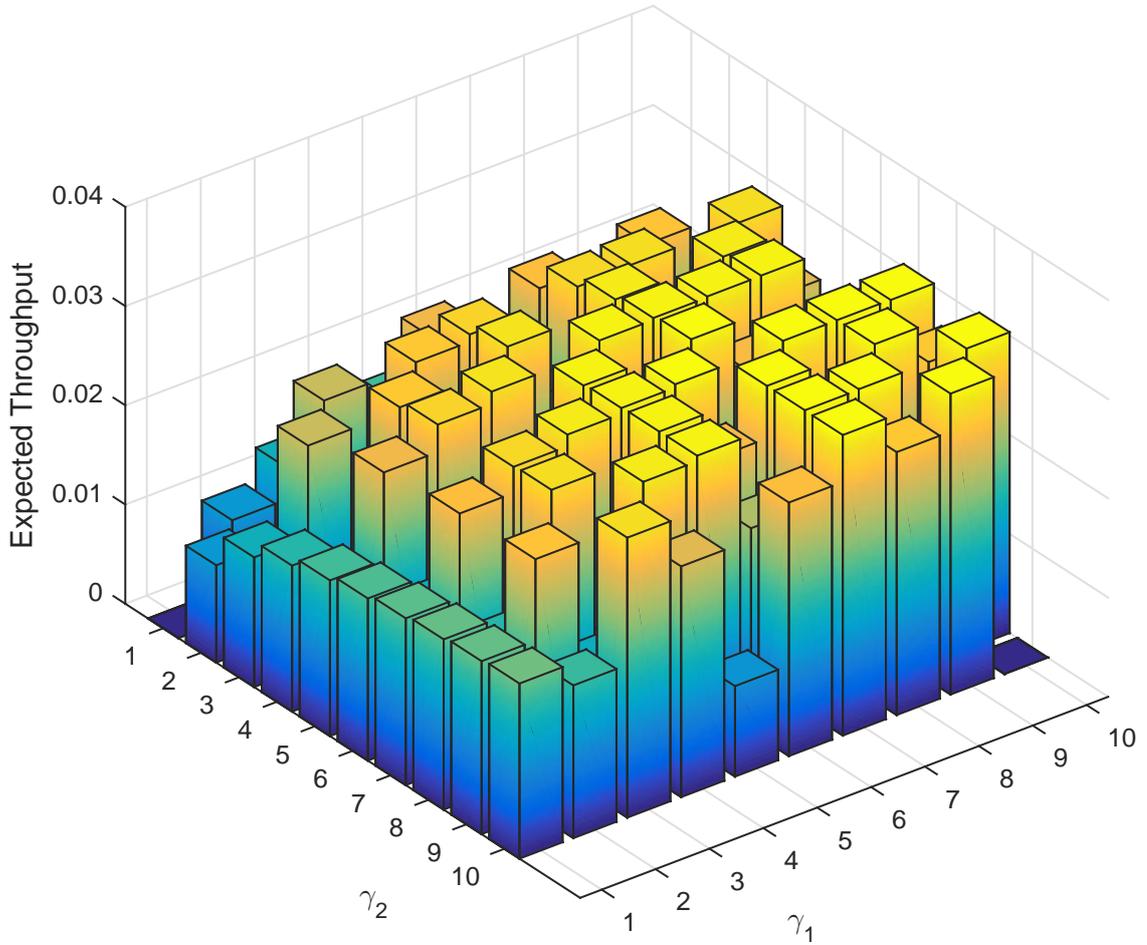


Figure 3.6: Expected total throughput for high positive correlation with $\delta' = 0.04$.

the amount of the harvested energy, the performance of the proposed threshold policy is studied. The established lemmas in Section 3.3 show that different assumptions about the underlying EH processes and the amount of the harvested energy demonstrate completely different optimal threshold characteristics. As our future work, we will investigate the cases when data queues are not infinitely backlogged and when the channels exhibit fading properties.

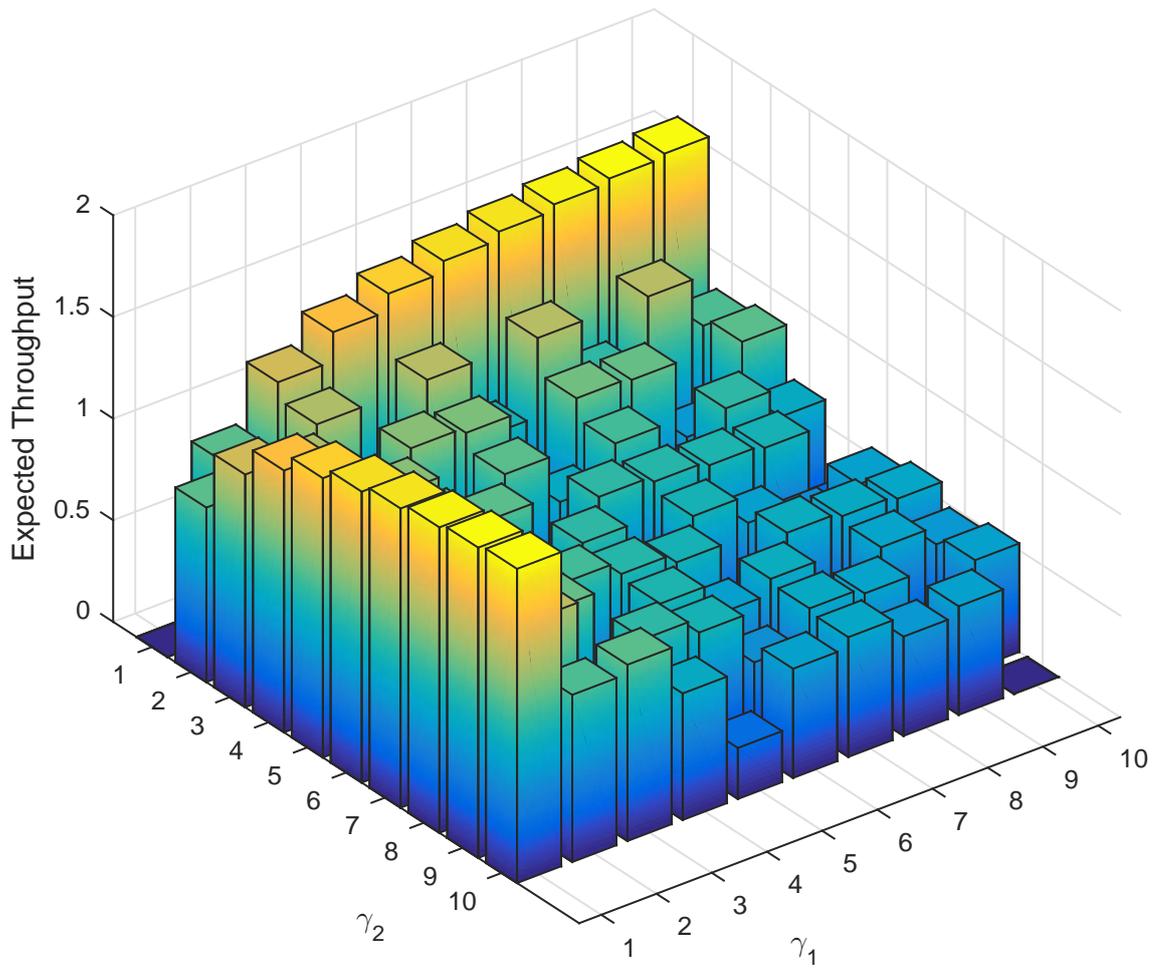


Figure 3.7: Expected total throughput for high positive correlation with $\delta' = 30$.

Chapter 4

Intelligent Channel Sensing Protocol over a Channel with Memory

We have addressed a MAC protocol design that utilizes correlation information of harvested energy across devices in Chapter 3. Note that upon successfully accessing the channel through a MAC protocol, a transmitter has to combat another source of randomness which is the channel conditions. To this end, we consider an energy harvesting (EH) transmitter communicating over a time-correlated wireless channel. The transmitter is capable of sensing the current channel state, albeit at the cost of both energy and transmission time. The EH transmitter aims to maximize its long-term throughput by choosing one of the following actions: *i*) defer its transmission to save energy for future use, *ii*) transmit reliably at a low rate, *iii*) transmit at a high rate, and *iv*) sense the channel to reveal the channel state at a cost of energy and transmission time, and then decide to defer or to transmit. The problem is formulated as a partially observable Markov decision process with a belief on the channel state. The optimal policy is shown to exhibit a threshold behavior on the belief state, with battery-dependent threshold values. The optimal threshold values and performance are characterized numerically via the value iteration algorithm as well as a policy search algorithm that exploits the threshold structure of the optimal policy. Our results demonstrate that, despite the associated time and energy cost, sensing the channel intelligently to track the channel state improves the achievable long-term throughput significantly as compared to the performance of those protocols lacking this ability as well as the one that always senses the channel.

4.1 Overview

Due to the tremendous increase in the number of battery-powered wireless communication devices over the past decade, replenishing the batteries of these devices by harvesting energy from natural resources has become an important research area [6]. Regardless of the type of energy harvesting (EH) device and the energy source employed, a main concern for such communication systems is the stochastic nature of the EH process [22, 53–55]. To model the uncertainty in the EH process, we consider a discrete-time system model in which random amount of energy is harvested by the transmitter at each time slot with independent and identically distributed (i.i.d.) values over time¹. We assume that the harvested energy is stored in a finite capacity rechargeable battery.

The communication takes place over a time-varying wireless channel with memory. The channel memory is modeled as a finite state Markov chain [56], such that the channel state in the next time slot depends only on the current state. A convenient and often-employed simplification is a two-state Markov chain, known as the Gilbert-Elliott channel [57]. This model assumes that the channel can be either in a *GOOD* or a *BAD* state. We assume that, by spending a certain amount of energy from its battery in a *GOOD* state, the transmitter can transmit R_2 bits of information within a time slot, while in a *BAD* state, it can only transmit R_1 bits, where $R_1 < R_2$.

In this chapter, differently from most of the literature on EH systems, we take into account the energy cost of acquiring channel state information (CSI). At the beginning of each time slot, without the current CSI, EH transmitter takes one of the following actions: *i*) defer the transmission to save its energy for future use; *ii*) transmit at a low rate of R_1 bits while guaranteeing successful delivery; *iii*) transmit at a high rate of R_2 bits and risk an unsuccessful transmission if the channel is in a *BAD* state, and *iv*) sense the channel state, with some time and energy cost, and then decide either to defer or transmit at a rate according to the revealed channel state. Our objective is to maximize the expected discounted sum of bits transmitted over an infinite time horizon.

Gilbert-Elliott channel model has been previously investigated in the context of scheduling an EH transmitter in [58], where the transmitter always has perfect CSI, ob-

¹Typically, the EH process is neither memoryless nor discrete, and the energy is accumulated continuously over time. However, in order to develop the analytical model underlying this chapter, we follow the common assumption in the literature [30, 53], and assume that the continuous energy arrival is accumulated in an intermediate energy storage device to form energy quantas.

tained by sensing at every time slot. The transmitter makes a decision to defer or to transmit based on the current CSI and the battery state. Similarly, without considering the channel sensing capability, [59] addresses the problem of optimal power management for an EH sensor over a multi-state wireless channel with memory. Unlike previous work, we take into account the energy cost of channel sensing which can be significant for a low-power EH transmitter. Therefore, in order to minimize the energy consumed for channel sensing, an EH transmitter does not necessarily sense the channel at every time slot, and instead, it keeps an updated belief of the channel state according to its past observations, and only occasionally senses the current channel state.

Channel sensing is an essential part of opportunistic and cognitive spectrum access. In [60], the authors investigate the problem of optimal access to a Gilbert-Elliot channel, wherein an energy-unlimited transmitter senses the channel at every time slot. In [61] channel sensing is done only occasionally. The transmitter can decide to transmit at a high or a low rate without sensing the channel, or it can first sense the channel and transmit at a reduced rate due to the time spent for sensing. However, the energy cost of sensing is ignored in [61]. Energy cost of channel sensing has been previously studied in [62] for a multiple-input single-output fading channel without memory when both the transmitter and the receiver harvest energy.

4.1.1 Organization of the chapter

In Section 4.2, we explain the channel and EH processes, and elaborate on the transmission protocol. In Section 4.3, we formulate the problem as a two-state partially observable MDP (POMDP) which is then converted to a continuous-state MDP by introducing a belief state. In Section 4.4, we show that the optimal policy is of threshold type and the optimal threshold values on the belief state depend on the state of the battery. In Section 4.5, we present the results of our Monte-Carlo simulations where we numerically obtain the optimal threshold values and the corresponding optimal performance. Finally, we conclude the chapter and present future research directions in Section 4.6.

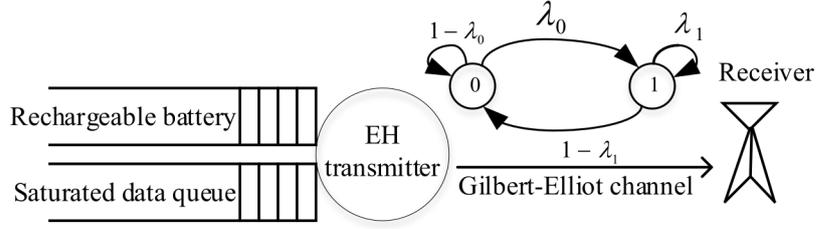


Figure 4.1: System model.

4.2 System Model

4.2.1 Channel and energy harvesting models

We consider the communication system illustrated in Fig. 4.1, where an EH transmitter communicates over a slotted Gilbert-Elliot channel. Let G_t denote the state of the channel at time slot t , which is modeled as a one-dimensional Markov chain with two states: GOOD state denoted by 1, and BAD state denoted by 0. Channel transitions occur at the beginning of each time slot. The transition probabilities are given by $\mathbb{P}[G_t = 1|G_{t-1} = 1] = \lambda_1$ and $\mathbb{P}[G_t = 1|G_{t-1} = 0] = \lambda_0$. We consider a simple constant-power transmitter which can employ error correcting codes at two different rates, each designed to achieve (almost) reliable transmission at one of the channel states. Accordingly, the transmitter is able to transmit R_2 bits per time slot if $G_t = 1$, and $R_1 < R_2$ bits if $G_t = 0$. We normalize the slot duration to one unit; and hence, R_1 and R_2 refer to both the transmission rate and the number of transmitted bits in a time slot. We assume that the transmitter has an infinitely backlogged data queue, and thus, it always has data to transmit.

We consider an energy *quanta*, representing the smallest energy unit, and assume that the energy arrivals and expenditures, both for transmission and channel sensing, are always integer multiples of this energy unit. At the end of time slot t , E_t units of energy arrive according to an i.i.d. random process², where $E_t \in \{0, 1, \dots, M - 1\}$ and $\mathbb{P}[E_t = m] = q_m$ for all t . The transmitter stores the energy packets in a battery with a capacity of B_{max} units of energy. We denote the state of the battery, i.e., the energy available in the battery at the beginning of time slot t , by B_t .

²There is an enormous body of the literature (see, for example, [58], [63], and references therein) which assumes i.i.d. EH processes. Nevertheless, results presented in this chapter can be extended to time-correlated EH processes by incorporating the EH process state into the state of the system. We restrict our attention to i.i.d. EH processes for the clarity of the exposition.

4.2.2 Transmission protocol

Once a transmission occurs, the receiver replies with an acknowledgment (ACK) if the transmission is successful, or with a negative acknowledgment (NACK) if the transmission fails. Note that, after a transmission at rate R_2 an ACK message informs the transmitter that the most recent state of the channel was GOOD, whereas a NACK message informs otherwise. No such information is acquired following a transmission at rate R_1 , which is successful independent of the channel state.

At the beginning of each time slot, the transmitter takes one of the following actions:

i) defer transmission, *ii)* transmit at rate R_1 , *iii)* transmit at rate R_2 , and *iv)* sense the channel and transmit or defer, based on the channel state.

i) Defer transmission (D): The transmitter remains idle, saving its energy to avoid future energy outages. If this action is chosen, there is no message exchange between the transmitter and the receiver. Hence, the transmitter does not obtain the current CSI³.

ii) Transmit at rate R_1 (L): The transmitter transmits at rate R_1 without sensing the channel. If this action is chosen, the transmitter uses a high redundancy coding scheme to guarantee the successful delivery of the message. Since the delivery of the information is guaranteed, the receiver always sends an ACK feedback, and thus, the transmitter does not obtain the current CSI.

iii) Transmit at rate R_2 (H): The transmitter transmits at rate R_2 without sensing the channel. If the channel is in a GOOD state, the transmission is successful and the receiver sends an ACK. Otherwise, the transmission fails, and the receiver sends a NACK. This feedback allows the transmitter to obtain the CSI for the completed time slot. We assume that the energy cost of both L and H actions is $\mathcal{E}_T \in \mathbb{Z}^+$ units of energy.

iv) Channel sensing/Defer at BAD state OD: The transmitter decides to sense the channel at the beginning of the time slot. Channel sensing operation is carried out by sending a control/probing packet, to which the receiver responds with a single bit indicating the channel state. We assume that sensing takes τ portion of a time slot, where $0 < \tau < 1$, and the transmitter consumes on average the same power as data transmission over the sensing period. Therefore, the energy cost of channel sensing is $\mathcal{E}_S = \tau\mathcal{E}_T$ units of energy, where $\mathcal{E}_S \in \mathbb{Z}^+$, and $\mathcal{E}_S < \mathcal{E}_T$. After sensing the channel, if the channel is

³The scenario in which the transmitter is informed about the current CSI even when it does not transmit any data packet is equivalent to the system model investigated in [58].

revealed to be in a GOOD state, in the remaining $1 - \tau$ portion of the time slot, the transmitter transmits at rate R_2 if it has more than $(1 - \tau)\mathcal{E}_T$ energy remaining in the battery. A total of $(1 - \tau)R_2$ bits can be transmitted by the end of the time slot. If the channel is revealed to be in a BAD state, then the transmitter defers transmission, saving the rest of the energy (i.e., $(1 - \tau)\mathcal{E}_T$).

v) Channel sensing/Transmit at BAD state OT: The transmitter again senses the channel initially, and transmits at rate R_2 if the channel is in a GOOD state. However, if the channel is in a BAD state, it transmits at rate R_1 in the remainder of the time slot.

Remark 4.1. *Note that, in both actions involving channel sensing (OD and OT) the transmitter transmits at rate R_2 if the channel is revealed to be in a GOOD state. This follows from the fact that transmitting at rate R_2 when the channel is known to be in a GOOD state has the highest reward for the amount of energy used. A more rigorous proof of this argument is provided in Section 4.7.*

Thanks to the channel sensing capability, the transmitter can adapt its behavior to the current channel state. As we show in this chapter, this proves to be an important capability to improve the efficiency in EH networks with scarce energy sources.

4.3 POMDP Formulation

At the beginning of each time slot, the transmitter chooses an action from the action set $\mathcal{A} \triangleq \{D, L, OD, OT, H\}$, based on the state of its battery and its belief about the channel state to maximize a long-term discounted reward to be defined shortly. Although the transmitter is perfectly aware of its battery state, it does not know the current channel state. Hence, the problem can be formulated as a POMDP.

Let the state of the system at time t be denoted by $S_t = (B_t, X_t)$, where X_t denotes the *belief* of the transmitter at time slot t about the channel state. The belief X_t , is the conditional probability that the channel is in a GOOD state at the beginning of the current slot, given the history \mathcal{H}_t , i.e., $X_t = \mathbb{P}[G_t = 1 | \mathcal{H}_t]$, where \mathcal{H}_t represents all past actions and observations of the transmitter up to, but not including, slot t . The belief of the transmitter constitutes a sufficient statistic to characterize its optimal actions [64]. Note that with this definition of a state, the POMDP problem is converted into an MDP with an uncountable state space $\{0, 1, 2, \dots, B_{max}\} \times [0, 1]$.

A transmission policy π describes a set of rules that dictate which action to take at each slot depending on the history. Let $V^\pi(b, p)$ be the expected infinite-horizon discounted reward with initial state $S_0 = (b, \mathbb{P}[G_0 = 1 | \mathcal{H}_0] = p)$ under policy π with discount factor $\beta \in [0, 1)$. The use of the expected discounted reward allows us to obtain a tractable solution, and one can gain insights into the optimal policy for the average reward when β is close to 1. β can be interpreted as the probability that the transmitter is allowed to use the channel, or the probability of the transmitter to remain active at each time slot as in [65]. For an initial belief p , the expected discounted reward has the following expression

$$V^\pi(b, p) = \mathbb{E} \left[\sum_{t=0}^{\infty} \beta^t R(S_t, A_t) | S_0 = (b, p) \right], \quad (4.1)$$

where t is the time index, $A_t \in \mathcal{A}$ is the action chosen at time t , and $R(S_t, A_t)$ is the expected reward acquired when action A_t is taken at state S_t . The expectation in (4.1) is over the state sequence distribution induced by the given transmission policy π . The expected reward when action A_t is chosen at state S_t is given as follows:

$$R(S_t, A_t) = \begin{cases} X_t R_2, & A_t = H, B_t \geq \mathcal{E}_T, \\ R_1, & A_t = L, B_t \geq \mathcal{E}_T, \\ (1 - \tau) X_t R_2, & A_t = OD, B_t \geq \mathcal{E}_T, \\ (1 - \tau)[(1 - X_t) R_1 \\ \quad + X_t R_2], & A_t = OT, B_t \geq \mathcal{E}_T, \\ 0, & \text{otherwise.} \end{cases} \quad (4.2)$$

Since \mathcal{E}_T energy units is required for transmission (with or without channel sensing), if the battery state is below \mathcal{E}_T , the reward becomes zero. Hence, in (4.2) we only consider actions taken when the battery state is at least \mathcal{E}_T . If the action of transmitting at rate R_2 without sensing is chosen, R_2 bits are transmitted successfully if the channel is in a GOOD state, and 0 bits otherwise. Since the belief, X_t , represents the probability of the channel being in a GOOD state, the expected reward is given by $X_t R_2$. It is guaranteed that transmitting at low rate is always successful, so the expected reward for this action is R_1 . If the action of channel sensing is chosen, $\mathcal{E}_S = \tau \mathcal{E}_T$ energy units is spent sensing

the channel with the remaining $(1 - \tau)\mathcal{E}_T$ energy units either being used for transmission, or saved in the battery. If the channel is in a GOOD state, $(1 - \tau)R_2$ bits are transmitted successfully. If the channel is in a BAD state, the transmitter either remains silent and receives no rewards, or utilizes $(1 - \tau)\mathcal{E}_T$ energy units and transmits $(1 - \tau)R_1$ bits in the rest of the time slot. Thus, the expected reward of action OD is $(1 - \tau)X_t R_2$, while the expected reward of OT is $(1 - \tau)[(1 - X_t)R_1 + X_t R_2]$. Finally, if the action of deferring (D) is taken, the transmitter neither senses the channel nor transmits data, so the reward is zero.

Define the value function $V(b, p)$ as

$$V(b, p) = \max_{\pi} V^{\pi}(b, p), \forall b \in \{0, 1, \dots, B_{max}\}, \forall p \in [0, 1]. \quad (4.3)$$

The optimal infinite-horizon expected reward can be achieved by a stationary policy, i.e., there exists a stationary policy π^* , mapping the state space $\{0, 1, \dots, B_{max}\} \times [0, 1]$ into the action space \mathcal{A} , such that $V(b, p) = V^{\pi^*}(b, p)$ [66]. The value function $V(b, p)$ satisfies the Bellman equation

$$V(b, p) = \max_{A \in \{D, L, OD, OT, H\}} \{V_A(b, p)\}, \quad (4.4)$$

where $V_A(b, p)$ is the action-value function, defined as the expected infinite-horizon discounted reward acquired by taking action A in state (b, p) , and is given by

$$\begin{aligned} V_A(b, p) = & R((b, p), A) \\ & + \beta \mathbb{E}_{(\acute{b}, \acute{p})} \left[V(\acute{b}, \acute{p}) | S_0 = (b, p), A_0 = A \right], \end{aligned} \quad (4.5)$$

where (\acute{b}, \acute{p}) denotes the next state when action A is taken at state $S_0 = (b, p)$. The expectation in (4.5) is over the distribution of next states. Below, we evaluate the action-value function $V_A(b, p)$, and how the system state evolves for each action.

Defer transmission (D): Since there is no transmission, there is no feedback; and thus, the transmitter does not learn the the channel state. Therefore, the belief is updated as the probability of finding the channel in a GOOD state given the current belief state. If

$X_t = p$ at time slot t , after taking action D , belief is updated as

$$J(p) = \lambda_0(1 - p) + \lambda_1 p. \quad (4.6)$$

After taking action D , the value function evolves as:

$$V_D(b, p) = \beta \sum_{m=0}^{M-1} q_m V(\min(b + m, B_{max}), J(p)). \quad (4.7)$$

Note that the term $\min(b + m, B_{max})$ is used to ensure that the battery state does not exceed the battery capacity, B_{max} .

Transmit at rate R_1 (L): This action can be taken only if⁴ $b \geq \mathcal{E}_T$. The transmission will be successful independent of the channel state. Hence, the ACK feedback from the receiver does not inform the transmitter about the channel state. Similarly to action D , the belief state is updated using (4.6), and the value function is given by:

$$V_L(b, p) = R_1 + \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), J(p)). \quad (4.8)$$

Transmit at rate R_2 (H): This action can only be chosen if $b \geq \mathcal{E}_T$. If the channel is in GOOD state, R_2 bits are successfully delivered to the receiver, the receiver sends back an ACK, and the belief for the next time slot is updated as λ_1 . Otherwise, the transmission fails, the receiver sends a NACK, and the belief is updated as λ_0 . Hence, the value function evolves as:

$$V_H(b, p) = p \left[R_2 + \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_1) \right] + (1 - p) \left[\beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_0) \right]. \quad (4.9)$$

Channel sensing/ Defer in BAD state (OD): If $b \geq \mathcal{E}_T$ and the transmitter decides to sense the channel, it consumes $\mathcal{E}_S = \tau \mathcal{E}_T$ units of energy to sense the current channel

⁴Note that in the generic MDP formulation, we have the same set of actions in every state. We can re-define the reward function by assigning $-\infty$ reward to those actions that are not possible to be taken in specific states to account for this. For the ease of exposition, we chose to present the formulation in this manner.

state. If the channel is found to be in a GOOD state, $(1 - \tau)\mathcal{E}_T$ units of energy is used to transmit $(1 - \tau)R_2$ bits, and the belief state is updated as λ_1 . Note that the transmitter always transmits if the channel is in a GOOD state, because this is the best state possible and saving energy for future cannot improve the reward. We refer the interested readers to Section 4.7 for a rigorous proof of this claim. In action *OD*, transmission is deferred if the channel is in a BAD state, and the transmitter saves $(1 - \tau)\mathcal{E}_T$ units of energy for possible future transmissions. The belief is updated as λ_0 for the next time slot. The action-value function for action *OD* is given by:

$$\begin{aligned}
V_{OD}(b, p) &= p \left[(1 - \tau)R_2 + \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_1) \right] \\
&+ (1 - p) \left[\beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \tau\mathcal{E}_T, B_{max}), \lambda_0) \right]. \tag{4.10}
\end{aligned}$$

Meanwhile, if $\tau\mathcal{E}_T \leq b < \mathcal{E}_T$, transmission is not possible since it requires at least \mathcal{E}_T units of energy. However, it is still possible to sense the channel, since it only requires $\tau\mathcal{E}_T$ units of energy. This case may arise when the transmitter believes that learning the channel state may help its decision in the future. Thus, for $\tau\mathcal{E}_T \leq b < \mathcal{E}_T$, the action-value function for action *OD* is given by:

$$\begin{aligned}
V_{OD}(b, p) &= p\beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \tau\mathcal{E}_T, B_{max}), \lambda_1) \\
&+ (1 - p)\beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \tau\mathcal{E}_T, B_{max}), \lambda_0). \tag{4.11}
\end{aligned}$$

Channel sensing/Transmit at BAD state (OT): The transmitter senses the channel, and transmits no matter what the channel state is. It transmits $(1 - \tau)R_2$ bits if it is in a GOOD state, and $(1 - \tau)R_1$ bits in a BAD state. The belief is updated as λ_1 (λ_0) if the channel is in a GOOD (BAD) state. The action-value function is given by:

$$\begin{aligned}
& V_{OT}(b, p) \\
&= p \left[(1 - \tau)R_2 + \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_1) \right] \\
&+ (1 - p) \left[(1 - \tau)R_1 + \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_0) \right]. \tag{4.12}
\end{aligned}$$

Based on the action-value functions presented above, the evolution of the battery state is as follows:

$$B_{t+1} = \begin{cases} \min(B_t + E_t, B_{max}), & A_t = D, \\ \min(B_t + E_t - \mathcal{E}_T, B_{max}), & A_t \in \{L, H, OT\}, B_t \geq \mathcal{E}_T, \\ \min(B_t + E_t - \tau\mathcal{E}_T \\ \quad - (1 - \tau)\mathcal{E}_T G_t, B_{max}), & A_t = OD, B_t \geq \mathcal{E}_T \\ \min(B_t + E_t - \tau\mathcal{E}_T, B_{max}), & A_t = OD, \tau\mathcal{E}_T \leq b < \mathcal{E}_T. \end{cases} \tag{4.13}$$

4.4 The Structure of The Optimal Policy

4.4.1 General Case

In this section, we show that the optimal policy has a threshold-type structure on the belief state. The belief state set, i.e., the interval $[0, 1]$, can be divided into mutually exclusive subsets where each subset is assigned to a distinct action. We begin to establish our main results by proving the convexity of the value function $V(b, p)$, with respect to p .

Lemma 4.1. *For any given $b \geq 0$, $V(b, p)$ is convex in p .*

Proof. Define $V(b, p, n)$ as the optimal value function for the finite-horizon problem spanning only n time slots. We will first prove the convexity of $V(b, p, n)$ in p by induction. Optimal value function can be written as follows,

$$\begin{aligned}
V(b, p, n) = \max \{ & V_D(b, p, n), V_L(b, p, n), V_{OD}(b, p, n), \\ & V_{OT}(b, p, n), V_H(b, p, n) \}, \tag{4.14}
\end{aligned}$$

where

$$V_D(b, p, n) = \beta \sum_{m=0}^{M-1} q_m V(\min(b + m, B_{max}), J(p), n - 1), \quad (4.15)$$

$$\begin{aligned} V_L(b, p, n) &= R_1 \\ &+ \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), J(p), n - 1), \end{aligned} \quad (4.16)$$

$$\begin{aligned} V_{OD}(b, p, n) &= p \left[(1 - \tau) R_2 \right. \\ &+ \left. \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_1, n - 1) \right] \\ &+ (1 - p) \left[\beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \tau \mathcal{E}_T, B_{max}), \lambda_0, n - 1) \right], \\ &\text{for } b \geq \mathcal{E}_T, \end{aligned} \quad (4.17)$$

$$\begin{aligned} V_{OT}(b, p, n) &= p \left[(1 - \tau) R_2 \right. \\ &+ \left. \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_1, n - 1) \right] \\ &+ (1 - p) \left[(1 - \tau) R_1 \right. \\ &+ \left. \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_0, n - 1) \right], \text{ for } b \geq \mathcal{E}_T, \end{aligned} \quad (4.18)$$

$$\begin{aligned} V_{OD}(b, p, n) &= p \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \tau \mathcal{E}_T, B_{max}), \lambda_1, n - 1) \\ &+ (1 - p) \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \tau \mathcal{E}_T, B_{max}), \lambda_0, n - 1), \\ &\text{for } \tau \mathcal{E}_T \leq b < \mathcal{E}_T, \end{aligned} \quad (4.19)$$

$$\begin{aligned}
V_H(b, p, n) &= p \left[R_2 \right. \\
&\quad \left. + \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_1, n - 1) \right] \\
&\quad + (1 - p) \left[\beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_0, n - 1) \right], \\
&\qquad\qquad\qquad \text{for } b \geq \mathcal{E}_T. \tag{4.20}
\end{aligned}$$

Note that when $b < \mathcal{E}_T$, we have $V(b, p, 1) = 0$, and when $b \geq \mathcal{E}_T$ we have $V(b, p, 1) = \max \{R_1, pR_2, (1 - \tau)pR_2, (1 - \tau)[pR_2 + (1 - p)R_1]\}$ which is a maximum of four convex functions. We see that $V(b, p, 1)$ is a convex function of p .

Now, let us assume that $V(b, p, n - 1)$ is convex in p for any $b \geq 0$, then for $a \in [0, 1]$ we can investigate the convexity of the value function for each action separately as follows.

For deferring the transmission, i.e., $A = D$, we can write:

$$\begin{aligned}
V_D(b, ap_1 + (1 - a)p_2, n) &= \beta \sum_{m=0}^{M-1} q_m V(\min(b + m, B_{max}), J(ap_1 + (1 - a)p_2), n - 1) \\
&= \beta \sum_{m=0}^{M-1} q_m V(\min(b + m, B_{max}), aJ(p_1) + (1 - a)J(p_2), n - 1) \\
&\leq a\beta \sum_{m=0}^{M-1} q_m V(\min(b + m, B_{max}), J(p_1), n - 1) \\
&\quad + (1 - a)\beta \sum_{m=0}^{M-1} q_m V(\min(b + m, B_{max}), J(p_2), n - 1) \\
&= aV_D(b, p_1, n) + (1 - a)V_D(b, p_2, n) \tag{4.21}
\end{aligned}$$

Hence, $V_D(b, p, n)$ is convex in p . Similarly, consider action L :

$$\begin{aligned}
V_L(b, ap_1 + (1-a)p_2, n) &= R_1 \\
&+ \beta \sum_{m=0}^{M-1} q_m V(\min(b+m-\mathcal{E}_T, B_{max}), J(ap_1 + (1-a)p_2), n-1) \\
&= R_1 + \beta \sum_{m=0}^{M-1} q_m V(\min(b+m-\mathcal{E}_T, B_{max}) \\
&\quad, aJ(p_1) + (1-a)J(p_2), n-1) \\
&\leq aR_1 + a\beta \sum_{m=0}^{M-1} q_m V(\min(b+m-\mathcal{E}_T, B_{max}), J(p_1), n-1) \\
&+ (1-a)R_1 \\
&+ (1-a)\beta \sum_{m=0}^{M-1} q_m V(\min(b+m-\mathcal{E}_T, B_{max}), J(p_2), n-1) \\
&= aV_L(b, p_1, n) + (1-a)V_L(b, p_2, n). \tag{4.22}
\end{aligned}$$

Thus, $V_L(b, p, n)$ is also convex in p . Note that $V_{OD}(b, p, n)$, $V_{OT}(b, p, n)$, and $V_H(b, p, n)$ are linear functions of p , thus they are also convex in p . Since the value function $V(b, p, n)$ is the maximum of five (or, in some cases two) convex functions when $b \geq \mathcal{E}_T$ ($\tau\mathcal{E}_T \leq b < \mathcal{E}_T$), it is also convex. By induction we can claim the convexity of $V(b, p, n)$ for all n . Since $V(b, p, n) \rightarrow V(b, p)$ as $n \rightarrow \infty$, $V(b, p)$ is also convex. \square

Next, we show that the value function is a non-decreasing function of the battery state, b . This lemma provides the intuition why deferring or sensing actions are advantageous in some states. The incentive of taking these actions is that the value function transitions into higher values without consuming any energy, or consuming only $\tau\mathcal{E}_T$ energy units.

Lemma 4.2. *Given an arbitrary belief $0 \leq p \leq 1$, $V(b_1, p) \geq V(b_0, p)$ if $b_1 > b_0$.*

Proof. We will again use induction to prove the claim for $V(b, p, n)$ defined as in the proof of Lemma 4.1 as the optimal value function when the decision horizon spans n stages. We have $V(b, p, 1) = 0$ if $b < \mathcal{E}_T$ and we have

$V(b, p, 1) = \max\{R_1, pR_2, (1-\tau)pR_2, (1-\tau)[pR_2 + (1-p)R_1]\}$ if $b \geq \mathcal{E}_T$. Hence, $V(b, p, 1)$ is trivially non-decreasing in b . Suppose that $V(b, p, n-1)$ is non-decreasing

in b . Each of the value functions given in (4.15), (4.16), (4.17), (4.18), (4.19) and (4.20) is the summation of positive weighted non-decreasing functions. Therefore, they are all non-decreasing in b . Since the optimal value function is the maximum of these non-decreasing functions, it is also non-decreasing in b for any n . By letting $n \rightarrow \infty$, we conclude that $V(b, p)$ is non-decreasing in b . \square

The next lemma states that the value function is non-decreasing with respect to the belief state as well.

Lemma 4.3. *For a given battery state $b \in \{0, 1, \dots, B_{max}\}$, if $p_1 > p_0$ then $V(b, p_1) \geq V(b, p_0)$.*

Proof. We employ induction on $V(b, p, n)$ once again. For $n = 1$, $V(b, p, 1)$ is 0 if $b < \mathcal{E}_T$, and $\max\{R_1, pR_2, (1 - \tau)pR_2, (1 - \tau)[pR_2 + (1 - p)R_1]\}$ if $b \geq \mathcal{E}_T$. Therefore, $V(b, p, 1)$ is non-decreasing in p for any b .

Assume that $V(b, p, n - 1)$ is non-decreasing in p . Since $J(p)$ is non-decreasing, it is easy to see that $V_D(b, p, n)$ in (4.15) and $V_L(b, p, n)$ in (4.16) are also non-decreasing.

Since $V_A(b, p, n)$ s for $A \in \{OD, OT, H\}$ are linear in p , we have $V_A(b, ap_1 + (1 - a)p_0, n) = aV_A(b, p_1, n) + (1 - a)V_A(b, p_0, n)$. Using this result, we have

$$\begin{aligned} & V_A(b, p_1, n) - V_A(b, p_0, n) \\ &= V_A(b, p_1 - p_0 + p_0, n) - V_A(b, p_0, n) \end{aligned} \tag{4.23a}$$

$$= V_A(b, p_1 - p_0, n) \geq 0, \quad A \in \{OD, OT, H\} \tag{4.23b}$$

Note that (4.23b) follows from the fact that $V_A(b, p_1 - p_0 + p_0, n) = V_A(b, p_1 - p_0, n) + V_A(b, p_0, n)$. Since the value function, $V(b, p, n)$, is the maximum of non-decreasing functions, it is also non-decreasing. Hence, by letting $n \rightarrow \infty$, we prove that $V(b, p)$ is non-decreasing in p . \square

Lemma 4.1 is necessary in proving the structure of the optimal policy. For each $b \in \{0, 1, \dots, B_{max}\}$ and $A \in \mathcal{A}$, we define:

$$\Phi_A^b \triangleq \{p \in [0, 1] : V(b, p) = V_A(b, p)\}. \tag{4.24}$$

For any $b \geq 0$, Φ_A^b characterizes the set of belief states for which it is optimal to choose action A . In Theorem 4.1, we show that the optimal policy has a threshold-type structure.

Theorem 4.1. *The optimal policy is a threshold-type policy on the belief state p , and the thresholds are functions of the battery state, b .*

Proof. This theorem states that the optimal policy has a threshold structure. Initially, we aim to prove that Φ_A^b for $A \in \{OD, OT, H\}$ is convex. It is easy to see that for $b = 0$, $V(b, p) = V_D(b, p)$, and hence, $\Phi_D^0 = [0, 1]$, and $\Phi_L^0 = \Phi_{OD}^0 = \Phi_{OT}^0 = \Phi_H^0 = \emptyset$. First, we consider battery states $\tau\mathcal{E}_T \leq b < \mathcal{E}_T$. We will prove that for any $\tau\mathcal{E}_T \leq b < \mathcal{E}_T$, Φ_{OD}^b is convex. Let $p_1, p_2 \in \Phi_{OD}^b$, and $a \in (0, 1)$. We have

$$V(b, ap_1 + (1 - a)p_2) \leq aV(b, p_1) + (1 - a)V(b, p_2), \quad (4.25)$$

$$= aV_{OD}(b, p_1) + (1 - a)V_{OD}(b, p_2), \quad (4.26)$$

$$= V_{OD}(b, ap_1 + (1 - a)p_2), \quad (4.27)$$

$$\leq V(b, ap_1 + (1 - a)p_2), \quad (4.28)$$

where (4.25) follows from Lemma 4.1; (4.26) is due to the fact that $p_1, p_2 \in \Phi_{OD}^b$; (4.27) follows from the linearity of V_{OD} in p ; and (4.28) holds due to the definition of $V(b, p)$. Consequently, $V(b, ap_1 + (1 - a)p_2) = V_{OD}(b, ap_1 + (1 - a)p_2)$, and it follows that $ap_1 + (1 - a)p_2 \in \Phi_{OD}^b$, which, in turn, proves the convexity of Φ_{OD}^b . Note also that $p = 0$ and $p = 1$ both belong to Φ_D^b for all $0 \leq b < \mathcal{E}_T$. Since no transmission is possible for $0 \leq b < \mathcal{E}_T$, we have $\Phi_L^b = \Phi_H^b = \emptyset$. Hence, for $0 \leq b < \mathcal{E}_T$, either $\Phi_{OD}^b = \emptyset$, or there exists $0 < \rho_1(b) \leq \rho_2(b) < 1$ such that $\Phi_{OD}^b = [\rho_1(b), \rho_2(b)]$. Consequently, we have $\Phi_D^b = [0, \rho_1(b)) \cup (\rho_2(b), 1]$, if $0 \leq b < \mathcal{E}_T$.

Next, consider $\mathcal{E}_T \leq b \leq B_{max}$. We will prove that Φ_H^b , Φ_{OD}^b , and Φ_{OT}^b are convex subsets of the belief state set. Let $p_1, p_2 \in \Phi_H^b$ and $a \in (0, 1)$. Similar to (4.25)-(4.28) we can argue

$$\begin{aligned} V(b, ap_1 + (1 - a)p_2) &\leq aV(b, p_1) + (1 - a)V(b, p_2), \\ &= aV_H(b, p_1) + (1 - a)V_H(b, p_2), \\ &= V_H(b, ap_1 + (1 - a)p_2), \\ &\leq V(b, ap_1 + (1 - a)p_2). \end{aligned} \quad (4.29)$$

Consequently, $V(b, ap_1 + (1 - a)p_2) = V_H(b, ap_1 + (1 - a)p_2)$; and hence, $ap_1 + (1 - a)p_2 \in \Phi_H^b$, which proves the convexity of Φ_H^b . Since it is always optimal to transmit at rate

R_2 if the channel is in a GOOD state (see [58], and Section 4.7) $1 \in \Phi_H^b$, and since the convex subsets of the real line are intervals, there exists $\rho_N(b) \in (0, 1]$ such that $\Phi_H^b = [\rho_N(b), 1]$. Note that N is the number of thresholds, which depends on the system parameters. Using the same technique we can prove that Φ_{OD}^b and Φ_{OT}^b are both convex, and hence, there exists $0 < \rho_{i_1}(b) \leq \rho_{i_2}(b) \leq \rho_{j_1}(b) \leq \rho_{j_2}(b) \leq \rho_N(b) \leq 1$, such that $\Phi_{OD}^b = [\rho_{i_1}(b), \rho_{i_2}(b)]$ and $\Phi_{OT}^b = [\rho_{j_1}(b), \rho_{j_2}(b)]$; or $\Phi_{OT}^b = [\rho_{i_1}(b), \rho_{i_2}(b)]$ and $\Phi_{OD}^b = [\rho_{j_1}(b), \rho_{j_2}(b)]$. However, since $V_A(b, ap_1 + (1 - a)p_2) \neq aV_A(b, p_1) + (1 - a)V_A(b, p_2)$ for $A \in \{D, L\}$, in general, Φ_D^b and Φ_L^b are not necessarily convex sets. \square

Although the optimal policy is of threshold-type, as shown in Theorem 4.1, the subsets of the belief space associated with actions D and L , i.e., Φ_D^b and Φ_L^b , are not necessarily convex. Each of these sets can be composed of infinitely many intervals; therefore, despite the threshold-type structure, characterizing the optimal policy may require identifying infinitely many threshold values. Finding the exact N and corresponding threshold values is elusive and out of the scope of this chapter.

4.4.2 Special Case: $R_1 = 0$

In order to further simplify the problem we assume that it is not possible to transmit any bits when the channel is in a BAD state, i.e., $R_1 = 0$ and $R_2 = R$. Hence, action L is no longer available, and the action for sensing the channel consists of only OD which is denoted by O in the rest of this section.

With this modified model, the expected reward function can be simplified as follows:

$$R(S_t, A_t) = \begin{cases} X_t R, & \text{if } A_t = H \text{ and } B_t \geq \mathcal{E}_T, \\ (1 - \tau)X_t R, & \text{if } A_t = O \text{ and } B_t \geq \mathcal{E}_T, \\ 0, & \text{otherwise.} \end{cases} \quad (4.30)$$

Since at least \mathcal{E}_T energy units is required for transmission, if $b < \mathcal{E}_T$, the reward in (4.30) becomes zero. If action H is taken, R bits are transmitted successfully if the channel is in a GOOD state, and 0 bits otherwise. If action O is taken, $\tau\mathcal{E}_T$ energy units is spent sensing the channel with the remainder of the energy being used for transmission if the channel is in a GOOD state. In this case, $(1 - \tau)R$ bits are transmitted successfully.

If the channel is in a BAD state, the transmitter remains silent in the rest of the time slot. Finally, if action D is taken the reward is zero.

Next, we prove that the optimal policy has a threshold-type structure on the belief state with a finite number of thresholds. Note that, in the modified model, the value function is still convex and Lemmas 4.1, 4.2 and 4.3 still hold. Theorem 4.2 below states that the optimal solution of the problem defined in (4.3) is a threshold-type policy with either two or three thresholds on the belief state. Threshold values depend on the state of the battery and system parameters.

Theorem 4.2. *Let $p \in [0, 1]$ and $b \geq 0$. There are thresholds $0 \leq \rho_1(b) \leq \rho_2(b) \leq \rho_3(b) \leq 1$, all of which are functions of the battery state b , such that for $\tau\mathcal{E}_T \leq b < \mathcal{E}_T$*

$$\pi^*(b, p) = \begin{cases} D, & \text{if } 0 \leq p < \rho_1(b) \text{ or } \rho_2(b) < p \leq 1, \\ O, & \text{if } \rho_1(b) \leq p \leq \rho_2(b). \end{cases} \quad (4.31)$$

and for $b \geq \mathcal{E}_T$,

$$\pi^*(b, p) = \begin{cases} D, & \text{if } 0 \leq p < \rho_1(b) \text{ or } \rho_2(b) < p < \rho_3(b) \\ O, & \text{if } \rho_1(b) \leq p \leq \rho_2(b), \\ H, & \text{if } \rho_3(b) \leq p \leq 1, \end{cases} \quad (4.32)$$

Proof. The proof follows similarly to the proof of Theorem 4.1. Consider the sets Φ_A^b defined in (4.24) for $A \in \{D, O, H\}$. Note that for $b = 0$, $V(b, p) = V_D(b, p)$, and hence, $\Phi_D^0 = [0, 1]$, and $\Phi_O^0 = \Phi_H^0 = \emptyset$. First, consider battery states $\tau\mathcal{E}_T \leq b < \mathcal{E}_T$. We prove that for any $\tau\mathcal{E}_T \leq b < \mathcal{E}_T$, Φ_O^b is convex, which implies the structure of the optimal policy in (4.31). Let $p_1, p_2 \in \Phi_O^b$, and $a \in (0, 1)$. We have

$$V(b, ap_1 + (1-a)p_2) \leq aV(b, p_1) + (1-a)V(b, p_2), \quad (4.33)$$

$$= aV_O(b, p_1) + (1-a)V_O(b, p_2), \quad (4.34)$$

$$= V_O(b, ap_1 + (1-a)p_2), \quad (4.35)$$

$$\leq V(b, ap_1 + (1-a)p_2), \quad (4.36)$$

where (4.33) follows from Lemma 4.1; (4.34) is due to the fact that $p_1, p_2 \in \Phi_O^b$; (4.35) follows from the linearity of V_O in p ; and (4.36) from the definition of $V(b, p)$. Hence,

$V(b, ap_1 + (1 - a)p_2) = V_O(b, ap_1 + (1 - a)p_2)$, and it follows that $ap_1 + (1 - a)p_2 \in \Phi_O^b$, which, in turn, proves the convexity of Φ_O^b . Note also that $p = 0$ and $p = 1$ both belong to Φ_D^b for all $0 \leq b < \mathcal{E}_T$. Hence, for $0 \leq b < \mathcal{E}_T$, either $\Phi_O^b = \emptyset$, or there exists $0 < \rho_1(b) \leq \rho_2(b) < 1$ such that $\Phi_O^b = [\rho_1(b), \rho_2(b)]$. Consequently, we have $\Phi_D^b = [0, \rho_1(b)) \cup (\rho_2(b), 1]$.

Next, consider $\mathcal{E}_T \leq b \leq B_{max}$. We prove that Φ_H^b and Φ_O^b are both convex, which implies the structure of the optimal policy in (4.32). Let $p_1, p_2 \in \Phi_H^b$ and $a \in (0, 1)$. Similarly to (4.25)-(4.28) we can argue

$$\begin{aligned}
V(b, ap_1 + (1 - a)p_2) &\leq aV(b, p_1) + (1 - a)V(b, p_2), \\
&= aV_H(b, p_1) + (1 - a)V_H(b, p_2), \\
&= V_H(b, ap_1 + (1 - a)p_2), \\
&\leq V(b, ap_1 + (1 - a)p_2).
\end{aligned} \tag{4.37}$$

Thus, $V(b, ap_1 + (1 - a)p_2) = V_H(b, ap_1 + (1 - a)p_2)$; and hence, $ap_1 + (1 - a)p_2 \in \Phi_H^b$, which proves the convexity of Φ_H^b . Since it is always optimal to transmit at rate R_2 if the channel is in a GOOD state, $1 \in \Phi_H^b$, and since the convex subsets of the real line are intervals, there exists $\rho_3(b) \in (0, 1]$ such that $\Phi_H^b = [\rho_3(b), 1]$. Using the same technique we can prove that Φ_O^b is convex; and hence, there exists $0 < \rho_1(b) \leq \rho_2(b) < 1$ such that $\Phi_O^b = [\rho_1(b), \rho_2(b)]$. The remaining segments belong to action D , and we have $\Phi_D = [0, \rho_1(b)) \cup (\rho_2(b), \rho_3(b))$. \square

Theorem 4.2 proves that at any battery state $b \geq \mathcal{E}_T$, at most three threshold values are sufficient to characterize the optimal policy; whereas two thresholds suffice for $0 \leq b < \mathcal{E}_T$. However the optimal policy can even be simpler for some battery states and some instances of the problem as it is possible to have $\rho_2(b) = \rho_3(b)$, or even $\rho_1(b) = \rho_2(b) = \rho_3(b)$. Since, Φ_D^b is not a convex set in general (see Theorem 4.1), the structure of the optimal policy may result in four different regions even though there are only three possible actions. This may seem counter intuitive since deferring the transmission should not be advantageous when the belief is relatively high. Nevertheless, in Section 4.5, we demonstrate that in some cases it is indeed optimal to have a three-threshold policy.

4.5 Numerical Results

In this section, we use numerical techniques to characterize the optimal policy, and evaluate its performance. We utilize the value iteration algorithm to calculate the optimal value function. We numerically identify the thresholds for the optimal policy for different scenarios. We also evaluate the performance of the optimal policy, and compare it with some alternative policies in terms of throughput.

4.5.1 Evaluating the optimal policy

In the following, we consider the modified system model introduced in Section 4.4.2 in which no data can be transmitted in a BAD channel state, i.e., $R_1 = 0$. Moreover, without loss of generality, we set $M = 11$, $\mathcal{E}_T = 10$, and $q_{10} = q = 1 - q_0$ and $q_m = 0$ for $m = 1, \dots, 9$. We assume that $B_{max} = 50$, $\tau = 0.2$, $\beta = 0.98$, $\lambda_1 = 0.9$, $\lambda_0 = 0.6$, $R = 3$ and $q = 0.1$. The optimal policy is evaluated using the value iteration algorithm. In Fig. 4.2, each state (b, p) is illustrated with a different color corresponding to the optimal policy at that state. In Fig. 4.2, the areas highlighted with blue correspond to those states at which deferring the transmission is optimal, green areas correspond to the states at which sensing the channel is optimal, and finally yellow areas correspond to the states at which transmitting at high rate is optimal. As seen in Fig. 4.2, depending on the battery state the optimal policy may have one, two, or three thresholds on the belief state. For example, when the battery state is $b = 20$, there is a single threshold; the transmitter defers transmission up to a belief state of $p = 0.8$, and starts transmitting without sensing beyond this value. For no value of the belief state it opts for sensing the channel. On the other hand, when the battery state is 38, the policy has two thresholds, and three thresholds when the battery state is 28. Considering the low probability of energy arrivals ($q = 0.1$) and the relative high cost of sensing ($\tau = 0.2$), the transmitter senses the channel even when its battery state is below the transmission threshold, i.e., $b < 10$.

Another interesting observation from Fig. 4.2 is the periodicity of the optimal policy with respect to the battery. This is particularly visible for action D taken when the battery state is an integer multiple of \mathcal{E}_T , which is then followed by action O for increasing beliefs when the battery state is more than 20. The value function corresponding to the parameters used to obtain Fig. 4.2 is depicted in Fig. 4.3. Note the staircase behavior

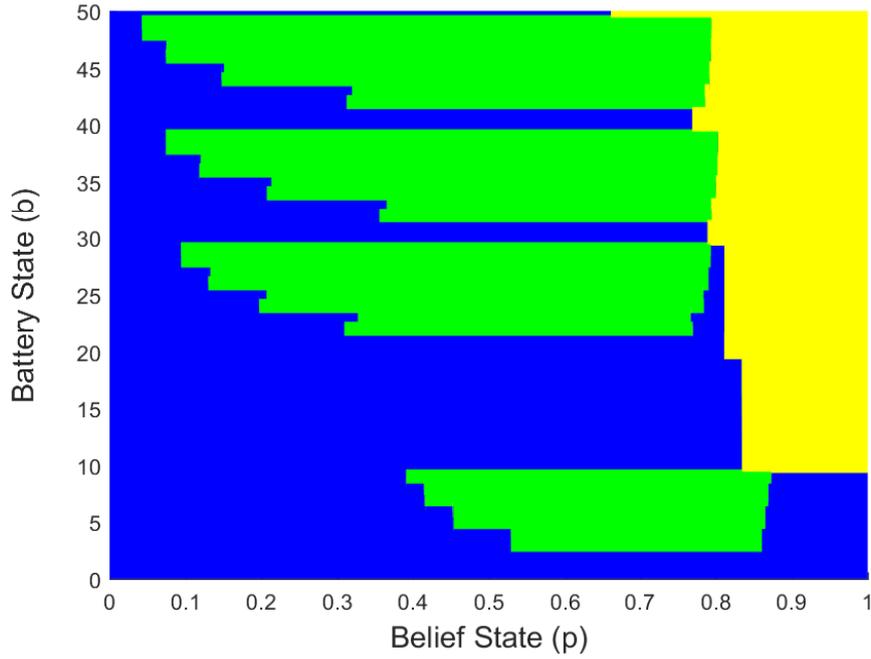


Figure 4.2: Optimal thresholds for taking the actions D (blue), O (green), H (yellow) for $B_{max} = 50$, $\mathcal{E}_T = 10$, $\tau = 0.2$, $\beta = 0.98$, $\lambda_1 = 0.9$, $\lambda_0 = 0.6$, $R = 3$ and $q = 0.1$.

of the value function. There is a jump in the value function when the battery state is an integer multiple of \mathcal{E}_T , while it approximately remains the same when the battery state is confined between two consecutive integer multiples of \mathcal{E}_T , i.e., $(n\mathcal{E}_T \leq b < (n+1)\mathcal{E}_T)$, where n is an integer. Hence, when the battery state of the transmitter is an integer multiple of \mathcal{E}_T , any action other than deferring will, with high probability, transition into a state with a relatively lower value. Thus, the transmitter chooses action D unless its belief is relatively high. However, when the battery state is between two consecutive integer multiples of \mathcal{E}_T , it is safe to sense the channel, since, in the worst case, the channel is in a BAD state and the transmitter loses only $\tau\mathcal{E}_T < \mathcal{E}_T$ units, but it makes a transition into a state which approximately has the same value. Thus, at those values of the battery, the transmitter senses the channel for moderate belief states.

To investigate the effect of the EH rate, q , on the optimal transmission policy, we consider the system parameters $B_{max} = 50$, $\tau = 0.1$, $\beta = 0.9$, $\lambda_1 = 0.8$, $\lambda_0 = 0.4$, and $R = 3$. We illustrate the optimal transmission policy for $q = 0.8$ and $q = 0.2$ in Fig. 4.4a and Fig. 4.4b, respectively. It can be observed by comparing those two figures that the yellow regions are much larger and blue areas are much more limited in Fig. 4.4a. This is because when the energy arrivals are more frequent, the EH node tends to

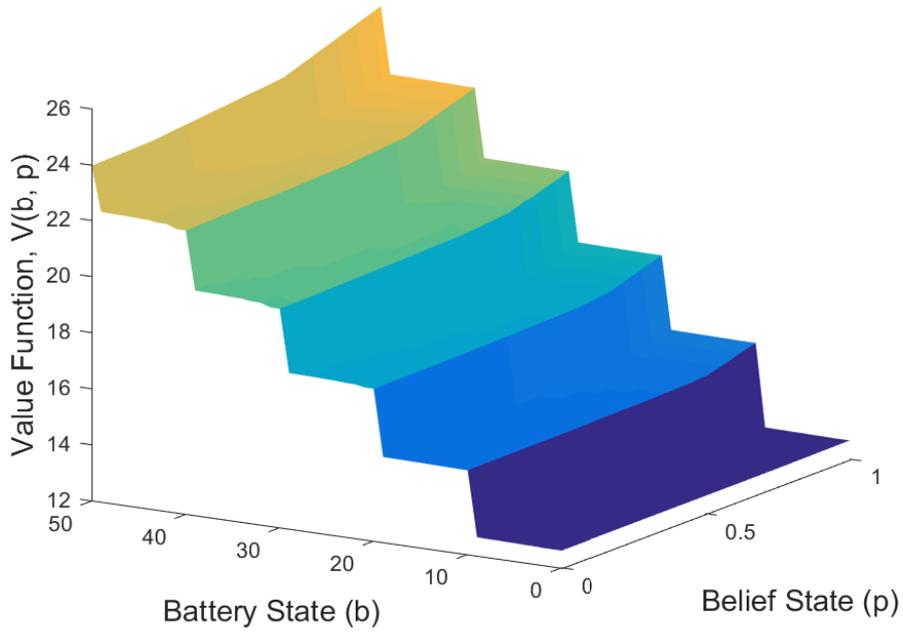


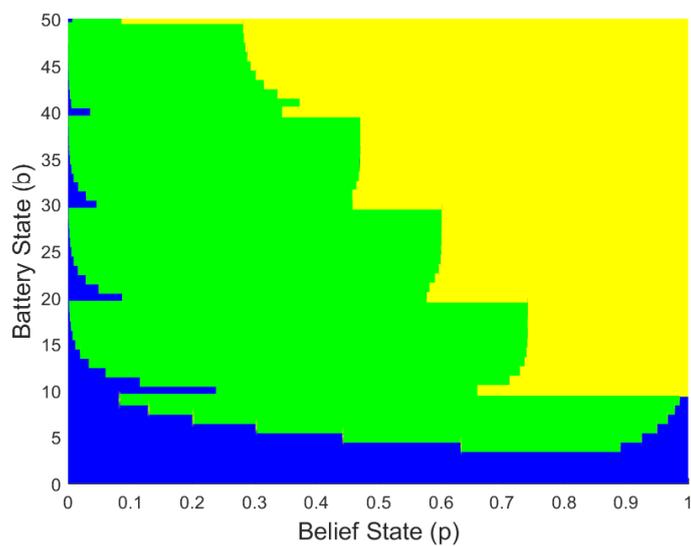
Figure 4.3: Value function associated with $B_{max} = 50$, $\mathcal{E}_T = 10$, $\tau = 0.2$, $\beta = 0.98$, $\lambda_1 = 0.9$, $\lambda_0 = 0.6$, $R = 3$ and $q = 0.1$.

consume its energy more generously. We also observe that the transmitter always defers its transmission for $b < 10$ when energy is limited (in Fig. 4.4b), whereas it may opt for sensing the channel when energy is more abundant.

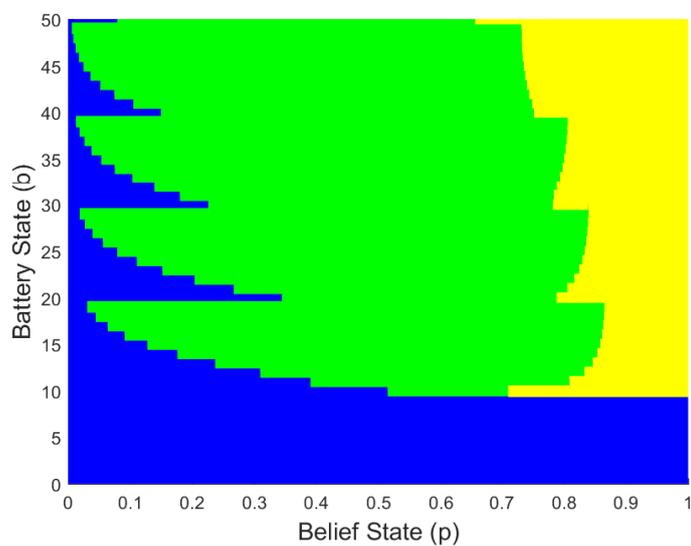
Next, we investigate the effect of the sensing cost, τ , on the optimal policy. We set the system parameters as $B_{max} = 50$, $\beta = 0.9$, $\lambda_1 = 0.8$, $\lambda_0 = 0.4$, $R = 3$ and $q = 0.8$. The regions for optimal actions are shown in Fig. 4.5a and Fig. 4.5b for sensing cost values $\tau = 0.2$ and $\tau = 0.3$, respectively. By comparing Fig. 4.5a and Fig. 4.5b, it is evident that a higher cost of sensing reduces the incentive for sensing the channel. We observe in Fig. 4.5b that the green areas have shrunk as compared to Fig. 4.5a, i.e., the transmitter is more likely to take a risk and transmit without sensing, or defer its transmission, when sensing consumes a significant portion of the available energy.

4.5.2 Throughput performance

In this section, we compare the performance of the optimal policy with three alternative policies, i.e., a greedy policy, a single-threshold policy and an opportunistic policy. For the optimal policy, as an alternative to the value iteration algorithm, we also employ *policy search* approach, which exploits the threshold structure of the optimal policy that we

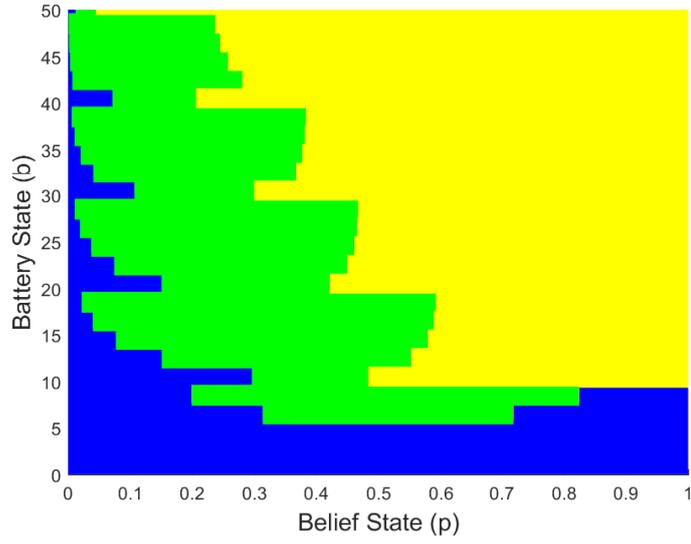


(a) $q = 0.8$.

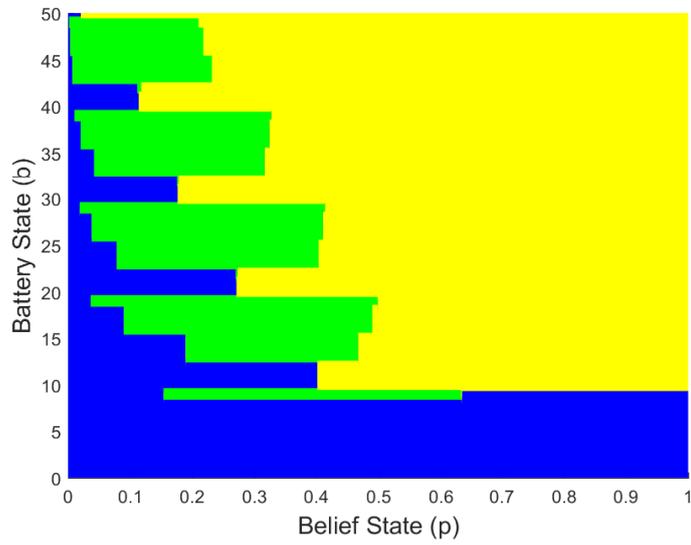


(b) $q = 0.2$.

Figure 4.4: Optimal thresholds for taking the actions D (blue), O (green), H (yellow) for $B_{max} = 50$, $\mathcal{E}_T = 10$, $\tau = 0.1$, $\beta = 0.9$, $\lambda_1 = 0.8$, $\lambda_0 = 0.4$, and $R = 3$.



(a) $\tau = 0.2$.

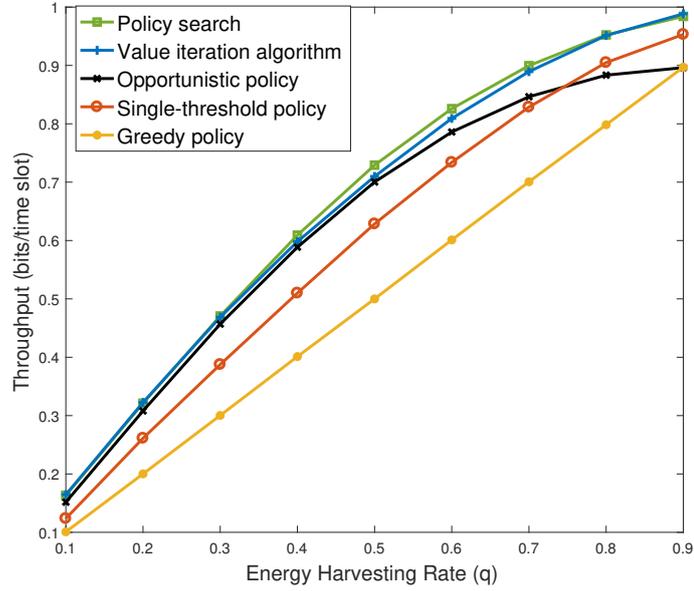


(b) $\tau = 0.3$.

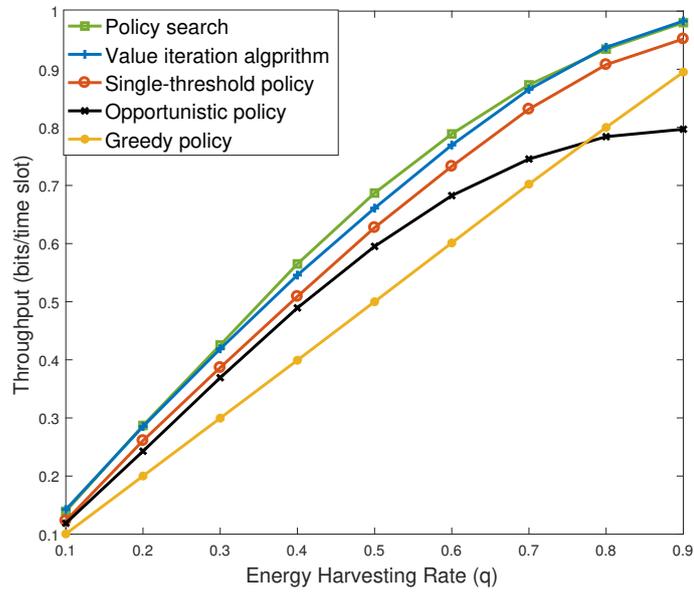
Figure 4.5: Optimal thresholds for taking the actions D (blue), O (green), H (yellow) for $B_{max} = 50$, $\beta = 0.9$, $\mathcal{E}_T = 10$, $\lambda_1 = 0.8$, $\lambda_0 = 0.4$, $R = 3$ and $q = 0.8$.

have proven. For the value iteration algorithm, the average discounted reward is evaluated with a discount value close to 1 ($\beta = 0.999$) to approximate the optimal average throughput. Note that, the value iteration algorithm does not exploit the structure of the optimal policy and uses action-value functions to maximize the discounted reward. The policy search method [67], on the other hand, uses the structure of the optimal policy, and the thresholds are directly optimized to maximize the average throughput (and not the discounted throughput). In the *greedy policy*, the EH node transmits whenever it has energy in its battery. In the *single-threshold policy*, there are only two actions: defer (D) or transmit (H). The belief of the transmitter on the current channel state depends only on the ACK/NACK feedback from the receiver, and channel sensing is not exploited at all. We optimize the threshold corresponding to each battery state for the single-threshold policy using the value iteration algorithm. Meanwhile, the *opportunistic policy* senses the channel at the beginning of every time slot, and transmits $(1 - \tau)R$ bits if the channel is in a GOOD state, and defers otherwise. By choosing the parameters $B_{max} = 50$, $\mathcal{E}_T = 10$, $\beta = 0.999$, $\lambda_1 = 0.8$, $\lambda_0 = 0.2$, $R = 2$, the throughput achieved by these four policies are plotted in Fig. 4.6 with respect to the EH rate q . Fig. 4.6a and Fig. 4.6b correspond to the sensing costs of $\tau = 0.1$ and $\tau = 0.2$, respectively.

As expected, the greedy policy performs much worse than the optimal policy as it does not exploit the transmitter's knowledge about the state of the channel. We can see that, by simply exploiting the ACK/NACK feedback from the receiver in order to defer transmission, the single-threshold policy already achieves a significantly higher throughput than the greedy policy at all values of the EH rate. Note that single-threshold and greedy policies do not have the sensing capability, and accordingly, the sensing cost, τ , has no effect on their performance. However, τ affects the optimal and opportunistic policies which have sensing capabilities. In particular, τ affects the opportunistic policy drastically, since this policy senses the channel at the beginning of each time slot. When the sensing cost is relatively low, it can be seen from Fig. 4.6a that the opportunistic policy achieves a near optimal throughput except when the EH rate, q , is high. For high values of q , the EH transmitter suffers less from energy deprivations and instead of sensing at each time slot, using the whole time slot for transmission becomes more beneficial. Hence, we observe that always sensing the channel performs poorly q is high. When τ is relatively high, it can be seen from Fig. 4.6b that the opportunistic policy performs worse



(a) $\tau = 0.1$.



(b) $\tau = 0.2$.

Figure 4.6: Throughputs by the optimal, greedy, single-threshold and opportunistic policies as a function of the EH rate, q .

than the single-threshold policy for all values of q , and even worse than the greedy policy for high values of q . On the other hand, the optimal policy, by intelligently utilizing the sensing capability, yields a superior performance for all the parameter values.

Remark 4.2. *We remark that the policy search achieves a better performance than the value iteration algorithm. This is because the latter maximizes the discounted reward rather than the average reward. To obtain the optimal average reward using value iteration algorithm, we need to set $\beta \rightarrow 1$. However, the value iteration algorithm is computationally demanding, and letting $\beta \rightarrow 1$ deteriorates its convergence rate to the point of infeasibility. On the other hand, policy search optimizes the thresholds directly to maximize the average throughput, and it is much faster compared to the value iteration algorithm. We owe this superior performance to the structure of the optimal policy that we have shown.*

4.5.3 Optimal policy evaluation with two different transmission rates

When the transmitter has the ability to transmit at two different rates, we proved that the optimal policy is a threshold-type policy; however, due to non-convexity of sets Φ_D^b and Φ_L^b it is not possible to characterize the optimal policy as we have done for a transmitter with a single rate in (4.32) and (4.31). Instead, we numerically evaluate the optimal policy as follows.

Let $B_{max} = 5$, $\mathcal{E}_T = 200$, $\mathcal{E}_S = 7$, $\beta = 0.7$, $\lambda_1 = 0.98$, $\lambda_0 = 0.81$, $R_1 = 2.91$, $R_2 = 3$ and $q_{201} = q = 1 - q_0$ and $q_m = 0$ for $m = 1, \dots, 200$. Note that these parameters are chosen in a way to show the non-convexity of the sets Φ_D^b and Φ_L^b and may not be relevant for a practical scenario. The optimal policy, obtained through the value iteration algorithm, is represented in Fig. 4.7. In the figure, the areas highlighted with blue correspond to the states at which deferring (D) is optimal, red correspond to states at which transmitting at the low rate (L) is optimal, green correspond to states at which sensing and deferring is optimal (OD), black correspond to states at which sensing and transmitting opportunistically (OT) is optimal, and yellow correspond to the states for which transmitting without sensing (H) is optimal.

As expected the optimal policy is again a battery-dependent threshold-type policy with respect to the belief state. The sets Φ_D^b and Φ_L^b (blue and red areas, respectively) are not convex. In theory, an optimal policy may have infinite threshold values if the sets

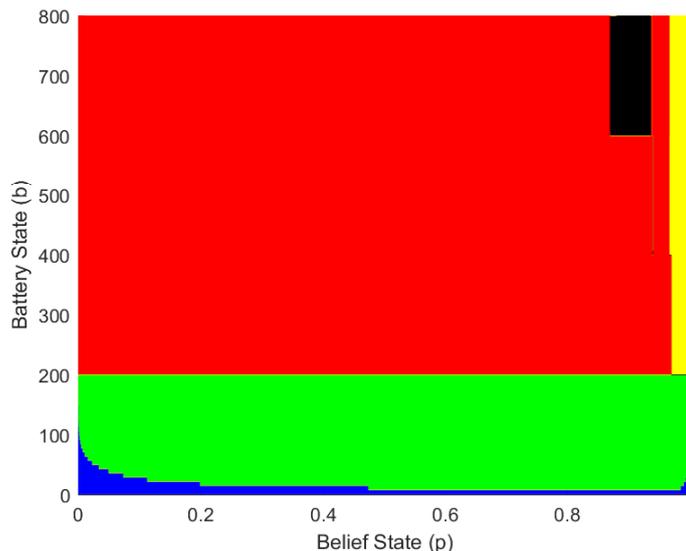


Figure 4.7: Optimal thresholds for taking the actions D (blue), L (red), OD (green), OT (black), H (yellow) for $B_{max} = 800$, $\mathcal{E}_T = 200$, $\tau = 0.035$, $\beta = 0.7$, $\lambda_1 = 0.98$, $\lambda_0 = 0.81$, $R_1 = 2.91$, $R_2 = 3$ and $q = 0.1$.

Φ_D^b and Φ_L^b are intertwined into infinitely many alternating intervals. We observe in Fig. 4.7 that, for the parameters considered here, this is not the case and the optimal policy consists of at most three-threshold policies.

4.6 Chapter Summary

In this chapter, we considered an EH transmitter equipped with a finite-capacity battery, operating over a time-varying finite-capacity channel with memory, modeled as a two-state Gilbert-Elliot channel. The transmitter receives ACK/NACK feedback after each transmission, which can be used to track the channel state. Additionally, the transmitter has the capability to sense the channel, which allows the transmitter to obtain the current channel state at a certain energy and time cost. Therefore, at the beginning of each time slot, the transmitter has the following possible actions to maximize the total expected discounted number of bits transmitted over an infinite time horizon: *i*) deferring transmission, *ii*) transmitting at a low rate of R_1 bits with guaranteed successful delivery, *iii*) transmitting at a high rate of R_2 bits, and *iv*) sensing the channel to reveal the channel state by consuming a portion of its energy and transmission time, and then deciding either to defer or to transmit at a suitable rate based on the channel state. We formulated the problem as a POMDP, which is then converted into an MDP with continuous state space

by introducing a belief parameter for the channel state. We have shown that the optimal transmission policy has a threshold structure with respect to the belief state, where the optimal threshold values depend on the battery state.

We then considered the simplified problem by assuming that it is not possible to transmit any information when the channel is in a BAD state, for which we were able to prove that the optimal policy has at most three thresholds. We calculated the optimal threshold values numerically using the value iteration and policy search algorithms. We compared the throughput achieved by the optimal policy to those achieved by a greedy policy and a single-threshold policy, which do not exploit the channel sensing capability, as well as an opportunistic policy, which senses the channel at every time slot. We have shown through simulations that the intelligent channel sensing capability improves the performance significantly, thanks to the increased adaptability to channel conditions.

4.7 Optimality of always transmitting in a GOOD state

After the sensing outcome is revealed to be in a GOOD state, the transmitter may defer, or transmit at low rate, instead of transmitting at high rate. It is easy to see that, it is suboptimal to transmit at low rate when the channel is in a GOOD state. Any low rate transmission can be replaced by a high rate transmission at no additional cost, resulting in a higher value function. To show that it is also suboptimal to defer when the channel is in a GOOD state, we need to define two new actions in addition to actions OD and OT . We define the action ODD , which defers transmission whatever the channel state is, and the action OTD , which defers transmission after sensing a GOOD channel state, but it transmits at a low rate in a BAD state. The action-value function for actions ODD and OTD evolve as follows:

$$\begin{aligned}
V_{ODD}(b, p) &= p \left[\beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \tau \mathcal{E}_T, B_{max}), \lambda_1) \right] \\
&+ (1 - p) \left[\beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \tau \mathcal{E}_T, B_{max}), \lambda_0) \right], \tag{4.38}
\end{aligned}$$

$$\begin{aligned}
V_{OTD}(b, p) &= p \left[\beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \tau \mathcal{E}_T, B_{max}), \lambda_1) \right] \\
&+ (1 - p) \left[(1 - \tau)R_1 + \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_0) \right]. \tag{4.39}
\end{aligned}$$

We will show that it is optimal to transmit after sensing a GOOD channel state by proving that $V_{OD}(b, p) > V_{ODD}(b, p)$ and $V_{OT}(b, p) > V_{OTD}(b, p)$, $\forall b, p$. First, we need the following lemma.

Lemma 4.4. For $b \geq 1$ and any $0 \leq p \leq 1$, $V(b + (1 - \tau)\mathcal{E}_T, p) - V(b, p) < (1 - \tau)R_2$.

Proof. We will use induction to prove the lemma, and define $V(b, p, n)$ as in the proof of 4.1. For $n = 1$, we have $V(b + (1 - \tau)\mathcal{E}_T, p, 1) - V(b, p, 1) = 0$. Assume that the lemma holds for $n - 1$. We need to show that the lemma also holds for n . We will prove that $V_{A_1}(b + (1 - \tau)\mathcal{E}_T, p, n) - V_{A_2}(b, p, n) \leq (1 - \tau)R_2$ for $A_1, A_2 \in \mathcal{A}_G$, where $\mathcal{A}_G = \{D, L, OD, ODD, OT, OTD, H\}$.

Let us assume that at both states $(b + (1 - \tau)\mathcal{E}_T, p, n)$ and (b, p, n) it is optimal to choose action D . We have

$$\begin{aligned}
&V(b + (1 - \tau)\mathcal{E}_T, p, n) - V(b, p, n) \\
&= V_D(b + (1 - \tau)\mathcal{E}_T, p, n) - V_D(b, p, n) \\
&= \beta \sum_{m=0}^{M-1} q_m V(\min(b + (1 - \tau)\mathcal{E}_T + m, B_{max}), J(p), n - 1) \\
&\quad - \beta \sum_{m=0}^{M-1} q_m V(\min(b + m, B_{max}), J(p), n - 1) \\
&< \beta \sum_{m=0}^{M-1} q_m (1 - \tau)R_2 = \beta(1 - \tau)R_2 < (1 - \tau)R_2. \tag{4.40}
\end{aligned}$$

Let us assume that at states $(b + (1 - \tau)\mathcal{E}_T, p, n)$ and (b, p, n) it is optimal to choose

the action L . We have

$$\begin{aligned}
& V(b + (1 - \tau)\mathcal{E}_T, p, n) - V(b, p, n) \\
&= V_L(b + (1 - \tau)\mathcal{E}_T, p, n) - V_L(b, p, n) \\
&= \beta \sum_{m=0}^{M-1} q_m V(\min(b + (1 - \tau)\mathcal{E}_T + m - \mathcal{E}_T, B_{max}), J(p), n - 1) \\
&\quad - \beta \sum_{m=0}^{M-1} q_m V(\min(b + m - \mathcal{E}_T, B_{max}), J(p), n - 1) \\
&< \beta \sum_{m=0}^{M-1} q_m (1 - \tau)R_2 = \beta(1 - \tau)R_2 < (1 - \tau)R_2. \tag{4.41}
\end{aligned}$$

Similarly, it follows that $V_A(b + (1 - \tau)\mathcal{E}_T, p, n) - V_A(b, p, n) \leq (1 - \tau)R_2$ for $A \in \{OD, ODD, OT, OTD, H\}$.

Next, we consider cases when different actions are optimal for the two state. First we assume that it is optimal to choose action D at state $(b + (1 - \tau)\mathcal{E}_T, p, n)$, and action L at state (b, p, n) . We can write

$$\begin{aligned}
& V(b + (1 - \tau)\mathcal{E}_T, p, n) - V(b, p, n) \\
&= V_D(b + (1 - \tau)\mathcal{E}_T, p, n) - V_L(b, p, n) \\
&= V_D(b + (1 - \tau)\mathcal{E}_T, p, n) - V_D(b, p, n) \\
&\quad + V_D(b, p, n) - V_L(b, p, n) \\
&< (1 - \tau)R_2 + 0 = (1 - \tau)R_2, \tag{4.42}
\end{aligned}$$

where (4.42) follows since L is the optimal action at state (b, p, n) ; and hence, $V_D(b, p, n) - V_L(b, p, n) \leq 0$. Also, $V_D(b + (1 - \tau)\mathcal{E}_T, p, n) - V_D(b, p, n) < (1 - \tau)R_2$ as we have shown in (4.40).

Similar to the derivations of (4.42), we can easily prove that $V_{A_1}(b + (1 - \tau)\mathcal{E}_T, p, n) - V_{A_2}(b, p, n) \leq (1 - \tau)R_2$ for $A_1 \in \mathcal{A}_G$ and $A_2 \in \{\mathcal{A}_G \setminus A_1\}$.

Combining all the above results, we can finally state that $V(b + (1 - \tau)\mathcal{E}_T, p, n) - V(b, p, n) < (1 - \tau)R_2$. Since $V(b, p, n) \rightarrow V(b, p)$ as $n \rightarrow \infty$, we have $V(b + (1 - \tau)\mathcal{E}_T, p) - V(b, p) < (1 - \tau)R_2$. \square

In the following, we will show that $V_{OD}(b, p) > V_{ODD}(b, p)$. We have

$$\begin{aligned}
V_{OD}(b, p) - V_{ODD}(b, p) &= p(1 - \tau)R_2 \\
&+ p\beta \sum_{m=0}^{M-1} q_m [V(\min(b + m - \mathcal{E}_T, B_{max}), \lambda_1) \\
&\quad - V(\min(b + m - \tau\mathcal{E}_T, B_{max}), \lambda_1)] \tag{4.43a}
\end{aligned}$$

$$> p(1 - \tau)R_2 - p\beta \sum_{m=0}^{M-1} q_m(1 - \tau)R_2 = p(1 - \beta)(1 - \tau)R_2 > 0, \tag{4.43b}$$

where we use the result established in Lemma 4.4 to simplify (4.43a) into (4.43b). With the same outline in the above, it directly follows that $V_{OT}(b, p) > V_{OTD}(b, p)$. The intuition behind the above result is the fact that by saving $(1 - \tau)\mathcal{E}_T$ units of energy in the GOOD state, one cannot get a better reward than $(1 - \tau)R_2$ in the future. Hence, there is no reason to save the energy when we are sure that the channel is in a GOOD state.

Chapter 5

Reliable Communication for a SWIPT enabled Receiver

We consider a class of wireless powered devices employing Hybrid Automatic Repeat reQuest (HARQ) to ensure reliable end-to-end communications over a two-state time-varying channel. A receiver, with no power source, relies on the energy transferred by a Simultaneous Wireless Information and Power Transfer (SWIPT) enabled transmitter to *receive* and *decode* information. Under the two-state channel model, information is received at two different rates while it is only possible to harvest energy in one of the states. The receiver aims to decode its messages with minimum expected number of re-transmissions. Dynamic and continuous nature of the problem motivated us to use a novel Markovian framework to bypass the complexities plaguing the conventional approaches such as MDP. Using the theory of absorbing Markov chains, we show that there exists an optimal policy utilizing the incoming RF signal solely to harvest energy or to accumulate mutual information. Hence, we convert the original problem with continuous action and state space into an equivalent one with discrete state and action space. For independent and identically distributed channels, we prove the optimality of a simple-to-implement *harvest-first-store-later* type policy. However, for time-correlated channels, we demonstrate that statistical knowledge of the channel may significantly improve the performance over such policies.

5.1 Overview

5.1.1 Background and Motivation

In simultaneous wireless information and power transfer (SWIPT), the incoming RF signal is used for both energy harvesting and decoding of information bits. The inherent challenge of energy harvesting (EH) is the stochastic nature of the EH process, which dictates the amount and availability of harvested energy that is beyond the control of system designers. However, SWIPT may provide the network administrators a leverage on replenishing the remote devices for proper network operations.

In the seminal paper [68], the rates at which energy and reliable information can be transferred over a single point-to-point noisy link were characterized. This result was later extended to frequency-selective channels with additive white Gaussian noise (AWGN) in [69]. In [70], the authors examined separated and co-located information and energy receiver architectures in a multiple-input multiple-output (MIMO) wireless broadcast system. In separated architecture, both receivers have separate antennas, whereas in co-located architecture a single antenna is shared by both. In general, EH devices have small footprints necessitating a co-located architecture. This arises a resource allocation problem of sharing the RF signal among the two receivers. The incoming RF signal is fed to Information Decoding (ID) and Energy Harvesting (EH) circuitries by applying either time-switching (TS) or power splitting (PS) schemes. In TS, the RF signal is split over two different parts of the time slot, one for EH and the other for ID, whereas in PS the incoming RF signal is fed to both, proportional to a given factor. In this chapter, we consider the class of PS policies. In particular, we consider two types of PS policies: *splitting* and *no-splitting*. A splitting policy divides the RF signal into two parts with strictly non-zero power and feeds them to ID and EH circuitries, whereas no-splitting policy feeds the RF signal completely to either EH or ID.

In inherently error-prone wireless communications systems, re-transmissions triggered by decoding errors have a major impact on the energy consumption of wireless devices. Hybrid automatic repeat request (HARQ) schemes are frequently used in order to reduce the number of re-transmissions by employing various channel coding techniques [21]. Nevertheless, this comes at the expense of extra processing time and energy associated with the enhanced error-correction decoders. A receiver employing HARQ

encounters two major energy consuming operations: (1) sampling or Analog-to-Digital Conversion (ADC), which includes all RF front-end processing, and (2) decoding. The energy consumption attributed to sampling, quantization and decoding plays a critical role in energy-constrained networks which makes their study a non-trivial problem. The authors in [71] investigated the performance of HARQ over an RF-energy harvesting point-to-point link, where the power transfer occurs over the downlink and the information transfer over the uplink. The authors studied the use of a TS policy when two HARQ mechanisms are used for information transfer; Simple HARQ (SH) and HARQ with Chase Combining (CC) [72]. Also, the authors in [73] studied the performance of HARQ in RF energy harvesting receivers, where heuristic TS policies are proposed to reduce the number of re-transmissions.

In this chapter, we consider a point-to-point link where an energy-abundant transmitter employs HARQ to deliver a message reliably to an EH receiver. The receiver has no energy source, so it relies on harvesting energy from the information-bearing RF signal. The channel is time-varying where the amount of energy harvested and information collected varies depending on the quality of the channel. The receiver aims to split the incoming RF signal between EH and ID so that the expected number of re-transmissions is minimized. Unlike prior works, e.g., [74], we do not assume the availability of the channel state information (CSI) at the receiver¹.

5.1.2 Contributions

Our main contributions in this chapter are summarized as follows:

- We formulate the problem of minimizing the expected number of re-transmissions using a Markov decision process (MDP).
- Due to the excessive number of states and actions in the MDP formulation, we use the special features of the EH HARQ framework to recast the MDP as a problem of minimizing the expected time to absorption in an absorbing Markov chain, significantly reducing the complexity associated with the MDP, when the wireless channel exhibits independent and identically distributed (i.i.d.), and time-correlated properties, respectively.

¹Due to the time and energy cost, the acquisition of CSI in EH networks is challenging. Some interesting ideas along this line, such as limited CSI feedback, have been discussed in [75].

- For i.i.d. channels, we prove that there is an optimal policy that does not split the incoming RF energy and uses it solely either for ID or EH. As a result, we convert the original problem whose states and actions take over continuous values into discrete ones, enabling a tractable solution.
- The numerical solution of the MDP identifies multiple distinct policies that achieve the minimum expected number of re-transmissions, implying that the optimal policy is not unique. Hence, we later completely characterize a class of simple-to-implement optimal policies. Among those, harvest-first-store-later is an optimal policy lending itself for simple implementation on low complexity devices.
- For a time-correlated channel, we once again show that there is an optimal policy that does not split the incoming RF energy. We develop a low complexity algorithm to determine the EH/ID decision for each state of the receiver. Note that unlike the i.i.d. case, a simple policy such as harvest-first-store-later is no longer optimal for correlated channels as demonstrated in our numerical analysis.
- We provide extensive numerical simulations to verify the analytical results established in the chapter.

5.1.3 Related Work

Early works on wireless energy transfer [76] considered a point-to-point single antenna communication system and studied its rate-energy trade-off. Single antenna systems are extended to single-input-multiple-output (SIMO) in [77], multiple-input-single-output (MISO) in [78] and multiple-input-multiple-output (MIMO) system in [79].

Note that EH devices harvest energy only in minuscule amounts (orders of μW s), so the energy consumption of the receiver circuitry to perform simple sampling and decoding can no longer be neglected. The authors in [40] addressed the energy consumption of sampling and decoding operations over a point-to-point link where the receiver harvests energy at a constant rate. In [41], a decision-theoretic approach is developed to optimally manage the transmit energy of an EH transmitter transmitting to an EH receiver, where both the transmitter and the receiver harvests energy independently from a Bernoulli energy source. The receiver uses selective sampling (SS) and informs the transmitter about

the SS information and its delayed battery state by feedback. Based on this feedback, the transmitter adjusts its transmission policy to minimize the packet error probability.

Meanwhile, in [42], the performance of different HARQ schemes for an EH receiver harvesting energy from a deterministic energy source with a constant energy rate was studied. In [43], the impact of the battery's internal resistance at the receiver was analyzed for an EH receiver with imperfect battery, with the aim of maximizing the amount of information decoded by the EH receiver. While ignoring the sampling energy cost at the receiver, [44] investigates the performance of TS policies to maximize the amount of information decoded at the receiver operating over a binary symmetric channel (BSC), by optimizing the fraction of time used for harvesting energy and for extracting information. For an EH transmitter and an EH receiver pair both harvesting ambient environmental energy with possible spatial correlation, [45] addresses the problem of outage minimization over a fading wireless channel with ACK-based re-transmission scheme by optimizing the power allocation at the transmitter. In [46], for a pair of EH transmitter-receiver employing ARQ and HARQ with binary EH process, packet drop probability over fading channels is minimized by optimally allocating power over different rounds of re-transmissions. In [47], an adaptive feedback mechanism for an EH receiver is proposed by taking into account the energy cost of sampling and decoding is proposed. The receiver is allowed to transmit a delayed feedback with the aim of efficiently utilizing the harvested energy in order to minimize the packet drop probability in the long run. In [48], the outage probability for an EH receiver powered by RF transmissions is minimized by implementing HARQ. In particular, the transmitter optimally allocates two different power levels in charging and information transmission periods so that the probability of the event that information is not correctly received by the receiver due to either unsuccessful message decoding or lack of minimum energy at the receiver is minimized. Although [48] is the most similar study to our work, it assumes that the channel stays constant during re-transmissions and it is known by the receiver. Differently, we assume that the wireless channel, with and without memory, varies over different instances of re-transmissions which calls for an online framework rather than an offline framework as in [48]. The problem of throughput optimization for an EH receiver operating in a multi-access network was studied in [80] where the receiver takes samples from the incoming RF signal to calculate the probability of a collision event and based on that decides to either utilize the incoming RF energy to

replenish its battery or to extract information bits.

In [81], maximization of long term weighted sum throughput, in an uplink scenario, for two RF EH transmitters is studied. The AP has the complete knowledge of the state of the network, i.e., battery levels, uplink and downlink CSI, and it calculates the optimal EH period, and the uplink durations of each transmitter at the beginning of each time slot. The finite horizon uplink throughput maximization for an EH transmitter with imperfect CSI and random EH process is studied in [82], and the optimal power allocation problem at each time slot is formulated using dynamic programming (DP). [83] studies the rate-energy (R-E) region of separated and co-located SWIPT architectures where R-E region characterizes all the achievable rate and harvested energy pairs under a given transmit power constraint. A strategy achieving the optimal R-E region is developed for the case of separated architecture. For the case of co-located architecture, two policies namely power splitting and time switching is investigated in terms of their achievable R-E region. In [84], for a network with a transmitter, a relay and a destination node, two relaying protocols namely power splitting based relaying (PSR) and time switching based relaying (TSR) protocols are proposed. Analytical expressions for outage probability of delay limited transmission mode and ergodic capacity of delay tolerant transmission mode are derived. In contrast to [83, 84], we show that there exists an optimal policy that does not split the incoming RF energy when HARQ mechanism is employed.

Differently from the available literature, we study the reliability of transmission by an HARQ mechanism in a SWIPT scenario, over time varying channels with unknown CSI and by considering an accurate model of energy consumption of the EH receiver. We develop a novel Markovian framework for the analysis which facilitates characterizing the optimal decision at any given time. A major contribution of this work is that we prove that there exists an optimal no-splitting policy that minimizes the number of re-transmissions. This finding enables a tractable optimal solution by reducing a two dimensional uncountable state MC into a countable state MC. In particular, for i.i.d. channels, we show that policies such as harvest-first-store-later are optimal enabling simple-to-implement optimal policies suitable for low power EH devices. However, for the case of correlated channels, we show that an intelligent algorithm that utilizes the correlation information of the channel states, can significantly outperform those simple-to-implement policies.

5.2 System Model and Preliminaries

5.2.1 Channel Model and Receiver Architecture

Consider a point-to-point time varying wireless link between a transmitter-receiver pair. The wireless channel is modeled according to a two-state block fading model where the states are GOOD and BAD². Let $G_t \in \{0, 1\}$ be the state of the channel at time slot t where BAD and GOOD states are denoted by 0 and 1, respectively. The CSI is neither available at the transmitter nor at the receiver due to the high computational and energy costs of transmitting and receiving a pilot signal necessary for measuring the CSI. We consider a communication scheme where the transmitter is connected to a power source with an unlimited energy supply. The receiver is equipped with a separate rectifier circuit for EH and a transceiver for ID, both connected to the same antenna.

Time is slotted and each slot has a length of N channel uses. We assume that N is sufficiently large so that we can apply information theoretic arguments. The instantaneous achievable rate of the receiver is the maximum achievable mutual information between the output symbols of the transmitter and input symbols at the receiver. Let the achievable rate of the receiver be $R(t)$ at time t . As $N \rightarrow \infty$, $R(t)$ approaches the Shannon rate, and it can be computed as:

$$R(t) = \log_2(1 + Pg(t)), \quad (5.1)$$

where $g(t) \in \{g_0, g_1\}$ is the channel power gain at time t and P is the noise-normalized transmit power of the transmitter. We assume that the transmitter power is fixed and known to the receiver. Let R_1 and R_0 be the achievable rates corresponding to the channel states GOOD and BAD, respectively:

$$R_1 = \log_2(1 + Pg_1), \quad (5.2)$$

$$R_0 = \log_2(1 + Pg_0). \quad (5.3)$$

²Note that the two-state channel process is an approximation of a more general multi-state time varying channel, where each state of the channel supports a maximum transmission rate. Here, we employ two-state channel process due to its analytical tractability.

The instantaneous channel states are not known a priori so we employ an HARQ scheme with incremental redundancy (IR) for providing reliability [85]. In the following, we give a brief overview of HARQ-IR.

5.2.2 Brief Overview of HARQ

HARQ is a well known method to provide reliable point to point communications [85]. There are several types of HARQ implementations, e.g., simple HARQ, HARQ with Chase Combining (CC), repetition time diversity and incremental redundancy (IR). Note that in EH devices, CSI acquisition is cost prohibitive due to the energy and temporal cost of probing the channel. Hence, here, the transmitter is blind to the instantaneous channel conditions and it cannot adapt the code rates according to a particular channel gain. Thus, in our system, we consider HARQ-IR due to its superior throughput performance [86] compared to other alternatives as well as its robustness against the absence of CSI [87]. Let us denote a message of the transmitter by $W \in \{1, 2, \dots, 2^{NC}\}$, where C denotes the rate of the information. Every incoming transport layer message into the transmitter is encoded by using a mother code of length MN channel uses. The encoded message, \mathbf{x} , is divided into M blocks, each of length N channel uses, with a variable redundancy and it is represented by $\mathbf{x} = [x^1, \dots, x^M]$. Let us assume that x^1 is transmitted at t_1 . If x^1 is successfully decoded, then the receiver sends a 1-bit, error-free, zero-delay, Acknowledgement (ACK) message, otherwise, the transmitter times out after waiting a certain time period. In case of no ACK received, the transmitter transmits x^2 at time slot t_2 and the receiver combines the previous block x^1 with x^2 . This procedure is repeated until the receiver accumulates C bits of mutual information or maximum blocks of information, M , is sent. We assume that, M is chosen sufficiently large so that the probability of decoding failure, due to exceeding the maximum number of re-transmissions, is approximately equal to zero. With HARQ-IR scheme, after r re-transmissions, the amount of accumulated mutual information at the receiver is $\sum_{k=1}^r R(t_k)$. The receiver, given that it has sufficient energy, can perform a successful decoding attempt after r re-transmissions, if the amount of accumulated mutual information exceeds the information rate of the transmitted message, i.e., $\sum_{k=1}^r R(t_k) \geq C$. We assume that each message is encoded at rate R_1 i.e., $C = R_1$ so that a transmission in a GOOD channel state carries all the information

needed for decoding³.

5.2.3 Energy Harvesting and Consumption Model

In the following, we assume that the receiver has a sufficiently large battery and memory, so there is no energy or information overflow. The receiver utilizes a PS policy, where $\rho(t) \in [0, 1]$ denotes the power splitting parameter at the beginning of time slot t . Note that $\rho(t) = 0$ indicates that the received signal is used solely for mutual information accumulation, and $\rho(t) = 1$ indicates that the received signal is used solely for harvesting energy. Any value of $\rho(t)$ between 0 and 1 refers to the case where the received signal is used for both harvesting energy and mutual information accumulation.

We incorporate a simplified energy harvesting model, which facilitates the formulation of a tractable optimization problem. In this model, the receiver harvests a maximum of $e \geq 1$ energy units in the GOOD channel state and zero units during the BAD channel state⁴. Typically, an EH device has two stages in its energy harvesting circuitry [88]: a rectifier stage that converts the incoming alternating current (AC) radio signals into direct current (DC); and a DC-DC converter that boosts the converted DC signal to a higher DC voltage value to produce the voltage required to charge the battery. The main limitation in an energy harvester is that every DC-DC converter has a minimum input voltage threshold below which it cannot operate. Hence, when the channel is in a BAD state, the input voltage is below the threshold of the DC-DC converter and no energy is harvested. Even though the receiver cannot harvest any RF energy in a BAD channel state, it can still accumulate mutual information since ID circuit operates at a lower power sensitivity, e.g., -10 dBm for EH and -60 dBm for ID circuits [89].

The energy consumption of HARQ was recently investigated in [90], and it was identified that the energy is consumed at the start up of the receiver, during decoding, for operating passband receiver elements (low-noise amplifiers, mixers, filters, frequency synthesizers, etc.), and for providing feedback to the transmitter. In order to develop a tractable optimization framework, we consider the model in [90], and combine the individual costs of energy into two parameters only: the receiver consumes $E_d \geq 1$ energy

³Note that this assumption is practically reasonable, since a time slot is typically defined as the duration of time necessary for transmission of a single information packet.

⁴The maximum energy is harvested if the received signal is completely directed to the energy harvester, i.e., $\rho(t) = 1$.

units for a decoding attempt and 1-energy unit for each mutual information accumulation event per time slot⁵, i.e., operating the passband receiver elements.

5.3 The Minimum Expected Number of Re-transmissions For I.I.D. Channels

In this section, we calculate the minimum expected number of re-transmissions needed for successful decoding for time varying channels. We first consider an i.i.d. channel, and in Section VI, we will investigate the system under a time correlated channel model. Note that the receiver requires at least E_d units of energy and R_1 bits of information before it can successfully decode the transmitted packet. Let the system states be (b, m) , where b is the total residual battery level and m is the total accumulated mutual information normalized by R_0 . For clarity of presentation, in the rest of the chapter we assume that $R_0 = 1$. Our objective is to optimally determine a scheduling policy $\rho(t)$ so that the transmission is successfully decoded with a minimum delay at the receiver. We formally define $\rho(t)$ next.

Definition 5.1. *A scheduling policy $\pi = (\rho(1), \rho(2), \dots)$ is a sequence of decision rules as such the k th element of π determines the power splitting ratio at k th time slot based on the observed system state (b, m) at the beginning of this time-slot for $t \in \{1, 2, \dots\}$. Similarly, a tail scheduling policy $\pi_t = (\rho(t), \rho(t+1), \dots)$ is a sequence of decision rules that determines the power splitting ratios for the time slots from t to ∞ .*

Let the probability that the channel is in GOOD state be λ , i.e., $\mathbb{P}[G_t = 1] = \lambda$. The problem can be mathematically modeled as a two-state Markov chain (MC). Also, let the states of the MC be (b, m) . It should be noted that the receiver is blind to the CSI before choosing the power splitting ratio. However, after it decides to sample the incoming RF signal for mutual information accumulation, the amount of the information in the sampled portion of the RF signal is revealed to the receiver. Because the scheduling policy is blind to the CSI, its decision only depends on (b, m) .

⁵One energy unit is normalized to the energy cost of operating the RF transceiver circuit during one time slot.

5.3.1 Markov Decision Process (MDP) Formulation

At any given time t , the next state of the system only depends on the current state, (b, m) , and the power split ratio $\rho(t)$. Hence, we can formulate the problem as an MDP. Let $f^\pi(t) \in \{-1, 0\}$ be an indicator function taking a value of 0 if the message can be decoded at the end of slot t under policy π , and a value of -1 otherwise. Then, the optimization problem we aim to solve is given as,

$$\max_{\pi} \sum_{t=0}^{\infty} f^\pi(t). \quad (5.4)$$

Let $V^\pi(b, 0)$ be the expected discounted reward with initial state $S_0 = (b, 0)$ under policy π with discount factor $\beta \in [0, 1)$. The expected discounted reward has the following expression

$$V^\pi(b, 0) = \mathbb{E}^\pi \left[\sum_{t=0}^{\infty} \beta^t U(S_t, \rho(t)) | S_0 = (b, 0) \right], \quad (5.5)$$

where \mathbb{E}^π is the expectation with respect to the policy π , t is the time index, $\rho(t) \in [0, 1]$ is the action chosen at time t , and $U(S_t, \rho(t))$ is the instantaneous reward acquired when the current state is S_t . In the rest of the chapter, we use $\rho(t)$ and $\rho(b, m)$ interchangeably by assuming that at time slot t , the system is at state (b, m) . The battery is recharged with incoming RF signal depending on the value of the power split ratio $\rho(t)$. Meanwhile, one unit of energy is consumed in order to accumulate non-zero bits of mutual information. Hence, the evolution of the battery state is characterized as follows:

$$B(t) = \begin{cases} B(t-1) + \rho(t)e - \mathbb{1}_{\rho(t) \neq 1}, & \text{if } G_t = 1 \\ B(t-1) - \mathbb{1}_{\rho(t) \neq 1}, & \text{if } G_t = 0 \end{cases}, \quad (5.6)$$

where $\mathbb{1}_{\rho(t) \neq 1} = 0$, if $\rho(t) = 1$, and $\mathbb{1}_{\rho(t) \neq 1} = 1$, otherwise.

According to (5.2) and (5.3), the transmit power is equal to $P = \frac{2^{R_1} - 1}{g_1} = \frac{2^{R_0} - 1}{g_0}$. At the power splitter, $1 - \rho(t)$ portion of the received power is directed into the ID, so the achievable mutual information accumulation is:

$$R(t) = \log_2(1 + g(t)P(1 - \rho(t))). \quad (5.7)$$

Note that the maximum value of the mutual information is attained by setting $\rho = 0$. Inserting the value of P in (5.7) for GOOD and BAD channel states gives the mutual information accumulation in these states respectively for a given power splitting ratio ρ as

$$R^H(\rho) = \log_2(\rho + (1 - \rho)2^{R_1}), \quad (5.8)$$

$$R^L(\rho) = \log_2(\rho + (1 - \rho)2^{R_0}). \quad (5.9)$$

Thus, the accumulated mutual information, $I(t)$, evolves as:

$$I(t) = \begin{cases} \min(I(t-1) + R^H(\rho(t)), R_1), & \text{if } G_t = 1 \\ \min(I(t-1) + R^L(\rho(t)), R_1), & \text{if } G_t = 0 \end{cases}. \quad (5.10)$$

Note that (5.10) follows from the operation of HARQ-IR which is described in Section 5.2.2 where the received messages over different time slots are combined in such a way that the mutual information of the combined messages is the summation of the individual mutual information of the messages. The instantaneous reward is zero if the message can be correctly decoded, and it is minus one otherwise. Recall that the decoding operation is successful if and only if the accumulated mutual information is above a certain threshold, and the battery level is sufficient to decode the message. Hence, the instantaneous reward is given as follows:

$$U(S_t, \rho(t)) = \begin{cases} 0, & \text{if } B_t \geq E_d, \text{ and } I(t) \geq R_1, \\ -1, & \text{if otherwise.} \end{cases}. \quad (5.11)$$

Define the value function $V(b, m)$ as

$$V(b, m) = \max_{\pi} V^{\pi}(b, m), \quad \forall b \in [0, \infty), \quad \forall m \in [0, R_1]. \quad (5.12)$$

The value function $V(b, m)$ satisfies the Bellman equation

$$V(b, m) = \max_{0 \leq \rho \leq 1} V_{\rho}(b, m), \quad (5.13)$$

where $V_{\rho}(b, m)$ is the expected reward achieved by taking action ρ when the state is (b, m) and is given by

$$V_{\rho}(b, m) = U((b, m), \rho) + \beta \mathbb{E} \left[V(\acute{b}, \acute{m}) | S = (b, m) \right], \quad (5.14)$$

where (\acute{b}, \acute{m}) is the next visited state and the expectation is over the distribution of the next state. The use of expected discounted reward allows us to obtain a tractable solution, and one can gain insights into the optimal policy when β is close to 1. Value iteration algorithm (VIA) is a standard tool to solve Bellman equations such as the one in (5.13). However, this problem suffers from the curse of dimensionality [14]. Note that from (6) and (10), the problem is a two dimensional uncountable state MDP with continuous actions at every state. Also, letting $\beta \rightarrow 1$, to approximate the average reward, slows down the algorithm to the point of infeasibility [30]. Hence, in the following, we take advantage of the special structure of our problem to derive an important characteristic of the optimal policy. The flow of the chapter is depicted in Figure 5.1.

5.3.2 Absorbing Markov Chain Formulation

Note that the MC describing the operation of our system is an *absorbing* MC, where all states except those (b, m) where $b \geq E_d$, and $m \geq R_1$ are transient states. The absorbing states are those where the receiver has both sufficient energy and information accumulated to correctly decode. In an absorbing chain, starting from a transient state, the chain makes a finite number of visits to some transient states before its eventual absorption into one of the absorbing states. Hence, the mean time to absorption of the chain, starting from

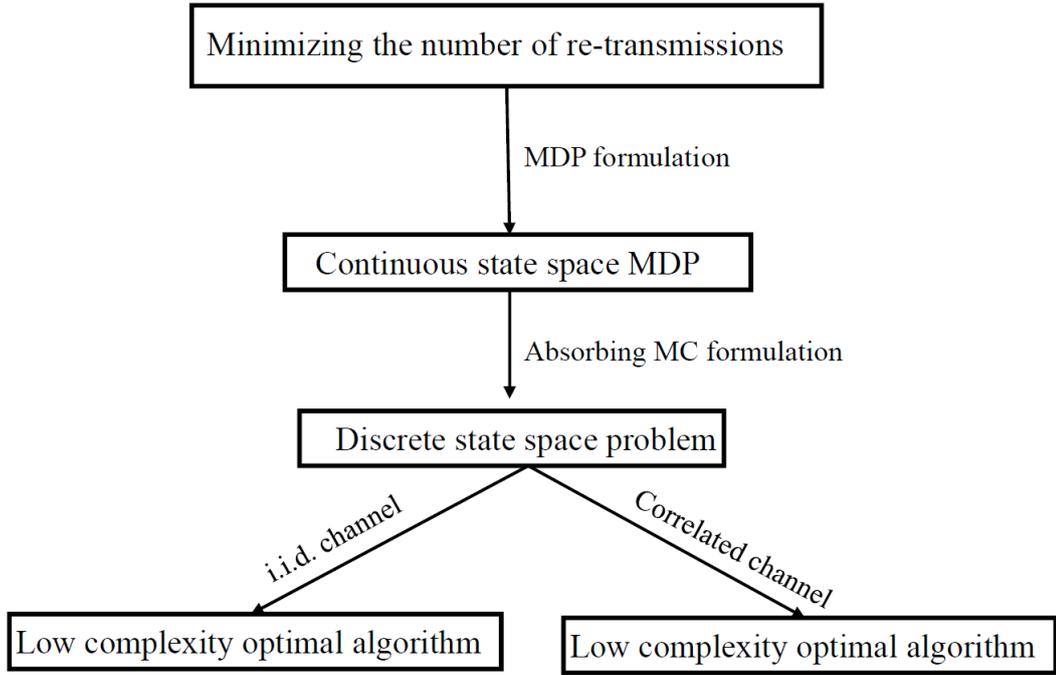


Figure 5.1: A brief overview of the chapter.

transient state i initially, is the sum of the expected numbers of visits made to transient states. In an absorbing MC, the expected number of steps taken before being absorbed in an absorbing state characterizes the *mean time to absorption*. Hence, the mean time to absorption starting from a given transient state (b, m) provides the number of re-transmissions until successful decoding when the battery has b units of energy and the memory contains m bits of information.

After establishing the ρ dependent state evolution of $B(t)$ and $I(t)$, we can formally introduce the state transition probabilities of the Markov chain as follows:

$$\rho = 1 \Rightarrow \begin{cases} \mathbb{P}((B, I), (B + l, I)) = \lambda \\ \mathbb{P}((B, I), (B, I)) = 1 - \lambda \end{cases}, \quad (5.15)$$

$$\rho = 0 \Rightarrow \begin{cases} \mathbb{P}((B, I), (B - 1, R_1)) = \lambda \\ \mathbb{P}((B, I), (B - 1, I + 1)) = 1 - \lambda \end{cases}, \quad (5.16)$$

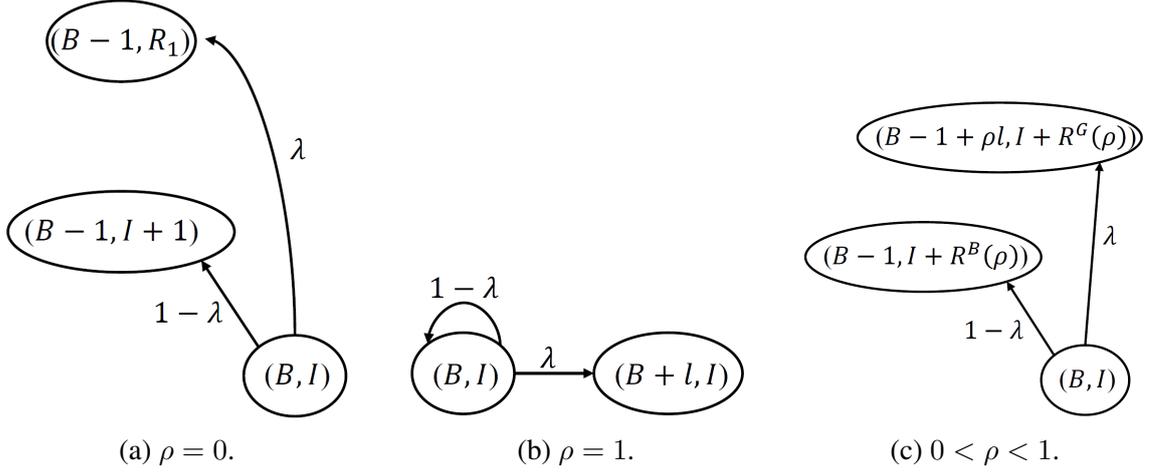


Figure 5.2: State transition probabilities of the Markov chain associated with ρ .

$$0 < \rho < 1 \Rightarrow \begin{cases} \mathbb{P}((B, I), (B - 1 + \rho l, I + R^H(\rho))) = \lambda \\ \mathbb{P}((B, I), (B - 1, I + R^L(\rho))) = 1 - \lambda \end{cases}, \quad (5.17)$$

where $\mathbb{P}(\mathbf{x}, \mathbf{y})$ is the transition probability from state \mathbf{x} into state \mathbf{y} , $B \in [0, \infty)$ and $I \in [0, R_1]$. The state transition probabilities of the Markov chain associated with ρ is depicted in Figure 5.2.

In the following, we perform *first-step analysis*, by conditioning on the first step the chain makes after moving away from a given initial state to obtain the mean time to absorption. Let $k_{b,m}$ be the expected number of transitions needed to hit an absorbing state when the MC starts from state (b, m) . The analysis is performed by assuming that the MC is in steady-state.

Let us first consider two trivial cases; when the battery has less than one unit of energy, i.e., $b < 1$, in which case the receiver has no option but harvest the incoming RF signal, and when the amount of accumulated mutual information is R_1 , in which case there is no point in further accumulating mutual information since the receiver has sufficient mutual information to decode the incoming packet. For these cases, the mean time to absorption starting from an initial state (b, m) is

$$\begin{aligned} k_{b,m} &= 1 + \lambda k_{b+e,m} + (1 - \lambda)k_{b,m} \\ &= \frac{1}{\lambda} + k_{b+e,m}, \quad \text{if } b < 1 \text{ or } m = R_1. \end{aligned} \quad (5.18)$$

Note that in (5.18), one slot is needed to harvest energy, and depending on the channel state in that slot, the battery state either transitions to $b + e$ or remains the same. The following lemma plays an important role in establishing the structure of the optimal policy.

Lemma 5.1. *For any $E_d - i \cdot e \leq b < E_d - (i - 1) \cdot e$ such that $i = 1, \dots, E_d$, given that $m = R_1$, the mean time to absorption is given by, $k_{b,R_1} = \frac{i}{\lambda}$.*

Proof. The proof is by induction.

1. Base case: Let us consider the smallest possible value for i , i.e., $i = 1$, such that $E_d - e \leq b < E_d$. Note that since $m = R_1$, the optimal decision is to use incoming RF signal only for harvesting energy, i.e., $\rho^*(b, R_1) = 1$. Thus, we get

$$k_{b,R_1} = 1 + \lambda k_{b+e,R_1} + (1 - \lambda)k_{b,R_1}. \quad (5.19)$$

For $E_d - e \leq b < E_d$, if the channel is GOOD then the MC transitions into state $(b + e, R_1)$, which is an absorbing state, so $k_{b+e,R_1} = 0$. Hence, $k_{b,R_1} = \frac{1}{\lambda}$ and thus, the lemma holds for $i = 1$.

2. Induction step: assume that the lemma is true for some $i = n$, i.e., $k_{b,R_1} = n/\lambda$ for $E_d - n \cdot e \leq b < E_d - (n - 1) \cdot e$.
3. Proof for case $i = n + 1$: Let us calculate the mean time to absorption for the case $n + 1$:

$$k_{b,R_1} = 1 + \lambda k_{b+e,R_1} + (1 - \lambda)k_{b,R_1},$$

$$\text{for } E_d - (n + 1)e \leq b < E_d - ne, \quad (5.20)$$

which reduces to $k_{b,R_1} = \frac{n+1}{\lambda}$ for $E_d - (n + 1) \cdot e \leq b < E_d - n \cdot e$.

Thus, the lemma holds by induction. □

We will use Lemma 5.1 to show that the optimal policy minimizing the mean time to absorption *does not* need to split the incoming RF signal. In order to show this, let us

define two tail policies $\pi_t^i = (a_i, \pi_{t+1})$, $i = \text{split}, \text{no-split}$ taking different actions a_i , in the current slot, but following the same set of actions, π_{t+1} afterwards⁶. Let policy $\pi_t^{\text{split}} = (\rho, \pi_{t+1})$ be a tail policy that always splits the incoming RF energy, i.e., $0 < \rho < 1$, except when $B(t) < 1$ or $I(t) = R_1$, when it only harvests energy. Assume that the state of the system is (b, m) at time slot t . Then, the mean time to absorption for tail policy π_t^{split} is:

$$k_{b,m}^{\pi^{\text{split}}} = 1 + \lambda k_{b-1+\rho e, m+R^H(\rho)} + (1 - \lambda) k_{b-1, m+R^L(\rho)}, \quad (5.21)$$

where $k_{x,y}$ is the mean time to absorption of policy π_{t+1} beginning at state (x, y) . Note that with probability λ the channel is in GOOD state, and thus, $\rho \cdot e$ units of energy is harvested⁷. However, one unit of energy is spent by operating the transceiver to accumulate $R^H(\rho)$ bits of mutual information. Meanwhile, with probability $1 - \lambda$ the channel is in BAD state, and no energy is harvested, but the transceiver still consumes one unit of energy to accumulate $R^L(\rho)$ bits of mutual information.

Under tail policy $\pi_t^{\text{no-split}}$ the RF signal is never split at time slot t , but rather, it is completely used for mutual information accumulation except when $B(t) < 1$ or $I(t) = R_1$ when it harvests energy only. In a similar way as before, we may calculate $k_{b,m}^{\pi^{\text{no-split}}}$ as follows:

$$k_{b,m}^{\pi^{\text{no-split}}} = 1 + \lambda k_{b-1, R_1} + (1 - \lambda) k_{b-1, m+R_0}. \quad (5.22)$$

Theorem 5.1. *Policy $\pi_t^{\text{no-split}}$ in (5.22) achieves an expected number of re-transmission that is never worse than that of policy π_t^{split} in (19), i.e., $k_{b,m}^{\pi^{\text{no-split}}} \leq k_{b,m}^{\pi^{\text{split}}}$ for every $b = 0, 1, \dots$ and $m = 0, 1, \dots, R_1$.*

Proof. Assume that at time slot t the system is at state (b, m) . Consider policy π^{split} which always chooses $0 < \rho < 1$. Hence, it follows that $R^H(\rho) < R_1$, $R^L(\rho) < R_0$ and, from (5.10), we have $I(t) \leq R_1$. Also, it is easy to verify that for any b , we have

⁶Note that (a_i, π_{t+1}) defines a tail policy obtained by concatenating action a_i in the current slot with tail policy π_{t+1} .

⁷We assume that the energy harvesting circuit is generating energy linearly proportional to the energy of the incoming RF signal.

$k_{b,m_1} \leq k_{b,m_2}$ whenever $m_1 \geq m_2$. Thus, a lower bound on $k_{b,m}^{\pi^{split}}$ in (5.21) can be established as,

$$k_{b,m}^{\pi^{split}} \geq 1 + \lambda k_{b-1+\rho e, R_1} + (1 - \lambda) k_{b-1, m+R_0}. \quad (5.23)$$

Furthermore, since $b-1 < b-1+\rho \cdot e < b-1+e$, from Lemma 5.1, we know that $k_{b-1+\rho e, R_1} = k_{b-1, R_1}$. Hence, the lower bound in (5.23) is exactly the same as $k_{b,m}^{\pi^{no-split}}$ given in (5.22), i.e., $k_{b,m}^{\pi^{no-split}} \leq k_{b,m}^{\pi^{split}}$. \square

Theorem 5.1 proves that a no-splitting policy can achieve the minimum number of re-transmissions. Hence, in the latter part of the chapter, we focus on characterizing the optimal no-splitting policy by determining the scheduling decision between EH or ID for each state of the MC. Therefore, the state space of the discrete MC associated with the optimal no-splitting policy is $b = 0, 1, \dots, \infty$, and $m = 0, 1, \dots, R_1$ ⁸.

Remark 5.1. *Theorem 5.1 plays an important role in simplifying the original problem by reducing the two dimensional uncountable state MDP with continuous action space into a two dimensional countable state MDP with binary decision space. This significantly reduces the complexity of numerical methods such as VIA. However, as we shall see in Section 5.4, the absorbing MC framework helps prove the optimality of a class of simple-to-implement algorithms that is more suitable for resource-deficient EH devices.*

Since the class of policies that we are interested in does not observe the channel, but make a decision based only on (b, m) , the time of the decision is irrelevant. Hence, given (b, m) , time t and $t+1$ are stochastically identical. Therefore, in the rest of the chapter we will omit the time index and optimize the scheduling decisions for any given state (b, m) . Define π^* as the optimal policy minimizing the mean time to absorption beginning at any given state (b, m) . Let $k_{b,m}^{\pi^*}$ be the minimum mean time to absorption obtained by policy π^* ⁹. Define the tail policy $\pi^i(b, m) = (i, \pi^*(\acute{b}, \acute{m}))$, $i = 0, 1$ such that it chooses $\rho = i$ at state (b, m) but follows policy π^* after transitioning into the new state (\acute{b}, \acute{m}) . Let $k_{b,m}^{\pi^i}$ be the mean time to absorption of policy $\pi^i(b, m)$, $i = 0, 1$. We can characterize $k_{b,m}^{\pi^0}$ and $k_{b,m}^{\pi^1}$ as follows:

⁸Note that in the original problem the states of the MC are $[0, \infty) \times [0, R_1]$.

⁹Note that the mean time to absorption calculated in Lemma 5.1 is the smallest possible value, i.e., $k_{b,R_1}^{\pi^*} = k_{b,R_1}$ for $b = 0, 1, \dots, E_d - 1$.

$$k_{b,m}^{\pi^0} = 1 + \lambda k_{b-1,R_1}^{\pi^*} + (1 - \lambda)k_{b-1,m+1}^{\pi^*}, \quad (5.24)$$

$$\begin{aligned} k_{b,m}^{\pi^1} &= 1 + \lambda k_{b+e,m}^{\pi^*} + (1 - \lambda)k_{b,m}^{\pi^1} \\ &= \frac{1}{\lambda} + k_{b+e,m}^{\pi^*}. \end{aligned} \quad (5.25)$$

Note that by evaluating and then comparing the values of $k_{b,m}^{\pi^0}$ and $k_{b,m}^{\pi^1}$, at all possible states (b, m) for $b = 0, 1, \dots, \infty$, and $m = 0, 1, \dots, R_1$, one can obtain the optimal policy π^* and its associated $k_{b,m}^{\pi^*}$.

Theorem 5.2. For states $(b, m) = (E_d + j, R_1 - j)$ for $j = 1, 2, \dots, R_1$, the minimum mean time to absorption, $k_{b,m}^{\pi^*}$ is given by

$$k_{E_d+j,R_1-j}^{\pi^*} = k_{E_d+j,R_1-j}^{\pi^0} = \sum_{i=1}^j (1 - \lambda)^{i-1}. \quad (5.26)$$

Furthermore, $k_{b,R_1-j}^{\pi^*} = k_{E_d+j,R_1-j}^{\pi^0}$ for $b = E_d + j + 1, E_d + j + 2, \dots$

Proof. The proof is by induction. For the base case consider the initial case when $j = 1$ so that $b = E_d + 1, E_d + 2, \dots$ and $m = R_1 - 1$. We have

$$k_{E_d+1,R_1-1}^{\pi^0} = 1 + \lambda k_{E_d,R_1}^{\pi^*} + (1 - \lambda)k_{E_d,R_1}^{\pi^*} = 1, \quad (5.27)$$

$$\begin{aligned} k_{E_d+1,R_1-1}^{\pi^1} &= \frac{1}{\lambda} + k_{E_d+e+1,m}^{\pi^*} \\ &> k_{E_d+1,R_1-1}^{\pi^0}. \end{aligned} \quad (5.28)$$

Note that when $b = E_d + 1, E_d + 2, \dots$, by choosing $\rho = 0$, regardless of the channel state, the next state, $(b - 1, R_1)$, is an absorbing state so $k_{b,R_1-1}^{\pi^0} = 1$. Thus, the lemma holds for $j = 1$. In the induction step assume that the theorem holds for $j = n - 1$, i.e., $k_{b,R_1-n+1}^{\pi^*} = k_{E_d+n-1,R_1-n+1}^{\pi^0} = \sum_{i=1}^{n-1} (1 - \lambda)^{i-1}$ for $b = E_d + n - 1, E_d + n, \dots$. Now, we prove that the claim is also true for $j = n$.

$$\begin{aligned}
k_{E_d+n, R_1-n}^{\pi^0} &= 1 + (1 - \lambda)k_{E_d+n-1, R_1-n+1}^{\pi^*} \\
&= 1 + \sum_{i=1}^{n-1} (1 - \lambda)^i \\
&= \sum_{i=1}^n (1 - \lambda)^{i-1}, \tag{5.29}
\end{aligned}$$

$$\begin{aligned}
k_{E_d+n, R_1-n}^{\pi^1} &= \frac{1}{\lambda} + k_{E_d+n+e, R_1-n}^{\pi^*} \\
&> \frac{1}{\lambda} + k_{E_d+n+e, R_1-n+1}^{\pi^*} \\
&= \frac{1}{\lambda} + k_{E_d+n-1, R_1-n+1}^{\pi^*} \\
&= \frac{1}{\lambda} + \frac{1 - (1 - \lambda)^{n-1}}{\lambda} \tag{5.30}
\end{aligned}$$

Furthermore,

$$\begin{aligned}
k_{E_d+n, R_1-n}^{\pi^0} &= \frac{1 - (1 - \lambda)^n}{\lambda} \\
&= 1 + (1 - \lambda) \frac{1 - (1 - \lambda)^{n-1}}{\lambda} < k_{E_d+n, R_1-n}^{\pi^1} \tag{5.31}
\end{aligned}$$

For the last part of the proof, we need to show that $k_{b, R_1-n}^{\pi^*} = k_{E_d+n, R_1-n}^{\pi^0}$ for $b = E_d + n + 1, E_d + n + 2, \dots$. We may write:

$$\begin{aligned}
k_{b, R_1-n}^{\pi^*} &= 1 + (1 - \lambda)k_{b-1, R_1-n+1}^{\pi^0} \\
&= 1 + (1 - \lambda)k_{E_d+n-1, R_1-n+1}^{\pi^0} = k_{E_d+n, R_1-n}^{\pi^0} \tag{5.32}
\end{aligned}$$

□

Theorem 5.2 states that if the receiver has $R_1 - n$ bits of mutual information accumulated and more than $E_d + n$ units of energy in its battery, then it should use the incoming RF signal for mutual information accumulation only. For any given state (b, m) , we exploit Lemma 5.1 and Theorem 5.2 to develop Algorithm 1 for calculating the minimum mean time to absorption, $k_{b, m}^{\pi^*}$, and the optimal scheduling decision at every state.

The idea of Algorithm 1 is to use Lemma 5.1 and Theorem 5.2 as boundary condi-

tions and to recursively calculate the mean time to absorption $k_{b,m}^{\pi^0}$ and $k_{b,m}^{\pi^1}$ starting from $(b, m) = (E_d, R_1 - 1)$. Note that $k_{E_d, R_1 - 1}^{\pi^0}$ and $k_{E_d, R_1 - 1}^{\pi^1}$ depend on the values of $k_{E_d - 1, R_1}^{\pi^*}$ and $k_{E_d + 1, R_1 - 1}^{\pi^*}$, which are obtained in the initialization step, and the optimal scheduling decision at state $(E_d, R_1 - 1)$ is given by $\arg \min_{i \in \{0, 1\}} k_{b,m}^{\pi^i}$. The procedure in Algorithm 1 continues by decrementing the value of b by 1 at each iteration, until $b = 0$ at which time the value of m is decremented by 1, b is initialized to $E_d + n$ and the procedure is repeated. The aforementioned order of spanning the states of the MC ensures that at each iteration the mean time to absorption can be calculated from the values determined in the previous iterations. We have shown in Theorem 5.3, that Algorithm 1 minimizes the expected number of re-transmissions starting from any state (b, m) .

Theorem 5.3. *Algorithm 1 calculates the minimum mean time to absorption starting from an arbitrary state (b, m) for which $b = 0, \dots, \infty$ and $m = 0, 1, \dots, R_1$.*

Proof. In Lemma 5.1, we characterized the minimum mean time to absorption for all states (b, R_1) , for $b = 0, \dots, E_d - 1$. Also, in Theorem 5.2, we characterized the minimum mean time to absorption for states, $(b, R_1 - j)$ where, $b = E_d + j, E_d + j + 1, \dots$ and $j = 1, \dots, R_1$. Furthermore, Theorem 5.1 proves that at any state (b, m) , the receiver should either choose to harvest energy or accumulate mutual information. Note that the iterations are ordered in Algorithm 1 (line 4-8) so that $k_{b,m}^{\pi^0}$ and $k_{b,m}^{\pi^1}$ only depend on $k_{b-1, R_1}^{\pi^*}$, $k_{b-1, m+1}^{\pi^*}$, and $k_{b+1, m}^{\pi^*}$ which are obtained at the previous rounds of the algorithm. □

Algorithm 1 Calculating the minimum mean time to absorption for an i.i.d. channel

- 1: Initialize $k_{b, R_1}^{\pi^*}$ for $b = 0, \dots, E_d - 1$ using Lemma 5.1.
 - 2: Initialize $k_{E_d + j, R_1 - j}^{\pi^*}$ for $j = 1, \dots, R_1$ using Theorem 5.2.
 - 3: $n \leftarrow 0$
 - 4: **for** $m = R_1 - 1 : 0$ **do**
 - 5: **for** $b = E_d + n : 0$ **do**
 - 6: Calculate $k_{b,m}^{\pi^0}$, $k_{b,m}^{\pi^1}$ from (5.24) and (5.25), respectively.
 - 7: $k_{b,m}^{\pi^*} = \min \left(k_{b,m}^{\pi^0}, k_{b,m}^{\pi^1} \right)$.
 - 8: $\rho^*(b, m) = \arg \min_i k_{b,m}^{\pi^i}$ for $i = 0, 1$
 - 9: $n \leftarrow n + 1$
-

5.4 Optimal Class of Policies for i.i.d. Channels

In the previous section, we have given a procedure to obtain the optimal scheduling decision of a TS policy, once we established that there exists a TS policy achieving the minimum number of re-transmissions. In this section, we formally determine the optimal class of scheduling policies minimizing the number of re-transmissions until successful decoding. In the following, we obtain our analytical results for $e = 1$ and $R_2 = 1$. However, our analysis holds in general for different values of e and R_2 , which is demonstrated by the numerical results presented in Section 5.6. The following theorem states that once the battery has sufficient charge to decode the packet, it is better to use the incoming RF signal only for information accumulation.

Theorem 5.4. *If $b = E_d + 1, E_d + 2, \dots$, the optimal decision is to accumulate mutual information, i.e., $\rho^*(b, m) = 0$ for $b = E_d + 1, E_d + 2, \dots$*

Proof. We need to show that $k_{E_d+j-i, R_1-j}^{\pi^0} < k_{E_d+j-i, R_1-j}^{\pi^1}$ for all $j = 1, \dots, R_1$ and $i = 0, 1, \dots, j - 1$. The proof is by induction. For the base case, we need to show that the theorem holds for $i = 0$ and all $j = 1, \dots, R_1$. We know from Theorem 5.2 that $k_{E_d+j, R_1-j}^{\pi^0} < k_{E_d+j, R_1-j}^{\pi^1}$ and, hence, the Theorem is true for $i = 0$ and all $j = 1, \dots, R_1$. Next, in the induction step, assume that the theorem is true for $i = n$ and all $j = 1, \dots, R_1$ i.e., $k_{E_d+j-n, R_1-j}^{\pi^0} < k_{E_d+j-n, R_1-j}^{\pi^1}$. We need to show that the theorem also holds for $i = n + 1$ and all $j = 1, \dots, R_1$.

$$\begin{aligned} k_{E_d+j-(n+1), R_1-j}^{\pi^1} &= \frac{1}{\lambda} + k_{E_d+j-n, R_1-j}^{\pi^*} \\ &= \frac{1}{\lambda} + k_{E_d+j-n, R_1-j}^{\pi^0} \end{aligned} \quad (5.33a)$$

$$= \frac{1}{\lambda} + 1 + (1 - \lambda)k_{E_d+(j-1)-n, R_1-(j-1)}^{\pi^*}, \quad (5.33b)$$

where (5.33a) follows because of the induction hypothesis, i.e., $k_{E_d+j-n, R_1-j}^{\pi^0} < k_{E_d+j-n, R_1-j}^{\pi^1}$.

Also,

$$\begin{aligned} k_{E_d+j-(n+1),R_1-j}^{\pi^0} &= 1 + (1 - \lambda)k_{E_d+j-(n+1)-1,R_1-j+1}^{\pi^*} \\ &\leq 1 + (1 - \lambda)k_{E_d+j-(n+1)-1,R_1-j+1}^{\pi^1} \end{aligned} \quad (5.34a)$$

$$= \frac{1}{\lambda} + (1 - \lambda)k_{E_d+(j-1)-n,R_1-(j-1)}^{\pi^*}, \quad (5.34b)$$

where (5.34a) is due to $k_{x,y}^{\pi^*} = \min(k_{x,y}^{\pi^0}, k_{x,y}^{\pi^1})$. From (5.33b) and (5.34b), we have $k_{E_d+j-(n+1),R_1-j}^{\pi^0} \leq k_{E_d+j-(n+1),R_1-j}^{\pi^1} - 1$, which in turn proves the following inequality:

$$k_{E_d+j-(n+1),R_1-j}^{\pi^0} < k_{E_d+j-(n+1),R_1-j}^{\pi^1}. \quad (5.35)$$

Since the theorem is also true for $i = n + 1$ and all j , by induction, the theorem holds for all $j = 1, \dots, R_1$ and $i = 0, 1, \dots, j - 1$. \square

Since it is optimal to accumulate mutual information whenever $b = E_d + 1, E_d + 2, \dots$ (i.e., $\rho^*(b, m) = 0$ for all $m = 0, \dots, R_1$ and $b = E_d + 1, E_d + 2, \dots$), we can calculate $k_{b,m}^{\pi^*} = k_{b,m}^{\pi^0}$ for those states. Using (5.24), we have:

$$k_{E_d+i,R_1-j-i}^{\pi^*} = 1 + (1 - \lambda)k_{E_d+i-1,R_1-j-i+1}^{\pi^*} \quad \text{for } j = 1, \dots, R_1 - 1, \quad i = 1, \dots, R_1 - j \quad (5.36)$$

Using the recursion in (5.36), it is possible to show that:

$$k_{E_d+i,R_1-j-i}^{\pi^*} = \frac{1 - (1 - \lambda)^i}{\lambda} + (1 - \lambda)^i k_{E_d,R_1-j}^{\pi^*} \quad (5.37)$$

Note that in order to calculate the minimum mean time to absorption using (5.37), one needs to know the values of $k_{E_d,R_1-j}^{\pi^*}$, $j = 1, \dots, R_1 - 1$. However, from (5.24), we know that $k_{E_d,R_1-j}^{\pi^0}$ depends on the unknown values of $k_{E_d-1,R_1-j+1}^{\pi^*}$, so it is not possible to compare $k_{E_d,R_1-j}^{\pi^0}$ and $k_{E_d,R_1-j}^{\pi^1}$ just yet.

Hence, let us first calculate $k_{E_d,R_1-1}^{\pi^0}$ and $k_{E_d,R_1-1}^{\pi^1}$. Using Lemma 5.1 and 5.2, as

well as (5.24) and (5.25), it can be seen that

$$k_{E_d, R_1-1}^{\pi^0} = k_{E_d, R_1-1}^{\pi^1} = 1 + \frac{1}{\lambda}. \quad (5.38)$$

Note that whenever the receiver is at state $(E_d, R_1 - 1)$, the decision to choose either energy harvesting or mutual information accumulation, does not alter the mean time to absorption at that specific state. The following theorem generalizes this observation to other states as well.

Theorem 5.5. *At state (b, m) where $b = 1, 2, \dots, E_d$, and $m = 0, 1, \dots, R_1 - 1$, any time switching decision is optimal.*

Proof. We have to show that $k_{i, R_1-j}^{\pi^0} = k_{i, R_1-j}^{\pi^1}$ for $i = 1, \dots, E_d$ and $j = 1, \dots, R_1$. The outline of the induction proof is as follows:

- For the base case we show that $k_{i, R_1-1}^{\pi^0} = k_{i, R_1-1}^{\pi^1}$ for all $i = 1, \dots, E_d$.
- In the induction step, we assume the the lemma is true for $j = n$ and all $i = 1, \dots, E_d$.
- Using the induction step, we prove that the theorem also holds for $j = n + 1$ and all $i = 1, \dots, E_d$.

Let us consider the base case of $j = 1$. From (5.38), we know that the theorem holds for $i = E_d$, i.e., $k_{E_d, R_1-1}^{\pi^0} = k_{E_d, R_1-1}^{\pi^1}$. Assume that $k_{i, R_1-1}^{\pi^0} = k_{i, R_1-1}^{\pi^1}$ and calculate:

$$\begin{aligned} k_{i-1, R_1-1}^{\pi^1} &= \frac{1}{\lambda} + k_{i, R_1-1}^{\pi^*} \\ &= \frac{1}{\lambda} + k_{i, R_1-1}^{\pi^0} \\ &= \frac{1}{\lambda} + 1 + k_{i-1, R_1}^{\pi^*} \\ &= 1 + \frac{E_d - i + 2}{\lambda} \end{aligned} \quad (5.39)$$

$$k_{i-1, R_1-1}^{\pi^0} = 1 + k_{i-2, R_1}^{\pi^*} = 1 + \frac{E_d - i + 2}{\lambda} \quad (5.40)$$

Hence, $k_{i-1, R_1-1}^{\pi^1} = k_{i-1, R_1-1}^{\pi^0}$ and the theorem holds for $j = 1$ and all $i = 1, \dots, E_d$. Next, for the induction step assume that the theorem is true for $j = n$ and all values of $i = 1, \dots, E_d$, i.e., $k_{i, R_1-n}^{\pi^0} = k_{i, R_1-n}^{\pi^1}$. To show that the theorem is also true for $j = n + 1$

and all values of $i = 1, \dots, E_d$, we have to start by first showing that the theorem holds for the state $(E_d, n+1)$ and work our way to show that it also holds for all states $(i, n+1)$. Let us calculate and compare values of $k_{E_d, R_1-(n+1)}^{\pi^0}$ and $k_{E_d, R_1-(n+1)}^{\pi^1}$.

$$\begin{aligned} k_{E_d, R_1-(n+1)}^{\pi^1} &= \frac{1}{\lambda} + k_{E_d+1, R_1-(n+1)}^{\pi^*} \\ &= \frac{1}{\lambda} + k_{E_d+1, R_1-(n+1)}^{\pi^0} \\ &= \frac{1}{\lambda} + 1 + (1 - \lambda)k_{E_d, R_1-n}^{\pi^*} \end{aligned} \quad (5.41)$$

$$k_{E_d, R_1-(n+1)}^{\pi^0} = 1 + (1 - \lambda)k_{E_d-1, R_1-n}^{\pi^*} + \lambda k_{E_d-1, R_1}^{\pi^*} \quad (5.42)$$

$$\begin{aligned} &= 1 + (1 - \lambda)k_{E_d-1, R_1-n}^{\pi^1} \\ &= 1 + (1 - \lambda)\left(\frac{1}{\lambda} + k_{E_d, R_1-n}^{\pi^*}\right) + 1 \\ &= 1 + \frac{1}{\lambda} + (1 - \lambda)k_{E_d, R_1-n}^{\pi^*}. \end{aligned} \quad (5.43)$$

Thus, $k_{E_d, R_1-(n+1)}^{\pi^0} = k_{E_d, R_1-(n+1)}^{\pi^1}$. Next, we assume that $k_{i, R_1-(n+1)}^{\pi^0} = k_{i, R_1-(n+1)}^{\pi^1}$ and prove that $k_{i-1, R_1-(n+1)}^{\pi^0} = k_{i-1, R_1-(n+1)}^{\pi^1}$. We have:

$$\begin{aligned} k_{i-1, R_1-(n+1)}^{\pi^0} &= 1 + (1 - \lambda)k_{i-2, R_1-n}^{\pi^*} + \lambda k_{i-2, R_1}^{\pi^*} \\ &= 1 + (1 - \lambda)k_{i-2, R_1-n}^{\pi^1} + E_d - i + 2 \\ &= 1 + (1 - \lambda)\left(\frac{1}{\lambda} + k_{i-1, R_1-n}^{\pi^*}\right) + E_d - i + 2 \\ &= \frac{1}{\lambda} + E_d - i + 2 + (1 - \lambda)k_{i-1, R_1-n}^{\pi^*}. \end{aligned} \quad (5.44)$$

$$\begin{aligned} k_{i-1, R_1-(n+1)}^{\pi^1} &= \frac{1}{\lambda} + k_{i, R_1-(n+1)}^{\pi^*} \\ &= \frac{1}{\lambda} + k_{i, R_1-(n+1)}^{\pi^0} \\ &= \frac{1}{\lambda} + 1 + (1 - \lambda)k_{i-1, R_1-n}^{\pi^*} + \lambda k_{i-1, R_1}^{\pi^*} \\ &= \frac{1}{\lambda} + 1 + (1 - \lambda)k_{i-1, R_1-n}^{\pi^*} + E_d - i + 1 = k_{i-1, R_1-(n+1)}^{\pi^0}. \end{aligned} \quad (5.45)$$

Hence, the theorem holds for $j = n+1$ and all $i = 1, \dots, E_d$. Therefore, the theorem is true by induction. \square

Theorem 5.5, in essence, proves that there is no unique optimal policy. Instead, there exists a family of optimal policies achieving the minimum mean time to absorption. We summarize our findings so far in the following theorem by formally characterizing the

family of optimal policies.

Theorem 5.6. *Optimal policy, π^* , satisfies the following properties.*

1. *If $b = 0$ or $m = R_1$, it chooses $\rho = 1$.*
2. *If $b = E_d + 1, E_d + 2, \dots$, it chooses $\rho = 0$.*
3. *If $b = 1, 2, \dots, E_d$ and $m = 0, 1, \dots, R_1 - 1$, chooses either $\rho = 0$ or $\rho = 1$.*

Proof. The proof of the theorem is straightforward and proceeds as follows:

1. When $b = 0$, the receiver has no energy to activate the RF transceiver and should first recharge its battery. When $m = R_1$, the receiver collected sufficient mutual information to decode, but needs energy to perform the decoding operation. Hence, it harvests energy.
2. This part of the theorem is proven in Theorems 5.2 and 5.4.
3. Theorem 5.5 states that whenever $b = 1, 2, \dots, E_d$, and $m = 0, 1, \dots, R_1 - 1$, then $k_{b,m}^{\pi^0} = k_{b,m}^{\pi^1}$. Consider a policy β which satisfies part 1 and 2 of the theorem. Whenever $b = 1, 2, \dots, E_d$ and $m = 0, 1, \dots, R_1 - 1$, the policy chooses $\rho = 0$ with probability p . The mean time to absorption of policy β , $k_{b,m}^\beta$ can be calculated as follows

$$k_{b,m}^\beta = pk_{b,m}^{\pi^0} + (1-p)k_{b,m}^{\pi^1} = k_{b,m}^{\pi^0} = k_{b,m}^{\pi^1} \quad (5.46)$$

□

Simple examples of such optimal policies that belong to the optimal family of policies characterized in Theorem 5.6, are:

- **Battery First (BF):** the receiver harvests energy until it acquires E_d units of energy and then starts accumulating the mutual information.
- **Information First (IF):** the receiver always accumulates mutual information unless $b = 0$ or $m = R_1$.
- **Coin Toss (CT):** the receiver harvests energy when $b = 0$ or $m = R_1$, while it accumulates mutual information when $b = E_d + 1, E_d + 2, \dots$. Otherwise, it tosses a fair coin to choose between harvesting energy or accumulating mutual information.

5.5 Expected Number of Re-Transmissions for a Correlated Channel

In many wireless systems, the wireless channel cannot be modeled as an i.i.d. channel. In this section, we investigate optimal scheduling policies under a time-correlated channel model. Our analysis for a correlated channel follows a similar approach to our analysis for i.i.d. channels. However, due to correlation between the subsequent channel states, the receiver can improve its decision by incorporating its knowledge of the current state. Let the transition probabilities of the channel states be $\mathbb{P}[G_t = 1|G_{t-1} = 1] = \lambda_1$ and $\mathbb{P}[G_t = 1|G_{t-1} = 0] = \lambda_0$. Note that due to time correlation, the previous state of the channel provides information about the current channel state to the receiver. Hence, although once again we model the system as a MC, this time the state space of MC is extended where the states are (b, m, G) with G being the previous state of the channel¹⁰. The resulting MC is still an absorbing MC, and the mean time to absorption is equivalent to the minimum expected number of re-transmissions until successful decoding. Define π^* as the optimal policy minimizing the mean time to absorption at any given state (b, m, G) . Let $k_{b,m,G}^{\pi^*}$ be the mean time to absorption obtained by policy π^* at state (b, m, G) .

Lemma 5.2. *For any $E_d - i \cdot e \leq b < E_d - (i - 1) \cdot e$ such that $i = 1, \dots, E_d$, and given that $m = R_1$, the minimum mean time to absorption is given by*

$$k_{b,R_1,1}^{\pi^*} = i \frac{1 + \lambda_0 - \lambda_1}{\lambda_0}, \quad i = 1, \dots, E_d, \quad (5.47)$$

$$k_{b,R_1,0}^{\pi^*} = \frac{1}{\lambda_0} + (i - 1) \frac{1 + \lambda_0 - \lambda_1}{\lambda_0}, \quad i = 1, \dots, E_d. \quad (5.48)$$

Proof. The proof is by induction. Let us consider $i = 1$ as the base case such that $E_d - e \leq b < E_d$. Note that since $m = R_1$, the optimal decision is to harvest energy, i.e., $\rho = 1$.

¹⁰Note that the receiver becomes aware of the channel state after it decides to sample the incoming RF signal.

We have:

$$k_{b,R_1,0}^{\pi^*} = 1 + \lambda_0 k_{b+e,R_1,1}^{\pi^*} + (1 - \lambda_0) k_{b,R_1,0}^{\pi^*} = \frac{1}{\lambda_0}, \quad (5.49)$$

$$k_{b,R_1,1}^{\pi^*} = 1 + \lambda_1 k_{b+e,R_1,1}^{\pi^*} + (1 - \lambda_1) k_{b,R_1,0}^{\pi^*} = \frac{1 + \lambda_0 - \lambda_1}{\lambda_0}. \quad (5.50)$$

Hence lemma holds for $i = 1$. Next, for induction step assume that the lemma is true for $i = n$, i.e., $k_{b,R_1,0}^{\pi^*} = \frac{1}{\lambda_0} + (n - 1) \frac{1 + \lambda_0 - \lambda_1}{\lambda_0}$ and $k_{b,R_1,1}^{\pi^*} = n \frac{1 + \lambda_0 - \lambda_1}{\lambda_0}$ for $E_d - n \cdot e \leq b < E_d - (n - 1) \cdot e$. Let us consider the case $n + 1$:

$$\begin{aligned} k_{b,R_1,0}^{\pi^*} &= 1 + \lambda_0 k_{b+e,R_1,1}^{\pi^*} + (1 - \lambda_0) k_{b,R_1,0}^{\pi^*} \\ &= \frac{1}{\lambda_0} + k_{b+e,R_1,1}^{\pi^*} \\ &= \frac{1}{\lambda_0} + n \frac{1 + \lambda_0 - \lambda_1}{\lambda_0}, \quad \text{for } E_d - (n + 1)e \leq b < E_d - ne, \end{aligned}$$

$$\begin{aligned} k_{b,R_1,1}^{\pi^*} &= 1 + \lambda_1 k_{b+e,R_1,1}^{\pi^*} + (1 - \lambda_1) k_{b,R_1,0}^{\pi^*} \\ &= 1 + \lambda_1 n \frac{1 + \lambda_0 - \lambda_1}{\lambda_0} + (1 - \lambda_1) \left(\frac{1}{\lambda_0} + n \frac{1 + \lambda_0 - \lambda_1}{\lambda_0} \right) \\ &= (n + 1) \frac{1 + \lambda_0 - \lambda_1}{\lambda_0}, \quad \text{for } E_d - (n + 1)e \leq b < E_d - ne. \end{aligned}$$

□

Similar to Theorem 5.1, by exploiting Lemma 5.2, we can prove that the optimal policy should either choose energy harvesting or information accumulation at any given state (b, m, G) . Therefore, MC associated with the optimal strategy has discrete states in which $b = 0, 1, \dots, \infty$, $m = 0, 1, \dots, R_1$ and $G = 0, 1$. Define the tail policy $\pi^i(b, m, G) = (i, \pi^*(\acute{b}, \acute{m}, \acute{G}))$, $i = 0, 1$ that chooses $\rho = i$ at state (b, m, G) but follows policy π^* after transitioning into the new state $(\acute{b}, \acute{m}, \acute{G})$. Let $k_{b,m,G}^{\pi^i}$ be the mean time to absorption of policy $\pi^i(b, m, G)$, $i = 0, 1$. We can calculate $k_{b,m,G}^{\pi^0}$ and $k_{b,m,G}^{\pi^1}$ as follows:

$$k_{b,m,0}^{\pi^0} = 1 + \lambda_0 k_{b-1,R_1,1}^{\pi^*} + (1 - \lambda_0) k_{b-1,m+1,0}^{\pi^*}, \quad (5.51)$$

$$k_{b,m,1}^{\pi^0} = 1 + \lambda_1 k_{b-1,R_1,1}^{\pi^*} + (1 - \lambda_1) k_{b-1,m+1,0}^{\pi^*}, \quad (5.52)$$

$$k_{b,m,0}^{\pi^1} = 1 + \lambda_0 k_{b+1,m,1}^{\pi^*} + (1 - \lambda_0) k_{b,m,0}^{\pi^1} = \frac{1}{\lambda_0} + k_{b+1,m,1}^{\pi^*}, \quad (5.53)$$

$$k_{b,m,1}^{\pi^1} = 1 + \lambda_1 k_{b+1,m,1}^{\pi^*} + (1 - \lambda_1) k_{b,m,0}^{\pi^*}. \quad (5.54)$$

Similar to the outline of the Theorem 5.2, in the following, we consider states $(b, m, G) = (E_d + j, R_1 - j, G)$ for $j = 1, \dots, R_1$ and derive the optimal strategy for those states.

Lemma 5.3. *The optimal strategy in states $(E_d + j, R_1 - j, G)$ for $j = 1, \dots, R_1$ and $G = 0, 1$ is to accumulate mutual information ($\rho^*(E_d + j, R_1 - j, G) = 0$) and also $k_{b,R_1-j,G}^{\pi^*} = k_{E_d+j,R_1-j,G}^{\pi^0}$ for $b = E_d + j + 1, E_d + j + 2, \dots$*

Proof. The proof is by induction. Let us first consider $(E_d + j, R_1 - j, 0)$. For $j = 1$ we have

$$k_{E_d+1,R_1-1,0}^{\pi^0} = 1 + (1 - \lambda_0) k_{E_d,R_1,0}^{\pi^*} = 1 \quad (5.55)$$

$$k_{E_d+1,R_1-1,0}^{\pi^1} = 1 + \lambda_0 k_{E_d+1+e,R_1-1,1}^{\pi^*} + (1 - \lambda_0) k_{E_d+1,R_1-1,0}^{\pi^*} > 1, \quad (5.56)$$

where it also follows that $k_{b,R_1-1,0}^{\pi^0} = 1 + (1 - \lambda_0) k_{b-1,R_1,0}^{\pi^*} = 1$ for $b = E_d + 1, E_d + 2, \dots$. Hence the theorem holds for $j = 1$. Let us assume that the theorem holds for $j = n - 1$ i.e., $k_{b,R_1-n+1,0}^{\pi^*} = k_{E_d+n-1,R_1-n+1,0}^{\pi^0}$ for $b = E_d + n, E_d + n + 1, \dots$. We have

$$\begin{aligned} k_{E_d+n,R_1-n,0}^{\pi^0} &= 1 + (1 - \lambda_0) k_{E_d+n-1,R_1-n+1,0}^{\pi^*} \\ &= 1 + (1 - \lambda_0) k_{E_d+n-1,R_1-n+1,0}^{\pi^0}, \end{aligned} \quad (5.57)$$

$$\begin{aligned} k_{E_d+n,R_1-n,0}^{\pi^1} &= 1 + \lambda_0 k_{E_d+n+e,R_1-n,1}^{\pi^*} + (1 - \lambda_0) k_{E_d+n,R_1-n,0}^{\pi^*} \\ &> 1 + (1 - \lambda_0) k_{E_d+n,R_1-n,0}^{\pi^*} \\ &\geq 1 + (1 - \lambda_0) k_{E_d+n,R_1-n+1,0}^{\pi^*} \\ &= 1 + (1 - \lambda_0) k_{E_d+n-1,R_1-n+1,0}^{\pi^*} = k_{E_d+n,R_1-n,0}^{\pi^0}. \end{aligned} \quad (5.58)$$

And we need to show that $k_{b,R_1-n,0}^{\pi^0} = k_{E_d+n,R_1-n,0}^{\pi^0}$ for $b = E_d + n + 1, E_d + n + 2, \dots$

We have

$$\begin{aligned} k_{b,R_1-n,0}^{\pi^0} &= 1 + (1 - \lambda_0)k_{b-1,R_1-n+1,0}^{\pi^*} \\ &= 1 + (1 - \lambda_0)k_{E_d+n-1,R_1-n+1,0}^{\pi^0} = k_{E_d+n,R_1-n,0}^{\pi^0}. \end{aligned} \quad (5.59)$$

Next we will prove the lemma for states $(E_d + j, R_1 - j, 1)$. Let us consider the base case $j = 1$.

$$k_{E_d+1,R_1-1,1}^{\pi^0} = 1 + (1 - \lambda_1)k_{E_d,R_1,0}^{\pi^*} = 1 \quad (5.60)$$

$$k_{E_d+1,R_1-1,1}^{\pi^1} = 1 + \lambda_1 k_{E_d+1+e,R_1-1,1}^{\pi^*} + (1 - \lambda_1)k_{E_d+1,R_1-1,0}^{\pi^*} > 1. \quad (5.61)$$

Also, $k_{b,R_1-1,1}^{\pi^0} = 1 + (1 - \lambda_1)k_{b-1,R_1,0}^{\pi^*} = 1$ for $b = E_d + 1, E_d + 2, \dots$. Hence the lemma holds for $j = 1$. Let us assume that the lemma holds for $j = n - 1$ i.e, $k_{b,R_1-n+1,1}^{\pi^0} = k_{E_d+n-1,R_1-n+1,0}^{\pi^0}$ for $b = E_d + n, E_d + n + 1, \dots$. We have

$$\begin{aligned} k_{E_d+n,R_1-n,1}^{\pi^1} &= 1 + \lambda_1 k_{E_d+n+e,R_1-n,1}^{\pi^*} + (1 - \lambda_1)k_{E_d+n,R_1-n,0}^{\pi^*} \\ &> 1 + (1 - \lambda_1)k_{E_d+n,R_1-n,0}^{\pi^*} \\ &\geq 1 + (1 - \lambda_1)k_{E_d+n,R_1-n+1,0}^{\pi^*} \\ &= 1 + (1 - \lambda_1)k_{E_d+n-1,R_1-n+1,0}^{\pi^*} = k_{E_d+n,R_1-n,1}^{\pi^0}. \end{aligned} \quad (5.62)$$

Finally we conclude the proof by showing that $k_{b,R_1-n,1}^{\pi^0} = k_{E_d+n,R_1-n,1}^{\pi^0}$ for $b = E_d + n + 1, E_d + n + 2, \dots$. We have

$$\begin{aligned} k_{b,R_1-n,1}^{\pi^0} &= 1 + (1 - \lambda_1)k_{b-1,R_1-n+1,0}^{\pi^*} \\ &= 1 + (1 - \lambda_0)k_{E_d+n-1,R_1-n+1,0}^{\pi^0} = k_{E_d+n,R_1-n,1}^{\pi^0}. \end{aligned} \quad (5.63)$$

□

Now that we know the optimal policy for states $(E_d + j, R_1 - j, G)$, we can calculate the minimum mean time to absorption for those states as follows:

$$k_{E_d+j, R_1-j, 0}^{\pi^*} = k_{E_d+j, R_1-j, 0}^{\pi^0} = \sum_{i=1}^j (1 - \lambda_0)^{i-1}, \quad j = 1, \dots, R_1, \quad (5.64)$$

$$k_{E_d+j, R_1-j, 1}^{\pi^*} = k_{E_d+j, R_1-j, 1}^{\pi^0} = 1 + (1 - \lambda_1) \sum_{i=1}^{j-1} (1 - \lambda_0)^{i-1},$$

$$j = 2, \dots, R_1, \quad (5.65)$$

$$k_{E_d+j, R_1-j, 1}^{\pi^*} = 1, \quad j = 1. \quad (5.66)$$

Algorithm 2 calculates the $k_{b,m,G}^{\pi^*}$ and the corresponding ρ^* for any b , m , and G . Proving the optimality of Algorithm 2 is similar to the outline of the optimality proof of Algorithm 1 and hence it is omitted here. Note that the knowledge of the previous channel state, G , enables the receiver to be able to fully utilize the information yielded by the correlation. However, it also results in four coupled equations, (5.51)-(5.54), over numerous states which makes the analysis extremely hard. For this reason, we omit the full characterization of the structure of the optimal policy. Nevertheless, note that Algorithm 2 provides a recursive method to determine the optimal scheduling decisions for each state (b, m, G) . In fact, we use these optimal decisions in the numerical experiments discussed in Section 5.6 to calculate the minimum number of re-transmissions.

Algorithm 2 Calculating the minimum mean time to absorption for correlated channel

- 1: Initialize $k_{b, R_1, G}^{\pi^*}$ for $b = 0, \dots, E_d - 1$ using (5.47) and (5.48).
 - 2: Initialize $k_{E_d+j, R_1-j, G}^{\pi^*}$ for $j = 1, \dots, R_1$ using (5.64), (5.65) and (5.66).
 - 3: $n \leftarrow 0$
 - 4: **for** $m = R_1 - 1 : 0$ **do**
 - 5: **for** $b = E_d + n : 0$ **do**
 - 6: Calculate $k_{b,m,G}^{\pi^0}$ for $G = 0, 1$ using (5.51) and (5.52), respectively.
 - 7: Calculate $k_{b,m,G}^{\pi^1}$ for $G = 0, 1$ using (5.53) and (5.54), respectively.
 - 8: $k_{b,m,G}^{\pi^*} = \min \left(k_{b,m,G}^{\pi^0}, k_{b,m,G}^{\pi^1} \right)$.
 - 9: $\rho^*(b, m, G) = \arg \min_i k_{b,m,G}^{\pi^i}$ for $i = 0, 1$
 - 10: $n \leftarrow n + 1$
-

5.6 Numerical Results

In this section, we provide numerical evidence to support the analytical results established in the chapter. VIA is a standard tool for solving the bellman equations in (5.14).

However, VIA iterates for numerous passes over each state, which is increasing in β , before converging to a steady solution, whereas Algorithm 1 and 2 needs a single iteration. Moreover, VIA achieves exactly the same performance as Algorithm 1 and 2. Thus, we omit the results obtained by VIA.

We will divide our attention to validate the optimal policy for i.i.d. and correlated channel models. Although the framework discussed is sufficiently general to determine the number of re-transmissions starting from any residual battery level, in this section for the clarity of presentation, we consider that the initial battery level is zero. We use a simple ARQ mechanism as a baseline for understanding the performance merits of the HARQ mechanism. In the following, we formally define the simple ARQ scheme for i.i.d. and correlated channels.

5.6.1 Simple ARQ

In simple ARQ, the packet is transmitted successfully whenever the channel is in a GOOD state and the receiver has sufficient energy to decode the packet. Otherwise, the receiver drops the packet and awaits re-transmissions. When the receiver employs simple ARQ, before any decoding attempt, it has to make sure that its battery has at least $E_d + 1$ units of energy. Otherwise, after consuming 1 unit of energy for sampling, it will not have sufficient energy to decode the data packet and it will drop the packet. It is easy to prove that the optimal simple ARQ policy minimizing the mean time to absorption first harvests $E_d + 1$ units of energy and then attempts decoding. If the decoding attempt is not successful, it harvests energy until its battery state reaches $E_d + 1$ units again before attempting to decode.

5.6.2 i.i.d. Channel States

In this section, we evaluate the minimum mean time to absorption obtained from Algorithm 1, and compare it to that of the following three simple policies. The studied policies are as follows: *i*) Battery First (BF), *ii*) Information First (IF), and *iii*) Coin Toss (CT). Also, we compare the performance of the receiver equipped with HARQ mechanism with the case of a receiver equipped with simple ARQ mechanism. We determine the mean number of re-transmissions by Monte Carlo simulations, and compare them with that of analytical calculation described in Algorithm 1. Note that Monte Carlo simulations

Table 5.1: Mean time to absorption for $R_1 = 10$, $e = 1$ and $E_d = 5$ vs. R_0

	$R_0 = 1$	$R_0 = 2$	$R_0 = 3$	$R_0 = 4$	$R_0 = 5$	$R_0 = 6$	$R_0 = 7$	$R_0 = 8$	$R_0 = 9$
Optimal analytical	15.9941	15.8125	15.6250	15.2500	14.5000	14.5000	14.5000	14.5000	14.5000
Optimal Monte-Carlo	15.9910	15.8103	15.6235	15.2490	14.4992	14.5001	14.4998	14.5000	14.4983
BF	15.9938	15.8116	15.6259	15.2504	14.4999	14.4995	14.4993	14.5012	14.5000
IF	15.9941	15.8143	15.6245	15.2508	14.4987	14.4997	14.5017	14.4989	14.5003
CT	15.9966	15.8140	15.6266	15.2491	14.5020	14.5007	14.5009	14.4984	14.5001
Simple ARQ	15.9992	15.9992	15.9992	16.0006	16.0007	15.9995	15.9996	16.0008	16.0011

provide only *sample* mean time which is a random variable. The mean of this random variable is equal to the mean time to absorption and its variance decreases with the number of samples and becomes zero only if the number of iterations go to infinity. Hence, we expect to see small differences between the results obtained by the Monte Carlo simulations and analytical results, which is the reason why some policies have slightly smaller mean time to absorption than the optimal analytical value.

Table 5.1 summarizes the mean time to absorption for $R_1 = 10$, $e = 1$, $\lambda = 0.5$ and $E_d = 5$ with respect to R_0 associated with different policies. For IF, BF, CT and simple ARQ policies, we run Monte Carlo simulations for 10^7 iterations and evaluate the sample mean. It can be seen from Table 5.1 that all policies have almost the same performance. This observation confirms our major finding that the optimal policy achieving the minimum mean time to absorption is not unique.

The effect of quality of the channel on the mean time to absorption for $R_0 = 5$, $R_1 = 10$, $E_d = 5$ and $e = 2$ with respect to λ is summarized in Table 5.2. As expected, it can be seen that the mean time to absorption decreases as the channel quality improves. Also, the performance gap between the HARQ and simple ARQ mechanism becomes smaller as the channel quality improves. This is because as the channel quality improves, the probability of harvesting energy and accumulating R_1 bits of mutual information also increases. Finally, the mean time to absorption for $R_0 = 5$, $R_1 = 10$, $E_d = 10$ and $\lambda = 0.3$ with respect to e is summarized in Table 5.3. We observe that the mean time to absorption is approximately the same for all policies and it is decreasing with respect to the amount of harvested energy, e .

The results presented in Table 5.1, 5.2 and 5.3 confirm our theoretical results that, indeed, the optimal policy harvests energy whenever $b = 0$ or $m = R_1$ and accumulates mutual information whenever $b > E_d$. For the rest of the states it does not matter what the receiver does, as long as, it does not split the received RF signal.

Table 5.2: Mean time to absorption for $R_1 = 10$, $R_0 = 5$, $e = 2$ and $E_d = 5$ vs. λ

	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$	$\lambda = 0.8$	$\lambda = 0.9$
Optimal analytical	40.9000	20.8000	14.0333	10.6000	8.5000	7.0667	6.0143	5.2000	4.5444
Optimal Monte-Carlo	40.8904	20.7979	14.0320	10.5985	8.4989	7.0659	6.0140	5.1999	4.5443
BF	40.8920	20.7962	14.0337	10.6002	8.4995	7.0666	6.0153	5.1998	4.5445
IF	40.8978	20.7960	14.0331	10.5991	8.5002	7.0667	6.0132	5.1998	4.5443
CT	40.8961	20.8006	14.0333	10.5973	8.4986	7.0665	6.0137	5.2001	4.5444
Simple ARQ	87.3286	31.1145	17.9077	12.3428	9.3310	7.4591	6.1846	5.2607	4.5568

Table 5.3: Mean time to absorption for $R_1 = 10$, $R_0 = 5$, $\lambda = 0.3$ and $E_d = 10$ vs. e

	$e = 1$	$e = 2$	$e = 3$	$e = 4$	$e = 5$	$e = 6$	$e = 7$	$e = 8$	$e = 9$
Optimal analytical	40.7000	21.7000	15.0333	11.7000	11.7000	8.3667	8.3667	8.3667	8.3667
Optimal Monte-Carlo	40.6956	21.6999	15.0320	11.6995	11.7009	8.3648	8.3651	8.3675	8.3670
BF	40.7015	21.6987	15.0381	11.6986	11.6995	8.3677	8.3653	8.3667	8.3672
IF	40.7023	21.7020	15.0308	11.7030	11.7010	8.3667	8.3658	8.3671	8.3654
CT	40.6980	21.7006	15.0345	11.6995	11.6992	8.3674	8.3663	8.3657	8.3670
Simple ARQ	47.7832	26.5340	19.1515	15.4839	14.0076	11.8479	10.8730	10.4191	10.2021

5.6.3 Correlated Channel

In this section, we investigate the performance of the optimal policy presented in Algorithm 2 for the case of correlated channel and compare its performance to the three baseline policies that employ HARQ mechanism as well as a simple ARQ mechanism. We also consider a randomized policy, which we call *Bernoulli* policy which harvests energy with probability, p , unless its battery state is less than one unit or it has accumulated sufficient mutual information during when it solely harvests energy. In the following, we study the effects of the encoding rate, the time correlation, and the EH rate. Note that the mean time to absorption is determined by calculating $k_{b,m,0}$ and $k_{b,m,1}$ and then averaging them with respect to the steady-state distribution of the channel states, i.e., $k_{b,m} = \phi(0)k_{b,m,0} + \phi(1)k_{b,m,1}$, where $\phi(0) = 1 - \phi(1) = \frac{1-\lambda_1}{1+\lambda_0-\lambda_1}$.

Remark 5.2. *Note that, in this section, we do not calculate the mean time to absorption by Algorithm 2 (i.e., $k_{b,m,0}^*$ and $k_{b,m,1}^*$). Instead, we use the optimal scheduling decisions dictated by Algorithm 2 for each state (b, m, G) to determine the mean time to absorption by Monte-Carlo simulations. This is because both methods yield the same mean time to absorption for the optimal policy and illustrating both on the same figure distinctly is not possible.*

To investigate the effect of the encoding rate on the mean time to absorption, we set the simulation parameters as $R_1 = 10$, $e = 1$, $E_d = 5$ and $p = 0.1$. The mean time to absorption with respect to R_0 , for negatively and positively correlated channel states, are depicted in Figures 5.3a and 5.3b, respectively. Unlike the i.i.d. case the knowledge of the channel state makes a significant difference in the performance of the proposed optimal policy as compared to the baseline policies. Hence, when the channel is correlated, a

simple scheduling policy is not sufficient to achieve a low number of re-transmissions.

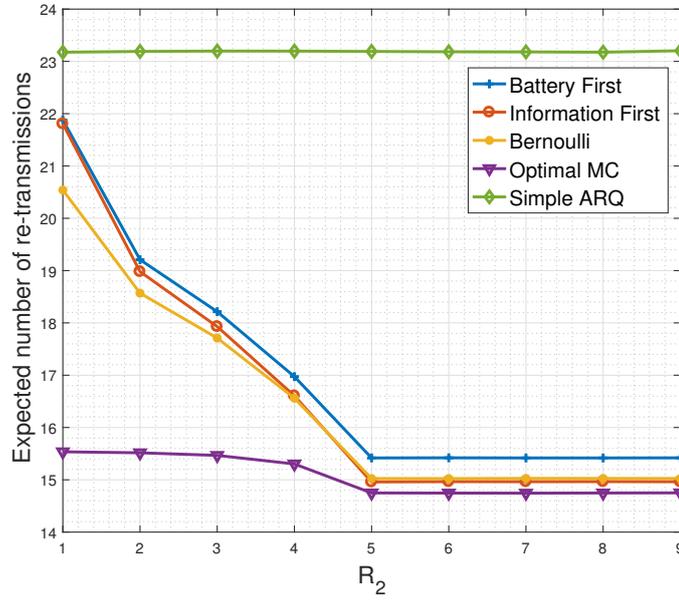
Next, we study the effect of the channel quality and the correlation on the mean time to absorption. We set $R_1 = 10$, $e = 1$, $E_d = 5$ and $p = 0.1$. We fix $\lambda_1 = 0.2$ and by varying λ_0 , we calculate the mean time to absorption as illustrated in Figure 5.4a. Similarly, we fix $\lambda_0 = 0.2$ and by varying λ_1 , we calculate the mean time to absorption by the aforementioned baseline policies and illustrate the results in Figure 5.4b. Note that when the channel is negatively correlated, as in Figure 5.4b, the gap between the optimal policy and the baseline policies is high. However, when the channel is positively correlated, as in Figure 5.4b, the gap disappears as λ_1 increases. This is because, when the channel is positively correlated, the channel tends to stay in the same state for a longer time before changing its state. On the contrary, in negatively correlated channel states, the channel is more likely to change its state at any time. This rapid change in state transition in the case of negatively correlated channel states requires a more adaptive policy rather than the case of the positively correlated channel state which rarely changes its state. Thus, the performance gain of Algorithm 2 is more evident in negatively correlated channels.

Finally the effect of EH rate, e , on the mean time to absorption for negatively and positively correlated channel states is depicted in Figure 5.5a and 5.5b, respectively. The results are obtained by setting $R_1 = 10$, $R_0 = 5$, $E_d = 10$, $p = 0.1$, $\lambda_0 = 0.7$ and $\lambda_1 = 0.2$ for negatively correlated channel states; and $\lambda_0 = 0.2$ and $\lambda_1 = 0.7$ for positively correlated channel states. We, again, observe that the optimal policy outperforms the baseline policies and the performance gain is more evident for negatively correlated channel states for the same reason we provided for the results in Figure 5.4.

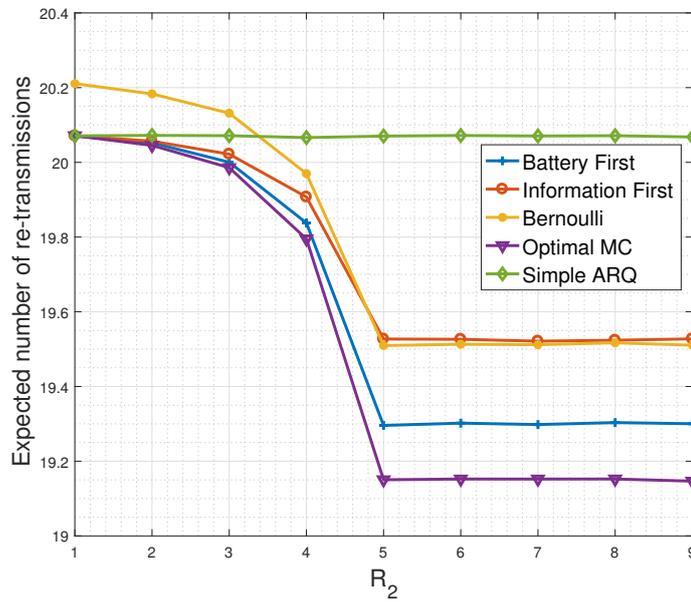
It should be noted that when the channel states are correlated, the knowledge about the future channel states plays a major role in making decision about the power splitting ratio. On the contrary, when the channel states evolve i.i.d. over time, there exist a class of optimal policies instead of a single optimal policy.

5.7 Chapter Summary

We analyzed a point-to-point wireless link employing HARQ for reliable transmission, where the receiver can only empower itself via the transmitter's RF signal. We modeled the problem of optimal power splitting using a Markovian framework, and developed an

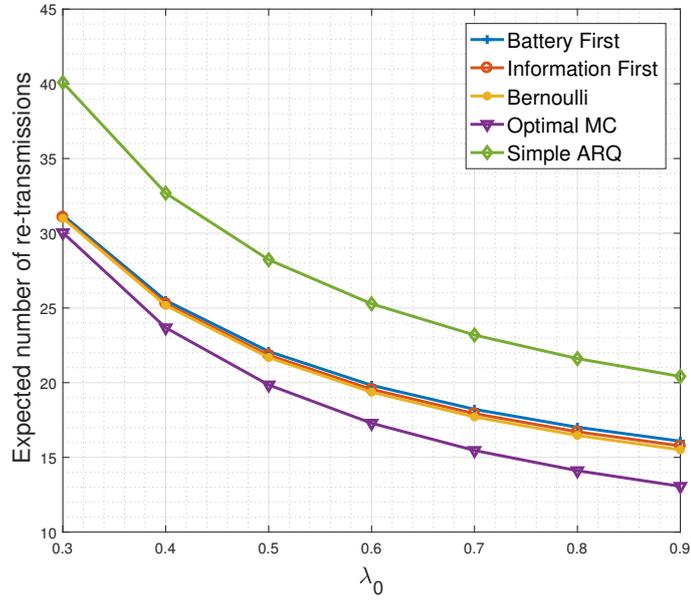


(a) Negatively correlated channel, $\lambda_0 = 0.7$ and $\lambda_1 = 0.2$.

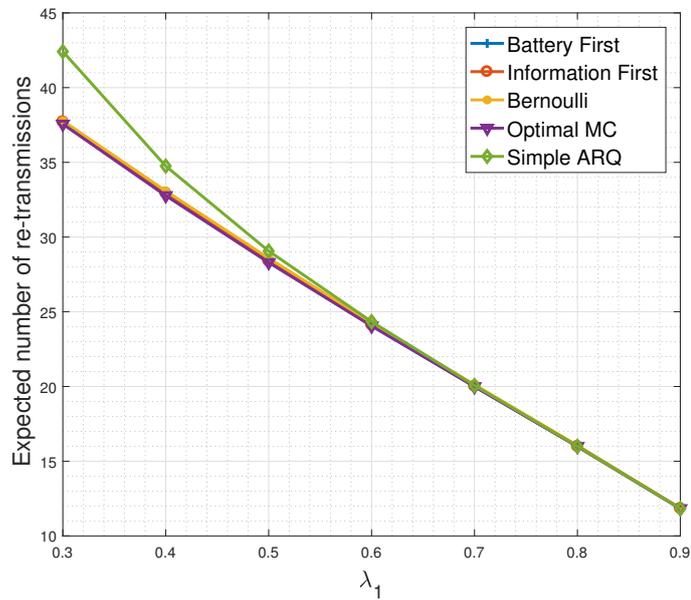


(b) Positively correlated channel $\lambda_0 = 0.2$ and $\lambda_1 = 0.7$.

Figure 5.3: The effect of the encoding rate on the minimum expected number of re-transmissions for $R_1 = 10$, $e = 1$, $E_d = 5$ and $p = 0.1$.

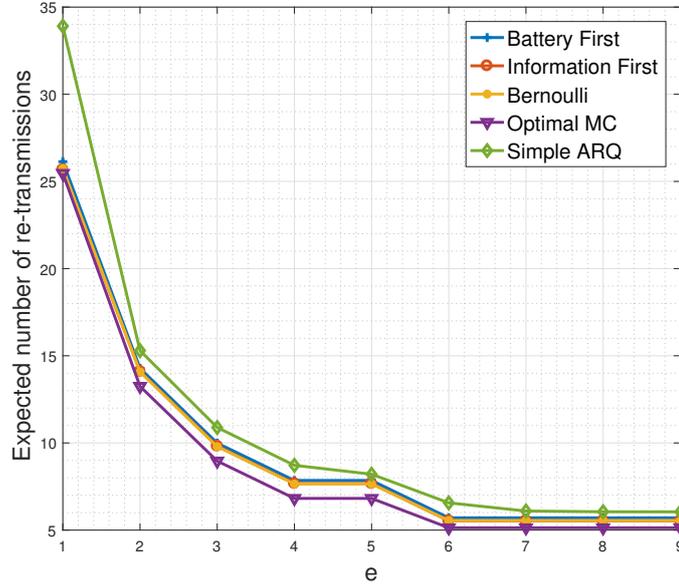


(a) Negatively correlated channel, $\lambda_1 = 0.2$.

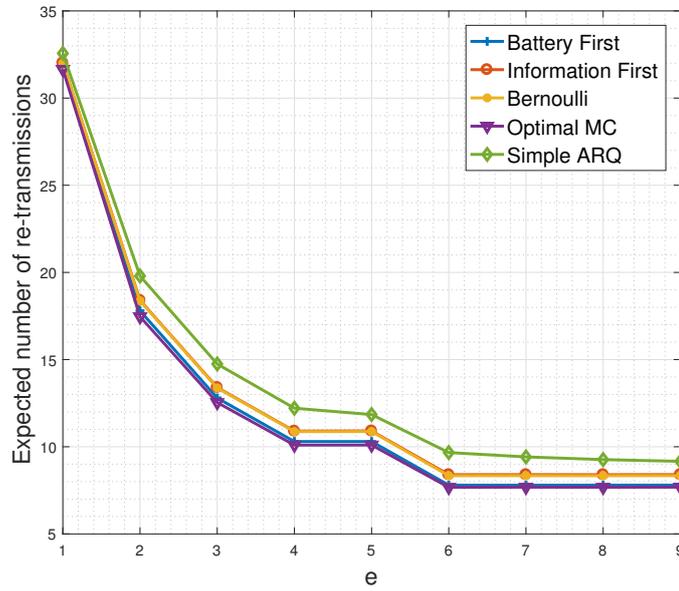


(b) Positively correlated channel, $\lambda_0 = 0.2$.

Figure 5.4: The effect of the channel quality and correlation on the minimum expected number of re-transmissions for $R_1 = 10$, $R_0 = 3$, $e = 1$, $E_d = 5$ and $p = 0.1$.



(a) $\lambda_0 = 0.7$ and $\lambda_1 = 0.2$.



(b) $\lambda_0 = 0.2$ and $\lambda_1 = 0.7$.

Figure 5.5: The effect of the EH rate on the minimum expected number of re-transmissions for $R_1 = 10$, $R_0 = 5$, $E_d = 10$ and $p = 0.1$.

optimal algorithm achieving the minimum mean time to absorption for both time varying i.i.d. and correlated channels. We developed computationally inexpensive algorithms to calculate the minimum mean time to absorption and optimize the power splitting ratio starting at any arbitrary state.

We proved that the optimal policy in case of i.i.d. channel states is not unique, and indeed the optimal policy belongs to the optimal family of policies. For correlated channel, we observed that it is only possible to achieve the optimal performance by intelligently utilizing the information offered by channel's correlation information. Finally, we numerically validated the analytical results established in the chapter by providing extensive number of simulations.

It is worth mentioning that the two-state model, adopted here, is an approximation of a more general multi-state wireless channel. As a future work, we aim to extend this work for a more general setting where we will consider multi-rate information transmission, multi-state EH process, and non-linear EH efficiency. Due to analytical complexity, it is uncertain that the optimality result of no-split policy carries over to the more general setting. In this case, deep reinforcement learning techniques can be used as a promising approach to address the aforementioned extensions.

Chapter 6

Optimal Sensing Strategy for Wirelessly Powered Devices

We are witnessing a significant advancements in the sensor technologies which has enabled a broad spectrum of applications. Often, the resolution of the produced data by the sensors significantly affects the output quality of an application. We study a sensing resolution optimization problem for a wireless powered device (WPD) that is powered by wireless power transfer (WPT) from an access point (AP). We study a class of harvest-first-transmit-later type of WPT policy, where an AP first employs RF power to recharge the WPD in the down-link, and then, collects the data from the WPD in the up-link. The WPD optimizes the sensing resolution, WPT duration and dynamic power control in the up-link to maximize an application dependant utility at the AP. The utility of a transmitted packet is only achieved if the data is delivered successfully within a finite time. Thus, we first study a finite horizon throughput maximization problem by jointly optimizing the WPT duration and power control. We prove that the optimal WPT duration obeys a time-dependent threshold form depending on the energy state of the WPD. In the subsequent data transmission stage, the optimal transmit power allocations for the WPD is shown to posses a channel-dependent fractional structure. Then, we optimize the sensing resolution of the WPD by using a Bayesian inference based multi armed bandit problem with fast convergence property to strike a balance between the quality of the sensed data and the probability of successfully delivering it.

6.1 Overview

6.1.1 Motivation

With the rapid increase in the number of battery-powered devices, energy harvesting (EH) technology provides a convenient window of opportunity to bypass the challenging, and in some cases infeasible task of replacing batteries. Traditional approaches in EH technologies harvest energy from natural resources such as wind, solar, etc. The inherent challenge of EH from natural resources is the stochastic nature of the EH process, which dictates the amount and availability of harvested energy that is beyond the control of system designers. Towards this end, wireless power transfer (WPT) [91] is considered as a promising technology to provide the network administrators a leverage on replenishing the remote devices for proper network operations, by utilizing the RF signals as a mean to transfer power to wireless powered devices (WPDs).

WPT brings forth a new dimension of optimization of the performance of sensor networks. In [92], a poll based medium access protocol (MAC) is proposed to collaboratively aide the energy request messages of those sensors that are low on energy. In [93], multiple sensors aim to estimate a parameter of interest in a distributed manner while an Access Point (AP) optimizes the WPT strategy in order to minimize the mean-square error (MSE). In [94], power-splitting and time-splitting schemes utilized in simultaneous wireless information and power transfer (SWIPT) are optimized to maximize the throughput of multiple wireless sensors. In [95], a feasibility analysis of wireless powered sensors under various scenarios is studied to ensure the reliability of energy autonomous critical infrastructure monitoring applications.

WPDs are utilized mainly for collecting and transmitting information for further processing to data collecting units. Traditionally, the scope for the application of sensors were limited to sensing and transmitting fixed-size data packets such as the information regarding temperature, humidity and etc. With the rapid development of hardware technologies for sensors many emerging applications require the transmission of a much broader type of information. On-body sensors and wearables are examples of these applications where audio, video and gesture information are captured and transmitted to an AP for further processing. The processing includes but not limited to audio, image and video where the resolution of the data points is an important factor in determining the quality of

an output produced by an application at hand. For example, the WPD could be an image sensor that transmits images to the AP, tracking the eye movement, i.e., estimating the *gaze* location of a person [49]. The accuracy of estimating the gaze depends on the number of pixels per frame. A gaze error varies from 10 – 15 pixels at 77 pixels/frame to 0 – 3 pixels at 1984 pixels/frame [96]. Hence, high resolution sensing provides a better utility in the application layer. However, high resolution sensing compromises the performance of the WPD in two main aspects; first, a high-resolution sensing typically consumes more energy. Second, it generates more data bits per sensing event which may then increase the packet drop probability. Our main objective is to strike a balance between the utility achieved by a sensing configuration and the probability of successfully delivering the sensed packet to the AP.

Optimizing the sensing resolution efficiently requires first addressing the design of WPT scenario. In wireless powered communication networks (WPCNs) [17–19], WPT occurs in the down-link (DL) to replenish the battery of WPDs which in turn is used for information transmission (IT) in the up-link (UL). A fundamental question inherited in WPCNs is the optimum duration for WPT period and power allocation in the IT period. We consider a delay sensitive sensing application scenario where the sensed packet needs to be delivered to the AP with a delay that cannot be tolerated beyond the duration of a finite horizon window. The term finite horizon corresponds to a maximum tolerable delay for the involved application. [17–19] perform a single-time-slot optimization assuming that the channel stays constant and all the harvested energy in a slot is totally used in the same time slot. Differently, [97] assumes an infinite horizon throughput maximization problem where the harvested energy is allowed to be used in later times. It was shown that this strategy significantly improves the throughput albeit having high computational complexity.

In the aforementioned works, it is assumed that in a single WPT instance, i.e., transmission of energy in the DL and reception of information in the UL, the channel state stays constant. However, in practice, this assumption is usually not valid, for example due to the body blocking the wearable sensors. In this work, we aim to optimize the sensing resolution of the WPD while jointly optimizing the WPT duration and power allocation in the IT period to maximize the chance of delivering the sensed packet by the WPD to the AP. Particularly, we first study the sub-problem of finite horizon throughput maxi-

mization, where both WPT and IT period is exposed to multiple random realizations of channel. The objective is to judiciously determine the optimal WPT duration and power allocations in the IT period. Throughput maximization problem maximizes the chance of delivering the sensed data to AP allowing to simplify the sensing optimization problem. The CSI is available causally and only in the IT period. The availability of causal CSI, makes the problem investigated here challenging, since any decision at any time slot has a cascading effect on the future outcomes.

For the throughput maximization problem, we study the problem under both offline and online settings. In the offline case, CSI is available to the WPD prior to transmission. In other words, at $t = 1$, the WPD knows the CSI for $t = 1, \dots, T$. In the online case, CSI is available only causally, i.e., the WPD only knows CSI for time t and not for any future time instants. For the offline case, we obtain closed form expressions to find the optimal WPT duration and power allocation in the IT period. We use the insights gained from the offline case, to develop an optimal online policy that maximizes the expected finite horizon throughput by optimally determining the WPT duration and power allocation in the IT period. Specifically, we formulate the problem of optimal WPT duration using the theory of stopping times. A stopping time is a random variable whose value maximizes a certain property of interest in a stochastic process. We show that there exist a time-dependent threshold on the energy level of the WPD in which it is optimal to stop WPT and start the IT period. Then, we show that the optimal power allocation in the IT period follows a fractional structure in which the WPD at each time slot allocates a fraction of its energy that depends on the current channel state as well as a specific measure of future channel expectations.

The optimal policy for determining the WPT period and power allocations in the IT period is used by the WPD to maximize its chance of delivering the sensed packet to the AP for gaining the application specific utility. Hence, as the last part of the solution, we aim to provide a framework where the WPD is able to determine the sensing resolution of the data to be sent to the AP for further processing. A high resolution data increases the performance of the application at the AP; however, a high resolution data has more bits compared to a lower resolution data which may compromise the probability of successfully delivering the data. Therefore, an optimal sensing resolution is required to balance the quality of the sensed data and the probability of successfully delivering it.

Due to the dynamic and online nature of the problem, i.e., availability of only causal information, instead of conventional optimization methods, we use Bayesian inference as a reinforcement learning method to provide a mean for the WPD in learning to balance the sensing resolution. We illustrate the benefits of the Bayesian inference over the traditional approaches such as ϵ -greedy algorithm using numerical evaluations.

6.1.2 Contributions

The contributions of the paper are summarized as follows:

- We formulate the problem of finite horizon sensing utility optimization for a WPD. The optimization problem is first addressed by maximizing the throughput of the WPD and then optimizing the sensing resolution of the sensed data.
- To maximize the throughput, we study the optimization of WPT duration and dynamic power allocation in offline and online settings.
- For the offline problem, where CSI is known non-causally, we derive a closed form expressions that enable a tractable framework to optimize both the WPT duration and power allocation in the IT period. We show that the optimal power allocation has a fractional structure depending on the current channel state as well as future channel states.
- Motivated by the results obtained from the offline problem, we formulate the online problem by assuming that the CSI is available only causally.
- We show that the optimal WPT duration for the online case has a time dependent threshold structure on the available energy of the WPD. We provide an easy to implement method to numerically calculate the thresholds.
- Similar to the offline case, we show that the optimal power allocation for the online counterpart also follows a fractional structure. The WPD allocates a fraction of its available energy in each time slot. Unlike the offline case, optimal fractions in the online case depends on the current channel state and a measure of the future channel state expectations.

- After developing an algorithm capable of maximizing the packet delivery chance, we then focus on optimizing the sensing resolution to maximize a given utility. We employ Bayesian framework based multi-armed bandit problem to learn to determine the resolution of the sensing to balance the quality of the sensed data and the probability of successfully delivering it. We show that the Bayesian framework converges much faster, by judiciously exploring in the action space of the problem, than its classic counterpart ϵ -greedy algorithm.

6.1.3 Related Work

WPCN has been studied in the literature under different settings . [98] studies a heterogeneous WPCN with the presence of EH and non-EH devices to find out how the presence of non-harvesting nodes can be utilized to enhance the network performance, compared to pure WPCNs. In [99], problem of throughput maximization in the presence of an EH relay is studied where the relay cooperatively help the source node in relaying its messages. The outage problem for a three node WPCN is analyzed in [100, 101] where both source and relay harvest energy for a certain duration, and then the source transmits to destination by using the relay. Approximate closed-form expressions for outage probability and ergodic capacity in a SWIPT scenario for multiple deployed sensors in [94]. In [102], for a multiuser orthogonal frequency division multiple (OFDM) system employing SWIPT, power-splitting and time splitting modes along with the allocation of the subcarriers are optimized so that the average outage across all users are minimized. Aforementioned works assume a known and time-invariant channel which is unlike our case where we consider a time varying channel with causal CSI. User cooperation is also studied in multiple works [18, 103, 104] to improve the performance of the WPCN by exploiting the cooperative diversity. Multiple works also studied the WPCN in the context of cloud computing [105–108]. Throughput maximization for WPCN is studied in [17, 81, 97, 109]. Per time slot throughput maximization is studied in [17]. By allowing the storage of the energy in a battery by the WPD, [97] studies infinite horizon throughput maximization in HD mode and the results are extended to FD mode in [81]. By adopting a NOMA strategy and under non-causal CSI, [109] studies the problem of finite horizon throughput maximization.

Finite horizon throughput maximization has been extensively addressed from a

communication perspective in the literature for non-RF EH techniques. For example, [23] aims at maximizing the finite horizon throughput by dynamically adjusting the transmission power in an offline setting where CSI and the EH information (EHI) is non-causally available at the transmitter for the duration of the deadline. Packet transmission time minimization over a finite horizon with non-causal EHI and a static channel is studied in [110]. However, in practice, the finite horizon spans over multiple time slots, and the CSI and EHI are not usually available. For time varying scenarios where EHI or CSI (or both) are available only causally, the problem needs to be solved dynamically. In [25, 82, 111, 112] under different EHI and CSI assumptions, the problem of finite horizon throughput maximization is formulated as a dynamic program (DP) and the optimal policy is evaluated by numerically solving the DP. The solution is later stored in the devices as a *look-up table*. However, the DP solutions are computationally expensive, and they require large memory space to store the solutions, which is usually prohibitive for resource-constrained IoT devices. Moreover, calculating and disseminating the optimal look up tables in a network consisting of large number of WPDs is inherently challenging and introduces large overheads [113]. Finally, the complexity of the numerical solutions increase exponentially with respect to the number of states in the DP formulation. A common way in dealing with such complexity is to reduce the size of the state space (action) of the problem by gaining insight into the dynamic problem as demonstrated in Chapter 4 and Chapter 5. Recently, [114] studied the problem of energy efficient scheduling for a non-RF EH over a finite horizon by developing a low complexity online heuristic policy that is built upon the offline solution and it can achieve close performance with respect to the offline policy. However, albeit the good performance, it is not evident how the algorithm would incorporate the optimal duration of the WPT period. Finally, in [115], we addressed the optimization of the WPT duration and power allocation under a simplified model. Unlike [115], here, we derive an optimal upper bound on the performance of the WPD in terms of the expected throughput over the finite horizon. We extend our results to incorporate a smart sensing application in the WPT scenario where we balance the quality of the sensed data and the probability of successful transmission using reinforcement learning. In [102], WPT is used as an incentive for motivating user involvement in a mobile crowd sensing scenario, where the users store a fraction of the received power as reward and use the rest to sense, compress and transmit a packet back to the AP for

maximizing data utility. However, the optimization problem is formulated for a single time slot with constant channel gain, enabling an offline solution approach in contrast to this work. Throughput maximization of WPT devices was previously considered in [116] where offline and online policies were presented in the context of a cognitive radio (CR) setting.

In this work, we investigate the problem of sensing optimization over a finite horizon in a WPCN where a WPD harvests energy from WPT of the AP to sense a data packet at a specific resolution and then allocates the harvested energy in the subsequent time slots to transmit its data. Unlike the previous works, we consider a scenario where the CSI evolves randomly over the duration of the deadline, and CSI is only causally available at the transmitter which necessitates an online optimization framework. We avoid the complexity of the tabular methods (such as value iteration algorithm [14]) by deriving closed form solutions for the optimal WPT duration and power allocations in the IT period. We show how the simple closed-form expressions simplify addressing the sensing optimization problem. We address the sensing optimization problem in a reinforcement learning framework, where the optimum sensing resolution is learned by the WPD in a sequence of actions and observations. Finally, we conduct extensive simulations to verify our analytical findings.

6.1.4 Outline

The paper is organized as follows: In Section 5.2, we formally present the system model and all relevant assumptions. In Section 6.3, we formulate the problem of sensing optimization. In Section 6.4, we formulate the sub-problem of finite horizon throughput maximization. In Section 6.4.1, we provide an upper bound on the maximum achievable throughput by assuming non-causal information. In Section 6.4.2, we solve the online counter-part of the problem by assuming only causal information. In Section 6.5, we address the sensing optimization problem and in Section 6.6, we provide Monte-Carlo simulations to verify our findings. Finally, we conclude the paper in Section 6.7.

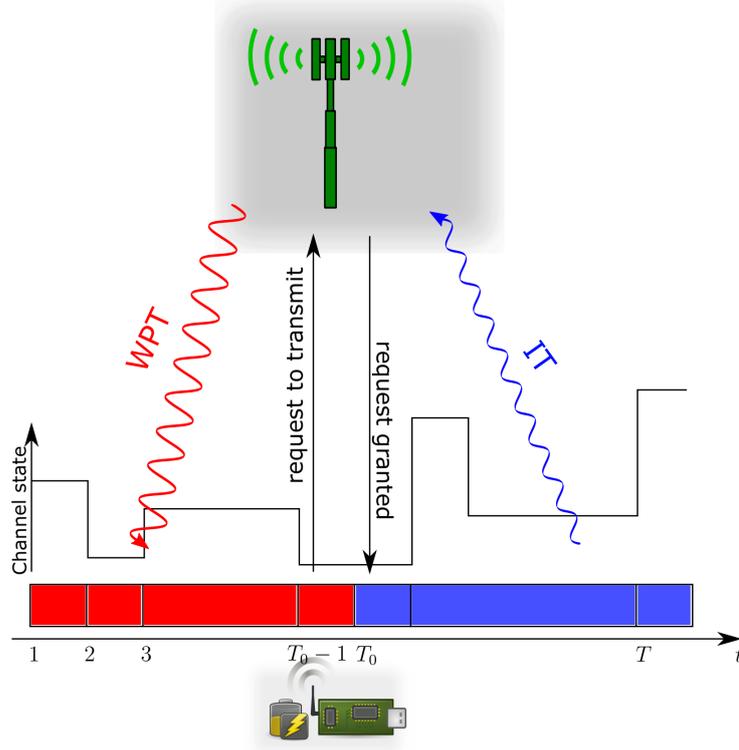


Figure 6.1: System model.

6.2 System Model

We consider a point-to-point communication wireless channel where a WPD sends its sensory data to an AP by dynamically allocating power as shown in Figure 6.1. The AP uses WPT to replenish the battery of the WPD. The WPT and information transmission (IT) periods are non-overlapping in time, assuming a half-duplex transmission scenario. We consider a harvest-first-transmit-later policy where the WPD harvests energy for a certain duration and utilizes it to sense and transmit data to the AP. Such a policy eliminates the need for signaling between the sensor and the AP at each time slot and, hence, is more suitable for energy deprived sensors. The sensory unit of the WPD is capable of capturing data at K distinct resolution settings, each representing a quality point which is described by the number of bits used. Let L_k be the size of the type $k = 1, \dots, K$ sensed data in bits. The duration of WPT and IT periods is governed by the channel gain process which jointly affects the amounts of the harvested energy and transmitted data. We assume a discrete time scenario over a finite horizon. The time is slotted $t = 1, \dots, T$ and $T < \infty$ denotes the frame length in units of slots. Let $g(t)$, $E_h(t)$ be the channel gain, and the amount of harvested energy at time slot t , respectively. Specifically, the amount

of harvested energy at time slot t is available at the beginning of slot $t + 1$. The wireless channel is modeled as a multi state independent and identically distributed (iid) random process with N levels. The channel gain remains constant for a duration of a time slot but changes randomly from one time slot to another, e.g., a wearable sensor exposed to blockage due to the movement of a person. Let $g(t) \in \{g_1, \dots, g_N\}$ be the channel power gain at slot t . We set $\mathbb{P}(g(t) = g_n) = q_n$ ¹. The WPD only has causal CSI and only during the IT period.

The AP transmits a power beacon of P watts over the wireless channel for a duration of $T_0 - 1$ time slots. Assuming channel reciprocity, the amount of energy harvested by the WPD at time t is $E_h(t) = \eta\delta g(t)P$, where η is a constant representing the efficiency of the EH process² and δ is the duration of a time slot. The energy state of the WPD at time slot t is denoted by $E(t)$. Let us denote $e_n = \eta\delta g_n P$ as the amount of harvested energy when the channel state is at level n . At the beginning of the T_0 -th time slot, the WPD consumes \mathcal{E}_k Joules to sense L_k bits of data to be sent to the AP. Immediately after sensing the data, IT period starts.

At time slot $t \geq T_0$, the WPD transmits with power $p(t)$, and the received power at the AP is $p(t)g(t)$. In order to develop a tractable analytical solution, we assume a widely used empirical transmission energy model as in [1, 117–121]. Specifically, the instantaneous rate of transmitting with power $p(t)$ when the channel gain is $g(t)$ is calculated by

$$r(t) = \sqrt[m]{\frac{p(t)g(t)}{\lambda}} \quad (6.1)$$

where λ denotes the energy coefficient incorporating the effects of bandwidth and noise power and m is the monomial order determined by the adopted coding scheme [1]. Figure 6.2 [1], compares the actual transmission rate with the monomial model described in (6.1). The approximated energy rate model, although may not be general for all cases, provides closed-form solutions for a challenging dynamic problem that gives insights to a practical and emerging problem.

Each type k data corresponds to a application specific utility upon being delivered

¹Note that g_n 's can be obtained by discretizing a continuous time channel process.

²Note that η in practice is a function of the received power and cannot be assumed to be a constant. We will show in Section 6.4 how to extend the results to account for an η when it is a function of the received power.

Modulation	Bits/symbol	SNR/symbol
2 PAM	1	$0.25 d^2$
4 QAM	2	$0.50 d^2$
16 QAM	4	$1.25 d^2$
64 QAM	6	$5.25 d^2$

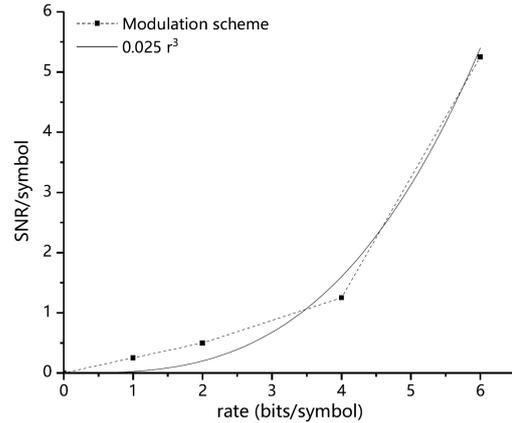


Figure 6.2: The comparison of monomial and actual transmission rate and required signal-to-noise (SNR) ratio per symbol for $m = 3$ and $\lambda = 0.025$ as given in [1]. d represents the minimum distance between signal points.

to the AP. If the WPD successfully delivers a type k data, it receives a known utility of $Z(L_k)$, and zero otherwise. We emphasize that providing a high resolution input data provides a higher utility. However, the increased utility in the application layer comes at a price of reduced chance of delivering the input data to the AP due to the finite time horizon and the dynamic nature of the wireless link. Hence, there exists an optimal trade-off in balancing the quality of input data and probability of delivering it successfully to the AP for processing. The WPD aims at maximizing its utility by jointly determining the optimal sensing resolution; optimal WPT period duration, T_0 ; and optimal power allocation in the IT period, $p(t)$ for $t = T_0, \dots, T$ in a decentralized fashion.

6.3 Problem Formulation

In this section, we formulate a joint utility optimization problem that aims at finding the optimal sensing resolution, the optimal trade-off between the EH and IT periods, and the dynamic control of transmission power during the IT period. More specifically, we aim at

solving the following optimization problem.

$$\max_{L_k, T_0, \{p(t)\}_{t=T_0}^T} Z(L_k) \mathbb{P} \left(\sum_{t=T_0}^T \sqrt{\frac{g(t)p(t)}{\lambda}} > L_k \right) \quad (6.2)$$

$$p(t) \leq E(t)/\delta, \quad t = T_0, \dots, T, \quad (6.3)$$

$$E(t+1) = E(t) + E_h(t), \quad t = 1, \dots, T_0 - 1, \quad (6.4)$$

$$E(t+1) = E(t) - p(t)\delta - \mathcal{E}_k \mathbb{1}_{t=T_0}, \quad t = T_0, \dots, T, \quad (6.5)$$

$$L_{min} \leq L_k \leq L_{max}. \quad (6.6)$$

Note that (6.2) is the expected utility of delivering data type k , (6.3) ensures that the consumed energy does not exceed the available energy, (6.4) and (6.5) are the battery dynamics in the WPT and IT periods, and (6.6) corresponds to the number of available resolution settings, respectively. Note that, in general, providing an explicit equation for $Z(L_k)$ may render infeasible as in the case of relating the error of estimating the gaze location to the number of pixels per frame. However, as we demonstrate in Section V, there is no need to have an explicit formulation for the utility function to optimize the sensing resolution. As long as there is a quantifiable mapping, either empirically or analytically, between L_k and the utility, we can find the optimal solution.

The above optimization problem consists of three sub-problems; choosing the size of the input data L_k , determining the optimal WPT duration T_0 , and optimal power allocations in the IT period $p(t)$, $t = T_0, \dots, T$. Note that a policy which maximizes the expected throughput of the WPD, by optimizing the optimal WPT duration and power allocation in the IT period, has a better probability of success compared to any alternative policy. Thus, in the following, we first consider finite horizon throughput maximization by optimizing the WPT duration as well as power allocation in the IT period.

6.4 Finite Horizon Throughput Maximization

In this section, we jointly optimize the WPT duration and power allocation in order to maximize the expected throughput of the WPD. Explicitly, We aim at solving the follow-

ing optimization problem³:

$$\max_{T_0, \{p(t)\}_{t=T_0}^T} \sum_{t=T_0}^T \sqrt[m]{\frac{g(t)p(t)}{\lambda}} \quad (6.7)$$

$$p(t) \leq E(t)/\delta, \quad t = T_0, \dots, T, \quad (6.8)$$

$$E(t+1) = E(t) + E_h(t), \quad t = 1, \dots, T_0 - 1 \quad (6.9)$$

$$E(t+1) = E(t) - p(t)\delta, \quad t = T_0, \dots, T. \quad (6.10)$$

Note that the objective function (6.7) is the total number of transmitted bits in the IT period, (6.8) ensures that the consumed energy does not exceed the available energy, (6.9) and (6.10) are the battery dynamics in the WPT and IT periods, respectively. We first solve the offline version of the optimization problem by assuming that the channel gains are available prior to the optimization. Using the insights from the offline problem, we will design an optimal online policy, where the channel gains are only available causally.

6.4.1 Optimal Offline Policy

We consider the offline counterpart of the optimization problem in (6.7). Thus, we assume that values of $g(t)$ are known non-causally for $t = 1, \dots, T$. Assuming that the optimal value of T_0 is given, we first aim at optimizing the power allocation in the IT period. We are interested in maximizing the following function

$$\max_{p(t)} \sum_{t=T_0}^T r(t)$$

$$0 \leq p(t) \leq E(t)/\delta.$$

In Theorem 6.1, we show that the optimal policy, that maximizes the total number of bits transmitted in the IT period, allocates at each time slot a fraction of the available energy which depends on the current channel realization as well as a measure of future channel expectations.

³For clarity of the presentation, we neglect the energy consumption of sensing, i.e., \mathcal{E}_k s, without affecting the main results. We consider them in the numerical evaluations.

Theorem 6.1. For a given T_0 and realizations of $g(t)$ for $t = 1, \dots, T$, the optimal dynamic power allocation for the offline problem is calculated by

$$p^*(t) = \frac{g(t)^{\frac{1}{m-1}}}{g(t)^{\frac{1}{m-1}} + G(t+1)^{\frac{1}{m-1}}} \frac{E(t)}{\delta} \quad (6.11)$$

where

$$G(t) = \begin{cases} \left[g(t)^{\frac{1}{m-1}} + G(t+1)^{\frac{1}{m-1}} \right]^{m-1}, & \text{if } t \leq T \\ 0, & \text{if } t > T \end{cases}, \quad (6.12)$$

and the maximum number of transmitted bits is calculated as

$$\sum_{t=T_0}^T r^*(t) = \sqrt[m]{\frac{E(T_0)}{\delta \lambda}} G(T_0) \quad (6.13)$$

Proof. Consider the following concave optimization of the throughput at time $T - 1$ and T , given that the amount of available energy at time $T - 1$ is $E(T - 1)$

$$\begin{aligned} \max_{p(T-1), p(T)} & \sqrt[m]{\frac{g(T-1)p(T-1)}{\lambda}} \\ & + \sqrt[m]{\frac{g(T)(E(T-1)/\delta - p(T-1))}{\lambda}} \\ & p(T-1) \leq E(T-1)/\delta, \\ & p(T) \leq E(T-1)/\delta - p(T-1). \end{aligned}$$

The WPD at the last time slot should utilize all the available energy before the transmission frame expires. Hence, we set $p(T) = E(T - 1)/\delta - p(T - 1)$. The optimization problem becomes

$$\begin{aligned} \max_{p(T-1)} & \sqrt[m]{\frac{g(T-1)p(T-1)}{\lambda}} + \sqrt[m]{\frac{g(T)(E(T-1)/\delta - p(T-1))}{\lambda}} \\ & 0 \leq p(T-1) \leq E(T-1)/\delta. \end{aligned}$$

The Lagrangian of the above problem can be written as

$$\begin{aligned}
\mathcal{L}(p(T-1), \mu_1, \mu_2) &= \sqrt[m]{\frac{g(T-1)p(T-1)}{\lambda}} \\
&+ \sqrt[m]{\frac{g(T)(E(T-1)/\delta - p(T-1))}{\lambda}} \\
&- \mu_1(p(T-1) - E(T-1)/\delta) \\
&+ \mu_2 p(T-1)
\end{aligned}$$

The derivative of the Lagrangian is calculated as follows

$$\begin{aligned}
\frac{\partial \mathcal{L}(p(T-1), \mu_1, \mu_2)}{\partial p(T-1)} &= \frac{1}{m} \sqrt[m]{\frac{g(T-1)}{\lambda}} p(T-1)^{\frac{1}{m}-1} \\
&- \frac{1}{m} \sqrt[m]{\frac{g(T)}{\lambda}} (E(T-1)/\delta - p(T-1))^{\frac{1}{m}-1} \\
&+ (\mu_2 - \mu_1)
\end{aligned}$$

Prior to equating the Lagrangian to zero, we assume that the optimal power allocation satisfies the constraint, i.e., $0 \leq p^*(T-1) \leq E(T-1)/\delta$, and set $\mu_1 = \mu_2 = 0$. By solving the derivative of the relaxed Lagrangian, we get

$$p^*(T-1) = \frac{g(T-1)^{\frac{1}{m-1}}}{g(T-1)^{\frac{1}{m-1}} + g(T)^{\frac{1}{m-1}}} E(T-1)/\delta$$

Note that since $0 \leq \frac{g(T-1)^{\frac{1}{m-1}}}{g(T-1)^{\frac{1}{m-1}} + g(T)^{\frac{1}{m-1}}} \leq 1$, the constraint is satisfied. Let us calculate the optimum sum throughput at time $T-1$ and T :

$$\begin{aligned}
r(T-1) + r(T) &= \sqrt[m]{\frac{g(T-1)p^*(T-1)}{\lambda}} \\
&+ \sqrt[m]{\frac{g(T)(E(T-1)/\delta - p^*(T-1))}{\lambda}} \\
&= \sqrt[m]{\frac{E(T-1)/\delta}{\lambda}} \left[\sqrt[m]{\frac{g(T-1)g(T-1)^{\frac{1}{m-1}}}{g(T-1)^{\frac{1}{m-1}} + g(T)^{\frac{1}{m-1}}}} \right. \\
&\quad \left. + \sqrt[m]{\frac{g(T)g(T)^{\frac{1}{m-1}}}{g(T-1)^{\frac{1}{m-1}} + g(T)^{\frac{1}{m-1}}}} \right] \\
&= \sqrt[m]{\frac{E(T-1)/\delta}{\lambda}} \frac{g(T-1)^{\frac{1}{m-1}} + g(T)^{\frac{1}{m-1}}}{\sqrt[m]{g(T-1)^{\frac{1}{m-1}} + g(T)^{\frac{1}{m-1}}}} \\
&= \sqrt[m]{\frac{E(T-1)/\delta G(T-1)}{\lambda}},
\end{aligned}$$

where $G(T-1) = [g(T-1)^{\frac{1}{m-1}} + g(T)^{\frac{1}{m-1}}]^{m-1}$. To generalize the results, we use induction. Suppose that the above results are true for some time $t+1$. Next consider the optimization of sum throughput from time t to T :

$$\max_{p(t)} \sqrt[m]{\frac{g(t)p(t)}{\lambda}} + \sqrt[m]{\frac{(E(t)/\delta - p(t))G(T-1)}{\lambda}}$$

Similar to the above analysis, it follows that

$$\begin{aligned}
p^*(t) &= \frac{g(t)^{\frac{1}{m-1}}}{g(t)^{\frac{1}{m-1}} + G(t+1)^{\frac{1}{m-1}}} E(t) \\
\sum_{\tau=t}^T r(\tau) &= \sqrt[m]{\frac{E(t)G(t)}{\lambda}},
\end{aligned}$$

where $G(t) = [g(t)^{\frac{1}{m-1}} + G(t+1)^{\frac{1}{m-1}}]^{m-1}$. □

The offline optimization problem becomes:

$$\begin{aligned}
\max_{T_0} \sqrt[m]{\frac{E(T_0)}{\delta\lambda}} G(T_0) & \tag{6.14} \\
2 \leq T_0 \leq T. &
\end{aligned}$$

The above maximization problem has only one integer variable and hence, the optimal value for T_0 can be easily calculated numerically. In Figure 6.3, we illustrate a sample realization of the battery of the WPD. The time frame has 10 time slots, each with a duration of $1ms$. The WPD accumulates energy until $t = 2$. At $t = 3$, since the available energy is larger than the threshold, the WPT period is stopped and the IT period began⁴.

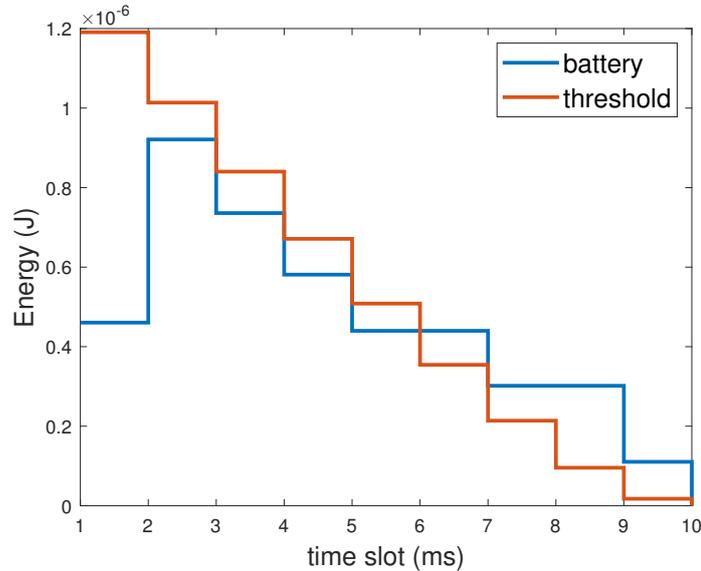


Figure 6.3: An illustrative example of the battery evolution, $E(t)$, where $T = 10$.

6.4.2 Optimal Online Policy

Note that, in the online case, $g(t)$ is only available causally. Therefore, the optimization problem in (6.7)-(6.10) cannot be solved using offline optimization tools and an online algorithm is required for its solution. A common approach to solve similar problems is to use dynamic programming (DP) [15] to find the solution numerically, and store the optimal decisions in a look-up table for the WPD. However, solving a DP and storing the result is prohibitive for resource constrained WPDs. In the following, we extend the insights gained in the offline case to the online counterpart of the optimization problem in (6.7).

At each time slot $t \geq T_0$, the WPD allocates a fraction of its remaining energy and allocates $p(t) = \alpha(t)E(t)/\delta$ as its transmit power. Hence, the optimization problem

⁴In Section 6.4.2, we show how to calculate the optimal WPT duration and power allocations in the IT period.

converts to:

$$\max_{T_0, \{\alpha(t)\}_{t=T_0}^T} \sum_{t=T_0}^T \sqrt[m]{\frac{g(t)\alpha(t)E(t)}{\delta\lambda}} \quad (6.15)$$

$$0 \leq \alpha(t) \leq 1, \quad t = T_0, \dots, T, \quad (6.16)$$

$$E(t+1) = E(t) + E_h(t), \quad t = 1, \dots, T_0 - 1 \quad (6.17)$$

$$E(t+1) = (1 - \alpha(t))E(t), \quad t = T_0, \dots, T. \quad (6.18)$$

Dynamic Energy Allocation

In this section, we first optimize the values of $\alpha(t)$ by conditioning on T_0 . Then using the obtained result, we will give a criteria for stopping the EH process, i.e., optimizing the value of T_0 .

Let the IT period begin at T_0 and aim to maximize the throughput over $T - T_0$ time slots by using DP. The problem is recursively solved starting at the last time slot T , and the result is propagated by recursion until it reaches $t = T_0$. We denote the instantaneous reward of choosing $\alpha(t)$ by $U_{\alpha(t)}(E(t), g(t))$ which is the instantaneous number of bits transmitted to the AP, when the amount of available energy at time t , is $E(t)$ and the channel power gain is at state $g(t)$. Thus,

$$U_{\alpha(t)}(E(t), g(t)) = \sqrt[m]{\frac{\alpha(t)g(t)E(t)}{\delta\lambda}}. \quad (6.19)$$

We denote the action-value function by $V_{\alpha}(E(t), g(t))$ which is equal to the instantaneous reward of choosing $\alpha(t)$ plus the expected number of bits that can be transmitted in the future. Hence, the action-value function evolves as,

$$\begin{aligned} V_{\alpha(t)}(E(t), g(t)) = & U_{\alpha(t)}(E(t), g(t)) \\ & + \sum_{i=1}^N q_i V(E(t+1), g_i), \end{aligned} \quad (6.20)$$

where, $V(E(t), g(t))$ is the value function defined as,

$$V(E(t), g(t)) = \max_{\alpha(t)} V_{\alpha(t)}(E(t), g(t)). \quad (6.21)$$

Note that at the last time slot, i.e., $t = T$, all the energy in the battery will be used

for transmission, i.e., $\alpha(T) = 1$. Thus, it follows that,

$$\begin{aligned}
V(E(T), g(t)) &= U_1(E(T), g(T)) \\
&= \sqrt[m]{\frac{g(T)E(T)}{\delta\lambda}} \\
&= \sqrt[m]{\frac{g(T)(1 - \alpha(T-1))E(T-1)}{\delta\lambda}}.
\end{aligned} \tag{6.22}$$

We maximize the action-value function at $t = T - 1$ by optimizing $\alpha(T - 1)$ as follows,

$$\begin{aligned}
V_\alpha(E(T-1), g(T-1)) &= U_\alpha(E(T-1), g(T-1)) \\
&+ \sum_{i=1}^N q_i V((1 - \alpha(T-1))E(T-1), g_i) \\
&= \sqrt[m]{\frac{g(T-1)\alpha(T-1)E(T-1)}{\delta\lambda}} \\
&+ \sum_{i=1}^N q_i \sqrt[m]{\frac{g_i((1 - \alpha(T-1))E(T-1))}{\delta\lambda}}.
\end{aligned} \tag{6.23}$$

It is easy to see that (6.23) is concave with respect to $\alpha(T - 1)$. Therefore, by differentiating (6.23), the optimal $\alpha(T - 1)$ can be calculated as follows:

$$\alpha^*(T-1) = \frac{g(T-1)^{\frac{1}{m-1}}}{g(T-1)^{\frac{1}{m-1}} + Q(T-1)^{\frac{m}{m-1}}}, \tag{6.24}$$

where,

$$Q(T-1) = \sum_{i=1}^N q_i \sqrt[m]{g_i}. \tag{6.25}$$

The corresponding value function can also be calculated as

$$\begin{aligned}
V(E(T-1), g(T-1)) &= \sqrt[m]{\frac{E(T-1)}{\delta\lambda}} (g(T-1))^{\frac{1}{m-1}} \\
&+ Q(T-1)^{\frac{m}{m-1}})^{\frac{m-1}{m}}.
\end{aligned} \tag{6.26}$$

In a similar manner as above, we can recursively calculate the optimal $\alpha(t)$ for $t = T - 2, \dots, T_0$. The result is summarized in the following theorem.

Theorem 6.2. For any $t = T - 1, \dots, T_0$, the optimal decision is to choose

$$\alpha^*(t) = \frac{g(t)^{\frac{1}{m-1}}}{g(t)^{\frac{1}{m-1}} + Q(t)^{\frac{m}{m-1}}}, \quad (6.27)$$

where

$$Q(t) = \sum_{i=1}^N q_i (g_i^{\frac{1}{m-1}} + Q(t+1)^{\frac{m}{m-1}})^{\frac{m-1}{m}}. \quad (6.28)$$

The corresponding value function is

$$V(E(t), g(t)) = \sqrt[m]{\frac{E(t)}{\delta\lambda}} (g(t)^{\frac{1}{m-1}} + Q(t)^{\frac{m}{m-1}})^{\frac{m-1}{m}} \quad (6.29)$$

Proof. The proof is by induction. We have shown in (6.24), (6.25), and (6.26), that the case for $k = 1$ is true. By assuming the the case for $k - 1$ is true, let us calculate the case k . The value function is given as

$$\begin{aligned} V_\alpha(E(T-k), g(T-k)) &= U_\alpha(E(T-k), g(T-k)) \\ &\quad + \sum q_i V(E(T-(k-1)), g_i) \end{aligned} \quad (6.30)$$

Note that $E(T-(k-1)) = (1 - \alpha(T-k))E(T-k)$ and since the case is true for $k - 1$, from (6.29), we have

$$\begin{aligned} V(E(T-(k-1)), g_i) &= \sqrt[m]{\frac{(1 - \alpha(T-k))E(T-k)/\delta}{\lambda}} (g_i^{\frac{1}{m-1}} \\ &\quad + Q(T-k+1)^{\frac{m}{m-1}})^{\frac{m-1}{m}} \end{aligned} \quad (6.31)$$

By substituting (6.31) in (6.30) we get

$$\begin{aligned}
V_\alpha(E(T-k), g(T-k)) &= \sqrt[m]{\frac{g(T-k)\alpha(T-k)E(T-k)/\delta}{\lambda}} \\
&+ \sum q_i \sqrt[m]{\frac{(1-\alpha(T-k))E(T-k)/\delta}{\lambda}} \times \\
&\quad (g_i^{\frac{1}{m-1}} + Q(T-k+1)^{\frac{m}{m-1}})^{\frac{m-1}{m}}
\end{aligned} \tag{6.32}$$

By differentiating with respect to $\alpha(T-k)$ and equating to zero, we obtain:

$$\alpha^*(T-k) = \frac{g(T-k)^{\frac{1}{m-1}}}{g(T-k)^{\frac{1}{m-1}} + Q(T-k)^{\frac{m}{m-1}}}, \tag{6.33}$$

where

$$Q(T-k) = \sum q_i (g_i^{\frac{1}{m-1}} + Q(T-k+1)^{\frac{m}{m-1}})^{\frac{m-1}{m}} \tag{6.34}$$

Hence, (6.27) and (6.28) hold by induction. For the last part, let us calculate $V(E(T-k), g(T-k))$

$$\begin{aligned}
&V(E(T-k), g(T-k)) \\
&= \sqrt[m]{\frac{g(T-k)g(T-k)^{\frac{1}{m-1}}E(T-k)/\delta}{\lambda(g(T-k)^{\frac{1}{m-1}} + Q(T-k)^{\frac{m}{m-1}})}} \\
&+ \sum q_i \sqrt[m]{\frac{Q(T-k)^{\frac{m}{m-1}}E(T-k)}{\lambda(g(T-k)^{\frac{1}{m-1}} + Q(T-k)^{\frac{m}{m-1}})}} \\
&\quad \times (g_i^{\frac{1}{m-1}} + Q(T-k+1)^{\frac{m}{m-1}})^{\frac{m-1}{m}} \\
&= \sqrt[m]{\frac{E(T-k)/\delta}{\lambda(g(T-k)^{\frac{1}{m-1}} + Q(T-k)^{\frac{m}{m-1}})}} \\
&\quad \times (g(T-k)^{\frac{1}{m-1}} + Q(T-k)^{\frac{m}{m-1}}) \\
&= \sqrt[m]{\frac{E(T-k)/\delta}{\lambda}} (g(T-k)^{\frac{1}{m-1}} + Q(T-k)^{\frac{m}{m-1}})^{\frac{m-1}{m}}.
\end{aligned} \tag{6.35}$$

Thus, (6.29) also holds by induction. \square

Theorem 6.2 gives a framework to dynamically allocate energy at each time slot $t \geq T_0$. Instead of numerically solving the DP and storing it in a large look up table, WPD needs to just calculate and store an array of values with a maximum dimension of T . The closed form expressions derived in (6.27)-(6.29) significantly simplify the procedure to optimize T_0 . We will use these results to find an structure for the optimal stopping time problem in the subsequent section.

Optimal Stopping time for the WPT duration

In the following, we derive the optimal stopping time for the WPT duration, i.e., optimizing T_0 in (6.7)-(6.10). Recall that the WPD accumulates energy up to some time t , and then stops the WPT to start transmitting its data bits. Also, recall that during WPT, the WPD is blind to the channel conditions. If the WPD stops the WPT at time t , then the expected number of bits that can be transmitted is

$$\begin{aligned} \sum_{i=1}^N q_i V(E(t), g_i) &= \sum_{i=1}^N q_i \sqrt[m]{\frac{E(t)}{\delta\lambda}} (g_i^{\frac{1}{m-1}} + Q(t)^{\frac{m}{m-1}})^{\frac{m-1}{m}} \\ &= \sqrt[m]{\frac{E(t)}{\delta\lambda}} Q(t-1). \end{aligned} \quad (6.36)$$

Note that (6.36) follows from the definition of $Q(t)$ given in (6.28).

Let $J_t(E(t))$, $t = 1, \dots, T$ be the maximum expected number of bits that can be transmitted if the WPT is stopped at time t , and the amount of available energy is $E(t)$. At any time t , the WPD will either stop or continue the WPT. The optimal stopping time for the WPT can be formulated as

$$\max_{t \leq T} J_t(E(t)), \quad (6.37)$$

where

$$\begin{aligned} J_t(E(t)) &= \max \left(\sqrt[m]{\frac{E(t)}{\delta\lambda}} Q(t-1) \right. \\ &\quad \left. , \mathbb{E}(J_{t+1}(E(t+1)) \mid E(t)) \right). \end{aligned} \quad (6.38)$$

The problem can be formulated as a DP and recursively solved for every possible

$E(t)$ and t . Before proceeding, we need the following lemma.

Lemma 6.1. $Q(t)$, defined in (6.28) is a monotonically decreasing function in t .

Proof.

$$\begin{aligned} \frac{Q(t)}{Q(t+1)} &= \frac{\sum_{i=1}^N q_i (g_i^{\frac{1}{m-1}} + Q(t+1)^{\frac{m}{m-1}})^{\frac{m-1}{m}}}{Q(t+1)} \\ &= \sum_{i=1}^N q_i \left(1 + \frac{g_i^{\frac{1}{m-1}}}{Q(t+1)^{\frac{m}{m-1}}}\right)^{\frac{m-1}{m}} > 1. \end{aligned} \quad (6.39)$$

It readily follows that $Q(t) > Q(t+1)$. \square

Note that at $t = T$, the best strategy is to stop the WPT and start the IT period, since otherwise no bits can be transmitted to the AP. Thus,

$$J_T(E(T)) = \sqrt[m]{\frac{E(T)}{\delta\lambda}} Q(T-1). \quad (6.40)$$

We continue the recursive evaluation at time slot $t = T - 1$. We have,

$$\begin{aligned} &J_{T-1}(E(T-1)) \\ &= \max\left(\sqrt[m]{\frac{E(T-1)}{\delta\lambda}} Q(T-2), \mathbb{E}(J_T(E(T))|E(T-1))\right) \\ &= \max\left(\sqrt[m]{\frac{E(T-1)}{\delta\lambda}} Q(T-2), \sum_{i=1}^N q_i \sqrt[m]{\frac{E(T-1) + e_i}{\delta\lambda}} Q(T-1)\right) \end{aligned} \quad (6.41)$$

Since $Q(T-2) > Q(T-1)$ as proven in Lemma 6.1, if $E(T-1) \geq \gamma(T-1)$, then

$$\sqrt[m]{\frac{E(T-1)}{\delta\lambda}} Q(T-2) \geq \sum_{i=1}^N q_i \sqrt[m]{\frac{E(T-1) + e_i}{\delta\lambda}} Q(T-1), \quad (6.42)$$

where $\gamma(T-1)$ is the solution to the following equation

$$\sum_{i=1}^N q_i \sqrt[m]{1 + \frac{e_i}{\gamma(T-1)}} = \frac{Q(T-2)}{Q(T-1)}. \quad (6.43)$$

Note that $\gamma(T-1)$ admits a unique solution because the left hand side of (6.43) is a strictly decreasing function in $\gamma(T-1)$ and its range belongs to $(1, \infty)$. Also, from

Lemma 6.1, we know that $\frac{Q(T-2)}{Q(T-1)} > 1$. Hence, it is optimal to stop the WPT at time $T - 1$ if $E(T - 1) \geq \gamma(T - 1)$. This suggests that the optimal stopping times are governed by a time varying threshold type structure, where at any given time t , it is optimal to stop the WPT if $E(t) \geq \gamma(t)$. Before, proving this observation, we need the following lemma.

Lemma 6.2. *For any $k = 1, \dots, T - 1$, we have*

$$\frac{Q(T - k - 1)}{Q(T - k)} < \frac{Q(T - k)}{Q(T - k + 1)} \quad (6.44)$$

Proof. By using (6.28), we have

$$\begin{aligned} \frac{Q(T - k - 1)}{Q(T - k)} &= \frac{\sum_{i=1}^N q_i (g_i^{\frac{1}{m-1}} + Q(T - k)^{\frac{m}{m-1}})^{\frac{m-1}{m}}}{Q(T - k)} \\ &= \sum_{i=1}^N q_i \left(1 + \frac{g_i^{\frac{1}{m-1}}}{Q(T - k)^{\frac{m}{m-1}}}\right)^{\frac{m-1}{m}}, \end{aligned} \quad (6.45)$$

and,

$$\begin{aligned} \frac{Q(T - k)}{Q(T - k + 1)} &= \frac{\sum_{i=1}^N q_i (g_i^{\frac{1}{m-1}} + Q(T - k + 1)^{\frac{m}{m-1}})^{\frac{m-1}{m}}}{Q(T - k + 1)} \\ &= \sum_{i=1}^N q_i \left(1 + \frac{g_i^{\frac{1}{m-1}}}{Q(T - k + 1)^{\frac{m}{m-1}}}\right)^{\frac{m-1}{m}}. \end{aligned} \quad (6.46)$$

From Lemma 6.1, we have $Q(T - k) > Q(T - k + 1)$ and thus the lemma holds. \square

In the following theorem, we give the structure of the optimal stopping policy.

Theorem 6.3. *At each time slot t , the optimal decision is to stop the WPT if $E(t) \geq \gamma(t)$, where $\gamma(t)$ is the solution to the following equation,*

$$\sum_{n=1}^N q_n \sqrt[m]{1 + \frac{e_n}{\gamma(t)}} = \frac{Q(t - 1)}{Q(t)} \quad (6.47)$$

Proof. The proof is by induction. We will show that the result of the theorem is true for $J_t(E(t))$ for all $t = 1, \dots, T - 1$. The result of the theorem is verified for $t = T - 1$ in (6.43). Let us assume that the theorem holds for $t + 1$, i.e., if $E(t + 1) \geq \gamma(t + 1)$, it is optimal to stop the EH process, where $\gamma(t + 1)$ is the solution to the following equation,

$$\sum q_i \sqrt[m]{1 + \frac{e_i}{\gamma(t+1)}} = \frac{Q(t)}{Q(t+1)} \quad (6.48)$$

At time slot t we have:

$$J_t(E(t)) = \max \left(\sqrt[m]{\frac{E(t)}{\delta\lambda}} Q(t-1), \mathbb{E}(J_{t+1}(E(t+1))|E(t)) \right) \quad (6.49)$$

First, let us assume that $E(t) \geq \gamma(t+1)$. Since $E(t+1) \geq E(t)$, it readily follows that $E(t+1) \geq \gamma(t+1)$. Thus, we have

$$\mathbb{E}(J_{t+1}(E(t+1))|E(t)) = \sum q_i \sqrt[m]{\frac{E(t) + e_i}{\delta\lambda}} Q(t) \quad (6.50)$$

Hence,

$$J_t(E(t)) = \max \left(\sqrt[m]{\frac{E(t)}{\delta\lambda}} Q(t-1), \sum q_i \sqrt[m]{\frac{E(t) + e_i}{\delta\lambda}} Q(t) \right) \quad (6.51)$$

Since, $Q(t-1) > Q(t)$, if $E(t) \geq \gamma(t)$, then it is optimal to stop the EH process, and $\gamma(t)$ is the solution of,

$$\sum q_i \sqrt[m]{1 + \frac{e_i}{\gamma(t)}} = \frac{Q(t-1)}{Q(t)}. \quad (6.52)$$

Note that the left hand side of (6.52) is strictly decreasing with respect to $\gamma(t)$ and its range is $(1, \infty)$. Since $\frac{Q(t-1)}{Q(t)} > 1$ is proved in Lemma 6.1, there is a unique solution for $\gamma(t)$ satisfying (6.52). Thus, if $E(t) \geq \gamma(t+1)$, then the theorem is also true for case k . In the following, we will generalize the proof for any value of $E(t)$. Note that if $\gamma(t) > \gamma(t+1)$, then the proof will include any $E(t)$. Because, if $E(t) \geq \gamma(t)$, then,

$$E(t+1) \geq E(t) \geq \gamma(t) > \gamma(t+1), \quad (6.53)$$

and (6.50) will hold. Using the results of Lemma 6.2 we have

$$\sum q_i \sqrt[m]{1 + \frac{e_i}{\gamma(t)}} < \sum q_i \sqrt[m]{1 + \frac{e_i}{\gamma(t+1)}} \quad (6.54)$$

Hence, $\gamma(t) > \gamma(t+1)$, and the theorem holds. \square

Note that the results of Theorem 6.3 can be easily extended to account for the dependability of EH efficiency, η , on the received power. More specifically, when the amount of harvested energy at fading state n is defined to be $e_n = \eta(g_n P)g_n P$, where $\eta(g_n P)$ is the EH efficiency when the received power at the WPD is $g_n P$, all the derivations given in the paper remain valid.

The results established in Theorem 6.2 and 6.3 enables us to develop an online low complexity optimal algorithm that maximizes the expected throughput. The procedure is summarized in Algorithm 3.

Algorithm 3 Online policy

```

1: Initialize  $Q(t)$  for  $t = 0, \dots, T - 1$  using (6.28),
2: Initialize  $\gamma(t)$  for  $t = 1, \dots, T - 1$  using (6.47),
3: for  $t = 1 : T$  do
4:   if  $E(t) < \gamma(t)$  then
5:     continue the WPT
6:   else
7:      $T_0 = t$ ,
8:     Stop the WPT,
9:     Break
10: for  $t = T_0 : T$  do
11:   Calculate  $\alpha(t)$  using (6.27),
12:   Transmit using  $\alpha(t)E(t)$ .
```

The time complexity of Algorithm 1 only depends on line 1 and 2 and the rest of the algorithm has a constant time complexity with respect to N , and T . Line 1 solves (6.28) where a constant time operation (i.e., the term inside the summation) is evaluated N times for any given $t = 1, \dots, T - 1$. Since (6.28) is evaluated T times, the complexity of line 1 is at most $O(NT)$. Line 2 calculates the thresholds by solving (6.47). Consider a root finding algorithm which solves (6.47) by evaluating the function at different points (e.g., bisection method). Since (6.47) involves summation of N nonlinear functions, the root finding algorithm needs to evaluate values of N non-linear functions. Thus, for a given t the complexity is $O(N)$. Moreover, since it is calculated at most T times, the overall

complexity is $O(NT)$. Thus the overall complexity of Algorithm 1 is $O(NT)$. It is worth mentioning that if the statics of the channel do not change over time, line 1 and 2 need to be calculated only once.

Remark 6.1. *Note that that the monomial rate function have enabled a closed form solution of the optimal power allocations and WPT duration. However, it is also possible to extend this work beyond the monomial rate function to the Shannon rate function. The optimal solutions for power allocations and WPT duration can be derived with the same recursive approach presented in this section. However, due to the logarithmic nature of the Shannon rate function, it is no longer possible to derive closed form solutions, and thus, we have to resort for tabular methods to store the optimal solutions. For each possible state, $(E(t), g(t), t)$, the optimal power allocation $p(t)$ should be calculated and stored in the table. A similar table is also required for storing the optimal duration of WPT. An obvious drawback of the tabular method is that the WPD endures significant computational complexities as well as memory requirement due to the large number of states.*

6.5 Optimal Sensing

Thus far, we have developed a policy that maximizes the expected finite horizon throughput of the WPD by determining the optimal WPT duration and dynamic power allocation in a distributed manner. Recall that the ultimate goal is to maximize the sensing utility of the WPD by optimizing the sensing resolution. Algorithm 3, is a framework that maximizes the chance of successful delivery of the data to the AP. Thus the last quantity to be optimized is the sensing resolution.

Remark 6.2. *Note that it is possible to increase the efficiency of the sensing utility by further compressing the sensed packets prior to the transmission as in [102]. Compressing the sensed packets decreases the number bits per packet and thus, increases the chance of delivering the packet. At the same time, due to utilizing the CPU for a number of cycles, the energy consumption of the WPD increases because of the compression. Hence, there exists a trade-off between the size of the sensed packet, compression ratio and the extra energy consumption. Compression can be easily accounted for in the learning model by simply extending the action space of the WPD to account for the compression ratio. We*

note that the inherent trade-off in compression is similar to the sensing resolution, and it can be incorporated in the formulation in a straightforward manner.

Let the event of successfully delivering a packet of L_k bits be χ_k . More specifically:

$$\chi_k = \begin{cases} 1 & \text{if } \sum_{t=T_0}^T r(t) > L_k, \\ 0 & \text{otherwise.} \end{cases} \quad (6.55)$$

We rewrite the optimization problem of interest as follows⁵

$$\max_{\{L_k\}_{k=1}^K} Z(L_k) \mathbb{E}(\chi_k) \quad (6.56)$$

The WPD in the beginning of each transmission frame chooses a L_k that optimizes the above optimization problem⁶. The unknown quantities in the optimization problem are $\mathbb{E}(\chi_k)$, $k = 1, \dots, K$. We aim to learn these quantities using a reinforcement learning (RL) technique. The RL framework interacts with the environment and learns the values of the parameters of interest by observing the outcomes of its decisions. Note that the observation feedbacks are limited and only the feedback associated with the chosen decision in a time slot is observed. This problem can be efficiently formulated in the context of *multi armed bandit* (MAB) problem. The parameters of interest in the MAB are denoted by $\theta_k = \mathbb{E}(\chi_k) = \mathbb{P}(\chi_k = 1)$. We aim to efficiently infer each θ_k by interacting with the environment and observing the outcomes. In a MAB there are multiple arms (i.e., actions) each generating a random reward according to a probability distribution function (PDF). An agent sequentially chooses an action $x_t = k$ for $t = 1, \dots$ and readjusts its strategy by observing the reward with the hope of maximizing its expected reward. In our problem, there are K actions. The WPD keeps initial estimates of $\hat{\theta}_k$ about the unknown parameters θ_k . The WPD chooses an action $x_t = k$ and observes the event $Z(L_k) \cdot \chi_k$. Based on the observation, it updates $\hat{\theta}_k$ until the algorithm converges to the optimal value. The typical method for optimizing a MAB problem is by the well known ϵ -greedy algorithm presented in Algorithm 4. The ϵ -greedy algorithm consists of two steps; exploration and exploitation. Exploration improves the estimate of non-greedy actions' values while

⁵The sensing formulation can be generalized beyond the indicator function for a utility function generating rewards with a support in $[0, 1]$.

⁶Note that a better strategy is to choose the size of the data after observing the amount of harvested energy and the duration of IT period. Since the amount of harvested energy is independent upon each observation, we can easily extend the framework by considering a contextual multi armed bandit problem.

exploitation is favorable when we reach a sufficient knowledge about the estimate of actions. ϵ -greedy algorithm, with probability (w.p.) $1 - \epsilon$, greedily chooses an action k that maximizes $Z(L_k)\hat{\theta}_k$ and w.p. ϵ randomly chooses an action. In other words, w.p. ϵ the algorithm explores in the action space of the MAB while w.p. $1 - \epsilon$ the algorithm exploits what it already knows. Although such an approach is guaranteed to approach the optimal performance [14], provided that ϵ is sufficiently small, the convergence rate of the algorithm is poor. This is because ϵ -greedy algorithm does not judiciously explore in the parameter space. To speed up the convergence, we use a Bayesian inference method to judiciously explore in the action space of the MAB problem. The augmentation of the Bayesian framework in MAB is known as Thompson sampling (TS)⁷ [124]. To see how TS works, let us model the uncertainty θ_k by assuming a prior distribution for it. Each θ_k is distributed according to a Beta distribution with parameters a_k and b_k . In particular, for each arm k , the prior probability density function of θ_k is:

$$\mathbb{P}(\theta_k) = \frac{\Gamma(a_k + b_k)}{\Gamma(a_k)\Gamma(b_k)} \theta_k^{a_k-1} (1 - \theta_k)^{b_k-1}, \quad (6.57)$$

where $\Gamma(\cdot)$ denotes the gamma function. The reason for choosing Beta as prior distribution is the conjugacy property of Beta distribution with Bernoulli distribution. In other words, if prior is Beta distributed and the likelihood is Bernoulli distributed, then the posterior distribution is also Beta distributed. This facilitates the process of sampling from the posterior distribution⁸. Given a sample realization of χ_k , we are interested in updating the posterior distribution of θ_k . We have:

$$\begin{aligned} \mathbb{P}(\theta_k | \chi_k) &\propto \mathbb{P}(\theta_k) \mathbb{P}(\chi_k | \theta_k) \\ &= \frac{\theta_k^{a_k-1} (1 - \theta_k)^{b_k-1}}{B(a_k, b_k)} \theta_k^{\chi_k} (1 - \theta_k)^{1-\chi_k} \\ &\propto \theta_k^{a_k-1+\chi_k} (1 - \theta_k)^{b_k-1+1-\chi_k} \end{aligned} \quad (6.58)$$

Hence, the posterior distribution is also Beta distributed with parameters, $a_k + \mathbb{1}_{\{\chi_k=1\}}$ and $b_k + \mathbb{1}_{\{\chi_k=0\}}$. Note that at any given time, only a single observation regarding the chosen

⁷See [122, 123] for the optimality analysis of TS.

⁸Note that the conjugacy property only makes it easier to sample from the posterior distribution. In case where the posterior distribution does not admit any known PDF, efficient Monte-Carlo methods such as Markov chain Monte-Carlo (MCMC) [125] method and its variants such as Gibbs sampling can be used to efficiently sample from the posterior.

action is revealed. Hence, after retrieving the observation about an action, the parameters of the posterior distribution is updated as:

$$(a_k, b_k) \leftarrow \begin{cases} (a_k, b_k) & \text{if } x_t \neq k, \\ (a_k + \chi_k, b_k + 1 - \chi_k) & \text{if } x_t = k. \end{cases} \quad (6.59)$$

The TS algorithm is given in Algorithm 5. Note that the only difference between the TS and ϵ -greedy algorithms in the exploration phase of the problem. TS judiciously explores by modeling the uncertainty of each action using a distribution with decreasing variance in the number of observations explored. This prevents the TS from exploring the actions that are believed to be sub-optimal. Meanwhile ϵ -greedy explores the action space randomly, reducing the efficiency of the exploration phase.

Algorithm 4 ϵ -greedy

- 1: **for** $t = 1, 2, \dots$ **do**
 - 2: With probability ϵ
 - 3: **for** $k=1, \dots, K$ **do**
 - 4: $\hat{\theta}_k = \frac{a_k}{a_k + b_k}$
 - 5: $x_t \leftarrow \begin{cases} \arg \max_k Z(L_k) \hat{\theta}_k & \text{with prob. } 1 - \epsilon, \\ \text{choose a random action} & \text{with prob. } \epsilon. \end{cases}$
 - 6: Apply x_t and observe χ_k
 - 7: update the posterior using (6.59)
-

Algorithm 5 Thompson Sampling (TS)

- 1: **for** $t = 1, 2, \dots$ **do**
 - 2: Sample from the posterior
 - 3: **for** $k=1, \dots, K$ **do**
 - 4: Sample $\hat{\theta}_k \sim \text{beta}(a_k, b_k)$
 - 5: $x_t \leftarrow \arg \max_k Z(L_k) \hat{\theta}_k$
 - 6: Apply x_t and observe χ_k
 - 7: update the posterior using (6.59)
-

6.6 Numerical Results

In this section, we compare the performance of the optimal online policy with that of the offline as well as two benchmark policies, namely *uniform* and *power-halving* policies. In uniform policy, the amount of harvested energy is uniformly distributed in the IT period.

Power-halving policy allocates half of its available energy in each time slot in the IT period. The WPT duration for both uniform and power-halving policies is optimized using exhaustive search method. We also evaluate the performance of TS algorithm in the sensing utility maximization problem developed in Section 6.5 and compare it with that of ϵ -greedy.

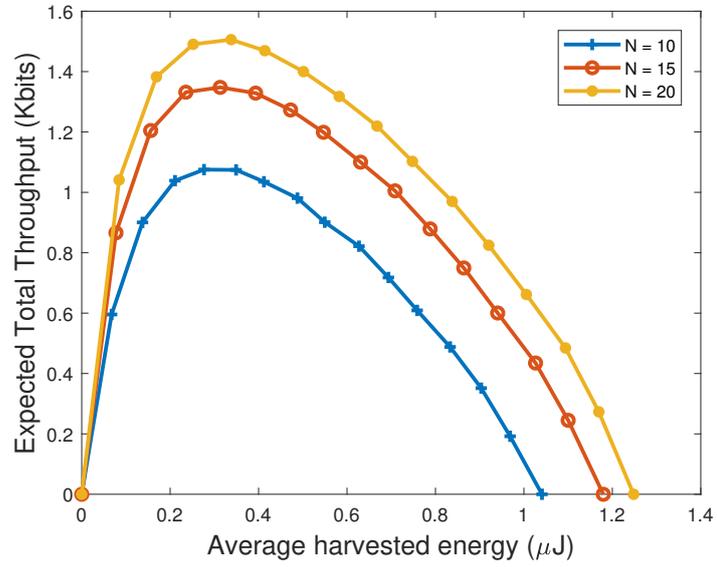
For the channel state, we assume two different channel models based on Rayleigh and Gilbert-Elliot (G-E) fading models. For Rayleigh fading, we assume an average channel gain of 1. For G-E model, we assume that there are two state; good and bad. The gain of good state is 1 and that of bad state is 0. The good and bad states occur with probability of 0.6 and 0.4, respectively. We assume that the AP transmits with power $P = 20\text{dBm}$ which is normalized with respect to distance and EH efficiency. Time slot duration is 1ms, the bandwidth is assumed to be 2KHz, and the noise power density is 176 dBm/Hz.

6.6.1 Rate-Energy Trade-off

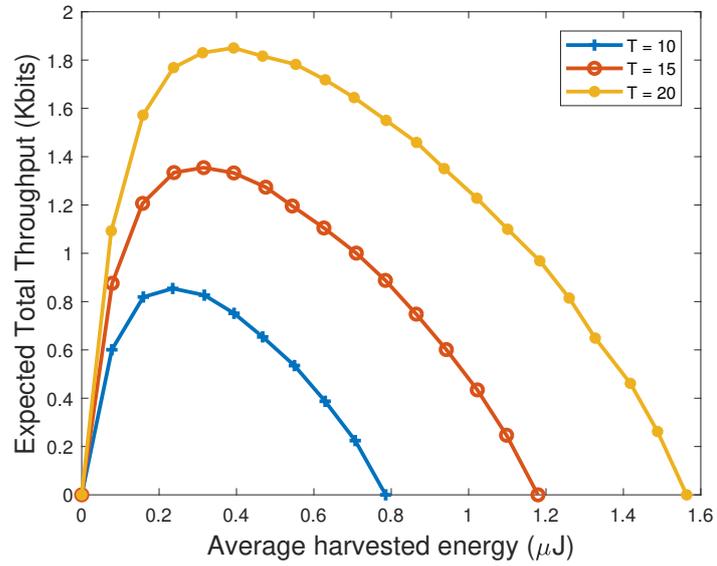
We first evaluate the rate-energy trade-off of the online policy which is the expected total number of bits transmitted with respect to the amount of harvested energy in a finite duration of T . In Figure 6.4a, for different values of channel discretization level, N , and a frame length of 15 time slots, the rate-energy trade-off is depicted. For different values of T , and $N = 15$, Figure 6.4b, illustrates the rate-energy trade-off. We observe from the figures that, spending too much time for transmitting more energy in the EH period reduces the time for IT period which in turn reduces the throughput. On the other hand, if we reduce the EH period, there would be less energy in the IT period resulting in a reduced throughput. Hence, an optimal balance is required.

6.6.2 Performance Evaluation

In Figure 6.5, when the fading is Rayleigh, the expected total number of bits that are transmitted in 100 time slots is depicted with respect to the number of channel discretization levels, N . We observe that as the number of channel levels increases, the discretization error decreases and hence the throughput of the all policies improve. The online policy achieves a throughput close to the upper-bound by optimally determining the WPT duration and power allocation in the IT period. Although the uniform and power-halving



(a) Expected throughput with respect to N .



(b) Expected throughput with respect to T .

Figure 6.4: The effect of channel discretization and deadline duration on the expected throughput.

policies harvest energy for an optimum duration, they considerably perform poor due to the blind power allocation in the IT period.

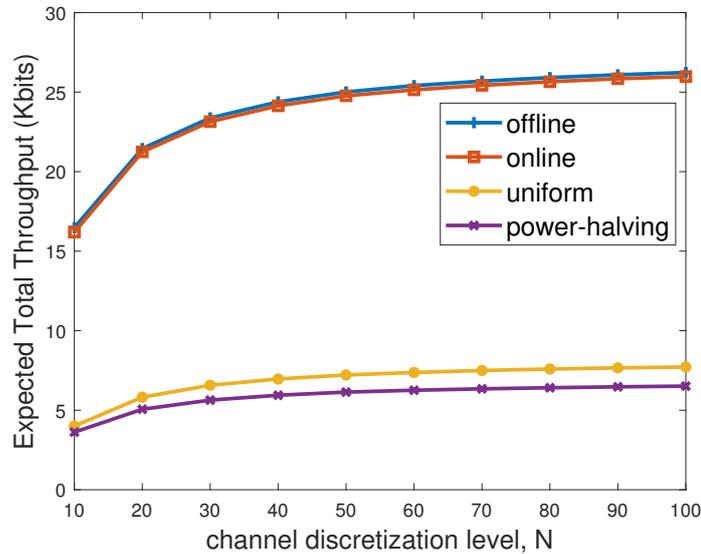
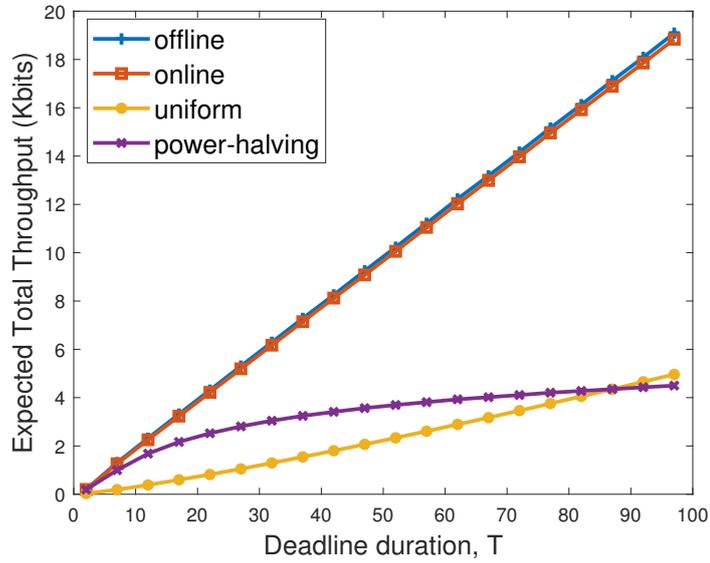


Figure 6.5: Expected total throughput of the WPD with respect to the number of channel discretization levels in $T = 100$ time slots.

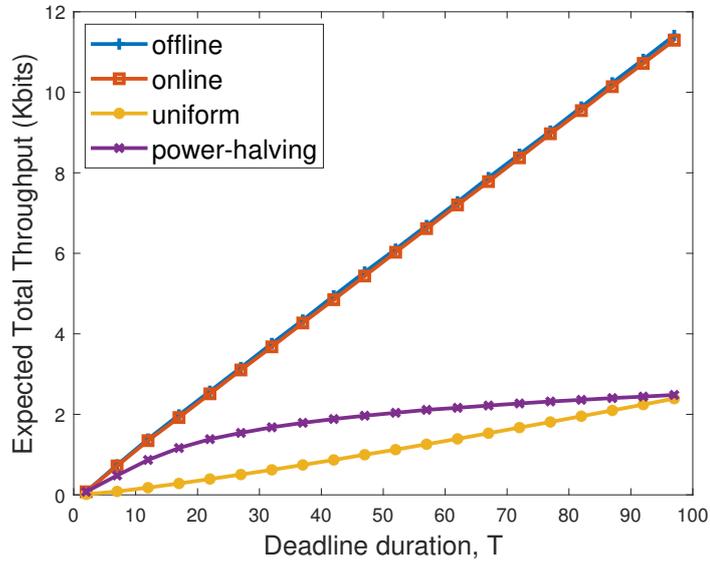
Next, we plot the expected total throughput of the WPD under Rayleigh and G-E fading models in Figure 6.6a and Figure 6.6b, respectively. Again, the online policy, for all values of T , achieves an outstanding performance compared to the offline policies. For smaller values of T , the power-halving policy achieves a good performance. However, as T increases, due to the concave nature of the rate-power function, the power-halving strategy becomes significantly inefficient. On the other hand, uniform policy is able to perform better, for larger values of T , with respect to power-halving policy by allocating the harvested energy uniformly across the IT period. Finally, we illustrate the transmission rate of the WPD in units of bits per seconds (bits/sec) in Figure 6.7. It can be seen from both Figure 6.7a and Figure 6.7b that the online policy has a significantly higher rate than the uniform and power-halving policies. It is also evident that on the average, the online policy achieves a significantly good performance with respect to the offline policy.

6.6.3 MAB

Here, we evaluate the performance of TS and ϵ -greedy algorithms and compare their performance. In Figure 6.8, we plot the per-period regret of both algorithms. For plots, we use the following synthetic parameters; $T = 15$, $N = 30$, $L = 1000, 2500, 3000$ bits,

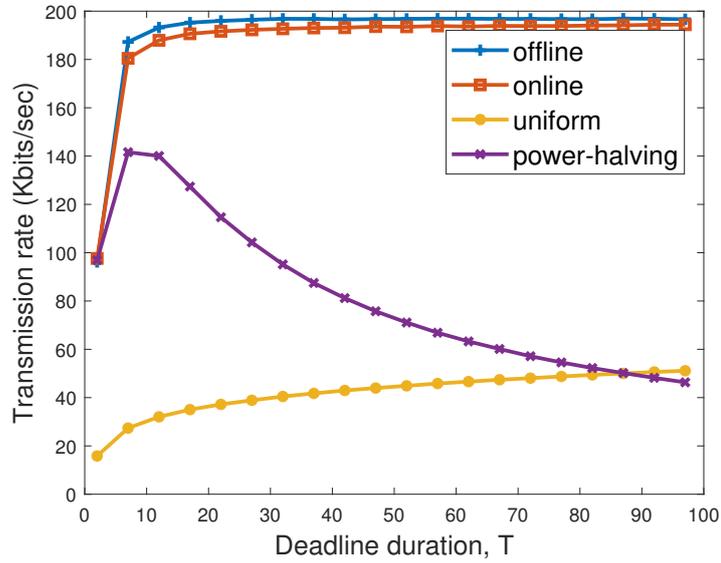


(a) Expected total throughput with respect to T under Rayleigh fading.

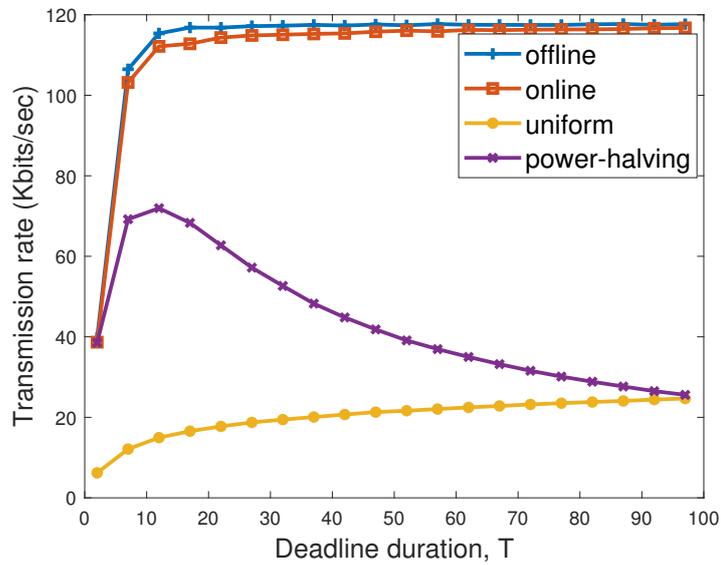


(b) Expected total throughput with respect to T under G-E model.

Figure 6.6: Expected total throughput of the WPD with $N = 20$ channel levels with respect to the frame length, T .



(a) Expected transmission rate of the WPD with respect to T under Rayleigh fading.



(b) Expected transmission rate of the WPD with respect to T under G-E model.

Figure 6.7: Expected transmission rate of the WPD with $N = 20$ channel levels with respect to the frame length, T .

$Z = 500, 700, 750$, and $\mathcal{E} = 1, 3, 4 \mu\text{Joules}$. Per-period regret is the gap between the optimal utility and the utility achieved by the given algorithm. We obtain the value of the optimal utility by exhaustive search for comparison purposes only. Each point in Figure 6.8 is averaged over 10^5 samples.

The greedy algorithm ($\epsilon = 0$) has the worst performance as it does not explore at all. By giving non-zero values for ϵ , we can see that 0.05-greedy and 0.1-greedy greatly improve upon the greedy algorithm by performing explorations. However, we see a poor performance regarding their convergence rate. TS improves the convergence rate significantly by simply adding intelligence to the exploration phase. This makes the TS algorithm to approach a per-period regret of 0 considerably faster than the ϵ -greedy algorithm.

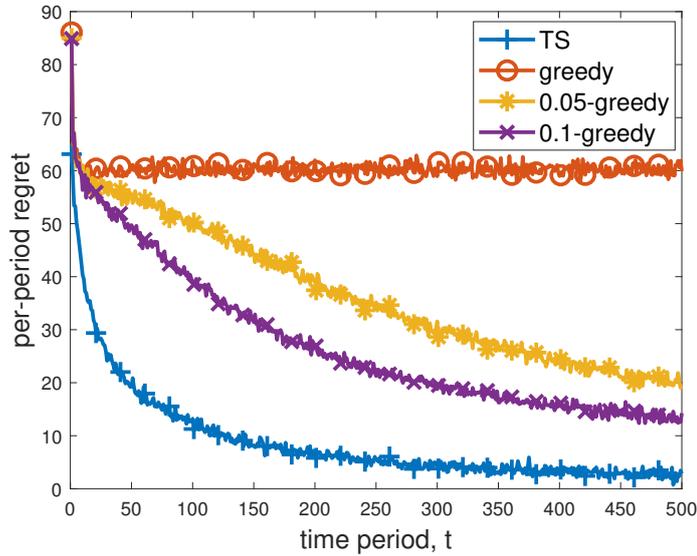


Figure 6.8: Per-period regret comparison of TS and ϵ -greedy algorithms for $\epsilon = 0, 0.05, 0.1$.

6.7 Chapter Summary

In this work, we studied the a WPCN scenario operating in a finite horizon. An AP transmits an RF signal to energize an WPD for a certain duration (WPT period) and then it stops sending energy and collects data from the WPD in the remainder of the horizon (IT period). The wireless channel varies randomly over the horizon and it is only available to the WPD causally and only in the IT period. We first derived an upper-bound

for the performance of the network in an offline manner by assuming that the channel realizations are available non-causally at the AP. We then studied the online counterpart of the problem by assuming that the channel realization are available only causally and in the IT period. We show that there exist a time-dependent threshold on the energy level of the WPD in which it is optimal to stop WPT and start the IT period. Then, we show that the optimal power allocation in the IT period follows a fractional structure in which the WPD at each time slot allocates a fraction of its energy that depends on the immediate channel state as well as certain measure of future expectations. The numerical results show that the online policy achieves a performance significantly close to the upper-bound. We then extended the model by embedding a MEC unit at the AP capable of performing heavy computational tasks. We formulated a Bayesian inference reinforcement learning problem to address the dependency of the application performance coupled with that of physical layer. We show that the Bayesian inference achieves a convergence rate that is much faster than that of the ϵ -greedy algorithm. In the future, we aim to extend the results to the case of multiple WPDs.

Chapter 7

Conclusions and Future Works

In this thesis, we have studied a series of fundamental problems of wireless communication in EH networks and have addressed the vision of future wireless technologies. More specifically, challenges associated with multiple access scheme, random nature of wireless channel, and reliability has been investigated in Chapter 3, Chapter 4, and Chapter 5, respectively. The scope of Chapter 6 serves to fulfill the service based vision of next generation wireless technologies.

In this thesis, we have aimed at revisiting some of the most fundamental challenges of wireless communication in energy harvesting networks. The inherent randomness in the amount and availability of energy introduces many challenges in the performance of the energy constrained devices. To this end, we first studied a multiple access strategy for an EH network when the harvested energy is correlated across the devices. We aimed at utilizing this information in order to design a low complexity random access protocol suitable for resource limited energy harvesting devices. For a low complexity threshold policy based on the battery state of the devices, we developed analytical expressions for the sum throughput of the network. Through deriving analytical expressions and the subsequent optimization, we showed that the correlation across devices has an important impact on the performance of the random access protocol. More specifically, when the EH process across the two EH sensors are negatively correlated, such that at each time slot only one of them harvest energy, we saw that the random access policy is optimal as it allows for the most frequent transmission opportunity without any collisions. On the other hand, when the EH process is positively correlated, so that either both sensors harvest energy at a given time or none of them, the threshold based access policy results in unfavorable

properties by maximizing the throughput of only one device while inducing collision on the other device. This effect will become more severe when the number of nodes increase in the network. In conclusion, when the EH process is negatively correlated it is favorable to use threshold based policy while in positively correlated scenario it is better to consume extra energy to avoid collisions. A future study direction is to design hybrid policies that could dynamically choose an access scheme based on the correlation information. Such policies can be learned using the machine learning techniques. Another direction is to study beam sweeping optimization in WPT systems to generate negatively correlated EH process across different sensors by carefully taking into account the delay requirements of the sensors to schedule energy transfer.

The effect of the correlation information in the performance of the study in Chapter 3 motivated us to further study the utilization of correlated information to improve the performance of the wireless communication for EH devices. Thus, we studied channel state acquisition policy that efficiently uses the time correlated information about the channel to only sense the channel when it is necessary. We studied an EH transmitter that aims at maximizing its throughput by dynamically adjusting its transmission rate. If the chosen rate is higher than the channel capacity the transmission fails. Thus, the channel state plays an important role in adapting transmission rate albeit inducing energy and time overheads. To balance the trade-off between the costs and benefits of channel sensing, we formulated the problem as a partially observable Markov decision process (POMDP) and converted it to an ordinary MDP by introducing a belief parameter on the channel state. We observed that due to the continuous nature of the belief the standard numerical solutions such as value iteration algorithm become too complicated. Thus, we analytically proved that a threshold policy on the belief about the channel state is optimal. Thus, at a given battery state, we need to optimize at most three thresholds which significantly reduces the complexity of the solution. We compared the throughput achieved by the optimal policy to those achieved by a greedy policy and a single-threshold policy, which do not exploit the channel sensing capability, as well as an opportunistic policy, which senses the channel at every time slot. We have shown through simulations that the intelligent channel sensing capability improves the performance significantly, thanks to the increased adaptability to channel conditions. The gain of the intelligent channel sensing policy become more substantial as the sensing cost increases. The adopted two state rate

model has enabled a deep insight into the dynamic channel sensing problem by enabling a simple threshold policy to be optimal. Note that in general multi state channel models it may be impossible to find a simple but optimal channel sensing mechanism. As a continuation of this work, it is interesting to consider an intelligent rate adaptive algorithm by introducing multidimensional belief vector on multiple channel states. As a promising solution approach, deep reinforcement learning (DRL) can be used to train the transmitter to choose an optimal action by observing the dynamic belief vector.

Next, we addressed enabling reliability within the EH networks by considering a HARQ enabled EH receiver. To this end, we considered a SWIPT system in which an energy abundant transmitter transmits energy using information bearing signals to an EH receiver. For a reliable transmission of information, we specifically considered widely used HARQ with incremental redundancy. We aimed at minimizing the number of re-transmissions triggered by erroneous transmission by optimally splitting the incoming RF signal for EH or ID purposes. We considered energy aspects of employing HARQ on the EH receiver by modeling the energy consumption induced on the receiver. Due to continuity of action and state spaces, we aimed at finding a structure for an optimal policy rather than using expensive numerical solutions. First, we reduced the state and action spaces of the problem to discrete ones, thus, greatly reducing the complexity. Second, we characterized a family of low complexity policies that can achieve minimum number of retransmission on the average. The results show that simple to implement policies suitable for EH devices can be implemented for enabling a reliable link over unreliable wireless medium. For a conclusive study, we believe two main extensions should be addressed. First, the general non-linear energy consumption models are required to reflect a more general setting. Second, the wireless channel should be extended for multi state fading channel.

In the second part of the thesis, we took a holistic view of the next generation of wireless technology to address service-based optimization for a wireless powered device (WPD). We considered a WPD that provides data at adjustable resolution settings for an application that exists at a remote access point (AP). Depending on the quality of the data as measured by its resolution setting, different utility metrics is achieved at the remote application. The WPD experiences a high energy consumption profile and reduced chance of packet delivery when sensing the data at a higher resolution while the application en-

joys a better utility with high resolution data. We aimed at optimizing this trade-off in a finite horizon of time slots over a fading channel by optimizing the sensing resolution of the WPD. Towards this, for a given packet with arbitrary length, we maximized the packet delivery chance by optimizing the charging duration and power allocation for transmission. Next, we formulated the problem of optimal sensing resolution as a multi-armed bandit problem and used Thompson sampling with fast convergence properties to solve the problem in an iterative framework. This approach addresses the vision of an intelligent edge in the wireless networks. Considering the privacy issues of sharing personal data with a remote AP, an important extension is to study the on-device computation for energy limited devices. Recently, neural networks (NNs) has shown a promising performance in many difficult tasks. The structure of the NNs is naturally suitable for parallel computation enabling distributed intelligence across many devices. With this, another interesting extension is to develop networking for a distributed intelligence comprised of EH devices without a need for centralized entities.

Bibliography

- [1] C. You, Y. Zeng, R. Zhang, and K. Huang, “Asynchronous mobile-edge computation offloading: Energy-efficient resource management,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7590–7605, Nov 2018.
- [2] V. Rajendran, J. J. Garcia-Luna-Aveces, and K. Obraczka, “Energy-efficient, application-aware medium access for sensor networks,” in *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, 2005.*, Nov 2005, pp. 8 pp.–630.
- [3] S. C. Ergen, C. Fischione, D. Marandin, and A. Sangiovanni-Vincentelli, “Duty-cycle optimization in unslotted 802.15.4 wireless sensor networks,” in *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, Nov 2008, pp. 1–6.
- [4] T. Arici and Y. Altunbasak, “Adaptive sensing for environment monitoring using wireless sensor networks,” in *2004 IEEE Wireless Communications and Networking Conference (IEEE Cat. No.04TH8733)*, vol. 4, March 2004, pp. 2347–2352 Vol.4.
- [5] D. Anthony, W. Bennett, M. Vuran, M. Dwyer, S. Elbaum, A. Lacy, M. Engels, and W. Wehtje, “Sensing through the continent: Towards monitoring migratory birds using cellular sensor networks,” in *Information Processing in Sensor Networks (IPSN), 2012 ACM/IEEE 11th International Conference on*, April 2012, pp. 329–340.
- [6] J. A. Paradiso and T. Starner, “Energy scavenging for mobile and wireless electronics,” *IEEE Pervasive Computing*, vol. 4, no. 1, pp. 18–27, Jan. 2005.

- [7] D. Niyato, E. Hossain, M. Rashid, and V. Bhargava, “Wireless sensor networks with energy harvesting technologies: a game-theoretic approach to optimal energy management,” *Wireless Communications, IEEE*, vol. 14, no. 4, pp. 90–96, August 2007.
- [8] G. Park, T. Rosing, M. D. Todd, C. R. Farrar, and W. Hodgkiss, “Energy harvesting for structural health monitoring sensor networks,” *Journal of Infrastructure Systems*, vol. 14, no. 1, pp. 64–79, Mar. 2008.
- [9] M. Anteppli, E. Uysal-Biyikoglu, and H. Erkal, “Optimal packet scheduling on an energy harvesting broadcast link,” *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 8, pp. 1721–1731, September 2011.
- [10] B. Devillers and D. Gündüz, “A general framework for the optimization of energy harvesting communication systems with battery imperfections,” *Journal of Communications and Networks*, vol. 14, no. 2, pp. 130–139, April 2012.
- [11] Z. Wang, A. Tajer, and X. Wang, “Communication of energy harvesting tags,” *Communications, IEEE Transactions on*, vol. 60, no. 4, pp. 1159–1166, April 2012.
- [12] A. Aprem, C. Murthy, and N. Mehta, “Transmit power control policies for energy harvesting sensors with retransmissions,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 7, no. 5, pp. 895–906, Oct 2013.
- [13] J. Lei, R. Yates, and L. Greenstein, “A generic model for optimizing single-hop transmission policy of replenishable sensors,” *Wireless Communications, IEEE Transactions on*, vol. 8, no. 2, pp. 547–551, Feb 2009.
- [14] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998. [Online]. Available: <http://www.cs.ualberta.ca/~Eesutton/book/ebook/the-book.html>
- [15] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Athena Scientific, 2000.
- [16] P. Blasco, D. Gündüz, and M. Dohler, “A learning theoretic approach to energy harvesting communication system optimization,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1872–1882, April 2013.

- [17] H. Ju and R. Zhang, "Throughput maximization in wireless powered communication networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 1, pp. 418–428, January 2014.
- [18] X. Di, K. Xiong, P. Fan, H. Yang, and K. B. Letaief, "Optimal resource allocation in wireless powered communication networks with user cooperation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 7936–7949, Dec 2017.
- [19] D. Xu and Q. Li, "Joint power control and time allocation for wireless powered underlay cognitive radio networks," *IEEE Wireless Communications Letters*, vol. 6, no. 3, pp. 294–297, June 2017.
- [20] J. Trinnaman and A. Clarke, *2004 Survey of energy resources*. Elsevier, 2004.
- [21] J. Huang, R. A. Berry, and M. L. Honig, "Wireless scheduling with hybrid ARQ," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2801–2810, Nov 2005.
- [22] J. Yang and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system," *IEEE Transactions on Communications*, vol. 60, no. 1, pp. 220–230, January 2012.
- [23] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: Optimal policies," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1732–1743, September 2011.
- [24] F. M. Ozcelik, G. Uctu, and E. Uysal-Biyikoglu, "Minimization of transmission duration of data packets over an energy harvesting fading channel," *IEEE Communications Letters*, vol. 16, no. 12, pp. 1968–1971, December 2012.
- [25] M. L. Ku, Y. Chen, and K. J. R. Liu, "Data-driven stochastic models and policies for energy harvesting sensor communications," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 8, pp. 1505–1520, Aug 2015.

- [26] K. J. Prabuchandran, S. K. Meena, and S. Bhatnagar, "Q-learning based energy management policies for a single sensor node with finite buffer," *IEEE Wireless Communications Letters*, vol. 2, no. 1, pp. 82–85, February 2013.
- [27] D. Zhao, C. Huang, Y. Chen, F. Alsaadi, and S. Cui, "Resource allocation for multiple access channel with conferencing links and shared renewable energy sources," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 423–437, March 2015.
- [28] Q. Sun, G. Zhu, C. Shen, X. Li, and Z. Zhong, "Joint beamforming design and time allocation for wireless powered communication networks," *IEEE Communications Letters*, vol. 18, no. 10, pp. 1783–1786, Oct 2014.
- [29] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in ofdma systems with hybrid energy harvesting base station," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3412–3427, July 2013.
- [30] A. M. Ibrahim, O. Ercetin, and T. ElBatt, "Stability analysis of slotted Aloha with opportunistic RF energy harvesting," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1477–1490, May 2016.
- [31] M. L. Ku, Y. Chen, and K. J. R. Liu, "Data-driven stochastic models and policies for energy harvesting sensor communications," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 8, pp. 1505–1520, Aug. 2015.
- [32] N. Michelusi, K. Stamatiou, and M. Zorzi, "Transmission policies for energy harvesting sensors with time-correlated energy supply," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2988–3001, Jul. 2013.
- [33] A. Hentati, F. Abdelkefi, and W. Ajib, "Energy allocation for sensing and transmission in WSNs with energy harvesting Tx/Rx," in *IEEE Vehicular Technology Conf. (VTC Fall)*, Sep. 2015, pp. 1–5.
- [34] S. Mao, M. H. Cheung, and V. W. S. Wong, "Joint energy allocation for sensing and transmission in rechargeable wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 6, pp. 2862–2875, Jul. 2014.

- [35] P. Blasco and D. Gündüz, “Multi-access communications with energy harvesting: A multi-armed bandit model and the optimality of the myopic policy,” *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 585–597, Mar. 2015.
- [36] C. K. Ho and R. Zhang, “Optimal energy allocation for wireless communications with energy harvesting constraints,” *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4808–4818, Sep. 2012.
- [37] B. T. Bacinoglu and E. Uysal-Biyikoglu, “Finite-horizon online transmission scheduling on an energy harvesting communication link with a discrete set of rates,” *Journal of Communications and Networks*, vol. 16, no. 3, pp. 393–300, Jun. 2014.
- [38] R. Srivastava and C. E. Koksal, “Basic performance limits and tradeoffs in energy-harvesting sensor nodes with finite data and energy storage,” *IEEE/ACM Trans. Netw.*, vol. 21, no. 4, pp. 1049–1062, Aug. 2013. [Online]. Available: <http://dx.doi.org/10.1109/TNET.2012.2218123>
- [39] D. Shaviv and A. Özgür, “Universally near optimal online power control for energy harvesting nodes,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3620–3631, Dec 2016.
- [40] R. D. Yates and H. Mahdavi-Doost, “Energy harvesting receivers: Packet sampling and decoding policies,” *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 558–570, March 2015.
- [41] A. Yadav, M. Goonewardena, W. Ajib, O. A. Dobre, and H. Elbiaze, “Energy management for energy harvesting wireless sensors with adaptive retransmission,” *IEEE Transactions on Communications*, vol. 65, no. 12, pp. 5487–5498, Dec 2017.
- [42] H. Mahdavi-Doost and R. D. Yates, “Hybrid ARQ in block-fading channels with an energy harvesting receiver,” in *2015 IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, June 2015, pp. 1144–1148.
- [43] Z. Ni, R. V. Bhat, and M. Motani, “Performance of energy-harvesting receivers with batteries having internal resistance,” in *2017 IEEE Wireless Communications*

- and Networking Conference Workshops (WCNCW)*, San Francisco, USA, March 2017, pp. 1–6.
- [44] Z. Ni and M. Motani, “Transmission schemes and performance analysis for time-switching energy harvesting receivers,” in *2016 IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.
- [45] S. Zhou, T. Chen, W. Chen, and Z. Niu, “Outage minimization for a fading wireless link with energy harvesting transmitter and receiver,” *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 496–511, March 2015.
- [46] M. K. Sharma and C. R. Murthy, “On the design of dual energy harvesting communication links with retransmission,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 4079–4093, June 2017.
- [47] Y. Mao, J. Zhang, and K. B. Letaief, “ARQ with adaptive feedback for energy harvesting receivers,” in *2016 IEEE Wireless Communications and Networking Conference*, Doha, Qatar, April 2016, pp. 1–6.
- [48] B. Makki, T. Svensson, and M. Zorzi, “A joint power and information transfer system using retransmissions,” in *2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–7.
- [49] A. Mayberry, P. Hu, B. Marlin, C. Salthouse, and D. Ganesan, “ishadow: Design of a wearable, real-time mobile gaze tracker,” in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '14. New York, NY, USA: ACM, 2014, pp. 82–94. [Online]. Available: <http://doi.acm.org/10.1145/2594368.2594388>
- [50] N. Michelusi and M. Zorzi, “Optimal random multiaccess in energy harvesting wireless sensor networks,” in *Communications Workshops (ICC), 2013 IEEE International Conference on*, June 2013, pp. 463–468.
- [51] H. Li, C. Huang, P. Zhang, S. Cui, and J. Zhang, “Distributed opportunistic scheduling for energy harvesting based wireless networks: A two-stage probing approach,” *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1618–1631, June 2016.

- [52] P. Blasco and D. Gündüz, “Multi-access communications with energy harvesting: A multi-armed bandit model and the optimality of the myopic policy,” *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 585–597, March 2015.
- [53] D. Gündüz, K. Stamatiou, N. Michelusi, and M. Zorzi, “Designing intelligent energy harvesting communication systems,” *IEEE Communications Magazine*, vol. 52, no. 1, pp. 210–216, Jan. 2014.
- [54] H. Li, C. Huang, P. Zhang, S. Cui, and J. Zhang, “Distributed opportunistic scheduling for energy harvesting based wireless networks: A two-stage probing approach,” *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1618–1631, Jun. 2016.
- [55] B. Devillers and D. Gündüz, “A general framework for the optimization of energy harvesting communication systems with battery imperfections,” *Journal of Communications and Networks*, vol. 14, no. 2, pp. 130–139, Apr. 2012.
- [56] Q. Zhang and S. A. Kassam, “Finite-state Markov model for Rayleigh fading channels,” *IEEE Trans. on Communs*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.
- [57] E. N. Gilbert, “Capacity of a burst-noise channel,” *The Bell System Technical Journal*, vol. 39, no. 5, pp. 1253–1265, Sep. 1960.
- [58] M. Kashef and A. Ephremides, “Optimal packet scheduling for energy harvesting sources on time varying wireless channels,” *Journal of Communications and Networks*, vol. 14, no. 2, pp. 121–129, Apr. 2012.
- [59] A. Aprem, C. R. Murthy, and N. B. Mehta, “Transmit power control policies for energy harvesting sensors with retransmissions,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 895–906, Oct. 2013.
- [60] Q. Zhao, L. Tong, A. Swami, and Y. Chen, “Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

- [61] A. Laourine and L. Tong, “Betting on Gilbert-Elliot channels,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 723–733, Feb. 2010.
- [62] R. Gangula, D. Gesbert, and D. Gündüz, “Optimization of energy harvesting miso communication system with feedback,” *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 396–406, Mar. 2015.
- [63] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, “Optimal energy management policies for energy harvesting sensor nodes,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1326–1336, Apr. 2010.
- [64] W. S. Lovejoy, “A survey of algorithmic methods for partially observed Markov decision processes,” *Annals of Operations Research*, vol. 28, no. 1, pp. 47–65, Dec. 1991.
- [65] P. Blasco, D. Gündüz, and M. Dohler, “A learning theoretic approach to energy harvesting communication system optimization,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1872–1882, Apr. 2013.
- [66] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1994.
- [67] M. P. Deisenroth, G. Neumann, and J. Peters, “A survey on policy search for robotics,” *Found. Trends Robot*, vol. 2, no. 2, pp. 1–142, Aug. 2013. [Online]. Available: <http://dx.doi.org/10.1561/23000000021>
- [68] L. R. Varshney, “Transporting information and energy simultaneously,” in *2008 IEEE International Symposium on Information Theory*. Auckland, New Zealand: IEEE, 2008, pp. 1612–1616.
- [69] P. Grover and A. Sahai, “Shannon meets tesla: Wireless information and power transfer,” in *International Symposium on Information Theory*, Austin, Texas, 2010, pp. 2363–2367.
- [70] R. Zhang and C. K. Ho, “MIMO broadcasting for simultaneous wireless information and power transfer,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 1989–2001, 2013.

- [71] F. A. de Witt, R. D. Souza, and G. Brante, "On the performance of hybrid ARQ schemes for uplink information transmission with wireless power transfer in the downlink," in *2014 IFIP Wireless Days (WD)*, Rio de Janeiro, Brazil, 2014, pp. 1–6.
- [72] H. Chen, R. G. Maunder, and L. Hanzo, "A survey and tutorial on low-complexity turbo coding techniques and a holistic hybrid ARQ design example," *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 1546–1566, Fourth 2013.
- [73] M. Zohdy, T. ElBatt, M. Nafie, and O. Ercetin, "RF energy harvesting in wireless networks with HARQ," in *2016 IEEE Globecom Workshops*, Washington, DC USA, Dec 2016, pp. 1–6.
- [74] B. Makki, T. Svensson, and M. Zorzi, "Wireless energy and information transmission using feedback: Infinite and finite block-length analysis," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 5304–5318, Dec 2016.
- [75] Y. Zeng, B. Clerckx, and R. Zhang, "Communications and signals design for wireless power transmission," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 2264–2290, May 2017.
- [76] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-energy tradeoff," *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4754–4767, November 2013.
- [77] L. Liu, R. Zhang, and K. C. Chua, "Wireless information and power transfer: A dynamic power splitting approach," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3990–4001, Sept. 2013.
- [78] C. Shen, W. C. Li, and T. H. Chang, "Wireless information and energy transfer in multi-antenna interference channel," *IEEE Transactions on Signal Processing*, vol. 62, no. 23, pp. 6249–6264, Dec 2014.
- [79] J. Park and B. Clerckx, "Transmission strategies for joint wireless information and energy transfer in a two-user MIMO interference channel," in *2013 IEEE International Conference on Communications Workshops (ICC)*, Budapest, Hungary, June 2013, pp. 591–595.

- [80] Y. Sarikaya and O. Ercetin, "Self-sufficient receiver with wireless energy transfer in a multi-access network," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 442–445, Aug 2017.
- [81] M. A. Abd-Elmagid, A. Biazon, T. ElBatt, K. G. Seddik, and M. Zorzi, "On optimal policies in full-duplex wireless powered communication networks," in *2016 14th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2016, pp. 1–7.
- [82] M. R. Zenaïdi, Z. Rezkî, and M. S. Alouini, "Performance limits of online energy harvesting communications with noisy channel state information at the transmitter," *IEEE Access*, vol. 5, pp. 1239–1249, 2017.
- [83] R. Zhang and C. K. Ho, "Mimo broadcasting for simultaneous wireless information and power transfer," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 1989–2001, May 2013.
- [84] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Relaying protocols for wireless energy harvesting and information processing," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3622–3636, July 2013.
- [85] S. B. Wicker, *Error control systems for digital communication and storage*. Prentice hall Englewood Cliffs, 1995, vol. 1.
- [86] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the gaussian collision channel," *IEEE Transactions on Information Theory*, vol. 47, no. 5, pp. 1971–1988, Jul 2001.
- [87] N. Varnica, E. Soljanin, and P. Whiting, "LDPC code ensembles for incremental redundancy hybrid ARQ," in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, Adelaide, Australia, Sept 2005, pp. 995–999.
- [88] V. Talla, B. Kellogg, B. Ransford, S. Naderiparizi, S. Gollakota, and J. R. Smith, "Powering the next billion devices with wi-fi," in *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, ser. CoNEXT '15. New York, NY, USA: ACM, 2015, pp. 4:1–4:13. [Online]. Available: <http://doi.acm.org/10.1145/2716281.2836089>

- [89] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, “Wireless networks with rf energy harvesting: A contemporary survey,” *IEEE Communications Surveys Tutorials*, vol. 17, no. 2, pp. 757–789, 2015.
- [90] F. Rosas, R. D. Souza, M. E. Pellenz, C. Oberli, G. Brante, M. Verhelst, and S. Pollin, “Optimizing the code rate of energy-constrained wireless communications with HARQ,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 191–205, 2016.
- [91] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, “Wireless networks with rf energy harvesting: A contemporary survey,” *IEEE Communications Surveys Tutorials*, vol. 17, no. 2, pp. 757–789, Secondquarter 2015.
- [92] M. S. I. Khan, J. Mii, and V. B. Mii, “A polling mac with reliable rf recharging of sensor nodes,” in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, pp. 831–836.
- [93] Y. . P. Hong, T. Hsu, and P. Chennakesavula, “Wireless power transfer for distributed estimation in wireless passive sensor networks,” *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5382–5395, Oct 2016.
- [94] G. Pan, H. Lei, Y. Yuan, and Z. Ding, “Performance analysis and optimization for swipt wireless sensor networks,” *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 2291–2302, May 2017.
- [95] Y. Zhang, H. Pflug, H. J. Visser, and G. Dolmans, “Wirelessly powered energy autonomous sensor networks,” in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2014, pp. 2444–2449.
- [96] M. Rostami, J. Gummeson, A. Kiaghadi, and D. Ganesan, “Polymorphic radios: A new design paradigm for ultra-low power communication,” in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, ser. SIGCOMM ’18. New York, NY, USA: ACM, 2018, pp. 446–460. [Online]. Available: <http://doi.acm.org/10.1145/3230543.3230571>
- [97] A. Biazon and M. Zorzi, “Battery-powered devices in wpcns,” *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 216–229, Jan 2017.

- [98] M. A. Abd-Elmagid, T. ElBatt, and K. G. Seddik, "Optimization of wireless powered communication networks with heterogeneous nodes," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–7.
- [99] C. Zhong, G. Zheng, Z. Zhang, and G. K. Karagiannidis, "Optimum wirelessly powered relaying," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1728–1732, Oct 2015.
- [100] H. Chen, Y. Li, J. L. Rebelatto, B. F. Ucha-Filho, and B. Vucetic, "Harvest-then-cooperate: Wireless-powered cooperative communications," *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1700–1711, April 2015.
- [101] Y. Gu, H. Chen, Y. Li, and B. Vucetic, "An adaptive transmission protocol for wireless-powered cooperative communications," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 4223–4228.
- [102] D. Xu and H. Zhu, "Outage minimized resource allocation for multiuser ofdm systems with swipt," *IEEE Access*, vol. 7, pp. 79 714–79 725, 2019.
- [103] H. Ju and R. Zhang, "User cooperation in wireless powered communication networks," in *2014 IEEE Global Communications Conference*, Dec 2014, pp. 1430–1435.
- [104] M. Zhong, S. Bi, and X. Lin, "User cooperation for enhanced throughput fairness in wireless powered communication networks," in *2016 23rd International Conference on Telecommunications (ICT)*, May 2016, pp. 1–6.
- [105] S. Bi and Y. J. Zhang, "Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 4177–4190, June 2018.
- [106] C. You and K. Huang, "Wirelessly powered mobile computation offloading: Energy savings maximization," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.
- [107] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–1, 2017.

- [108] F. Wang and J. Xu and X. Wang and S. Cui, “Joint offloading and computing optimization in wireless powered mobile-edge computing systems,” in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.
- [109] M. A. Abd-Elmagid and A. Biazon and T. ElBatt and K. G. Seddik and M. Zorzi, “Non-orthogonal multiple access schemes in wireless powered communication networks,” in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.
- [110] J. Yang and S. Ulukus, “Optimal packet scheduling in an energy harvesting communication system,” *IEEE Transactions on Communications*, vol. 60, no. 1, pp. 220–230, January 2012.
- [111] Z. Wang, V. Aggarwal, and X. Wang, “Power allocation for energy harvesting transmitter with causal information,” *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 4080–4093, Nov 2014.
- [112] R. Ma and W. Zhang, “Optimal power allocation for energy harvesting communications with limited channel feedback,” in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2014, pp. 193–197.
- [113] W. Du, J. C. Liando, H. Zhang, and M. Li, “Pando: Fountain-enabled fast data dissemination with constructive interference,” *IEEE/ACM Transactions on Networking*, vol. 25, no. 2, pp. 820–833, April 2017.
- [114] B. T. Bacinoglu, E. Uysal-Biyikoglu, and C. E. Koksall, “Finite-horizon energy-efficient scheduling with energy harvesting transmitters over fading channels,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 6105–6118, Sept 2017.
- [115] M. S. H. Abad and O. Ercetin, “Finite horizon throughput maximization for a wirelessly powered device over a time varying channel,” in *2018 IEEE Globecom Workshops (forthcoming)*, 2018.
- [116] D. Xu and Q. Li, “Cooperative resource allocation in cognitive wireless powered communication networks with energy accumulation and deadline requirements,”

Science China Information Sciences, vol. 62, no. 8, p. 82302, Jul 2019. [Online]. Available: <https://doi.org/10.1007/s11432-018-9813-9>

- [117] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, “Energy-optimal mobile cloud computing under stochastic wireless channel,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4569–4581, September 2013.
- [118] S. Ko, K. Huang, S. Kim, and H. Chae, “Live prefetching for mobile computation offloading,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3057–3071, May 2017.
- [119] J. Lee and N. Jindal, “Energy-efficient scheduling of delay constrained traffic over fading channels,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1866–1875, April 2009.
- [120] M. Zafer and E. Modiano, “Delay-constrained energy efficient data transmission over a wireless fading channel,” in *2007 Information Theory and Applications Workshop*, Jan 2007, pp. 289–298.
- [121] C. You, Y. Zeng, R. Zhang, and K. Huang, “Resource management for asynchronous mobile-edge computation offloading,” in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2018, pp. 1–6.
- [122] S. Agrawal and N. Goyal, “Analysis of thompson sampling for the multi-armed bandit problem,” in *Proceedings of the 25th Annual Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, S. Mannor, N. Srebro, and R. C. Williamson, Eds., vol. 23. Edinburgh, Scotland: PMLR, 25–27 Jun 2012, pp. 39.1–39.26.
- [123] E. Kaufmann, N. Korda, and R. Munos, “Thompson sampling: An asymptotically optimal finite-time analysis,” in *Algorithmic Learning Theory*, N. H. Bshouty, G. Stoltz, N. Vayatis, and T. Zeugmann, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 199–213.
- [124] D. J. Russo, B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen, “A tutorial on thompson sampling,” *Found. Trends Mach. Learn.*, vol. 11, no. 1, pp. 1–96, Jul. 2018. [Online]. Available: <https://doi.org/10.1561/22000000070>

[125] G. C. Christian Robert, *Monte Carlo Statistical Methods*, 2nd ed., ser. Springer Texts in Statistics. Springer, 2004.