Università degli Studi di Padova
Dipartimento di Ingegneria dell'Informazione

Scuola di Dottorato in Ingegneria dell'Informazione
Indirizzo in Scienza e Tecnologia dell'Informazione

XXX Ciclo

Stochastic Optimization of Energy Harvesting Wireless Communication Networks

Dottorando
Alessandro Biason

Supervisore:
Chiar.mo Prof. Michele Zorzi

Coordinatore del Corso:
Chiar.mo Prof. Andrea Neviani

Anno Accademico 2017/2018

Table of contents

Abstract

Energy harvesting from environmental energy sources (e.g., sunlight) or from man-made sources (e.g., RF energy) has been a game-changing paradigm, which enabled the possibility of making the devices in the Internet of Things or wireless sensor networks operate autonomously and with high performance for years or even decades without human intervention. However, an energy harvesting system must be correctly designed to achieve such a goal and therefore the *energy management* problem has arisen and become a critical aspect to consider in modern wireless networks. In particular, in addition to the hardware (e.g., in terms of circuitry design) and application point of views (e.g., sensor deployment), also the communication protocol perspective must be explicitly taken into account; indeed, the use of the wireless communication interface may play a dominant role in the energy consumption of the devices, and thus must be correctly designed and optimized. This analysis represents the focus of this thesis.

Energy harvesting for wireless system has been a very active research topic in the past decade. However, there are still many aspects that have been neglected or not completely analyzed in the literature so far. Our goal is to address and solve some of these new problems using a common stochastic optimization setup based on dynamic programming. In particular, we formulate both the centralized and decentralized optimization problems in an energy harvesting network with multiple devices, and discuss the interrelations between these two schemes; we study the combination of environmental energy harvesting and wireless energy transfer to improve the transmission rate of the network and achieve a balanced situation; we investigate the long-term optimization problem in wireless powered communication networks, in which the receiver supplies wireless energy to the terminal nodes; we deal with the energy storage inefficiencies of the energy harvesting devices, and show that traditional policies may be strongly suboptimal in this context; finally, we investigate how it is possible to increase secrecy in a wireless link where a third malicious party eavesdrops the information transmitted by an energy harvesting node.

List of Acronyms

ACRDA Asynchronous Contention Resolution Diversity Aloha.

AP Access Point.

App-VIA Approximate Value Iteration Algorithm.

BP Balanced Policy.

CDMA Code Division Multiple Access.

CEIP Constrained Energy Independent Policy.

CSI Channel State Information.

DAMA Demand Assignment Multiple Access.

DC Direct Current.

Dec-MDP Decentralized Markov Decision Process.

Dec-POMDP Decentralized Partially Observable Markov Decision Process.

DL Downlink.

E-SSA Enhanced Spread Spectrum Aloha.

ECL Energy Conversion Losses.

EEH Environmental Energy Harvesting.

EH Energy Harvesting.

EHD Energy Harvesting Device.

EIP Energy Independent Policy.

ERN Energy Rich Node.

FDMA Frequency Division Multiple Access.

GP Greedy Policy.

HCEIP Heuristic-CEIP.

i.i.d. independent and indentically distributed.

IC Interference Cancellation.

IoT Internet of Things.

LCD Low Complexity Decoding.

LCP Low Complexity Policy.

LRTA* Learning Real Time A*.

MAC Medium Access Control.

MC Markov Chain.

MDP Markov Decision Process.

MIMO Multiple-Input-Multiple-Output.

mmWave millimeter wave.

MPS Markov Policy Search.

NOMA Non-Orthogonal Multiple Access.

OCEIP Optimal-CEIP.

OEIP Optimal-EIP.

OFDM Orthogonal Frequency-Division Multiplexing.

OP Optimal Policy.

OP-OFF Optimal Offline Policy.

OP-ON Optimal Online Policy.

OSP Optimal Secrecy Policy.

OSP-FULL OSP with Full CSI.

PIA Policy Iteration Algorithm.

POMDP Partially Observable Markov Decision Process.

PS Power Splitting.

RF Radio-Frequency.

RF-WET Radio-Frequency Wireless Energy Transfer.

RX Receiver.

SCMR Strongly Coupled Magnetic Resonances.

SCMR-WET Strongly Coupled Magnetic Resonances Wireless Energy Transfer.

SICD Successive Interference Cancellation Decoding.

SoC State of Charge.

SQP Sequential Quadratic Programming.

SWIPT Simultaneous Wireless Information and Power Transfer.

TDMA Time Division Multiple Access.

TS Time Splitting.

TX Transmitter.

UL Uplink.

VIA Value Iteration Algorithm.

w.p. with probability.

WBAN Wireless Body Area Network.

WCSP Weighted Constraint Satisfaction Problem.

WET Wireless Energy Transfer.

WPCN Wireless Powered Communication Network.

WSN Wireless Sensor Network.

CHAPTER 1

Introduction

In the context of energy-constrained and green networking, the design of low-power systems and the use of renewable energy sources in wireless networks are prominent areas of investigation. If equipped with harvesting capabilities, the nodes in the Internet of Things (IoT) or in a Wireless Sensor Network (WSN) can recharge their batteries with renewable energy sources. Since one of the main goals in this context is to keep the network operational for a very long time (even decades), Energy Harvesting (EH) is a promising technique to achieve such a target.

From a networking perspective, one of the main challenges in EH scenarios consists in correctly *managing* the available energy in order to optimize the network performance (e.g., in terms of throughput, transmission rate, etc.). Indeed, EH adds a new level of complexity to the system, which must be correctly analyzed in order to really exploit its benefits. The goal of this thesis is to address, formalize and solve some of the problems related to the energy management.

The three main topics that will be covered are:

- **Energy Harvesting Networks with Multiple Devices.** Studying an Energy Harvesting Device (EHD) alone is useful to understand the key trade-offs and limits of an energy harvesting system. However, when multiple devices are considered, new network concerns (e.g., coordination, scheduling and channel access) need to be addressed and may significantly change the behavior of the EHDs. In the first part of the thesis, we will focus on the *centralized* and *decentralized* network optimization schemes. In the first case, the Receiver (RX) knows the state of the system (i.e., the energy level of the devices) and decides the actions the EHDs should perform (e.g., whether to transmit or not) in every time instant. Instead, in the latter, the receiver knows the state of the system sporadically and tries to compute the best possible series of actions using partial information only.

- **Wireless Energy Transfer.** In the second part of the thesis, we will study more recent applications of the energy harvesting technologies. We focus on *man-made* energy (and not only on the traditional environmental energy sources as, for example,

sunlight) and define how the capabilities of controlling and transfer the energy wirelessly bring new life to wireless systems. More specifically, we focus on the *energy cooperation* paradigm, in which users transfer energy to each other to improve the overall performance of the system, and on Wireless Powered Communication Networks (WPCNs), in which an Energy Rich Node (ERN) (e.g., the receiver) transfers Radio-Frequency (RF) energy to the other nodes in the network.

- **Peculiar Aspects of Energy Harvesting Systems.** Although EHDs have already been studied in the literature from different points of view, many aspects have been neglected or only marginally addressed so far. One of the contributions of this thesis is the study of the effects of *inefficiencies* of the EHDs (e.g., in terms of non-ideal batteries or limited State of Charge (SoC) knowledge) on the network performance and some *security*-related aspects from a physical layer perspective. First, we will show that neglecting these inefficiencies may lead to wrong energy management policies and consequently threaten the system performance; finally, we characterize how the maximum achievable secrecy rate changes according to the various system parameters and describe the importance of adapting the transmission power and coding rates.

Although we focus on slightly different applications in every chapter, throughout this thesis we use similar techniques and methodologies to model and optimize the networks. In particular, the key tools we will use are based on dynamic programming [14] and Markov decision processes [122].

1.1 Related Work

Energy Harvesting (EH) refers to the process of gathering energy from an external energy source. There are two main categories that can be considered in this context: Environmental Energy Harvesting (EEH) and *man-made* EH (i.e., Wireless Energy Transfer (WET) in this thesis). The first one represents the traditional idea of an EH system and summarizes the energy sources available in the ambient, e.g., sunlight [41, 106, 123], vibrational [13, 127], piezoelectric [39], indoor lighting [29, 139, 155], biological, chemical, thermal [139], electromagnetic [120], acoustic noise, etc. On the other hand, the latter represents the sources explicitly designed to deliver energy to the terminals, e.g., Radio-Frequency Wireless Energy Transfer (RF-WET) [38, 67, 99, 104, 116, 142, 159], Strongly Coupled Magnetic Resonances Wireless Energy Transfer (SCMR-WET) [71], inductive coupling [152], etc. We will discuss these techniques in more detail in the next sections.

Practical examples of EH scenarios include: a network of temperature sensing EHDs, where a high temperature measurement can be an indicator of overheating or fire; a sensor network which routes different priority packets [10]; data transmission over a fading channel where the EHDs adjust the transmitted redundancy according to the instantaneous channel realization [132].

1.1.1 Environmental Energy Harvesting

Environmental energy harvesting was also dubbed "energy scavenging," since the devices scavenge or harvest unused energy from the external environment. This is a very powerful technique, since an Energy Harvesting Device (EHD) does not need to interact with other devices or people to be operational, and its lifetime only depends on its own hardware

failures. Nevertheless, the main drawback of EEH is the intrinsic *stochastic* nature of the energy source, as it may not be possible to predict how much energy will be harvested in the future, if any. As a consequence, there is not a single solution that fits all possible scenarios, and the optimal series of actions to perform will depend on the environment [45]. Defining the correct way of managing the EH uncertainty is the common fundamental problem addressed in every chapter of the thesis.

There are many different ambient sources that have been used to design an EEH system ([169] presented a survey on the several different environmental energy harvesting technologies for WSNs). Two important parameters that are common to most technologies are the *size* of the energy harvester module and the *power density* delivered from the source (see [127] for a survey of many energy scavenging methods, with particular focus on vibrational energy). The size is generally constrained by the applications, since nodes in the IoT and WSNs are typically very compact. Instead, the power density strongly depends on the used technology (this can be clearly seen in Table 1.1) and on the environment; for example, using a solar panel of size 1 cm^2, the sunlight irradiation power may oscillate from hundreds of μW to tens of mW, depending on the time of the day, the day of the year, the geographic position, the weather conditions, etc. The analytic characterization of the energy sources have been addressed in the literature but is not the focus of this thesis, in which we will assume to know the stochastic model of the energy arrivals.

Table 1.1. Energy Sources [120, 123, 127].

Energy Source	**Power Density**
Solar outdoors - sunny day	15 mW/cm^2
Piezoelectric	330 $\mu W/cm^3$
Vibrations	200 $\mu W/cm^3$
Solar outdoors - cloudy day	150 $\mu W/cm^2$
Ambient Radio-Frequency (in London)	6.39 $\mu W/cm^2$
Acoustic Noise (at 100 dB)	0.96 $\mu W/cm^3$

Many previous papers studied energy harvesting communication systems from a networking perspective because of their ability to increase the network lifetime, provide self-sustainability and, ideally, allow perpetual operations [150]. [41] studied the network performance when solar cells are used to receive energy, showing how the harvested energy changes as a function of the latitude, time of the day and season. Analytically, [74] formulated the problem of maximizing the average value of the reported data using a node with a rechargeable battery. In [132, 133], Sharma *et al.* studied heuristic delay-minimizing policies and sufficient stability conditions for a single EHD with a data queue. Ozel *et al.* set up the offline throughput optimization problem from an information theoretic point of view in [113], where they derived the information-theoretic capacity of the AWGN channel and presented two schemes that achieve such capacity ("save-and-transmit" and "best-effort-transmit"). Some researchers focused on batteryless devices [112, 138]. In particular, [112] considered a traditional EH system with amplitude constraints and found the channel capacity under causal channel information knowledge. The throughput optimization problem with finite batteries in an EH system was studied in [93, 145]. A common technique to model the batteries is to approximate them with finite energy queues, in which energy arrives and departs over time. Markov models are suitable for these cases [6]

and were largely adopted in the literature so far [17, 74, 93, 91, 95]. Energy harvesting receivers were analyzed in [85, 86, 165], with a focus on the optimization of the sampling strategies. Also, [8] considered a transmitter-receiver pair with harvesting capabilities. For more references about energy harvesting, we refer the readers to the surveys in [84, 150].

Nevertheless, despite the extensive literature about EH, many different aspects still have to be analyzed and are an active research topic nowadays.

1.1.2 Wireless Energy Transfer

There are three main techniques used to transfer energy wirelessly: Radio-Frequency Wireless Energy Transfer (RF-WET), Strongly Coupled Magnetic Resonances (SCMR), and inductive coupling (this technology operates at distances less than a wavelength; clearly, although this mechanism is very efficient, it cannot be used in many WSNs or IoT scenarios because of the very short operating distance [128]). In this thesis, we focus on the first two techniques.

1.1.2.1 Radio-Frequency Wireless Energy Transfer

This paradigm has been studied for several decades (see [19] for a brief history of RF energy transmission) and, in the last few years, RF-WET was also considered in WSNs [61, 65, 84]. Via dedicated components, namely *rectifiers* [38] (which, for example, can be composed of a diode [49], a bridge of diodes or a voltage rectifier multiplier), the devices are able to convert the input RF signal into Direct Current (DC) voltage, which can be used to refill their batteries. In many papers, the authors assume to have an Energy Rich Node (ERN) (a typical example is the *Powercaster Transmitter* [121]) that supplies energy to several passive sensor nodes (equipped with a *Powerharvester Receiver*, for example). One of the main problems studied in this area so far has been the combination of energy and information transmission. Indeed, even if it would be theoretically possible to transmit energy and data simultaneously using the same signal, this is not feasible with current technology [44]. Therefore, two techniques were developed: Time Splitting (TS) and Power Splitting (PS) [167]. In the first case the transferred energy and data are sent at different times. In the second case, the transmit power is split: part of it is used for data and the rest for energy. Articles such as [116] or [142] studied the optimal power splitting for the PS technique. TS was used in [67], where transmission policies for a relay in a topology composed of three nodes (source, relay and destination) were studied. [65] proposed a medium access control mechanism based on WET that achieves a high degree of fairness among the devices. [107] studied a network composed of one access point that transmits RF energy to several nodes, with the aim to design an admission control mechanism. In [62], the authors studied the interleaving problems related to transmitting and receiving energy simultaneously, introducing a polling-based MAC protocol. [26] studied the case where some devices (energy rich nodes) move through the network and refill the batteries of the sensors with RF radiation. In [99] and [105] the authors introduced a Medium Access Control (MAC) protocol, where nodes request energy from some transmitters, and these cooperate by sending RF energy to those nodes.

Another important class of networks studied in the literature are the Wireless Powered Communication Networks (WPCNs) [58], in which an ERN supplies energy to several devices using RF-WET, and these transmit data in uplink. In these scenarios, a *doubly near-far* problem is present, since users far away from the base station experience, on average, worse channel conditions than the others both in uplink and in downlink. The *doubly near-far* problem was initially studied in [58]. The authors introduced the "harvest-

then-transmit" scheme in which the time horizon is divided in slots and every slot is divided in two phases: first, the energy rich node transfers energy to the devices and, secondly, the devices use the harvested energy to transmit data in the uplink channel. The trade-offs between the times to use for transferring energy and transmitting data were investigated and the optimal scheduling scheme was provided. The authors extended their work in [59], where user cooperation was taken into account in a two-device system. It was shown that cooperation is a powerful technique which can effectively improve the system performance. Nevertheless, because of the additional complexity demanded to compute the scheduling scheme, and the unavoidable coordination and physical proximity required among devices, the cooperation solution may not be suitable for every scenario. [24] described the "harvest-then-cooperate" protocol, in which source and relay work cooperatively in the uplink phase for the source's information transmission. The authors also derived an approximate closed-form expression for the average throughput of the proposed protocol. [1] generalized conventional Time Division Multiple Access (TDMA) wireless networks (no energy harvesting) to a new type of wireless networks named generalized-WPCNs (g-WPCNs), where nodes are assumed to be equipped with RF energy harvesting circuitries along with energy supplies. It was shown that both conventional TDMA wireless networks and WPCNs with only RF energy harvesting nodes provide lower bounds on the performance of g-WPCNs in terms of maximum sum-throughput and max-min throughput. [64] studied the case of devices with energy and data queues and described a Lyapunov approach to derive the stochastic optimal control algorithm which minimizes the expected energy downlink power and stabilizes the queues. Optimization over multiple slots was considered in an Orthogonal Frequency-Division Multiplexing (OFDM) based WPCN in [171], where non-causal (offline) and causal (online) strategies are proposed to maximize the average transmission rate. The long-term performance of a single-user system for a simple transmission scheme was presented in closed form in [96]. [52] modeled a WPCN with a Decentralized Partially Observable Markov Decision Process (Dec-POMDP) and minimized the total number of waiting packets in the network. Similarly to [58], a WPCN was studied in [80], where the energy rich node has also the capability of beamforming the transferred RF signal in order to serve the most disadvantaged users and to guarantee throughput fairness. The authors managed to convert a non-convex optimization problem into a spectral radius minimization problem, which can be efficiently solved. [162] studied the applicability of the massive Multiple-Input-Multiple-Output (MIMO) technology to a WPCN. With massive MIMO it becomes possible to receive data from several different devices simultaneously (thanks to spatial multiplexing), but also to improve the downlink performance by using sharp beams. Most previous works describe a half-duplex system in which uplink and downlink cannot be operated simultaneously. Instead, the full duplex case was studied in [57, 60]. [57] optimized the time allocations for WET and data transmission for different users in order to maximize the weighted sum rate of the uplink transmission. The authors considered perfect as well as imperfect self-Interference Cancellation (IC) at the access point and showed that, when IC is performed effectively, the full-duplex case outperforms the half-duplex one. A survey of recent advances and future perspectives in the WPCN field can be found in [16].

Although it has been used in many applications, it is important to note that RF wireless energy transfer, due to the radiative nature of the mechanism, has a very low energy efficiency [53] and requires line-of-sight to achieve high performance.

1.1.2.2 Strongly Coupled Magnetic Resonances Wireless Energy Transfer

Wireless energy transfer based on SCMR is a compromise between inductive coupling and RF-WET: it can be used in mid-range applications (order of $2-3$ meters) and has a high efficiency.

In [71], it was shown that it is possible to power a 60 W light bulb at a distance of 2 m with an efficiency of 40% using SCMR-WET. The authors also extended this work in [72], showing that SCMR wireless energy transfer can be used to power several devices at the same time with high efficiency. This is possible because non-radiative wireless energy transfer is used, which relies on *near-field* magnetic coupling of conductive loops. In [128], the authors showed that it is possible to achieve the maximum possible energy transfer efficiency regardless of the orientation of the device, as long as the receiver is in the working range of the transmitter. The two main problems related to SCMR-WET are that: 1) it is necessary to use coils of large size (order of 20 cm) and 2) the transmission range is limited only to a few meters. For these reasons, using SCMR-WET in WSNs may be feasible only in very particular applications in which the positions of the transmitter and receiver are stationary (e.g., we may consider two devices in adjacent rooms of a building).

Although SCMR-WET seems promising, only a few applications can be found in the literature so far. [136] considered a vehicle that travels inside a WSN, periodically recharging the nodes (one at a time) wirelessly, and showed that through periodic charges the network may ideally remain operational for an unlimited amount of time. The authors extended the study to multiple transmissions in [160], and a similar technique was also discussed in [168]. Some applications can be found in biomedical implants, e.g., [124], and a wireless charger prototype based on SCMR-WET was proposed in [164].

1.1.3 Environmental Energy Harvesting and Wireless Energy Transfer

Equipping an EEH system with a WET interface can be a powerful solution to avoid energy outages and increase the capabilities of the devices. However, only a few papers investigated the optimization strategies for EEH system with WET capabilities. In particular, in [46–48], Gurakan *et al.* introduced the concept of *energy cooperation*, unifying the study of environmental energy harvesting and wireless energy transfer techniques. They considered a system composed of a few nodes and investigated optimal offline communication schemes. However, none of these papers considered the effects of finite batteries, which instead will be a key feature of our models. Also [146, 147] studied the combination of WET and EEH with infinite batteries and bi-directional energy transfer, whereas in [148] the authors presented the case of two transmitters with finite batteries. A model that considers the circuitry cost was published in [103], where a transmitter and a receiver powered by the same power source with infinite batteries can exchange energy.

1.2 Organization of the Thesis

The thesis is subdivided into five chapters. Chapter 2 defines the generic system model analyzed in the remainder of the thesis. Chapters 3-5 can be read separately using Chapter 2 as a baseline. In particular:

- In Chapter 3 we analyze the networking problem in which multiple energy harvesting devices are considered in the same network. The main results of this chapter are based on **[C1]**, **[C13]** and **[J6]** (see page 155 for a list of publications).
- In Chapter 4 we investigate the benefits of using wireless energy transfer to further boost the network performance. The main results of this chapter are based on **[C3]**, **[C7]**, **[C11]**, **[J1]** and **[J3]**.
- In Chapter 5 we study the inefficiencies that can be found in real devices and set up a physical layer security optimization problem. The main results of this chapter are based on the conference papers **[C4]**-**[C6]** and on the journal paper **[J2]**.

Finally, Chapter 6 concludes this thesis.

Notation. Table 1.2 summarizes the variables we consider in this thesis. In general, subscripts refer to the time indices or to the states of the system whereas superscripts denote the indices of the devices or of the sub-carriers. Boldface letters indicate vectors. $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ refer to probabilities and expectations, respectively.

Table 1.2. Notation and parameters.

	Symbol	**Meaning**
Indices	k	Slot index (subscript)
	b or $\mathbf{s}$	MC state (subscript)
	$i = 1, \ldots, N$	EHD index (superscript)
	$r = 1, \ldots, M$	Sub-carrier index (superscript), Section 5.2
Energy related	b_k	Battery level [e.q.]
	$b_{\max}$	Maximum battery level [e.q.]
	P_k	Transmission power [W]
	$q(P_k)$	Transmission energy [e.q.]
	e_k^{EEH} (or e_k)	Harvested energy from environmental sources [e.q.]
	e_k^{RF}	Harvested energy with RF-WET [e.q.]
	e_k^{SCMR}	Harvested energy with SCMR-WET [e.q.]
	z_k	Transferred energy (e.g., using SCMR-WET) [e.q.]
Slot	T	Slot duration
	$\tau_k^{(i)}$	Duration of the data transmission phase
	τ_k^{rx}	Duration of the WET downlink phase
	β_k	Probability of the "Extraction of the SoC" phase
Losses	η_{ECL}	Energy Conversion Losses
	η_{ESL}	Energy Storage Losses
	η_{SCMR}	Efficiency of the SCMR-WET link (including all losses)
Markov chains	$\mathbf{S}_k$, $\mathbf{s}$	State of the system
	r	Reward in a single slot
	G	Reward in the long term

CHAPTER 2

Preliminaries and System Model

In this chapter we present a generic energy harvesting model that will be used in the remainder of the thesis. Chapters 3, 4 and 5 will always refer to this model, possibly with some extensions or changes according to the aspects being analyzed.

2.1 Generic Model for an Energy Harvesting Device

The network is always composed of one Receiver (RX) and one or more terminal nodes (also referred as *users*, *nodes*, *sensors*, or *Energy Harvesting Devices* (EHDs)) with Energy Harvesting (EH) capabilities. The receiver may coincide with the Access Point (AP) or with a terminal node itself, depending on the scenario. Users gather environmental data and report it to RX via a point-to-point wireless communication channel. Multi-hop networks are not the focus of this thesis, although they have been investigated in the literature (e.g., see the routing protocols described in [56, 87]). In the following we present the model for a single node; this will be properly extended in Chapters 3 and 4 for the cases with multiple users.

2.1.1 Slot Division

The time horizon of our framework is typically *infinite* in order to consider the steady-state conditions of the network. Indeed, EH systems generally operate in the same working conditions for a long time and usually reach stationary conditions. The time is slotted (see [34, 111, 145] for continuous time models) and slot k corresponds to the time interval $[k\,T, (k+1)T)$, $k = 0, 1, \ldots$, where T is the common duration of all slots (in several cases, we will consider a normalized slot duration $T = 1$ s without loss of generality). In every slot, the following operations are performed:

- **Battery Charging.** During every slot, the node gathers energy from the environment and/or from man-made sources and store it in its finite size battery. When ambient EH or SCMR-WET are considered, this phase generally lasts for the whole slot duration since the harvesting circuitry is independent of the other parts of the device (e.g., consider a solar panel). Instead, with Radio-Frequency Wireless Energy Transfer (RF-WET), the node must reserve a portion of time for the energy

reception,[1] as transmit/receiving data and energy simultaneously using the same antenna is not considered here.

- **Extraction of the State of Charge (SoC).** The node inspects its battery level and communicate it to RX, which plays the role of a *controller* that computes the *policy*, i.e., the series of actions that defines how the node should behave, according to the SoC.

  In a fully centralized approach (Section 3.1 and Chapters 4 and 5), this phase is repeated every slot with a probability $\beta_k = 1$. Instead, in the decentralized scheme of Section 3.2, this task will be performed with a probability $\beta_k \in [0, 1]$.

  We also remark that although in an ideal scenario the SoC is perfectly known by the EHD and RX, in practice, because of hardware limitations, the state of charge may be known with a finite accuracy only. We will consider this particular case in Section 5.1.

- **Policy Computation and Dissemination.** Using the available information (e.g., the SoC), RX computes the actions the EHD should use in the current or future slots (e.g., to transmit or not, the transmission power to use, the transmission duration, etc.) and communicate it to the node. In the centralized case, a new action is computed in every slot according to the state of charge of the user, whereas in decentralized scenarios the actions are computed only when the SoC is extracted (see the previous bullet point). The series of actions is called *policy*.

  The policies we consider can always be decomposed in two steps. First, a policy is fully evaluated *offline*: the actions to use are defined in every case of interest (i.e., for every possible state of the system). Then, the policy is used *online*: in every slot, given the state of the system, RX decides the action to use according to the precomputed offline actions. The online phase is very lightweight and does not require any optimization, since the actions to use have been previously computed in the offline phase, therefore it can be implemented in real devices with low computing capabilities. We will mainly deal with dynamic programming and Markov Decision Processes (MDPs) [14] to define the policy.

- **Data Transmission.** This is the main phase in which the node, if necessary (i.e., if it has data packets to send) and possible (i.e., if it has enough energy), transmits data packets to RX in uplink. When multiple EHDs are considered, this may be performed in many different ways; some of them will be separately studied in the next chapters (e.g., random access in Section 3.2, Time Division Multiple Access (TDMA) in Section 4.2), whereas others (e.g., Code Division Multiple Access (CDMA)) can be found in the literature (e.g., see paper **[C14]** at page 155). These schemes have pros and cons, and may also depend on the scenario under investigation.

The previous phases will be further discussed in the following chapters, where relevant. We now give more details about the battery charging and discharging processes and the battery modeling.

[1] There is also another technique, named Simultaneous Wireless Information and Power Transfer (SWIPT) (see [63, 135]), that studies how to transmit data and energy simultaneously. However, this is not the focus of this thesis.

2.1.2 Energy Arrival Process

The energy arrives at the node from environmental (i.e., EEH) and/or man-made sources (i.e., WET). Quantity E_k^{x} denotes the harvested energy in slot k using technology x.

2.1.2.1 Environmental Energy Harvesting

In general, we will model the environmental energy using a random process independent and indentically distributed (i.i.d.) over time $\{E_k^{\mathrm{EEH}}\}$ (and among nodes, when multiple EHDs are considered), e.g., deterministic, Bernoulli or truncated geometric (see [92, 134, 156], or [43] for a characterization of the energy gathered from the light). This is a particular case of the Generalized Markov model described in [51], where E_k^{EEH} statistically depends on the amount of energy $(E_{k-1}^{\mathrm{EEH}}, ..., E_{k-L}^{\mathrm{EEH}})$ harvested in the previous L time slots as well as on a scenario parameter. In [51] it is shown that different energy sources can be efficiently modeled by means of different values of L. In particular, piezoelectric energy is well described using $L = 0$, while solar is better characterized by $L = 1$. In this thesis, we use $L = 0$ to maintain the analysis simpler [90]. However, the models can be extended to the more general, temporally correlated (e.g., via an underlying common Markov model as in [93]) or spatial correlated cases (see [40] for a discussion about temporal and spatial correlation).[2] More detailed models have been proposed in the literature (e.g., see [106]), which however are strictly related with the energy source under investigation and may be more computationally demanding.

Finally, we note that using Bernoulli energy arrival processes is a common choice in the literature so far, e.g., [92, 134, 156] and it was shown that for a traditional EH system, the optimal policy for Bernoulli arrivals is approximately optimal for generic i.i.d. arrival processes as well [134].

2.1.2.2 Radio-Frequency Wireless Energy Transfer

Man-made sources present different characteristics and strongly depend on the technology (see Section 1.1.2). For modeling the RF energy, we use the following formula:

$$E_k^{\mathrm{RF}} = \tau_k^{\mathrm{rx}} P_k^{\mathrm{rx}} \eta_{\mathrm{ECL}} g_k = \tau_k^{\mathrm{rx}} P_k^{\mathrm{rx}} \eta_{\mathrm{ECL}} g_0 d_k^{-\delta_k} \kappa_k, \tag{2.1.1}$$

where τ_k^{rx} is the duration of the WET phase, η_{ECL} is a constant in $(0, 1]$ that models the Energy Conversion Losses (ECL) at the device, P_k^{rx} is the transmission power of the energy rich node (which always coincides with the receiver in our model), and $\{g_k\}$ is the random process related to the channel conditions, and on the specific beamforming technique (if any). The term g_k can be explicitly written as $g_k = \tilde{g}_k \kappa_k$, where $\{\kappa_k\}$ is a random process that represents the fading and $\tilde{g}_k$ is the average channel gain, obtained by considering the path loss effects as $\tilde{g}_k = g_0 d_k^{-\delta_k}$. The term g_0 is the signal power gain at a reference distance of 1 m, d_k is the distance between the device and rx expressed in meters, and δ_k is the path loss exponent. Note that the parameters depend on k and may change over time. We note that, because of fading, the channel conditions may not be known in advance, thus the exact amount of transferred energy is unknown a priori.

2.1.2.3 Strongly Coupled Magnetic Resonances Wireless Energy Transfer

In this thesis, we consider only a very particular example of wireless energy transfer based on SCMR in which we know the efficiency of the link, namely η_{SCMR}, in advance, so that

[2] The outcomes of this thesis are not strictly related with the model nor with the harvesting technology we consider, but are general design considerations that can be applied to many EH systems.

the received energy is

$$E_k^{\mathrm{SCMR}} = \tau_k^{\mathrm{rx}} P_k^{\mathrm{rx}} \eta_{\mathrm{SCMR}} \tag{2.1.2}$$

where τ_k^{rx} and P_k^{rx} are defined as before.

Note that the 40% efficiency claimed in [71] for a distance of 2 m using SCMR is only referred to the transmission itself. Indeed, the effective wall-to-load efficiency (ratio between the power extracted from the wall power outlet and the received power) was 15%, which will be used as a baseline for our numerical evaluations. However, we remark that this technology presents many pitfalls for IoT applications and may not always be feasible in practice.

2.1.3 Energy Consumption Process

The two main sources of energy consumption we consider for an EHD are due to data transmission and to energy transfer to other devices (e.g., using WET based on SCMR). When a device is in sleep mode, it is assumed to consume negligible energy.

2.1.3.1 Data Transmission

In general, modeling the data transmission energy costs is a difficult task: in addition to the energy directly injected into the channel (which would be equal to $\tau_k \times P_k$ when the device transmits for a time interval of duration τ_k with a constant power P_k), also the costs of sensing, pre-processing (coding) and (eventually) compressing the data [144] have to be considered (see papers **[J4]**, **[C12]**, and **[J5]** at page 155). In addition to that, inefficiencies should be taken into account in real devices and generally degrade the performance of the transmission phase (e.g., non-linear behavior, and thus inefficiencies, of the power amplifier, reflections, mismatched circuits, etc.). Formally, we describe the energy consumption with a generic continuous, increasing and concave[3] function $Q(P_k)$, where P_k is the transmit power and $Q(0) = 0$ (sleep mode). More details about the structure of $Q(P_k)$ will be given in the next chapters.

2.1.3.2 Energy Transfer

In some scenarios, the node may receive much more energy and/or consume less energy than some of its neighbors. In these cases, it is reasonable to transfer energy from the node to other devices in order to balance the energy levels. This operation may be performed using SCMR-WET, since RF-WET would involve much higher losses. We will consider this scenario in Section 4.1, in which the receiver itself is an EHD with WET capabilities (and thus will be modeled similarly to a terminal node).

Throughout this thesis, we use the notation Z_k to indicate the transferred energy.

2.1.4 Battery Modeling

The energy harvesting node is equipped with a rechargeable battery, so that the energy stored in slot k, namely B_k (expressed in joules), can be used in a later slot. According to the energy arrival process (environmental and/or man-made), to the energy consumption process, and to the parameters set in the decision phase (i.e., the policy), the battery level

[3] In this thesis, the term "concave" will be used to designate concave functions, e.g., functions with non-positive second derivative.

changes dynamically in every slot as follows

$$
\begin{aligned}
B_{k+1} = \min\{ & B_{\max}, && \text{(battery size [J])} && \text{(2.1.3a)}\\
& B_k && \text{(previous battery level [J])} && \text{(2.1.3b)}\\
& - Q(P_k) && \text{(data transmission, uplink [J])} && \text{(2.1.3c)}\\
& + E_k^{\mathrm{EEH}} && \text{(Environmental EH, received [J])} && \text{(2.1.3d)}\\
& + E_k^{\mathrm{RF}} && \text{(RF-WET, received [J])} && \text{(2.1.3e)}\\
& + E_k^{\mathrm{SCMR}} && \text{(SCMR-WET, received [J])} && \text{(2.1.3f)}\\
& - Z_k\} && \text{(WET, transferred [J])} && \text{(2.1.3g)}
\end{aligned}
$$

The "min" operator is used to consider overflow situations. The energy losses due to inefficiencies are already included in the term $Q(P_k)$. We note that the battery update formula may slightly change when only a partial SoC knowledge if available (see Equation (5.1.3) in Section 5.1 for more details).

The finite battery capacity entails the following.

Definition 2.1.1 (Energy Outage and Overflow)**.** *In slot k, energy outage occurs if $B_k < Q(P_{\min})$, where $P_{\min}$ is the minimum available transmission power (greater than zero), and energy overflow occurs if $B_k - Q(P_k) + E_k^{\mathrm{EEH}} + E_k^{\mathrm{RF}} + E_k^{\mathrm{SCMR}} - Z_k > B_{\max}$.*

Energy outage prevents any transmission from being executed, due to the fact that the battery is empty, and hence the data transmission energy, $Q(P_k)$, must be equal to 0, regardless of the policy or other parameters. Differently, when energy overflow occurs, the new harvested energy cannot be stored in the energy buffer of the EHD, due to its finite storage capacity: consequently, as energy is lost, this event may potentially represent a future lost transmission opportunity. In summary, energy outage and overflow degrade the system performance and have to be considered in the design of a transmission policy [92].

2.2 Stochastic Optimization Tools

Different optimization problems related to EH can be found in the literature so far. Generally, the main focus is on optimizing some network parameters, e.g., throughput, delay, packet drop rate, etc. A lot of research focused on the *offline* optimization, where it was assumed that RX (or, eventually, the EHD itself) knows everything about the environment of the EHD (future energy arrivals, past history, channel status, etc.) [163]. Some relevant papers were produced by Sharma *et al.* [132], where the authors found throughput and mean delay optimal policies when the considered device has an infinite data queue. Also, Ozel *et al.* in [113] set up the offline throughput optimization problem and presented schemes to achieve the AWGN capacity. Tutuncuoglu *et al.* studied offline policies for more than one device [144]. The authors also presented several works on the imperfect storage capabilities of energy harvesting devices [149].

Instead, another class of problems regards *online* optimization. In this context, only a statistical knowledge of the environment is required. Generally, a Markov approach is used to model and solve the problem [17, 74, 93, 95]. [93] considered the case of a correlated energy generation process. [76] studied the use of relays in an energy harvesting communication system. The interactions among multiple devices were analyzed in [52, 94].

2.2.1 Battery Quantization

In order to perform the optimization, we model the battery of an EHD with a discrete Markov Chain (MC). In particular, we discretize the battery of the device in $b_{\max}+1$ levels, where $b_{\max}$ represents the maximum amount of energy quanta (e.q.) that can be stored in the battery and one energy quantum corresponds to $\delta \triangleq B_{\max}/b_{\max}$ J. By doing so, the battery is modeled as a finite size energy buffer. There exists a trade-off between the precision of the discrete approximation and the corresponding numerical complexity of the model. In general, if $b_{\max}$ is sufficiently high, the discrete model can be considered as a good approximation of the continuous system. The accuracy of the discrete approximation can always be improved by using a finer quantization, which however results in a model with more states and therefore higher complexity. Equation (2.1.3) can be rewritten in terms of energy quanta ($b_k \in \mathcal{B} \triangleq \{0,\dots,b_{\max}\}$):

$$
\begin{aligned}
b_{k+1} = \min\{ & b_{\max}, && \text{(battery size [e.q.])} && \text{(2.2.1a)}\\
& b_k && \text{(previous battery level [e.q.])} && \text{(2.2.1b)}\\
& -\lceil Q(P_k)/\delta \rceil && \text{(data transmission, uplink [e.q.])} && \text{(2.2.1c)}\\
& +\lfloor E_k^{\mathrm{EEH}}/\delta \rfloor && \text{(Environmental EH, received [e.q.])} && \text{(2.2.1d)}\\
& +\lfloor E_k^{\mathrm{RF}}/\delta \rfloor && \text{(RF-WET, received [e.q.])} && \text{(2.2.1e)}\\
& +\lfloor E_k^{\mathrm{SCMR}}/\delta \rfloor && \text{(SCMR-WET, received [e.q.])} && \text{(2.2.1f)}\\
& -\lceil Z_k/\delta \rceil \}. && \text{(WET, transferred [e.q.])} && \text{(2.2.1g)}
\end{aligned}
$$

Note that, in every slot, only an integer amount of energy quanta can be extracted from the battery. Similarly, only an integer amount of energy quanta can be harvested. Our choice of the floor and ceiling operations will lead to a lower bound to the real performance (however, if the quantization is fine, this almost coincides with the continuous model).

Equation (2.2.1) can be rewritten using the new lowercase notation as follows:

$$
b_{k+1} = \min\{b_{\max}, b_k - q(P_k) + e_k^{\mathrm{EEH}} + e_k^{\mathrm{RF}} + e_k^{\mathrm{SCMR}} - z_k\}, \tag{2.2.2}
$$

where all terms are expressed in terms of energy quanta.

2.2.2 Markov Decision Processes for Energy Harvesting

The model described so far can be studied as a MDP [14] with the following properties (this section is generic and presents the features of the MDP without a precise formulation, which however will be properly given in each chapter).

State of the MC. The state of the Markov chain is given by the battery level, which, thanks to Equation (2.1.3) or (2.2.2), satisfy the Markov property (i.e., it is conditionally independent of the past, given the current level). This means that the behavior of the EHD depends on its SoC: intuitively, when the battery is almost fully charged, the policy will be more aggressive and consume more energy. Vice-versa, when the battery is almost in outage, the policy will be more conservative and consume less resources.

Transition Probabilities. The transition probabilities from one state of charge to another depend on the energy harvesting process and on the policy according to Equation (2.2.2). In general, the transitions are not deterministic. Indeed, with EEH, the future energy arrivals are known only statistically because of the intrinsic random

nature of the environment, whereas with RF-WET the channel conditions (in particular the fading effects) make the energy arrivals unknown a priori. On the other hand, the *policy*, which will be formally specified in the next chapters, directly influences the energy consumption (and thus the transition probabilities of the MC) of the device.

Actions. The actions of the MDP are defined by the *policy*. Depending on the scenario under investigation, the policy may decide the transmission power, the transmission duration, the coding, the probability of accessing the channel, etc.

Reward. The reward is generally given by the total amount of bits that the device transmits in a given amount of time, i.e., the transmission rate, although also some variations of that are presented (e.g., in Section 5.2). Different actions lead to different rewards.

The above model was proposed in the literature [92] for traditional EH systems; our results are based on that with some extensions (e.g., in Section 3.2 we will use a Decentralized-MDP instead of a standard MDP) and different considerations.

Some common tools used to solve MDPs are the Value Iteration Algorithm (VIA) or the Policy Iteration Algorithm (PIA) [14]. Both algorithms entail a common phase, namely *policy improvement step*, in which the Bellman equation is recursively solved until convergence (see [14] for more details). In addition to that, PIA also adds a *value determination step* to speed up the convergence of the procedure. In the next chapters, we will use and, eventually, extend these two methods.

CHAPTER 3

Energy Harvesting Networks with Multiple Devices

Many of the protocols proposed in the Energy Harvesting (EH) literature considered isolated nodes and did not account for the interactions among devices. However, when multiple nodes are considered, new network considerations may arise and new transmission policies must be considered. For example, simultaneous transmissions toward the same Receiver (RX) may lead to unrecoverable collisions and thus degrade the system performance. This aspect should be explicitly taken into account in the system design.

Coordinated and uncoordinated approaches have been followed in the literature so far to deal with the Medium Access Control (MAC) scheduling problem. Protocols such as Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), etc. fall into the first category, since all nodes in the network only use a predefined amount of resources assigned by RX. Instead, Aloha-like protocols constitute the set of uncoordinated schemes, in which multiple nodes take individual actions to achieve a common goal (e.g., maximizing the throughput of the network). When EH is considered, both coordinated and uncoordinated schemes change according to the energy availability of the devices. The study of these two paradigms represents the focus of this chapter.

Structure of the Chapter. The chapter is divided into the following two parts. The two sections can be read separately using Chapter 2 as a baseline.

Section 3.1 focuses on the fully centralized (i.e., coordinated) problem in which nodes are managed by RX and analyzes low-complexity policies for such a scenario. In particular, we consider the case of a pair of Energy Harvesting Devices (EHDs), with the main goal of maximizing the long-term aggregate average potential reward associated with the transmitted data. The devices, at each time instant, have data of different potential rewards to be transmitted, as well as different battery energy levels. In order to avoid collisions, the Receiver (RX) allows at most the transmission of a single EHD per time slot. Assuming a negligible processing cost in terms of energy, our objective is to identify low-complexity transmission policies, that achieve good performance with respect to the optimal one. We numerically show that two policies, namely the Balanced Policy (BP) and the Heuristic-CEIP (HCEIP), despite being independent from the battery energy levels, achieve near optimal performance in most cases of interest, and can be easily found with an adaptation to the ambient energy supply. Moreover, we derive analytically an

approximation of BP and we show that this policy can be considered a good lower bound for the performance of the optimal policy.

Section 3.2 studies the decentralized optimization in which the RX cannot communicate with the nodes all the time and thus introduces a compromise between coordinated and uncoordinated schemes. Designing decentralized policies for wireless communication networks is a crucial problem, which has only been partially solved in the literature so far. In the second part of this chapter, we propose a Decentralized Markov Decision Process (Dec-MDP) framework to analyze a wireless sensor network with multiple users which access a common wireless channel. We consider devices with energy harvesting capabilities, that aim at balancing the energy arrivals with the data departures and with the probability of colliding with other nodes. Over time, the receiver triggers a SYNC slot, wherein it recomputes the optimal transmission parameters of the whole network, and distributes this information. Every node receives its own policy, which specifies how it should access the channel in the future, and, thereafter, proceeds in a *fully decentralized* fashion, with no interactions with other entities in the network. We propose a multi-layer Markov model, where an external MDP manages the jumps between SYNC slots, and an internal Dec-MDP computes the optimal policy in the short term. We numerically show that, because of the harvesting, stationary policies are suboptimal in energy harvesting scenarios, and the optimal trade-off lies between an orthogonal and a random access system.

3.1 Centralized Approach

3.1.1 Introduction

In this section, as a first step towards the study and design of multi-user EH systems, we study a system model consisting of two EHDs that report data with different potential rewards to a common Receiver (RX), with the aim of maximizing the long-term aggregate average potential reward of the reported data. In order to avoid collisions, we consider a centralized case, in which the receiver allows at most the transmission of one EHD at the same time. RX knows the battery energy level and the potential reward of each EHD. This approach can be considered as an upper bound to the decentralized scenario described in the second part of this chapter in terms of overall performance.

Contributions. Using the structure of the optimal policy, we derive different suboptimal policies which are shown to perform closely to the optimal one. First, we introduce the Energy Independent Policy (EIP), i.e., a threshold policy which, on average, transmits with a constant probability, independent of the battery energy levels. Furthermore, in accordance to the values of the average energy harvesting rates, we define the Balanced Policy (BP) and the Heuristic-CEIP (HCEIP), that are particular cases of EIP. These are *low-complexity* policies, that do not require any optimization process to compute the transmission probabilities. The main implication of these results is that near-optimal performance can be obtained *without* precise knowledge of the energy stored in the sensor batteries at any given time and with a *simple adaptation* to the ambient energy supply. Moreover, we find an analytical approximation of the balanced policy, numerically showing that this can be considered as a good analytical lower bound of the optimal policy in most cases of interest.

Structure. The first part of the chapter is organized as follow. In Section 3.1.2 we present the system model. In Section 3.1.3 we discuss the general optimization problem

and introduce the analyzed policies. Section 3.1.4 studies the balanced policy and an analytical approximation of its performance. In Section 3.1.5 we derive the heuristic constrained energy independent policy. Section 3.1.6 presents the numerical results. Finally, Section 3.1.7 concludes this part of the thesis.

References. This section is based on the conference paper **[C1]**.

3.1.2 System Model

Consider a WSN composed of two EHDs, reporting data to a common Receiver (RX) via a shared wireless channel. Time is slotted, ans slot k corresponds to the time interval $[k, k+1)$, $k = 0, 1, \ldots$. We assume that in every time slot k each device has a packet to transmit, i.e., its data queue is nonempty. Packet transmission occupies one slot and, if not sent, the packet is lost. The receiver always allows only the transmission of one EHD, so as to avoid collisions and, hence, the loss of both packets. The battery of each EHD is modeled as an energy buffer, and we assume that each position in the buffer contains exactly one energy quantum (e.q.) (see the discrete model of Section 2.2.1). In addition, as assumed in [55, 129], the transmission of one data packet requires the expenditure of one energy quantum. The maximum amount of energy that can be stored in the battery of EHD i is denoted as $b_{\max}^{(i)}$, and the set of its possible energy levels is $\mathcal{B}^{(i)} = \{0, 1, \ldots, b_{\max}^{(i)}\}$, $i = 1, 2$. If the energy level of EHD i in slot k is $b_k^{(i)} \in \mathcal{B}^{(i)}$, the evolution of $b_k^{(i)}$ is described by Equation (2.2.2), which in this scenario can be simplified to

$$b_{k+1}^{(i)} = \min\{b_{\max}^{(i)}, b_i^{(k)} - q_i^k + e_k^{(i)}\}, \tag{3.1.1}$$

where $\{e_k^{(i)}\}$ is the energy arrival process (we omit the superscript "EEH" because we only consider environmental energy sources), modeling the randomness in the amount of energy that can be harvested from the environment. In the following, we assume that $\{e_k^{(i)}\}$ is an independent and indentically distributed (i.i.d.) Bernoulli process with mean $\bar{e}^{(i)} \in (0, 1]$, independent across EHDs, and that a quantum harvested in slot k can only be exploited at a subsequent time instant $> k$. Furthermore, the amount of energy devoted to transmission by EHD i in slot k is represented by $\{q_k^{(i)}\}$, the action process, which is one if the current data packet is transmitted (and one energy quantum is drawn from the buffer and consumed), and zero otherwise (note that $q_k^{(i)}$ of Equation (3.1.1) coincides with $q^{(i)}(P_k^{(i)})$ of Equation (2.2.2) when the value of $P_k^{(i)}$ can only be 0 or equal to a fixed value). Clearly, due to the adopted collision model, the elements of the pair $\mathbf{q}_k = (q_k^{(1)}, q_k^{(2)})$ cannot be simultaneously positive.

The assumption of two users can be extended in order to consider larger networks, but this would require a higher computational cost. However, conceptually, the step from the single user to the case with two users is the most interesting one because channel contention problems and collision problems firstly arise.

In slot k, the state of the system is $\mathbf{S}_k = (\mathbf{b}_k, \boldsymbol{\nu}_k) = (b_k^{(1)}, b_k^{(2)}, \nu_k^{(1)}, \nu_k^{(2)})$, where $\nu_k^{(1)}$ and $\nu_k^{(2)}$ are the potential rewards of the current data packets at both EHDs. The term $\nu_k^{(i)}$ can be interpreted as the realization of a positive continuous random variable $V_k^{(i)}$ with probability density function (pdf) $f_{V^{(i)}}(\nu)$, $\nu \geq 0$, and we assume that $\{V^{(i)}\}$ are i.i.d. across time and EHDs.

3.1.3 Policy Definition and General Optimization Problem

For a given value of $\mathbf{S}_k$, a policy μ decides on the amounts of energy $(q_k^{(1)}, q_k^{(2)}) \in \{(0,0),(0,1),(1,0)\}$ to devote to transmission at time k. In detail, μ is a probability measure on the action space $\{(0,0),(0,1),(1,0)\}$, parameterized by state $\mathbf{S}_k$: given $\mathbf{S}_k = (\mathbf{b}_k, \boldsymbol{\nu}_k) \in \mathcal{B}^{(1)} \times \mathcal{B}^{(2)} \times (\mathbb{R}^+)^2$, $\mu(\boldsymbol{q}_k; \mathbf{b}_k, \boldsymbol{\nu}_k)$ is the probability of drawing $q_k^{(1)}$ and $q_k^{(2)}$ energy quanta from the two EHDs. Clearly, to avoid collisions, the central controller prevents $q_k^{(1)}$ and $q_k^{(2)}$ from being simultaneously equal to one.

Using policy μ, the long-term average reward is defined as

$$G(\mu, \mathbf{S}_0) = \liminf_{K\to\infty} \frac{1}{K} \mathbb{E}\left[\sum_{k=0}^{K-1} \left(q_k^{(1)}\nu_k^{(1)} + q_k^{(2)}\nu_k^{(2)}\right) \middle| \mathbf{S}_0\right] \tag{3.1.2}$$

where $\mathbf{S}_0$ is the initial state of the system, $\mathbf{q}_k$ is selected according to μ, and the expectation is taken with respect to the energy arrival, the potential packet reward and action processes. Consequently, the optimization problem is to determine the optimal policy μ^* such that

$$\mu^* = \arg\max_{\mu} G(\mu, \mathbf{S}_0). \tag{3.1.3}$$

As stated in [33], the optimal policy μ^*, must have a threshold structure with respect to the importance of the current data packet: a pair of thresholds $(\nu_{\text{th}}^{(1)}(\mathbf{b}), \nu_{\text{th}}^{(2)}(\mathbf{b}))$ is associated to every pair of joint energy levels $\mathbf{b} \in \mathcal{B}^{(1)} \times \mathcal{B}^{(2)}$, such that

$$\begin{cases} \mu((1,0);\mathbf{b},\boldsymbol{\nu}) = 1, & \nu^{(1)} > \nu_{\text{th}}^{(1)}(\mathbf{b}) \quad \text{and} \quad \nu^{(1)} - \nu_{\text{th}}^{(1)}(\mathbf{b}) \geq \nu^{(2)} - \nu_{\text{th}}^{(2)}(\mathbf{b}), \\ \mu((0,1);\mathbf{b},\boldsymbol{\nu}) = 1, & \nu^{(2)} > \nu_{\text{th}}^{(2)}(\mathbf{b}) \quad \text{and} \quad \nu^{(2)} - \nu_{\text{th}}^{(2)}(\mathbf{b}) > \nu^{(1)} - \nu_{\text{th}}^{(1)}(\mathbf{b}), \\ \mu((0,0);\mathbf{b},\boldsymbol{\nu}) = 1, & \nu^{(1)} \leq \nu_{\text{th}}^{(1)}(\mathbf{b}) \quad \text{and} \quad \nu^{(2)} \leq \nu_{\text{th}}^{(2)}(\mathbf{b}). \end{cases} \tag{3.1.4}$$

Consequently, in the following we consider only the subset of policies with such threshold structure.

It is now possible to define the marginal transmission probability of EHD i, when the joint energy energy level state is $\mathbf{b} = (b^{(1)}, b^{(2)})$

$$\omega^{(i)}(\mathbf{b}) \triangleq \mathbb{E}[q_k^{(i)} = 1 | \mathbf{b}_k = \mathbf{b}], \quad i = 1, 2, \tag{3.1.5}$$

whereas the probability of no EHD transmitting is denoted by $\omega^{(0)}(\mathbf{b}) = 1 - \omega^{(1)}(\mathbf{b}) - \omega^{(2)}(\mathbf{b})$. The expected reward can thus be defined as a function of the marginal probabilities $\omega^{(i)}(\mathbf{b})$:

$$r(\omega^{(1)}(\mathbf{b}), \omega^{(2)}(\mathbf{b})) = \mathbb{E}\left[q_k^{(1)}\nu_k^{(1)} + q_k^{(2)}\nu_k^{(2)} | \mathbf{b}_k = \mathbf{b}\right]. \tag{3.1.6}$$

The threshold structure allows a one-to-one mapping between $v_{\text{th}}^{(i)}$, $\omega^{(i)}$ and μ, and hence both the policy μ and the transition probabilities of the time-homogeneous Markov Chain (MC) related to the energy states, can be reformulated as a function of the pair $(\omega^{(1)}, \omega^{(2)})$. Consequently, if $(\omega^{(1)}, \omega^{(2)})$ induces an irreducible MC, the long-term reward

of Equation (3.1.2) does not depend on the initial state $\mathbf{S}_0$, and can be formulated as

$$G(\omega^{(1)},\omega^{(2)}) = \sum_{b^{(1)}=0}^{b^{(1)}_{\max}} \sum_{b^{(2)}=0}^{b^{(2)}_{\max}} \pi_{\omega}(b^{(1)},b^{(2)}) r(\omega^{(1)}(b^{(1)},b^{(2)}),\omega^{(2)}(b^{(1)},b^{(2)})), \tag{3.1.7}$$

with $\pi_{\omega}(b^{(1)},b^{(2)})$ being the steady-state distribution of the joint energy levels given by policy $\boldsymbol{\omega} = (\omega^{(1)},\omega^{(2)})$. The optimization problem (3.1.3) becomes

$$\boldsymbol{\omega}^{\star} = \arg\max_{\boldsymbol{\omega}} G(\boldsymbol{\omega}), \tag{3.1.8}$$

and can be solved via standard stochastic optimization techniques, like the Value Iteration Algorithm (VIA) or the Policy Iteration Algorithm (PIA) (see Section 2.2.2).

3.1.3.1 Maximization of the Transmission Rate

A practically important case is the optimization of Equation (3.1.3) for the case in which the goal is the maximization of the long-term average transmission rate from the EHDs to the receiver. Denoting the normalized channel gain random variables as $H^{(1)}$ and $H^{(2)}$, and assuming them i.i.d. across EHDs and over time and exponentially distributed with unit mean (i.e., with pdf $f_{H^{(i)}}(\hbar) = e^{-\hbar},\ \hbar > 0$, where e is the Napier's constant), the total SNR enjoyed by device i in slot k is $\Lambda^{(i)}\hbar_k^{(i)}$, where $\Lambda^{(i)}$ is the average SNR of link i at RX and $\hbar_k^{(i)}$ is the realization of the random variable $H^{(i)}$. Hence, the rate achievable by EHD i in slot k is proportional to

$$V_k^{(i)} = \log(1+\Lambda^{(i)}\hbar_k^{(i)}), \tag{3.1.9}$$

and a threshold on the importance level $\nu_{\text{th}}^{(i)}$ corresponds to a threshold on the normalized channel gain $\hbar_{\text{th}}^{(i)} = (e^{\nu_{\text{th}}^{(i)}}-1)/\Lambda^{(i)}$. As derived in [33], in the practical case in which the sensor nodes are energy constrained and $\Lambda^{(i)} \ll 1$ (low SNR regime [132]), the Shannon capacity expression (3.1.9) can be approximated as $V_k^{(i)} \simeq \Lambda^{(i)}\hbar_k^{(i)}$ and the formulations for the marginal probabilities and the reward function (3.1.6) become, respectively:

$$\omega^{(0)}(\mathbf{b}) = \left(1-e^{-\hbar_{\text{th}}^{(1)}(\mathbf{b})}\right)\left(1-e^{-\hbar_{\text{th}}^{(2)}(\mathbf{b})}\right), \tag{3.1.10}$$

$$\omega^{(1)}(\mathbf{b}) = e^{-\hbar_{\text{th}}^{(1)}(\mathbf{b})}\left(1-\frac{\Lambda^{(2)}}{\Lambda^{(1)}+\Lambda^{(2)}}e^{-\hbar_{\text{th}}^{(2)}(\mathbf{b})}\right), \tag{3.1.11}$$

$$\omega^{(2)}(\mathbf{b}) = e^{-\hbar_{\text{th}}^{(2)}(\mathbf{b})}\left(1-\frac{\Lambda^{(1)}}{\Lambda^{(1)}+\Lambda^{(2)}}e^{-\hbar_{\text{th}}^{(1)}(\mathbf{b})}\right) \tag{3.1.12}$$

and

$$\begin{aligned} r(\omega^{(1)}(\mathbf{b}),\omega^{(2)}(\mathbf{b})) = &\sum_{i=1}^{2}\Lambda^{(i)}e^{-\hbar_{\text{th}}^{(i)}(\mathbf{b})}(\hbar_{\text{th}}^{(i)}(\mathbf{b})+1) \\ &+\frac{\Lambda^{(1)}\Lambda^{(2)}}{\Lambda^{(1)}+\Lambda^{(2)}}e^{-(\hbar_{\text{th}}^{(1)}(\mathbf{b})+\hbar_{\text{th}}^{(2)}(\mathbf{b}))}(\hbar_{\text{th}}^{(1)}(\mathbf{b})+\hbar_{\text{th}}^{(2)}(\mathbf{b})+1). \end{aligned} \tag{3.1.13}$$

In addition, it is possible to derive the channel thresholds $\hbar_{\text{th}}^{(1)}(\mathbf{b})$ and $\hbar_{\text{th}}^{(2)}(\mathbf{b})$ as functions of the marginal probabilities $\omega^{(1)}(\mathbf{b})$ and $\omega^{(2)}(\mathbf{b})$ by performing the inversion of Equations (3.1.11) and (3.1.12):

$$\hbar_{\text{th}}^{(1)}(\mathbf{b}) = \log\left(\frac{-(\omega^{(1)}(\mathbf{b})+\omega^{(2)}(\mathbf{b}))\Lambda^{(2)}+(\Lambda^{(1)}+\Lambda^{(2)})(\omega^{(1)}(\mathbf{b})+1)+\Delta}{2(\Lambda^{(1)}+\Lambda^{(2)})\omega^{(1)}(\mathbf{b})}\right), \tag{3.1.14}$$

$$\hbar_{\text{th}}^{(2)}(\mathbf{b}) = \log\left(\frac{2\Lambda^{(2)}}{(\omega^{(1)}(\mathbf{b})+\omega^{(2)}(\mathbf{b}))\Lambda^{(2)}+(\Lambda^{(1)}+\Lambda^{(2)})(1-\omega^{(1)}(\mathbf{b}))-\Delta}\right), \tag{3.1.15}$$

where

$$\Delta \triangleq \sqrt{[(\omega^{(1)}(\mathbf{b})+\omega^{(2)}(\mathbf{b}))\Lambda^{(2)}+(\Lambda^{(1)}+\Lambda^{(2)})(1-\omega^{(1)}(\mathbf{b}))]^2-4\Lambda^{(2)}\omega^{(2)}(\mathbf{b})(\Lambda^{(1)}+\Lambda^{(2)})}. \tag{3.1.16}$$

3.1.3.2 Definition of the Analyzed Policies

In Table 3.1 the following policies are compared, according to the value of $\bar{e}^{(1)}+\bar{e}^{(2)}$, where $\bar{e}^{(i)}$ is the average energy arrival rate of device i expressed in energy quanta.

- **Optimal Policy (OP).** The optimal policy that, for each $\bar{e}^{(1)}$, $\bar{e}^{(2)}$, $b_{\max}^{(1)}$ and $b_{\max}^{(2)}$, identifies the values of $\omega^{(1)}(\boldsymbol{b})$, $\omega^{(2)}(\boldsymbol{b})$ maximizing Equation (3.1.7). This policy can be calculated using VIA or PIA.

- **Energy Independent Policy (EIP).** A policy in which $\omega^{(1)}(\boldsymbol{b}) = \omega^{(1)}\chi\{b^{(1)}>0\}$ and $\omega^{(2)}(\boldsymbol{b}) = \omega^{(2)}\chi\{b^{(2)}>0\}$, where $\chi\{\cdot\}$ is the indicator function and $\omega^{(1)}, \omega^{(2)}$ are scalar values. With this policy, the values of $\omega^{(1)}(\boldsymbol{b})$ and $\omega^{(2)}(\boldsymbol{b})$ do not depend on the battery status, provided that the batteries are not empty. In particular, we are interested in the Optimal-EIP (OEIP). Special cases of EIP include:

  - **Balanced Policy (BP).** A particular case of EIP, only defined for $\bar{e}^{(1)}+\bar{e}^{(2)} \leq 1$, where $\omega^{(1)} = \bar{e}^{(1)}$ and $\omega^{(2)} = \bar{e}^{(2)}$, which can be shown to be asymptotically optimal for large batteries. An approximation of the performance obtained with this policy will be discussed in Section 3.1.4.

  - **Constrained Energy Independent Policy (CEIP).** A particular case of EIP, only defined for $\bar{e}^{(1)}+\bar{e}^{(2)} > 1$, where $\omega^{(1)}+\omega^{(2)} = 1$ and $\min\{\omega^{(1)}, \omega^{(2)}\} \leq \min\{\bar{e}^{(1)}, \bar{e}^{(2)}\}$. In particular, we are interested in the Optimal-CEIP (OCEIP).

  - **Heuristic-CEIP (HCEIP).** A heuristic approximation of OCEIP, which will be discussed in Section 3.1.5.

Table 3.1. Available policies for different values of $\bar{e}^{(1)}+\bar{e}^{(2)}$.

EH condition	Available policies
$\bar{e}^{(1)}+\bar{e}^{(2)} \leq 1$	OP, EIP, BP
$\bar{e}^{(1)}+\bar{e}^{(2)} > 1$	OP, EIP, CEIP, HCEIP

Our objective is to analyze suboptimal low-complexity policies, avoiding the computationally demanding optimization processes needed to compute OP, but still achieving good performance with respect to it. In particular, BP and HCEIP do not require any optimization, and hence are ideal for low-performing EHDs.

The comparison of the performance achieved by the introduced policies is based on the following metric.

Definition 3.1.1 (Reward Precision)**.** *The reward precision of two policies A and B is defined as:*

$$\mathcal{R}_A^B \triangleq \frac{G_A - G_B}{G_A}, \tag{3.1.17}$$

where G_A and G_B are the long-term reward of policies A and B, respectively, as defined in Equation (3.1.7)*.*

For example, a policy B can be considered a good lower bound of policy A if $\mathcal{R}_A^B \ll 1$.

In Section 3.1.4, we consider the case $\bar{e}^{(1)} + \bar{e}^{(2)} \leq 1$, first discussing the balanced policy, and then deriving an analytical approximation of its reward function G_{BP}. Section 3.1.5, instead, describes an energy independent policy for the case $\bar{e}^{(1)} + \bar{e}^{(2)} \geq 1$, defining a heuristic low-complexity policy HCEIP, for which no optimization is required.

3.1.4 Balanced Policy

The Balanced Policy (BP) is a particular case of energy independent policy and is defined only when $\bar{e}^{(1)} + \bar{e}^{(2)} \leq 1$, as $\omega^{(i)}(\mathbf{b}) = \bar{e}^{(i)}\chi\{b_i > 0\}$.

As an analytical formulation for G_{BP} cannot be easily computed, we now derive its approximation $\hat{G}_{\text{BP}}$, so as to characterize the performance obtained by BP in a closed-form expression. In addition, we will numerically show that $G_{\text{BP}} \geq \hat{G}_{\text{BP}}$ and, since $G_{\text{OP}} \geq G_{\text{BP}}$, it would be possible to find an approximate analytical lower bound to the optimum reward G_{OP}.

The basic idea is to divide the set of energy states in four classes and force the steady-state probability of all states in the same class to be equal. Also, solving a reduced system of steady-state equations, it will be possible to find explicitly the value of the *approximate steady-state distribution* $\hat{\pi}_\omega(\mathbf{b})$, from which the *approximate reward function* $\hat{G}_{\text{BP}}$ can be computed. After the analysis of some properties of $\hat{G}_{\text{BP}}$, in Section 3.1.6 we numerically evaluate the goodness of this approximation with respect to the balanced policy G_{BP}.

3.1.4.1 Computation of the Approximate Reward Function

The computation of the exact steady state distribution would involve the solution of a system of $(b_{\max}^{(1)} + 1) \times (b_{\max}^{(2)} + 1)$ equations.

In order to simplify the problem, we use the following working assumption

$$\hat{\pi}_\omega(b^{(1)}, b^{(2)}) = \begin{cases} \pi_{00}, & \text{if } b^{(1)} = 0,\ b^{(2)} = 0, \\ \pi_{b^{(1)}}, & \text{if } b^{(1)} > 0,\ b^{(2)} = 0, \\ \pi_{b^{(2)}}, & \text{if } b^{(1)} = 0,\ b^{(2)} > 0, \\ \pi_{\mathbf{b}}, & \text{if } b^{(1)} > 0,\ b^{(2)} > 0 \end{cases} \tag{3.1.18}$$

and we find the previous values solving the reduced system of equations involving states $(0,1)$, $(1,0)$ and $(1,1)$ and the normalization equation

$$\sum_{b^{(1)}=0}^{b_{\max}^{(1)}} \sum_{b^{(2)}=0}^{b_{\max}^{(2)}} \hat{\pi}_{\omega}(b^{(1)}, b^{(2)}) = 1. \tag{3.1.19}$$

This analysis requires $b_{\max}^{(1)} > 1$ and $b_{\max}^{(2)} > 1$, otherwise the structure of the system of steady-state equations would be different. Thus, we neglect the case $b_{\max}^{(1)} = 1$ or $b_{\max}^{(2)} = 1$, which seems of less practical importance, and can be solved in a similar way. In particular, we obtain:

$$\begin{pmatrix} \pi_{00} \\ \pi_{b^{(1)}} \\ \pi_{b^{(2)}} \\ \pi_{\mathbf{b}} \end{pmatrix} = \frac{1}{D} \begin{pmatrix} 1 - \bar{e}^{(1)} - \bar{e}^{(2)} + 2\,\bar{e}^{(1)}\bar{e}^{(2)} \\ 1 - \bar{e}^{(2)} \\ 1 - \bar{e}^{(1)} \\ 1 \end{pmatrix}, \tag{3.1.20}$$

where

$$D = (b_{\max}^{(1)} + 1)(b_{\max}^{(2)} + 1) - \bar{e}^{(1)}(b_{\max}^{(2)} + 1) - \bar{e}^{(2)}(b_{\max}^{(2)} + 1) + 2\,\bar{e}^{(1)}\bar{e}^{(2)}. \tag{3.1.21}$$

The approximate reward function $\hat{G}_{\mathrm{BP}}$ is thus derived as:

$$\hat{G}_{\mathrm{BP}} = b_{\max}^{(1)}\pi_{b^{(1)}} r^{(1)} + b_{\max}^{(2)}\pi_{b^{(2)}} r^{(2)} + b_{\max}^{(1)} b_{\max}^{(2)} \pi_{\mathbf{b}}\, r^{(1,2)}, \tag{3.1.22}$$

where we have defined $r^{(1)} \triangleq r(\bar{e}^{(1)}, 0)$, $r^{(2)} \triangleq r(0, \bar{e}^{(2)})$ and $r^{(1,2)} \triangleq r(\bar{e}^{(1)}, \bar{e}^{(2)})$.

3.1.4.2 Approximate Reward Function Properties

Limit values. We want to investigate the values that $\hat{G}_{\mathrm{BP}}$ assumes when $b_{\max}^{(1)} \to \infty$ or $b_{\max}^{(2)} \to \infty$ or both.

$$\lim_{b_{\max}^{(1)} \to \infty} \hat{G}_{\mathrm{BP}} = \frac{(1 - \bar{e}^{(2)}) r^{(1)} + b_{\max}^{(2)} r^{(1,2)}}{b_{\max}^{(2)} + 1 - \bar{e}^{(2)}}, \tag{3.1.23}$$

$$\lim_{b_{\max}^{(2)} \to \infty} \hat{G}_{\mathrm{BP}} = \frac{(1 - \bar{e}^{(1)}) r^{(2)} + b_{\max}^{(1)} r^{(1,2)}}{b_{\max}^{(1)} + 1 - \bar{e}^{(1)}}, \tag{3.1.24}$$

$$\lim_{\substack{b_{\max}^{(1)} \to \infty \\ b_{\max}^{(2)} \to \infty}} \hat{G}_{\mathrm{BP}} = r^{(1,2)}. \tag{3.1.25}$$

From the previous equations, it is possible to derive the gap between the asymptotic reward (when both $b_{\max}^{(i)} \to \infty$) and the one achieved in a scenario where only a single device i has a large battery, as $r^{(1,2)} - \lim_{b_{\max}^{(i)} \to \infty} \hat{G}_{\mathrm{BP}}$. For example, if $b_{\max}^{(1)} \to \infty$, we have

$$r^{(1,2)} - \lim_{b_{\max}^{(1)} \to \infty} \hat{G}_{\mathrm{BP}} = (1 - \bar{e}^{(2)}) \frac{r^{(1,2)} - r^{(1)}}{b_{\max}^{(2)} + 1 - \bar{e}^{(2)}}, \tag{3.1.26}$$

which allows us to justify the reward loss in the non symmetric case.

OP bounds. In Section 3.1.6, we will show that the approximate reward function $\hat{G}_{\mathrm{BP}}$ numerically results to be a lower bound for the balanced policy. Furthermore, as $G_{\mathrm{BP}} \leq G_{\mathrm{OP}}$ for every parameter choice and, if $\bar{e}^{(1)} + \bar{e}^{(2)} \leq 1$, $G_{\mathrm{OP}} \leq r^{(1,2)}$ (see [33]), the following inequality chain holds:

$$\hat{G}_{\mathrm{BP}} \leq G_{\mathrm{BP}} \leq G_{\mathrm{OP}} \leq r^{(1,2)}, \tag{3.1.27}$$

i.e., $\hat{G}_{\mathrm{BP}}$ can be used as an analytical lower bound for the optimal reward G_{OP}.

3.1.5 Heuristic Constrained Energy Independent Policy

In this section we consider the case $\bar{e}^{(1)} + \bar{e}^{(2)} > 1$, first discussing the Constrained Energy Independent Policy (CEIP) and then introducing its heuristic low-complexity version (HCEIP). The latter is interesting because:

- For a given set of $\bar{e}^{(1)}$, $\bar{e}^{(2)}$, $b^{(1)}_{\max}$ and $b^{(2)}_{\max}$, it provides the values of $\omega^{(1)}$ and $\omega^{(2)}$ with no optimization needed.
- It achieves near optimal performance among the CEIPs, i.e., $\mathcal{R}^{\mathrm{OCEIP}}_{\mathrm{HCEIP}} \ll 1$.

Note that in the case $\bar{e}^{(1)} + \bar{e}^{(2)} > 1$, the balanced policy where $\omega^{(1)} = \bar{e}^{(1)}$, $\omega^{(2)} = \bar{e}^{(2)}$ cannot be defined, otherwise $\omega^{(1)} + \omega^{(2)} > 1$, which would be infeasible. Therefore, we must have $\omega^{(1)} + \omega^{(2)} < \bar{e}^{(1)} + \bar{e}^{(2)}$; however the choice of the marginal transmission probabilities is not obvious.

A constrained energy-independent policy can be seen as a particular EIP, with the following additional constraints:

- $\omega^{(1)} + \omega^{(2)} = 1$, i.e., $\omega^{(0)}$ is forced to be zero;
- $\min\{\omega^{(1)}, \omega^{(2)}\} \leq \min\{\bar{e}^{(1)}, \bar{e}^{(2)}\}$. This constraint will be useful to design HCEIP described next.

CEIP can be defined for every $\bar{e}^{(1)} + \bar{e}^{(2)}$, but we focus on the case $\bar{e}^{(1)} + \bar{e}^{(2)} > 1$, as, if $\bar{e}^{(1)} + \bar{e}^{(2)} \leq 1$, the simpler balanced policy already achieves good performance. Although the Optimal-CEIP (OCEIP) behaves differently from the Optimal-EIP, i.e., $\omega^{(i)}_{\mathrm{OCEIP}} \neq \omega^{(i)}_{\mathrm{OEIP}}$, a CEIP has the peculiarity of allowing to reduce the number of the variables from two to one and hence, to compute the OCEIP, only one parameter needs to be optimized. Exploiting this fact, in the next subsection we define the Heuristic-CEIP, an approximation of OCEIP, whose objective is to achieve good performance avoiding the resource-demanding optimization process. Also, we consider only the case $b_{\max} \triangleq b^{(1)}_{\max} = b^{(2)}_{\max}$, leaving the asymmetric case as future work.

3.1.5.1 Computation of the Heuristic-CEIP

The design of HCEIP is based on the analysis of the marginal transmission probabilities of the Optimal-CEIP. In particular we approximate $\omega^{(1)}_{\mathrm{OCEIP}}$ and $\omega^{(2)}_{\mathrm{OCEIP}}$ with $\omega^{(1)}_{\mathrm{HCEIP}}$ and $\omega^{(2)}_{\mathrm{HCEIP}}$, which are simple functions of the system parameters.

By evaluating OCEIP, for $\bar{e}^{(1)}, \bar{e}^{(2)} \in (0, 1]$, with $\bar{e}^{(2)} \leq \bar{e}^{(1)}$ and $\bar{e}^{(1)} + \bar{e}^{(2)} > 1$, for a given value of $\bar{e}^{(1)}$ and $b_{\max}$ the behavior of $\omega^{(2)}_{\mathrm{OCEIP}}$ as a function of $\bar{e}^{(2)}$ can be divided into three regions: a first linear zone with slope equal to one ($\omega^{(2)}_{\mathrm{OCEIP}} = \bar{e}^{(2)}$); a second non-linear part; a last constant zone ($\omega^{(2)}_{\mathrm{OCEIP}} = 0.5$). Defining the thresholds dividing the

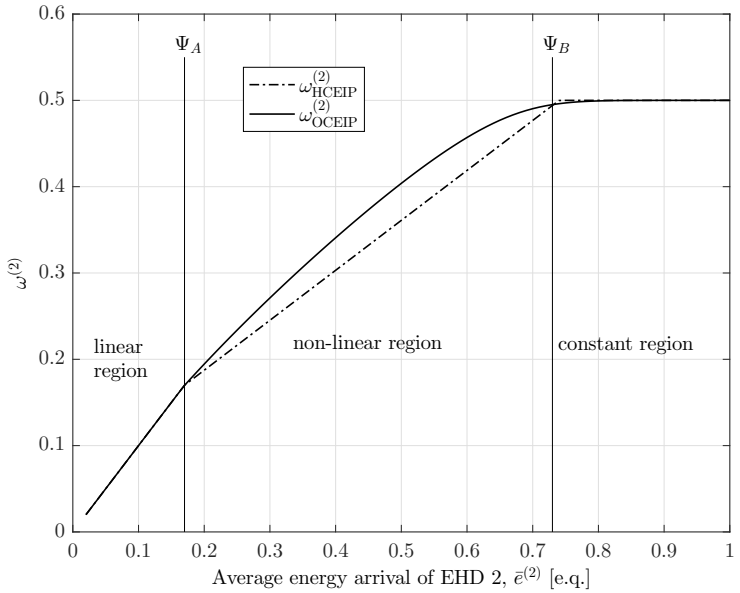

Figure 3.1. $\omega^{(2)}_{\text{HCEIP}}$ and $\omega^{(2)}_{\text{OCEIP}}$ as a function of $\bar{e}^{(2)}$ in the range $[0.02, 1]$ with $\bar{e}^{(1)} = 0.98$ and $b_{\max} = 6$.

three regions as Ψ_A and Ψ_B (with $\Psi_A < \Psi_B$), as in Figure 3.1, we now introduce the heuristic marginal transmission probability $\omega^{(2)}_{\text{HCEIP}}$ (note that $\omega^{(1)}_{\text{OCEIP}} = 1 - \omega^{(2)}_{\text{OCEIP}}$).

Given Ψ_A and Ψ_B for each $\bar{e}^{(1)}$ and $b_{\max}$, we deduce $\omega^{(2)}_{\text{HCEIP}}$ as a function of these thresholds, approximating the central region of Figure 3.1 as a straight line. With the previous assumptions, we have:

$$\omega^{(2)}_{\text{HCEIP}} = \begin{cases} \bar{e}^{(2)}, & \text{if } \bar{e}^{(2)} \leq \Psi_A, \\ \dfrac{0.5 - \Psi_A}{\Psi_B - \Psi_A}(\bar{e}^{(2)} - \Psi_A) + \Psi_A, & \text{if } \Psi_A < \bar{e}^{(2)} < \Psi_B, \\ 0.5, & \text{if } \Psi_B \leq \bar{e}^{(2)}. \end{cases} \tag{3.1.28}$$

Thresholds. In the general case, each threshold is a function of both $b_{\max}$ and $\bar{e}^{(1)}$. However, it can be numerically verified that:

$$\Psi_A \triangleq \Psi_A(\bar{e}^{(1)}, b_{\max}) \approx \Psi_A(\bar{e}^{(1)}), \tag{3.1.29}$$

$$\Psi_B \triangleq \Psi_B(\bar{e}^{(1)}, b_{\max}) \approx \min\{\Psi_B(b_{\max}), \bar{e}^{(1)}\}, \tag{3.1.30}$$

that is, Ψ_A only depends on $\bar{e}^{(1)}$, while Ψ_B also depends on a function $\Psi_B(b_{\max})$, defined in (3.1.32).

In Figure 3.1 we show, as an example, both $\omega^{(2)}_{\text{OCEIP}}$ and $\omega^{(2)}_{\text{HCEIP}}$ as a function of $\bar{e}^{(2)}$, with the thresholds Ψ_A and Ψ_B when $\bar{e}^{(1)} = 0.98$ and $b_{\max} = 6$.

We now discuss how the derivation of Ψ_A and Ψ_B has been accomplished. Ψ_A has been computed numerically, for different values of $b_{\max}$ and $\bar{e}^{(1)}$, as the value for which $\omega^{(2)}$ is no longer equal to $\bar{e}^{(2)}$. Since it can be observed that Ψ_A is approximately independent of $b_{\max}$, for each value of $\bar{e}^{(1)}$ we have dropped this dependence averaging Ψ_A over all possible

values of $b_{\max}$. Finally, by a linear interpolation with respect to $\bar{e}^{(1)}$, Equation (3.1.31) has been derived.

The threshold Ψ_B has been determined as the value for which $\omega^{(2)}$ saturates to 0.5. We have found that Ψ_B has the structure in (3.1.30), where Ψ_B has been determined with a technique similar to the one used to find Ψ_A, first averaging on $\bar{e}^{(1)}$ and then interpolating in $b_{\max}$, resulting in (3.1.32).

The resulting heuristic expressions of the thresholds $\Psi_A(\bar{e}^{(1)})$ and $\Psi_B(b_{\max})$ are given by

$$\Psi_A(\bar{e}^{(1)}) = -0.6875 \times \bar{e}^{(1)} + 0.84375, \tag{3.1.31}$$

$$\Psi_B(b_{\max}) = 0.5159e^{-0.1775\times b_{\max}} + 0.5624 \tag{3.1.32}$$

and are valid for every possible choice of $b_{\max}$ and $\bar{e}^{(1)}$ and hence represent a general result for this scenario.

From Equations (3.1.29)-(3.1.32), it is possible to compute Ψ_A and Ψ_B, used to derive $\omega^{(2)}_{\text{HCEIP}}$ from (3.1.28) and $\omega^{(1)}_{\text{HCEIP}} = 1 - \omega^{(2)}_{\text{HCEIP}}$, for all values of $\bar{e}^{(1)}$, $\bar{e}^{(2)}$ and $b_{\max}$. Finally, with the marginal probabilities, G_{HCEIP} is obtained.

3.1.6 Numerical Results

In our numerical evaluation we use the following parameters: $b^{(1)}_{\max}, b^{(2)}_{\max} \in \{1, \ldots, 20\}$, $\bar{e}^{(1)}, \bar{e}^{(2)} \in \{0.05, 0.10, \ldots, 1\}$ and $\Lambda^{(1)} = \Lambda^{(2)}$. Since there are twenty cases for each parameter, we have a total number of cases equal to $20^4 = 160000$.

In the following, we show that the reward functions $\hat{G}_{\text{BP}}$ and G_{HCEIP} are good lower bounds for G_{OP} in most cases of interest.[1]

Table 3.2. Policies comparison in the worst-case scenarios.

EH condition	Policies		$\max\{\mathcal{R}^B_A\}$
$\bar{e}^{(1)} + \bar{e}^{(2)} \gtreqless 1$	OP	OEIP	5.36%
$\bar{e}^{(1)} + \bar{e}^{(2)} \leq 1$	OEIP	BP	10.17%
	OP	BP	10.18%
	BP	$\hat{\text{BP}}$	4.14%
$\bar{e}^{(1)} + \bar{e}^{(2)} > 1$	OEIP	OCEIP	1.59%
	OCEIP	HCEIP	0.45%
	OP	HCEIP	4.56%

We now comment the results shown in Table 3.2, based on the value of $\bar{b}_1 + \bar{b}_2$.

3.1.6.1 Case $\bar{e}^{(1)} + \bar{e}^{(2)} \lesseqgtr 1$

OP and OEIP comparison. As $\max\{\mathcal{R}^{\text{OEIP}}_{\text{OP}}\} = 5.36\%$, OEIP is a good lower bound for OP. This is an important result, which implies that only few parameters need to be optimized to achieve near-optimal performance.

[1] In the following comparison, $\hat{\text{BP}}$ refers to the approximate performance of BP, obtained with Equation (3.1.22).

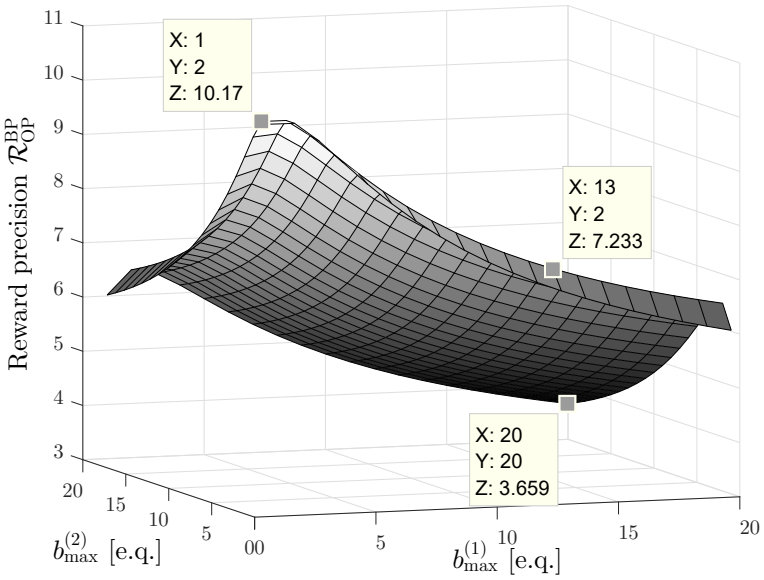

Figure 3.2. Percentage reward precision $\mathcal{R}_{\mathrm{OP}}^{\mathrm{BP}}$ when $\bar{e}^{(1)} = \bar{e}^{(2)} = 0.05$ (worst case) and $b_{\max}^{(1)}$ and $b_{\max}^{(2)}$ range between 1 and 20.

3.1.6.2 Case $\bar{e}^{(1)} + \bar{e}^{(2)} \leq 1$

OEIP and BP comparison. The balanced policy is not always a good lower bound for OEIP, as $\max\{\mathcal{R}_{\mathrm{OEIP}}^{\mathrm{BP}}\} = 10.17\%$. However, in general, BP is rather close to OEIP, especially if $b_{\max}^{(1)}$ and/or $b_{\max}^{(2)}$ are not too small. In particular, we have noticed that $\mathcal{R}_{\mathrm{OEIP}}^{\mathrm{BP}}$ is high when both $\bar{e}^{(1)}$ and $\bar{e}^{(2)}$ are close to 0, which is not a very practical scenario. For example, if $b_{\max}^{(1)} = b_{\max}^{(2)} = 1$ and $\bar{e}^{(1)} = 0.4$, $\bar{e}^{(2)} = 0.2$, $\mathcal{R}_{\mathrm{OEIP}}^{\mathrm{BP}} = 1.2\%$, but if $b_{\max}^{(1)} = b_{\max}^{(2)} = 5$, this value decreases to 0.14%. We can finally say that BP is a good lower bound of OEIP in most cases of interest.

OP and BP comparison. From the previous paragraphs, we can state that in most cases of interest BP is a good lower bound for OP. As an example, in Figure 3.2, we plot the worst case scenario ($\bar{e}^{(1)} = \bar{e}^{(2)} = 0.05$) for $\mathcal{R}_{\mathrm{OP}}^{\mathrm{BP}}$, with different values of $b_{\max}^{(i)}$. It can be seen that $\mathcal{R}_{\mathrm{OP}}^{\mathrm{BP}}$ decreases quickly when $b_{\max}^{(1)}$ and/or $b_{\max}^{(2)}$ increase, i.e., the balanced policy is better for high values of $b_{\max}$.

$\hat{\mathrm{BP}}$ and BP comparison. For each pair of $b_{\max}^{(1)}$ and $b_{\max}^{(2)}$, the maximum of $\mathcal{R}_{\mathrm{BP}}^{\hat{\mathrm{BP}}}$ is reached when $\bar{e}^{(1)} = \bar{e}^{(2)} = 0.5$ and, in particular, in the worst possible case $\mathcal{R}_{\mathrm{BP}}^{\hat{\mathrm{BP}}}$ is equal to 4.14%. Moreover, the lower bound of $\mathcal{R}_{\mathrm{BP}}^{\hat{\mathrm{BP}}}$ is 0, i.e., $\hat{G}_{\mathrm{BP}}$ is always lower than G_{BP}. Since $\mathcal{R}_{\mathrm{BP}}^{\hat{\mathrm{BP}}} \ll 1$, we can state that the approximate balanced policy can be considered a good lower bound for BP. Note that this result is not obvious, as $\hat{G}_{\mathrm{BP}}$ has been derived as an approximation of G_{BP}. In Figure 3.3, we depict the reward precision $\mathcal{R}_{\mathrm{BP}}^{\hat{\mathrm{BP}}}$ for $b_{\max}^{(1)} = b_{\max}^{(2)} = 2$, for different values of $\bar{e}^{(1)}$ and $\bar{e}^{(2)}$. It can be seen that $\bar{e}^{(1)} = \bar{e}^{(2)} = 0.5$ is the worst case, but for low values of $\bar{e}^{(1)}$ or $\bar{e}^{(2)}$, the two rewards are comparable.

3.1.6.3 Case $\bar{e}^{(1)} + \bar{e}^{(2)} > 1$

OEIP and OCEIP comparison. The difference between OCEIP and OEIP is mainly due to the constraint $\omega^{(1)} + \omega^{(2)} = 1$. However, we have verified that $\omega_{\mathrm{OEIP}}^{(1)} + \omega_{\mathrm{OEIP}}^{(2)}$ is

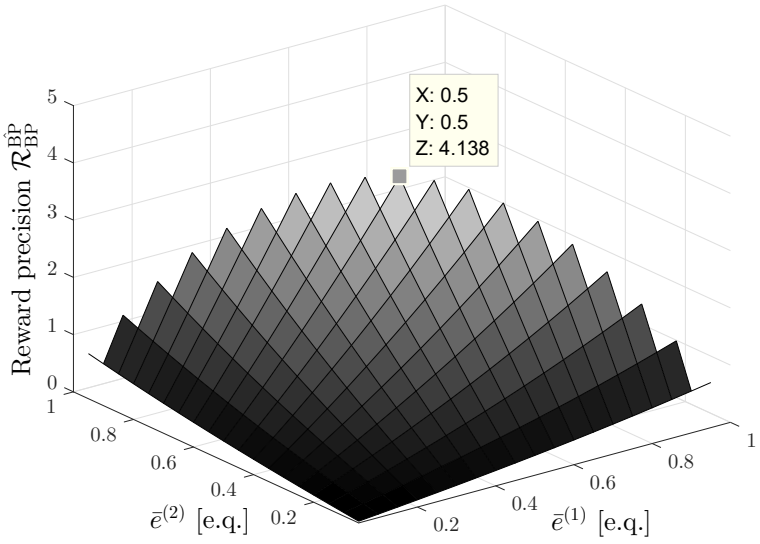

Figure 3.3. Percentage reward precision $\mathcal{R}_{\mathrm{BP}}^{\hat{\mathrm{BP}}}$ when $b_{\max}^{(1)} = b_{\max}^{(2)} = 2$ and $\bar{e}^{(1)}$ and $\bar{e}^{(2)}$ range between 0.05 and 1 with $\bar{e}^{(1)} + \bar{e}^{(2)} \leq 1$.

approximately equal to one, even for the optimal unconstrained EIP. The results show that $\mathcal{R}_{\mathrm{OEIP}}^{\mathrm{OCEIP}}$ is always lower than 1.59%, and the worst case occurs when $\bar{e}^{(1)} + \bar{e}^{(2)}$ is close to 1. Indeed, in this case, OEIP is more conservative than OCEIP, i.e., $\omega_{\mathrm{OEIP}}^{(1)} < \omega_{\mathrm{OCEIP}}^{(1)}$ and $\omega_{\mathrm{OEIP}}^{(2)} < \omega_{\mathrm{OCEIP}}^{(2)}$, because OEIP generally attempts to avoid energy outage, whose probability increases if $\bar{e}^{(1)}$ and $\bar{e}^{(2)}$ decrease.

OCEIP and HCEIP comparison. We have verified that the quantity $\mathcal{R}_{\mathrm{OCEIP}}^{\mathrm{HCEIP}}$ has a maximum that is less than 0.45%, i.e., with the considered parameters, the heuristic approximation can be considered as a good lower bound of OCEIP. Clearly, the approximation performs worse when $\bar{e}^{(2)} \in (\Psi_A, \Psi_B)$, where we have fitted a non-linear function with a linear one. In Figure 3.4, we show the reward precision $\mathcal{R}_{\mathrm{OCEIP}}^{\mathrm{HCEIP}}$ as a function of $\bar{e}^{(2)}$ for the same parameters of Figure 3.1. It can be seen that $\mathcal{R}_{\mathrm{OCEIP}}^{\mathrm{HCEIP}}$ is low even in the non-linear zone, where the approximation is worse.

OP and HCEIP comparison. By the aforementioned results, it follows that

$$G_{\mathrm{OP}} \gtrsim G_{\mathrm{OEIP}} \gtrsim G_{\mathrm{OCEIP}} \gtrsim G_{\mathrm{HCEIP}}, \tag{3.1.33}$$

i.e., HCEIP is a good lower bound for OP. In particular, from Table 3.2, $\max\{\mathcal{R}_{\mathrm{OP}}^{\mathrm{HCEIP}}\} = 4.56\%$: this is an interesting result, as HCEIP, differently from OP, can be analytically formulated with a closed form expression.

Finally, in Figures 3.5 and 3.6 we compare G_{OP}, G_{OEIP}, $\hat{G}_{\mathrm{BP}}$, and G_{HCEIP}, when $b_{\max}^{(1)} = b_{\max}^{(2)} \in \{1, \ldots, 20\}$. It can be seen that all the policies approach OP for high values of $b_{\max}$ and are very close already for $b_{\max} = 20$.

3.1.7 Conclusions of Section 3.1

The first part of this chapter focused on the case of two energy harvesting devices which report data associated to different rewards to a common access point, and are managed by a central controller. We distinguished two cases, depending on the global energy harvesting

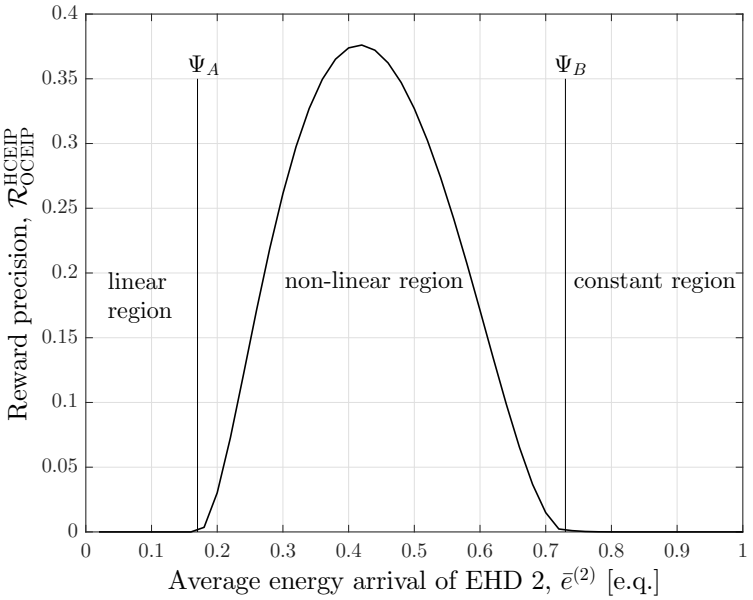

Figure 3.4. Percentage reward precision $\mathcal{R}_{OCEIP}^{\text{HCEIP}}$ as a function of $\bar{e}^{(2)}$ that range in $[0.02, 1]$ with $\bar{e}^{(1)} = 0.98$ and $b_{\max} = 6$.

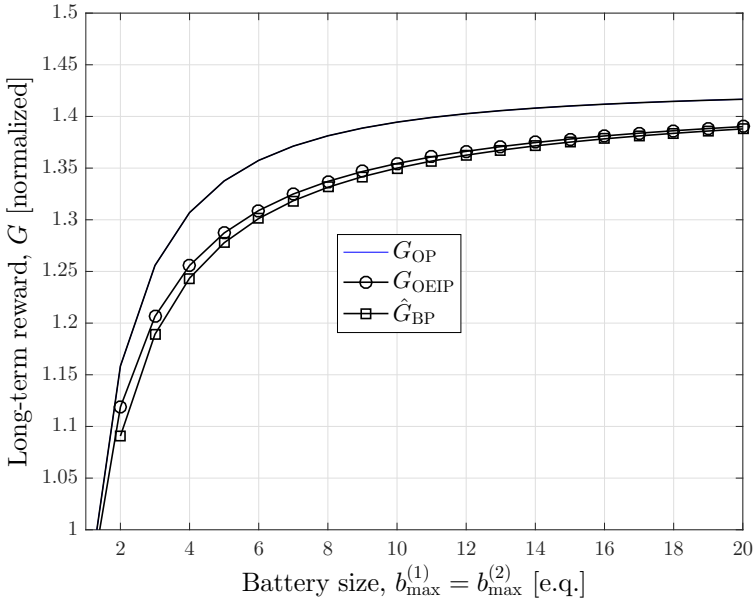

Figure 3.5. Comparison of different reward functions G when $b_{\max}^{(1)} = b_{\max}^{(2)} \in \{1, \ldots, 20\}$ and $\bar{e}^{(1)} = \bar{e}^{(2)} = 0.4$.

rate of the system. In the first one, when $\bar{e}^{(1)} + \bar{e}^{(2)} \leq 1$, we used the balanced policy and we derived an analytic approximation of its performance. We showed that BP can be considered a good lower bound for the optimal policy when $\bar{e}^{(1)}$ and $\bar{e}^{(2)}$ are not too close to zero. Furthermore, we numerically derived an approximate reward function of BP, that is always lower than or equal to the real reward function. In this way, we found an analytical lower bound to BP and, consequently, to the optimal policy. Future work

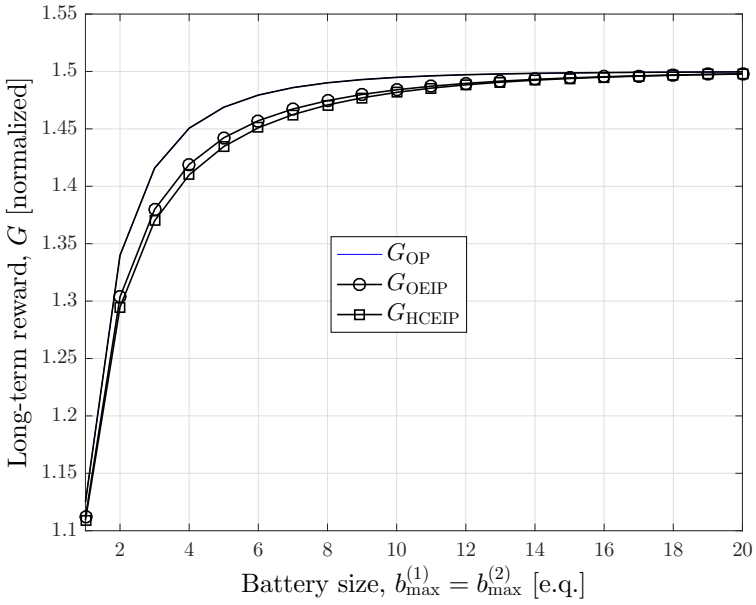

Figure 3.6. Comparison of different reward functions G when $b_{\max}^{(1)} = b_{\max}^{(2)} \in \{1, \dots, 20\}$ and $\bar{e}^{(1)} = \bar{e}^{(2)} = 0.55$.

may complete the analysis of BP in the particular case $b_{\max}^{(1)} = 1$ or $b_{\max}^{(2)} = 1$. In the case $\bar{e}^{(1)} + \bar{e}^{(2)} > 1$, we introduced a chain of policies that leads to the heuristic constrained energy independent policy, which is computable with no optimization which nevertheless achieves good performance with respect to the optimal policy. Future work may also include the case $b_{\max}^{(1)} \neq b_{\max}^{(2)}$.

3.2 Decentralized Approach

In this section we introduce a *decentralized* optimization framework for an energy harvesting network. Differently from Section 3.1, in this case the receiver is unaware of the state of the system in every slot, and thus it must compute the policy using partial information only.

3.2.1 Introduction

Environmental Energy Harvesting (EEH) has been established as one of the most prominent solutions for prolonging the lifetime and enhancing the performance of wireless networks. Although this topic has been widely investigated in the literature so far, finding proper energy management schemes is still an open issue in many cases of interest. In particular, using *decentralized* policies, in which every node in the network acts autonomously and independently of the others, is a major problem of practical interest in Wireless Sensor Networks (WSNs) where a central controller may not be used all the time. Many decentralized communication schemes (e.g., Aloha-like) can be found in the literature; however, most of them were designed without a principle of *optimality*, i.e., without explicitly trying to maximize the network performance. Instead, in this section we characterize the optimal decentralized policy in a WSN with EH constraints and describe the related computational issues. Although this approach intrinsically leads to a more complex protocol definition, it

also characterizes the maximum performance a network can achieve, and may serve as a benchmark for defining quasi-optimal low-complexity protocols.

Differently from the approach we investigate in this section, most of the protocols proposed in the literature considered isolated nodes and did not account for the interactions among devices, or focused on centralized policies, in which a controller coordinates all nodes and knows the global state of the system over time. [94] analyzed decentralized policies with a particular focus on symmetric systems, and proposed a game theoretic approach for solving the problem. Instead, in this section we use a different framework based on decentralized Markov decision processes, which can handle asymmetric scenarios. Decentralized theory was also used in [52] for a wireless powered communication network. However, the scenario proposed therein is different from ours and the authors only focused on a narrow subclass of policies. [91] considered a scenario similar to ours and used a low-complexity approach to solve access problem; in particular, they replaced the battery dynamics with average power constraints, therefore simplifying the system model but introducing suboptimal policies.

To model our energy harvesting system, we use the results about decentralized control theory developed by Dibangoye *et al.* [36, 37, 35]. In particular, [35] presented a detailed study of Decentralized Partially Observable Markov Decision Processes (Dec-POMDPs) and proposed different approaches to solve them. The notion of *occupancy state* was introduced as a fundamental building block for Dec-POMDPs, and it was shown that, differently from classic statistical descriptions (e.g., belief states), it represents a sufficient statistic for control purposes. Using the occupancy state, we can convert the Dec-POMDP in an equivalent MDP with a continuous state space, called *occupancy-MDP*. Then, standard techniques to solve POMDPs and MDPs can be applied; for example, an approach to solve a continuous state space MDP is to define a grid of points (see Lovejoy's grid approximation [83]) and solve the MDP only in a subset of states. Although several papers introduced more advanced techniques to refine the grid [170], this approach may still be inefficient and difficult to apply. Instead, in this section we use a different scheme, namely the Learning Real Time A* (LRTA*) algorithm [66], which has the key advantage of exploring only the states which are actually visited by the process, without the difficulty of defining a grid of points.

Converting the Dec-POMDP to an occupancy-MDP produces a simpler formulation of the problem, which however does not reduce its complexity. Indeed, for every occupancy state, it is still required to perform an *exhaustive backup* operation, i.e., to compute a decentralized control policy. This is the most critical operation in decentralized optimization, since it involves solving a non-convex problem with many variables. Dibangoye *et al.* proposed an alternate formulation of the exhaustive backup operation as a Constraint Program [31], which can be solved, e.g., using the bucket-elimination algorithm [30]. The problem can be further simplified by imposing a predefined structure to the policy [52], so that only few parameters need to be optimized. While this may lead to suboptimal solutions, it greatly simplifies the numerical evaluation and, if correctly designed, produces close to optimal results. In this section, we explore and compare both these directions.

Contributions. We consider a decentralized network with multiple devices and a receiver that computes and distributes to all nodes the randomized transmission policy. A multi-layer Markov model, in which an internal Dec-MDP is nested inside an external MDP, is proposed and solved. The external layer models the time instants, namely SYNC slots, at which RX computes the policy, whereas the internal layer models the system

evolution between consecutive SYNC slots. To solve the external layer, we use the Value Iteration Algorithm (VIA) (see Section 2.2.2) as in [17] or our paper **[J3]** (see page 155). However, differently from these papers, in our model the transition probabilities between states are derived from the optimization of the internal layer; moreover, the sojourn times in every state are not deterministic. Instead, the internal layer is solved using the Markov Policy Search algorithm [36]. Because of the complexity of the optimal approach, we introduce two simpler schemes, which still exploit the structure of the optimal policy but can be computed in practice. In our numerical results we compare centralized and decentralized approaches, and discuss the performance loss of using a decentralized scheme.

Our main contributions lie in the design and analysis of decentralized policies and of the tools to construct them for a star topology network; in particular, they can be summarized as follows

1. We present a decentralized random access transmission scheme derived using a principle of *optimality*, and discuss which are the computational pitfalls of this approach;

2. We introduce two suboptimal policies, which are closely related to the structure of the optimal policy but can be numerically computed with reasonable complexity. These can be used as a baseline for developing heuristic schemes and real-time protocols. Moreover, although we present these approaches for an EH scenario, they may also be used in other contexts;

3. We show that a decentralized scheme, if correctly designed, may achieve high performance, comparable with that of centralized solutions, while greatly reducing the signaling in the network;

4. Finally, we show that, differently from traditional networks (i.e., without energy constraints), where using an orthogonal resource allocation all the time is optimal, the best transmission policy with energy harvesting is a hybrid approach between random and orthogonal access.

Structure. The second part of this chapter is organized as follows. Section 3.2.2 presents the system model. The internal layer is described in Section 3.2.3, whereas the external layer and the optimization problem are shown in Section 3.2.4. Optimal and suboptimal solutions are derived in Sections 3.2.5 and 3.2.6, respectively. The numerical results are shown in Section 3.2.7. Finally, Section 3.2.8 concludes the section.

References. This section is based on the conference paper **[C13]** and on the journal paper **[J6]**.

3.2.2 System Model

The network is composed of one Receiver (RX) and N harvesting nodes. We focus on an infinite time horizon framework, where a time slot k corresponds to the time interval $[kT, (k+1)T)$, $k = 0, 1, \ldots$, and T is the common duration of all slots (see Section 2.1.1). During a slot, every node independently decides whether to access the uplink channel and transmit a message to RX, or to remain idle. We adopt an on/off collision model in which overlapping packet transmissions are always unrecoverable.

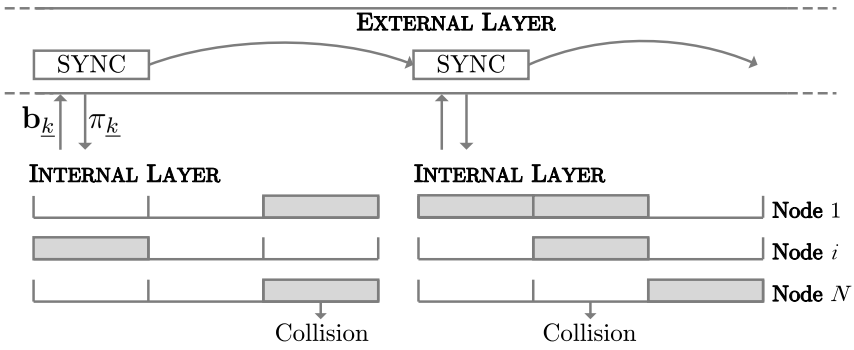

Figure 3.7. Time evolution of the system. After the SYNC slots every user acts independently of the others.

In slot k, node i harvests energy from the environment according to a distribution $E_k^{\mathrm{EEH},i}$, independent of the other nodes. In this section, we consider Bernoulli energy arrival processes.

Every node stores the harvested energy in a rechargeable battery. We assume that the energy scavenged in slot k can be used only in a later slot. The global energy level vector in slot k is $\mathbf{b}_k = \langle b_k^{(1)}, \dots, b_k^{(N)} \rangle$. Vector $\mathbf{b}_k$ is not known to the devices, which only see their own battery status, and is known to RX only in certain slots, namely the "SYNC slots". In particular, in slot k, with probability (w.p.) $\beta_k \in [0,1]$, RX may trigger a SYNC slot and request all nodes to share their energy levels so that it can acquire $\mathbf{b}_k$. We neglect the energy costs of these synchronization messages, which however, if considered, would make the benefits of using decentralized policies even higher. For example, when $\beta_k = 1$, $\forall k$, RX has full knowledge of the battery levels in every time slot, and our model degenerates to [33] (see Section 3.2.4.2); instead, if $\beta_k = \beta_0$, $\forall k$, where β_0 is a constant value in $(0,1]$, RX uses a fully stochastic approach and asks for $\mathbf{b}_k$ with the same probability in every slot. Another possibility may be $\beta_k = 1$ for some predefined k, and $\beta_k = 0$ for others, so that the SYNC slots are deterministically defined a priori.

RX uses the information about $\mathbf{b}_k$ to initialize the transmission parameters of the whole network. Therefore, every time $\mathbf{b}_k$ is acquired (a SYNC slot), a coordination phase is performed and RX disseminates the *policy* to all nodes (the policy is decentralized, so every node receives only its own policy).[2] Thereafter, every device acts independently of the others until the next SYNC slot.

Although the proposed framework is very simple, modeling and solving it formally requires a complex mathematical structure. In particular, we decompose the system in two nested layers (see Figure 3.7 for a graphical illustration):

- The external layer considers the jumps between consecutive SYNC slots. Indeed, since the global battery level $\mathbf{b}_k$ is completely known in every SYNC state, the system follows a Markov evolution;
- The internal layer models the actions to take between two SYNC slots and requires to compute a decentralized policy given $\mathbf{b}_k$. This will be modeled as a Dec-MDP, since multiple devices indirectly collaborate to achieve a common goal.

[2]We note that, although a user may also receive the policy of other devices, this information would not be useful. Indeed, the decentralized transmission policy is jointly designed by RX, and therefore implicitly considers the contributions of all nodes.

The two layers will be separately analyzed in the following sections.

3.2.3 Internal Layer

We first consider the internal layer and present a mathematical tool to model the actions of the devices between two SYNC slots. In particular, we adopt the Decentralized Markov Decision Process (Dec-MDP) framework [37], which in our context is formally defined as follows.

3.2.3.1 Decentralized–MDPs for EH Systems

An N-user Dec-MDP $\mathcal{M} = (\underline{k}, \boldsymbol{\mathcal{B}}, \boldsymbol{\Omega}, p_{\text{int}}, r, \eta_{\underline{k}}, \beta)$ is specified by

- **Initial Index.** $\underline{k}$ represents the index of the SYNC slot that triggers the beginning of the internal layer. Thus, all the slots $k < \underline{k}$ are of no interest in this section. Note that, since β_k is a probability, the position of the next SYNC slot is unknown a priori, therefore the time horizon of the Dec-MDP $\mathcal{M}$ begins at $\underline{k}$ and may extend to arbitrarily large time indices;

- **Battery Level.** $\boldsymbol{\mathcal{B}} = \mathcal{B}^{(1)} \times \cdots \times \mathcal{B}^{(N)}$ is the set of global battery levels $\mathbf{b}_k = \langle b_k^{(1)}, \ldots, b_k^{(N)} \rangle$, with $b_k^{(i)} \in \mathcal{B}^{(i)} \triangleq \{0, \ldots, b_{\max}^{(i)}\}$ (device i can store up to $b_{\max}^{(i)}$ discrete energy quanta). Throughout, the terms "battery level" or "state" will be used interchangeably;

- **Action.** $\boldsymbol{\Omega} = \Omega^{(1)} \times \cdots \times \Omega^{(N)}$ is the set of global actions $\boldsymbol{\omega}_k = \langle \omega_k^{(1)}, \ldots, \omega_k^{(N)} \rangle$, where $\omega_k^{(i)} \in \Omega^{(i)} \triangleq [0, 1]$ denotes the transmission probability. Although $\omega_k^{(i)}$ should assume continuous values, we only consider S_a uniformly distributed samples of the interval $[0, 1]$ for numerical tractability. Action $\omega_k^{(i)}$ is chosen by user i in slot $k \geq \underline{k}$ through a function $\sigma_k^{(i)} : \mathcal{B}^{(i)} \to \Omega^{(i)}$, and depends only on the local state $b_k^{(i)}$. Finding $\sigma_k^{(i)}$ will be the objective of the optimization problem;

- **Reward.** r is the reward function $r : \boldsymbol{\mathcal{B}} \times \boldsymbol{\Omega} \to \mathbb{R}^+$ that maps the global action $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ to the reward $r(\mathbf{b}, \boldsymbol{\omega})$ when the global state is $\mathbf{b} \in \boldsymbol{\mathcal{B}}$;

- **Transition Probability.** p_{int} is the probability transition function $p_{\text{int}} : \boldsymbol{\mathcal{B}} \times \boldsymbol{\Omega} \times \boldsymbol{\mathcal{B}} \to [0, 1]$ which defines the probability $p_{\text{int}}(\bar{\mathbf{b}}|\mathbf{b}, \boldsymbol{\omega})$ of moving from a global battery level $\mathbf{b} = \langle b^{(1)}, \ldots, b^{(N)} \rangle \in \boldsymbol{\mathcal{B}}$ to a global battery level $\bar{\mathbf{b}} = \langle \bar{b}^{(1)}, \ldots, \bar{b}^{(N)} \rangle \in \boldsymbol{\mathcal{B}}$ under the global action $\boldsymbol{\omega} \in \boldsymbol{\Omega}$. When user i performs a transmission, it consumes $m^{(i)} \geq 1$ energy quanta;

- **Initial State Distribution.** $\eta_{\underline{k}}$ is the initial state distribution. In our scenario we take

$$\eta_{\underline{k}}(\mathbf{b}) = \begin{cases} 1, & \text{if } \mathbf{b} = \mathbf{b}_{\underline{k}}, \\ 0, & \text{if } \mathbf{b} \in \boldsymbol{\mathcal{B}} \setminus \{\mathbf{b}_{\underline{k}}\}, \end{cases} \tag{3.2.1}$$

  where $\mathbf{b}_{\underline{k}}$ is the global state in correspondence of the initial SYNC slot and is fully known by RX;

- **SYNC Probability.** β represents the sequence $\beta_{\underline{k}}$, $\beta_{\underline{k}+1}, \ldots$, which are the probabilities that a SYNC slot occurs.

In Section 3.2.4.1 we will describe the optimization problem related to $\mathcal{M}$. Its solution provides a *decentralized control policy*, which will be discussed in Sections 3.2.5 and 3.2.6.

Before presenting in more detail the previous bullet points, it is important to emphasize the following key characteristics of the Dec-MDP under investigation:

- $\mathcal{M}$ is *jointly fully observable*, i.e., if all nodes collaborated and shared their local energy levels, the global state would be completely known (actually, this is what differentiates Dec-MDPs from Dec-POMDPs [7]);
- $\mathcal{M}$ is a *transition independent* Dec-MDP, i.e., the action taken by node i influences only its own battery evolution in that slot and *not* the others. Formally, the transition probability function p_{int} can be decomposed as

$$p_{\text{int}}(\bar{\mathbf{b}}|\mathbf{b}, \boldsymbol{\omega}) = \prod_{i=1}^{N} p_{\text{int}}^{(i)}(\bar{b}^{(i)}|b^{(i)}, \omega^{(i)}). \tag{3.2.2}$$

  This feature is important to develop compact representations of the transmission policies, and in particular to derive Markovian policies as discussed in our Section 3.2.3.3 and in [36, Theorem 1].

3.2.3.2 Battery Level

We adopt a discrete model for the energy-related quantities, so that every battery can be referred to as an energy queue, in which arrivals coincide with the energy harvesting process, and departures with packet transmissions. In particular, a battery of $B_{\max}^{(i)}$ J is quantized in $b_{\max}+1$ uniform levels $\{0, 1, \ldots, b_{\max}^{(i)}\}$; in this context, the basic energy unit is called *energy quantum* and corresponds to $B_{\max}^{(i)}/b_{\max}^{(i)}$ J of energy. This model has been widely used in the EH literature [17, 94, 156], and represents a good approximation of a real battery when $b_{\max}^{(i)}$ is sufficiently high, i.e., the battery quantization is fine (see Section 2.2.1).

In particular, the battery level of node i in slot k is $B_k^{(i)}$, and it corresponds to $b_k^{(i)}$ energy quanta. The battery status evolves as

$$b_{k+1}^{(i)} = \min\{b_{\max}^{(i)}, b_k^{(i)} - q_k^{(i)} + e_k^{(i)}\}, \tag{3.2.3}$$

where the "min" accounts for the finite battery size, $q_k^{(i)}$ is the energy used for transmission and $e_k^{(i)}$ is the energy arrived in slot k. $q_k^{(i)}$ is equal to 0 w.p. $1-\omega_k^{(i)}$, and to $m^{(i)}$ w.p. $\omega_k^{(i)}$.

3.2.3.3 Action and Reward

We will use the term "global reward" to indicate the overall performance of the system in a slot, and simply "single-user reward" to refer to the performance of individual users.

Single-User Reward. Assume to study isolated users, which do not suffer from interference, as in [93]. Data messages are associated with a *potential reward*, described by a random variable $V^{(i)}$ which evolves independently over time and among nodes. The realization $\nu_k^{(i)}$ is perfectly known only at a time $t \geq kT$ and only to node i; for $t < kT$, only a statistical knowledge is available. Every node can decide to transmit

(and accrue the potential reward $\nu_k^{(i)}$) or not in the current slot k according to its value $\nu_k^{(i)}$ (see [92]). Our model is very general and can be specialized to different practical cases. For example, in our numerical evaluation we will consider Shannon capacity and set $V^{(i)} = \log(1+\Lambda^{(i)}H^{(i)})$, where $H^{(i)}$ is the fading random variable. Another example would be to consider transmissions with different levels of *importance*, according to a random distribution.

It can be shown that a threshold transmission model is optimal for this system [93]; thus, node i always transmits when $\nu_k^{(i)} \geq \nu_{\text{th}}^{(i)}(b^{(i)})$ and does not otherwise. Note that $\nu_{\text{th}}^{(i)}(b^{(i)})$ depends on the underlying state (battery level) of user i but not on the time index k (thus a *stationary* scheduler can be developed).

On average, the reward of user i in a single slot when the battery level is $b^{(i)}$ will be

$$g(\nu_{\text{th}}^{(i)}(b^{(i)})) \triangleq \mathbb{E}[\chi(V^{(i)} \geq \nu_{\text{th}}^{(i)}(b^{(i)}))V^{(i)}] = \int_{\nu_{\text{th}}^{(i)}(b^{(i)})}^{\infty} \nu f_V^{(i)}(\nu)\ \mathrm{d}\nu, \tag{3.2.4}$$

where $\chi(\cdot)$ is the indicator function and $f_V^{(i)}(\cdot)$ is the pdf of the potential reward, $V^{(i)}$. It is now clear that the transmission probability $\omega^{(i)}$ is inherently dependent on the battery level as[3]

$$\omega^{(i)} = \sigma^{(i)}(b^{(i)}) = \int_{\nu_{\text{th}}^{(i)}(b^{(i)})}^{\infty} f_V^{(i)}(v)\ \mathrm{d}v = \bar{F}_V^{(i)}(\nu_{\text{th}}^{(i)}(b^{(i)})), \tag{3.2.5}$$

where we explicitly introduced the function $\sigma^{(i)}(b^{(i)})$, which maps local observations $(b^{(i)})$ to local actions $\sigma^{(i)}(b^{(i)}) = \omega^{(i)}$. Note that the complementary cumulative distribution function $\bar{F}_V^{(i)}(\cdot)$ is strictly decreasing and thus can be inverted. Therefore, there exists a one-to-one mapping between the threshold values and the transmission probabilities. In the following, we will always deal with $\omega^{(i)}$ instead of $\nu_{\text{th}}^{(i)}(\cdot)$, and write $g(\omega^{(i)})$ with a slight abuse of notation.

It follows from basic analysis that $g(\omega^{(i)})$ is increasing and concave in $\omega^{(i)}$ [92, Lemma 1], i.e., transmitting more often leads to higher rewards, but with diminishing returns. Finally, note that this model is quite general and, depending on the meaning of $V^{(i)}$, can be adapted to different scenarios. For example, in a standard communication system in which the goal is the throughput maximization, $V^{(i)}$ can be interpreted as the transmission rate subject to fading fluctuations [92].

Global Reward. The global reward is zero when multiple nodes transmit simultaneously, whereas it is equal to $w^{(i)}\nu_k^{(i)}$ if only node i transmits in slot k ($w^{(i)}$ is the weight of node i). On average, since the potential rewards are independent among nodes, we have

$$\begin{aligned} r(\nu_{\text{th},k}(\mathbf{b}_k)) = \mathbb{E}\Bigg[& \sum_{i=1}^{N} w^{(i)}\, V_k^{(i)}\, \chi(V_k^{(i)} \geq \nu_{\text{th},k}^{(i)}(b_k^{(i)})) \\ & \times \prod_{j\neq i} \chi(V_k^j < \nu_{\text{th},k}^j(b_k^j))\Bigg], \end{aligned} \tag{3.2.6}$$

[3] We adopt a probabilistic approach because, as will be clear from the numerical results, a binary choice that forces $\omega_k^{(i)} \in \{0,1\}$ is suboptimal in general.

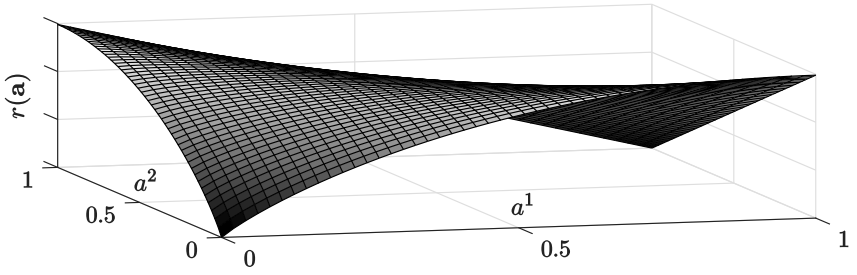

Figure 3.8. Global reward $r(\boldsymbol{\omega})$ when $N = 2$.

which can be rewritten as

$$r(\boldsymbol{\omega}_k) = r(\boldsymbol{\sigma}_k(\mathbf{b}_k)) = \sum_{i=1}^{N} w^{(i)} g(\omega_k^{(i)}) \prod_{j \neq i} (1 - \omega_k^j), \tag{3.2.7}$$

where we used $\boldsymbol{\omega}_k$ instead of $\nu_{\text{th},k}(\mathbf{b}_k)$ for ease of notation, and we introduced the vector function $\boldsymbol{\sigma}_k \triangleq \langle \sigma_k^{(1)}, \ldots, \sigma_k^{(N)} \rangle$. We highlight that $\boldsymbol{\sigma}_k$ summarizes the actions of all users given every battery level in slot k, i.e., it specifies all the following quantities

$$\begin{matrix} \sigma_k^{(1)}(0) & \ldots & \sigma_k^{(1)}(b_{\max}^{(1)}), \\ \vdots & & \\ \sigma_k^{(N)}(0) & \ldots & \sigma_k^{(N)}(b_{\max}^{(N)}). \end{matrix} \tag{3.2.8}$$

Finding $\boldsymbol{\sigma}_{\underline{k}}, \boldsymbol{\sigma}_{\underline{k}+1}, \ldots$ represents the biggest challenge when solving a Dec-MDP. Since, when a transmission is performed, $m^{(i)}$ energy quanta are drained from the battery, if we had $b_k^{(i)} < m^{(i)}$, no transmission could be performed and $\omega_k^{(i)} = \sigma_k^{(i)}(b_k^{(i)}) = 0$.

An important observation is that the reward (3.2.7) is not necessarily increasing nor convex in $\boldsymbol{\omega}$, which significantly complicates the solution. An example of $r(\boldsymbol{\omega})$ for the case with two users can be seen in Figure 3.8. Note that the maximum is achieved when only one device transmits w.p. 1 and the other does not transmit. This implies that, when the devices are not energy constrained (i.e., they have enough energy for transmitting and the current transmission policy does not influence the future), the optimal user allocation should follow an orthogonal approach so as to avoid collisions (the corner points $\langle \omega^{(1)}, \omega^{(2)} \rangle = \langle 1, 0 \rangle$ and $\langle \omega^{(1)}, \omega^{(2)} \rangle = \langle 0, 1 \rangle$ achieve the maximum reward). Although Figure 3.8 describes a system with two users only, this observation holds for a generic number of users, as shown in Lemma 3.2.2. However, as we will discuss later, when EH is taken into account, maximizing $r(\boldsymbol{\omega})$ alone is not optimal because an action in the current slot influences the future energy levels and, consequently, the future rewards. Therefore, the user allocation is not guaranteed to be orthogonal.

Note that, in the previous expressions, we have implicitly restricted our study to Markovian policies, which map local observations to local actions (i.e., $\boldsymbol{\sigma}_k(\mathbf{b}_k) = \boldsymbol{\omega}_k$). In general decentralized frameworks, tracking previous observations can be used to optimally decide the current action (i.e., $\boldsymbol{\omega}_k$ depends on $\mathbf{b}_{\underline{k}}, \ldots, \mathbf{b}_k$, $k \geq \underline{k}$). However, it can be proved [36] that under transition independent conditions (which hold in our case, see Section 3.2.3.1), Markovian policies are optimal and thus keeping track of previous states is not necessary.

3.2.3.4 Transition Probability

The transition probability function of user i, namely $p_{\text{int}}^{(i)}$ (see Equation (3.2.2)), is defined as follows

$$p_{\text{int}}^{(i)}(\bar{b}^{(i)}|b^{(i)},\omega^{(i)}) = \begin{cases} (1-\mathbb{P}_E^{(i)})\omega^{(i)}, & \text{if } \bar{b}^{(i)} = b^{(i)} - m^{(i)}, \\ (1-\mathbb{P}_E^{(i)})(1-\omega^{(i)}) + \mathbb{P}_E^{(i)}\omega^{(i)}, & \text{if } \bar{b}^{(i)} = b^{(i)} - m^{(i)} + 1, \\ \mathbb{P}_E^{(i)}(1-\omega^{(i)}), & \text{if } \bar{b}^{(i)} = b^{(i)} + 1, \\ 0, & \text{otherwise,} \end{cases} \tag{3.2.9}$$

when $\bar{b}^{(i)} < b_{\max}^{(i)}$ and $b^{(i)} \geq m^{(i)}$, whereas analogous formulas hold in the other cases. The term $\mathbb{P}_E^{(i)}$ is the probability that user i harvests one energy quantum. More sophisticated models, in which an arbitrary number of energy quanta can be simultaneously extracted can be integrated into our model (involving, however, higher computational costs).

3.2.3.5 Occupancy State

Before formulating the optimization problem in the next section, we first introduce the concept of *occupancy state*.

The occupancy state η_k is defined as

$$\eta_k(\bar{\mathbf{b}}) \triangleq \mathbb{P}(\mathbf{b}_k = \bar{\mathbf{b}}|\eta_{\underline{k}}, \boldsymbol{\sigma}_{\underline{k}}, \ldots, \boldsymbol{\sigma}_{k-1}), \quad k \geq \underline{k} \tag{3.2.10}$$

and represents a probability distribution over the battery levels given the initial distribution $\eta_{\underline{k}}$ (introduced in Equation (3.2.1)) and all decentralized decision rules prior to k.

It can be shown that the occupancy state represents a sufficient statistic for control purposes in Dec-MDPs, and can be easily updated in every slot using old occupancy states ($k > \underline{k}$):

$$\eta_k(\bar{\mathbf{b}}) = \iota(\eta_{k-1}, \boldsymbol{\sigma}_{k-1}) = \sum_{\mathbf{b}} p_{\text{int}}(\bar{\mathbf{b}}|\mathbf{b}, \boldsymbol{\sigma}_{k-1}(\mathbf{b}))\eta_{k-1}(\mathbf{b}), \tag{3.2.11}$$

where ι is the occupancy update function.

Occupancy-MDP. Dibangoye *et al.* [35] developed a technique to solve Dec-MDPs by recasting them in equivalent continuous state MDPs. Similarly to the reduction techniques of POMDPs, in which the belief is used as the state in an equivalent MDP for solution, for Dec-MDPs the occupancy state will represent the building block of the equivalent MDP (called *occupancy-MDP*). Intuitively, there is an analogy between traditional states in MDPs, beliefs in POMDPs and occupancy states in Dec-MDPs, since they all represent a sufficient statistic for control purposes.

The state space of the occupancy-MDP is the occupancy simplex, the transition rule is given by Equation (3.2.11), the action space is $\boldsymbol{\Omega}$, and the instantaneous reward corresponding to decentralized decision rule $\boldsymbol{\sigma}_k$ is

$$\rho(\eta_k, \boldsymbol{\sigma}_k) = \sum_{\bar{\mathbf{b}} \in \mathcal{B}} \eta_k(\bar{\mathbf{b}}) r(\boldsymbol{\sigma}_k(\bar{\mathbf{b}})), \tag{3.2.12}$$

Note that $\rho(\eta_k, \boldsymbol{\sigma}_k) \leq \max_{\bar{\mathbf{b}}} r(\boldsymbol{\sigma}_k(\bar{\mathbf{b}}))$, i.e., the loss of information corresponds to a lower reward. Moreover, note that if k were a SYNC slot, we would have $\rho(\eta_k, \boldsymbol{\sigma}_k) = r(\boldsymbol{\sigma}_k(\mathbf{b}_k))$.

The complete structure of the occupancy-MDP will be given in Section 3.2.5.1.

3.2.4 External Layer and Optimization Problem

So far, we have described how the system evolves between two SYNC slots. We now introduce the *external Markov Chain*, which models the long-term evolution of the network by considering the subset of all slots composed only by the SYNC slots.

Assume that, without loss of generality, the first SYNC slot occurs at $\underline{k} = 0$, and that the state of the system is $\mathbf{b}_0$. According to Section 3.2.3, RX uses $\mathbf{b}_0$ to compute and distribute to all nodes a decentralized policy $\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \ldots$. Moreover, the initial occupancy state η_0 is defined in (3.2.1), whereas the occupancy states $\eta_1, \eta_2, \ldots$ are evaluated as in Section 3.2.3.5.

Assume now that the first SYNC slot after $\underline{k} = 0$ is slot k'; thus, when the new SYNC slot occurs, we know that the transition probability from the initial global state $\mathbf{b}_0$ to the final global state $\bar{\mathbf{b}}$ is $\eta_{k'}(\bar{\mathbf{b}})$ (i.e., it is given by the occupancy state by its definition), with $\eta_{\underline{k}} = \eta_0 = \chi\{\mathbf{b} = \mathbf{b}_0\}$.

Using $\eta_{k'}(\bar{\mathbf{b}})$, we can compute the probability of going from $\mathbf{b}_0$ to $\bar{\mathbf{b}}$. Since k' is a random quantity (β represents a sequence of probabilities), we provide the expression of the probability averaged over k' (this models the jumps of the external Markov Chain (MC) between SYNC states)

$$p_{\text{ext}}(\bar{\mathbf{b}}|\mathbf{b}_0) = \sum_{k'=1}^{\infty} \beta_{k'} \left(\prod_{k''=1}^{k'-1} (1 - \beta_{k''}) \right) \eta_{k'}(\bar{\mathbf{b}}). \tag{3.2.13}$$

In the previous expression, k' represents the index > 0 of the first SYNC slot; $\beta_{k'}$ is the probability that k' is a SYNC slot; the product $\prod_{k''=1}^{k'-1}(1 - \beta_{k''})$ is the probability that no slots prior to k' are SYNC slots; $\eta_{k'}(\bar{\mathbf{b}})$ is the transition probability, which implicitly depends on $\mathbf{b}_0$. We remark that, to evaluate (3.2.13), we need the sequence $\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \ldots$ in order to compute all the future occupancy states $\eta_1, \eta_2, \ldots$. In Section 3.2.4.1, we will specify how to compute $\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \ldots$.

Observation The sequence of SYNC slots satisfies the Markov property, i.e., if k' is a SYNC slot, the system evolution for $k \geq k'$ is conditionally independent of past states given $\mathbf{b}_{k'}$.

We now formally define the optimization problem and link the internal and external layers.

3.2.4.1 Optimization Problem

Define $\rho_{k|\underline{k}} \triangleq \rho(\eta_k, \boldsymbol{\sigma}_k)$ as the decentralized reward given the SYNC slot $\underline{k}$. Then, prior to the next SYNC slot, the reward of the system will be: $\rho_{\underline{k}|\underline{k}}$ w.p. $\beta_{\underline{k}+1}$, $\rho_{\underline{k}|\underline{k}} + \rho_{\underline{k}+1|\underline{k}}$ w.p. $(1-\beta_{\underline{k}+1})\beta_{\underline{k}+2}$, $\rho_{\underline{k}|\underline{k}} + \rho_{\underline{k}+1|\underline{k}} + \rho_{\underline{k}+2|\underline{k}}$ w.p. $(1-\beta_{\underline{k}+1})(1-\beta_{\underline{k}+2})\beta_{\underline{k}+3}$, and so forth. Summing together the previous terms, and taking the average over the energy harvesting processes, we obtain the normalized average reward

$$R_{\underline{k}} \triangleq \mathbb{E}\left[\sum_{k=\underline{k}}^{\infty} \rho_{k|\underline{k}} \sum_{k'=k+1}^{\infty} \beta_{k'} \prod_{k''=\underline{k}+1}^{k'-1} (1 - \beta_{k''}) \right]. \tag{3.2.14}$$

The final goal of the system is to maximize the cumulative weighted undiscounted long-term reward, defined as

$$G(\Pi, \mathbf{b}_0) = \liminf_{K\to\infty} \frac{1}{K} \sum_{\underline{k}=0}^{K-1} \beta_{\underline{k}} \times R_{\underline{k}}(\mu_{\underline{k}}, \mathbf{b}_{\underline{k}}). \tag{3.2.15}$$

$R_{\underline{k}}(\mu_{\underline{k}}, \mathbf{b}_{\underline{k}})$ is given in (3.2.14) when the initial state of the system is $\mathbf{b}_{\underline{k}}$ and a policy $\mu_{\underline{k}} \triangleq (\boldsymbol{\sigma}_{\underline{k}}^{\mu_{\underline{k}}}, \boldsymbol{\sigma}_{\underline{k}+1}^{\mu_{\underline{k}}}, \ldots)$ is employed. Policy $\mu_{\underline{k}}$ is decentralized, and it is drawn from Π, which includes all the decentralized transmission policies $\mu_{\underline{k}}, \mu_{\underline{k}+1}, \ldots$. Equation (3.2.15) focuses on the long run because nodes in wireless sensor networks generally operate in the same condition for long times (e.g., in an environmental monitoring application). Moreover, the problem is *weighted* because of the terms $w^{(1)}, \ldots, w^{(N)}$ included in Equation (3.2.7). Using weights gives more flexibility to the system, since, by tuning them, it is possible to assign priority to one user or another, according to the network requirements, or to achieve fairness in asymmetric networks.

Since the sequence β is a design parameter and its choice is arbitrary, we restrict our attention to the following case.

Assumption The SYNC probability sequence is periodic with period τ (i.e., $\beta_k = \beta_{k+\tau},\ \forall k$).

For example, the simplest case is $\tau = 1$, and $\beta_k = \beta_0$ for every k. Under Assumption 3.2.4.1, it can be shown that (3.2.15) is equivalent to [126]

$$G(\Pi) = \sum_{\underline{k}=0}^{\tau-1} \sum_{\mathbf{b}\in\mathcal{B}} \beta_{\underline{k}} \times R_{\underline{k}}(\mu_{\mathbf{b}}, \mathbf{b}) \times \pi_{\underline{k}}(\mathbf{b}), \tag{3.2.16}$$

where $\pi_{\underline{k}}(\mathbf{b})$ is the steady-state probability of the global energy level $\mathbf{b}$ associated with $\beta_{\underline{k}}$, and, instead of iterating over all k, we take the sum over the energy levels (i.e., we iterate over the states of the external MC). Note that, in this case, the long-term undiscounted reward does not depend on the initial state of the system, therefore $G(\Pi, \mathbf{b}_0) = G(\Pi)$ for every $\mathbf{b}_0$. The optimal solution of the external problem will be

$$\Pi^\star = \arg\max_{\Pi} G(\Pi), \tag{3.2.17}$$

which is a Markov Decision Process (MDP). The underlying MC states are all the elements of $\mathcal{B}$, whereas the actions, which influence the transition probabilities (Equation (3.2.13)), are given by the evolution of the internal Dec-MDP. We note that it is fundamental to consider the two layers of optimization together, so as to avoid suboptimal approaches (e.g., focusing on the internal layer only would correspond to a greedy optimization). In the following, for the sake of presentation simplicity, we impose $\beta_k = \beta_0,\ \forall k$ (i.e., $\tau = 1$). However, the results can be straightforwardly extended to the more general case.

Value Iteration Algorithm (VIA). The optimization problem of Equation (3.2.17) can be solved using VIA [14, Vol. 1, Sec. 7.4]. Since we focus on $\tau = 1$, thanks to Equation (3.2.16) we only examine $\underline{k} = 0$ (in the more general case $\tau > 1$, the procedure is analogous but with τ different equations). The Bellman equation (see [14]) to iteratively

solve is

$$z^{\text{iter}}(\mathbf{b}) \leftarrow \max_{\mu_{\mathbf{b}}} \Big\{ \beta_0 \, R_0(\mu_{\mathbf{b}}, \mathbf{b}) + \sum_{\bar{\mathbf{b}} \in \mathcal{B}} p_{\text{ext}}(\bar{\mathbf{b}}|\mathbf{b}) z^{\text{iter}-1}(\bar{\mathbf{b}}) \Big\}, \tag{3.2.18}$$

where "iter" is the index of VIA; $\beta_0 \, R_0(\mu_{\mathbf{b}}, \mathbf{b})$ represents the initial reward, whereas the other term is the expected future reward (this is derived from the old values of the Bellman equation, $z^{\text{iter}-1}(\cdot)$). After a number of iterations (typically only a few), VIA converges, and the Bellman equations yield the optimal solution $G(\Pi^\star)$.

3.2.4.2 Centralized Optimization

In a centralized system, RX knows the state of the batteries and communicates with the users in every time slot and thus can compute the policy using this information. In our model, the centralized scheme is obtained when every slot is a SYNC slot, i.e., $\beta_{\underline{k}} = 1$, $\forall \underline{k}$. The long-term reward (3.2.15) can be rewritten in the centralized scenario as

$$G^{\text{cent}}(\Pi, \mathbf{b}_0) = \liminf_{K \to \infty} \frac{1}{K} \sum_{\underline{k}=0}^{K-1} \mathbb{E}\Big[\rho(\eta_{\underline{k}}, \boldsymbol{\sigma}_{\underline{k}}^{\mu_{\underline{k}}})\Big], \tag{3.2.19}$$

where $\rho(\eta_{\underline{k}}, \boldsymbol{\sigma}_{\underline{k}}^{\mu_{\underline{k}}})$ degenerates to $r(\boldsymbol{\sigma}_{\underline{k}}^{\mu_{\underline{k}}}(\mathbf{b}_{\underline{k}}))$ because $\eta_{\underline{k}}(\bar{\mathbf{b}}) = \chi\{\bar{\mathbf{b}} = \mathbf{b}_{\underline{k}}\}$ according to Equation (3.2.12). In this case, policy $\mu_{\underline{k}} = (\boldsymbol{\sigma}_{\underline{k}}^{\mu_{\underline{k}}})$ defines the action to perform in any state deterministically.

The centralized scheme can be modeled with a Markov decision process and Function (3.2.19) can be maximized using VIA (more details about the centralized optimization can be found in Section 3.1). The centralized scheme always represents an upper bound to the performance of the decentralized one, as it exploits more information about the system to compute the policy.

3.2.5 Optimal Solution of the Internal Layer

In the previous section we discussed the external optimization problem and its solution via the value iteration algorithm. However, every iteration of VIA requires to solve the "max" in Equation (3.2.18). This is equivalent to solving the Dec-MDP of the internal layer, since (3.2.18) depends on the decentralized policy sequence $\mu_{\mathbf{b}}$. In this section, we discuss how to do that optimally, whereas in Section 3.2.6 we discuss suboptimal solutions.

3.2.5.1 Bellman Equation

The right-hand side of the Bellman equation (3.2.18) can be rewritten by replacing the terms $p_{\text{ext}}(\bar{\mathbf{b}}|\mathbf{b})$ and $R_0(\cdot)$ with their definitions given in Equations (3.2.13) and (3.2.14), respectively. After algebraic manipulations, we obtain:

$$\max_{\mu_{\mathbf{b}}} \left\{ \beta_0 \, \mathbb{E}\left[\sum_{k=0}^{\infty} \phi_k(\eta_k, \boldsymbol{\sigma}_k^{\mu_{\mathbf{b}}}) \middle| \mathbf{b} \right] \right\}, \tag{3.2.20}$$

$$\phi_k(\eta, \boldsymbol{\sigma}) \triangleq (1 - \beta_0)^k \Big(\rho(\eta, \boldsymbol{\sigma}) + \sum_{\bar{\mathbf{b}} \in \mathcal{B}} \eta_{+1}(\bar{\mathbf{b}}) z^{\text{iter}-1}(\bar{\mathbf{b}}) \Big). \tag{3.2.21}$$

Equation (3.2.20) represents the occupancy-MDP under investigation. For ease of notation, we used $\eta_{+1} \triangleq \iota(\eta, \boldsymbol{\sigma})$. Therefore, it is formally possible to find the optimal $\mu_{\mathbf{b}}$ by applying VIA and exploit the policy stationarity with respect to the occupancy states [126]. However,

this operation is impossible in practice because of the huge size of the occupancy state space. Therefore, we resort to the following alternative approach.

Since $(1-\beta_0)^k$ decreases with k and all the other terms are bounded, $\phi_k(\cdot)$ decreases with k. In particular, by definition of "max", we have $\phi_k(\eta, \boldsymbol{\sigma}) \leq \Delta(1-\beta_0)^k$, with $\Delta \triangleq \max_{\eta,\boldsymbol{\sigma}}(\rho(\eta,\boldsymbol{\sigma}) + \sum_{\bar{\mathbf{b}}\in\mathcal{B}} \eta_{+1}(\bar{\mathbf{b}}) z^{\text{iter}-1}(\bar{\mathbf{b}}))$. Therefore, once we select $\epsilon > 0$, there always exists $K = \lceil \log_{1-\beta_0}(\epsilon/\Delta) - 1 \rceil$ such that:

$$\beta_0 \sum_{k=K+1}^{\infty} \Delta(1-\beta_0)^k \leq \epsilon, \tag{3.2.22}$$

Since the left-hand side of (3.2.22) is an upper bound to the tail of (3.2.20), the following expression

$$\max_{\mu_{\mathbf{b}}} \left\{ \beta_0 \, \mathbb{E}\left[\sum_{k=0}^{K} \phi_k(\eta_k, \boldsymbol{\sigma}_k^{\mu_{\mathbf{b}}}) \middle| \mathbf{b} \right] \right\}, \tag{3.2.23}$$

differs from the optimal solution (3.2.20) by at most ϵ. Thus, in the following we focus on ϵ-optimal policies and examine only the first $K+1$ slots.

We can solve the "max" by rewriting it in recursive form as:

$$v_k(\eta_k) = \max_{\boldsymbol{\sigma}} \begin{cases} \phi_k(\eta_k, \boldsymbol{\sigma}) + v_{k+1}(\iota(\eta_k, \boldsymbol{\sigma})), & \text{if } k < K, \\ \phi_K(\eta_K, \boldsymbol{\sigma}), & \text{if } k = K, \end{cases} \tag{3.2.24}$$

where $v_k(\cdot)$ is the *cost-to-go function*. Equation (3.2.23) is equivalent to $\beta_0 \times v_0(\eta_0)$.

The trivial solution to find $\mu_{\mathbf{b}}$ is to apply VIA in the finite horizon; however, this would require, for every k, to specify $v_k(\eta_k)$ for *every* η_k, which is impossible in practice.

An alternate solution is to use techniques originally developed for POMDPs which were later used for Dec-POMDPs. In particular, the Learning Real Time A* (LRTA*) algorithm is suitable for our case, since it explores only the occupancy states which are actually visited during the planning horizon and avoids grid-based approaches (e.g., as used in [83]). In [36], the Markov Policy Search (MPS) algorithm was introduced as an adaptation of LRTA* to decentralized scenarios.

In summary, MPS operates as follows

1. It starts at $\underline{k} = 0$ and, for every $k \geq 0$, it computes the LHS of (3.2.24) with LRTA*, i.e., the maximization problem is solved only for the occupancy states which are actually visited and not for every η_k;

2. It replaces v_{k+1} in the right-hand side with an upper bound, which can be computed using the convexity of the cost-to-go function. In Section 3.2.5.2 we will further discuss this point;

3. When $k = K$ is reached, a lower bound of the optimal cost-to-go function is evaluated in a backward direction (see [36, Section 5.1]).

The procedure is repeated until upper and lower bounds converge to the optimal solution. We refer the readers to [36, 37] for a full description of the algorithm. In the following, we discuss how to find the upper bound of the cost-to-go function, which will be used as a building block in Section 3.2.5.3.

3.2.5.2 Upper Bound of the Cost-to-go Function

It can be shown by induction that the optimal cost-to-go function $v_k^\star$ is a convex function of the occupancy states and can be approximated by piecewise linear functions [37, Theorem 4.2]. The upper bound $\bar{v}_k$ of $v_k^\star$ can be written as

$$\bar{v}_k(\eta_k) = \max_{\boldsymbol{\sigma}} \{\phi_k(\eta_k, \boldsymbol{\sigma}) + C(\Upsilon_k, \iota(\eta_k, \boldsymbol{\sigma}))\}, \tag{3.2.25}$$

where C interpolates the occupancy state $\iota(\eta_k, \sigma)$ using the point set Υ_k, which contains the visited occupancy states along with their upper bound values. Every time (3.2.25) is solved, a new point $(\eta_k, \bar{v}_k(\eta_k))$ is added to Υ_k. The first points to be put in Υ_k are the corners of the occupancy simplex (i.e., the $|\mathcal{B}|$ points $[1, 0, \ldots, 0], \ldots, [0, \ldots, 0, 1]$) with their upper bound values obtained solving the following full knowledge MDP:

$$\overline{R}_0(\mu_{\mathbf{b}}, \mathbf{b}) \triangleq \mathbb{E}\left[\sum_{k=0}^{\infty} r(\boldsymbol{\sigma}_k^{\mu_{\mathbf{b}}}(\mathbf{b}_k)) \, (1-\beta_0)^k \middle| \mathbf{b} \right], \tag{3.2.26}$$

which is equivalent to (3.2.14) but with $r(\cdot)$ instead of $\rho_{k|0}$ and with $\tau = 1$. Expression (3.2.26) implicitly assumes that the state of the system $\mathbf{b}_k$ is globally known in slot k. Since this is a standard MDP, it can be easily solved with VIA.

Sawtooth Projection. Ideally, we could use a linear interpolation as the function C (i.e., map η_k on the convex hull of point set Υ_k). However, the complexity of mapping η_k into the convex hull would increase polynomially with the number of points in Υ_k (it requires to solve a linear program), making this approach infeasible in practice. Because of that, in the literature so far, different interpolation methods have been proposed as a replacement (e.g., see [35, 50]). An approach that has shown good performance in many applications is to replace C with the sawtooth projection:[4]

$$\textbf{sawtooth}(\Upsilon_k, \eta) = y^0(\eta) - \max_{\ell \in \mathcal{L}} \{(y^0(\eta^\ell) - v^\ell)\xi^\ell\}, \tag{3.2.27}$$

where η is the occupancy state to interpolate, (η^ℓ, v^ℓ) is the ℓ-th element of Υ_k, $\mathcal{L}$ is the set of indices of Υ_k, ξ^ℓ is the interpolation coefficient, and $y^0(\cdot)$ is the upper bound computed using the corner points of Υ_k, i.e.,

$$y^0(\eta) = \sum_{\mathbf{b} \in \mathcal{B}} \eta(\mathbf{b}) \Upsilon_k(\mathbf{b}). \tag{3.2.28}$$

In the previous expression, with a slight abuse of notation, $\Upsilon_k(\mathbf{b})$ indicates the upper bound value at the corner $\mathbf{b}$ of the simplex. The interpolation coefficient is defined as

$$\xi^\ell \triangleq \min_{\mathbf{b} \,:\, \eta^\ell(\mathbf{b}) > 0} \frac{\eta(\mathbf{b})}{\eta^\ell(\mathbf{b})}, \tag{3.2.29}$$

and can be derived geometrically (see Figure 3.9). Note that we use the "max" in (3.2.27) so as to obtain the lowest (i.e., best) upper bound. We now rewrite the sawtooth projection

[4] The term "sawtooth" comes from the shape of the interpolating function in the two-dimensional case (see Figure 3.9). The idea of the approach is to interpolate a point η using $|\mathcal{B}| - 1$ corner points of the simplex, and one point taken from Υ_k (ℓ in Equation (3.2.27)) [50].

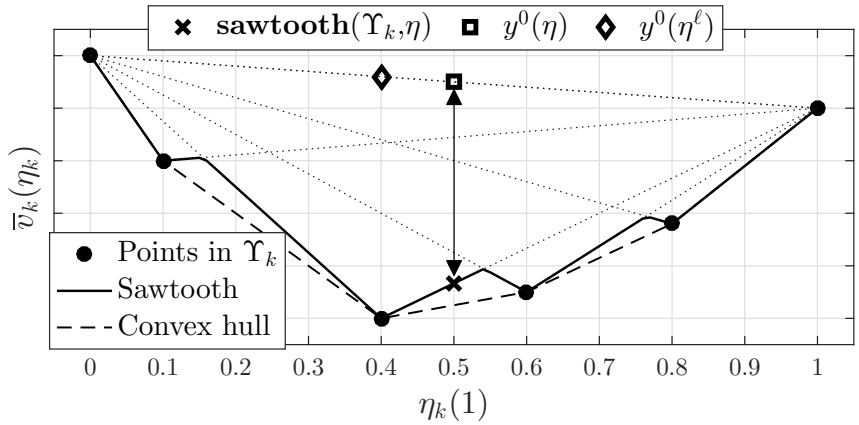

Figure 3.9. Sawtooth projection in the two-dimensional case. The arrow identifies the quantity $\max_{\ell\in\mathcal{L}}\{(y^0(\eta^\ell)-v^\ell)\xi^\ell\}$ when $\eta=\langle 0.5,0.5\rangle$.

in a simpler form:

$$\begin{aligned}\textbf{sawtooth}(\Upsilon_k,\eta) &\overset{(a)}{=} y^0(\eta)+\min_{\ell\in\mathcal{L}}\left\{(v^\ell-y^0(\eta^\ell))\min_{\mathbf{b}\,:\,\eta^\ell(\mathbf{b})>0}\frac{\eta(\mathbf{b})}{\eta^\ell(\mathbf{b})}\right\}\\ &\overset{(b)}{=} y^0(\eta)+\min_{\ell\in\mathcal{L}}\max_{\mathbf{b}\,:\,\eta^\ell(\mathbf{b})>0}\left\{\frac{\eta(\mathbf{b})}{\eta^\ell(\mathbf{b})}(v^\ell-y^0(\eta^\ell))\right\}\\ &\overset{(c)}{=} \min_{\ell\in\mathcal{L}}\left\{y^0(\eta)+\max_{\mathbf{b}\,:\,\eta^\ell(\mathbf{b})>0}\left\{\frac{\eta(\mathbf{b})}{\eta^\ell(\mathbf{b})}(v^\ell-y^0(\eta^\ell))\right\}\right\}\\ &\overset{(d)}{=} \min_{\ell\in\mathcal{L}}\textbf{sawtooth}^\ell(\Upsilon_k,\eta).\end{aligned}\tag{3.2.30}$$

Step (a) coincides with Definition (3.2.27); step (b) holds because v^ℓ is not greater than $y^0(\eta^\ell)$, since $y^0(\eta^\ell)$ represents the interpolation using only the corner points (see Figure 3.9); in step (c), we move $y^0(\eta)$ inside the "$\min_\ell$", since it does not depend on ℓ; finally, we define $\textbf{sawtooth}^\ell(\Upsilon_k,\eta)$ in step (d).
We also introduced $\textbf{sawtooth}^\ell(\Upsilon_k,\eta)$, which will be used in the next subsection, as the sawtooth projection obtained using the ℓ-th element of Υ_k.

The sawtooth projection produces higher (i.e., worse) upper bounds than the convex hull projection and thus MPS may require more iterations to converge (however, a single iteration can be performed much more quickly), but convergence to the optimum is still guaranteed [35, Section 5.2].

3.2.5.3 Constraint Programming Formulation

The key step to perform to find the policy is solving the "max" in (3.2.25). Although this would be possible by performing the *exhaustive backup*, i.e., by inspecting all possible choices of $\boldsymbol{\sigma}$, it is more practical to introduce faster solutions. Constraint Programming [31] is a technique to express hard and soft constraints as an optimization problem. We now use it to reformulate (3.2.25) with $C=\textbf{sawtooth}$.

Using the notation of Dibangoye *et al.* [35], we define

$$W_k(\eta_k,\boldsymbol{\sigma},\ell)\triangleq\phi_k(\eta_k,\boldsymbol{\sigma})+\textbf{sawtooth}^\ell(\Upsilon_k,\iota(\eta_k,\boldsymbol{\sigma})).\tag{3.2.31}$$

Combining (3.2.30) and (3.2.31), we can rewrite Equation (3.2.25) as

$$\begin{aligned}\bar{v}_k(\eta_k) &= \max_{\boldsymbol{\sigma}}\{\phi_k(\eta_k,\boldsymbol{\sigma}) + \mathbf{sawtooth}(\Upsilon_k, , \iota(\eta_k,\boldsymbol{\sigma}))\}\\ &= \max_{\boldsymbol{\sigma}}\{\phi_k(\eta_k,\boldsymbol{\sigma}) + \min_{\ell\in\mathscr{L}} \mathbf{sawtooth}^{\ell}(\Upsilon_k, \iota(\eta_k,\boldsymbol{\sigma}))\}\\ &= \max_{\boldsymbol{\sigma}} \min_{\ell\in\mathscr{L}} W_k(\eta_k,\boldsymbol{\sigma},\ell). \end{aligned} \tag{3.2.32}$$

To solve the previous equation, we split it in $|\mathscr{L}|$ separate mixed-integer programs:

$$\bar{v}_k(\eta_k) = \max_{\ell\in\mathscr{L}} \bar{v}_k^{\ell}(\eta_k), \tag{3.2.33}$$

with

$$\bar{v}_k^{\ell}(\eta_k) \triangleq \max_{\boldsymbol{\sigma}} W_k(\eta_k,\boldsymbol{\sigma},\ell), \tag{3.2.34a}$$

$$\text{subject to:}\quad W(\eta_k,\boldsymbol{\sigma},\ell) \le W_k(\eta_k,\boldsymbol{\sigma},l), \quad \forall l\in\mathscr{L}. \tag{3.2.34b}$$

Weighted Constraint Satisfaction Problem (WCSP). Focus now on the optimization $\max_{\boldsymbol{\sigma}} W(\eta_k,\boldsymbol{\sigma},\ell)$ *without* any constraints. This can be formulated as a WCSP as follows. First, we rewrite $W(\eta_k,\boldsymbol{\sigma},\ell)$ using (3.2.12), (3.2.21), (3.2.28), (3.2.30), and (3.2.31):

$$\begin{aligned} W_k(\eta_k,\boldsymbol{\sigma},\ell) = {}& (1-\beta_0)^k \sum_{\mathbf{b}\in\mathscr{B}} \eta_k(\mathbf{b}) r(\boldsymbol{\sigma}(\mathbf{b}))\\ &+ (1-\beta_0)^k \sum_{\mathbf{b}'\in\mathscr{B}} \eta_{k+1}(\mathbf{b}') z^{\text{iter}-1}(\mathbf{b}')\\ &+ \sum_{\mathbf{b}'\in\mathscr{B}} \eta_{k+1}(\mathbf{b}')\Upsilon_k(\mathbf{b}') + \max_{\mathbf{b}'':\eta^{\ell}(\mathbf{b}'')>0} \left\{\frac{\eta_{k+1}(\mathbf{b}'')}{\eta^{\ell}(\mathbf{b}'')}(v^{\ell} - y^0(\eta^{\ell}))\right\}, \end{aligned} \tag{3.2.35}$$

where $\eta_{k+1} = \iota(\eta_k,\boldsymbol{\sigma})$, and we used $\mathbf{b}$, $\mathbf{b}'$ and $\mathbf{b}''$ to differentiate the indices. Note that, since ℓ is given, the term $y^0(\eta^{\ell})$ is fixed. Using the occupancy update formula of Equation (3.2.11), we get

$$\begin{aligned} W(\eta_k,\boldsymbol{\sigma},\ell) = {}& (1-\beta_0)^k \sum_{\mathbf{b}\in\mathscr{B}} \eta_k(\mathbf{b}) r(\boldsymbol{\sigma}(\mathbf{b}))\\ &+ (1-\beta_0)^k \sum_{\mathbf{b}'\in\mathscr{B}}\sum_{\mathbf{b}\in\mathscr{B}} \eta_k(\mathbf{b}) p(\mathbf{b}'|\mathbf{b},\boldsymbol{\sigma}(\mathbf{b})) z^{\text{iter}-1}(\mathbf{b}')\\ &+ \sum_{\mathbf{b}'\in\mathscr{B}}\sum_{\mathbf{b}\in\mathscr{B}} \eta_k(\mathbf{b}) p(\mathbf{b}'|\mathbf{b},\boldsymbol{\sigma}(\mathbf{b}))\Upsilon_k(\mathbf{b}')\\ &+ \max_{\mathbf{b}'':\eta^{\ell}(\mathbf{b}'')>0} \left\{\sum_{\mathbf{b}\in\mathscr{B}} \frac{\eta_k(\mathbf{b}) p(\mathbf{b}''|\mathbf{b},\boldsymbol{\sigma}(\mathbf{b}))}{\eta^{\ell}(\mathbf{b}'')}(v^{\ell} - y^0(\eta^{\ell}))\right\}. \end{aligned} \tag{3.2.36}$$

Since the first terms of (3.2.36) do not depend on $\mathbf{b}''$, we move them inside the "$\max_{\mathbf{b}''}$" and take the common sum over $\mathbf{b}$; then, we note that all terms multiply $\eta_k(\mathbf{b})$. Thus, by introducing a variable $u_{\ell}(\cdot)$, we obtain

$$W(\eta_k,\boldsymbol{\sigma},\ell) = \max_{\mathbf{b}'':\eta^{\ell}(\mathbf{b}'')>0} \sum_{\mathbf{b}\in\mathscr{B}} \eta_k(\mathbf{b}) u_{\ell}(\mathbf{b},\boldsymbol{\sigma}(\mathbf{b}),\mathbf{b}''). \tag{3.2.37}$$

For every fixed ℓ, the WCSP is formally defined as follows. The variables are defined by $\boldsymbol{\sigma}$ as in (3.2.8) (i.e., the actions $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ for every $\mathbf{b} \in \boldsymbol{\mathcal{B}}$) plus the index $\mathbf{b}''$. The domains are the same as in the original problem, i.e., $\boldsymbol{\Omega}$ for $\boldsymbol{\sigma}$ and $\boldsymbol{\mathcal{B}}$ for $\mathbf{b}''$. A WCSP is fully specified by its constraints, which are of the form

$$\mathbf{constraint}_\ell(\mathbf{b}) = M - \eta_k(\mathbf{b}) u_\ell(\mathbf{b}, \boldsymbol{\sigma}(\mathbf{b}), \mathbf{b}''). \tag{3.2.38}$$

The total number of constraints is at most $|\boldsymbol{\mathcal{B}}|$, one for every possible battery level. M is a large number used to cast the WCSP in its standard form. Standard WCSP solvers compute the following quantity:

$$\min_{\sigma_k^{(i)}(b^{(i)}), \forall b^{(i)}, \forall i, \text{ and } \mathbf{b}'' \in \boldsymbol{\mathcal{B}}} \sum_{\mathbf{b} \in \boldsymbol{\mathcal{B}}} \mathbf{constraint}_\ell(\mathbf{b}), \tag{3.2.39}$$

whose solution is equal to the solution of (3.2.34a). In practice, for every $\mathbf{b} \in \boldsymbol{\mathcal{B}}$, the quantity $\mathbf{constraint}_\ell(\mathbf{b})$ is evaluated for all the combinations of $\boldsymbol{\sigma}(\mathbf{b}) \in \boldsymbol{\Omega}$ and $\mathbf{b}'' \in \boldsymbol{\mathcal{B}}$ independently of the others constraints. Then, they are summed together and the minimum among all the solutions is chosen. However, a solution referred to a constraint $\mathbf{constraint}_\ell(\mathbf{b})$ may be related to other solutions. For example, different states $\mathbf{b}$ have some common entries (e.g., both $\langle 0, 0 \rangle \in \boldsymbol{\mathcal{B}}$ and $\langle 0, 1 \rangle \in \boldsymbol{\mathcal{B}}$ have the first entry equal to 0); in this case, the corresponding actions must have some common element even if they are referring to different constraints (e.g., $\boldsymbol{\sigma}(\langle 0, 0 \rangle) = \langle \sigma^0(0), \sigma^{(1)}(0) \rangle$ and $\boldsymbol{\sigma}(\langle 0, 1 \rangle) = \langle \sigma^0(0), \sigma^{(1)}(1) \rangle$ must have the first entry in common).

The main advantage of using a WCSP solver is that the decentralized policy does not need to be computed as a whole, but can be divided in constraints which are later combined together.

So far, we have only focused on (3.2.34a) (i.e., on a single ℓ). To solve (3.2.33), we need to compute $|\mathcal{L}|$ different WCSPs (one for every $\ell \in \mathcal{L}$) and take the maximum among all solutions. However this approach presents one major drawback: for a fixed ℓ, solving a single WCSP may not be sufficient. Indeed, the solution must also satisfy (3.2.34b), which has been completely neglected in the definition of the WCSP. If a constraint were violated, then the solution of the WCSP would be infeasible and should be discarded. In this case, Dibangoye *et al.* (see [37, Section 3.4.2]) propose to formulate a new WCSP in which the previous infeasible solution results in a very high cost constraint (and thus it is never chosen in the solution process). If the new solution also turns out to be infeasible, the procedure is repeated. The iterations stop if a feasible solution is found or if all decentralized actions have been examined.

Although the previous method formally leads to the optimal solution, it may often degenerate in an exhaustive search (i.e, examining all the decentralized policies). The corresponding complexity would be $\mathcal{O}((S_a)^{b_{\max} \times N})$ if all users had the same battery size $b_{\max}$ (see the structure of $\boldsymbol{\sigma}(\mathbf{b})$ in Equation (3.2.8)), i.e., exponential in N. This operation is computationally infeasible when lots of possibilities are involved. Thus, optimally solving a Dec-MDP with guarantees on the worst case performance is still an open issue. In Section 3.2.6, we propose two suboptimal approaches for handling the problem based on the previous results.

3.2.5.4 Stationary Policies

We define a stationary policy as follows.

Definition 3.2.1. (Stationary Policy) *A policy* Π *is* stationary *with respect to the battery levels if:*

1. *(External Layer) The decentralized policies of two different SYNC slots with the same state are equal, i.e.,* $\mu_{\underline{k}'} = \mu_{\underline{k}''}$ *for any* $\underline{k}'$, $\underline{k}''$, *when* $\mathbf{b}_{\underline{k}'} = \mathbf{b}_{\underline{k}''}$.
2. *(Internal Layer) Given the SYNC slot* $\underline{k}$, *the actions to use in two different slots with the same state are equal, i.e.,* $\boldsymbol{\sigma}_{k'}^{\mu_{\underline{k}}}(\mathbf{b}) = \boldsymbol{\sigma}_{k''}^{\mu_{\underline{k}}}(\mathbf{b})$, *for any* $\mathbf{b}$, k', k'', $\underline{k}$, *with* $k', k'' \geq \underline{k}$.

In practice, a stationary policy always uses the same action over time when the battery levels are the same. Note that, thanks to [126], point 1) of the previous definition is always satisfied in our model, since the external layer is formulated and solved as an MDP. This is also the reason why using (3.2.16) is optimal. However, in general, point 2) is not satisfied (this will be very clear from our numerical evaluation, e.g., see Figure 3.10) because the internal layer is not an MDP but a Dec-MDP. In particular, Equations (3.2.20) and (3.2.21) depend on the occupancy states and not on the battery levels only; thus, there is no guarantee that a stationary scheduler is optimal in EH scenarios, in general. This is not the case for traditional systems in which EH is not considered, as shown in the next lemma.

Lemma 3.2.2. *In a traditional system with an infinite energy supply and no EH, a stationary orthogonal policy is optimal.*

Proof. A system with an infinite energy supply and no EH can be modeled as an EH system in which the harvested energy is deterministically equal to one in every slot. In this case, the term $\eta_{+1}(\bar{\mathbf{b}})$ in Equation (3.2.21) is equal to $\chi\{\bar{\mathbf{b}} = \mathbf{b}_{\max}\}$. Moreover, according to (3.2.12), the instantaneous reward $\rho(\eta, \boldsymbol{\sigma})$ degenerates to $r(\boldsymbol{\sigma}(\mathbf{b}_{\max}))$. Thus, we rewrite (3.2.20) as

$$\begin{aligned}&\arg\max_{\mu_{\mathbf{b}}} \left\{\beta_0\, \mathbb{E}\Big[\sum_{k=0}^{\infty}(1-\beta_0)^k\Big(r(\boldsymbol{\sigma}_k^{\mu_{\mathbf{b}}}(\mathbf{b}_{\max})) + z^{\text{iter}-1}(\mathbf{b}_{\max})\Big)\Big|\mathbf{b}\Big]\right\}\\ &= \arg\max_{\mu_{\mathbf{b}}} \left\{\sum_{k=0}^{\infty}(1-\beta_0)^k r(\boldsymbol{\sigma}_k^{\mu_{\mathbf{b}}}(\mathbf{b}_{\max}))\right\},\end{aligned}$$

where in the second equality we removed all the constant terms and the expectation. Since the actions performed in slot k do not influence any other slot (the batteries are always charged at $\mathbf{b}_{\max}$), the optimization can be performed for each k independently: $\arg\max_{\boldsymbol{\sigma}_k^{\mu_{\mathbf{b}}}}\{r(\boldsymbol{\sigma}_k^{\mu_{\mathbf{b}}}(\mathbf{b}_{\max}))\}$, which leads to the same solution for every k and $\mathbf{b}$. Therefore, a stationary policy is optimal.

Moreover, since a system without EH is deterministic, it can be modeled as a centralized system [33], as all the quantities are known to RX in every slot. [33] discusses why an optimal policy is orthogonal in this scenario; intuitively, the best option is to schedule only (at most) one user in every slot so as to avoid collisions, which would degrade the system performance. ■

In summary, using a fully orthogonal policy all the time in a traditional environment is optimal because the system evolves deterministically and thus collisions can be avoided. Instead, with EH, the unpredictable future energy arrivals introduce a stochastic component in the system so that full knowledge is not possible. As a consequence, the use of stochastic policies is optimal, even though collisions may be incurred in this case. This is a major difference between decentralized schemes with and without EH.

3.2.6 Sub-Optimal Solutions of the Internal Layer

Since the main issue of the exhaustive search is that the space of variables in Problem (3.2.33)-(3.2.34) is exceedingly large, we aim at reducing this space, so that $\boldsymbol{\sigma}$ cannot take all possible values but is constrained to lie in a smaller subset. The problem now is to define such a subset. In the next subsection we present an approach based on WCSPs, whereas in Section 3.2.6.2 we introduce a different scheme based on parametric policies.

3.2.6.1 WCSP-Based Policies

In this case, we exploit the results about WCSPs presented in Section 3.2.5.3 to find a suboptimal policy. The proposed algorithm is as follows.

Algorithm 1 (Suboptimal policy using WCSPs)

1: **for** $\ell \in \mathcal{L}$ **do**
2: $\quad \boldsymbol{\sigma}_\ell \leftarrow$ Solve WCSP for ℓ given η_k
3: $\quad$ **for** $l \in \mathcal{L}$ **do**
4: $\quad\quad$ Evaluate $W(\eta_k, \boldsymbol{\sigma}_\ell, l)$
5: $\quad l^\star \leftarrow \arg\min_{l \in \mathcal{L}} W(\eta_k, \boldsymbol{\sigma}_\ell, l)$
6: $\quad \bar{v}_k^\ell(\eta_k) \leftarrow W(\eta_k, \boldsymbol{\sigma}_\ell, l^\star)$
7: $\ell^\star \leftarrow \max_\ell \bar{v}_k^\ell(\eta_k)$
8: **return** $\boldsymbol{\sigma}_{\ell^\star}$

As required by the "max" in (3.2.33), we look at every $\ell \in \mathcal{L}$ (Line 1) and, at the end of the algorithm, we return the solution with the maximum value (Lines 7-8). Lines 2-6 solve Problem (3.2.34) suboptimally as described in the following.

First, we solve the WCSP for a fixed ℓ (i.e., we solve (3.2.34a)), and find the corresponding solution $\boldsymbol{\sigma}_\ell$. Then, using $\boldsymbol{\sigma}_\ell$, we evaluate $W(\eta_k, \boldsymbol{\sigma}_\ell, l)$ for every index l. Two cases should now be examined, which can be handled in a fully equivalent way, but have different meanings:

- If $W(\eta_k, \boldsymbol{\sigma}_\ell, \ell) \leq W(\eta_k, \boldsymbol{\sigma}_\ell, l)$, $\forall l \in \mathcal{L}$, then $\boldsymbol{\sigma}_\ell$ would be an optimal solution of (3.2.34), since it maximizes (3.2.34a) (Line 2) and satisfies (3.2.34b). In this case, $\ell \equiv l^\star$ (Line 5), and $\bar{v}_k^\ell(\eta_k)$ is the solution of (3.2.34);
- Instead, when there exists $l \neq \ell$ such that $W(\eta_k, \boldsymbol{\sigma}_\ell, l) < W(\eta_k, \boldsymbol{\sigma}_\ell, \ell)$, then $\boldsymbol{\sigma}_\ell$ *is not optimal* for index ℓ. In this case, the optimal approach would require the execution of a new WCSP discarding the previous solution (in the new WCSP, $\boldsymbol{\sigma}_\ell$ would become a very high cost solution). Instead, in Algorithm 1, we implicitly make the following observation: $\boldsymbol{\sigma}_\ell$ is a feasible solution (i.e., it satisfies (3.2.34b)) of $\bar{v}_k^{l^\star}(\eta_k)$, where $l^\star$ is such that $W(\eta_k, \boldsymbol{\sigma}_\ell, l^\star) \leq W(\eta_k, \boldsymbol{\sigma}_\ell, l)$, $\forall l \in \mathcal{L}$ (Line (5)). Therefore, solution $W(\eta_k, \boldsymbol{\sigma}_\ell, l^\star)$ is *feasible*; for simplicity, we improperly save its value in $\bar{v}_k^\ell(\eta_k)$ (Line 6). By doing so, at the end of Line 6, $\bar{v}_k^\ell(\eta_k)$ does not represent the solution of (3.2.34) associated to index ℓ, but it contains a feasible solution for some index $l^\star$.

In practice, while executing Algorithm 1, the space of variables of Problem (3.2.33) is defined by the solutions of WCSPs at every iteration (thus, it is not determined a priori).

The proposed approach is faster than the optimal one, since it completely avoids the exhaustive search; however, in general, it is suboptimal and thus achieves worse performance.

3.2.6.2 Parametric Policies

Another possibility to avoid the exhaustive search step is to use parametric policies and thus reduce the number of optimization variables to few parameters. In particular, we force the actions of user i to follow a predetermined structure:

$$\sigma^{(i)}(b^{(i)}) = f^{(i)}_{\text{par}}(\Theta^{(i)}, b^{(i)}) \tag{3.2.40}$$

where $b^{(i)}$ is the independent variable and $\Theta^{(i)}$ is a set of parameters which specify the structure of $f^{(i)}_{\text{par}}$. For example, if we used $\Theta^{(i)} = \{\theta^{(i)}\}$, and a simple linear function $f^{(i)}_{\text{par}}(\Theta^{(i)}, b^{(i)}) = \theta^{(i)} b^{(i)}$, the only optimization variable of user i would be $\theta^{(i)}$, and not $\sigma^{(i)}(0), \ldots, \sigma^{(i)}(b^{(i)}_{\max})$ as in the original problem. In this case, for a symmetric scenario, the complexity of the exhaustive search step goes from $\mathcal{O}((S_a)^{b_{\max} \times N})$ to $\mathcal{O}((S_\theta)^N)$, therefore it remains exponential in N but with a much smaller coefficient in the exponent. S_θ is the number of values that $\theta^{(i)}$ can assume.[5]

In our scenario we force $f^{(i)}_{\text{par}}(\Theta^{(i)}, b^{(i)})$ to be a non-decreasing function of $b^{(i)}$ as in [93], which implies that higher energy levels cannot correspond to lower transmission probabilities.

3.2.7 Numerical Results

The numerical evaluation is performed using two nodes, since the complexity grows super-exponentially with the number of users. Indeed the size of the occupancy state (i.e., the number of all possible probabilities $\eta_k(\bar{\mathbf{b}})$, as defined in Equation (3.2.10)) evolves exponentially with N, and the exhaustive search operation (exponential in N), or a suboptimal approach, is to be performed for every element of the occupancy state. If not otherwise stated, we adopt the following parameters: the batteries can contain up to $b^{(1)}_{\max} = b^{(2)}_{\max} = 8$ energy quanta; the energy arrival processes are i.i.d. over time and the probability of receiving one energy quantum is $\mathbb{P}^{(1)}_E = \mathbb{P}^{(2)}_E$ in every slot; when a transmission is performed a reward $V^{(i)} = \log(1+\Lambda^{(i)} H^{(i)})$ is accrued, where $V^{(i)}$ represents the normalized transmission rate in a slot, and $H^{(i)}$ is an exponentially distributed random variable with mean 1 (see [92]); the transmission probabilities in $[0, 1]$ are uniformly quantized with $S_a = 19$ samples; the average normalized SNRs are $\Lambda^{(1)} = 6$ and $\Lambda^{(2)} = 3$; both devices have the same weight; to perform a transmission $m^{(1)} = m^{(2)} = 2$ energy quanta are drawn from the battery; finally, β has period $\tau = 1$ (i.e., it is constant over time). All the numerical evaluations were written in C++ and, for the solution of the weighted constraint satisfaction problems, we used ToulBar2 [143], a highly efficient solver of WCSPs. We first focus on the solution of the internal layer (i.e., we only look at $R_0(\mu_{\mathbf{b}_0}, \mathbf{b}_0)$), and discuss later how the external layer performs.

Transmission Probabilities. In Figures 3.10 (low energy arrival rates) and 3.11 (high energy arrival rates) we show the transmission probabilities of the parametric decentralized policy of Section 3.2.6.2, where f_{par} is a linear function, $\Theta^{(i)} = \{\theta^{(i)}\}$ and $\theta^{(i)}$ is such that $\theta^{(i)} b^{(i)}_{\max} \in \Omega^{(i)}$. The dashed lines have been slightly manually shifted to the right only for graphical purposes to avoid that superposition of the two curves makes it hard to

[5] We note that, although we did not reduce the theoretical complexity of the exhaustive search (which is still exponential), using smaller coefficients allows much faster numerical evaluations. The more general problem of developing heuristic schemes with lower complexity is still open.

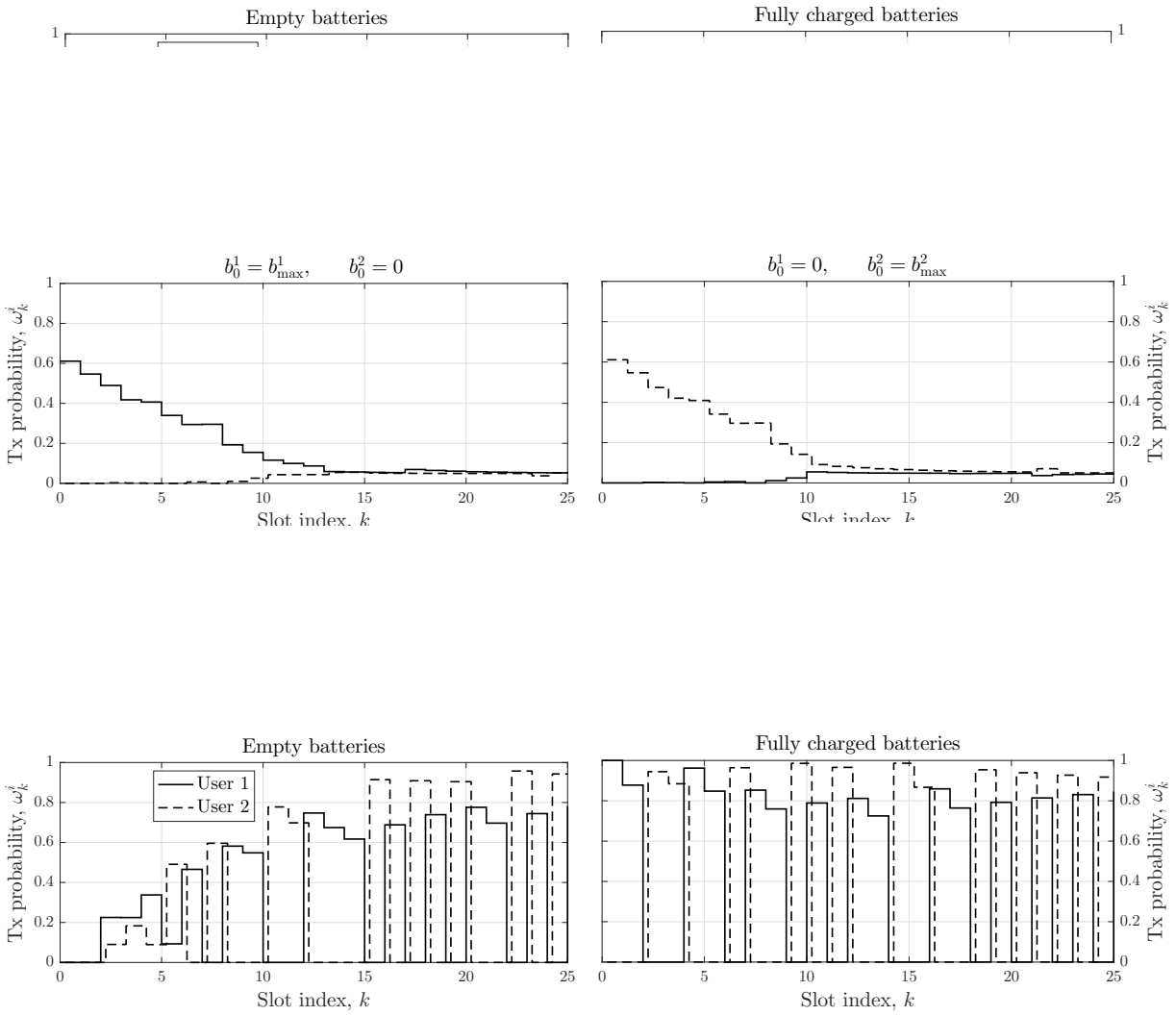

Figure 3.11. Transmission probabilities as a function of time for two users when $\beta_0 = 0.05$ and $\mathbb{P}_E^{(1)} = \mathbb{P}_E^{(2)} = 0.9$ for different initial battery levels.

distinguish between them. In these figures, we only focus on the internal layer, thus we study the system behavior between two consecutive SYNC slots.

The main difference between Figures 3.10 and 3.11 is that, after many slots, the transmission probabilities are both greater than zero in the first case, whereas an almost pure time-orthogonal approach is used in the latter, regardless of the initial energy levels. Thus, when the energy resources are scarce (i.e., low energy arrivals) then, if k is large, an orthogonal scheme in which collisions are avoided is suboptimal. The trade-off between orthogonal and random access schemes can be intuitively explained as follows. As time goes on, nodes lose information about the global state of the system, thus a device does not know the energy level of the other. In this case, an orthogonal scheme might be highly inefficient: if a slot were assigned only to user 1, but this did not have enough energy for transmission, then the slot would be unused. Since the energy resources are scarce, it is likely that such a case happens. Instead, when both have a lot of energy, it is easier to estimate the energy level of the nodes, and thus pre-assigning the slots is possible (and in

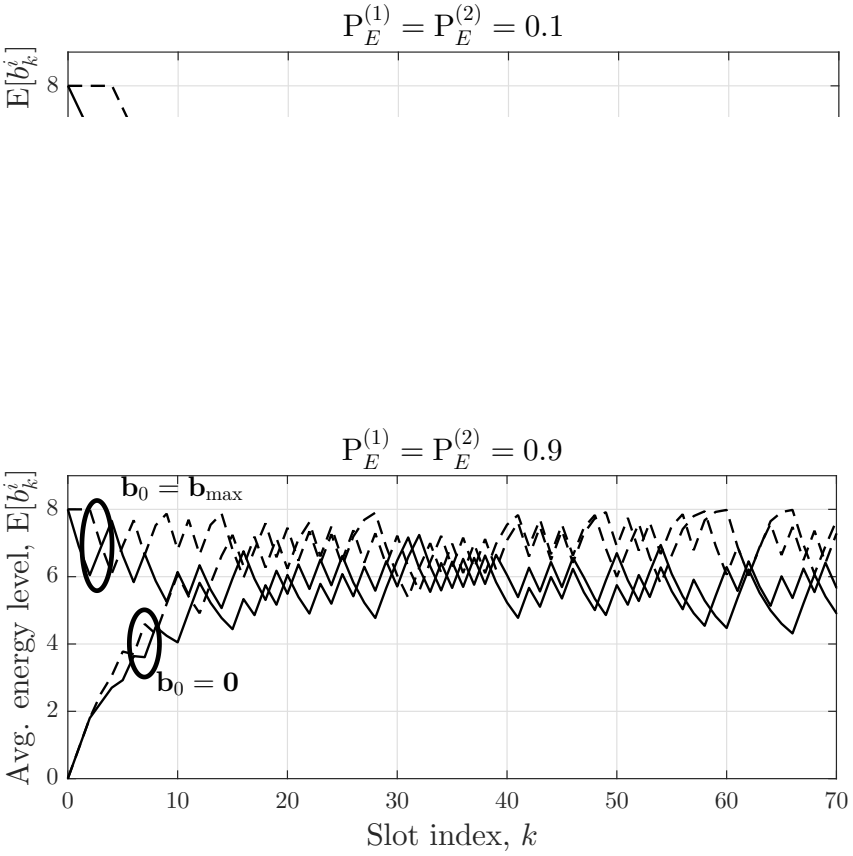

Figure 3.12. Battery level evolution as a function of time for two users when $\beta_0 = 0.05$, and $\mathbb{P}_E^{(1)} = \mathbb{P}_E^{(2)} = 0.1$ or 0.9 for different initial battery levels.

fact optimal). This means that, in the low energy case, it is better to transmit even if the collision probability is non-negligible, to compensate the (likely) lack of energy of the other user. This represents the main novelty of EH scenarios over traditional ones in which, as explained in Section 3.2.3.3, an orthogonal decentralized access scheme is optimal.

The same effect can be observed in Figure 3.10 when the initial energy levels are high. Indeed, when k is small, nodes still have information about the state of the other device; therefore, an orthogonal approach is optimal in this case also, and collisions are avoided since it is very likely that a node has enough energy to perform the transmission.

Note that when an orthogonal access scheme is employed, user 1 is advantaged with respect to user 2 (this can be clearly seen in Figure 3.11, where more slots are allocated to the first node). Indeed, the SNR of the first node ($\Lambda^{(1)}$) is greater than the SNR of the other ($\Lambda^{(2)}$), and therefore higher returns are obtained by user 1 when a transmission is performed. In this case, fairness is not achieved because of the near-far effect (a node with a better channel is advantaged over the other); however, the network could be rebalanced by changing the weights $w^{(1)}$, $w^{(2)}$.

Finally, note that in Figure 3.10 the average transmission probabilities in the long run almost coincide with the energy arrival rate divided by $m^{(i)}$, so as to achieve energy neutrality.

Energy Levels. Figure 3.12 shows another interesting, though predictable, result: despite the initial energy level, in the long run all the energy levels of the same device converge, approximately, to the same value. This is because all the initial fluctuations

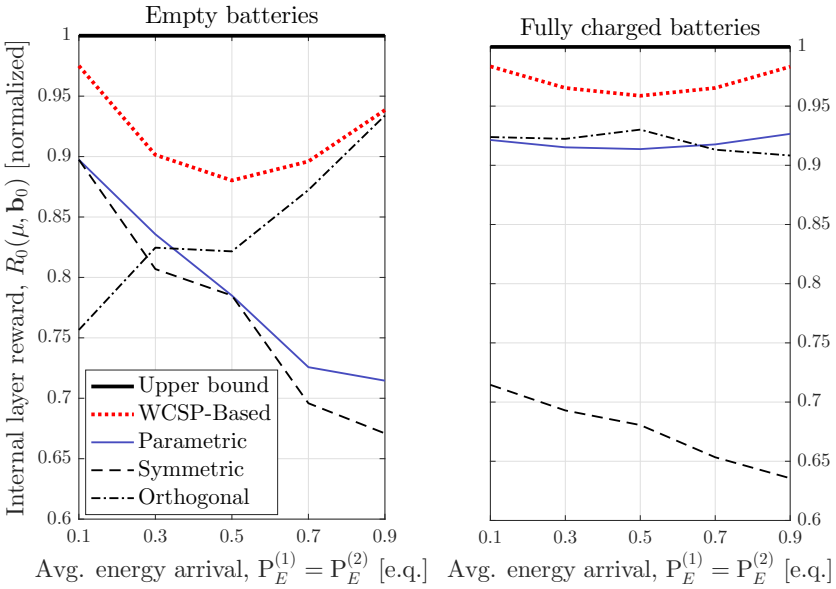

Figure 3.13. Long-term rewards of the internal layer as a function of the energy arrival rates $\mathbb{P}_E^{(1)} = \mathbb{P}_E^{(2)}$ for two users when $\beta_0 = 0.2$ for batteries initially empty (left) or fully charged (right).

have been absorbed by the batteries. Note that the energy levels of user 2 are higher because $\Lambda^{(2)} < \Lambda^{(1)}$, thus user 2 transmits less frequently than the other and consumes less energy, on average.

Internal Reward. In Figure 3.13 we show the long-term discounted reward as a function of the energy arrival rate for the decentralized scheme solved using WCSPs (Section 3.2.6.1), the decentralized parametric scheme (Section 3.2.6.2), a fully orthogonal approach, and a fully symmetric scheme. The curves are normalized with respect to the upper bound, given by the full knowledge scheme of Equation (3.2.26). To understand the trend of the curves, it is important to remark that the first slots after the initial SYNC slot are the most important ones for two reasons:

1. The decentralized reward $\phi_k(\cdot)$ defined in Equation (3.2.21) decreases with k. Therefore, the initial slots have higher weights and contribute more to the global reward;

2. There is more information about the state of the other device in the initial slots.

Therefore, when the initial batteries are fully charged, the decentralized schemes are close to the upper bound. Indeed, since $\mathbf{b}_0 = \mathbf{b}_{\max}$, the reward of the decentralized scheme and the upper bound are very close. Instead, if the batteries are initially fully discharged, the gap between the two is wider. In this case, the first slots do not play a fundamental role, since there is not a lot of energy to exploit. Therefore, most of the gain is obtained for higher k, which in turn leads to a less informative situation about the state of the other device. In this case, the full knowledge scheme (which represents our upper bound) may perform much better than the decentralized one because of this lack of information.

Although it is not visible from Figure 3.13 because of the normalization, the curves increase with $\mathbb{P}_E^{(1)} = \mathbb{P}_E^{(2)}$, because more energy can be harvested (see Section V of our

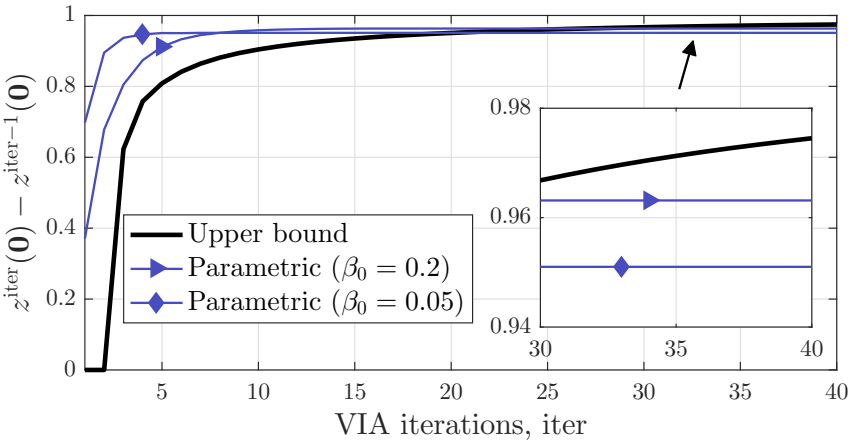

Figure 3.14. Transmission probabilities as a function of time for two users with batteries initially empty when $\mathbb{P}_E^{(1)} = \mathbb{P}_E^{(2)} = 0.05$.

paper [**C13**] at page 155). The lowest normalized reward for the WCSP-based policy is obtained around $\mathbb{P}_E^{(1)} = \mathbb{P}_E^{(2)} = 0.5$; indeed, this corresponds to the least informative case, since the battery fluctuations are not predictable at all.

Finally, note that the decentralized policy obtained using WCSP outperforms the parametric policy in almost every case. However, this is also strongly influenced by the number of parameters $\Theta^{(i)}$ we used, and using more parameters would lead to better performance. Moreover, it can be clearly seen that the fully orthogonal and the symmetric policies are strongly suboptimal in this scenario, therefore using an optimized approach significantly improves the throughput of the network. An interesting outcome of Figure 3.13 is that the gap between upper bound and decentralized scheme may be small ($< 15\%$), when the decentralized policy is correctly designed. Therefore, using low values of β (e.g., $\beta = 0.05$) we are able to achieve the twofold goal of greatly reducing the overhead in the network, since the policy is distributed only during the SYNC slots, and of achieving good performance.

External Reward. We now describe the performance of the complete system, i.e., of the external layer. First, we show the iterations of VIA (Equation (3.2.18)) in Figure 3.14. Note that the y-axis represents the difference between two consecutive steps of VIA, which converges to $G(\Pi^\star)$ (see the Relative Value Iteration Algorithm in [14, Vol. 1, Sec. 7.4]). The case iter $= 1$ corresponds to the internal layer only (i.e., $\beta_0 \times R_0(\mu_{\mathbf{0}}, \mathbf{0})$), since $z^0(\mathbf{b}) = 0,\ \forall \mathbf{b} \in \mathcal{B}$. We highlight that, to solve a single step of VIA, many decentralized optimization steps are performed (one for every $\mathbf{b} \in \mathcal{B}$). From the figure, it can be clearly seen that only few iterations are required for convergence, especially for lower β_0 (which are the most computationally expensive, since the intervals between SYNC slots are longer). Note that, when VIA converges, the higher β_0, the higher the reward, as expected; moreover, the upper bound curve (i.e., the centralized case of Section 3.2.4.2) outperforms the others (in this case, there is less uncertainty and more global information). In Figure 3.15, we plot the reward $G(\Pi)$ of the external layer as a function of the energy arrival rates. The values reported here are derived from the last step of VIA. As expected, $G(\Pi)$ increases with $\mathbb{P}_E^{(1)} = \mathbb{P}_E^{(2)}$ (more energy available). It is interesting to observe that the parametric policy is very close to the upper bound, whereas the rewards of the other, simpler policies are much lower.

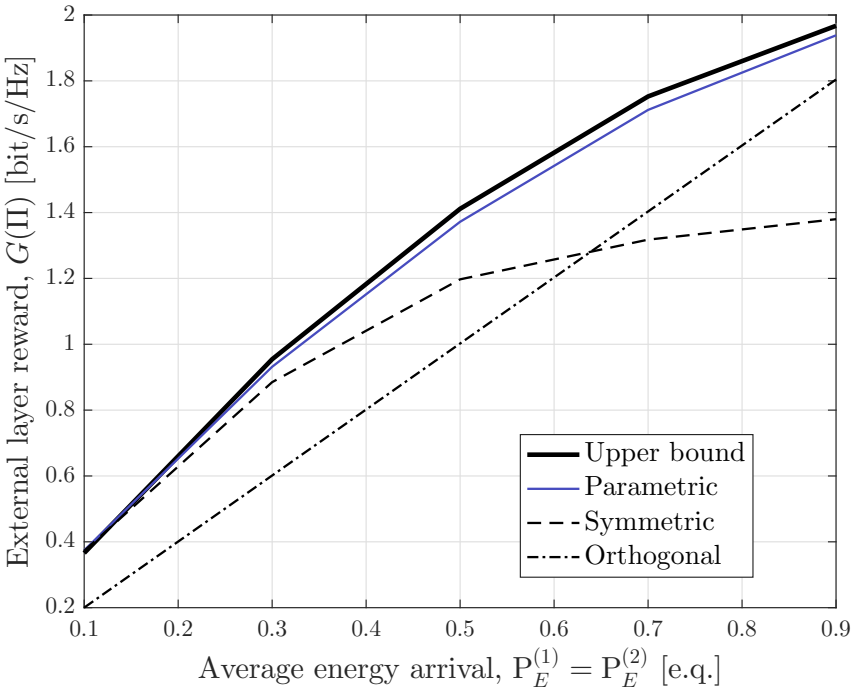

Figure 3.15. Long-term rewards as a function of the energy arrival rates for two users when $\mathbb{P}_E^{(1)} = \mathbb{P}_E^{(2)} = 0.05$.

3.2.8 Conclusions of Section 3.2

In the part of this chapter, we studied a decentralized optimization framework for a star topology energy harvesting communication network with collisions. We used a multi-layer Markov setup to model the system. An external layer, that models the jumps between SYNC slots (the time instants at which the policy is computed) is optimized as a Markov Decision Process (MDP) whose actions are given by an internal layer, modeled as a decentralized-MDP. To solve the external layer we used the Value Iteration Algorithm, whereas the Markov Policy Search algorithm was employed for the internal layer. Because of the exponential complexity of the exhaustive search used in the optimal policy, we presented two simpler schemes, namely WCSP-based and parametric approaches. In our numerical evaluations we described the trade-off between accessing the channel and energy arrivals, and we showed that a stationary access mechanism (e.g., fully orthogonal) is not guaranteed to be optimal under harvesting constraints. Moreover, we noted that decentralized schemes, if correctly designed, may achieve high performance while significantly reducing the signaling overhead in the network.

Our work introduces a principle of optimality in decentralized energy harvesting networks. However, because of the complexity of the optimal policy, additional studies are required for proposing more practical schemes which inherit the key properties of our framework while being less computationally demanding. Our future work will also include the investigation of different network topologies.

CHAPTER 4

Wireless Energy Transfer

In the previous chapters we have extensively discussed the importance of prolonging the lifetime of Wireless Sensor Networks (WSNs) and Internet of Things (IoT) networks. In this chapter we introduce the concept of Wireless Energy Transfer (WET) as a technique to achieve such a goal. Indeed, WET, differently from Environmental Energy Harvesting (EEH), has the major benefit of being *controllable*, and thus can be conveniently used when the devices are running out of energy.

Wireless energy transfer techniques have experienced a renewed research interest in the last few years [159] and several applications can be found in the WSN field, where low-power devices are fed with the transferred energy and use it for transmission or computation and sensing purposes. Different aspects of WET have been studied by both industry and academia, e.g., in terms of circuit and rectenna design [104] but also in terms of transmission protocols by the communication and networking community. In this field, three major research areas can be identified so far: Simultaneous Wireless Information and Power Transfer (SWIPT), energy cooperation and Wireless Powered Communication Networks (WPCNs). In SWIPT systems, the trade-offs between information and energy transfer are investigated [44]. Nowadays, because of hardware constraints of the current technology, a real *simultaneous* data and energy transmission is not possible yet, and therefore the Time Splitting (TS) and Power Splitting (PS) approaches were introduced [167]. The TS approach, in which WET and data transmission are temporally interleaved, was studied in [67, 79], whereas PS was analyzed in [100, 116, 135, 142]. The energy cooperation paradigm and WPCNs are the focus of this chapter.

Structure of the Chapter. The chapter is composed of two main parts which deal with different problems related to WET and can be read separately. In particular, Section 4.1 is devoted to the study of the energy cooperation paradigm in a network composed of two Energy Harvesting Devices (EHDs). We show that WET and EEH can be used jointly to improve the system performance. We consider a transmitter-receiver pair, showing how the WET improvement depends upon the statistics of the energy arrivals and the energy consumption of the devices. With the aim of maximizing a reward function, e.g., the average transmission rate, we find performance upper bounds with and without WET, define both online and offline optimization problems, and present results based on realistic

energy arrivals in indoor and outdoor environments. We show that WET can significantly improve the system performance even when a sizable fraction of the transmitted energy is wasted and that, in some scenarios, the online approach can obtain close to optimal performance.

Instead, Section 4.2 deals with wireless powered communication networks, which are becoming an effective solution for improved self sustainability of mobile devices. In this context, a hybrid access point transfers energy to a group of nodes, which use the harvested energy to perform computation or transmission tasks. While the availability of the wireless energy transfer mechanism opens up new frontiers, an appropriate choice of the network parameters (e.g., transmission powers, transmission duration, amount of transferred energy, etc.) is required in order to achieve high performance. In Section 4.2, we study the throughput optimization problem in a system composed of a receiver which recharges the batteries of two devices at different distances. In the literature, the main focus so far has been on the *slot-oriented* optimization, in which all the harvested energy is used in the same slot in which it is harvested. However, this approach is strongly suboptimal because it does not exploit the possibility to store the energy and use it at a later time. Thus, instead of considering the slot-oriented case, we address the *long-term* maximization. This assumption greatly increases the optimization complexity, as it requires to consider, e.g., the channel state statistics and the batteries evolution. Our objective is to find the best scheduling scheme, both for the energy transferred by the access point and for the data sent by the two nodes. We discuss how to perform the maximization with optimal as well as approximate techniques and show that the slot-oriented policies proposed so far are strongly suboptimal in the long run.

4.1 Energy Cooperation

4.1.1 Introduction

In the first part of this chapter we discuss the combination of two different techniques: Environmental Energy Harvesting (EEH), that allows a device to refill its battery gathering energy from the environment, and Wireless Energy Transfer (WET), that makes it possible to exchange energy among different devices. We show how WET and EEH can be jointly used to improve the overall system performance and prolong network lifetime. Indeed, in some scenarios, a node may receive much more energy and/or consume less energy than some of its neighbors. In these cases, it is reasonable to transmit energy from the rich energy source to other nodes in order to balance the energy levels. WET enables this possibility, and combining it with environmental energy harvesting is interesting because it allows to better exploit the renewable energy source and avoid energy overflows (see Definition 2.1.1). An example of application is the design of energy-aware routing algorithms that exploit the possibility of sharing energy.

As a first step to understand the key trade-offs before addressing more complex scenarios, in this part of the chapter we consider a network composed of two devices (here we focus on a transmitter and a receiver but the model can be readily extended to the case of two transmitters) equipped with EEH and WET interfaces. We explicitly take into account the effects of *finite batteries* and, differently from most of the related literature, model the devices energy consumption with *generic functions*. We show that, in the cases where the scenario is unbalanced, i.e., a device harvests much more energy than the other, it is possible to use wireless energy transfer to balance the energy levels of the two devices

and, as a consequence, to achieve higher rewards even when a significant fraction of the transmitted energy is wasted. We initially find analytical performance upper bounds with and without WET. Then, we investigate both online and offline approaches and compare them. We present two scenarios with realistic irradiation data showing that WET can be used to increase the average transmission rate. We also describe the effects of finite batteries on the system performance.

The model that we use in this section is based on Chapter 2, and is also similar to the one proposed in [90] for the optimization of an energy harvesting system *without* WET.

Several different technologies for Energy Transfer have been considered so far (see Section 1.1.2.1), e.g., Radio-Frequency Wireless Energy Transfer (RF-WET), Strongly Coupled Magnetic Resonances Wireless Energy Transfer (SCMR-WET), or inductive coupling. In this section we will set up a generic model which does not rely on any particular technology; however, our results are based on the parameters of a typical WET performed via SCMR.

Contributions. For a transmitter/receiver pair, we present performance upper bounds with and without WET when the energy costs are general functions that can include, e.g., the circuitry costs. The optimal online and offline policies are introduced and characterized. In particular, we use the offline case as a benchmark for our online policies in the finite horizon setting. We show that WET can significantly improve the system performance and that, in some scenarios, the online policies are close to optimal. We also consider the effects of finite batteries, showing that, although the reward improvement depends upon the battery size, it is not necessary to have very large batteries to obtain high gains.

Structure. The first part of this chapter is organized as follows. Section 4.1.2 defines the system model we analyzed, and Section 4.1.3 provides the performance upper bounds. In Sections 4.1.4 and 4.1.5 we introduce the online and offline policies, respectively. Section 4.1.6 presents the numerical evaluation for the online policies. In Section 4.1.7 we analyze two practical examples using realistic irradiation data. Finally, Section 4.1.8 draws the conclusions.

References. This section is based on the conference paper **[C3]** and on the journal paper **[J1]**.

4.1.2 System Model and Optimization Problems

We study Energy Harvesting Devices (EHDs) that, in addition to the capability of gathering energy from the environment through an EEH interface, are also able to *transmit* and/or *receive energy* via a WET mechanism. To characterize this technique, we will deal with a pair of EHDs where one device is the Transmitter (TX) that sends data to a Receiver (RX), whereas RX can send energy to TX (we will comment on the extension to bi-directional WET in Section 4.1.3.4).[1]

We assume a slotted-time system, where slot k corresponds to the time interval $[k, k+1)$ (see Section 2.1.1), with $k = 0, 1, \ldots$. Both devices are equipped with some interface that can harvest ambient energy, e.g., from solar light, indoor light, or vibrations (see Section 1.1.1).

TX transmits data packets toward RX and, in every slot, has a new data packet to send. The contributions to consider for data transmission are described in Section 2.1.3.

[1] The model we consider in this part is analogous to the one described in Section 2.1 when only one terminal node is considered and RX has energy harvesting capabilities.

For packet reception, instead, the main contributions are sampling (demodulation, filtering, quantization), processing (decoding) and storage [165]. We simplify the energy consumption models as follows. For reliable communications at rate R, TX needs to provide an SNR (thus a transmit power) that depends upon R. Similarly, also the reception power depends upon R because of sampling and processing. By combining these concepts, it is possible to establish a relationship between the reception power and the transmit power (see [8]). Formally, we describe the energy consumptions with two generic continuous, increasing and concave downward functions $Q^{\mathrm{tx}}(P)$, $Q^{\mathrm{rx}}(P)$, where P is the transmit power and $Q^{\mathrm{tx}}(0) = Q^{\mathrm{rx}}(0) = 0$ (sleep mode). The transmit power used in slot k, $P_k \in \mathcal{P} \triangleq [0, P_{\max}]$, is decided at the beginning of each slot.

Example 4.1.1. *For a transmitter, a common model for the energy function is [47, 132]*

$$Q^{\mathrm{tx}}(P) = \sigma^{\mathrm{tx}} P. \tag{4.1.1}$$

For the receiver, instead, a reasonable approximation is to assume that the energy function is proportional to the transmission rate:

$$Q^{\mathrm{rx}}(P) = \alpha^{\mathrm{rx}} \log(1 + \Lambda P). \tag{4.1.2}$$

This model is a good approximation when the circuitry costs are negligible. Note that in the low-SNR regime, we can approximate $Q^{\mathrm{rx}}(\cdot)$ as $Q^{\mathrm{rx}}(P) \approx \sigma^{\mathrm{rx}} P$.
σ^{tx}, σ^{rx}, α^{rx} are proper constants and Λ is an SNR scaling factor. The contributions of the circuitry costs can be included in this model by adding to Equations (4.1.1) *and* (4.1.2) *two terms $\zeta^{\mathrm{tx}}(P)$ and $\zeta^{\mathrm{rx}}(P)$ that, starting from* 0*, increase quickly until constant values in order to preserve the continuity and concavity of $Q^{\mathrm{tx}}(P)$ and $Q^{\mathrm{rx}}(P)$. Note that, in the general case, our model allows the circuitry costs for TX and RX to be different.*

The amount of energy to be sent with the WET mechanism, $Z_k^{\mathrm{rx}\to\mathrm{tx}} \geq 0$, is decided in every slot. The energy received in slot k can be exploited only in a later slot. We mainly focus on uni-directional energy transfer from RX to TX and discuss in Section 4.1.3.4 how to extend this hypothesis to the bi-directional case. We assume that only a fraction $\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}}$ of the transmitted energy is received, where $\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} \in [0, 1]$ is the *wireless energy transfer efficiency* (when SCMR technology is used to perform WET, we have $\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} \equiv \eta_{\mathrm{SCMR}}$ defined in Section 2.1.2). In this section, although our analysis is general, we will use a transfer efficiency $\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} = 0.15$ as a baseline.

The devices have finite batteries that can store at most $B_{\max}^{\mathrm{tx}}$ and $B_{\max}^{\mathrm{rx}}$ joule of energy. The randomness of the energy arrivals is described through two independent processes $\{E_k^{\mathrm{tx}}\}$ and $\{E_k^{\mathrm{rx}}\}$ with some statistics, e.g., deterministic, Bernoulli or truncated geometric (see Section 2.1.2). The energy arrival processes have means $\bar{E}^{\mathrm{tx}} > 0$ and $\bar{E}^{\mathrm{rx}} > 0$ and the energy harvested in a slot can be exploited only in a later slot.

With the introduced quantities, the evolutions of the two batteries can be described as:

$$B_{k+1}^{\mathrm{tx}} = \min\{B_{\max}^{\mathrm{tx}}, B_k^{\mathrm{tx}} - Q^{\mathrm{tx}}(P_k) + E_k^{\mathrm{tx,EEH}} + E_k^{\mathrm{tx,WET}}\}, \tag{4.1.3}$$

$$B_{k+1}^{\mathrm{rx}} = \min\{B_{\max}^{\mathrm{rx}}, B_k^{\mathrm{rx}} - Q^{\mathrm{rx}}(P_k) + E_k^{\mathrm{rx,EEH}} - Z_k^{\mathrm{rx}\to\mathrm{tx}}\}, \tag{4.1.4}$$

where B_k^{tx}, B_k^{rx} are the energy levels in slot k and $E_k^{\mathrm{tx,WET}} \equiv \eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} Z_k^{\mathrm{rx}\to\mathrm{tx}}$. Since we consider slots of fixed length, in this section we refer to power or energy interchangeably.

4.1.2.1 Optimization Problems

The *state of the system* $\mathbf{S}_k = (B_k^{\text{tx}}, B_k^{\text{rx}})$ is known to RX at the beginning of every slot. A policy μ defines which action to perform in every slot k, i.e., how much energy should be used to transmit data (which is equivalent to define the transmission powers $\mathbf{P} \triangleq \{P_0, P_1, \ldots\}$) and how much energy should be transferred (i.e., $\mathbf{Z}^{\text{rx}\to\text{tx}} \triangleq \{Z_0^{\text{rx}\to\text{tx}}, Z_1^{\text{rx}\to\text{tx}}, \ldots\}$).[2]

We consider as metrics the average unconstrained rewards in K slots and in the long term, defined as

$$G_\mu^K \triangleq \frac{1}{K}\sum_{k=0}^{K-1} r(P_k), \tag{4.1.5}$$

$$G_\mu \triangleq \liminf_{K\to\infty} G_\mu^K, \tag{4.1.6}$$

where $r(x)$ is a non-decreasing and concave function of x. As a baseline, we focus on the *average normalized transmission rate*, obtained when $r(x) = \log(1 + \Lambda x)$, where Λ is an SNR scaling factor.

We consider the following optimization problem

$$\mu^\star = \arg\max_\mu G_\mu, \tag{4.1.7}$$

subject to appropriate feasibility constraints (i.e., the transmission power and the transferred energy must be non-negative and must not exceed the energy available in the batteries). Since the optimization variables and the specific constraints depend upon the chosen approach, this problem will be discussed in more detail in Sections 4.1.4 and 4.1.5.

4.1.2.2 Optimization Approaches

To solve the optimization problem set up in the previous subsections, we proceed as follows. Initially, we introduce some performance upper bounds (Section 4.1.3), i.e., upper bounds to G_μ. These do not depend upon the optimization technique. Then, we discuss Problem (4.2.8) with the following approaches (Sections 4.1.4 and 4.1.5).

Online approach. In this case, in every time slot k, the policy chooses an action that depends upon the current state of the system $\mathbf{S}_k$ and upon the energy arrival statistics. In the online case, the output of the optimization process is a set of rules (one for every state of the system) that, given $\mathbf{S}_k$, can be applied to choose the action to perform. In order to model the system as a Markov Decision Process, in Section 4.1.4 we approximate the continuous model with a discrete one.

Offline approach. In this case, the policies are found by exploiting the non-causal knowledge of the energy arrivals (not only the statistics). In the offline case, the output of the optimization process is a pair of sequences $(\mathbf{P}, \mathbf{Z}^{\text{rx}\to\text{tx}})$ that define in every slot from 0 to $K-1$ which action to use.

The main focus of our work is on online policies which, though performing worse than offline policies in general, have the important advantage of not requiring non-causal knowledge of the energy arrivals, and are therefore practically usable. Offline policies will be used in Section 4.1.7 as a benchmark, showing that in some relevant cases the performance loss incurred by the online approach can be quite small.

[2] The specific structure of μ depends upon the considered scenario and will be discussed in more detail in Sections 4.1.4 and 4.1.5.

4.1.3 Upper Bounds

In this section we introduce upper bounds to G_μ for the cases with and without WET. This is an interesting problem because the presented upper bounds are closely approached in several cases of interest and provide an easy characterization of the system reward without performing any optimization.

They are derived in the infinite horizon case, but can be simply reformulated in the finite horizon case by changing the long-term means $\bar{E}^{\mathrm{tx}}$ and $\bar{E}^{\mathrm{rx}}$ with the means in K slots.[3] In particular, we will generalize the following intuitive results. As an example, consider $Q^{(i)}(P) = P$ (in the following, $i \in \{\mathrm{tx}, \mathrm{rx}\}$) and $\bar{E}^{\mathrm{rx}} > \bar{E}^{\mathrm{tx}}$ (RX harvests more energy than TX). An upper bound to G_μ without WET is given by $r(\bar{E}^{\mathrm{tx}})$ and is achievable if the devices consume in every slot (except possibly for a vanishing fraction of them) an amount equal to the average harvested energy. This can happen if the battery sizes are infinite or if the batteries are finite and the energy arrivals are deterministic. Moreover, since $\bar{E}^{\mathrm{rx}} > \bar{E}^{\mathrm{tx}}$, it may be interesting to use WET to improve the performance (we recall that RX can send energy to TX). In this case an upper bound is given by a balanced combination of the transmitter and receiver average energy arrivals: $r((\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}}\bar{E}^{\mathrm{rx}} + \bar{E}^{\mathrm{tx}})/(\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} + 1))$. Note that in this last expression both $\bar{E}^{\mathrm{tx}}$ and $\bar{E}^{\mathrm{rx}}$ contribute to increasing the upper bound. Also, we remark that the transfer efficiency $\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}}$ needs to be considered. These considerations are formalized in the general case in the following (note that, unlike in the above example, we do not impose any constraints on $\bar{E}^{\mathrm{tx}}$ and $\bar{E}^{\mathrm{rx}}$).

4.1.3.1 Upper Bound without WET

We first focus on the case without WET. We have the following result.

Theorem 4.1.2 (Upper Bound without WET)**.** *If there exist two continuous and increasing functions $\Psi^{\mathrm{tx}}(P)$, $\Psi^{\mathrm{rx}}(P)$ such that*

1. $0 \le \Psi^{\mathrm{tx}}(P) \le Q^{\mathrm{tx}}(P)$ *and* $0 \le \Psi^{\mathrm{rx}}(P) \le Q^{\mathrm{rx}}(P)$, $\forall P \in \mathcal{P}$, *and*
2. $r(\Psi^{\mathrm{tx}^{-1}}(\cdot))$ *and* $r(\Psi^{\mathrm{rx}^{-1}}(\cdot))$ *are concave functions,*

then an upper bound for the reward is

$$G_{\mathrm{u.b.}}^{\mathrm{noET}} = r\left(\min\left\{\Psi^{\mathrm{tx}^{-1}}(\bar{E}^{\mathrm{tx}}), \Psi^{\mathrm{rx}^{-1}}(\bar{E}^{\mathrm{rx}})\right\}\right). \tag{4.1.8}$$

If only $\Psi^{(i)}(P)$ exists, $i \in \{\mathrm{tx}, \mathrm{rx}\}$, then an upper bound is $G_{\mathrm{u.b.}}^{\mathrm{noET}} = r\left(\Psi^{i^{-1}}(\bar{E}^{(i)})\right)$.
If neither $\Psi^{\mathrm{tx}}(P)$ nor $\Psi^{\mathrm{rx}}(P)$ exists, then the optimal reward is infinite.

Proof. See Appendix 4.A. ■

Note that, in the previous theorem, we convert a *energy consumption* $\bar{E}^{(i)}$ (or, equivalently, a power consumption, since the slot length is fixed) to a *reward* in two steps. First, we apply the inverse function $\Psi^{i^{-1}}$ to convert the power consumption into a *transmission power*. Then, we apply the function $r(\cdot)$ to the transmission power in order to obtain the corresponding reward.

[3] In the following, we find upper bounds based on the *means* of the harvesting processes. Thus, even if we do not explicitly take into account the specific random behavior of the energy arrivals, we are still considering the fact that energy is gathered over time, which is a fundamental feature of EEH.

In practice, $\Psi^{(i)}(\cdot)$ is an optimistic auxiliary energy consumption function that makes it possible to mathematically obtain (4.1.8). Intuitively, the closer $\Psi^{(i)}(\cdot)$ and $Q^{(i)}(\cdot)$, the tighter the upper bound.

Remark If $\mathcal{P}$ is bounded, i.e., $P_{\max} < \infty$, then the conditions of Theorem 4.1.2 only need to be satisfied for a finite range of P, and therefore it is always possible to find $\Psi^{(i)}(\cdot)$.

Note that a particular case of the previous remark is obtained when the battery sizes are finite. In this case $P_{\max}$ is bounded by the maximum battery size (in particular $Q^{(i)}(P_{\max}) \leq B^{(i)}_{\max}$).

As shown in the following corollaries, there exist cases in which the upper bound of Theorem 4.1.2 can be achieved.

Corollary 4.1.3. *If $Q^{\text{tx}}(\cdot) = \Psi^{\text{tx}}(\cdot)$ and $Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}}) \leq Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}})$ (TX is the bottleneck) then, in the deterministic energy arrivals case,[4] the upper bound* (4.1.8) *is achievable with finite batteries $B^{\text{tx}}_{\max} \geq \bar{E}^{\text{tx}}$, $B^{\text{rx}}_{\max} \geq Q^{\text{rx}}(Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}}))$. An optimal policy is*

$$P_k = \begin{cases} Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}}), & \text{if } \bar{E}^{\text{tx}} \leq B^{\text{tx}}_k \text{ and } Q^{\text{rx}}(Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}})) \leq B^{\text{rx}}_k, \\ 0, & \text{otherwise.} \end{cases} \tag{4.1.9}$$

The same holds if the roles of TX and RX are exchanged.

Proof. Let $v = Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}})$.
Assume that at the beginning $E^{\text{tx}}_0 = E^{\text{rx}}_0 = 0$. The batteries evolution is the following: $E^{\text{tx}}_1 = \bar{E}^{\text{tx}}$, $E^{\text{rx}}_1 = \bar{E}^{\text{rx}}$. Note that $Q^{\text{tx}}(v) = \bar{E}^{\text{tx}}$ by definition and $Q^{\text{rx}}(v) \leq \bar{E}^{\text{rx}}$ by hypothesis. At $k = 2$, we have: $E^{\text{tx}}_2 = 2\bar{E}^{\text{tx}} - Q^{\text{tx}}(v) = \bar{E}^{\text{tx}}$ (transmit with power v and then harvest an amount of energy exactly equal to $\bar{E}^{\text{tx}}$) and $E^{\text{rx}}_2 = 2\bar{E}^{\text{rx}} - Q^{\text{rx}}(v) \geq \bar{E}^{\text{rx}}$. Thus, in every slot, excluding an initial transient, TX can transmit data with power v and RX is always able to receive them, thus the reward per slot is $r(v)$. In the long-term, the upper bound in Equation (4.1.8) is achieved. With different initial states the reasoning is the same.
Note that in the previous considerations we implicitly used the hypotheses $B^{\text{tx}}_{\max} \geq \bar{E}^{\text{tx}}$, $B^{\text{rx}}_{\max} \geq Q^{\text{rx}}(Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}}))$, that are necessary to obtain the thesis. ■

The policy of Equation (4.1.9), possibly excluding an initial transient, consumes all the energy that is received in every slot, and thus achieves the upper bound $r(\bar{E}^{\text{tx}})$.

When the battery sizes are infinite, Corollary 4.1.3 generalizes to any energy arrival process.

Corollary 4.1.4. *If $Q^{\text{tx}}(\cdot) = \Psi^{\text{tx}}(\cdot)$, $Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}}) \leq Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}})$ (TX is the bottleneck) and the battery sizes are infinite then the upper bound* (4.1.8) *is achievable for any statistics of the energy arrivals. The same holds if the roles of TX and RX are exchanged.*

A formal proof of Corollary 4.1.4 is given in [113] for the special case of a linear energy consumption model in a single EHD, but can be extended to our case. To show that a reward arbitrarily close to the upper bound can be achieved, a *Save-and-Transmit Scheme* was introduced, where the device does not transmit in an initial transient in order to accumulate enough energy to absorb energy fluctuations, so as to avoid energy outage and manage to consume and receive, on average, the same energy.

[4] Note that, since we consider i.i.d. energy arrivals, *deterministic* is equivalent to *constant*.

4.1.3.2 Upper Bound with WET

We now derive similar results for the case where WET is considered. We introduce two new functions $\bar{C}^{\mathrm{tx}}(\cdot)$ and $\bar{C}^{\mathrm{rx}}(\cdot)$ defined as follows:

$$\bar{C}^{\mathrm{tx}}(\xi) = \bar{E}^{\mathrm{tx}} + \eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} \bar{E}^{\mathrm{rx}}(1-\xi), \tag{4.1.10}$$

$$\bar{C}^{\mathrm{rx}}(\xi) = \bar{E}^{\mathrm{rx}}\xi, \tag{4.1.11}$$

where $\xi \in [0,1]$ is a constant that represents the average fraction of the harvested energy that is transferred with WET under a policy μ. $\bar{C}^{(i)}(\xi)$ represents the average amount of energy that can be exploited at device $i \in \{\mathrm{tx}, \mathrm{rx}\}$ to transmit or receive. In particular, RX transfers part of the harvested energy, thus the residual energy that it can exploit is, on average, only a fraction ξ of the harvested one ($\bar{E}^{\mathrm{rx}}$). TX, in addition to its own harvested energy ($\bar{E}^{\mathrm{tx}}$), receives the energy that RX transferred (scaled by the energy transfer efficiency $\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}}$). One of the key results of this section is stated in the following theorem.

Theorem 4.1.5 (Upper Bound with WET)**.** *Under the hypotheses of Theorem 4.1.2, when WET is used, an upper bound to G_μ is*

$$G_{\mathrm{u.b.}}^{\mathrm{WET}} = r\left(\Psi^{\mathrm{rx}^{-1}}(\bar{C}^{\mathrm{rx}}(\xi^\star))\right), \tag{4.1.12}$$

where

- *if* $\Psi^{\mathrm{rx}^{-1}}(\bar{C}^{\mathrm{rx}}(1)) \leq \Psi^{\mathrm{tx}^{-1}}(\bar{C}^{\mathrm{tx}}(1))$*, then* $\xi^\star = 1$*;*
- *otherwise,* $\xi^\star$ *is such that* $\Psi^{\mathrm{tx}^{-1}}(\bar{C}^{\mathrm{tx}}(\xi^\star)) = \Psi^{\mathrm{rx}^{-1}}(\bar{C}^{\mathrm{rx}}(\xi^\star))$*.*

Proof. From Theorem 4.1.2, an upper bound is given using $\bar{E}^{\mathrm{tx}}$ and $\bar{E}^{\mathrm{rx}}$ inside the "min" operation. When WET is used, the average amounts of incoming energy at TX and RX are $\bar{C}^{\mathrm{tx}}(\xi)$ and $\bar{C}^{\mathrm{rx}}(\xi)$, respectively. Thus, when ξ is fixed, an upper bound is

$$G_{\mathrm{u.b.}}^{\mathrm{WET}}(\xi) = r\left(\min\left\{\Psi^{\mathrm{tx}^{-1}}(\bar{C}^{\mathrm{tx}}(\xi)), \Psi^{\mathrm{rx}^{-1}}(\bar{C}^{\mathrm{rx}}(\xi))\right\}\right). \tag{4.1.13}$$

In practice, we replaced $\bar{E}^{\mathrm{tx}}$ and $\bar{E}^{\mathrm{rx}}$ with $\bar{C}^{\mathrm{tx}}(\xi)$ and $\bar{C}^{\mathrm{rx}}(\xi)$ because, with WET, the energy that the devices can exploit is described by $\bar{C}^{\mathrm{tx}}(\xi)$ and $\bar{C}^{\mathrm{rx}}(\xi)$ (see the description of Equation (4.1.10)-(4.1.11)).

Note that $\Psi^{i^{-1}}(\cdot)$ is an increasing and continuous function because $\Psi^{(i)}(\cdot)$ is increasing and continuous. Moreover, $\partial\bar{C}^{\mathrm{tx}}(\xi)/\partial\xi < 0$ and $\partial\bar{C}^{\mathrm{rx}}(\xi)/\partial\xi > 0$. Thus, the first argument of the minimum in Equation (4.1.13) is decreasing in ξ, whereas the second one is increasing. The minimum of the two is maximized when they are equal, if this is possible, or otherwise for the maximum value of ξ, i.e., $\xi^\star = 1$. Note that, since $\Psi^{\mathrm{tx}^{-1}}(\bar{C}^{\mathrm{tx}}(0)) > \Psi^{\mathrm{rx}^{-1}}(\bar{C}^{\mathrm{rx}}(0)) = 0$, $\xi^\star$ is equal to one if and only if at $\xi = 1$ we have $\Psi^{\mathrm{rx}^{-1}}(\bar{C}^{\mathrm{rx}}(1)) \leq \Psi^{\mathrm{tx}^{-1}}(\bar{C}^{\mathrm{tx}}(1))$, i.e., $\Psi^{\mathrm{tx}^{-1}}(\bar{C}^{\mathrm{tx}}(\xi))$ and $\Psi^{\mathrm{rx}^{-1}}(\bar{C}^{\mathrm{rx}}(\xi))$ do not have an intersection point in $[0,1)$. ■

Corollaries 4.1.3 and 4.1.4 can be generalized as follows.

Corollary 4.1.6. *If* $Q^{\text{tx}}(\cdot) = \Psi^{\text{tx}}(\cdot)$ *and* $Q^{\text{rx}}(\cdot) = \Psi^{\text{rx}}(\cdot)$ *then, in the deterministic energy arrivals case, the upper bound* (4.1.12) *is achievable with finite batteries* $B^{\text{tx}}_{\max} \geq Q^{\text{tx}}(Q^{\text{rx}^{-1}}(\bar{C}^{\text{rx}}(\xi^\star)))$, $B^{\text{rx}}_{\max} \geq \bar{E}^{\text{rx}}$. *An optimal policy is*

$$P_k = \begin{cases} Q^{\text{rx}^{-1}}(\bar{C}^{\text{rx}}(\xi^\star)), & \text{if } \bar{C}^{\text{rx}}(\xi^\star) \leq B^{\text{rx}}_k \text{ and } Q^{\text{tx}}(Q^{\text{rx}^{-1}}(\bar{C}^{\text{rx}}(\xi^\star))) \leq B^{\text{tx}}_k, \\ 0, & \text{otherwise}, \end{cases} \tag{4.1.14}$$

$$Z^{\text{rx}\to\text{tx}}_k = \begin{cases} \bar{E}^{\text{rx}} - Q^{\text{rx}}(P_k), & \text{if } B^{\text{rx}}_k \geq \bar{E}^{\text{rx}}, \\ 0, & \text{otherwise}. \end{cases} \tag{4.1.15}$$

Proof. The proof is similar to that of Corollary 4.1.3. Let $v = Q^{\text{rx}^{-1}}(\bar{C}^{\text{rx}}(\xi^\star))$. At $k = 1$, $E^{\text{tx}}_1 = \bar{E}^{\text{tx}}$ and $E^{\text{rx}}_1 = \bar{E}^{\text{rx}}$.

If $\bar{E}^{\text{tx}}$ is greater than or equal to $Q^{\text{tx}}(v)$, then the policy chooses $P_1 = v$ because $\bar{C}^{\text{rx}}(\xi^\star) \leq \bar{E}^{\text{rx}}$ by definition of $\bar{C}^{\text{rx}}(\cdot)$ and $D_1 = \bar{E}^{\text{rx}} - Q^{\text{rx}}(v)$ because $E^{\text{rx}}_1 \geq \bar{E}^{\text{rx}}$. Note that the sum $Q^{\text{rx}}(P_1) + D_1$ is equal to $\bar{E}^{\text{rx}}$, thus, at $k = 2$, $E^{\text{rx}}_2 = \bar{E}^{\text{rx}}$. Instead, for TX, $E^{\text{tx}}_2 = \bar{E}^{\text{tx}} - Q^{\text{tx}}(v) + \bar{E}^{\text{tx}} + \eta^{\text{rx}\to\text{tx}}_{\text{WET}}(\bar{E}^{\text{rx}} - Q^{\text{rx}}(v)) = \bar{E}^{\text{tx}} - Q^{\text{tx}}(v) + \bar{C}^{\text{tx}}(\xi^\star)$. If $\xi^\star < 1$, then $E^{\text{tx}}_2 = \bar{E}^{\text{tx}}$ because $v = Q^{\text{tx}^{-1}}(\bar{C}^{\text{tx}}(\xi^\star))$, otherwise $E^{\text{tx}}_2 \geq \bar{E}^{\text{tx}}$ (see Theorem 4.1.5).

If instead $\bar{E}^{\text{tx}} < Q^{\text{tx}}(v)$, the policy chooses $P_1 = 0$ and $D_1 = \bar{E}^{\text{rx}}$. Note that, if $\xi^\star = 1$, we have $Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}}) \leq Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}})$ and the inequality chain becomes $Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}}) \leq Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}}) < Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}})$, which is not possible. Thus $\xi^\star$ must be less than 1 and $Q^{\text{tx}^{-1}}(\bar{C}^{\text{tx}}(\xi^\star)) = Q^{\text{rx}^{-1}}(\bar{C}^{\text{rx}}(\xi^\star))$ implies that $Q^{\text{tx}}(v) = \bar{C}^{\text{tx}}(\xi^\star) > \bar{E}^{\text{tx}}$. At $k = 2$, $E^{\text{tx}}_2 = 2\bar{E}^{\text{tx}} + \eta^{\text{rx}\to\text{tx}}_{\text{WET}}\bar{E}^{\text{rx}} > \bar{C}^{\text{tx}}(\xi^\star) = Q^{\text{tx}}(v)$ and $E^{\text{rx}}_2 = \bar{E}^{\text{rx}}$. For $k \geq 2$, TX always has enough energy to transmit with power v.

The previous considerations hold if the battery sizes satisfy the hypotheses of the theorem. Thus, after an initial transient, the devices always have enough energy to transmit and receive with power v and in the long term the upper bound (4.1.12) is achieved. ■

Corollary 4.1.7. *If* $Q^{\text{tx}}(\cdot) = \Psi^{\text{tx}}(\cdot)$, $Q^{\text{rx}}(\cdot) = \Psi^{\text{rx}}(\cdot)$ *and the battery sizes are infinite, then the upper bound* (4.1.8) *is achievable for any statistics of the energy arrivals.*

Proof. See Corollary 4.1.4. ■

The following result establishes when it is beneficial to use WET.

Proposition 4.1.8. *If* $Q^{\text{tx}}(\cdot) = \Psi^{\text{tx}}(\cdot)$ *and* $Q^{\text{rx}}(\cdot) = \Psi^{\text{rx}}(\cdot)$, *WET always improves the upper bound (i.e.,* $G^{\text{WET}}_{\text{u.b.}} > G^{\text{noET}}_{\text{u.b.}}$*) if and only if* $\xi^\star < 1$.

Proof. WET improves the performance if $G^{\text{noET}}_{\text{u.b.}} < G^{\text{WET}}_{\text{u.b.}}$, i.e.,

$$r\left(\min\left\{Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}}), Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}})\right\}\right) < r\left(Q^{\text{rx}^{-1}}(\bar{C}^{\text{rx}}(\xi^\star))\right). \tag{4.1.16}$$

Since $r(\cdot)$ is an increasing function, the previous condition is equivalent to

$$\min\left\{Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}}), Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}})\right\} < Q^{\text{rx}^{-1}}(\bar{C}^{\text{rx}}(\xi^\star)). \tag{4.1.17}$$

- *(if)* $\xi^\star < 1$ means that (see Theorem 4.1.5) $Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}}) > Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}})$, thus the condition becomes $Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}}) < Q^{\text{rx}^{-1}}(\bar{C}^{\text{rx}}(\xi^\star))$. Thanks to Theorem 4.1.5 and to

(4.1.10)-(4.1.11), and since $Q^{\text{tx}^{-1}}(\cdot)$ is increasing, if $\xi^\star < 1$, then $Q^{\text{rx}^{-1}}(\bar{C}(\xi^\star)) = Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}} + \eta_{\text{WET}}^{\text{rx}\to\text{tx}}\bar{E}^{\text{rx}}(1-\xi^\star)) > Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}})$;

- *(only if)* Assume $\xi^\star = 1$. In this case $Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}}) \leq Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}})$, which implies $G_{\text{u.b.}}^{\text{noET}} = r(Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}}))$ and $G_{\text{u.b.}}^{\text{WET}} = r(Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}}))$, thus WET does not improve the performance upper bound.

■

Note that when $Q^{\text{tx}}(\cdot) = \Psi^{\text{tx}}(\cdot)$ and $Q^{\text{rx}}(\cdot) = \Psi^{\text{rx}}(\cdot)$, $\xi^\star < 1$ is equivalent to $Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}}) < Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}})$. Thus, independently of the transfer efficiency $\eta_{\text{WET}}^{\text{rx}\to\text{tx}}$, if the average amount of energy harvested per slot at RX ($\bar{E}^{\text{rx}}$) corresponds to a transmission power ($Q^{\text{rx}^{-1}}(\bar{E}^{\text{rx}})$) greater than what is used at TX ($Q^{\text{tx}^{-1}}(\bar{E}^{\text{tx}})$), then the use of WET results in an increased upper bound. When $\xi^\star = 1$, WET cannot provide any improvement because RX is the energy bottleneck and therefore is unable to cooperate with TX. Also, note that the previous considerations also apply to the actual performance for the deterministic energy arrival case (in which the upper bounds can be achieved).

According to the above results, we can identify three main reasons why the upper bounds may not be achieved: 1) The functions $Q^{(i)}(\cdot)$ and $\Psi^{(i)}(\cdot)$ do not coincide. In this case, the only chance to obtain a better upper bound is to redefine $\Psi^{(i)}(\cdot)$, if possible. In the following examples we show how to derive $\Psi^{(i)}(\cdot)$ in several cases of interest. 2) The batteries are small (see Corollaries 4.1.4 and 4.1.7). As the battery sizes grow, the performance gets closer to the upper bounds. 3) The time horizon is finite. Indeed, the save and transmit scheme of Corollary 4.1.4 can be applied only if an infinite number of slots are available.

4.1.3.3 Examples

Example 4.1.9. *Consider the low-SNR regime (see Example 4.1.1). In this case the energy consumptions of both the transmitter and the receiver are linear in P. The functions $\Psi^{(i)}(\cdot)$ can then be taken equal to $Q^{(i)}(\cdot)$ and the upper bounds are*

$$G_{\text{u.b.}}^{\text{noET}} = r\left(\min\left\{\frac{\bar{E}^{\text{tx}}}{\sigma^{\text{tx}}}, \frac{\bar{E}^{\text{rx}}}{\sigma^{\text{rx}}}\right\}\right), \quad G_{\text{u.b.}}^{\text{WET}} = r\left(\frac{\bar{E}^{\text{rx}}}{\sigma^{\text{rx}}}\xi^\star\right),$$
$$\xi^\star = \min\left\{1, \frac{\sigma^{\text{rx}}}{\bar{E}^{\text{rx}}}\frac{\eta_{\text{WET}}^{\text{rx}\to\text{tx}}\bar{E}^{\text{rx}} + \bar{E}^{\text{tx}}}{\eta_{\text{WET}}^{\text{rx}\to\text{tx}}\sigma^{\text{rx}} + \sigma^{\text{tx}}}\right\}.$$

$\xi^\star$ is a linear combination of the average energy arrivals and is used to balance $\bar{C}^{\text{tx}}$ and $\bar{C}^{\text{rx}}$.

Example 4.1.10. *Another interesting case is $Q^{\text{tx}} = \sigma^{\text{tx}}P$, $Q^{\text{rx}} = \alpha^{\text{rx}}\log(1+\Lambda P)$ (Equation (4.1.2)) and $r(x) = \log(1+\Lambda P)$. Note that $r(\cdot)$ and $Q^{\text{rx}}(\cdot)$ are proportional and $r(Q^{\text{rx}^{-1}}(x)) = x/\alpha^{\text{rx}}$ is concave. Also in this example the functions $\Psi^{(i)}(\cdot)$ can be taken equal to $Q^{(i)}(\cdot)$. The upper bounds become*

$$G_{\text{u.b.}}^{\text{noET}} = \min\left\{r\left(\frac{\bar{E}^{\text{tx}}}{\sigma^{\text{tx}}}\right), \frac{\bar{E}^{\text{rx}}}{\alpha^{\text{rx}}}\right\}, \quad G_{\text{u.b.}}^{\text{WET}} = \frac{\bar{C}^{\text{rx}}(\xi^\star)}{\alpha^{\text{rx}}},$$

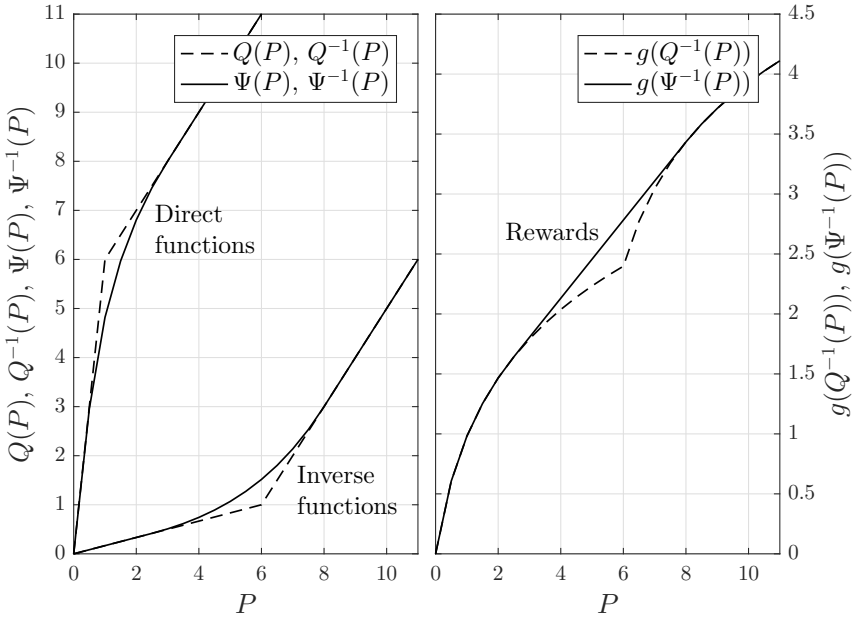

Figure 4.1. $Q(P)$, $\Psi(P)$, their inverse functions, and $r(Q^{-1}(P))$, $r(\Psi^{-1}(P))$ of Example 4.1.11 as a function of P.

where $\xi^\star$ is the unique solution of

$$\frac{\bar{E}^{\mathrm{tx}} + \eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} \bar{E}^{\mathrm{rx}}(1-\xi)}{\sigma^{\mathrm{tx}}} = \frac{e^{\xi \bar{E}^{\mathrm{rx}}/\alpha^{\mathrm{rx}}} - 1}{\Lambda},$$

if $\Lambda \bar{E}^{\mathrm{tx}}/\sigma^{\mathrm{tx}} < e^{b^{\mathrm{rx}}/\alpha^{\mathrm{rx}}} - 1$, and $\xi^\star = 1$ otherwise.

Example 4.1.11. *We now want to show a case where $\Psi^{(i)}(\cdot)$ and $Q^{(i)}(\cdot)$ are not the same. Consider $r(x) = \log(1+\Lambda x)$, $\bar{E} \triangleq \bar{E}^{\mathrm{tx}} = \bar{E}^{\mathrm{rx}}$ and $Q(\cdot) \triangleq Q^{\mathrm{tx}}(\cdot) = Q^{\mathrm{rx}}(\cdot)$ with*

$$Q(P) = \begin{cases} \dfrac{\zeta + P_n}{P_n} P, & \text{if } P < P_n, \\ \zeta + P, & \text{if } P \geq P_n, \end{cases} \tag{4.1.18}$$

with P_n arbitrarily close to 0. Note that this energy consumption model is suitable for the cases where the circuitry costs are not negligible. If we choose $\Psi(P) = Q(P)$, then it can be verified that there exist values of ζ and $\bar{E}$ such that $r(Q^{-1}(\cdot))$ is not concave. In this case $r(Q^{-1}(\bar{E}))$ is not guaranteed to be an upper bound.

However, an upper bound can be found by considering a function $\Psi(P)$ defined as in Theorem 4.1.2. In Figure 4.1 we plot $Q(P)$, $\Psi(P)$ and their inverse functions when $\Lambda = 10$, $\zeta = 5$, $B_{\max} = 11$ (normalized quantities). For the purpose of illustration, we arbitrarily set $P_n = 1$. Note that $r(Q^{-1}(P))$ is piece-wise concave whereas $r(\Psi^{-1}(P))$ is always concave. The function $\Psi(P)$ is such that $r(\Psi^{-1}(P))$ is divided in three regions. The two external regions are equal to two concave portions of $r(Q^{-1}(P))$. The central region is designed to be concave where $r(Q^{-1}(P))$ is not, and is obtained considering the straight line that is tangent to $r(Q^{-1}(P))$ in two points without intersecting it. In Section 4.1.6 we show that the upper bound given by this choice of $\Psi(P)$ is close to the real performance.

4.1.3.4 Extension: Bi-directional Energy Transfer

In the following, we present how our model can be extended to the bi-directional WET case. In this context, also TX is able to send part of its stored energy to RX when appropriate. In slot k, TX sends an amount of energy $Z_k^{\text{tx}\to\text{rx}}$ to RX, whereas RX sends $Z_k^{\text{rx}\to\text{tx}}$ to TX. The second term inside the minimum of Equation (4.1.3) has to be changed to

$$B_k^{\text{tx}} - Q^{\text{tx}}(P_k) + E_k^{\text{tx,EEH}} + E_k^{\text{tx,WET}} - Z_k^{\text{tx}\to\text{rx}} \tag{4.1.19}$$

and similarly for Equation (4.1.4) by switching "tx" and "rx".

The optimization of Equation (4.2.8) in this case provides three quantities, i.e., the transmission power, the energy sent from RX to TX and vice-versa.[5]

The upper bound of Equation (4.1.8) does not change because it does not depend upon WET. Theorem 4.1.5 can be reformulated by changing Equations (4.1.10)-(4.1.11) as follows

$$\bar{C}^{\text{tx}}(\xi^{\text{tx}\to\text{rx}}, \xi^{\text{rx}\to\text{tx}}) = \bar{E}^{\text{tx}}\xi^{\text{tx}\to\text{rx}} + \eta_{\text{WET}}^{\text{rx}\to\text{tx}}\bar{E}^{\text{rx}}(1-\xi^{\text{rx}\to\text{tx}}), \tag{4.1.20}$$

$$\bar{C}^{\text{rx}}(\xi^{\text{tx}\to\text{rx}}, \xi^{\text{rx}\to\text{tx}}) = \bar{E}^{\text{rx}}\xi^{\text{rx}\to\text{tx}} + \eta_{\text{WET}}^{\text{tx}\to\text{rx}}\bar{E}^{\text{tx}}(1-\xi^{\text{tx}\to\text{rx}}), \tag{4.1.21}$$

where $\xi^{i\to\text{j}}$ represents the average fraction of the harvested energy that is sent from device i to device j.

In our work we decided to focus on the uni-directional case and outline in this section how to extend it for presentation simplicity. Moreover, uni-directional WET can be effectively used in the practically relevant cases where one device harvests more energy than the other. Finally, uni-directional WET can be seen as a simpler lower bound for the bi-directional case.

4.1.4 Online Optimization

We now discuss the online approach and focus on long-term optimization. According to Section 4.1.2.2, the aim of an online policy is to define a set of rules that, given the state of the system in a slot, specifies which action (transmission power and transferred energy) should be used in that slot. The online approach is interesting because it requires only a statistical knowledge of the energy arrival process, thus may be effectively used in practice.

In order to formulate the problem as a discrete Markov Decision Process (MDP) (for which there exist efficient solving algorithms, see Section 2.2.2), we introduce the notion of energy quanta (e.q.), i.e., we discretize the amounts of energy (energy arrivals, energy consumptions, energy stored, energy exchanged). The batteries have integer sizes $b_{\max}^{\text{tx}}$, $b_{\max}^{\text{rx}}$ and can be considered as buffers (see Section 2.2.1). In order to obtain a consistent formulation, the values of $b_{\max}^{\text{tx}}$ and $b_{\max}^{\text{rx}}$ are chosen such that $B_{\max}^{\text{tx}}/b_{\max}^{\text{tx}} = B_{\max}^{\text{rx}}/b_{\max}^{\text{rx}}$. Under this assumption, one energy quantum corresponds to $\delta \triangleq B_{\max}^{(i)}/b_{\max}^{(i)}$ J. Therefore, when we deal with the online model, all the energy values are expressed as a (not necessarily integer) number of energy quanta.

With the above formulation, we will model the system as a finite two-dimensional Markov Chain (MC). When the MC is in state $\mathbf{b} \triangleq (b^{\text{tx}}, b^{\text{rx}})$, TX and RX have b^{tx} and b^{rx} energy quanta stored in their batteries, respectively. In every state of the MC, a decision

[5]Note that for any realistic system, in which $\eta_{\text{WET}}^{\text{rx}\to\text{tx}} < 1$ and $\eta_{\text{WET}}^{\text{tx}\to\text{rx}} < 1$, under the optimal policy we must have $Z_k^{\text{tx}\to\text{rx}} Z_k^{\text{rx}\to\text{tx}} = 0$, i.e., transferring non-zero energy in both directions simultaneously is strictly suboptimal.

is made on the transmission power $P(\mathbf{b}) \in \mathscr{P}$ of TX and on how many energy quanta RX transfers to TX, namely $z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b}) \in \{0, 1, \ldots, b^{\mathrm{rx}}_{\max}\}$.

Following the approach of [92, 149], in this part we only consider deterministic policies. Therefore, an online policy η specifies a mapping between the current state of the system, $\mathbf{b}$, and the corresponding action (transmitted power $P(\mathbf{b})$ and transferred energy $z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b})$), i.e., $\eta = \{(P(\mathbf{b}), z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b})), \forall \mathbf{b}\}$. Through an online policy η, a specific sequence of energy arrivals can be simply mapped to a sequence of actions $(\mathbf{P}, \mathbf{Z}^{\mathrm{rx}\to\mathrm{tx}})$.

The batteries evolution (4.1.3)-(4.1.4) can be rewritten in terms of energy quanta where, instead of $\eta^{\mathrm{rx}\to\mathrm{tx}}_{\mathrm{WET}} Z^{\mathrm{rx}\to\mathrm{tx}}_k$ and $Q^{(i)}(P_k)$, we use $\lfloor \eta^{\mathrm{rx}\to\mathrm{tx}}_{\mathrm{WET}} z^{\mathrm{rx}\to\mathrm{tx}}_k \rfloor$ and $q^{(i)}(P_k) \triangleq \lceil Q^{(i)}(P_k)/\delta \rceil$, respectively (see Section 2.2.1).

We restrict our study to the set of feasible policies, i.e., those in which, for every $\mathbf{b}$, we have $P(\mathbf{b}) \geq 0$, $z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b}) \geq 0$, $q^{\mathrm{tx}}(P(\mathbf{b})) \leq b^{\mathrm{tx}}$, $q^{\mathrm{rx}}(P(\mathbf{b})) + z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b}) \leq b^{\mathrm{rx}}$.

The reward of Equation (4.1.6) does not depend upon the starting state when the underlying MC has a unique recurrent class [75, 126]. Under this assumption, the long-term reward can be rewritten as

$$G_\eta = \sum_{b^{\mathrm{tx}}=0}^{b^{\mathrm{tx}}_{\max}} \sum_{b^{\mathrm{rx}}=0}^{b^{\mathrm{rx}}_{\max}} \pi_\eta(\mathbf{b}) r(P(\mathbf{b})), \tag{4.1.22}$$

where $\pi_\eta(\mathbf{b})$ is the steady-state probability of being in state $\mathbf{b}$ under policy η. The optimization variables of Problem (4.2.8) become $(P(\mathbf{b}), z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b})), \forall \mathbf{b}$ and the maximization is performed over all the feasible policies.

The Optimal Online Policy (OP-ON), namely $\eta^\star$, that maximizes G_η can be found numerically with the Policy Iteration Algorithm (PIA) (see Section 2.2.2) by exploiting the full energy arrivals statistics. The algorithm starts with an initial policy (thus we arbitrarily initialize $P(\mathbf{b})$ and $z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b})$) and then performs the *policy evaluation* and *improvement* steps in order to iteratively find a new policy, until the reward function G_η converges (for additional details see [14, Section. 7.2]).

4.1.4.1 Low Complexity Policies

In addition to the optimal online policy OP-ON, here we also introduce some simple heuristic policies, that will be used in the numerical evaluations in Sections 4.1.6 and 4.1.7 to show that, even when suboptimal policies are adopted, the system reward can be improved using WET.

When EEH is combined with WET, the structure of the optimal policy is complex and, moreover, depends upon the energy arrival processes and the energy consumption functions. For these reasons, it is difficult to introduce a simple policy that approximates the optimal one in a broad range of values.

In the general case, we define the Greedy Policy (GP) as follows[6]

$$\begin{aligned} P(\mathbf{b}) &= \min\left\{q^{\mathrm{tx}^{-1}}(b^{\mathrm{tx}}), q^{\mathrm{rx}-1}(b^{\mathrm{rx}})\right\}, \\ z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b}) &= b^{\mathrm{rx}} - q^{\mathrm{rx}}(P(\mathbf{b})). \end{aligned} \tag{4.1.23}$$

GP is a simple policy that empties at least one battery in every slot and is independent of the energy arrivals. Consider now the case where both TX and RX have $Q^{(i)}(\cdot) = \Psi^{(i)}(\cdot)$. We introduce two other policies, namely BP and LCP, as extensions of GP.

Balanced Policy (BP). The balanced policy is defined as the solution of the following system (note that BP does not depend upon the energy arrival statistics, a useful feature when the harvesting process is unknown)

$$\begin{cases} b_k^{\mathrm{tx}} + \eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} z_k^{\mathrm{rx}\to\mathrm{tx}} - q^{\mathrm{tx}}(P_k) = b_k^{\mathrm{rx}} - z_k^{\mathrm{rx}\to\mathrm{tx}} - q^{\mathrm{rx}}(P_k), \\ P_k = \min\left\{q^{\mathrm{tx}^{-1}}(b_k^{\mathrm{tx}}), q^{\mathrm{rx}^{-1}}(b_k^{\mathrm{rx}} - z_k^{\mathrm{rx}\to\mathrm{tx}})\right\}. \end{cases} \tag{4.1.24}$$

At the end of slot k, neglecting outage and overflow and the floor and ceiling operations that should be considered in the battery update formulas in the discrete model, the energy levels of the two devices are: $b_k^{\mathrm{tx}} - q^{\mathrm{tx}}(P_k) + e_k^{\mathrm{tx,EEH}} + \eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} z_k^{\mathrm{rx}\to\mathrm{tx}}$ and $b_k^{\mathrm{rx}} - q^{\mathrm{rx}}(P_k) + e_k^{\mathrm{rx,EEH}} - z_k^{\mathrm{rx}\to\mathrm{tx}}$. We impose that at the beginning of the next slot these two quantities be equal. Note that $e_k^{\mathrm{tx,EEH}}$ and $e_k^{\mathrm{rx,EEH}}$ are not known a priori,[7] thus we neglect them as well (it is possible to include only the *means* of the energy arrivals, but we verified that this refinement would not provide any significant benefit). Also, since we need to specify both P_k and $z_k^{\mathrm{rx}\to\mathrm{tx}}$, we need an additional equation. We impose that one of the two batteries is emptied in every slot, and therefore choose P_k as the minimum between $q^{\mathrm{tx}^{-1}}(b_k^{\mathrm{tx}})$ and $q^{\mathrm{rx}^{-1}}(b_k^{\mathrm{rx}} - z_k^{\mathrm{rx}\to\mathrm{tx}})$.

Assume that an acceptable solution of (4.1.24) exists and name it $(\bar{P}, \bar{z}^{\mathrm{rx}\to\mathrm{tx}})$. Two cases have to be considered:

1. $q^{\mathrm{tx}^{-1}}(b_k^{\mathrm{tx}}) < q^{\mathrm{rx}^{-1}}(b_k^{\mathrm{rx}} - \bar{z}^{\mathrm{rx}\to\mathrm{tx}}) \Leftrightarrow \bar{P} = q^{\mathrm{tx}^{-1}}(b_k^{\mathrm{tx}})$. In this case, the first equation can be simplified and we find $\bar{z}^{\mathrm{rx}\to\mathrm{tx}} = \dfrac{b_k^{\mathrm{rx}} - q^{\mathrm{rx}}(q^{\mathrm{tx}^{-1}}(b_k^{\mathrm{tx}}))}{\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} + 1}$.
2. $q^{\mathrm{tx}^{-1}}(b_k^{\mathrm{tx}}) \geq q^{\mathrm{rx}^{-1}}(b_k^{\mathrm{rx}} - \bar{z}^{\mathrm{rx}\to\mathrm{tx}}) \Leftrightarrow \bar{P} = q^{\mathrm{rx}^{-1}}(b_k^{\mathrm{rx}} - \bar{z}^{\mathrm{rx}\to\mathrm{tx}})$. In this case $\bar{P}$ and $\bar{z}^{\mathrm{rx}\to\mathrm{tx}}$ can be numerically found.

Also, it may happen that the system does not have acceptable solutions, i.e., $\bar{P}$ or $\bar{z}^{\mathrm{rx}\to\mathrm{tx}}$ is negative or exceeds the current battery levels. In this case we proceed as follows. First, we substitute the second equation into the first one. Then, we find the solution of the first equation, namely $\bar{z}^{\mathrm{rx}\to\mathrm{tx}}$, following the previous reasoning, i.e., considering the two possible cases. Finally, if $\bar{z}^{\mathrm{rx}\to\mathrm{tx}}$ is negative, we set $\bar{z}^{\mathrm{rx}\to\mathrm{tx}} = 0$. Instead, if $\bar{z}^{\mathrm{rx}\to\mathrm{tx}} > b_k^{\mathrm{rx}}$, we set $\bar{z}^{\mathrm{rx}\to\mathrm{tx}} = b_k^{\mathrm{rx}}$. $\bar{P}$ is then derived from the second equation.

[6] Note that, differently from $Q^{(i)}(\cdot)$, the function $q^{(i)}(\cdot)$ may not be bijective. In this context we define $q^{(i)^{-1}}(x) \triangleq \max_{P:q^{(i)}(P)=x} P$, i.e., $q^{(i)^{-1}}(x)$ is the greatest element of $\mathcal{P}$ that is mapped to x. This is a reasonable choice because, for all values of P such that $q^{(i)}(P) = x$, the energy consumption is the same but the reward $r(P)$ is different and, since $r(P)$ increases with P, we choose the greatest value in order to obtain the highest reward.

[7] It is possible to relax this hypothesis if the arrival process is predictable or partially predictable.

Once $(\bar{P}, \bar{z}^{\mathrm{rx}\to\mathrm{tx}})$ is specified, we extract the online policy as $P(\mathbf{b}) = \bar{P}$ and $z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b}) = \lfloor \bar{z}^{\mathrm{rx}\to\mathrm{tx}} \rfloor$. We used the *floor* operation in order to guarantee $q^{\mathrm{rx}}(P(\mathbf{b})) + z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b}) \leq b^{\mathrm{rx}}$ (with the *round* operation, the condition might not be satisfied).

The balanced policy, obtained according to the above procedure, is designed with the goal to balance the energy levels of the two devices.

Low Complexity Policy (LCP). The low complexity policy is defined as follows

$$\begin{aligned} P(\mathbf{b}) &= \min\left\{ {q^{\mathrm{tx}}}^{-1}(b^{\mathrm{tx}}), {q^{\mathrm{rx}}}^{-1}(b^{\mathrm{rx}}), {Q^{\mathrm{rx}}}^{-1}(\bar{C}^{\mathrm{rx}}(\xi^{\star})) \right\}, \\ z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b}) &= \min\{b^{\mathrm{rx}} - q^{\mathrm{rx}}(P(\mathbf{b})), Round\left(\bar{E}^{\mathrm{rx}}/\delta\right) - q^{\mathrm{rx}}(P(\mathbf{b}))\}, \end{aligned} \tag{4.1.25}$$

where $\bar{E}^{\mathrm{rx}}$ is the mean of the energy arrival process at the receiver. Consider the last terms of the two "min" operations. It can be seen that they are the discretized versions of Equations (4.1.14)-(4.1.15) (we applied the *round* operation in order to obtain an integer value). Note that the policy in (4.1.14)-(4.1.15) does not transmit when the batteries cannot support the use of a power ${Q^{\mathrm{rx}}}^{-1}(\bar{C}^{\mathrm{rx}}(\xi^{\star}))$, whereas in this case LCP would instead always use the maximum transmit power allowed by the status of the two batteries, which results in the full discharge of at least one of them. Although different from (4.1.14)-(4.1.15), LCP can achieve optimality in some cases, e.g., in the presence of deterministic arrivals.

LCP is a policy that, except for the "min" operators, does not depend upon the energy status. When the distribution has a small standard deviation, then it is expected that LCP provides good results and moreover, in the deterministic case, it degenerates to an optimal policy.

4.1.5 Offline Optimization

We now focus on offline optimization. One of the key aspects of this approach is that the energy arrival sequence is assumed to be known a priori (a statistical knowledge of the arrival process is not sufficient). Therefore, we restrict the study to the finite horizon case, considering separately the two cases of infinite and finite batteries. In this context, the aim is to find the Optimal Offline Policy (OP-OFF), namely $\mu^{\star}$, i.e., the sequence of actions $(\mathbf{P}, \mathbf{Z}^{\mathrm{rx}\to\mathrm{tx}})$ that maximize G_{μ}^{K} (Equation (4.1.5)). In Section 4.1.7 we will use OP-OFF as a benchmark for the online ones in the finite horizon case.[8]

4.1.5.1 Optimal Offline Policy with Infinite Batteries

We first set up the offline optimization problem (4.2.8) by clearly specifying the constraints that have to be satisfied and the optimization variables, in the case where the battery sizes are infinite. A formulation for the case with finite batteries will be given in Section 4.1.5.2.

In this case, the optimization problem in (4.2.8) can be explicitly written as follows (we start with empty batteries):

$$\mu = \underset{(\mathbf{P}, \mathbf{Z}^{\mathrm{rx}\to\mathrm{tx}})}{\min} \sum_{k=0}^{K-1} -r(P_k) \tag{4.1.26a}$$

[8] In this case, we simply apply to the finite horizon scenario the optimal online policy for infinite horizon derived in Section 4.1.4.

subject to:

$$Q^{\text{tx}}(P_k) \le B_k^{\text{tx}}, \qquad k = 0, \dots, K-1, \tag{4.1.26b}$$

$$Q^{\text{rx}}(P_k) + Z_k^{\text{rx}\to\text{tx}} \le B_k^{\text{rx}}, \qquad k = 0, \dots, K-1, \tag{4.1.26c}$$

$$P_k \ge 0, \qquad Z_k^{\text{rx}\to\text{tx}} \ge 0, \qquad k = 0, \dots, K-1, \tag{4.1.26d}$$

$$B_{k+1}^{\text{tx}} = B_k^{\text{tx}} - Q^{\text{tx}}(P_k) + E_k^{\text{tx,EEH}} + E_k^{\text{tx,WET}}, \qquad k = 0, \dots, K-2, \tag{4.1.26e}$$

$$B_{k+1}^{\text{rx}} = B_k^{\text{rx}} - Q^{\text{rx}}(P_k) + E_k^{\text{rx,EEH}} - Z_k^{\text{rx}\to\text{tx}}, \qquad k = 0, \dots, K-2, \tag{4.1.26f}$$

$$E_1^{\text{tx}} = E_1^{\text{rx}} = 0. \tag{4.1.26g}$$

Note that the battery evolutions include neither "min" operations (because the batteries are infinite) nor "max" operations (thanks to (4.1.26b) and (4.1.26c)). We recall that the energy harvested in slot k can be exploited only in a later slot and similarly for the energy transferred with WET ($E_k^{\text{tx,WET}} = \eta_{\text{WET}}^{\text{rx}\to\text{tx}} Z_k^{\text{rx}\to\text{tx}}$).

Lemma 4.1.12. $\left\{\mu = (\mathbf{P}, \mathbf{Z}^{\text{rx}\to\text{tx}}) : (4.1.26\text{b}) - (4.1.26\text{d}) \textit{ are satisfied}\right\}$ *is a convex set.*

Proof. The set is convex if $Q^{\text{tx}}(P_k) - B_k^{\text{tx}}$, $Q^{\text{rx}}(P_k) + Z_k^{\text{rx}\to\text{tx}} - B_k^{\text{rx}}$, $-P_k$ and $-Z_k^{\text{rx}\to\text{tx}}$ are concave function of $(P_k, Z_k^{\text{rx}\to\text{tx}})$ for every $k = 0, \dots, K-1$. These conditions are satisfied because $Q^{(i)}(P_k)$ are defined as concave functions and the other constraints are linear. ■

Since the reward function is convex (sum of convex functions) and Lemma 4.1.12 holds, (4.1.26) is a convex problem and can be solved using standard optimization techniques.

4.1.5.2 Optimal Offline Policy with Finite Batteries

When the battery sizes are finite, the optimization problem is the same of Equations (4.1.26a)-(4.1.26d), with the battery update formulas (4.1.26e)-(4.1.26f) replaced by

$$B_{k+1}^{\text{tx}} = \min\{B_{\max}^{\text{tx}}, B_k^{\text{tx}} - Q^{\text{tx}}(P_k) + E_k^{\text{tx,EEH}} + E_k^{\text{tx,WET}}\}, \tag{4.1.27}$$

$$B_{k+1}^{\text{rx}} = \min\{B_{\max}^{\text{rx}}, B_k^{\text{rx}} - Q^{\text{rx}}(P_k) + E_k^{\text{rx,EEH}} - Z_k^{\text{rx}\to\text{tx}}\}. \tag{4.1.28}$$

The problem can be formulated in a *standard form* (convex function to minimize plus inequality and equality constraints) by adding an inequality constraint for every possible condition imposed by the "min" operations. For example, for the receiver, the first four inequalities that have to be satisfied are (set $Q_k^{(i)} \triangleq Q^{(i)}(P_k)$ for ease of notation)

$$Q_0^{\text{rx}} + Z_0^{\text{rx}\to\text{tx}} \le 0, \tag{4.1.29a}$$

$$Q_1^{\text{rx}} + Z_1^{\text{rx}\to\text{tx}} \le \begin{cases} B_{\max}^{\text{rx}}, \\ B_0^{\text{rx}} - Q_0^{\text{rx}} - Z_0^{\text{rx}\to\text{tx}}, \end{cases} \tag{4.1.29b}$$

$$Q_2^{\text{rx}} + Z_2^{\text{rx}\to\text{tx}} \le \begin{cases} B_{\max}^{\text{rx}}, \\ B_{\max}^{\text{rx}} + B_1^{\text{rx}} - Q_1^{\text{rx}} - Z_1^{\text{rx}\to\text{tx}}, \\ B_0^{\text{rx}} - Q_0^{\text{rx}} - Z_0^{\text{rx}\to\text{tx}} + B_1^{\text{rx}} - Q_1^{\text{rx}} - Z_1^{\text{rx}\to\text{tx}}, \end{cases} \tag{4.1.29c}$$

$$Q_3^{\text{rx}} + Z_3^{\text{rx}\to\text{tx}} \le \begin{cases} B_{\max}^{\text{rx}}, \\ B_{\max}^{\text{rx}} + B_2^{\text{rx}} - Q_2^{\text{rx}} - Z_2^{\text{rx}\to\text{tx}}, \\ B_{\max}^{\text{rx}} + B_1^{\text{rx}} - Q_1^{\text{rx}} - Z_1^{\text{rx}\to\text{tx}} + B_2^{\text{rx}} - Q_2^{\text{rx}} - Z_2^{\text{rx}\to\text{tx}}, \\ B_0^{\text{rx}} - Q_0^{\text{rx}} - Z_0^{\text{rx}\to\text{tx}} + B_1^{\text{rx}} - Q_1^{\text{rx}} - Z_1^{\text{rx}\to\text{tx}} + B_2^{\text{rx}} - Q_2^{\text{rx}} - Z_2^{\text{rx}\to\text{tx}}, \end{cases} \tag{4.1.29d}$$

and similar constraints have to be considered for the transmitter. The total number of constraints scales as K^2.

The general expressions for the transmitter and receiver constraints can be written in compact form as ($\ell = 0, \dots, k-1$ and $k = 0, \dots, K-1$)

$$\sum_{j=\ell}^{k-1} Q_j^{\mathrm{tx}} - \sum_{j=\ell}^{k-2} \eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} Z_j^{\mathrm{rx}\to\mathrm{tx}} \leq B_{\max}^{\mathrm{tx}} \chi\{\ell > 0\} + \sum_{j=\ell}^{k-2} B_j^{\mathrm{tx}}, \tag{4.1.30}$$

$$\sum_{j=\ell}^{k-1} (Q_j^{\mathrm{rx}} + Z_j^{\mathrm{rx}\to\mathrm{tx}}) \leq B_{\max}^{\mathrm{rx}} \chi\{\ell > 0\} + \sum_{j=\ell}^{k-2} B_j^{\mathrm{rx}}, \tag{4.1.31}$$

where $\chi\{\cdot\}$ is the indicator function. The four cases in (4.1.29) can be obtained from (4.1.31) for $k = 0, 1, 2, 3$ (note that there are k constraints in each case, obtained for $\ell = 0, \dots, k-1$). For example, when $\ell = 0$ or $\ell = k-1$, the last and the first lines of (4.1.29) are obtained, respectively.

In practice, techniques such as the interior-point algorithm or a Sequential Quadratic Programming (SQP) method can be used to find the optimal solution. However, if the time horizon is large, the computational time can be long. Moreover, to run the algorithms $E_k^{\mathrm{tx,EEH}}$ and $E_k^{\mathrm{rx,EEH}}$ must be known in advance. Thus, even if the offline optimization gives the policy with the highest reward among all, in practice it can rarely be used. On the other hand, finding the optimal offline policy is still useful, as it makes it possible to understand what are the limits of the energy transfer mechanism, and can be used as a benchmark for all other policies.

4.1.6 Numerical Results - Online Optimization

In this section we present some numerical results for the online policies. In order to understand their properties, here we consider some analytical examples in the infinite horizon case. In Section 4.1.7 we discuss how these policies can be applied to a realistic scenario, with finite horizon and real data.

In addition to studying the optimal policy OP-ON, we present the performance of suboptimal policies in several settings. We remark that, since we focus on the online case, all energies are expressed in terms of energy quanta.

We consider the long-term maximization of G_η (Equation (4.1.22) or, equivalently, (4.1.6)) when the reward function is the transmission rate $r(x) = \log(1 + \Lambda x)$, where Λ is a scaling factor. The term $\eta^\star$ represents the optimal policy obtained when WET is used, whereas $\eta_0^\star$ is the optimal policy without WET.

The numerical results strongly depend upon the system parameters, and on the structure of $r(\cdot)$ and $Q^{(i)}(\cdot)$. In the following we focus on a particular energy consumption model, but similar considerations can be made in other cases as well. Consider the following energy consumption functions ($\zeta > 0$) expressed in energy quanta (e.q.)

$$Q^{\mathrm{rx}}(P) = \begin{cases} \dfrac{\zeta + P_n}{P_n} P, & \text{if } P < P_n, \\ \zeta + P_n - \alpha^{\mathrm{rx}} \log(1 + \Lambda P_n) + \alpha^{\mathrm{rx}} \log(1 + \Lambda P), & \text{if } P \geq P_n \end{cases} \tag{4.1.32}$$

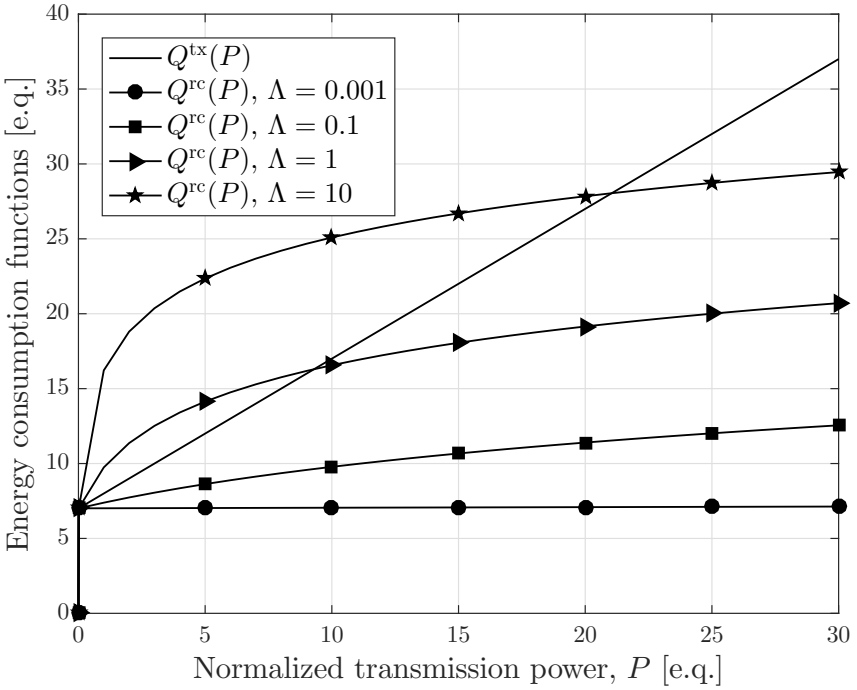

Figure 4.2. Energy consumptions $Q^{\mathrm{tx}}(P)$ and $Q^{\mathrm{rx}}(P)$ as a function of P for several values of Λ. The transmission power is expressed in terms of energy quanta because the slot length is fixed and thus we refer to power or energy interchangeably.

and $Q^{\mathrm{tx}}(P)$ is piece-wise linear as in Equation (4.1.18) with $P_n = 1/100$. ζ and α^{rx} are parameters that depend upon the considered technology. Both devices have a fixed energy cost ζ plus a linear or logarithmic curve.[9]

If not otherwise specified, we consider $b_{\max} \triangleq b_{\max}^{\mathrm{tx}} = b_{\max}^{\mathrm{rx}} = 30$ e.q., truncated geometric arrivals with parameters $\bar{E}^{\mathrm{tx}} = 2$ e.q., $b_{\max} = 5$ e.q. for TX, uniform energy arrivals with parameters $\bar{e}^{\mathrm{rx}} = 12.5$ e.q., $e_{\max}^{\mathrm{rx}} = 25$ e.q. for RX, $\zeta = 7$ e.q., $\alpha^{\mathrm{rx}} = 4$ e.q., $\Lambda = 0.1$, $\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} = 0.15$, a unit slot length and $P_{\max} = e_{\max}$ (in a slot, potentially, all the stored energy can be consumed).

In Figure 4.2, the bold curve represents the energy consumption $Q^{(i)}(\cdot)$ considered in this example. Note that in the online optimization we consider $q^{(i)}(\cdot) = \left\lceil Q^{(i)}(\cdot) \right\rceil$ as described in Section 4.1.4.

We define the following functions exploiting the technique introduced in Example 4.1.11

$$\Psi^{\mathrm{tx}}(P) = \begin{cases} \dfrac{1}{m}\log(1+\Lambda P), & \text{if } P < \bar{x} - \zeta, \\ \zeta + P, & \text{otherwise,} \end{cases} \qquad (4.1.33)$$
$$\Psi^{\mathrm{rx}}(P) = \frac{q^{\mathrm{rx}}(P_{\max})}{P_{\max}} P,$$

with $P_{\max} = b_{\max} - \zeta$, $\bar{x} = 20.99$ and $m = 0.0417$. It can be verified that these functions satisfy the hypotheses of Theorem 4.1.2 and the upper bounds are $G_{\mathrm{u.b.}}^{\mathrm{noET}} = 0.0834$ and $G_{\mathrm{u.b.}}^{\mathrm{WET}} = 0.1561$.

[9] We decided to focus on the case $\zeta^{\mathrm{tx}} = \zeta^{\mathrm{rx}}$ for presentation simplicity, but this assumption is not restrictive.

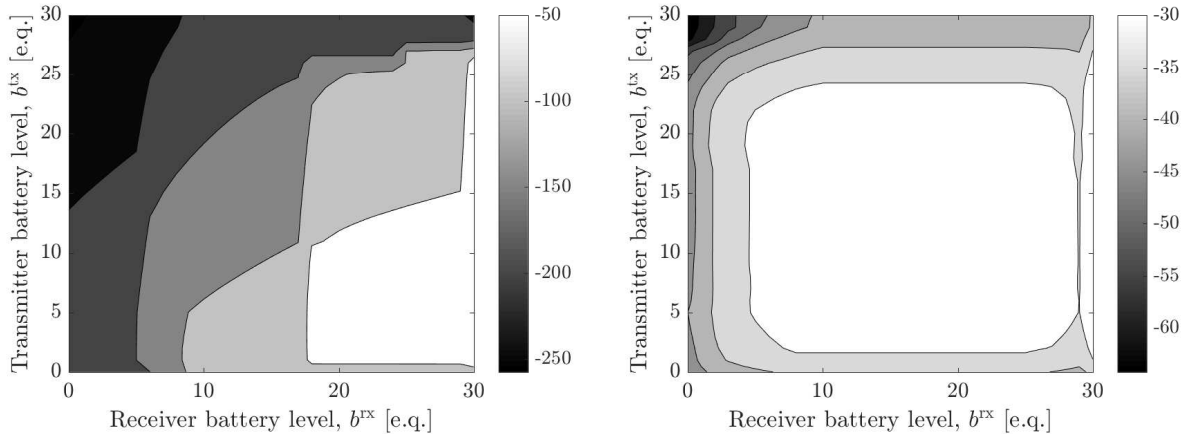

Figure 4.3. Steady-state probabilities ($10\log_{10}(\cdot)$ scale) without (left) and with (right) WET as a function of the batteries energy status b^{tx}, b^{rx}.

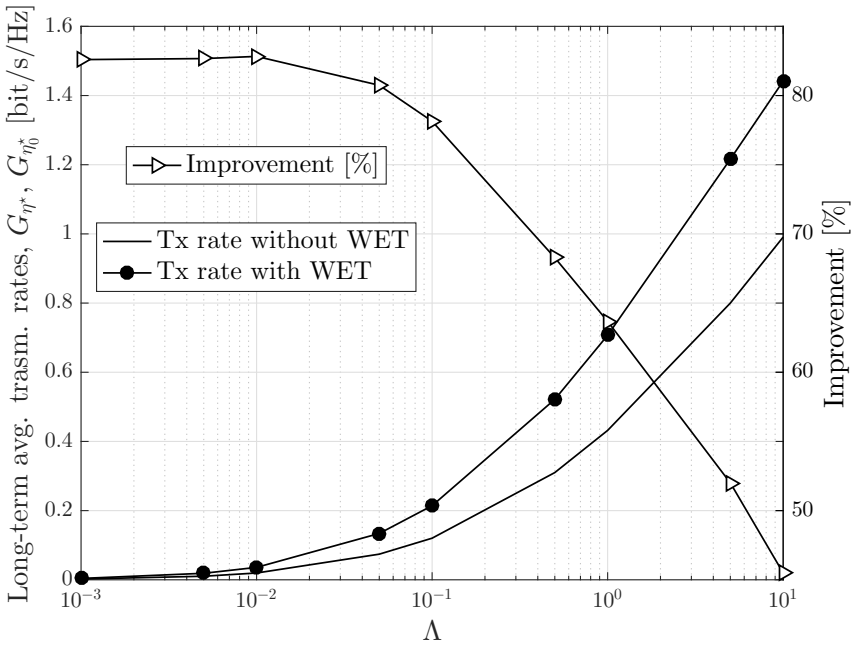

Figure 4.4. Long-term average transmission rates $G_{\eta_0^\star}$, $G_{\eta^\star}$ (optimal rewards) and corresponding improvement as a function of Λ when $b^{\text{tx}}_{\max} = b^{\text{rx}}_{\max} = 30$ e.q. and $\zeta = 7$ e.q..

In Figure 4.3 we show the steady-state distribution of the system state when using the optimal policies with and without WET. As expected, when energy transfer is not used, the energy levels are highly unbalanced and the receiver is almost always in overflow. With energy transfer, instead, the overflow probability becomes lower. In this case, even in the presence of a relatively low efficiency, $\eta_{\text{WET}}^{\text{rx}\to\text{tx}}$ (85% of the energy sent is wasted), energy transfer provides a reward improvement of 78%, see Figure 4.4. Note that the improvement is due to the fact that RX can send part of its energy to TX and this is particularly effective when RX receives more energy and/or consumes less energy than TX. A comparison with the upper bounds shows that $G_{\eta_0^\star} > 0.99\, G_{\text{u.b.}}^{\text{noET}}$ and $G_{\eta^\star} > 0.95\, G_{\text{u.b.}}^{\text{WET}}$. The reward without WET and its upper bound are very close (this happens because the batteries are large). Instead, with WET the distance from the upper bound is wider

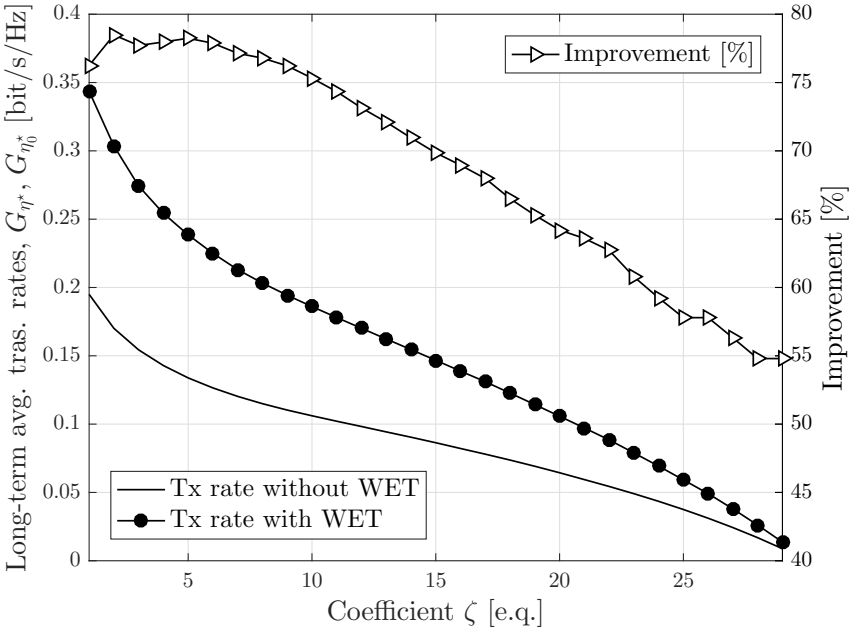

Figure 4.5. Long-term average transmission rates $G_{\eta_0^\star}$, $G_{\eta^\star}$ (optimal rewards) and corresponding improvement as a function of ζ when $b_{\max}^{\mathrm{tx}} = b_{\max}^{\mathrm{rx}} = 30$ e.q. and $\Lambda = 0.1$.

because the function $\Psi^{\mathrm{rx}}(\cdot)$ is distant from $Q^{\mathrm{rx}}(\cdot)$ and the batteries are not sufficiently large. When $\Lambda \in \{0.001, 1, 10\}$ the improvements provided by the use of WET become $\{83, 64, 45\}\%$, thus the performance is significantly increased in a wide range of values of Λ. This can be observed in Figure 4.4, where we plot the rewards with and without WET, along with the corresponding improvement, defined as $\left(G_{\eta^\star} - G_{\eta_0^\star}\right)/G_{\eta_0^\star}$. WET works better in the low SNR regime because $r(\cdot)$ tends to be linear, thus smart energy transmission techniques (e.g., delay a transmission in order to transmit with more power) do not improve the reward significantly.

Figure 4.5 shows how the two rewards (with and without energy transfer) change as a function of ζ. When ζ is very high, in both cases the value of the reward is very small in absolute terms (see Figure 4.5), but the use of energy transfer may provide a significant reward improvement in relative terms as pointed out by the improvement curve $(G_{\eta^\star} > 1.5\, G_{\eta_0^\star})$. Thus, it is better to use Energy Transfer even when ζ is high. Even if we present our results for $\zeta^{\mathrm{tx}} = \zeta^{\mathrm{rx}}$, similar results can be found in the general case. In particular, if either energy consumption $\zeta^{(i)}$ decreases, then the reward improvement and the reward itself increase (similarly to Figure 4.5) and vice-versa.

Also, in Figure 4.6 we plot the reward when $b_{\max}^{\mathrm{rx}} = 30$ e.q. is fixed and $b_{\max}^{\mathrm{tx}}$ changes (a similar curve can be obtained switching $b_{\max}^{\mathrm{tx}}$ and $b_{\max}^{\mathrm{rx}}$). The WET improvement increases with the battery size. The abscissa values start from 7 e.q. since, for $b_{\max}^{\mathrm{tx}} \leq 7$ e.q., the reward is zero because of the circuitry costs.

As an additional interesting example, consider the case $\zeta = 0$ e.q., where $q^{\mathrm{tx}}(\cdot) = \Psi^{\mathrm{tx}}(\cdot)$ and $q^{\mathrm{rx}}(\cdot) = \Psi^{\mathrm{rx}}(\cdot)$. The energy consumption functions are

$$q^{\mathrm{tx}}(P) = P, \qquad q^{\mathrm{rx}}(P) = 4\log(1 + \Lambda P). \tag{4.1.34}$$

In this case $\Psi^{(i)}(\cdot) = q^{(i)}(\cdot)$. The distances from $G_{\mathrm{u.b.}}^{\mathrm{noET}}$ and $G_{\mathrm{u.b.}}^{\mathrm{WET}}$ are 0.25% and 3.3%, respectively. For larger batteries the upper bound gaps are even smaller. We also computed

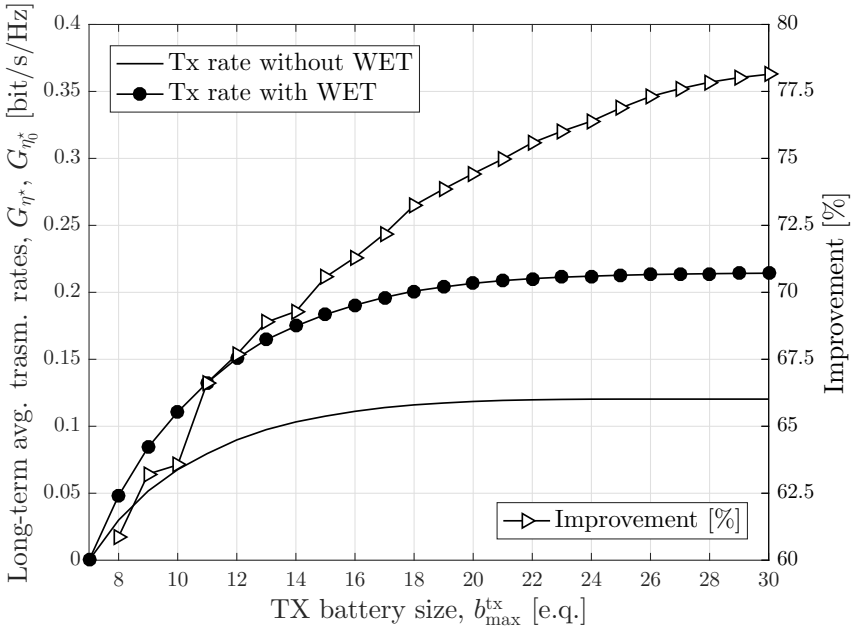

Figure 4.6. Long-term average transmission rates $G_{\eta_0^\star}$, $G_{\eta^\star}$ (optimal rewards) and corresponding improvement as a function of $b_{\max}^{\text{tx}}$ when $b_{\max}^{\text{rx}} = 30$ e.q., $\zeta = 7$ e.q. and $\Lambda = 0.1$.

the rewards of policies GP, BP and LCP and we found $G_{\text{GP}} = 0.88\, G_{\eta^\star}$, $G_{\text{BP}} = 0.88\, G_{\eta^\star}$, $G_{\text{LCP}} = 0.82\, G_{\eta^\star}$, i.e., in this particular case, the simpler policies provide almost as good a performance as OP, while being significantly faster to compute.

4.1.7 Real Data Analysis

In this section we want to apply the policies found so far to some realistic examples. Since in reality only a finite sequence of energy arrivals can be available, we focus on the optimization of G_μ^K (Equation (4.1.5)). If we assume that the energy arrivals are known a priori, the offline optimal policy (OP-OFF) provides the best reward among all. Instead, to compute the online policies, only the statistics of the energy arrivals is required. In this section, in addition to discussing the benefits of WET, we compare the offline and online approaches. As in Section 4.1.5, we consider separately the cases of infinite and finite batteries.

4.1.7.1 Infinite Batteries

Consider a scenario with two devices in two different rooms of a building, where energy harvesting is based on indoor light.

At enhants.ee.columbia.edu, a collection of light energy data traces is available.[10] The authors took measurements of the irradiance in different indoor rooms during an extended period of time. We use part of this data in our performance evaluation.

We assume that TX is located on a bookshelf in an office (Setup A) and the receiver in another office (Setup B). The receiver, generally, harvests more energy than the transmitter because it gets more sunshine. We show in Figure 4.7 the irradiance arrivals for the two devices (measured on 09 January 2010). It can be seen that, in this case, RX receives significantly more energy than the transmitter, therefore it may be interesting to use

[10] These data were discussed in [43] by Gorlatova *et al.*.

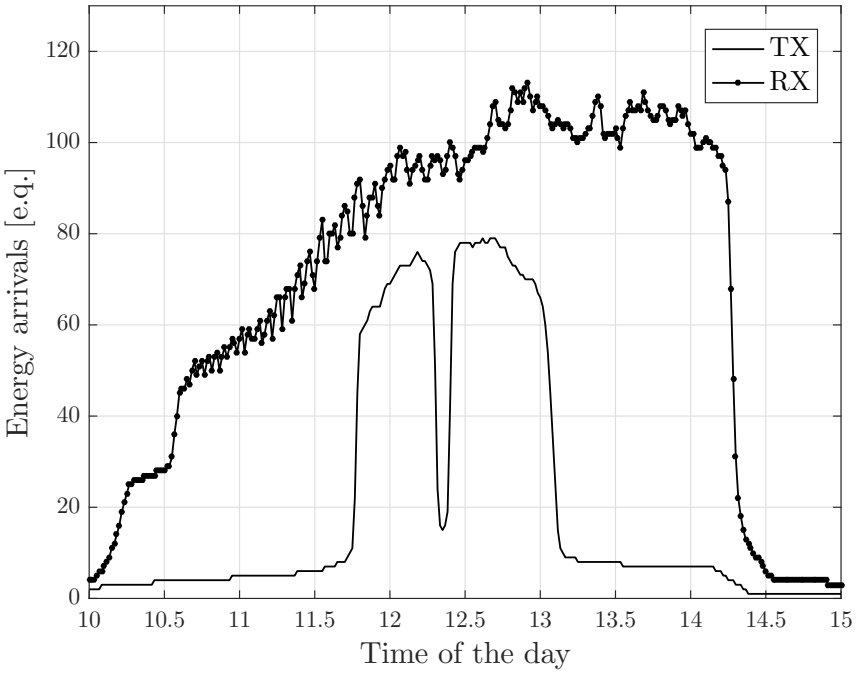

Figure 4.7. Indoor light energy arrivals as a function of the time of the day.

energy transfer to try to balance the system. In this setup, the harvested power is at most 113 μW/cm^2, i.e., very low. In an indoor environment, an ultra low power sensor network should be deployed, otherwise the energy costs would be too high to be sustained by the renewable energy source. Therefore, we assume that the transmitter can choose its transmit power to be even lower than 1 mW. In this case, it can be verified that the effects of a finite battery can be neglected (even if a very small battery is used, e.g., 0.16 J [43]), thus in this section we can consider infinite batteries with no loss of generality.

Time is divided in slots of 60 s each, and in every slot a new $(P_k, Z_k^{\text{rx}\to\text{tx}})$ is chosen. The maximum energy that can arrive in 60 s is 60 s $\times$ 113 μW/cm^2 $\times$ S cm^2 where S is the solar panel size (assumed equal for the two devices). We compute the reward using $r(x) = \log(1 + \Lambda x)$ in a low SNR regime ($\Lambda = 0.002$). In order to highlight the system behavior, we present the results for $Q^{\text{tx}}(P) = Q^{\text{rx}}(P) = P$. The model can be extended, e.g., using the energy consumption model of Equation (4.1.34), which would result in an even better improvement because RX would consume less energy.

We use two approaches to apply WET to the system: 1) online low complexity balanced policy (BP), which is very easy to compute and can be used in practice, and 2) offline optimal policy (OP-OFF) (presented in Section 4.1.5.1). We selected 1 μW/cm^2 as the minimum non negligible power that can be harvested. In this case, one energy quantum corresponds to the minimum energy that can arrive in 60 s, i.e.,

$$1 \text{ e.q.} \equiv 60 \text{ s} \times 1 \ \mu\text{W/cm}^2 \times S \text{ cm}^2. \tag{4.1.35}$$

Figure 4.8 shows the sent data and energy (expressed in energy quanta) for BP and OP-OFF. In Figure 4.9, the corresponding energy evolutions are presented.

BP is designed in order to balance the energy of the two devices. Indeed, when the transmitter battery is low, $Z_k^{\text{rx}\to\text{tx}}$ (transfer energy from RX to TX) is high, i.e., WET is better exploited when the difference between the energy arrivals is high. Analytically, it

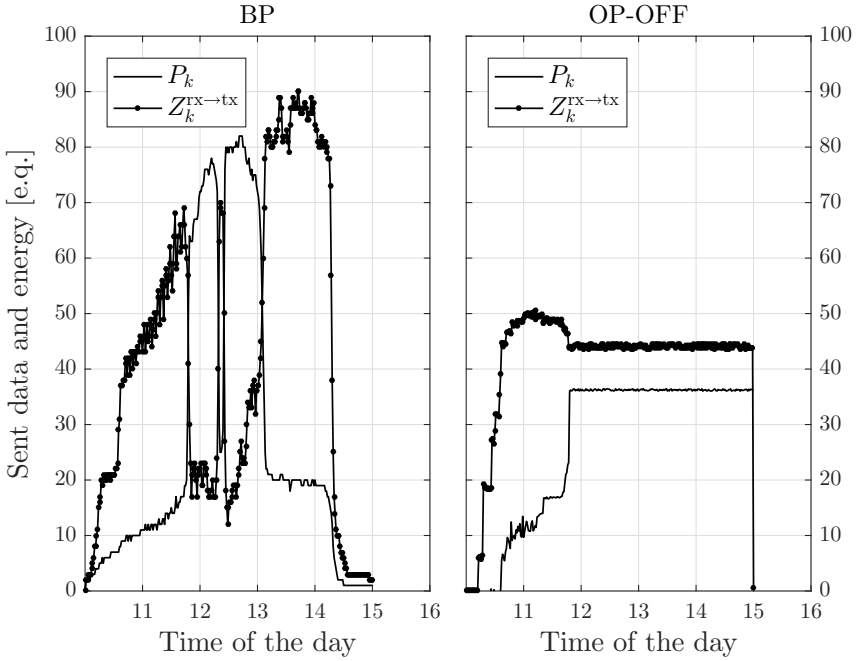

Figure 4.8. Policies BP (left) and OP-OFF (right) as a function of the time of the day.

can be verified that in the linear energy consumption case, BP degenerates in the following policy:

$$z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b}) = \left(\left\lfloor \frac{b^{\mathrm{rx}} - b^{\mathrm{tx}}}{1+\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}}} \right\rfloor \right)^{+}, \tag{4.1.36}$$

$$P(\mathbf{b}) = \min\{b^{\mathrm{tx}}, b^{\mathrm{rx}} - z^{\mathrm{rx}\to\mathrm{tx}}(\mathbf{b})\}, \tag{4.1.37}$$

where $(\cdot)^+ \triangleq \max\{\cdot, 0\}$.

On the left side of Figure 4.9 we depicted B_k^{tx} and $B_k^{\mathrm{rx}} - Z_k^{\mathrm{rx}\to\mathrm{tx}}$ in order to compare the two arguments of Equation (4.1.37). It can be seen that B_k^{tx} is always lower than $B_k^{\mathrm{rx}} - Z_k^{\mathrm{rx}\to\mathrm{tx}}$, thus Equation (4.1.37) becomes $P(\mathbf{b}) = b^{\mathrm{tx}}$ (indeed the curves of P_k and B_k^{tx} are the same), i.e., the transmitter battery is emptied in every slot. Moreover, note that $B_{k+1}^{\mathrm{tx}} = B_k^{\mathrm{tx}} + \lfloor \eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} Z_k^{\mathrm{rx}\to\mathrm{tx}} \rfloor$, i.e., the status of the transmitter battery is similar to E_k^{tx}, but higher (thanks to energy transfer).

Instead, OP-OFF chooses the initial values of P_k and $Z_k^{\mathrm{rx}\to\mathrm{tx}}$ in order to reach a situation where P_k and $Z_k^{\mathrm{rx}\to\mathrm{tx}}$ can be kept constant. This is possible because we consider infinite batteries. The resulting battery trends are represented on the right side of Figure 4.9. Note that P_k and $Z_k^{\mathrm{rx}\to\mathrm{tx}}$ were chosen in order to have zero energy stored in the last plus one slot, i.e., all the available energy is exploited in the finite horizon of K slots. Differently from the previous case, B_k^{tx} is greater than B_k^{rx} in the central region because TX receives a lot of energy and RX transfers its energy to TX.

Note that, if WET is not employed, an upper bound for the performance is given by the minimum between the means of $\{E_k^{\mathrm{tx}}\}$ and $\{E_k^{\mathrm{rx}}\}$, whereas, if WET is used, the upper bound is given by Equation (4.1.12).[11]

[11] Theorems 4.1.2 and 4.1.5 can be reformulated using the temporal means in this case.

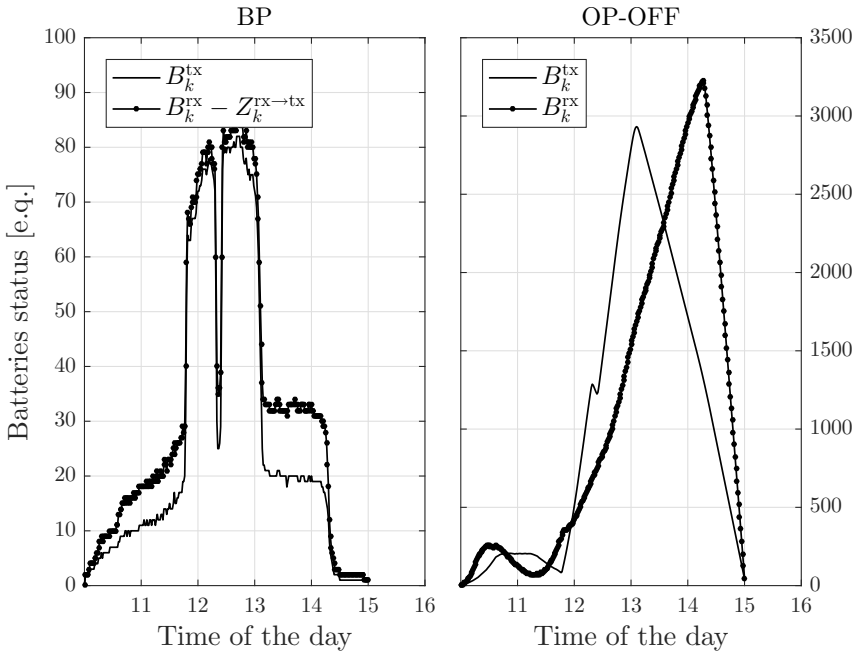

Figure 4.9. Battery energy status of BP (left) and OP-OFF (right) as a function of the time of the day.

BP gives a reward equal to 0.0512, whereas $G_{\mu^\star} = 0.0528$ (optimal offline reward with WET) and $G_{\mu_0^\star} = 0.0411$ (optimal offline reward without WET). The upper bound with and without WET are $G_{\text{u.b.}}^{\text{WET}} = 0.0532$ and $G_{\text{u.b.}}^{\text{noET}} = 0.0414$. Note that $G_{\mu^\star} = 0.99\, G_{\text{u.b.}}^{\text{WET}}$ and $G_{\mu_0^\star} = 0.99\, G_{\text{u.b.}}^{\text{noET}}$, i.e., OP-OFF is very close to but does not achieve the upper bounds even if the batteries are infinite and this is because we consider a finite time horizon. The reward improvement due to WET is 28%. Note that, even though BP is a suboptimal policy (much simpler to compute than OP-OFF) and only has a causal knowledge of the energy arrivals, its reward G_{BP} is very close to that of the optimal offline policy, $G_{\mu^\star}$.

4.1.7.2 Finite Battery Effects

In the previous section we assumed infinite batteries, which is legitimate in the indoor environment we considered. However, when the solar panel is powered with direct sunlight, it is likely that an inappropriate use of the energy may lead to battery overflow. At [101], a collection of solar light measurements in several locations over the past years is available and in Figure 4.10 we show the irradiance measured in Elizabeth City on 20 July 2014. The continuous lines represent all the measured data. We performed a sampling and considered only the points depicted with squares and circles. This is in order to perform the offline optimization in a reasonable computational time (we recall that with finite batteries the number of constraints grows quadratically with the number of samples). We considered the same energy arrival profile for both transmitter and receiver, but we assumed that the transmitter has a solar panel three times smaller than RX (in reality, the two devices could also receive different solar energy because of their position). We scaled the irradiance data in order to apply an MDP approach to solve the problem: the histograms of the two energy arrival profiles were assumed as empirical pdfs of the two arrival processes and we found OP-ON according to the model of Section 4.1.4. Since this

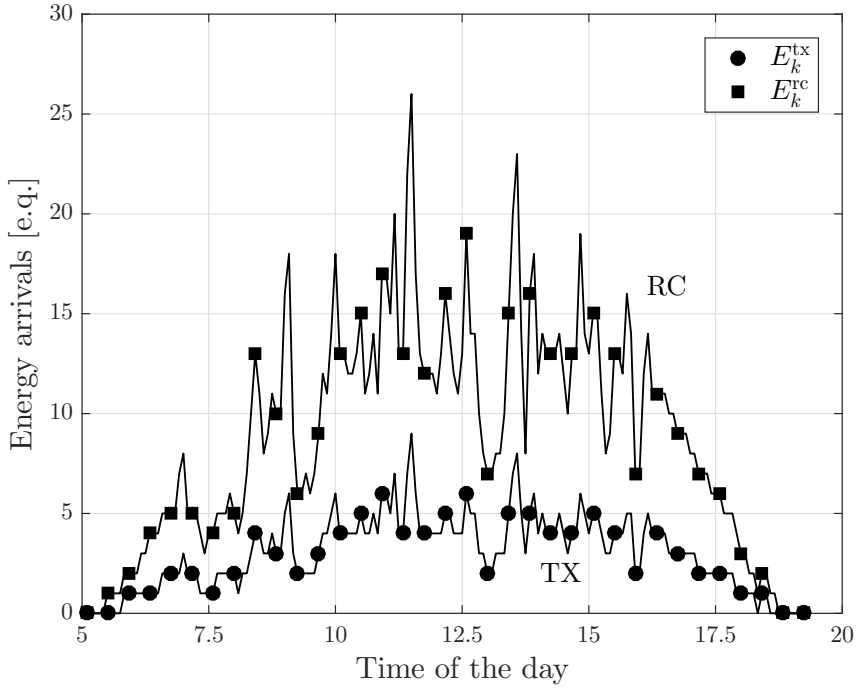

Figure 4.10. Solar energy arrivals as a function of the time of the day.

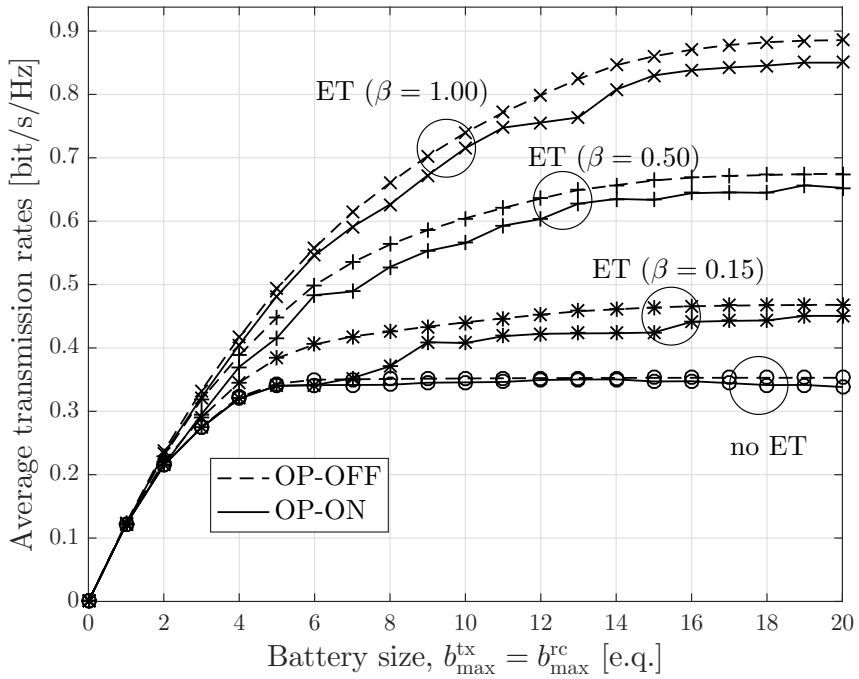

Figure 4.11. Optimal offline and online rewards as a function of $b_{\max}^{\mathrm{tx}} = b_{\max}^{\mathrm{rx}}$ when $\Lambda = 0.1$ and $\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} \in \{0.15, 0.50, 1.00\}$.

approach is suboptimal because it assumes i.i.d. energy arrivals, we compared it with OP-OFF, that gives the best possible results.

Figure 4.11 shows the (simulated) rewards with and without WET as a function of the battery sizes. We considered the model of Equation (4.1.34) with $\Lambda = 0.1$, $b_{\max} = b_{\max}^{\mathrm{tx}} = b_{\max}^{\mathrm{rx}}$ and $\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} \in \{0.15, 0.50, 1.00\}$. When $b_{\max}$ is low, even when $\eta_{\mathrm{WET}}^{\mathrm{rx}\to\mathrm{tx}} = 1$, WET does not improve the system reward. This is because the energy harvesting mechanism manages

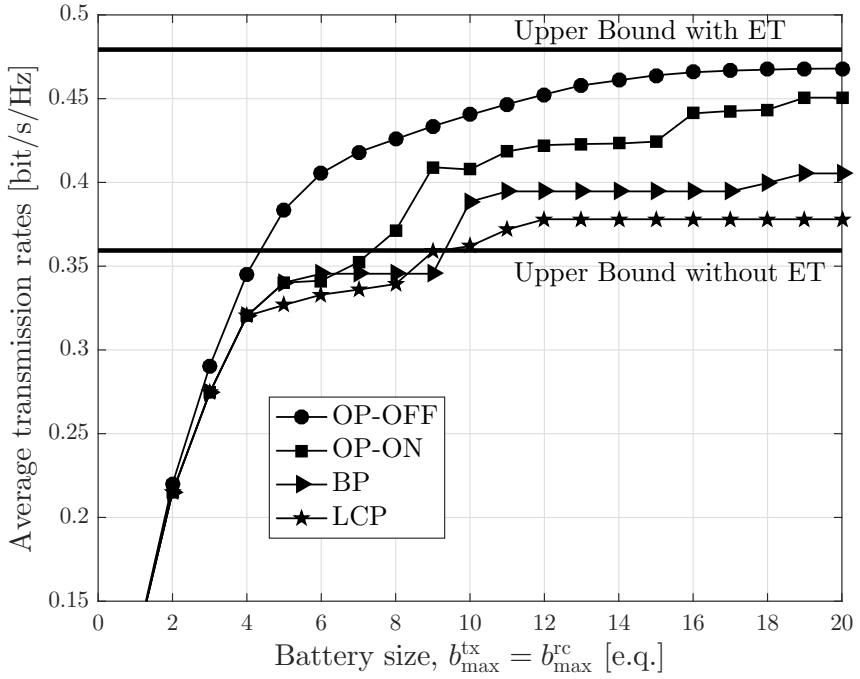

Figure 4.12. Rewards as a function of $b_{\max}^{\text{tx}} = b_{\max}^{\text{rx}}$ when $\Lambda = 0.1$ and $\eta_{\text{WET}}^{\text{rx}\to\text{tx}} = 0.15$ for several policies.

to fill up both batteries almost all the time, thus it is not necessary to exchange energy. Instead, when the size of the batteries grows, WET may significantly improve the reward: when $b_{\max} = 5$ e.q. the ratio $G_{\mu^\star}/G_{\mu_0^\star}$ for $\eta_{\text{WET}}^{\text{rx}\to\text{tx}} \in \{0.15, 0.50, 1.00\}$ is $\{1.12, 1.30, 1.44\}$, and becomes $\{1.33, 1.91, 2.51\}$ when $b_{\max} = 20$ e.q..

Beyond a certain value of $b_{\max}$, the rewards can be observed to saturate, thus it is not necessary to use very large batteries to achieve high rewards. This is because the effects of outage and overflow become negligible. Note that, because of the transmitter energy arrivals, without WET the system reward saturates very soon, whereas, with energy transfer, the saturation value is only reached for higher $b_{\max}$. Note that for $\eta_{\text{WET}}^{\text{rx}\to\text{tx}} = 0.15$ and $b_{\max} \leq 7$ e.q., $G_{\eta^\star}$ is low and this is due to the discretization ($\lfloor \eta_{\text{WET}}^{\text{rx}\to\text{tx}} b_{\max} \rfloor = 0$ e.q. for $b_{\max} < 7$ e.q.).

In this example OP-ON and OP-OFF are very close, which makes online policies very good candidates for application in real scenarios, because they are easier to implement while being almost optimal.

Finally, in Figure 4.12 we plot OP-OFF, OP-ON, the suboptimal online policies BP and LCP and the upper bounds. Note that with the online policies OP-ON may be lower than BP (at $b_{\max} = 6$ e.q. for example). This is because OP-ON is optimal in the *long-run*, thus in a particular realization it may turn out to be suboptimal. OP-OFF increases with $b_{\max}$ and almost reaches the upper bound (which is not achieved because the simulation time is finite). The balanced policy is generally better than the low complexity policy, because BP operates with the energy levels (see Equation (4.1.24)), whereas LCP operates with the average energy arrival statistics (see Equation (4.1.25)).

4.1.8 Conclusions of Section 4.1

In the first part of this chapter, we used the combination of environmental energy harvesting and Wireless Energy Transfer (WET) to improve the network performance. We studied a

scenario composed of two energy harvesting devices, a transmitter and its receiver, that can exchange energy through a WET interface. We considered two generic energy consumption functions and found performance upper bounds with and without WET, showing that, under some assumptions, they are achievable. Then we studied the online and offline optimization problems. In the first case we modeled the system with an MDP, studying numerically the optimal online policy and introducing some low complexity policies. For the offline optimization we set up the optimization problem and showed that it is convex. In our numerical evaluations we derived the optimal transmission policies, showing that WET can significantly improve the system performance and discussing how the system behaves as a function of the system parameters. For example, we noticed that the reward improvement increases with the battery sizes and remains high even for large values of the circuitry cost. Also, we analyzed two realistic examples of indoor and outdoor light radiation, showing the effects of finite batteries on the transmission strategies.

Possible extensions of our work are the exploitation of the predictability and correlation of the transmitter and receiver energy sources, and consideration of battery imperfections.

4.2 Wireless Powered Communication Networks

4.2.1 Introduction

New generation devices, e.g., in Wireless Sensor Networks (WSNs) or mobile cellular networks, are able to provide high communication performance in terms of throughput or delay at the cost of computational complexity and demanding power supplies. Wireless Energy Transfer (WET) has been recognized as one of the most appealing solutions for supplying energy to mobile devices when their batteries cannot be easily or cheaply replaced. Via WET it becomes possible to greatly extend the network lifetime and improve the devices performance by avoiding energy outage situations. Nowadays, it is possible to transfer powers of tens to hundreds of microwatts at distances of up to 10 m (see, for example, the Powercast company products [121]) and thus it becomes possible to supply ultra-low power mid-range networks. Differently from standard Environmental Energy Harvesting (EEH) techniques, WET has the major advantage of being fully controlled, as it does not rely on an external random phenomenon.

While most literature on WET has focused on the slot-oriented case, in which all the harvested energy is immediately used [58], in this thesis we consider the battery-powered case, in which the harvested energy can be stored and used at a later time. In this case, new considerations can be made (channel conditions, current battery level, battery size, future energy arrivals, etc.) and studying the system performance becomes more involved. The goal of the present work is to investigate such a problem.

We focus on Radio-Frequency Wireless Energy Transfer (RF-WET), since it is a far-field WET technique, and thus is suitable for powering several devices simultaneously in a distributed area. The RF signal can be harvested from the environment (e.g., this may be possible in a city where several electromagnetic sources are available), or from a dedicated source, i.e., a particular node (generally the access point) which emits RF signals to feed the devices (commercial products for RF-WET transmission/reception are already available, see [121]). This last kind of scenario is known as Wireless Powered Communication Network (WPCN).

In a WPCN where multiple devices harvest energy from the receiver or the access point and transmit data in uplink, a doubly near-far phenomenon is present: a user far away

from the receiver experiences, on average, a worse channel than the others both in uplink (therefore it has to use more energy to perform its transmission) and in downlink (thus it gathers less energy). This is substantially different from the classic energy harvesting scenarios [150], in which energy is gathered from the environment and thus does not decay over distance.

Contributions. In this section, we consider a WPCN composed of a Receiver (RX) and two distributed nodes. RX transfers energy in downlink to the nodes, which use the harvested energy for transmission purposes. Our system model is similar to that of [58, 80, 96] and [150, Section IV]. As in [96], we consider *battery-powered* devices and focus on the long-term performance. However, [96] considered only one device, whereas, in the present work, we consider the near-far effect problem when multiple devices are present. Moreover, differently from [96], we describe how to derive the optimal strategy to maximize the throughput of the system, whereas [96] focuses on the performance evaluation of a given strategy. We also take into account the imperfections of the devices in terms of circuitry costs and energy depletion over time (other sources of inefficiencies can be modeled as in Section 5.1 or [89, 149, 103]). [58, 80] describe a problem similar to what we analyze, but focus on the optimization in a single slot and *not* in the long term. This assumption turns out to be very restrictive in practice. Indeed, in our numerical evaluation we will describe the differences between these two approaches and show that focusing only on a greedy slot-oriented optimization is strongly suboptimal in the long run. We study the throughput maximization problem and solve it both optimally, via Markov Decision Process (MDP) theory [122], and approximately, exploiting the results we derived in the optimization section. Previous papers summarized in [150, Section IV] do not study the WPCN scenario and do not deal with the doubly near-far effect. Indeed, when the latter is considered, fairness between different devices should be explicitly addressed so as to avoid significant performance degradation. Moreover, most previous works did not focus on finding schemes for approximating the optimal reward. We explicitly study the trade-offs among battery size, amount of available energy, fading effects and performance. We show how fading and amount of dowlink energy are related and describe how the system changes when the power supply is scarce or abundant. This section can be considered as a first step to understand the key trade-offs and optimization problems in a WPCN with finite battery-powered devices.

Structure. The second part of this chapter is organized as follows. Section 4.2.2 defines the system model we analyzed and introduces the optimization problem, which is solved in Sections 4.2.3 and 4.2.4 (optimally and approximately, respectively). Some interesting extensions are outlined in Section 4.2.5. We briefly describe the slot-oriented maximization in Section 4.2.6. Section 4.2.7 presents our numerical results. Finally, Section 4.2.8 concludes the section.

References. This section is based on the conference papers **[C7]** and **[C11]** and on the journal paper **[J3]**.

4.2.2 System Model and Optimization Problem

The system is composed of three nodes: one Receiver (RX) with Wireless Energy Transfer (WET) capabilities and two devices, namely $\mathrm{D}^{(1)}$ and $\mathrm{D}^{(2)}$. Via a RF-WET mechanism, RX recharges the batteries (with finite capacities $B_{\max}^{(1)}$ J and $B_{\max}^{(2)}$ J) of the two devices.

It is assumed that RX has an unlimited amount of energy available. The devices use the energy transferred in downlink to upload data packets to the receiver.

An approach similar to the "harvest-then-transmit" protocol proposed in [58] is adopted to keep the devices operational. Under this scheme and according to Section 2.1.1, time is divided in slots of length T and slot k corresponds to the time interval $[kT, (k+1)T)$. Every slot is divided in two sequential phases:[12]

Uplink (UL). In the first phase, which lasts for $\tau^{(1)} + \tau^{(2)} \leq T$ seconds, the two devices transmit data to RX in a TDMA fashion using the energy stored in their batteries.

Downlink (DL). During the second $\tau^{\text{rx}} \leq T - \tau^{(1)} - \tau^{(2)}$ seconds, $\text{D}^{(1)}$ and $\text{D}^{(2)}$ harvest the energy transferred from the receiver and store it in their batteries.

RX is assumed to have multiple antennas and is able to perform *energy beamforming* in order to split the energy transferred to $\text{D}^{(1)}$ or $\text{D}^{(2)}$ during the DL phase, whereas $\text{D}^{(1)}$ and $\text{D}^{(2)}$ are assumed to be equipped with an omnidirectional antenna.

4.2.2.1 Uplink Phase

At the beginning of a slot, device $\text{D}^{(i)}$ ($i \in \{1, 2\}$) has $B^{(i)} \in [0, B^{(i)}_{\max}]$ J of energy stored. In a TDMA fashion, first device 1 and then device 2 occupy the channel to transmit data in the uplink for $\tau^{(1)}$ and $\tau^{(2)}$ seconds, respectively. The transmission powers $P^{(1)\to\text{rx}}$ and $P^{(2)\to\text{rx}}$ and the time allocations $\tau^{(1)}$ and $\tau^{(2)}$ can change dynamically in every slot and are the control variables of our optimization.

In the uplink phase, device $\text{D}^{(i)}$ is constrained to consume an amount of energy $Q^{(i)}$, with $Q^{(i)} \leq B^{(i)}$, decomposed as

$$Q^{(i)} \triangleq Q^{(i)}(\tau^{(i)}, P^{(i)\to\text{rx}}) = \zeta^{(i)} + \tau^{(i)} \times P^{(i)\to\text{rx}}, \tag{4.2.1}$$

where $\tau^{(i)} \times P^{(i)\to\text{rx}}$ is the energy used for the transmission and $\zeta^{(i)}$ represents a fixed energy cost consumed every time a transmission is performed. We also impose $P^{(i)\to\text{rx}} \in [P^{(i)\to\text{rx}}_{\min}, P^{(i)\to\text{rx}}_{\max}]$ when a transmission is performed. We assume that, in every slot, the devices always have enough data to transmit, i.e., the transmission data queue is always non-empty. This assumption is useful to characterize the maximum throughput of the system.

According to Shannon's formula, when a power $P^{(i)\to\text{rx}}$ is used, the noise power is σ_0^2 and the uplink channel gain is $\hbar^{(i)}$, the maximum reliable transmission rate of device $\text{D}^{(i)}$ is

$$r(P^{(i)\to\text{rx}}, \hbar^{(i)}) \propto \log\left(1 + \frac{\hbar^{(i)} P^{(i)\to\text{rx}}}{\sigma_0^2}\right). \tag{4.2.2}$$

Thus, during a single slot, the amount of transmitted data can be approximated as the time reserved for device $\text{D}^{(i)}$ multiplied by the transmission rate (i.e., $\tau^{(i)} r(P^{(i)\to\text{rx}}, \hbar^{(i)})$).

The uplink channel is affected by flat fading, which remains constant within every slot but may change from slot to slot. The channel gain $\hbar^{(i)}$ can be expressed as $\hbar^{(i)} = \tilde{\hbar}^{(i)}\theta^{(i)}$, where $\theta^{(i)}$ is a random variable which represents the fading and $\tilde{\hbar}^{(i)}$ is the average channel

[12] Unlike in [58], and without loss in generality, we chose to consider first the uplink and then the downlink phases in order to more easily track the energy level of the two devices when we set up the Markov Decision Process (MDP) formulation in Section 4.2.2.3.

gain, obtained by considering the path loss effects as

$$\tilde{h}^{(i)} = h_0^{(i)} \, {d^{(i)}}^{-\gamma^{(i)}}. \tag{4.2.3}$$

The term $h_0^{(i)}$ is the signal power gain at a reference distance of 1 m, $d^{(i)}$ is the distance between $\mathrm{D}^{(i)}$ and RX expressed in meters, and $\gamma^{(i)}$ is the path loss exponent.

4.2.2.2 Downlink Phase

The downlink period lasts for $\tau^{\mathrm{rx}} \leq T - \tau^{(1)} - \tau^{(2)}$ seconds. During this phase, the receiver transfers two energy beams (with powers $P^{\mathrm{rx}\to(1)}$ and $P^{\mathrm{rx}\to(2)}$) to the devices, under a maximum power constraint $P^{\mathrm{rx}\to(1)} + P^{\mathrm{rx}\to(2)} \leq P_{\max}^{\mathrm{rx}}$. The power received at $\mathrm{D}^{(i)}$ is $P^{\mathrm{rx}\to(i)}\eta_{\mathrm{ECL}}g^{(i)}$, where η_{ECL} is a constant in $(0, 1]$ that models the energy conversion losses at the devices and $g^{(i)}$ is a random variable related to the channel conditions, and on the specific beamforming technique (if any). The term $g^{(i)}$ can be explicitly written as $g^{(i)} = \tilde{g}^{(i)}\kappa^{(i)}$ in order to consider the flat fading effects, where $\tilde{g}^{(i)} = g_0^{(i)} {d^{(i)}}^{-\delta^{(i)}}$ and $\kappa^{(i)}$ are defined similarly to $\tilde{h}^{(i)}$ and $\theta^{(i)}$, respectively (see Section 4.2.2.1). In summary, when a power $P^{\mathrm{rx}\to(i)}$ is transferred to device i, the stored energy is

$$E^{(i),\mathrm{RF}} = \tau^{\mathrm{rx}} \, P^{\mathrm{rx}\to(i)} \, \eta_{\mathrm{ECL}} \, g_0^{(i)} \, {d^{(i)}}^{-\delta^{(i)}} \, \kappa^{(i)}. \tag{4.2.4}$$

The channel gain components in uplink $h^{(1)}$, $h^{(2)}$ and downlink $g^{(1)}$, $g^{(2)}$ can be assumed equal if the transmission is performed in the same frequency band, which is a common assumption in WPCNs [58]. Finally, note that the downlink channel of the user farther from RX is worse (on average), leading to a doubly near-far scenario.

4.2.2.3 Batteries

In every slot, the energy level of battery i is updated according to (see Equation (2.1.3))

$$B^{(i)} \leftarrow \min\{B_{\max}^{(i)}, [B^{(i)} - Q^{(i)} - Q_{\mathrm{d}}^{(i)}]^+ + E^{(i),\mathrm{RF}}\}, \tag{4.2.5}$$

where we defined $[\cdot]^+ \triangleq \max\{\cdot, 0\}$ to avoid negative energy levels. $Q_{\mathrm{d}}^{(i)}$ is a constant term that accounts for the energy depletion over time. Note that, when $Q_{\mathrm{d}}^{(i)} = 0$, the arguments of the "min" are always non-negative because the energy consumption $Q^{(i)}$ is chosen such that $Q^{(i)} \leq B^{(i)}$. The circuitry energy term is already enclosed in $Q^{(i)}$. We also highlight that $E^{(i),\mathrm{RF}}$ (defined in (4.2.4)) is a random variable because of the channel fading. The "min" operation is used to explicitly consider the effects of finite batteries. The battery evolution depends upon the choices of all parameters $\tau^{(1)}$, $\tau^{(2)}$, $P^{(1)\to\mathrm{rx}}$, $P^{(2)\to\mathrm{rx}}$, τ^{rx}, $P^{\mathrm{rx}\to(1)}$ and $P^{\mathrm{rx}\to(2)}$, which are the control variables of our optimization and will be analyzed in the next section.

In order to perform the optimization, we model the system with a discrete Markov Chain (MC). In particular, we discretize the battery of $\mathrm{D}^{(i)}$ in $b_{\max}^{(i)} + 1$ levels, where $b_{\max}^{(i)}$ represents the maximum amount of energy quanta (e.q.) that can be stored in the battery and one energy quantum corresponds to $B_{\max}^{(i)}/b_{\max}^{(i)}$ J (see Section 2.2.1). In our numerical evaluation, we always use a sufficiently high number of quantization levels. Equation (4.2.5) can be rewritten in terms of energy quanta: $b^{(i)} \leftarrow \min\{b_{\max}^{(i)}, [b^{(i)} - q^{(i)} - q_{\mathrm{d}}^{(i)}]^+ + e^{(i),\mathrm{RF}}\}$ (this is analogous to Equation (2.2.2)). In every slot, only an integer amount of energy

quanta $q^{(i)} = Q^{(i)} b^{(i)}_{\max}/B^{(i)}_{\max} \in \{0, \dots, b^{(i)}_{\max}\}$ can be extracted from the battery. Similarly, only an integer amount of energy quanta can be harvested, thus we define $e^{(i),\mathrm{RF}} = \lfloor E^{(i),\mathrm{RF}} b^{(i)}_{\max}/B^{(i)}_{\max} \rfloor$. Moreover, if the channel fading is described by a continuous random variable, we discretize it using a finite number of intervals.

In the rest of this section, the bold notation is used to identify a pair of values, e.g., $\mathbf{b} = \langle b^{(1)}, b^{(2)} \rangle$.

4.2.2.4 Optimization Problem

We define a *policy* μ as an action probability measure over the state set, namely $\mathcal{S}$. $\mathcal{S}$ represents all the combinations of battery levels $\mathbf{b}$ and channels $\boldsymbol{g}$, $\boldsymbol{\hbar}$. The policy is computed by a central controller (e.g., the receiver), which knows the state of the two batteries $\mathbf{b}$ and the joint channel state $(\boldsymbol{g}, \boldsymbol{\hbar})$, and distributed among nodes. We assume a perfect Channel State Information (CSI) in every slot (causal), and only a statistical knowledge for future slots, both at the devices and at RX.[13] Note that, while estimating the uplink channel is a standard task, downlink channel estimation may be more challenging due to the hardware limitations of the energy receivers. However, by exploiting innovative techniques, e.g., [161], it is possible to obtain accurate CSI for the downlink channel as well.

For every state $\mathbf{s} = (\mathbf{b}, \boldsymbol{g}, \boldsymbol{\hbar}) \in \mathcal{S}$, μ defines with which probability an action a is performed. Action a summarizes the data transmission duration $\boldsymbol{\tau} \triangleq \langle \tau^{(1)}, \tau^{(2)} \rangle$, the energy transfer duration τ^{rx}, the transmission powers $\mathbf{P}^{\to \mathrm{rx}} \triangleq \langle P^{(1)\to \mathrm{rx}}, P^{(2)\to \mathrm{rx}} \rangle$, and the amount of power $\mathbf{P}^{\mathrm{rx}\to} \triangleq \langle P^{\mathrm{rx}\to(1)}, P^{\mathrm{rx}\to(2)} \rangle$ to send over the two beams, i.e., $a = (\boldsymbol{\tau}, \tau^{\mathrm{rx}}, \mathbf{P}^{\to \mathrm{rx}}, \mathbf{P}^{\mathrm{rx}\to})$. Formally, μ defines $\mathbb{P}_\mu(a|\mathbf{s})$, with $\sum_{a \in \mathcal{A}_\mathbf{s}} \mathbb{P}_\mu(a|\mathbf{s}) = 1$, where $\mathcal{A}_\mathbf{s}$ is the set of the possible actions in state $\mathbf{s}$ (e.g., $\mathcal{A}_\mathbf{s}$ includes the energy constraints imposed by the battery levels).

For the sake of presentation simplicity, in the next sections we use a deterministic policy μ, i.e., $\mathbb{P}_\mu(a|\mathbf{s})$ is equal to 1 for $a = \bar{a}_\mathbf{s}$ and to 0 for $a \neq \bar{a}_\mathbf{s}$, where $\bar{a}_\mathbf{s}$ is an action in $\mathcal{A}_\mathbf{s}$.[14] However, in our numerical evaluation we consider a more general random policy.

Our focus is on the long-term throughput optimization problem. This is suitable for scenarios in which nodes operate in the same location for a sufficient amount of time (e.g., sensors), but can be easily extended to the finite-horizon case with similar techniques. Our goal is to maximize the minimum throughput value reached by both devices in order to increase the QoS. Formally, the reward G_μ is expressed as

$$G_\mu = \min\{G^{(1)}_\mu, G^{(2)}_\mu\}, \tag{4.2.6}$$

$$G^{(i)}_\mu \triangleq \liminf_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} \mathbb{E}\left[\tau^{(i)}_k r(P^{(i)\to \mathrm{rx}}_k, \hbar^{(i)}_k)\right], \qquad i \in \{1, 2\}, \tag{4.2.7}$$

where the expectation is taken with respect to the channel conditions. The maximization process is

$$\mu^\star = \arg\max_\mu G_\mu, \tag{4.2.8}$$

[13] In the cases in which CSI is only partially available, our model is useful to characterize a performance upper bound. A detailed analysis of the partial CSI case is left for future study.

[14] All our results can be extended to a general random policy by substituting a in Bellman's equation (4.2.12) with a probability distribution.

where $\mu^\star$ is the Optimal Policy (OP). Note that the optimal policy we find in this section may be useful to compare other suboptimal low-complexity policies, as well as to understand which is the maximum throughput a network can supply.

4.2.3 Optimal Solution

In this section, we will show how to solve the problem described in Section 4.2.2.4 and obtain OP. In particular, by exploiting MDP theory, the problem can be simplified by focusing on the optimization of $\bar{a}_{\mathbf{s}}$ for every fixed $\mathbf{s}$ instead of considering the whole function μ, i.e., the optimization can be parallelized (see Bellman's equation in [14]). Moreover, we will describe how it is possible to reduce the action $a = (\boldsymbol{\tau}, \tau^{\mathrm{rx}}, \mathbf{P}^{\to \mathrm{rx}}, \mathbf{P}^{\mathrm{rx}\to})$ to a simpler action with only four entries $\tilde{a} = (\tau^{\mathrm{rx}}, P^{\mathrm{rx}\to(1)}, \mathbf{q})$.

4.2.3.1 Max-min Problem

We now derive a simple technique to deal with the max-min optimization problem of Equation (4.2.8). Indeed, since standard dynamic programming techniques are designed for min or max (and not max-min) problems, we recast the problem in a standard form.

Consider a new optimization problem, similar to the previous one except for the objective function, which becomes $H_\mu(\alpha)$ instead of G_μ:

$$H_\mu(\alpha) = \alpha\, G_\mu^{(1)} + (1-\alpha) G_\mu^{(2)}, \tag{4.2.9}$$

where $\alpha \in [0,1]$ is a constant. Note that the new problem

$$\mu^\star(\alpha) = \arg\max_{\mu} H_\mu(\alpha) \tag{4.2.10}$$

is expressed in a max form, and thus is easier to solve. If $\alpha = 1$ [$\alpha = 0$], then we are maximizing the performance of device $\mathrm{D}^{(1)}$ [$\mathrm{D}^{(2)}$] only and neglecting the other device.

Let $\mu^\star(\alpha)$ be the policy which maximizes $H_\mu(\alpha)$ for a given α. Since $\mu^\star(\alpha)$ depends upon α, also $G^{(1)}_{\mu^\star(\alpha)}$ and $G^{(2)}_{\mu^\star(\alpha)}$ implicitly depend upon α. It is straightforward to show that $G^{(1)}_{\mu^\star(\alpha)}$ [$G^{(2)}_{\mu^\star(\alpha)}$] increases [decreases] as α increases. We now want to find the value $\bar{\alpha}$ such that the new problem coincides with the original one. Consider the following intuitive result.

Lemma 4.2.1. *The optimal solution of Problem* (4.2.8) *allocates the same throughput to both users.*

Therefore, we impose Lemma 4.2.1 as a design constraint for the new problem and name $\bar{\alpha}$ the value of α at which such condition is satisfied, i.e., $G^{(1)}_{\mu^\star(\bar{\alpha})} = G^{(2)}_{\mu^\star(\bar{\alpha})}$. Under this condition, we have

$$H_{\mu^\star(\bar{\alpha})}(\bar{\alpha}) = G^{(1)}_{\mu^\star(\bar{\alpha})} = G^{(2)}_{\mu^\star(\bar{\alpha})}. \tag{4.2.11}$$

As a consequence, at $\alpha = \bar{\alpha}$, we obtain $\mu^\star \equiv \mu^\star(\bar{\alpha})$, i.e., OP (solution of (4.2.8)) coincides with the new policy $\mu^\star(\bar{\alpha})$ which maximizes $H_\mu(\bar{\alpha})$. This procedure simplifies the numerical optimization because $\mu^\star(\bar{\alpha})$ can be found exploiting standard stochastic optimization algorithms, e.g., the Value Iteration Algorithm (VIA), or the Policy Iteration Algorithm (PIA) (see Section 2.2.2).

Practically, the value $\bar{\alpha}$ which satisfies (4.2.11) can be found with a bisection search as follows. First, arbitrarily fix $\alpha \in [0,1]$ and maximize $H_\mu(\alpha)$ with VIA or PIA. Using

the optimal solution, compute $G^{(1)}_{\mu^\star(\alpha)}$ and $G^{(2)}_{\mu^\star(\alpha)}$. If $G^{(1)}_{\mu^\star(\alpha)}$ is greater [less] than $G^{(2)}_{\mu^\star(\alpha)}$, then decrease [increase] α and repeat the procedure. The algorithm is repeated until the throughputs of the two nodes are within ϵ of each other, with ϵ a sufficiently small constant. In the next, we will (equivalently) deal with $H_\mu(\alpha)$ instead of G_μ.

4.2.3.2 Bellman's Equation Structure

The most suitable algorithms to solve our problem are VIA or PIA. In the next we describe the *policy improvement step* which is one of the basic operations of both algorithms (see [14, Sec. 7.4, Vol. 1]).

We define the cost-to-go function associated to state $\mathbf{s}$ as $J_\mathbf{s}$. The policy improvement step exploits Bellman's equation as follows

$$J_\mathbf{s} \leftarrow \max_{a\in\mathcal{A}_\mathbf{s}} \left\{ r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to\mathrm{rx}}|\boldsymbol{\hbar}) + \sum_{\mathbf{s}'} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a) J_{\mathbf{s}'} \right\}, \tag{4.2.12}$$

$$r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to\mathrm{rx}}|\boldsymbol{\hbar}) \triangleq \alpha\,\tau^{(1)} r(P^{(1)\to\mathrm{rx}}, \hbar^{(1)}) + (1-\alpha)\tau^{(2)} r(P^{(2)\to\mathrm{rx}}, \hbar^{(2)}). \tag{4.2.13}$$

The probability of going from state $\mathbf{s}$ to state $\mathbf{s}'$ given the action a can be expressed as

$$\begin{aligned}
\mathbb{P}(\mathbf{s}'|\mathbf{s}, a) &\overset{(a)}{=} \mathbb{P}(\mathbf{b}', \boldsymbol{g}', \boldsymbol{\hbar}'|\mathbf{b}, \boldsymbol{g}, \boldsymbol{\hbar}, a) && \text{(4.2.14a)}\\
&\overset{(b)}{=} \mathbb{P}(\mathbf{b}', \boldsymbol{g}', \boldsymbol{\hbar}'|\mathbf{b}, \boldsymbol{g}, a) && \text{(4.2.14b)}\\
&\overset{(c)}{=} f(\boldsymbol{g}', \boldsymbol{\hbar}')\mathbb{P}(\mathbf{b}'|\mathbf{b}, \boldsymbol{g}, a) && \text{(4.2.14c)}\\
&\overset{(d)}{=} f(\boldsymbol{g}', \boldsymbol{\hbar}')\mathbb{P}(b^{(1)\prime}|b^{(1)}, g^{(1)}, a)\mathbb{P}(b^{(2)\prime}|b^{(2)}, g^{(2)}, a), && \text{(4.2.14d)}
\end{aligned}$$

where $f(\boldsymbol{g}, \boldsymbol{\hbar})$ is the probability mass function of the channel state (note that the randomness is given by the fading components $\theta^{(i)}$ and $\kappa^{(i)}$ only). (a) holds by definition. (b) holds because the uplink channel does not influence the battery evolution (given the action). (c) holds because the channel is independent and indentically distributed (i.i.d.) over time and independent of other quantities. The last step holds because the states of the batteries evolve independently in the two devices, given a fixed action. Exploiting Equation (4.2.4) and the MDP formulation, the transition probabilities can be expressed as follows. If $b^{(i)\prime} < b^{(i)}_{\max}$,

$$\begin{aligned}
\mathbb{P}(b^{(i)\prime}|b^{(i)}, g^{(i)}, a) = \chi\Bigg\{ &\left[b^{(i)} - \left\lceil \left(\zeta^{(i)} + \tau^{(i)} P^{(i)\to\mathrm{rx}}\right) \frac{b^{(i)}_{\max}}{B^{(i)}_{\max}} \right\rceil - q^{(i)}_d \right]^+ && \text{(4.2.15)}\\
&+ \left\lfloor \tau^{\mathrm{rx}} P^{\mathrm{rx}\to(i)} \eta_{\mathrm{ECL}} g^{(i)} \frac{b^{(i)}_{\max}}{B^{(i)}_{\max}} \right\rfloor = b^{(i)\prime} \Bigg\}, && \text{(4.2.16)}
\end{aligned}$$

otherwise

$$\begin{aligned}
\mathbb{P}(b^{(i)}_{\max}|b^{(i)}, g^{(i)}, a) = \chi\Bigg\{ &\left[b^{(i)} - \left\lceil \left(\zeta^{(i)} + \tau^{(i)} P^{(i)\to\mathrm{rx}}\right) \frac{b^{(i)}_{\max}}{B^{(i)}_{\max}} \right\rceil - q^{(i)}_d \right]^+ && \text{(4.2.17)}\\
&+ \left\lfloor \tau^{\mathrm{rx}} P^{\mathrm{rx}\to(i)} \eta_{\mathrm{ECL}} g^{(i)} \frac{b^{(i)}_{\max}}{B^{(i)}_{\max}} \right\rfloor \geq b^{(i)}_{\max} \Bigg\}. && \text{(4.2.18)}
\end{aligned}$$

$\chi\{\cdot\}$ is the indicator function and the *floor* is used to discretize the energy and use the MDP approach. Equations (4.2.15)-(4.2.18) indicate that the battery transitions follow a deterministic scheme (given the action and the state of the system). Intuitively, this happens because the randomness of the channel fading is already included in $g^{(i)}$. Therefore, (4.2.12) can be reformulated as follows

$$J_{\mathbf{s}} \leftarrow \max_{a \in \mathcal{A}_{\mathbf{s}}} \left\{ r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to \mathrm{rx}} | \boldsymbol{\hbar}) + \sum_{\boldsymbol{g}', \boldsymbol{\hbar}'} f(\boldsymbol{g}', \boldsymbol{\hbar}') J_{(\mathbf{b}', \boldsymbol{g}', \boldsymbol{\hbar}')} \right\}, \tag{4.2.19}$$

with $\mathbf{b}'$ defined according to Equation (4.2.15)-(4.2.18). Note that, with this observation, we can avoid to iterate over $\mathbf{b}'$, saving computation time.

Another interesting point is that $\mathbf{b}'$ does not depend upon the particular values of $\boldsymbol{\tau}$ and $\mathbf{P}^{\to \mathrm{rx}}$ but only upon their products $\tau^{(1)} P^{(1) \to \mathrm{rx}}$ and $\tau^{(2)} P^{(2) \to \mathrm{rx}}$ because of the structure of $q^{(i)}$. We will use this property in the next section.

4.2.3.3 Variables Reduction

VIA or PIA requires to focus on the maximization of Equation (4.2.19) only, which can be formally written as (in this section we always refer to a fixed state $\mathbf{s} = (\mathbf{b}, \boldsymbol{g}, \boldsymbol{\hbar})$)

$$\max_{\boldsymbol{\tau}, \tau^{\mathrm{rx}}, \mathbf{P}^{\to \mathrm{rx}}, \mathbf{P}^{\mathrm{rx} \to}} \left\{ r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to \mathrm{rx}} | \boldsymbol{\hbar}) + \Delta(\boldsymbol{\tau} \circ \mathbf{P}^{\to \mathrm{rx}}, \tau^{\mathrm{rx}}, \mathbf{P}^{\mathrm{rx} \to} | \boldsymbol{g}) \right\}, \tag{4.2.20a}$$

s.t.:

$$\zeta^{(i)} + \tau^{(i)} P^{(i) \to \mathrm{rx}} \leq B^{(i)}, \qquad i \in \{1, 2\}, \tag{4.2.20b}$$

$$\tau^{(1)} + \tau^{(2)} + \tau^{\mathrm{rx}} \leq T, \tag{4.2.20c}$$

$$P^{\mathrm{rx} \to (1)} + P^{\mathrm{rx} \to (2)} \leq P^{\mathrm{rx}}_{\max}, \tag{4.2.20d}$$

$$\boldsymbol{\tau} \succeq 0, \ \tau^{\mathrm{rx}} \geq 0, \ \mathbf{P}^{\to \mathrm{rx}}_{\min} \preceq \mathbf{P}^{\to \mathrm{rx}} \preceq \mathbf{P}^{\to \mathrm{rx}}_{\max}, \ \mathbf{P}^{\mathrm{rx} \to} \succeq 0. \tag{4.2.20e}$$

Constraints (4.2.20b)-(4.2.20e) represent the set $\mathcal{A}_{\mathbf{s}}$.[15] $\succeq$ and $\preceq$ are the component-wise inequalities. $\Delta(\boldsymbol{\tau} \circ \mathbf{P}^{\to \mathrm{rx}}, \tau^{\mathrm{rx}}, \mathbf{P}^{\mathrm{rx} \to} | \boldsymbol{g})$ is a quantity that, as the second term in Equation (4.2.19), does not depend upon the individual values of $\boldsymbol{\tau}$ and $\mathbf{P}^{\to \mathrm{rx}}$ but only on their products ($\circ$ denotes the Hadamard, i.e., entry-wise, product). This happens because the battery update formulas consider only the overall energy consumption of a device in a slot, that depends on the transmission duration $\tau^{(i)}$ multiplied by the transmission power $P^{(i) \to \mathrm{rx}}$ (see Equations (4.2.15) and (4.2.18)). Without deriving particular properties of $J_{\mathbf{s}}$, the classic procedure to solve (4.2.20) is to perform an exhaustive search over all the seven optimization variables. However, this computation may be too demanding[16] and simpler optimization techniques are desirable. In particular, in this section we propose a method to simplify the optimization.

First, it can be shown that choosing $P^{\mathrm{rx} \to (1)} + P^{\mathrm{rx} \to (2)} = P^{\mathrm{rx}}_{\max}$ is optimal (otherwise the available resources would be underused). Similarly, using $\tau^{\mathrm{rx}} < T - \tau^{(1)} - \tau^{(2)}$ is suboptimal. Therefore, without loss of optimality, we can choose $P^{\mathrm{rx} \to (2)} = P^{\mathrm{rx}}_{\max} - P^{\mathrm{rx} \to (1)}$ and $\tau^{(2)} = T - \tau^{\mathrm{rx}} - \tau^{(1)}$ and avoid to iterate over $P^{\mathrm{rx} \to (2)}$ and $\tau^{(2)}$. Now, fix the products

[15] Technically, we should also consider the cases in which $P^{(1) \to \mathrm{rx}} = 0$ and/or $P^{(2) \to \mathrm{rx}} = 0$. However, these are trivial cases that can be easily analyzed separately.

[16] Note that Problem (4.2.20) must be solved for every combination of $\mathbf{b}$, $\boldsymbol{g}$, $\boldsymbol{\hbar}$ and for every step of PIA.

$\boldsymbol{\tau} \circ \mathbf{P}^{\to \mathrm{rx}} = \mathbf{Q} - \boldsymbol{\zeta}$, where $Q^{(i)}$ represents the energy consumed by device $\mathrm{D}^{(i)}$. In order to solve Problem (4.2.20), we consider the vector $\mathbf{Q}$ instead of $\boldsymbol{\tau}$ and $\mathbf{P}^{\to \mathrm{rx}}$.

Given $P^{\mathrm{rx}\to(1)}$, τ^{rx}, $\mathbf{Q}$, the particular values for the duration and the transmission power are extracted by solving the following sub-problem

$$\max_{\boldsymbol{\tau}, \mathbf{P}^{\to \mathrm{rx}}} \left\{ r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to \mathrm{rx}} | \boldsymbol{\hbar}) + \Delta(\mathbf{Q} - \boldsymbol{\zeta}, \tau^{\mathrm{rx}}, \mathbf{P}^{\mathrm{rx}\to} | \boldsymbol{g}) \right\}, \tag{4.2.21a}$$

s.t.:

$$\boldsymbol{\tau} \circ \mathbf{P}^{\to \mathrm{rx}} = \mathbf{Q} - \boldsymbol{\zeta}, \tag{4.2.21b}$$

$$\tau^{(1)} + \tau^{(2)} = T - \tau^{\mathrm{rx}}, \tag{4.2.21c}$$

$$\boldsymbol{\tau} \succeq 0, \ \mathbf{P}^{\to \mathrm{rx}}_{\min} \preceq \mathbf{P}^{\to \mathrm{rx}} \preceq \mathbf{P}^{\to \mathrm{rx}}_{\max}, \tag{4.2.21d}$$

where $\Delta(\mathbf{Q} - \boldsymbol{\zeta}, \tau^{\mathrm{rx}}, \mathbf{P}^{\mathrm{rx}\to} | \boldsymbol{g})$ is a constant term that can be removed from the "max" argument. Problem (4.2.21) can be rewritten as a function of $\tau^{(1)}$ only:

$$\max_{\tau^{(1)}} \left\{ \alpha \, \tau^{(1)} r \left(\frac{Q^{(1)} - \zeta^{(1)}}{\tau^{(1)}}, \hbar^{(1)} \right) + (1 - \alpha)(T - \tau^{\mathrm{rx}} - \tau^{(1)}) r \left(\frac{Q^{(2)} - \zeta^{(2)}}{T - \tau^{\mathrm{rx}} - \tau^{(1)}}, \hbar^{(2)} \right) \right\}, \tag{4.2.22a}$$

s.t.:

$$\tau^{(1)} \geq \tau^{(1)}_{\min} \triangleq \max \left\{ \frac{Q^{(1)} - \zeta^{(1)}}{P^{(1)\to \mathrm{rx}}_{\max}}, T - \tau^{\mathrm{rx}} - \frac{Q^{(2)} - \zeta^{(2)}}{P^{(2)\to \mathrm{rx}}_{\min}} \right\}, \tag{4.2.22b}$$

$$\tau^{(1)} \leq \tau^{(1)}_{\max} \triangleq \min \left\{ \frac{Q^{(1)} - \zeta^{(1)}}{P^{(1)\to \mathrm{rx}}_{\min}}, T - \tau^{\mathrm{rx}} - \frac{Q^{(2)} - \zeta^{(2)}}{P^{(2)\to \mathrm{rx}}_{\max}} \right\}, \tag{4.2.22c}$$

(4.2.22) is a one-dimensional maximization problem which (except in the trivial cases, e.g., $Q^{(1)} = 0$ or $Q^{(2)} = 0$ or no feasible solutions) can be easily solved by taking the derivative of the reward function, given by the following expression

$$\begin{aligned} &\alpha \left(\log \left(1 + \frac{\hbar^{(1)}}{\sigma_0^2} \frac{Q^{(1)} - \zeta^{(1)}}{\tau^{(1)}} \right) - \frac{(Q^{(1)} - \zeta^{(1)}) \hbar^{(1)}}{\tau^{(1)} \sigma_0^2 + (Q^{(1)} - \zeta^{(1)}) \hbar^{(1)}} \right) \\ &+ (1 - \alpha) \left(\frac{(Q^{(2)} - \zeta^{(2)}) \hbar^{(2)}}{(T - \tau^{\mathrm{rx}} - \tau^{(1)}) \sigma_0^2 + (Q^{(2)} - \zeta^{(2)}) \hbar^{(2)}} - \log \left(1 + \frac{\hbar^{(2)}}{\sigma_0^2} \frac{(Q^{(2)} - \zeta^{(2)})}{T - \tau^{\mathrm{rx}} - \tau^{(1)}} \right) \right), \end{aligned} \tag{4.2.23}$$

and setting it to zero. It can be shown that the previous expression has a unique zero in $(0, T - \tau^{\mathrm{rx}})$ that corresponds to the optimal value $\tau^{(1)}_{\mathrm{n.c.}}$ of Problem (4.2.22) without constraints. The optimal solution of (4.2.22), namely $\tau^{(1)\star}$, can be found as

$$\tau^{(1)\star} = \max\{\min\{\tau^{(1)}_{\mathrm{n.c.}}, \tau^{(1)}_{\max}\}, \tau^{(1)}_{\min}\}. \tag{4.2.24}$$

Given τ^{rx}, $\mathbf{Q}$ and $\tau^{(1)\star}$, the values of $\tau^{(2)\star}$, $P^{(1)\to \mathrm{rx}\star}$ and $P^{(2)\to \mathrm{rx}\star}$ can be derived from (4.2.21b)-(4.2.21c).

In summary, instead of performing an exhaustive search over seven variables, we just iterate over τ^{rx}, $P^{\mathrm{rx}\to(1)}$ and $\mathbf{Q}$, and recover the other parameters by solving (4.2.21) and

choosing $P^{\mathrm{rx}\to(2)} = P^{\mathrm{rx}}_{\max} - P^{\mathrm{rx}\to(1)}$, $\tau^{(2)} = T - \tau^{\mathrm{rx}} - \tau^{(1)}$. We also remark that $Q^{(1)}$, $Q^{(2)}$ must satisfy $q^{(i)} \triangleq Q^{(i)} b^{(i)}_{\max}/B^{(i)}_{\max} \in \{0, \ldots, b^{(i)}_{\max}\}$.

4.2.3.4 Low-SNR Regime

An interesting and practical case in which more analytical results can be developed and explained is the low-SNR regime. In this section we provide additional details for such a case. We assume $\hbar^{(1)} P^{(1)\to\mathrm{rx}}/\sigma_0^2 \ll 1$ and $\hbar^{(2)} P^{(2)\to\mathrm{rx}}/\sigma_0^2 \ll 1$ (low-SNR assumption), therefore $r(P^{(i)\to\mathrm{rx}}, \hbar^{(i)}) \approx \hbar^{(i)} P^{(i)\to\mathrm{rx}}/\sigma_0^2$. In this case, $r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to\mathrm{rx}}|\boldsymbol{\hbar})$ reduces to $\alpha(Q^{(1)} - \zeta^{(1)})\hbar^{(1)}/\sigma_0^2 + (1-\alpha)(Q^{(2)} - \zeta^{(2)})\hbar^{(2)}/\sigma_0^2$, i.e., it depends only upon the product $\boldsymbol{\tau} \circ \mathbf{P}^{\to\mathrm{rx}} = \mathbf{Q} - \boldsymbol{\zeta}$. Therefore, the best choice becomes to use the maximum transmission power $P^{(i)\to\mathrm{rx}}_{\max}$ and the minimum transmission duration $(Q^{(i)} - \zeta^{(i)})/P^{(i)\to\mathrm{rx}}_{\max}$ at both devices. In this way, the system achieves the same reward per slot and maximizes the downlink phase, thus more energy is harvested and stored. As a consequence, once $\mathbf{Q}$ is specified, the downlink duration τ^{rx} is uniquely determined as $\tau^{\mathrm{rx}} = T - (Q^{(1)} - \zeta^{(1)})/P^{(1)\to\mathrm{rx}}_{\max} - (Q^{(2)} - \zeta^{(2)})/P^{(2)\to\mathrm{rx}}_{\max}$.

4.2.3.5 Reducing State Space Complexity

In a general step of PIA or VIA, given the current policy, the corresponding cost-to-go function $J_{\mathbf{s}}$ has to be computed (policy evaluation step [14, Sec. 7.4, Vol. 1]). This process is challenging when the state space is large.

So far, the state of the system is the tuple $\mathbf{s} = (\mathbf{b}, \boldsymbol{g}, \boldsymbol{\hbar})$. However, since $\boldsymbol{g}$ and $\boldsymbol{\hbar}$ evolve independently over time, the state space can be reduced to $\mathbf{s} = \mathbf{b}$ only, as follows. Define a new cost-to-go function

$$K_{\mathbf{b}} \triangleq \sum_{\boldsymbol{g},\boldsymbol{\hbar}} f(\boldsymbol{g}, \boldsymbol{\hbar}) J_{(\mathbf{b},\boldsymbol{g},\boldsymbol{\hbar})}. \tag{4.2.25}$$

$K_{\mathbf{b}}$ substitutes $J_{(\mathbf{b},\boldsymbol{g},\boldsymbol{\hbar})}$ in the original problem. Indeed, we can rewrite the policy improvement step as

$$K_{\mathbf{b}} \leftarrow \sum_{\boldsymbol{g},\boldsymbol{\hbar}} f(\boldsymbol{g}, \boldsymbol{\hbar}) \max_{a \in \mathcal{A}_{(\mathbf{b},\boldsymbol{g},\boldsymbol{\hbar})}} \left\{ r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to\mathrm{rx}}|\boldsymbol{\hbar}) + \sum_{\mathbf{s}'} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a) J_{\mathbf{s}'} \right\} \tag{4.2.26a}$$

$$= \sum_{\boldsymbol{g},\boldsymbol{\hbar}} f(\boldsymbol{g}, \boldsymbol{\hbar}) \max_{a \in \mathcal{A}_{(\mathbf{b},\boldsymbol{g},\boldsymbol{\hbar})}} \left\{ r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to\mathrm{rx}}|\boldsymbol{\hbar}) + K_{\mathbf{b}'} \right\}, \tag{4.2.26b}$$

where $\mathbf{b}'$ is defined according to Equations (4.2.15)-(4.2.18).

This procedure further simplifies the numerical computation without loss of optimality because 1) it reduces the complexity of the policy evaluation step (there is a lower number of states) and 2) it reduces the number of elementary operations inside the "max" operation in the policy improvement step.

4.2.4 Approximate Scheme

Finding the optimal policy is practically feasible only for a relatively small number of discrete values, which however corresponds to a rough quantization. Therefore, in this section we propose a method which is based on the characteristics of the original solution but is faster to compute and achieves approximately the same performance of OP. This is particularly useful to characterize the system performance and identify the system trade-offs.

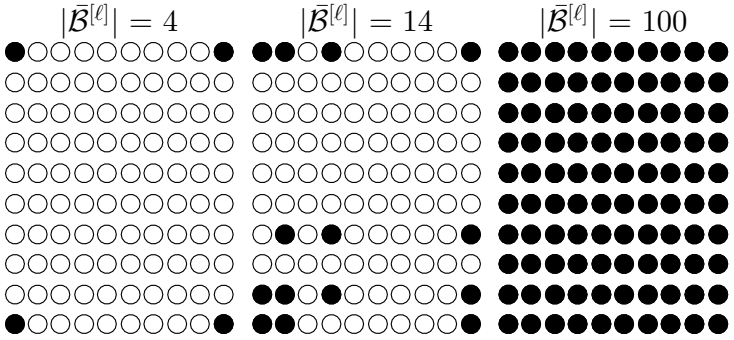

Figure 4.13. Different sets $\bar{\mathcal{B}}^{[\ell]}$ when $b_{\max}^{(1)} = b_{\max}^{(2)} = 9$. Rows and columns correspond to $b^{(1)}$ and $b^{(2)}$, respectively.

Even with the simplifications introduced in Section 4.2.3, the main challenge is to perform the policy improvement step, i.e., solving Equation (4.2.26) for all system states. To manage this problem, several different approximate techniques have been proposed in the literature so far. An interesting idea is to approximate the function $K_{\mathbf{b}}$ with another one that is simpler to compute. We follow this approach in the remainder of this section, and derive an Approximate Value Iteration Algorithm (App-VIA) (see [15, Sec. 6.5]).

4.2.4.1 Approximate Value Iteration Algorithm

In the classic VIA, the optimal policy is derived by iteratively solving Equation (4.2.26) until the cost-to-go function converges. In the approximate approach, we modify every iteration of VIA according to the following two steps:

1. compute $K_{\mathbf{b}}^{[\ell]}$ for every $\mathbf{b} \in \bar{\mathcal{B}}^{[\ell]}$ performing the policy improvement step (Equation (4.2.26)), with $\bar{\mathcal{B}}^{[\ell]} \subseteq \mathcal{B} = \{0, \dots, b_{\max}^{(1)}\} \times \{0, \dots, b_{\max}^{(2)}\}$. The superscript $[\ell]$ denotes the ℓ-th iteration of VIA and $\mathcal{B}$ is the set of all battery levels;

2. interpolate $K_{\mathbf{b}}^{[\ell]}$ for every $\mathbf{b} \in \mathcal{B} \backslash \bar{\mathcal{B}}^{[\ell]}$ using the values of $K_{\mathbf{b}}^{[\ell]}$ computed in the previous step.

The advantage is that the policy improvement is performed only for a subset $\bar{\mathcal{B}}^{[\ell]}$ rather than for every battery level in $\mathcal{B}$. See Figure 4.13 for a graphical interpretation. A black circle means that $\mathbf{b} \in \bar{\mathcal{B}}^{[\ell]}$. In the last case, all the battery levels are in $\bar{\mathcal{B}}^{[\ell]}$, i.e., $\bar{\mathcal{B}}^{[\ell]} = \mathcal{B}$. In general, $\bar{\mathcal{B}}^{[\ell]}$ can dynamically change at every step of the algorithm in a deterministic or stochastic manner. We further discuss our approach in the numerical evaluation section.

We now discuss in more detail the two previous points. The policy improvement step becomes, for every $\mathbf{b} \in \bar{\mathcal{B}}^{[\ell+1]}$,

$$\widehat{K}_{\mathbf{b}}^{[\ell+1]} = \sum_{\boldsymbol{g},\boldsymbol{\hbar}} f(\boldsymbol{g},\boldsymbol{\hbar}) \max_{a \in \mathcal{A}_{(\mathbf{b},\boldsymbol{g},\boldsymbol{\hbar})}} \left\{ r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to \mathrm{rx}} | \boldsymbol{\hbar}) + \widetilde{K}_{\mathbf{b}'}^{[\ell]} \right\}, \tag{4.2.27}$$

where $\mathbf{b}'$ is defined according to Equations (4.2.15)-(4.2.18). $\widehat{K}_{\mathbf{b}}^{[\ell+1]}$ represents the approximate value function at step $\ell+1$ and is defined only in subset $\bar{\mathcal{B}}^{[\ell+1]}$, whereas $\widetilde{K}_{\mathbf{b}}^{[\ell]}$ is such that

$$\widetilde{K}_{\mathbf{b}}^{[\ell]} = \widehat{K}_{\mathbf{b}}^{[\ell]}, \quad \text{if } \mathbf{b} \in \bar{\mathcal{B}}^{[\ell]}. \tag{4.2.28}$$

In the second phase of the algorithm, for all $\mathbf{b} \notin \bar{\mathcal{B}}^{[\ell]}$, $\widetilde{K}_{\mathbf{b}}^{[\ell]}$ is derived exploiting (4.2.28) with an interpolation process or using a mean square error approximation. In practice, $\widetilde{K}_{\mathbf{b}}^{[\ell]}$ is designed in order to approximate the true function $K_{\mathbf{b}}^{[\ell]}$. We remark that $\widehat{K}_{\mathbf{b}}^{[\ell+1]}$ is defined only in $\bar{\mathcal{B}}^{[\ell]}$, whereas $\widetilde{K}_{\mathbf{b}}^{[\ell+1]}$ is defined for every $\mathbf{b} \in \mathcal{B}$.

4.2.4.2 Properties

In the following we show that, provided that the approximation $\widetilde{K}_{\mathbf{b}}^{[\ell]}$ is sufficiently good, the long-term reward of App-VIA is a good approximation of VIA.

First, we introduce the notation $\mathcal{T}(\cdot)$ as follows. Define the two sets $\mathcal{K}^{[\ell]} \triangleq \{K_{\mathbf{b}}^{[\ell]},\ \forall \mathbf{b} \in \mathcal{B}\}$ and $\widetilde{\mathcal{K}}^{[\ell]} \triangleq \{\widetilde{K}_{\mathbf{b}}^{[\ell]},\ \forall \mathbf{b} \in \mathcal{B}\}$. Then, Equations (4.2.26) and (4.2.27) can be written as

$$K_{\mathbf{b}}^{[\ell+1]} = \mathcal{T}\left(\mathcal{K}^{[\ell]}, \mathbf{b}\right), \quad \forall \mathbf{b} \in \mathcal{B}, \tag{4.2.29}$$

$$\widehat{K}_{\mathbf{b}}^{[\ell+1]} = \mathcal{T}\left(\widetilde{\mathcal{K}}^{[\ell]}, \mathbf{b}\right), \quad \forall \mathbf{b} \in \bar{\mathcal{B}}^{[\ell+1]}, \tag{4.2.30}$$

respectively. Also, assume that the initial configurations are equal, i.e., $\mathcal{K}^{[0]} = \widetilde{\mathcal{K}}^{[0]}$. Note that $K_{\mathbf{b}}^{[\ell+1]}$ is evaluated for every $\mathbf{b}$, whereas we compute $\widehat{K}_{\mathbf{b}}^{[\ell+1]}$ only in subset $\bar{\mathcal{B}}^{[\ell+1]}$.

Proposition 4.2.2. *After I iterations, the cost-to-go functions of App-VIA and VIA differ by at most $I\epsilon$, i.e.,*[17]

$$\| \underbrace{\mathcal{K}^{[I]}}_{\text{VIA}} - \underbrace{\widetilde{\mathcal{K}}^{[I]}}_{\text{App-VIA}} \|_{\infty} \leq I\epsilon, \tag{4.2.31}$$

with

$$\epsilon \triangleq \max_{\ell=0,\ldots,I-1} \max_{\mathbf{b}\in\mathcal{B}} \left| \widetilde{K}_{\mathbf{b}}^{[\ell+1]} - \mathcal{T}\left(\widetilde{\mathcal{K}}^{[\ell]}, \mathbf{b}\right) \right|. \tag{4.2.32}$$

Proof. See Appendix 4.B. ■

We first remark that, because of Definition (4.2.32), Proposition 4.2.2 describes a worst case analysis. I corresponds to the number of iterations of VIA and, in our problem, it can be numerically verified that I is typically small, e.g., $I \approx 10$. The previous proposition provides a bound to the algorithm's performance. When the approximation of $K_{\mathbf{b}}^{[\ell+1]}$ is sufficiently good, since I is small, the proposed approach closely approximates the optimal case.

In Figure 4.14 we show the performance of this approximate approach as we increase $|\bar{\mathcal{B}}^{[\ell]}|$. When $|\bar{\mathcal{B}}^{[\ell]}| = |\mathcal{B}|$, we obtain the optimal scheme. It is interesting to note that using only few states already leads to close to optimal performance, thus it is not necessary to use very large $|\bar{\mathcal{B}}^{[\ell]}|$ to obtain a good approximation.

4.2.5 Extensions

The model we present in this section is quite general and can be adapted to different EH scenarios by changing the energy arrival and consumption models, the fading statistics, as well as the battery sizes and energy losses. Also, it is easy to include in the model other sources of inefficiencies (see Section 5.1) by changing the energy evolution model of Equation (4.2.5).

[17] We adopt the notation $\|\mathcal{K}^{[I]} - \widetilde{\mathcal{K}}^{[I]}\|_{\infty} \triangleq \max_{\mathbf{b}\in\mathcal{B}} |K_{\mathbf{b}}^{[I]} - \widetilde{K}_{\mathbf{b}}^{[I]}|$.

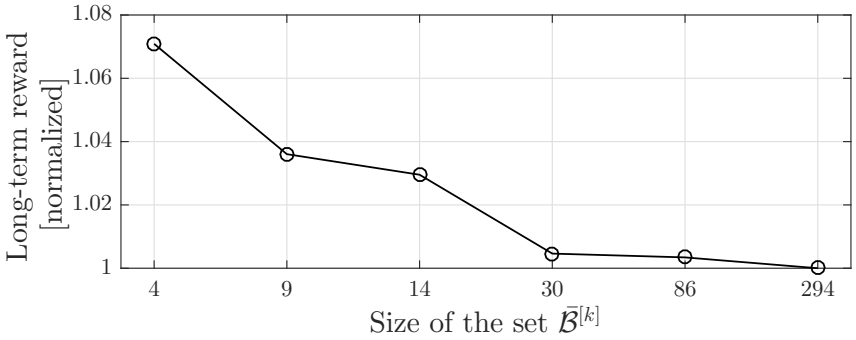

Figure 4.14. Accuracy of the approximate approach as a function of $|\bar{\mathcal{B}}^{[\ell]}|$.

An important extension of the model is the study of a system with a generic number of users > 2. The centralized model we have presented can be straightforwardly extended to such a case, but finding the *optimal* policy would also incur high computational complexity.[18] It can be shown that PIA or VIA can be performed with complexity $\mathcal{O}(|\mathcal{B}|^2|\mathcal{S}|) = \mathcal{O}((b_{\max}^{(1)} b_{\max}^{(2)})^3 \times n_{\text{ch}}^{(1)} n_{\text{ch}}^{(2)})$, as for every state of the system $\mathbf{s} \in \mathcal{S}$, we need to find the optimal strategy, which is $\mathcal{O}(b_{\max}^{(1)} b_{\max}^{(2)})$, to reach every other battery level.[19] Thus, when additional states are considered, the state dimension, and thus the complexity, grow significantly. Instead, the heuristic approach of Section 4.2.4 can be computed in $\mathcal{O}(|\bar{\mathcal{B}}^{[\ell]}|^2 \times b_{\max}^{(1)} b_{\max}^{(2)} \times n_{\text{ch}}^{(1)} n_{\text{ch}}^{(2)})$. Since $|\bar{\mathcal{B}}^{[\ell]}| \ll |\mathcal{B}|$ by construction, the heuristic scheme may be extended more easily to the case with more than two users. However, for a system with an arbitrarily large number of users, tracking the battery evolution of all users in a centralized fashion is a computationally demanding task, thus simpler or decentralized techniques should be considered. Two interesting approaches to extend our model are presented in [94], in which a game theoretic formulation is adopted, or [52], where a DEC-MDP framework is presented. Part of our future work agenda includes the investigation of these schemes for WPCNs.

4.2.6 Harvest-then-Transmit

In the literature on WPCN, the main focus so far has been on the optimization in a single time slot, which we briefly report in this section for the sake of completeness. In particular, we consider the "harvest-then-transmit" scheme, in which all the energy harvested in a slot is immediately used for transmission.

If $E^{(1),\text{RF}}$ and $E^{(2),\text{RF}}$ joule of energy are transferred at the beginning of the slot, in the uplink transmission phase $\text{D}^{(i)}$ is subject to the following constraint (for the sake of continuity with the previous sections, we also consider a limited battery)

$$P^{\text{rx}\to(i)} \leq \min\{E^{(i),\text{RF}}, B_{\max}^{(i)}\}, \tag{4.2.33}$$

i.e., it cannot consume more energy than what it received in the same slot nor can it exceed the maximum battery size. Note that $P^{\text{rx}\to(i)}$ accounts for both the circuitry and transmission energy. The optimization variable is a tuple of 7 elements. Formally, the optimization problem is (as in Section 4.2.3, we solve separately the trivial cases in which

[18] It is always possible to decrease the computational burden by focusing on simpler policies, which however do not necessarily achieve the maximum fair-throughput.

[19] We recall that $|\mathcal{B}| = (b_{\max}^{(1)} + 1) \times (b_{\max}^{(2)} + 1)$ and $|\mathcal{S}| = |\mathcal{B}| \times n_{\text{ch}}^{(1)} n_{\text{ch}}^{(2)}$, where we define $n_{i,\text{ch}}$ as the number of possible channel realizations for device i.

at least one device is unused):

$$\max_{\boldsymbol{\tau},\tau^{\mathrm{rx}},\mathbf{P}^{\to\mathrm{rx}},\mathbf{P}^{\mathrm{rx}\to}} \min\{\tau^{(1)} r(P^{(1)\to\mathrm{rx}},\hbar^{(1)}),\tau^{(2)} r(P^{(2)\to\mathrm{rx}},\hbar^{(2)})\}, \tag{4.2.34a}$$

s.t.:

$$\tau^{(1)}+\tau^{(2)}+\tau^{\mathrm{rx}} \le T, \tag{4.2.34b}$$

$$P^{\mathrm{rx}\to(1)}+P^{\mathrm{rx}\to(2)} \le P^{\mathrm{rx}}_{\max}, \tag{4.2.34c}$$

$$\boldsymbol{\zeta}+\boldsymbol{\tau}\circ\mathbf{P}^{\to\mathrm{rx}} \preceq \tau^{\mathrm{rx}}\,\eta_{\mathrm{ECL}}\,\mathbf{P}^{\mathrm{rx}\to}\circ\boldsymbol{g}, \tag{4.2.34d}$$

$$\boldsymbol{\zeta}+\boldsymbol{\tau}\circ\mathbf{P}^{\to\mathrm{rx}} \preceq \mathbf{B}_{\max}, \tag{4.2.34e}$$

$$\boldsymbol{\tau}\succeq 0, \quad \tau^{\mathrm{rx}}\ge 0, \quad \mathbf{P}^{\to\mathrm{rx}}_{\min} \preceq \mathbf{P}^{\to\mathrm{rx}} \preceq \mathbf{P}^{\to\mathrm{rx}}_{\max}, \quad \mathbf{P}^{\mathrm{rx}\to}\succeq 0 \tag{4.2.34f}$$

The solution of (4.2.34) is given in Proposition 4.2.3. Constraints (4.2.34e)-(4.2.34f) identify the *feasible region*.

Proposition 4.2.3. *The optimal* $\mathbf{P}^{\to\mathrm{rx}}$ *(solution of Problem* (4.2.34)*) can be derived as follows (the other parameters are obtained according to Equations* (4.C.3)-(4.C.5) *in Appendix 4.C).*

- *Name* $P_0^{(i)\to\mathrm{rx}}$ *the solution of*

$$\eta_{\mathrm{ECL}} g^{(i)} P^{\mathrm{rx}}_{\max} + P^{(i)\to\mathrm{rx}} = \left(\frac{\sigma_0^2}{\hbar^{(i)}}+P^{(i)\to\mathrm{rx}}\right)\log\left(1+\frac{\hbar^{(i)}}{\sigma_0^2}P^{(i)\to\mathrm{rx}}\right). \tag{4.2.35}$$

  If $\mathbf{P}_0^{\to\mathrm{rx}}$ *and the corresponding* $\boldsymbol{\tau}_0$, $\mathbf{Q}_0$, τ_0^{rx} *lie in the feasible region, then* $\mathbf{P}^{\to\mathrm{rx}^\star}=\mathbf{P}_0^{\to\mathrm{rx}}$*;*

- *otherwise the optimal solution lies on the boundary of the feasible region of* (4.C.2)*, as described in the proof.*

Proof. See Appendix 4.C. ■

Exploiting the results of the previous proposition, we can derive the optimal reward achieved in a single slot. By averaging over the channel gains, we obtain the corresponding long-term throughput

$$G_\omega \triangleq \sum_{\boldsymbol{g},\boldsymbol{\hbar}} f(\boldsymbol{g},\boldsymbol{\hbar})\,\tau^{(i)\star} r(P^{(i)\to\mathrm{rx}^\star},\hbar^{(i)}), \quad i=1,2, \tag{4.2.36}$$

where ω is the *slot-oriented* policy which solves (4.2.34). In the numerical evaluation we will compare G_ω and $G_{\mu^\star}$. Note that, differently from $\mu^\star$, the slot-oriented strategy is much simpler to compute but provides a lower reward, as expected.

4.2.6.1 Low-SNR Regime

In this section we provide additional details for the low-SNR regime in the case $\mathbf{P}^{\to\mathrm{rx}^\star}=\mathbf{P}_0^{\to\mathrm{rx}}$. Equation (4.2.35) can be solved in closed form as

$$P^{(i)\to\mathrm{rx}^\star} = \sqrt{\frac{\eta_{\mathrm{ECL}} g^{(i)} P^{\mathrm{rx}}_{\max}\sigma_0^2}{\hbar^{(i)}}}. \tag{4.2.37}$$

Note that the optimal transmission power of device i depends upon its own parameters only. If the downlink channel gain increases, more energy is harvested, therefore a higher transmission power can be used. Note that the better the uplink channel gain $\hbar^{(i)}$, the lower the transmission power. The corresponding $P^{\mathrm{rx}\to(i)\star}$ can be derived using Equation (4.C.5) (for simplicity, we neglect the circuitry costs)

$$P^{\mathrm{rx}\to(1)\star} = \frac{g^{(2)}\hbar^{(2)}P^{\mathrm{rx}}_{\max}}{g^{(1)}\hbar^{(1)} + g^{(2)}\hbar^{(2)}}, \quad P^{\mathrm{rx}\to(2)\star} = \frac{g^{(1)}\hbar^{(1)}P^{\mathrm{rx}}_{\max}}{g^{(1)}\hbar^{(1)} + g^{(2)}\hbar^{(2)}}. \tag{4.2.38}$$

In order to balance the system performance, $P^{\mathrm{rx}\to(i)\star}$ decreases if $g^{(i)}$ or $\hbar^{(i)}$ increases. In this case, it is better to allocate less resources to the node with a better channel and direct more energy to the other node.

A closed form expression for the reward in a single slot can be obtained. Starting from the equations of $\boldsymbol{\tau}^{\star}$, $\mathbf{P}^{\to\mathrm{rx}^{\star}}$ and $\mathbf{P}^{\mathrm{rx}\to^{\star}}$, we have

$$\begin{aligned}\tau^{(1)\star} r(P^{(1)\to\mathrm{rx}^{\star}}, \hbar^{(1)}) &= \tau^{(2)\star} r(P^{(2)\to\mathrm{rx}^{\star}}, \hbar^{(2)}) \\ &= \frac{\eta_{\mathrm{ECL}} g^{(1)} g^{(2)} \frac{\hbar^{(1)}\hbar^{(2)}}{\sigma_0^2} P^{\mathrm{rx}}_{\max} T}{g^{(2)}\hbar^{(2)}\left(\sqrt{\frac{\eta_{\mathrm{ECL}} g^{(1)}\hbar^{(1)}P^{\mathrm{rx}}_{\max}}{\sigma_0^2}}+1\right) + g^{(1)}\hbar^{(1)}\left(\sqrt{\frac{\eta_{\mathrm{ECL}} g^{(2)}\hbar^{(2)}P^{\mathrm{rx}}_{\max}}{\sigma_0^2}}+1\right)},\end{aligned} \tag{4.2.39}$$

which represents the highest reward that can be achieved in a single slot. The long-term reward can be obtained combining the previous expression with Equation (4.2.36), which can be easily solved numerically.

4.2.7 Numerical Results

We study how the achievable rate changes as a function of the system parameters in different scenarios. As in [58, 80], we assume channel reciprocity for uplink and downlink, thus $g^{(i)} = \hbar^{(i)}$ in every slot (however, we remark that our model is general and can be easily adapted to other cases). We consider an exponential random variable with unit mean for $\theta^{(i)}$ (Rayleigh fading) to model non line-of-sight links or Nakagami fading with parameter 5 when a strong line-of-sight component is present. We explicitly consider energy conversion losses by setting $\eta_{\mathrm{ECL}} = 0.8$. If not otherwise stated, we use the following parameters $\hbar_0^{(1)} = \hbar_0^{(2)} = 1.25 \times 10^{-3}$, $\gamma^{(1)} = \gamma^{(2)} = 2$ (path loss exponents), $\sigma_0^2 = -155$ dBm/Hz (noise power), a bandwidth of 1 MHz, $P^{\mathrm{rx}}_{\max} = 3$ W (maximum transfer power), $P^{(1)\to\mathrm{rx}}_{\min} = P^{(2)\to\mathrm{rx}}_{\min} = 1$ mW and $P^{(1)\to\mathrm{rx}}_{\max} = P^{(2)\to\mathrm{rx}}_{\max} = 10$ mW. Without loss of generality, we assume a unit slot length T. The battery sizes are important parameters which influence the performance of the system. In particular, since with large batteries the throughput of the system saturates, we choose to focus on the case of small batteries [Infinite Power Solutions]. Also, the relation between batteries and slot length should be taken into account. We express all the energy quantities as a function of the reference value $B_0 = T \times 10^{-3}$ J. We also assume equal batteries $B_{\max} \triangleq B^{(1)}_{\max} = B^{(2)}_{\max}$.

In Figure 4.15 we depict the slot division (obtained by averaging all the quantities with the steady-state probabilities) with and without throughput fairness when $d^{(1)} = 1$ m and $d^{(2)} = 3$ or 5 m. The first figure is obtained by setting $\alpha = 0.5$, i.e., the objective function is the unweighted sum of the rewards of the two devices. Since $\mathrm{D}^{(1)}$ is closer to RX and experiences, on average, a better channel, it spends more time transmitting. Moreover,

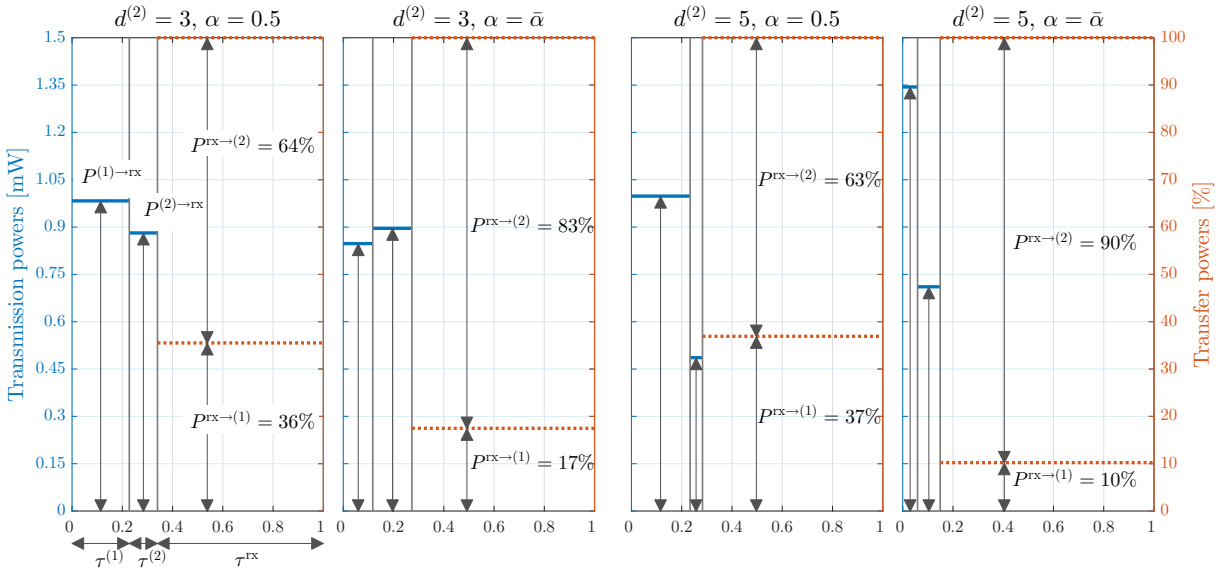

Figure 4.15. Average transmission powers $P^{(1)\to rx}$, $P^{(2)\to rx}$, transfer powers $P^{rx\to(1)}/P^{rx}_{max}$, $P^{rx\to(2)}/P^{rx}_{max}$ and duration $\tau^{(1)}$, $\tau^{(2)}$, τ^{rx} with ($\alpha = \bar{\alpha}$) and without ($\alpha = 0.5$) throughput balancing when $d^{(1)} = 1$ m, $d^{(2)} = 3$ or 5 m and $B_{max} = 0.25B_0$.

even if $P^{rx\to(1)} < P^{rx\to(2)}$, $D^{(1)}$ harvests much more energy than $D^{(2)}$ on average. While this scheme achieves the maximum system sum-throughput, it does not ensure fairness. In particular, the throughput of $D^{(1)}$ is 1.76 Mbps, whereas the throughput of $D^{(2)}$ turns out to be only 0.68 Mbps. It is also worth noting that, thanks to the convexity of (4.2.2), $D^{(2)}$ does contribute to the global performance, and a lot of resources are used to feed it ($P^{rx\to(2)} \gg P^{rx\to(1)}$). $P^{rx\to(1)}$ is smaller than $P^{rx\to(2)}$ because the downlink channel of $D^{(1)}$ is better and thus the first device harvests much more energy. When $d^{(2)}$ increases as in the third plot of Figure 4.15, the transmission duration of $D^{(2)}$ and its harvested energy become much lower. In this case, $D^{(2)}$ is so far from RX with respect to $D^{(1)}$ that it is not worth using a lot of resources to increase its throughput. Instead, the second plot of Figure 4.15 is obtained at the end of the algorithm described in Section 4.2.3.1, i.e., for α equal to $\bar{\alpha} = 0.91$. With this policy, fairness is achieved and the throughput of the two devices $G^{(1)}_{\mu^\star} = G^{(2)}_{\mu^\star}$ is 0.94 Mbps (which, as expected, leads to a smaller sum-throughput than in the unbalanced case). Note that to achieve this situation and to compensate the doubly near-far effect, $D^{(2)}$ must receive much more energy and transmit with much more power than $D^{(1)}$. This phenomenon is emphasized in the last plot, in which 90% of the transmission power is devoted to $D^{(2)}$.

We remark that we used a discrete model to approximate the continuous nature of the energy stored in the batteries (see Section 4.2.2.3), thus $b^{(1)}_{max}$ and $b^{(2)}_{max}$ play a key role in the computation of $\mu^\star$. In particular, for larger batteries higher $b^{(1)}_{max}$ and $b^{(2)}_{max}$ are required, incurring additional numerical complexity, whereas for small batteries the quantization can be coarser. Nevertheless, even with small batteries, computing the optimal policy $\mu^\star$ with PIA or VIA is a computationally intensive task. Therefore, in the following we present our results using the approximate App-VIA scheme introduced in Section 4.2.4. To justify the goodness of our approximation, focus on Figure 4.16, where we depict the throughput as a function of the distance $d^{(1)}$ for several different battery sizes. It can be seen that App-VIA closely approaches the optimal schemes, especially if the battery sizes are small.

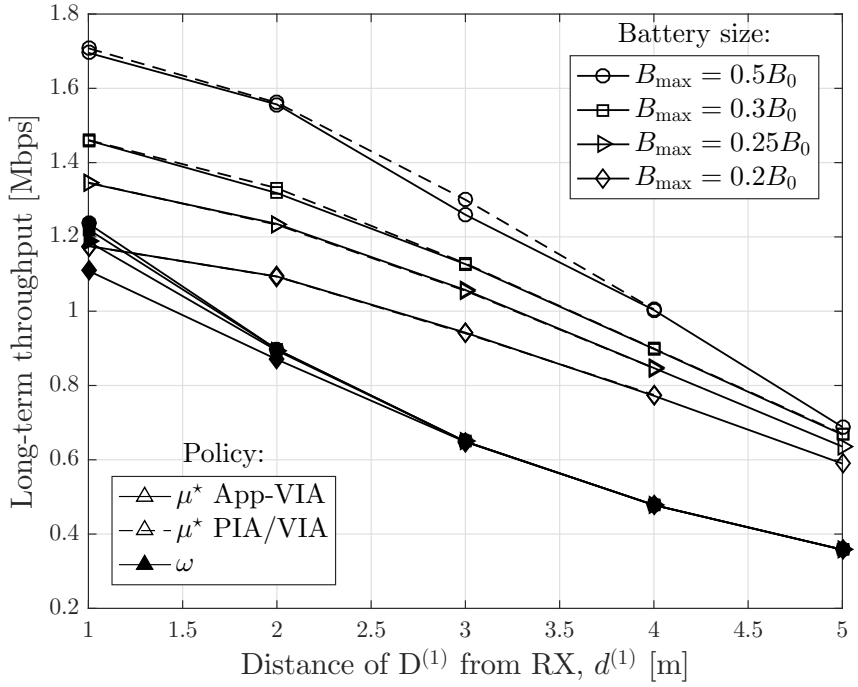

Figure 4.16. Long-term reward of $\mu^\star$ evaluated with PIA/VIA and App-VIA and of ω as a function of $d^{(1)}$ when $d^{(2)} = 3$ m and with Rayleigh fading.

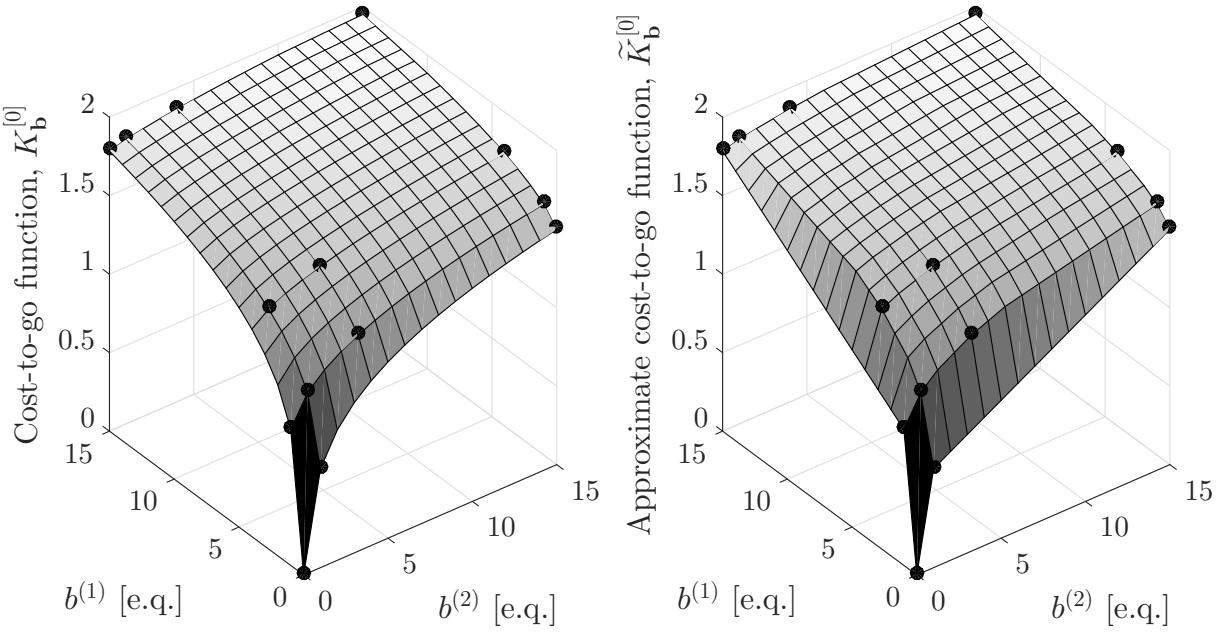

Figure 4.17. Cost-to-go function $K_{\mathbf{b}}^{[0]}$ (left) and its approximation $\widetilde{K}_{\mathbf{b}}^{(0)}$ (right). Black circles represent $\bar{\mathcal{B}}^{[0]}$.

In our numerical evaluation we derived $\bar{\mathcal{B}}^{[\ell]}$ as shown in Figure 4.17 (see the black circles). The left figure represents the optimal cost-to-go function $K_{\mathbf{b}}^{[0]}$, i.e., Problem (4.2.26) has been solved for every pair $(b^{(1)}, b^{(2)})$, whereas the right plot represents its approximation $\widetilde{K}_{\mathbf{b}}^{(0)}$ defined in Section 4.2.4. $\widetilde{K}_{\mathbf{b}}^{(0)}$ is obtained with a linear interpolator.

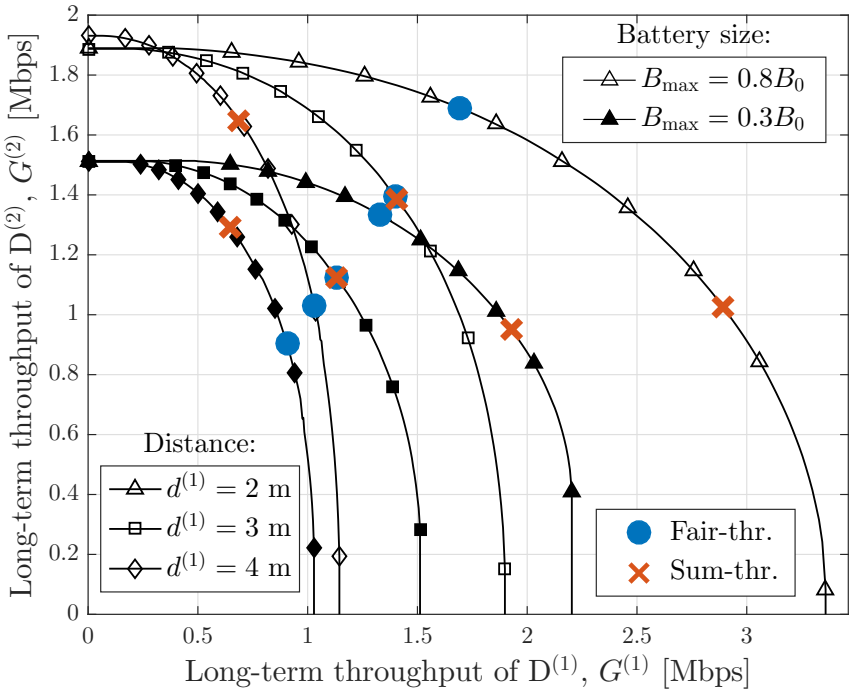

Figure 4.18. Long-term rewards $G^{(1)}$ and $G^{(2)}$ of $\mu^\star$ and ω when $d^{(2)} = 3$ m and with Rayleigh fading

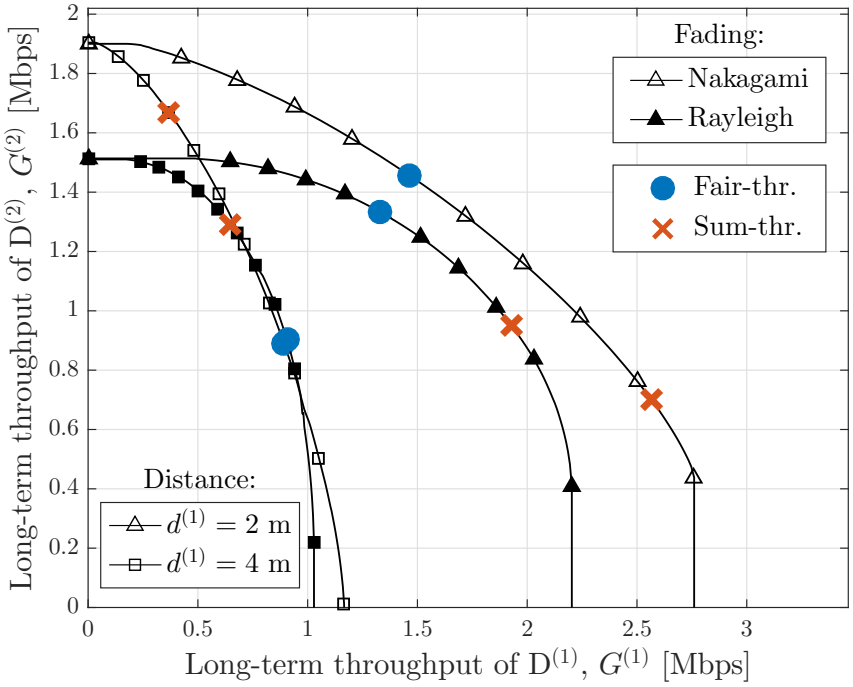

Figure 4.19. Long-term rewards $G^{(1)}$ and $G^{(2)}$ of $\mu^\star$ and ω with Rayleigh and Nakagami fading when $d^{(2)} = 3$ m and $B_{\max} = 0.30B_0$.

Figure 4.18 represents the throughput region of D$^{(1)}$ and D$^{(2)}$, obtained changing α in $[0, 1]$. Blue circles represent the fair-throughput optimal points, whereas the red crosses are the sum-throughput optimal points. They coincide only in the symmetric cases $d^{(1)} = d^{(2)}$. Otherwise, to balance the system performance, part of the throughput of one of the two

devices has be to reduced. Abscissa [ordinate] points are obtained when $\alpha = 1$ [$\alpha = 0$], i.e., $\mathrm{D}^{(2)}$ [$\mathrm{D}^{(1)}$] is completely neglected. Similar curves are depicted in Figure 4.19, where we compare Rayleigh and Nakagami fading. Even if on average the channel gains are the same in the two scenarios, when a strong line-of-sight component is present (as in Nakagami fading), better performance can be achieved because 1) it becomes easier to predict the future energy arrivals and thus to correctly manage the available energy, and 2) the system approaches the deterministic energy arrivals case, which represents an upper bound for energy harvesting scenarios (see Section 4.1).

We now describe how the throughput changes as a function of the distance of $\mathrm{D}^{(1)}$ from RX. Figures 4.20, 4.21 and 4.22 are obtained in the "high transmission power regime," i.e., $P_{\min}^{(1)\to\mathrm{rx}} = P_{\min}^{(2)\to\mathrm{rx}} = 1$ mW and $P_{\max}^{(1)\to\mathrm{rx}} = P_{\max}^{(2)\to\mathrm{rx}} = 10$ mW, whereas Figure 4.23 is determined in the "low transmission power regime," i.e., with $P_{\min}^{(1)\to\mathrm{rx}} = P_{\min}^{(2)\to\mathrm{rx}} = 0.01$ mW and $P_{\max}^{(1)\to\mathrm{rx}} = P_{\max}^{(2)\to\mathrm{rx}} = 0.5$ mW. When $d^{(2)}$ is small (Figure 4.20), the difference between the slot-oriented and the long-term approaches is smaller because a lot of energy is available at the two devices, thus even an inefficient use of it leads to high performance. Instead, as $d^{(2)}$ increases (see Figure 4.22), the difference between the two approaches is significant and this supports the need for a long-term optimization approach. As expected, in all cases the throughput decreases as $d^{(1)}$ increases. This is particularly emphasized when $d^{(2)}$ is small because, since it is farther from RX, $\mathrm{D}^{(1)}$ represents the performance bottleneck. On the contrary, when $d^{(2)} = 5$ m, $\mathrm{D}^{(2)}$ is the bottleneck, thus the system performance shows a weak dependence on the distance of $\mathrm{D}^{(1)}$ from RX. The differences between high and low transmission power regimes can be seen comparing Figures 4.21 and 4.23. It can be seen that with lower transmission powers it is possible to achieve higher rewards.

Indeed, in the analyzed scenario the distances are small, thus the uplink SNR is high even for low transmission powers. Therefore, because of the concavity of the reward function in Equation (4.2.2), with lower transmission powers it may be possible to achieve high throughput while consuming less energy, leading to an overall improvement of the system performance.

In Figure 4.24 we plot the long-term reward of $\mu^\star$ as a function of the battery size of the first device. When $B_{\max}$ is very small, the batteries represent the system bottleneck because $\mathrm{D}^{(1)}$ and/or $\mathrm{D}^{(2)}$ are not able to store and use all the incoming energy. As the battery sizes grow, the performance of the system saturates because the energy available at the receiver $P_{\max}^{\mathrm{rx}}$ is limited. The throughput difference between low and high $B_{\max}$ is larger when $d^{(1)}$ is small because, when the battery of $\mathrm{D}^{(1)}$ is small, the device is not able to fully exploit its channel potential, which instead could be fully used with larger batteries. Some artifacts can be noticed (e.g., at $B_{\max} = 0.45B_0$ for the curve $d^{(1)} = 3$ m) because we are using App-VIA and not the real optimal policy, whose throughput strictly increases with the battery sizes.

Finally, Figures 4.25 and 4.26 show how the performance of the system changes when the circuitry costs and the battery depletion over time are taken into account. In Figure 4.25, we imposed a fixed circuitry energy consumption when a transmission is performed $\zeta \triangleq \zeta^{(1)} = \zeta^{(2)} = 0.125B_0$, and a fixed energy loss at the end of every slot $Q_{\mathrm{d}} \triangleq Q_{\mathrm{d}}^{(1)} = Q_{\mathrm{d}}^{(2)} = 0.025B_0$. When losses are taken into account, the system performance may be greatly degraded, especially because of ζ. An interesting comparison is given by policy ω and $\mu^\star$ when $Q_{\mathrm{d}} > 0$. We remark that, in our model, policy ω does not suffer from energy depletion since energy is used in the slot in which it has been harvested. However, as can be seen in Figure 4.26, policy $\mu^\star$ still achieves higher rewards than ω.

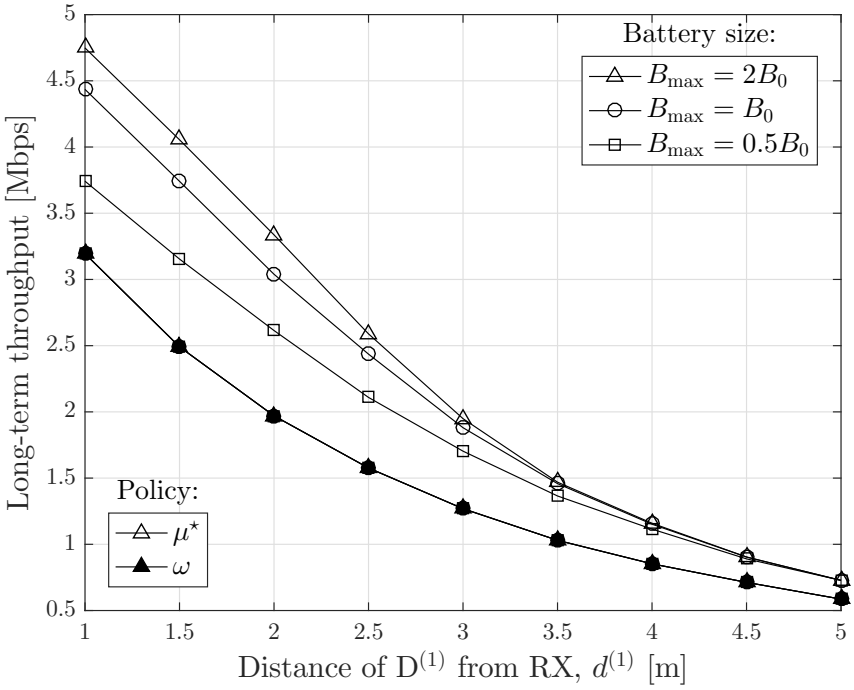

Figure 4.20. Long-term reward of $\mu^\star$ and ω as a function of $d^{(1)}$ with high transmission powers when $d^{(2)} = 1$ m.

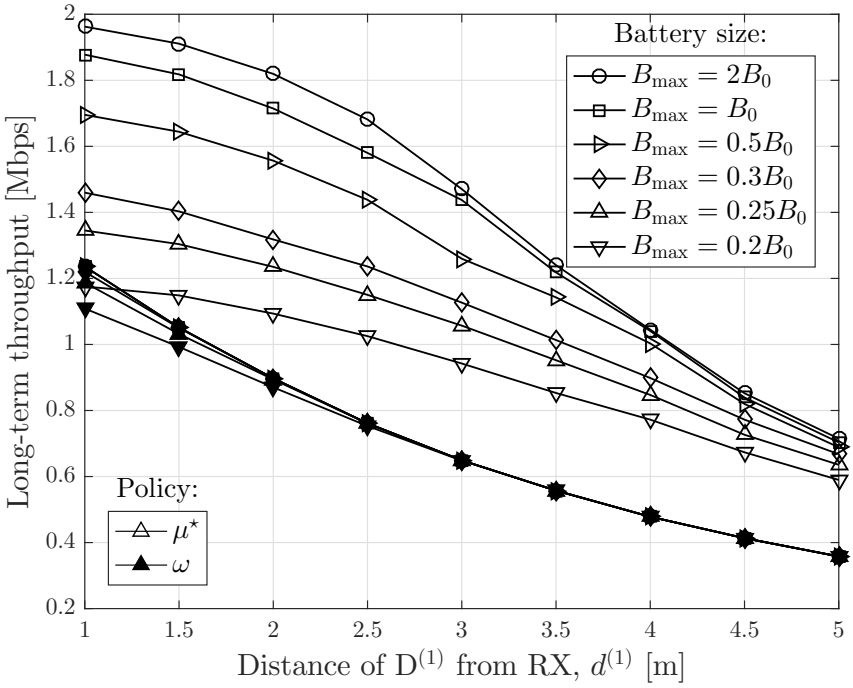

Figure 4.21. Long-term reward of $\mu^\star$ and ω as a function of $d^{(1)}$ with high transmission powers when $d^{(2)} = 3$ m.

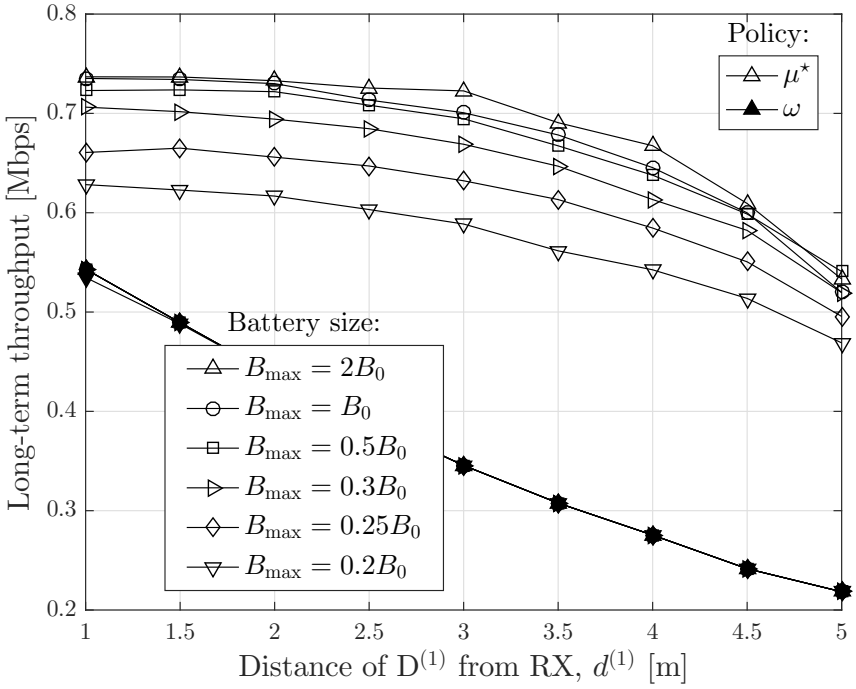

Figure 4.22. Long-term reward of $\mu^\star$ and ω as a function of $d^{(1)}$ with high transmission powers when $d^{(2)} = 5$ m.

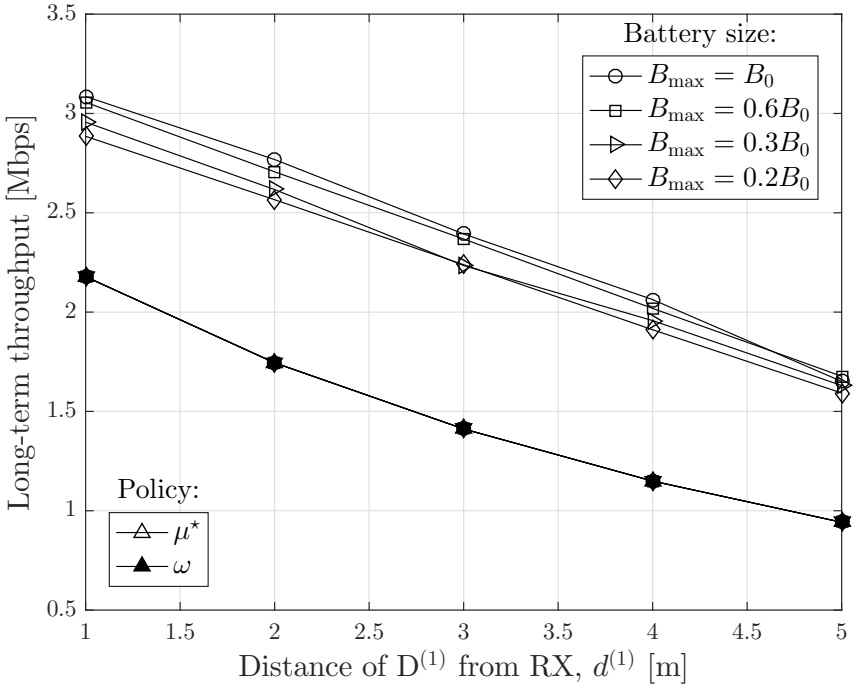

Figure 4.23. Long-term reward of $\mu^\star$ and ω as a function of $d^{(1)}$ with low transmission powers when $d^{(2)} = 3$ m.

Even if $Q_{\mathrm{d}} = \infty$ (i.e., all the remaining energy were lost at the end of the slot), looking at the long term would still provide much better results. Indeed, there is a substantial difference between the two approaches: while our scheme aims at achieving fairness *in the long run*, the slot-oriented scheme achieves fairness in every slot, which may be strongly

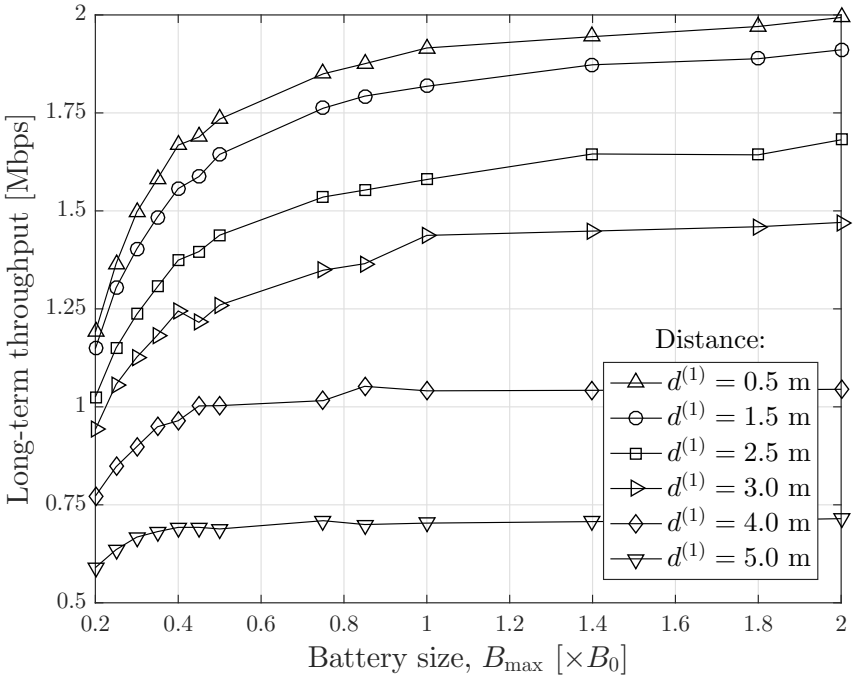

Figure 4.24. Long-term reward of $\mu^\star$ as a function of $B_{\max}$ when $d^{(2)} = 3$ m.

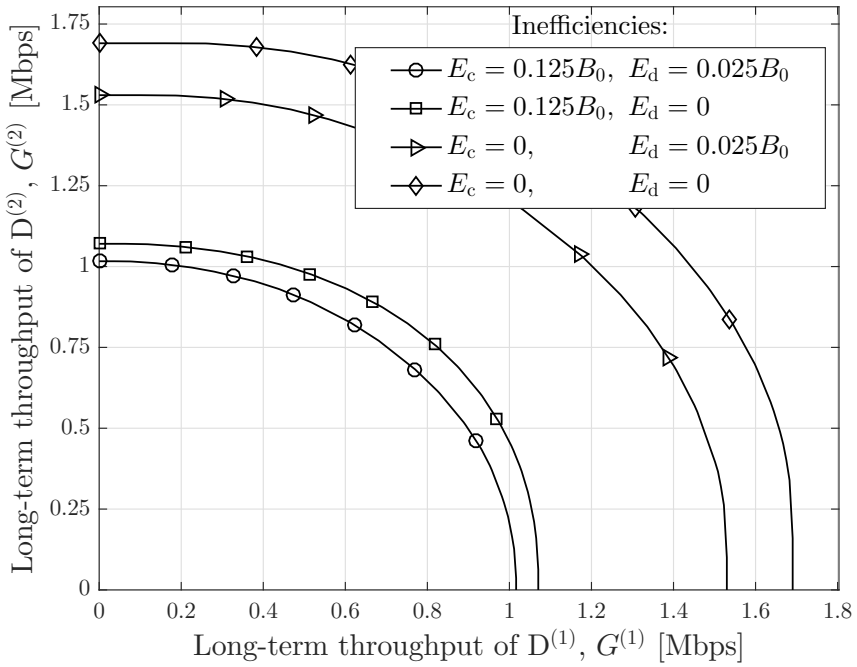

Figure 4.25. Long-term rewards $G^{(1)}$ and $G^{(2)}$ of $\mu^\star$ with and without circuitry costs and battery depletion when $d^{(1)} = d^{(2)} = 3$ m and $B_{\max} = 0.4B_0$.

suboptimal. Instead, the circuitry costs have a much stronger impact on the performance of the system and may push the throughput to zero if ζ is comparable with the battery sizes.

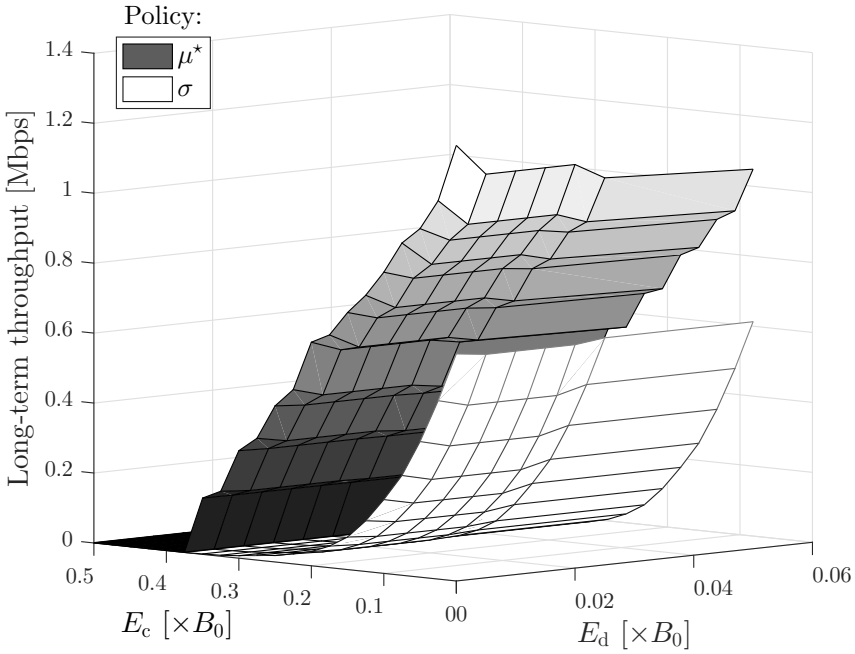

Figure 4.26. Long-term reward of $\mu^\star$ and ω as a function of the circuitry costs $\zeta^{(i)}$ and depletion energy $Q_{\mathrm{d}}^{(i)}$ when $d^{(1)} = d^{(2)} = 3$ m and $B_{\max} = 0.4B_0$.

4.2.8 Conclusions of Section 4.2

In this section we studied the long-term throughput optimization in a wireless powered communication network composed of a receiver and two distributed devices. The system alternates a downlink phase, in which RX recharges the batteries of the nodes via an RF-WET mechanism, and an uplink phase, in which both devices transmit data toward RX in a TDMA fashion. We explained how to solve the long-term throughput maximization problem optimally and approximately while explicitly considering the batteries evolution and the channel state information. We simplified the optimization by exploiting the structure of Bellman's equation. Finally, we compared the long-term approach with the slot-oriented one and noticed that, in terms of achievable performance, the traditional schemes proposed in the literature are strongly suboptimal. The schemes proposed here can be considered as an upper bound to the real performance of a network.

As part of our future work we would like to extend the model of the system in order to consider partial CSI or storage losses, extend the long-term optimization to the case with a generic number of nodes, and compare our results with those obtained using a distributed approach.

Appendix 4.A Proof of Theorem 4.1.2

The energy harvesting mechanism imposes

$$\liminf_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} Q^{(i)}(P_k) \leq \bar{E}^{(i)}.$$

Using Definitions (4.1.5)-(4.1.6) and the hypotheses, we have

$$\begin{aligned}
G_\mu &= \liminf_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} r(P_k) \\
&= \liminf_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} r(\Psi^{i^{-1}}(\Psi^{(i)}(P_k))) \\
&\leq \liminf_{K\to\infty} r\left(\Psi^{i^{-1}}\left(\frac{1}{K}\sum_{k=0}^{K-1}\Psi^{(i)}(P_k)\right)\right) \\
&\leq r\left(\Psi^{i^{-1}}\left(\liminf_{K\to\infty}\frac{1}{K}\sum_{k=0}^{K-1}Q^{(i)}(P_k)\right)\right) \leq r(\Psi^{i^{-1}}(\bar{E}^{(i)})).
\end{aligned}$$

The relation holds for both TX and RX, thus, since we deal with increasing functions, (4.1.8) is obtained. For the last point of the theorem we introduce the following proposition.

Proposition 4.A.1. *If $\Psi^{(i)}(P)$ does not exist, then the battery of device i is infinite.*

Proof. We will equivalently show that if the battery size is finite, then $\Psi^{(i)}(\cdot)$ always exists. Since the battery is finite, the transmission power is bounded by $P_{\max} < \infty$. Function $\Psi^{i}(\cdot)$ can be chosen as a linear function $\Psi^{(i)}(P) = mP$ where m is a slope such that $mP \leq Q^{(i)}(P)$, $\forall P \in \mathcal{P}$. Thus, since $\Psi^{(i)}(\cdot)$ is linear, also its inverse is linear. In this case $r(\Psi^{i^{-1}}(\cdot))$ is concave because $r(\cdot)$ is concave, therefore $\Psi^{(i)}(\cdot)$ can be correctly defined and therefore always exists. ■

Now, assume that both $\Psi^{\text{tx}}(P)$ and $\Psi^{\text{rx}}(P)$ do not exist. This implies that the battery sizes are infinite and in this case $r(Q^{i^{-1}}(P))$ for large P increases faster than P (otherwise $\Psi^{(i)}(P)$ can be found). To show that the reward tends to infinity, consider the following policy over a time horizon of K slots:

$$P_0 = P_1 = \ldots = P_{K-2} = 0, \qquad P_{K-1} = Q^{i^{-1}}\left(\sum_{k=0}^{K-2} E_k^{(i)}\right).$$

The corresponding reward is

$$G_\mu^K = \frac{1}{K}\, r\left(Q^{i^{-1}}\left(\sum_{k=0}^{K-2} E_k^{(i)}\right)\right)$$

and $\lim_{K\to\infty} G_\mu^K = \infty$ because the argument of $Q^{i^{-1}}(\cdot)$ grows linearly in K.

Appendix 4.B Proof of Proposition 4.2.2

The proof is by induction on I. If $I = 0$, Equation (4.2.31) holds because $\mathcal{K}^{[0]} = \widetilde{\mathcal{K}}^{[0]}$. Then, assume that (4.2.31) holds for some I. The inductive step is as follows

$$\|\mathcal{K}^{[I+1]} - \widetilde{\mathcal{K}}^{[I+1]}\|_\infty = \max_{\mathbf{b}\in\mathcal{B}} \left|K_{\mathbf{b}}^{[I+1]} - \widetilde{K}_{\mathbf{b}}^{[I+1]}\right| \tag{4.B.1a}$$

$$\leq \max_{\mathbf{b}\in\mathcal{B}} \left|K_{\mathbf{b}}^{[I+1]} - \mathcal{T}\left(\widetilde{\mathcal{K}}^{[I]}, \mathbf{b}\right)\right| + \max_{\mathbf{b}\in\mathcal{B}} \left|\mathcal{T}\left(\widetilde{\mathcal{K}}^{[I]}, \mathbf{b}\right) - \widetilde{K}_{\mathbf{b}}^{[I+1]}\right| \tag{4.B.1b}$$

$$\leq \max_{\mathbf{b}\in\mathcal{B}} \left|K_{\mathbf{b}}^{[I+1]} - \mathcal{T}\left(\widetilde{\mathcal{K}}^{[I]}, \mathbf{b}\right)\right| + \epsilon \tag{4.B.1c}$$

$$= \max_{\mathbf{b}\in\mathcal{B}} \left|\mathcal{T}\left(\mathcal{K}^{[I]}, \mathbf{b}\right) - \mathcal{T}\left(\widetilde{\mathcal{K}}^{[I]}, \mathbf{b}\right)\right| + \epsilon \tag{4.B.1d}$$

$$\leq \max_{\mathbf{b}\in\mathcal{B}} \left|K_{\mathbf{b}}^{[I]} - \widetilde{K}_{\mathbf{b}}^{[I]}\right| + \epsilon \tag{4.B.1e}$$

$$= \|\mathcal{K}^{[I]} - \widetilde{\mathcal{K}}^{[I]}\|_\infty + \epsilon \leq I\epsilon + \epsilon = [I+1]\epsilon. \tag{4.B.1f}$$

(4.B.1a) holds by definition. (4.B.1b) exploits the triangular inequality. (4.B.1c) uses the hypothesis of the proposition and in particular Definition (4.2.32). (4.B.1d) is by definition of $\mathcal{T}$. (4.B.1e) is formally proved in the next lemma. (4.B.1f) uses the inductive hypothesis. Thus, the proof is concluded by showing the following lemma.

Lemma 4.B.1. *Inequality* (4.B.1d)-(4.B.1e) *holds.*

Proof. Using Definitions (4.2.29)-(4.2.30), we obtain

$$\max_{\mathbf{b}\in\mathcal{B}} \left|\mathcal{T}\left(\mathcal{K}^{[I]}, \mathbf{b}\right) - \mathcal{T}\left(\widetilde{\mathcal{K}}^{[I]}, \mathbf{b}\right)\right| \tag{4.B.2}$$

$$= \max_{\mathbf{b}\in\mathcal{B}} \left| \sum_{\boldsymbol{g},\boldsymbol{\hbar}} f(\boldsymbol{g},\boldsymbol{\hbar}) \Big(\max_{a\in\mathcal{A}_{(\mathbf{b},\boldsymbol{g},\boldsymbol{\hbar})}} \left\{ r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to\mathrm{rx}}|\boldsymbol{\hbar}) + K_{\mathbf{b}'}^{[I]} \right\} - \max_{a\in\mathcal{A}_{(\mathbf{b},\boldsymbol{g},\boldsymbol{\hbar})}} \left\{ r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to\mathrm{rx}}|\boldsymbol{\hbar}) + \widetilde{K}_{\mathbf{b}'}^{[I]} \right\} \Big) \right|, \tag{4.B.3}$$

where we recall that $\mathbf{b}'$ is defined according to Equations (4.2.15)-(4.2.18). Using the triangular inequality, we work as follows

$$(4.B.3) \leq \max_{\mathbf{b}\in\mathcal{B}} \left| \sum_{\boldsymbol{g},\boldsymbol{\hbar}} f(\boldsymbol{g},\boldsymbol{\hbar}) \max_{a\in\mathcal{A}_{(\mathbf{b},\boldsymbol{g},\boldsymbol{\hbar})}} \left\{ r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to\mathrm{rx}}|\boldsymbol{\hbar}) + K_{\mathbf{b}'}^{[I]} - r_\alpha(\boldsymbol{\tau}, \mathbf{P}^{\to\mathrm{rx}}|\boldsymbol{\hbar}) - \widetilde{K}_{\mathbf{b}'}^{[I]} \right\} \right| \tag{4.B.4}$$

$$= \max_{\mathbf{b}\in\mathcal{B}} \left| \sum_{\boldsymbol{g},\boldsymbol{\hbar}} f(\boldsymbol{g},\boldsymbol{\hbar}) \max_{a\in\mathcal{A}_{(\mathbf{b},\boldsymbol{g},\boldsymbol{\hbar})}} \left\{ K_{\mathbf{b}'}^{[I]} - \widetilde{K}_{\mathbf{b}'}^{[I]} \right\} \right|. \tag{4.B.5}$$

Since $\sum_{\boldsymbol{g},\boldsymbol{\hbar}} f(\boldsymbol{g},\boldsymbol{\hbar}) = 1$, we can substitute the sum with a "max" to obtain the upper bound:

$$(4.B.5) \leq \max_{\mathbf{b}\in\mathcal{B}} \max_{\boldsymbol{g},\boldsymbol{\hbar}} \max_{a\in\mathcal{A}_{(\mathbf{b},\boldsymbol{g},\boldsymbol{\hbar})}} \left| K_{\mathbf{b}'}^{[I]} - \widetilde{K}_{\mathbf{b}'}^{[I]} \right|, \tag{4.B.6}$$

which is less than or equal to $\max_{\mathbf{b}\in\mathcal{B}} \left|K_{\mathbf{b}}^{[I]} - \widetilde{K}_{\mathbf{b}}^{[I]}\right|$. ∎

Appendix 4.C Proof of Proposition 4.2.3

First, similarly to Lemma 4.2.1, note that the optimal choice leads to $\tau^{(1)\star} r(P^{(1)\to \mathrm{rx}\star}, \hbar^{(1)}) = \tau^{(2)\star} r(P^{(2)\to \mathrm{rx}\star}, \hbar^{(2)})$. Also, in order not to underuse the available resources, (4.2.34b)-(4.2.34c) are satisfied with equality. Additionally, we have $\tau^{\mathrm{rx}\star}\eta_{\mathrm{ECL}} P^{\mathrm{rx}\to(i)\star} g^{(i)} \leq B^{(i)}_{\max}$, otherwise the transferred energy would be wasted, which is suboptimal because the battery could be equivalently filled by reducing τ^{rx} and increasing $\tau^{(1)}$ and/or $\tau^{(2)}$ (thus leading to a better solution). Based on this result, it also follows that Constraint (4.2.34d) is always satisfied with equality, i.e., all the energy harvested in a slot is also consumed in the same slot. The problem becomes (we define $r^{(i)} \triangleq r(P^{(i)\to \mathrm{rx}}, \hbar^{(i)})$ for notation simplicity)

$$\max_{\boldsymbol{\tau}, \tau^{\mathrm{rx}}, \mathbf{P}^{\to \mathrm{rx}}, \mathbf{P}^{\mathrm{rx}\to}} \tau^{(1)} r^{(1)} = \max_{\boldsymbol{\tau}, \tau^{\mathrm{rx}}, \mathbf{P}^{\to \mathrm{rx}}, \mathbf{P}^{\mathrm{rx}\to}} \tau^{(2)} r^{(2)}, \tag{4.C.1a}$$

s.t.:

$$\tau^{(1)} + \tau^{(2)} + \tau^{\mathrm{rx}} = T, \tag{4.C.1b}$$

$$P^{\mathrm{rx}\to(1)} + P^{\mathrm{rx}\to(2)} = P^{\mathrm{rx}}_{\max}, \tag{4.C.1c}$$

$$\boldsymbol{\zeta} + \boldsymbol{\tau} \circ \mathbf{P}^{\to \mathrm{rx}} = \tau^{\mathrm{rx}} \eta_{\mathrm{ECL}}\, \mathbf{P}^{\mathrm{rx}\to} \circ \boldsymbol{g}, \tag{4.C.1d}$$

$$\tau^{(1)} r^{(1)} = \tau^{(2)} r^{(2)}, \tag{4.C.1e}$$

$$\boldsymbol{\zeta} + \boldsymbol{\tau} \circ \mathbf{P}^{\to \mathrm{rx}} \preceq \mathbf{B}_{\max}, \tag{4.C.1f}$$

$$\boldsymbol{\tau} \succeq 0,\ \tau^{\mathrm{rx}} \geq 0,\ \mathbf{P}^{\to \mathrm{rx}}_{\min} \preceq \mathbf{P}^{\to \mathrm{rx}} \preceq \mathbf{P}^{\to \mathrm{rx}}_{\max},\ \mathbf{P}^{\mathrm{rx}\to} \succeq 0. \tag{4.C.1g}$$

By solving the previous equalities, we can write all the variables as a function of $\mathbf{P}^{\to \mathrm{rx}}$

$$\max_{\mathbf{P}^{\to \mathrm{rx}}} \tau^{(1)} r^{(1)} = \max_{\mathbf{P}^{\to \mathrm{rx}}} \tau^{(2)} r^{(2)}, \tag{4.C.2a}$$

s.t.:

$$\boldsymbol{\zeta} + \boldsymbol{\tau} \circ \mathbf{P}^{\to \mathrm{rx}} \preceq \mathbf{B}_{\max}, \tag{4.C.2b}$$

$$\boldsymbol{\tau} \succeq 0,\ \tau^{\mathrm{rx}} \geq 0,\ \mathbf{P}^{\to \mathrm{rx}}_{\min} \preceq \mathbf{P}^{\to \mathrm{rx}} \preceq \mathbf{P}^{\to \mathrm{rx}}_{\max},\ \mathbf{P}^{\mathrm{rx}\to} \succeq 0, \tag{4.C.2c}$$

where the other parameters are obtained as

$$\tau^{\mathrm{rx}} = \frac{g^{(1)}(\zeta^{(2)}(r^{(1)}+r^{(2)}) + Tr^{(1)}P^{(2)\to \mathrm{rx}}) + g^{(2)}(\zeta^{(1)}(r^{(1)}+r^{(2)}) + Tr^{(2)}P^{(1)\to \mathrm{rx}})}{\eta_{\mathrm{ECL}} g^{(1)} g^{(2)} P^{\mathrm{rx}}_{\max}(r^{(1)}+r^{(2)}) + g^{(1)} r^{(1)} P^{(2)\to \mathrm{rx}} + g^{(2)} r^{(2)} P^{(1)\to \mathrm{rx}}}, \tag{4.C.3}$$

$$\tau^{(1)} = \frac{(T\eta_{\mathrm{ECL}} g^{(1)} g^{(2)} P^{\mathrm{rx}}_{\max} - \zeta^{(2)} g^{(1)} - \zeta^{(1)} g^{(2)}) r^{(2)}}{\eta_{\mathrm{ECL}} g^{(1)} g^{(2)} P^{\mathrm{rx}}_{\max}(r^{(1)}+r^{(2)}) + g^{(1)} r^{(1)} P^{(2)\to \mathrm{rx}} + g^{(2)} r^{(2)} P^{(1)\to \mathrm{rx}}}, \tag{4.C.4}$$

$$P^{\mathrm{rx}\to(1)} = \frac{P^{\mathrm{rx}}_{\max}\eta_{\mathrm{ECL}}(\zeta^{(1)} g^{(2)}(r^{(1)}+r^{(2)}) + g^{(2)} P^{(1)\to \mathrm{rx}} r^{(2)} T) + \zeta^{(1)} P^{(2)} r^{(1)} - \zeta^{(2)} P^{(1)} r^{(2)}}{\eta_{\mathrm{ECL}}\Big(g^{(1)}(\zeta^{(2)}(r^{(1)}+r^{(2)}) + Tr^{(1)}P^{(2)\to \mathrm{rx}}) + g^{(2)}(\zeta^{(1)}(r^{(1)}+r^{(2)}) + Tr^{(2)}P^{(1)\to \mathrm{rx}})\Big)}, \tag{4.C.5}$$

and $\tau^{(2)}$, $P^{\mathrm{rx}\to(2)}$ can be found by switching the subscripts 1 and 2. To solve the problem, we can take the partial derivatives of $\tau^{(i)} r^{(i)}$ over $P^{(1)\to \mathrm{rx}}$ and $P^{(2)\to \mathrm{rx}}$ and set them to zero (see Expression (4.2.35)). Using [58, Lemma 3.2] it can be shown that there exists a unique pair of values $(P^{(1)\to \mathrm{rx}}_0, P^{(2)\to \mathrm{rx}}_0)$ that solves (4.2.35), which then corresponds to the global maximum of $\tau^{(i)} r^{(i)}$. Therefore, if $\mathbf{P}^{\to \mathrm{rx}}_0$ and the corresponding $\boldsymbol{\tau}_0$, $\mathbf{Q}_0$, τ^{rx}_0 obtained with Equations (4.C.3)-(4.C.5) satisfy Constraints (4.C.2b)-(4.C.2c), then $\mathbf{P}^{\to \mathrm{rx}}_0$ is the optimal solution (unique maximum).

Otherwise, the optimal solution must fall on the boundary of the admissible region (since there exists only one stationary point, starting from $\mathbf{P}^{\to \mathrm{rx}}_0$, the reward function

decreases in every direction). In this case, the bottleneck is given by the device which violates some constraint. For example, if (4.C.2b) were violated for device i, then we would impose $\zeta^{(i)} + \tau^{(i)} P^{(i)\to\mathrm{rx}} = B^{(i)}_{\max}$, derive $\tau^{(i)}$ as a function of $P^{(i)\to\mathrm{rx}}$ and solve

$$\max_{P^{(i)\to\mathrm{rx}} \in [P^{(i)\to\mathrm{rx}}_{\min}, P^{(i)\to\mathrm{rx}}_{\max}]} \frac{B^{(i)}_{\max} - \zeta^{(i)}}{P^{(i)\to\mathrm{rx}}} r^{(i)}. \tag{4.C.6}$$

CHAPTER 5

Peculiar Aspects of Energy Harvesting Systems

Traditional Energy Harvesting Devices (EHDs) have been widely studied in previous works [150]. However, despite this trend, some practical aspects have not been properly considered by the literature so far. In particular, the study of non-ideal devices has been conducted only partially; moreover, security and privacy issues have been addressed mostly by neglecting low-power design principles (except possibly for some attempts at limiting the computation and processing costs and/or the number of messages needed to implement a secure protocol). The goal of this chapter is to investigate how the performance and behavior of the networks change when these aspects are taken into account.

Structure of the Chapter. The chapter is divided in two main sections which can be read separately. We also refer the reader to Chapter 2 for more details about the system models. Section 5.1 deals with the inefficiencies of energy harvesting devices. Indeed, while in the literature the main focus is on EHDs with ideal batteries, in reality several inefficiencies have to be considered to correctly design the operating regimes of an EHD. In Section 5.1 we describe how the throughput optimization problem changes under *real battery* constraints. In particular, we consider imperfect knowledge of the state of charge of the battery and storage inefficiencies, i.e., part of the harvested energy is wasted in the battery recharging process. We formulate the problem as a Markov Decision Process (MDP), basing our model on some realistic observations about transmission, consumption and harvesting power. We find the performance upper bound with a real battery and numerically discuss the novelty introduced by the real battery effects. We show that using the *old* policies obtained without considering the real battery effects is strongly suboptimal and may even result in zero throughput.

In Section 5.2, we focus on the security aspects of a wireless communication link with energy constraints from a secrecy rate point of view. The secrecy rate represents the amount of information per unit time that can be securely sent on a communication link. We investigate an energy harvesting communication system composed of a transmitter, a receiver and a malicious eavesdropper. In particular, because of the energy constraints and the channel conditions, it is important to understand when a device should transmit and to optimize how much power should be used in order to improve security. Both full knowledge and partial knowledge of the channel are considered under a Nakagami fading

scenario. We show that high secrecy rates can be obtained only with power *and* coding rate adaptation. Moreover, we highlight the importance of optimally dividing the transmission power in the frequency domain, and note that the optimal scheme provides high gains in secrecy rate over the uniform power splitting case. Analytically, we explain how to find the optimal policy and prove some of its properties. In our numerical evaluation, we discuss how the maximum achievable secrecy rate changes according to the various system parameters. Furthermore, we discuss the effects of a finite battery on the system performance and note that, in order to achieve high secrecy rates, it is not necessary to use very large batteries.

5.1 Battery Imperfections in an Energy Harvesting Device

5.1.1 Introduction

Several aspects in terms of inefficiencies are frequently neglected when designing the optimal policy for energy harvesting devices, e.g., battery degradation [89] (although ideally with Energy Harvesting (EH) the network may operate for an unlimited amount of time, in practice the network lifetime is constrained by hardware failures, e.g., the batteries have a limited number of charging/discharging cycles), energy leakage [34], imperfect knowledge of the State of Charge (SoC) [90], storage losses [43, 149] or circuitry costs [103]. In reality, the battery of an EHD is affected by all these aspects and, in general, the optimal policy obtained without considering them is not necessarily optimal.

Contributions. In this section we redefine the traditional optimization problem of an EHD when the device 1) has only a limited knowledge of its SoC and 2) the battery has storage inefficiencies (possibly dependent upon the current SoC). Even if in the literature these problems were partially studied separately, the combination of the two represents a more realistic case. Moreover, we find that the existing policies for the two separate scenarios are not applicable to the combined case because they would provide very poor performance. To model the system, we use an online approach (see Section 2.2) and focus on the throughput optimization problem. In the definition of the system model we take into account some realistic and practical considerations (for example, using the energy consumption of a real device). We present two heuristic suboptimal policies and discuss how they can be applied to our case. In our numerical evaluation we remark the importance of computing a new optimal policy, explicitly designed for the imperfect SoC case with a real battery.

Structure. The first part of this chapter is organized as follows. Section 5.1.2 defines the system model. The optimization model, the performance upper bound and the suboptimal policies are discussed in Section 5.1.3. Section 5.1.4 presents our numerical results. Section 5.1.5 concludes the section.

References. This section is based on the conference papers **[C4]** and **[C6]**.

5.1.2 System Model

We focus on a single EHD with energy losses in the battery storage process and imperfect knowledge of its SoC.

Time is divided in slots where slot k corresponds to the time interval $[kT, (k+1)T)$. At the beginning of every slot, EHD decides whether to stay idle or become active according

to its SoC (see Section 2.1.1). In the idle phase, the energy consumption is assumed negligible. In the active phase, a stream of bits is sent to the access point during the first $\tau_k^{\text{tx}} < T$ seconds of the slot. We consider a fixed transmission duration, therefore we have $\tau^{\text{tx}} \triangleq \tau_k^{\text{tx}}$, $\forall k$. We assume that τ^{tx} is much lower than T. This is a realistic assumption that will be discussed in more detail in Section 5.1.2.3.[1]

5.1.2.1 State of Charge

The device has the capability of gathering energy from the environment and store it in a finite battery. We model the energy quantities as in Section 2.2.1. In every time slot, the energy stored in the battery is $b_k \in \mathcal{B} \triangleq \{0, \dots, b_{\max}\}$. As in [90], we assume that, in the general case, EHD cannot read all the values of $\mathcal{B}$, but can only observe the partition $\tilde{\mathcal{T}}$ (with $|\tilde{\mathcal{T}}|$ subsets). For example, the node may only know if the energy level is low or high ($|\tilde{\mathcal{T}}| = 2$). In this case, we would have $\tilde{\mathcal{T}} = \{\text{LOW}, \text{HIGH}\}$, with $\text{LOW} \triangleq \{0, \dots, \lfloor b_{\max}/2 \rfloor\}$ and $\text{HIGH} \triangleq \{\lfloor b_{\max}/2 \rfloor + 1, \dots, b_{\max}\}$. When the battery level is $b \in \mathcal{B}$, the node can observe only $\tilde{t} \in \tilde{\mathcal{T}}$, where $\tilde{t}$ is given by $\tilde{t} = \psi(b)$. In the previous example, $\psi(b) = \text{LOW},\ \forall b \leq \lfloor b_{\max}/2 \rfloor$ and $\psi(b) = \text{HIGH},\ \forall b \geq \lfloor b_{\max}/2 \rfloor + 1$.[2]

We focus on the case where $\tilde{\mathcal{T}}$ is a partition of $\mathcal{B}$, but future extensions may include the scenario where different subsets of $\tilde{\mathcal{T}}$ are partially overlapped. This may be useful to study the cases where only a noisy observation of b is available.

Note that it is meaningful to consider the imperfect SoC case because, in reality, the battery level cannot be known with absolute precision but can only be approximated (this strongly depends upon the considered technology). Moreover, we will find that, even restricting the study to the imperfect knowledge case, high performance can be achieved.

5.1.2.2 Harvested and Stored Energy

The energy arrivals are modeled as a i.i.d. discrete random process according to Sections 2.1.3 and 2.2.1. Thus, in slot k, e_k^{EEH} (in this chapter we will write $e_k = e_k^{\text{EEH}}$ for ease of notation) energy quanta arrive at the device, according to some energy arrival statistics with probability mass function (pmf) $\mathbb{P}(e)$, mean $\bar{e}$ and maximum arrivals $e_{\max}$. In an ideal battery (no losses), all the e_k energy quanta can be stored in the battery, if this is not fully charged. Instead, in this section we consider a *real battery* with losses in the energy storage process. Different energy loss models exist, e.g., we can assume that only a fixed fraction $\eta_{\text{Energy Storage Losses}} = \eta_{\text{ESL}}$ of the incoming power can be stored [149], or, more generally, that η_{ESL} depends upon the current state of charge of the battery. This is a realistic assumption, e.g., in a capacitor [43]. In this case, when the state of charge b_k is low or high, then only a small fraction of the incoming energy can be stored, whereas, if $b_k \approx b_{\max}/2$, then almost all the energy can be successfully stored. An example, that was proposed in [43] as an approximation for the storage losses in a capacitor and will be used in our numerical evaluations (however, our results are general and do not depend upon the particular structure of $\eta_{\text{ESL}}(b)$), is

$$\eta_{\text{ESL}}(b) = 1 - \frac{(b - b_{\max}/2)^2}{\phi_{\text{n.l.}}(b_{\max}/2)^2}, \tag{5.1.1}$$

[1] It is also possible to extend the model to the case $\tau^{\text{tx}} \approx T$ by redefining the function y_T of Equation (5.1.2).

[2] Note that the perfect SoC case is obtained when the partition $\tilde{\mathcal{T}}$ has $b_{\max} + 1$ subsets (one for every energy state) and the function $\psi(b) = \{b\}$.

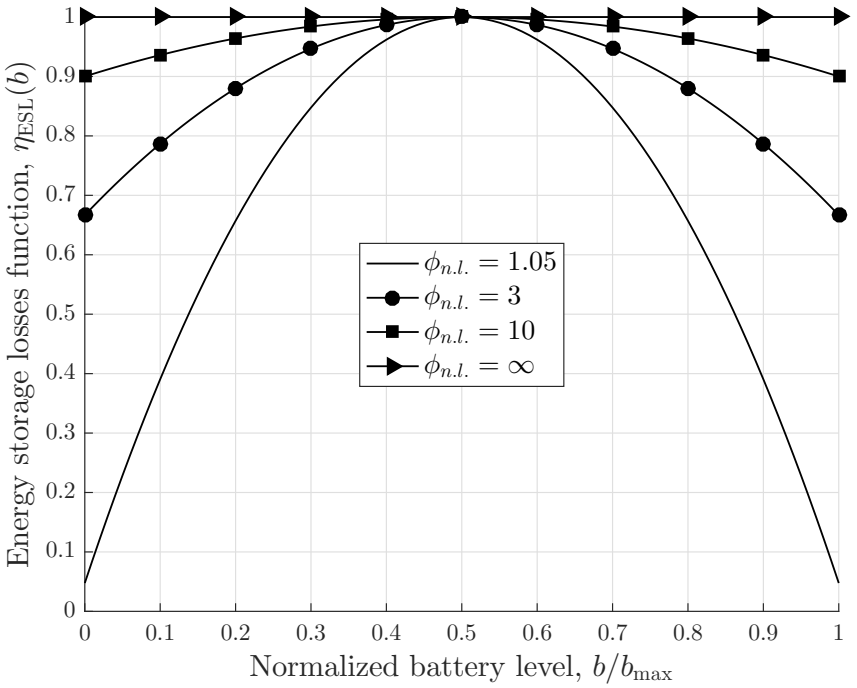

Figure 5.1. Function $\eta_{\mathrm{ESL}}(b)$ (Equation (5.1.1)) as a function of the normalized battery level when $b_{\max} = 100$ (however, the structure of $\eta_{\mathrm{ESL}}(b)$ is only weakly dependent on $b_{\max}$).

where $\phi_{\mathrm{n.l.}} > 1$ is a constant which strictly depends upon the used technology. An example, that we will use as a baseline in this thesis, is $\phi_{\mathrm{n.l.}} = 1.05$ (see [43]). In Figure 5.1 we plot $\eta_{\mathrm{ESL}}(e)$ as a function of the state of charge $b \in \mathcal{B}$, normalized with respect to $b_{\max}$. Note that, as $\phi_{\mathrm{n.l.}}$ increases, the storage losses decrease, until the asymptotic situation $\phi_{\mathrm{n.l.}} \to \infty$ where we obtain an ideal battery with no losses.

Assume that a harvesting power e_k/T is received (constant over a slot). In this case, only a power $\eta_{\mathrm{ESL}}(b)e_k/T$ can be converted and stored in the battery. Note that, at $b = b_{\max}/2$ there are no energy losses, i.e., $\eta_{\mathrm{ESL}}(b) = 1$, whereas when $b \in \{0, b_{\max}\}$, the losses are highest, i.e., $\eta_{\mathrm{ESL}}(b) = 1 - 1/\phi_{\mathrm{n.l.}}$.

We now describe how to evaluate the evolution of the stored energy during a single slot. Equation (5.1.1) is an instantaneous expression that has to be applied only to *powers*. Assume that during a slot the harvested power is constant and equal to e_k/T and the battery is charged at b_k. After $t \leq T$ seconds, the battery status will ideally be

$$y_t = y_t(e_k, b_k) = b_k + \frac{1}{T}\int_0^t e_k \eta_{\mathrm{ESL}}(y_\tau)\mathrm{d}\tau. \tag{5.1.2}$$

At $t = T$ we obtain the battery level after an entire slot, i.e., $b_{k+1} = y_T(e_k, b_k)$. Also, since the battery energy levels are discretized (Section 2.2.1), we approximate $y_T(e_k, b_k)$ with a $Round(\cdot)$ operation. In Equation (5.1.3) we will present the full expression of the battery evolution by also considering other phenomena (overflow and transmission energy).

Note that the procedure to evaluate $y_T(e_k, b_k)$ does not depend upon the particular structure of $\eta_{\mathrm{ESL}}(b)$. The only basic hypothesis that has to be made is that $Round(y_T(e_{\max}, b)) \geq 1$ for every b (otherwise, there may exist a state where recharging

the battery would not be possible). However, it is always possible to satisfy this hypothesis by redefining the notion of energy quantum or the slot length.

5.1.2.3 Energy Consumption

In slot k, a power $P_k \in \mathscr{P}$ (e.g., $\mathscr{P} \triangleq \{0, P_{\min}, \ldots, P_{\max}\}$) is used for transmitting a new stream of bits (we assume that in every slot there is always enough data to be transmitted). Transmitting with a power $P \geq 0$ provides a reward $r(P)$, where $r(\cdot)$ is the *instantaneous reward function*, positive, concave, increasing in P and with $r(0) = 0$. A typical example of $r(P)$, that we also use as a baseline in this section, is $r(P) = \log(1 + \hbar P)$, where $\hbar$ represents an SNR scaling factor [17, 93, 113, 132]. In this case, $r(\cdot)$ represents the transmission rate.

In slot k, the device spends $q(P_k)$ energy quanta to perform the transmission. In order to consider the circuitry costs (e.g., due to inefficiencies and processing), in general $q(P_k)/\tau^{\text{tx}} > P_k$ (i.e., the consumed power is higher than the transmission power). An example of power consumption and corresponding transmission power is given in Table 5.1.

Table 5.1. MSP430 SoC With RF Core [140].

Tx Power (P_k)	**Power Consumption** ($q(P_k)/\tau^{\text{tx}}$)			
	315 MHz	433 MHz	868 MHz	915 MHz
14 mW	79.2 mW	100.2 mW	106.5 mW	104.4 mW
10 mW	75.6 mW	86.4 mW	99.0 mW	96.3 mW
1 mW	43.8 mW	50.4 mW	53.4 mW	52.8 mW
0.25 mW	44.1 mW	52.5 mW	53.4 mW	52.8 mW

The previous table refers to an MSP430 device from Texas Instruments, a microcontroller designed with low power consumption operations and with transmission/reception capabilities. The "Tx Power" column represents, approximately, the power sent into the channel. The other columns describe the overall power consumption in different frequency bands. Note that, in general, $P_k \ll q(P_k)/\tau^{\text{tx}}$, regardless of the operating frequency.

5.1.2.4 Battery Evolution

In order to understand how the battery evolves, we briefly discuss how a slot is structured. During the first τ^{tx} seconds, a stream of bits may be sent. We assume $\tau^{\text{tx}} \ll T$ (i.e., the time required for a transmission is much smaller than the time required to obtain enough energy quanta for a transmission). This can be seen by comparing the values of Table 5.1 and Table 1.1 (copied below for simplicity).

In Table 5.2 we report some common energy sources with the corresponding power density levels. These values should be compared with the power consumption $q(P_k)/\tau^{\text{tx}}$ in Table 5.1. In general, except for the solar light in a sunny day, the harvested power is several orders of magnitude smaller than the transmission power when small devices are considered. Thus, to store enough energy to perform a transmission, a long time is required.

If a device with much smaller power consumption and a rich energy harvesting source were considered, then it might be possible to allow the device to directly send the harvested energy instead of storing it. This is particularly useful in the cases of imperfect real batteries in order to avoid the storage losses [149].

Table 5.2. Energy Sources [120, 123, 127].

Energy Source	**Power Density**
Solar outdoors - sunny day	15 mW/cm^2
Piezoelectric	330 μW/cm^3
Vibrations	200 μW/cm^3
Solar outdoors - cloudy day	150 μW/cm^2
Ambient Radio-Frequency (in London)	6.39 μW/cm^2
Acoustic Noise (at 100 dB)	0.96 μW/cm^3

Because of the presented situation, in our model we assume that the transmission is instantaneous and happens at the beginning of a slot (i.e., τ^{tx} is considered negligible with respect to T). Under this assumption, the value b_k of Equation (5.1.2) is replaced with $b_k - q(P_k)$. Following the previous reasoning, the battery evolution from slot k to $k+1$ can be computed as

$$b_{k+1} = \min\{b_{\max}, y_T(e_k, [b_k - q(P_k)]^+)\}. \tag{5.1.3}$$

Note that we explicitly take into account the effects of a finite battery. $[\cdot]^+ \triangleq \max\{0, \cdot\}$ has to be applied to $b_k - q(P_k)$ because, in the imperfect SoC knowledge case, it may happen that a too high transmission power is selected, and $q(P_k) > b_k$. In this case, we assume that the reward in the corresponding slot is zero (only a partial codeword is transmitted). In summary, the attained reward in slot k is

$$r(P_k) = \begin{cases} r(P_k), & \text{if } q(P_k) \leq b_k, \\ 0, & \text{otherwise.} \end{cases} \tag{5.1.4}$$

5.1.3 Optimization Problem

The system described so far can be modeled with a discrete Markov Chain (MC), whose states correspond to the actual level of charge of the battery (see Section 2.2.2). Since we assume that only partial SoC knowledge is available in general, the device cannot see the state of the system (partially observable MC).

Our goal is to maximize the average long-term reward G_μ obtained with the function $r(\cdot)$:

$$G_\mu(b_0) = \liminf_{K\to\infty} \frac{1}{K} \mathbb{E}\left[\sum_{k=0}^{K-1} r(P_k) \middle| b_0\right], \tag{5.1.5}$$

where the average is taken with respect to the harvesting process and the policy. If $r(\cdot)$ is the transmission rate, then G_μ represents the throughput of the system. We remark that in this case we are maximizing the throughput only over τ^{tx} (i.e., the device is always in idle mode for a fraction $(T - \tau^{\text{tx}})/T$ of the time). μ is the *policy*, i.e., the function that defines which transmission power should be used. b_0 is the initial charge of the battery (in general we fix $b_0 = 0$).

Since we focus on the long-term optimization (see Section 2.1.1), it can be shown [126] that the problem can be formulated by exploiting the MC structure:

$$G_\mu(b_0) = \sum_{b\in\mathcal{B}} \pi_\mu(b|b_0) j_\mu(b), \tag{5.1.6}$$

where

- $\pi_\mu(b|b_0)$ is the long-term probability of being in state b given the initial state of the system b_0.
- $j_\mu(b) \triangleq \mathbb{E}_{\Theta_b}[r(\Theta_b)]$. Θ_b is a random variable with probability mass function $f(P;b)$, $P \in \mathcal{P}$. The policy μ specifies the pmf of Θ_b for every $b \in \mathcal{B}$.

When perfect SoC is available, the problem can be formulated as a Markov Decision Process (MDP) and solved with standard algorithms as described in Section 2.2.2. In this case, the role of μ is to define, for every possible battery level $b \in \mathcal{B}$, the pmf $f(\cdot;b)$. It can be proved that the optimal policy is deterministic (see paper **[C4]** at page 155), i.e., given the state of the system, then $f(P;b) = \chi\{P = P^\star\}$, where $\chi\{\cdot\}$ is the indicator function and $P^\star$ is the optimal transmission power when the state is b.

In the general case (imperfect SoC knowledge), the problem is a Partially Observable Markov Decision Process (POMDP) [14]. This kind of problems can be solved using a belief state as described in [14, Chapter 5]. With this technique, in order to find the optimal policy, in general it is necessary to keep track of the energy arrivals in all past slots. However, this would incur a very high computational cost and, moreover, storing the optimal policy would require a large memory because all possible combinations of the past events have to be taken into account. Because of these, following [90], in this thesis we focus on a suboptimal approach, i.e., we do not consider the history and focus only on the current slot. In this context, the policy μ defines a pmf $f(\cdot;\tilde{t})$ for every partition $\tilde{t} \in \tilde{\mathcal{T}}$. Also, we restrict our study to deterministic policies as in the perfect SoC case for simplicity. Unfortunately, the problem is non-convex, thus standard optimization algorithms cannot be applied. Because of this, in our numerical evaluation, we resort to an exhaustive search (that is computationally affordable when the cardinality of $\mathcal{P}$ and the number of partitions $|\tilde{\mathcal{T}}|$ are not too high).

5.1.3.1 Upper Bound

We want to find an upper bound to the performance in the case $q(P_k) = \tau^{\text{tx}} P_k$ (no losses). When $q(P_k)/\tau^{\text{tx}} > P_k$, the upper bound is still valid but looser.

Assume that e energy quanta arrive. An interesting quantity to evaluate is

$$b^\star = \arg\max_b y_T(e,b), \tag{5.1.7}$$

$$y_T^\star(e) \triangleq y_T(e,b^\star). \tag{5.1.8}$$

In practice, for a given harvesting power e, $y_T^\star(e)$ represents the maximum energy that can be stored in the battery (maximized over all possible values of the current state of charge, b). For example, using Equation (5.1.1), we would have $b^\star < b_{\max}/2$. Instead, with a constant η_{ESL}, we obtain $y_T^\star(e) = \eta_{\text{ESL}} e$ for any b.

Using the previous equations, it is possible to derive an upper bound for the performance.

Proposition 5.1.1. *An upper bound to G_μ is given by*

$$G_{\text{u.b.}} = r\left(\frac{\bar{e}_s}{\tau^{\text{tx}}}\right), \qquad \bar{e}_s \triangleq \sum_{e=0}^{e_{\max}} \mathbb{P}(e) y_T^\star(e), \tag{5.1.9}$$

where $y_T^\star(e)$ is defined in (5.1.8) *and $\mathbb{P}(e)$ is the pmf of the energy arrivals.*

Proof. See Appendix 5.A. ■

Note that the upper bound may depend upon the battery size. For example, focus on Equation (5.1.1). As the battery size increases, the region where $\eta_{\text{ESL}}(b) \approx 1$ increases. Thus, for large batteries, $y_T^\star(e) \approx e$ and the upper bound degenerates in $r(\bar{e}/\tau^{\text{tx}})$.

Instead, if η_{ESL} is constant, the upper bound simply becomes $r(\eta_{\text{ESL}}\bar{e}/\tau^{\text{tx}})$ and does not depend upon the particular battery size.

5.1.3.2 Suboptimal Policies

In our numerical evaluation we will present results for the Optimal Policy (OP). However, in general, computing the optimal policy is a difficult task, especially if the number of partitions $|\tilde{\mathcal{T}}|$ is high (exhaustive search). Because of this, here we introduce some simpler policies that are easier to implement than the optimal one. The first, named Low Complexity Policy (LCP), provides good performance when the battery size is not too large, whereas the second one, named Balanced Policy (BP), works well for large batteries.

LCP is derived from the optimal policy with perfect SoC knowledge, namely OP_{RP} (Real battery and Perfect knowledge), which can be efficiently found solving a linear program. Once the optimal RP policy is computed for every $b \in \mathcal{B}$, we define LCP as follows.

The energy consumption of LCP in the subset $\tilde{t}$ is denoted by $q(P_{\text{LCP}}(\tilde{t}))$, where $P_{\text{LCP}}(\tilde{t})$ represents the transmission power in subset $\tilde{t}$, and is defined as the average energy consumption of RP over all the values $b \in \tilde{t}$, i.e.,[3]

$$q(P_{\text{LCP}}(\tilde{t})) = \frac{1}{|\tilde{t}|} \sum_{b \in \tilde{t}} q(P_{\text{RP}}(b)). \tag{5.1.10}$$

LCP can be seen as the adaptation of OP_{RP} to the imperfect knowledge case. We will show that this approximation is legitimate if the battery is small, but degenerates quickly as the battery size increases. This remarks the importance of using ad hoc policies for the cases of imperfect SoC knowledge with real batteries.

The aim of BP is to consume, on average, a power equal to $\bar{e}_s/\tau^{\text{tx}}$ (Equation (5.1.9)). If this were always possible, then BP would achieve the upper bound. The definition of BP strictly depends upon the number of partitions $|\tilde{\mathcal{T}}|$ and the battery inefficiencies (function $\eta_{\text{ESL}}(b)$). In the case $|\tilde{\mathcal{T}}| = 2$ (LOW-HIGH) and for $\eta_{\text{ESL}}(b)$ defined in Equation (5.1.1) we define BP as $q(P_{\text{BP}}(\text{LOW})) = 0$ and $q(P_{\text{BP}}(\text{HIGH})) = \bar{e}_s$. Note that it is important to set $P_{\text{BP}}(\text{LOW}) = 0$ in order to avoid to be trapped in the low energy states (we will discuss this effect in more detail in the next section).

[3] An alternative definition of LCP would consider a *weighted* mean (e.g., with the steady-state probabilities). However, we have numerically verified that this leads to worse performance.

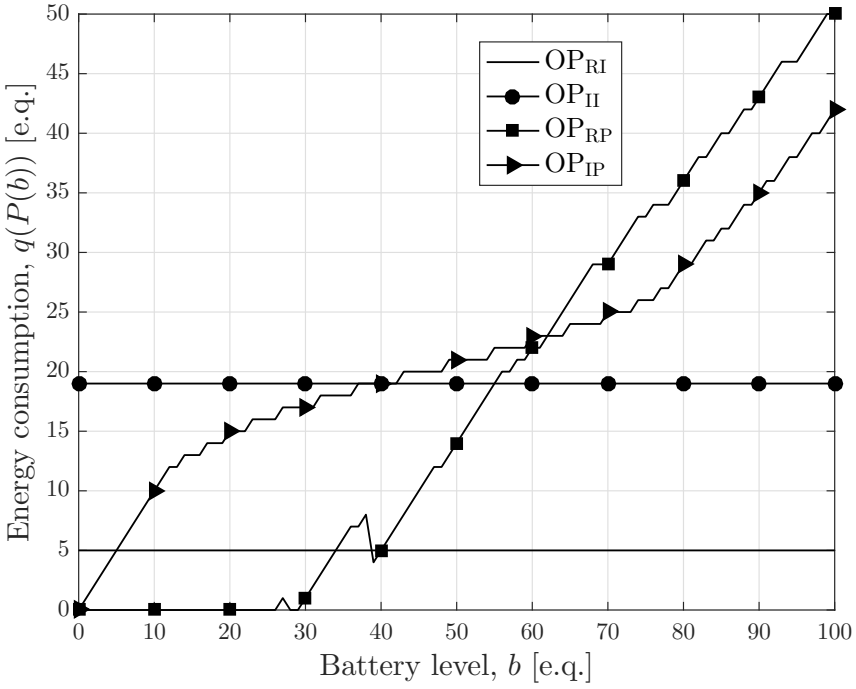

Figure 5.2. Energy consumption $q(P(b))$ (without circuitry costs $q(P(b))/\tau^{\mathrm{tx}} = P(b)$) of the optimal policies in Table 5.3 as a function of the battery level $b \in \mathcal{B}$.

5.1.4 Numerical Results

In our numerical results we discuss the behavior and compare the performance of the policies in Table 5.3.

Table 5.3. Policies.

	Perfect SoC	**Imperfect SoC**
Ideal Battery	$\mathrm{OP_{IP}}$	$\mathrm{OP_{II}}$
Real Battery	$\mathrm{OP_{RP}}$	$\mathrm{OP_{RI}}$, BP, LCP, $\mathrm{OP_{II}}$

We denote with $\mathrm{OP_x}$ the optimal policy in scenario x. Our focus is on the RI case.[4] Here, we also evaluate the performance of $\mathrm{OP_{II}}$ applied to this scenario and of BP and LCP, defined in Section 5.1.3.2.

5.1.4.1 Optimal Policies Structure

We now discuss the shape and the performance of the four optimal policies of Table 5.3. A channel gain $\hbar = 10^{-2}$ is chosen in order to operate in the low SNR regime [132]. Also we set $\tilde{\mathcal{T}} = \{\mathrm{LOW}, \mathrm{HIGH}\}$ ($|\tilde{\mathcal{T}}| = 2$) and neglect the circuitry costs, thus the power consumption in state b is equal to $P(b)$. The storage losses are modeled with Equation (5.1.1). The battery size is set to $b_{\max} = 100$ energy quanta and the energy arrival process is described by a truncated geometric random variable with mean $\bar{e} = 20$ and maximum arrivals $e_{\max} = 50$.

In Figure 5.2 we plot the transmission energy $q(P(b))$ as a function of the SoC b. Note that, because of the battery imperfections, the curve of $\mathrm{OP_{RP}}$ (Real battery with Perfect

[4] $\mathrm{OP_{IP}}$ was studied in [92], $\mathrm{OP_{II}}$ in [90] and $\mathrm{OP_{RP}}$ in **[C4]** (see page 155).

knowledge) is not as smooth as the one for the ideal case $\mathrm{OP}_{\mathrm{IP}}$ (Ideal battery with Perfect knowledge).

Interestingly, for the low energy states, the curves for a real battery are lower than their counterparts for an ideal battery. In practice, when the battery is almost empty, the energy losses are high, thus only a small fraction of the harvested energy can be stored. In this case, it is better not to transmit (or transmit with low power) and wait until the battery reaches a more favorable region.

We now want to highlight the importance of applying the right optimal policy. In particular, we show that applying $\mathrm{OP}_{\mathrm{II}}$ (Ideal battery with Imperfect knowledge) and $\mathrm{OP}_{\mathrm{RP}}$ (Real battery with Perfect knowledge) is suboptimal in the RI case.

5.1.4.2 $\mathrm{OP}_{\mathrm{II}}$ applied to the Case with Real Battery and Imperfect SoC

It can be verified that applying $\mathrm{OP}_{\mathrm{II}}$ to the RI case provides a reward equal to zero whereas $\mathrm{OP}_{\mathrm{RI}}$ provides a reward of 0.1655 (we discuss how the reward changes in Section 5.1.4.4). This result can be explained as follows. In state $b = 0$, the maximum storable energy is given by $y_T(e_{\max}, 0)$. In our example, $e_{\max} = 50$ and, using Equation (5.1.1), it can be verified that $y_T(e_{\max}, 0) = 6.3$, i.e., approximately 6 energy quanta. Starting from $b_0 = 0$, after the first slot we have $b_1 = 6$. In this slot, a transmission energy corresponding to 11 energy quanta is demanded (as shown by the red curve), but, since $b_1 < 11$, only 6 energy quanta are drawn from the battery. However, this results in a failed transmission (the transmission is interrupted at 55%), thus the corresponding reward is zero. In the successive slot, $b_2 = 0$ and the process repeats periodically, providing a global reward equal to zero. Note that this behavior is not influenced by the initial state: even if the battery were initially fully charged, there would exist a positive probability of reaching state 0 and being trapped there. A similar behavior can be noticed in the perfect SoC case, but with less disruptive effects. This example highlights an important characteristics of real batteries, that has to be accurately taken into account when a system is designed.

5.1.4.3 $\mathrm{OP}_{\mathrm{RP}}$ applied to the Case with Real Battery and Imperfect SoC

The optimal policy with perfect knowledge cannot be directly applied to the imperfect SOC case. Therefore we have to resort to the approximation described in Section 5.1.3.2, namely LCP. $\mathrm{OP}_{\mathrm{RP}}$ takes into account the fact that the low energy states should be avoided in order not to trap the battery level. However, by taking the average over every subset, it may happen that, as in the $\mathrm{OP}_{\mathrm{II}}$ case, a too high energy consumption is employed, resulting also in this case in zero reward. This behavior is better explained in Section 5.1.4.4.

While applying a suboptimal policy to the RI case may result in zero reward, $\mathrm{OP}_{\mathrm{RI}}$ takes into account this effect and avoids the inefficient energy states. This is easy to obtain when $|\tilde{\mathcal{T}}|$ is high. However, if $|\tilde{\mathcal{T}}|$ is very low ($|\tilde{\mathcal{T}}| = 1$, i.e., no SoC knowledge) this may result in a significant performance degradation. In Figure 5.3 we plot the transmission energy of $\mathrm{OP}_{\mathrm{RI}}$ when we use one, two or three partitions. Note that in the case $|\tilde{\mathcal{T}}| = 1$, in order to avoid the effect previously described (reward equal to zero if the power consumption is too high), $P(b)$ has to be very low. With $|\tilde{\mathcal{T}}| = 3$, any transmission is avoided for the low energy states in order not to operate in an inefficient battery region. The normalized rewards obtained in the three cases are 0.0488, 0.1655 and 0.1670, respectively. Also, the reward obtained with perfect SoC knowledge is 0.1714. We notice that even a rough idea of the battery SoC (quantization to two levels, LOW and HIGH) provides close-to-optimal performance, and therefore a more accurate representation is not necessary. On the other hand, if no information is provided ($|\tilde{\mathcal{T}}| = 1$) the performance is heavily suboptimal.

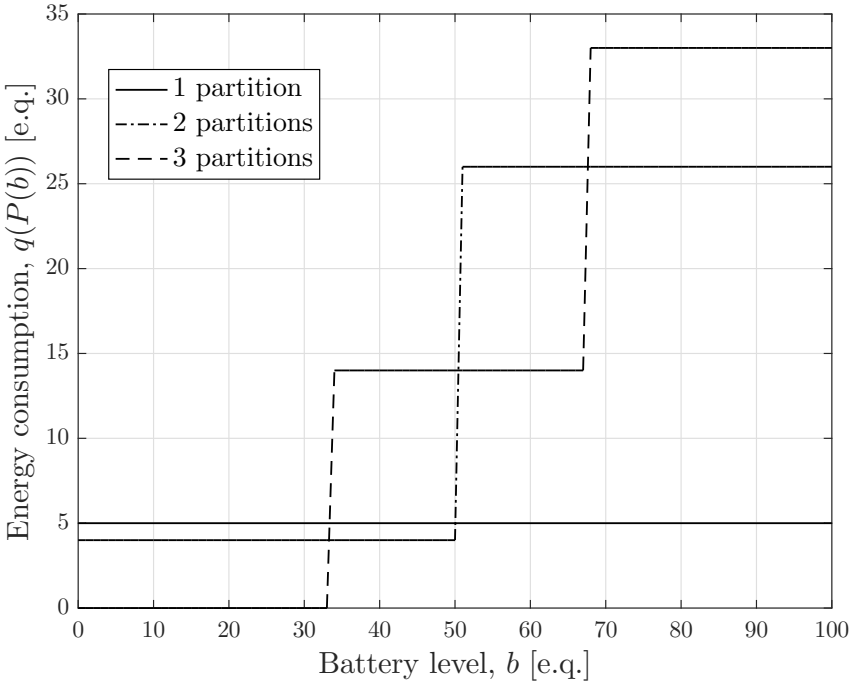

Figure 5.3. Energy consumption $q(P(b))$ (without circuitry costs $q(P(b))/\tau^{\text{tx}} = P(b)$) of OP_{RI} for $|\tilde{\mathcal{T}}| = 1, 2, 3$ as a function of the battery level $b \in \mathcal{B}$.

5.1.4.4 Throughput

Another interesting quantity to analyze is the throughput of the various policies when the battery size increases. In this case (see Figure 5.4) we also want to show the system upper bounds and the performance of LCP and BP. As expected, as the battery size increases, the optimal policy and the upper bound (5.1.9) converge to $r(\bar{e})$ (upper bound with an ideal battery). An interesting point is that $G_{\text{u.b.}}$ is much closer to the performance of the optimal policy than $r(\bar{e})$, especially for low batteries.

We remark that $G_{\text{u.b.}}$ is an upper bound also for the RP (Real battery with Perfect knowledge) case. Thus, since the upper bound and the performance of OP_{RI} are quite close, we can state that it is almost equal to considering $\tilde{\mathcal{T}}$ with a LOW-HIGH subsets, or the perfect knowledge case ($|\tilde{\mathcal{T}}| = b_{\max} + 1$), i.e., even considering $\tilde{\mathcal{T}}$ with very few subsets it is still possible to achieve close to optimal rewards.

The figure also reports the Balanced Policy. Its performance is very close to the reward of OP_{RI}, for both low and high batteries. BP also converges to the upper bound $G_{\text{u.b.}}$ for very large batteries. This is because, on average, BP transmits data with energy $\bar{e}_s$.

Finally, we applied the optimal II (Ideal battery with Imperfect knowledge) and RP (Real battery with Perfect knowledge) policies to the RI scenario. As pointed out in Section 5.1.4.3, in order to apply RP to the RI case, we introduced LCP. As can be seen from the figure, OP_{II} achieves a very low reward that quickly degenerates to zero. Instead, LCP seems to work well for low batteries but as soon as the battery size grows too much, its reward degenerates to zero as well. This further emphasizes the importance of using the proper policy when a real battery with imperfect knowledge is considered.

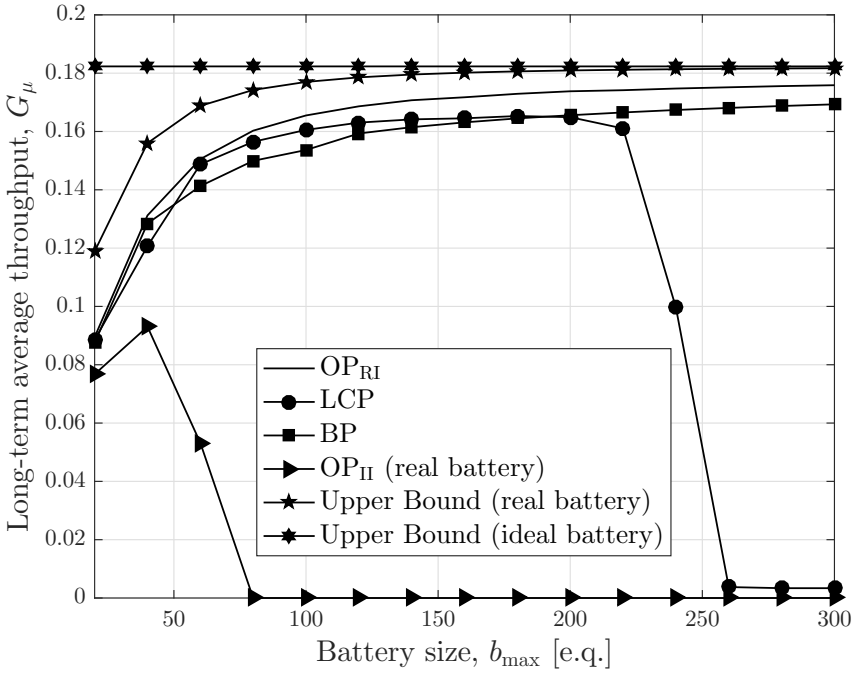

Figure 5.4. Long-term average throughput G_μ (Equation (5.1.6)) as a function of the battery size $b_{\max}$ (with $\bar{e} = 20$).

5.1.4.5 Real Case Example

Table 5.4. Numerical example parameters.

Parameter	**Value**
One energy quantum	10 μJ
Slot length (T)	1 s
Transmission duration (τ^{tx})	5 ms
Available bandwidth (W)	2 MHz
Noise power density (N_0)	$10^{-20.4}$ W/Hz
Channel state ($\hbar$)	$3 \cdot 10^{-13}$ W/Hz
Min/Max power consumption	44.1/106.5 mW (22/53 e.q.)
Min/Mean/Max harvested energy	10/300/500 μJ (1/30/50 e.q.)

We present numerical results for the device reported in Table 5.1. In Table 5.4 we summarize the parameters used for our example.

The considered energy source can be, e.g., piezoelectric or vibrations. In order to model the statistics of the energy source we use a truncated Poisson random variable in order to have a distribution centered around its mean. Note that $T \gg \tau^{\text{tx}}$. The instantaneous reward function is defined according to Shannon's formula

$$r(P) = \frac{\tau^{\text{tx}}}{T} W \log_2 \left(1 + \frac{\hbar}{W N_0} P\right). \tag{5.1.11}$$

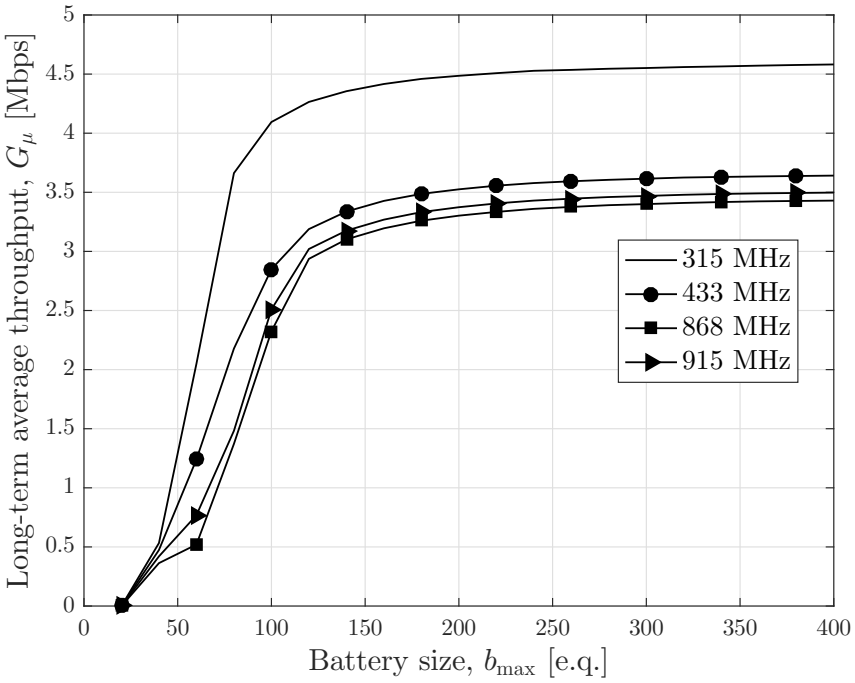

Figure 5.5. Long-term average throughput G_μ (Equation (5.1.6)) as a function of the battery size $b_{\max}$ (with $\bar{e} = 30$) for the various frequency bands of Table 5.1. The bandwidth is $W = 2$ MHz.

The term τ^{tx}/T takes into account the fact that the optimization is performed in the first τ^{tx} seconds of the slot.

We assume that the device fixes the transmission frequency according to the columns of Table 5.1 a priori. In Figure 5.5 we show the throughput for a 2 MHz bandwidth, expressed in kbps, at the four frequencies. At 315 MHz, the throughput is greater because, for every fixed transmission power P, the corresponding power consumption $q(P)/\tau^{\text{tx}}$ is lower.

5.1.5 Conclusions of Section 5.1

In the first part of this chapter, we studied the throughput optimization problem of an energy harvesting device with a Markov decision process approach. We explicitly considered the effects of imperfect batteries and in particular the imperfect SoC knowledge and the energy storage inefficiencies (with losses related to the SoC status). We based our model on some realistic consideration about the transmission/consumption/harvesting power. We found a performance upper bound and showed that, with a real battery, the upper bound and the performance are lower than in the ideal case. We proposed low complexity policies and showed that the balanced policy is a good approximation for the optimal case. We discussed the application of the policies derived in traditional settings and showed that they are strongly suboptimal and that, in some cases, the corresponding throughput could even be zero. This emphasizes the importance of considering the real characteristics of an EHD when the optimization process is performed.

As part of the future work we would like to set up a model where also an ultra-low power device can be studied, i.e., τ and T are similar. Also, other model imperfections can be included, e.g., energy leakage or battery degradation.

5.2 Physical Layer Security

5.2.1 Introduction

Security and privacy are becoming more and more important in communications and networking systems, and have key applications in the Wireless Sensor Network (WSN) and Internet of Things (IoT) worlds [82]. While most works in this area deal with security *protocols* [20, 114], implementing security mechanisms at the physical layer represents an interesting complement to those networking approaches [18], and has the potential to provide stronger (information-theoretic) secrecy properties [131]. In particular, the impact of power allocation policies and of system features related to energy harvesting has only been studied in some special cases [5, 102]. Since green aspects will play an increasingly large role in future networks, it is essential to bring low-power, energy-constrained and green considerations into this picture. In this section, we try to partly fill this gap, studying how the use of energy harvesting affects the design and performance of physical layer security methods.

We consider an Energy Harvesting Device (EHD) (i.e., a device with the capability of gathering energy from the environment [45], e.g., through a solar panel or a rectenna) that sends data to a receiver over an insecure communication channel. The goal is to transmit data securely, i.e., in such a way that an adversary (or *eavesdropper*) with access to the communication link is not able to gather useful information about the data sent. We study how the specific EH characteristics influence the achievable *secrecy rate* (i.e., the information rate at which the EHD can reliably send data to the receiver while keeping it secret from the eavesdropper). Deciding whether the EHD should transmit or not, how much power should be transmitted or how to divide the power among the different sub-carriers is not obvious, and all these aspects need to be appropriately optimized. Moreover, while in the classic throughput optimization problem if the available resources were used improperly the corresponding penalty would be a performance reduction, in the secrecy optimization problem an improper use of the resources may imply not only a reduced transmission rate, but also a security loss, possibly making sensitive data accessible to a malicious party.

Security aspects have been widely studied in the WSN literature [82, 114, 119]. Examples of relevant applications in a WSN/IoT context include health-care monitoring [3, 4], where the sensitive data of patients may be exposed to a malicious entity, or military use [151, 157], where a WSN can be used for monitoring or tracking enemy forces. In particular, in addition to higher layers [172], that are relatively insensitive to the physical characteristics of the wireless medium, physical layer can be used to strengthen the security of digital communication systems and improve already existing security measures. The basic idea behind the concept of physical layer secrecy is to exploit the randomness of the communication channel to limit the information that can be gathered by the eavesdropper at the signal level. Through channel coding techniques, it is possible to simultaneously allow the legitimate receiver to correctly decode a packet and prevent a potential third party malicious eavesdropper from decoding it and thus provide information-theoretic or *unconditional* security. Differently from computational security methods, that are based on the limited computational capabilities of the adversary (as in a cryptographic system), unconditional security is considered the strongest notion of security [88] because no limits on the adversary's computing power are assumed. Perfect secrecy [131] is achieved when there is zero mutual information between the information signal, s, and the signal received

by the eavesdropper, z, i.e., $I(s;z) = 0$ and z is useless when trying to determine s. In [158], Wyner showed that if the eavesdropper's channel is degraded with respect to the legitimate channel, then it is possible to exchange secure information at a non-zero rate while keeping the information leakage to the eavesdropper at a vanishing rate. This result was extended in [27] for non-degraded channels provided the eavesdropper channel is not less noisy than the legitimate channel. In [154], the secrecy capacity of fading channels in the presence of multiple eavesdroppers is studied. It was shown in [42] that in a fading scenario it is also possible to obtain a non-zero secure rate even if, on average, the eavesdropper's channel is better than the legitimate one. The authors also established the importance of variable rate coding (i.e., matching the code rate to the channel rate) in enabling secure communications. In [108], the authors compute the secrecy capacity of a MIMO wiretap channel with one receiver and one eavesdropper and an arbitrary number of antennas. A survey of physical layer security in modern networks is presented in [97].

The secrecy capacity paradigm in an energy harvesting communication system was studied in [109, 110], where the authors considered the case of a batteryless transmitter and found the rate-equivocation region. [98] studied the deployment of an energy harvesting cooperative jammer to increase physical layer security. In [102] the authors presented a resource allocation algorithm for a multiple-input single-output secrecy system for a communication system based on RF energy harvesting. Also [166] studied how to efficiently allocate power over several sub-carriers in an EH system with secrecy constraints. In [77] the authors employed a physical layer secrecy approach in a system with a transmitter that sends confidential messages to a receiver and transfers wireless energy to energy harvesting receivers. Our focus is substantially different from those, since we consider an EHD that harvests energy from an external, non-controllable and renewable energy source. Our goal is to maximize the achievable secrecy rate, i.e., to define how to correctly exploit the available (random) energy according to the device battery dynamics.

Contributions. Our main contribution lies in the definition of a new practical and challenging problem. As in [109, 110], we investigate the physical layer secrecy in an EH system. However, differently from those papers, we explicitly consider the effects of a finite battery and we focus on finding the transmission strategy that maximizes the secrecy rate, namely the Optimal Secrecy Policy (OSP). Since in a WSN the devices operate under the same conditions for long periods, the steady-state regime is generally reached, and thus we focus on the long-term optimization. Similarly to [93], we set up an optimization problem based on a Markov Decision Process (MDP) approach but, unlike in those works, we focus on the security aspects, considering the presence of a malicious eavesdropper and a generic number of sub-carriers. Thus, even if the proposed analytical framework is similar to those provided in the literature, since additional dimensions are considered, the optimization process is more challenging and different considerations and insights are derived. In particular, we prove several properties of OSP and describe a technique to compute it by decomposing the problem into two steps. We specify how to allocate the power over the different sub-carriers and remark that a smart power splitting scheme is important to achieve high secrecy rates. As in [42], we consider several degrees of knowledge of the channel state information, describing both variable and constant rate coding techniques and discussing how the achievable secrecy rate changes in these cases. However, unlike [42], we study an energy constrained system with M parallel sub-carriers, and accordingly formulate and solve an optimization problem to determine the maximum secrecy rate. Therefore, this section considers aspects that either have not yet been considered or have

been separately studied in the literature, and represents an advancement of the state of the art in the important areas of green networking and security, leading to novel insights about the interaction of many different system design aspects.

Structure. The second part of this chapter is organized as follows. Section 5.2.2 defines the system model we analyze and introduces the notion of secrecy rate. In Section 5.2.3 we introduce the secrecy rate optimization problem. Section 5.2.4 describes how to find OSP and some of its properties with full CSI. In Section 5.2.5 we study the case of imperfect CSI knowledge. Section 5.2.6 presents our numerical results. Finally, Section 5.2.7 concludes the chapter.

References. This section is based on the conference paper **[C5]** and on the journal paper **[J2]**.

5.2.2 System Model and Secrecy Rate

An EHD simultaneously transmits data in a wide frequency band composed of M narrow bands to RX. The transmission power can be different for every sub-carrier. The transmission model can be described as a set of M parallel Gaussian wiretap channels, affected by independent fading, as in [12]. The goal of the transmitter is to send data to the legitimate receiver (i.e., the receiver) with a positive secrecy rate in order to guarantee secure transmission. An eavesdropper attempts to intercept the transmitted data.

We initially assume that the EHD knows the Channel State Information (CSI) of all the sub-carriers toward the receiver and the eavesdropper instantaneously, and will relax this hypothesis in Section 5.2.5. Time is divided into slots of equal duration T, chosen according to the channel coherence time, in order to guarantee constant channel gains in every slot. The EHD is equipped with a battery of finite size $b_{\max}$ and in slot k the device has $b_k \in \mathcal{B} \triangleq \{0, \dots, b_{\max}\}$ energy quanta stored. Knowledge of the state of charge is useful at the transmitter side only to determine when to schedule a transmission. The harvesting process is described through an energy quanta arrival process $\{e_k^{\mathrm{EEH}}\}$ (see Sections 2.1.2 and 2.2.1), or simply $\{e_k\}$ in this section. The average harvesting rate is $\bar{e}$, the maximum (minimum) number of energy quanta harvested per slot is $e_{\max}$ ($e_{\min}$), and a quantum harvested in slot k can only be used in time slots $> k$. We assume that the device always has data to send and that the energy cost that the device sustains is mainly due to data transmission (Section 2.1.3).

The channel gains in slot k are $\boldsymbol{\hbar}_k = [\hbar_k^{(1)}, \dots, \hbar_k^{(M)}]$ and $\boldsymbol{\jmath}_k = [\jmath_k^{(1)}, \dots, \jmath_k^{(M)}]$ for the M legitimate and eavesdropper sub-carriers, respectively. $\boldsymbol{\hbar}_k$ and $\boldsymbol{\jmath}_k$ can be interpreted as realizations of two jointly random vectors $\boldsymbol{H} = [H^{(1)}, \dots, H^{(M)}]$ and $\boldsymbol{J} = [J^{(1)}, \dots, J^{(M)}]$ (i.i.d. over time) with supports $\mathcal{H}$ and $\mathcal{J}$. We assume that the receiver has complete CSI of its channel in order to decode the received signal. Instead, the eavesdropper has knowledge on every aspect of the system (this is a reasonable worst-case assumption, as the transmission strategy should not rely on assuming the eavesdropper's ignorance of any state). Nevertheless, we should point out that, for a passive eavesdropper, knowledge of the main channel state is totally immaterial. In this section, when we refer to "full" or "partial" CSI, we always refer to the transmitter side.

5.2.2.1 Secrecy Rates and Capacity

We refer to the notions of *secrecy rate* and *secrecy capacity* as known in the physical layer secrecy literature [18, 158] and their ergodic counterparts in the fading scenario [78]. Specifically, we define an $(\tilde{M}, M, \ell)$ code for the parallel wiretap channel as consisting of:

1) a message set $\mathcal{M}$ with cardinality $\tilde{M}$, 2) a probabilistic encoder f_ℓ^{enc} at the transmitter that maps each message $s \in \mathcal{M}$ (realization of the r.v. S) to each $M \times \ell$ codeword $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}^\ell$, with $\boldsymbol{\mathcal{X}} = \mathcal{X}^{(1)} \times \cdots \times \mathcal{X}^{(M)}$ according to some conditional distribution $p(\boldsymbol{x}|s)$, and 3) a (deterministic) decoder at the legitimate receiver that extracts $\hat{s}$ (realization of the r.v. $\hat{S}$) from the received message $\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}^\ell$, where $\boldsymbol{\mathcal{Y}} = \mathcal{Y}^{(1)} \times \cdots \times \mathcal{Y}^{(M)}$, i.e., $f_\ell^{\text{dec}} : \mathcal{Y}^\ell \to \mathcal{M}$.

The average error probability of an $(\tilde{M}, M, \ell)$ code is given by

$$P_{\text{err}}(\ell) \triangleq \frac{1}{\tilde{M}} \sum_{s \in \mathcal{M}} P\Big(\hat{S} \neq s | S = s\Big). \tag{5.2.1}$$

The equivocation rate at the eavesdropper is $R_e(\ell) = (1/\ell)\,\mathrm{H}(S|U^\ell)$, i.e., the conditional entropy rate of the transmitted message given the eavesdropper's channel output U^ℓ. $R_e(\ell)$ represents the level of ignorance on the target secret message at the eavesdropper. Perfect secrecy (unconditional security) would be obtained if $R_e(\ell) = R(\ell)$, where $R(\ell) = (1/\ell)\,\mathrm{H}(S)$ is the secret message rate. However, this is not possible in general with wiretap coding techniques, so we must settle for a weaker requirement, that holds asymptotically. Therefore, a *secrecy rate* R_s is said to be achievable if there exists a sequence of $(2^{\ell R_s}, M, \ell)$ codes, $\ell = 1, 2, \ldots$, such that

$$\lim_{\ell \to \infty} P_{\text{err}}(\ell) = 0, \qquad R_s \leq \lim_{\ell \to \infty} R_e(\ell) \tag{5.2.2}$$

and the secrecy capacity is defined as the supremum of the set of achievable secrecy rates.

5.2.2.2 Coding Strategy

The transmitter coding strategy influences the secrecy rate. In particular, in this section we consider *constant* and *variable* rate coding defined as follows (a construction procedure for these codes can be derived as explained in [42, Theorems 1 and 2]).

Variable rate coding. It consists in adapting the code rate to the main channel state. This can be accomplished by constructing a separate codeword x for every realization of the channel, i.e., $x = x(\text{current channel})$. In this case, in every slot k and on every sub-carrier $r = 1, \ldots, M$ the transmitter observes the channel and picks the symbols to be transmitted from the current codeword $x(\hbar_k^{(r)})$. We study the long-term regime and thus we consider the case of infinite length codewords. With variable rate coding, when the gain of the legitimate channel in a given sub-carrier is $\hbar$, the transmitter uses symbols from codewords at a rate proportional to $\log(1 + \hbar P)$ (where P is the transmission power, which will be the objective of our optimization). To achieve such a rate, it is required *to use a codeword specifically designed for this channel*, i.e., $x(\hbar)$. Then, if the eavesdropper's channel gain is $j > \hbar$, thanks to the chosen coding rate, the mutual information between the transmitter and the eavesdropper is upper-bounded by $\log(1 + \hbar P)$. Instead, when $j \leq \hbar$, the mutual information becomes $\log(1 + jP)$ (Shannon's theorem). We can summarize the two previous cases as $\log(1 + \min\{\hbar, j\}P)$. Therefore, even if $j > \hbar$, the eavesdropper does not receive more information than the legitimate receiver (they both experience the same rate $\log(1 + \hbar P)$). In the long run, the average rate of the main channel and the information accumulated at the eavesdropper are

$$\liminf_{K \to \infty} \frac{1}{K} \sum_{k=0}^{K-1} \sum_{r=1}^{M} \log(1 + \hbar_k^{(r)} P) \tag{5.2.3}$$

and

$$\liminf_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} \sum_{r=1}^{M} \log(1 + \min\{\hbar_k^{(r)}, \jmath_k^{(r)}\}P), \tag{5.2.4}$$

respectively. In this case, by constructing a code and the corresponding coding map, the long-term secrecy rate (amount of secret information that can be sent) is

$$\liminf_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} \sum_{r=1}^{M} \Big(\log(1 + \hbar_k^{(r)} P) - \log(1 + \min\{\hbar_k^{(r)}, \jmath_k^{(r)}\}P) \Big). \tag{5.2.5}$$

Constant rate coding. It consists in keeping the code rate constant, regardless of the legitimate and eavesdropper's channel states. In this case, a single codeword x is used in every fading condition. In every slot, the transmitter picks the symbols to be transmitted from the only available codeword x. In the long run, since we consider infinite length codewords, x spans the entire fading statistic of the channel. With constant rate coding, regardless of the current channel state, the transmitter uses codewords at a fixed rate R_{con} such that $R_{\text{con}} \geq \log(1 + \hbar P)$ for every $\hbar$ and P. In this case, if the current legitimate channel is $\hbar$, the mutual information between transmitter and receiver is upper bounded by Shannon's theorem as $\log(1+\hbar P)$. Similarly, the mutual information between transmitter and eavesdropper is given by $\log(1 + \jmath P)$. The secrecy rate can be expressed as

$$\left[\liminf_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} \sum_{r=1}^{M} \Big(\log(1 + \hbar_k^{(r)} P) - \log(1 + \jmath_k^{(r)} P) \Big) \right]^+, \tag{5.2.6}$$

where $[\cdot]^+ \triangleq \max\{0, \cdot\}$ is used to obtain a non-negative rate. Note that (5.2.6) is lower than (or equal to) (5.2.5), i.e., higher secrecy is achieved with variable rate coding. However, its implementation is more difficult as the code rate has to be changed frequently according to the legitimate channel state.

For simplicity, in the next we use $R_{\hbar,\jmath}(P)$ to indicate the terms of the sum in (5.2.5) if variable rate coding is considered, or (5.2.6) in the constant rate coding case, i.e.,

$$R_{\hbar,\jmath}(P) \triangleq \begin{cases} \log(1 + \hbar P) - \log(1 + \min\{\hbar, \jmath\}P), & \text{variable rate,} \\ \log(1 + \hbar P) - \log(1 + \jmath P), & \text{constant rate.} \end{cases} \tag{5.2.7}$$

$c(\boldsymbol{P}, \boldsymbol{\hbar}, \boldsymbol{\jmath})$ is its generalization with a generic number of sub-carriers M:

$$c(\boldsymbol{P}, \boldsymbol{\hbar}, \boldsymbol{\jmath}) = \sum_{r=1}^{M} R_{\hbar^{(r)},\jmath^{(r)}}(P^{(r)}), \tag{5.2.8}$$

and P^{tot} is the corresponding total transmission power, defined as

$$P^{\text{tot}} \triangleq \mathbf{1}_M^T \boldsymbol{P}. \tag{5.2.9}$$

The value of $c(\boldsymbol{P}, \boldsymbol{\hbar}, \boldsymbol{\jmath})$ depends on the choice of the power allocation over the several sub-carriers, $\boldsymbol{P} \triangleq [P_1, \ldots, P_M]^T$, the channel conditions $\boldsymbol{\hbar}$ and $\boldsymbol{\jmath}$, and the coding rate

strategy. $\mathbf{1}_M$ is a column vector consisting of M ones. In the general case, the choice of $\boldsymbol{P}$ that maximizes the secrecy rate, among those satisfying (5.2.9), will in turn depend upon the channel conditions $\boldsymbol{\hbar}$ and $\boldsymbol{\jmath}$.

5.2.3 Optimization Problem

The system state $\mathbf{S}_k$ in time slot k is defined by the $(2M+1)$-tuple $(b_k, \boldsymbol{\hbar}_k, \boldsymbol{\jmath}_k)$. A policy μ is a set of rules that, given the state of the system, specifies the power allocation over the M sub-carriers.

In the long run, the average *secrecy rate* under a policy μ is given by the average undiscounted reward C_μ

$$C_\mu(b_0) \triangleq \left[\liminf_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} c(\boldsymbol{\Sigma}_k, \boldsymbol{\hbar}_k, \boldsymbol{\jmath}_k)\right]^+, \tag{5.2.10}$$

where $c(\cdot,\cdot,\cdot)$ is the instantaneous partial contribution defined in (5.2.8), $\boldsymbol{\Sigma}_k$ is the power allocation vector defined by the policy[5] and b_0 is the energy in the initial time slot. A secure communication can be performed if $C_\mu(b_0) > 0$. (5.2.10) is a generalization of (5.2.5) and (5.2.6) for M sub-carriers and a dynamic transmission power.

The battery evolution of Equation (2.2.1) can be modified as follows (note that in this case we do not consider wireless energy transfer nor transmission inefficiencies)

$$b_{k+1} = \min\left\{b_{\max}, b_k - \sum_{r=1}^{M} \Sigma_k^{(r)} + E_k\right\}, \tag{5.2.11}$$

where $\Sigma_k^{(r)}$ is the r^{th} component of the vector $\boldsymbol{\Sigma}_k$, and the "min" is used to account for the finite battery. Note that $\boldsymbol{\Sigma}_k$ must satisfy $\sum_{r=1}^{M} \Sigma_k^{(r)} \leq b_k,\ \forall k$ and $\Sigma_k^{(r)} \geq 0,\ \forall k,\ \forall r$. Thus, Problem (5.2.10) is implicitly influenced by the evolution of b_k because of $\boldsymbol{\Sigma}_k$.

Our aim is to solve the following maximization problem

$$\mu^\star = \arg\max_{\mu} C_\mu(b_0). \tag{5.2.12}$$

A policy that solves (5.2.12) is an Optimal Secrecy Policy (OSP). In the next subsection we explain in more detail the optimization variables and the constraints of the above problem.

5.2.3.1 Markov Decision Process Formulation

Since we consider a *long-term* optimization, we recast the problem using a Markov Decision Process (MDP) formulation. In particular, we model our system by a Markov Chain (MC) with a finite number of states. For every MC state $\mathbf{s} = (b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) \in \mathcal{S} \triangleq \mathcal{B} \times \mathcal{H} \times \mathcal{J}$, a *power allocation policy* μ is the set of rules

$$\mu = \{\mu(\cdot; \mathbf{s}),\ \forall \mathbf{s} \in \mathcal{S}\}, \tag{5.2.13}$$

[5]Given a temporal sequence of energy arrivals and channel states, the policy μ can be applied to obtain the power allocation vector $\boldsymbol{\Sigma}_k$. In this case we use a deterministic policy for presentation simplicity, and prove later that this choice is optimal.

where $\mu(\cdot;\mathbf{s})$ is the conditional distribution (pmf) of the power allocation vector defined as follows

$$\mu(\boldsymbol{P};\mathbf{s})\Big(=\mu(\boldsymbol{P};b,\boldsymbol{\hbar},\boldsymbol{\jmath})\Big)\triangleq P\left(\text{using a power splitting vector }\boldsymbol{P}\,\middle|\,\text{the MC state }\mathbf{s}\right), \tag{5.2.14}$$

and, for every $\boldsymbol{\hbar}$, $\boldsymbol{\jmath}$, is subject to

$$\sum_{\boldsymbol{P}\in\mathscr{P}_{\leq}(b)}\mu(\boldsymbol{P};\mathbf{s})=1, \tag{5.2.15a}$$

$$\mu(\boldsymbol{P};\mathbf{s})\geq 0,\qquad \forall\boldsymbol{P}\in\mathscr{P}_{\leq}(b), \tag{5.2.15b}$$

$$\mathscr{P}_{\leq}(b)\triangleq\left\{\boldsymbol{P}\ :\ \boldsymbol{P}\succeq 0\ \cap\ P^{\text{tot}}\triangleq\mathbf{1}_M^T\boldsymbol{P}\leq b\right\}, \tag{5.2.15c}$$

$$\mathbf{s}=(b,\boldsymbol{\hbar},\boldsymbol{\jmath}). \tag{5.2.15d}$$

$\mathscr{P}_{\leq}(b)$ is the set of all feasible vectors $\boldsymbol{P}$ when the energy level is b. The reward function becomes

$$C_\mu(b_0)=\sum_{b\in\mathcal{B}}\pi_\mu(b|b_0)\times\int_{\mathcal{H}\times\mathcal{J}}\underbrace{\sum_{\boldsymbol{P}\in\mathscr{P}_{\leq}(b)}c(\boldsymbol{P},\boldsymbol{\hbar},\boldsymbol{\jmath})\mu(\boldsymbol{P};b,\boldsymbol{\hbar},\boldsymbol{\jmath})}_{\text{secrecy rate given the MC state }(b,\boldsymbol{\hbar},\boldsymbol{\jmath})}\ \mathrm{d}F(\boldsymbol{\hbar},\boldsymbol{\jmath}), \tag{5.2.16}$$

where $\pi_\mu(b|b_0)\in[0,1]$ is the steady-state probability of having b energy quanta stored starting from state b_0 under a policy μ and $F(\boldsymbol{\hbar},\boldsymbol{\jmath})$ is the joint cumulative distribution function of $\boldsymbol{H}$ and $\boldsymbol{J}$. $\pi_\mu(b|b_0)$ summarizes the battery evolution and is evaluated according to (5.2.11). The optimization variables in Problem (5.2.12) are the pmfs $\mu(\cdot;b,\boldsymbol{\hbar},\boldsymbol{\jmath})$. Also, it can be shown (see Section 5.2.4.1) that an OSP which admits steady-state distribution always exists. Therefore, without loss of optimality, we decided to restrict our study to the class of policies with steady-state distribution. For these policies, since we focus on the average long-term optimization, (5.2.16) is equivalent to (5.2.10).

It is possible to separate μ into the product of a *transmit power policy*, which specifies the conditional distribution of the total transmission power given the current state, namely $\gamma_\mu(P^{\text{tot}};\mathbf{s})$, and the conditional distribution of the power allocation given the total transmission power and the current state, namely $\phi_\mu(\boldsymbol{P};P^{\text{tot}},\mathbf{s})$:

$$\mu(\boldsymbol{P};\mathbf{s})=\phi_\mu(\boldsymbol{P};P^{\text{tot}},\mathbf{s})\gamma_\mu(P^{\text{tot}};\mathbf{s}). \tag{5.2.17}$$

or, equivalently, to

$$\mu(\boldsymbol{P};b,\boldsymbol{\hbar},\boldsymbol{\jmath})=\phi_\mu(\boldsymbol{P};P^{\text{tot}},b,\boldsymbol{\hbar},\boldsymbol{\jmath})\gamma_\mu(P^{\text{tot}};b,\boldsymbol{\hbar},\boldsymbol{\jmath}). \tag{5.2.18}$$

The above expression will be useful to decompose the problem into two steps in Theorem 5.2.6.

We highlight that μ performs a *power control* mechanism, i.e., it specifies how much power is used in every MC state but, in addition to power control, also the code rate can be changed according to Section 5.2.2.2.

5.2.3.2 Finite Model

In the previous paragraphs, we assumed that the policy can be defined for every possible value of the channel gains. This can be done by simple enumeration if $|\mathcal{H}| < \infty$ and $|\mathcal{J}| < \infty$. However, the channel gains may be continuous variables in the general case. Instead of defining a policy for a continuously infinite set of values, we want to find a set of points where the policy can be computed and optimized efficiently. The following approach can be followed. Consider the random variable $H^{(1)}$ (for the others the reasoning is similar). We discretize the support of $H^{(1)}$ in n intervals with an *equally likely* strategy ($\mathbb{P}(H^{(1)} \in [p_i, p_{i+1})) = 1/n$, $i = 1, \ldots, n$). Then, we specify the policy in the centroid of every interval. If the number of intervals n is sufficiently large, the approximation is very close to the continuous case.

Remark Since we consider a discrete channel, we focus without loss of generality on channel conditions with non-zero probability, i.e., $\mathbb{P}(\boldsymbol{H} = \boldsymbol{\hbar}, \boldsymbol{J} = \boldsymbol{\jmath}) > 0$, $\forall \boldsymbol{\hbar} \in \mathcal{H}, \boldsymbol{\jmath} \in \mathcal{J}$.

5.2.4 Optimal Secrecy Policy with Complete CSI

In this section we study the case when the transmitter has perfect CSI knowledge, and introduce a technique to compute OSP and some of its properties. All our results are useful to simplify the numerical evaluation. In particular: 1) we prove that there exists a deterministic OSP (Theorem 5.2.1); 2) we propose a technique to derive a unichain OSP (Section 5.2.4.1); 3) we decompose the optimization process in two steps (Theorem 5.2.6); and 4) we show that the transmission power increases (decreases) with the channel gain of the legitimate receiver's (eavesdropper's) sub-carriers (Theorem 5.2.9).

Theorem 5.2.1. *There exists a deterministic OSP, i.e., an optimal secrecy policy in which, for every MC state* $\mathbf{s} = (b, \boldsymbol{\hbar}, \boldsymbol{\jmath})$

$$\mu^\star(\boldsymbol{P}; \mathbf{s}) = \begin{cases} 1, & \textit{if } \boldsymbol{P} = \boldsymbol{P}^\star_{\mathbf{s}}, \\ 0, & \textit{otherwise}, \end{cases} \tag{5.2.19}$$

for some $\boldsymbol{P}^\star_{\mathbf{s}}$ *depending upon the current MC state in general.*

Proof. See Appendix 5.B. ■

By exploiting Equation (5.2.17), it also follows that $\exists P^{\text{tot}^\star}_{\mathbf{s}}$ such that the transmit power policy γ_μ defined in (5.2.17) satisfies

$$\gamma_\mu(P^{\text{tot}}; \mathbf{s}) = \begin{cases} 1, & \text{if } P^{\text{tot}} = P^{\text{tot}^\star}_{\mathbf{s}}, \\ 0, & \text{otherwise.} \end{cases} \tag{5.2.20}$$

Definition 5.2.2 (Deterministic Policy)**.** *Since a deterministic OSP always exists, we only need to study deterministic policies, thus* μ *can be redefined as*

$$\mu = \{\boldsymbol{P}_{\mathbf{s}} \in \mathcal{P}_{\leq}(b),\ \forall \mathbf{s} \in \mathcal{B} \times \mathcal{H} \times \mathcal{J}\}. \tag{5.2.21}$$

$\boldsymbol{P}_{\mathbf{s}} = [P^{(1)}_{\mathbf{s}}, \ldots, P^{(M)}_{\mathbf{s}}]$ *characterizes the transmission powers on different sub-carriers in state* $\mathbf{s} = (b, \boldsymbol{\hbar}, \boldsymbol{\jmath})$.

We also introduce the sub-policy μ^{tot} as

$$\mu^{\text{tot}} = \{P_{\mathbf{s}}^{\text{tot}},\ \forall \mathbf{s} \in \mathcal{B} \times \mathcal{H} \times \mathcal{J}\}, \tag{5.2.22}$$

which accounts for the total transmission powers only. μ^{tot} and μ are *consistent* if the sum of the elements of $\boldsymbol{P}_{\mathbf{s}}$ in μ is equal to $P_{\mathbf{s}}^{\text{tot}}$ in μ^{tot}, $\forall \mathbf{s} \in \mathcal{B} \times \mathcal{H} \times \mathcal{J}$.

The deterministic property is particularly useful to simplify the numerical evaluation because a policy needs to define only a scalar value for every state of the system and not a probability distribution.

5.2.4.1 Unichain Policies

We restrict our study to the class of *unichain* policies, i.e., those that induce a unichain MC (i.e., a MC with a single recurrent class). This is useful in order to apply the standard optimization algorithms in the next section.

Some sufficient conditions to obtain a unichain policy are presented in the following proposition (in this subsubsection we use deterministic policies for presentation simplicity, but the results can be easily extended).

Proposition 5.2.3. *If a policy satisfies one of the following conditions, then it is unichain. If it satisfies both conditions, the policy induces an irreducible, positive recurrent MC.*

1. *For every $b \in \mathcal{B}\backslash\{b_{\max}\}$ there exists a pair $(\boldsymbol{\hbar}', \boldsymbol{j}')$ such that $P^{\text{tot}}_{b,\boldsymbol{\hbar}',\boldsymbol{j}'} < e_{\max}$ (maximum number of energy arrivals).*
2. *For every $b \in \mathcal{B}\backslash\{0\}$ there exists a pair $(\boldsymbol{\hbar}'', \boldsymbol{j}'')$ such that $P^{\text{tot}}_{b,\boldsymbol{\hbar}'',\boldsymbol{j}''} > e_{\min}$.*

Proof. See Appendix 5.C. ■

In practice, the first and second points ensure that there is a positive probability that the battery moves from level b to higher and lower energy levels, respectively. When they are both verified, no transient state can exist, and the MC is irreducible.

When at least one point of Proposition 5.2.3 is satisfied, the corresponding policy is guaranteed to be unichain. However, in general, these conditions may not be satisfied. In addition, there may exist more than one policy with the same maximum achievable secrecy rate (the highest secrecy rate among $C_\mu(0), \ldots, C_\mu(b_{\max})$). Some of these are unichain, whereas others are not. Consider the following example to justify these claims.

Example 5.2.4. *We want to show a case in which 1) multiple policies with the same maximum reward exist and 2) some of them are not unichain.*

Assume that the harvesting process is deterministic and equal to $\bar{e} < b_{\max}/2$, $M = 1$, and the channel is constant $\hbar_1 > j_1$. Consider the following policies

$$\mu_1 = \{P_{1;b,\hbar_1,j_1} = \min\{b, \bar{e}\},\ \forall b,\ \forall \hbar_1,\ j_1\},$$
$$\mu_2 = \begin{Bmatrix} P_{1;b,\hbar_1,j_1} = 2\bar{e}, & b = b_{\max},\ \forall \hbar_1,\ j_1 \\ P_{1;b,\hbar_1,j_1} = \bar{e}, & b = \bar{e},\ \forall \hbar_1,\ j_1 \\ P_{1;b,\hbar_1,j_1} = 0, & \textit{otherwise} \end{Bmatrix}.$$

μ_1 is a unichain policy (the recurrent class is the battery level $\{\bar{e}\}$) that provides a long-term secrecy rate $c(\bar{e}, \hbar_1, j_1)$ (Equation 5.2.8). Instead, μ_2 is not unichain (the two recurrent classes are $\{\bar{e}\}$ and $\{b_{\max} - \bar{e}, b_{\max}\}$) and its long-term secrecy rate depends upon

the initial state (it can be $c(\bar{e}, \hbar_1, j_1)$ or $0.5c(2\bar{e}, \hbar_1, j_1)$). Also, note that because of the concavity of Equation (5.2.8), $c(\bar{e}, \hbar_1, j_1) > 0.5c(2\bar{e}, \hbar_1, j_1)$. Therefore, there exist more than one policy with the same maximum achievable reward $c(\bar{e}, \hbar_1, j_1)$. Moreover, in μ_2, there are two recurrent classes, and thus it is not unichain.

This example shows that the long-term secrecy rate for a non-unichain policy may depend upon the starting state. Also, it shows that in general there may exist different policies, unichain and not unichain, with the same maximum achievable secrecy rate. The following proposition establishes that there is no loss in generality in considering only unichain policies.

Proposition 5.2.5. *Given a generic policy, it is always possible to derive another policy which is unichain and attains the same maximum achievable secrecy rate as the original policy, regardless of the initial state.*

Proof. We provide a constructive proof in Appendix 5.D. ■

In the remainder of the section we always refer to unichain policies, for which $C_\mu(b_0)$ is independent of b_0 [75]. In particular, Proposition 5.2.5 holds for the optimal secrecy policies, i.e., there always exists a unichain OSP, and therefore we will focus on unichain policies with no loss in optimality. Note that, since we consider a finite MC (we discretized both the battery level and the channel gains), a unichain policy always implies the existence of a steady-state distribution as in Equation (5.2.16).

5.2.4.2 Computation of OSP

We now want to simplify the expression of C_μ by exploiting the results we have found so far. If μ and μ^{tot} are consistent, the long-term secrecy function C_μ can be rewritten as

$$C_\mu = \sum_{b \in \mathcal{B}} \pi_{\mu^{\text{tot}}}(b) \int_{\mathcal{H} \times \mathcal{G}} c(\overbrace{\boldsymbol{P}_{b,\boldsymbol{\hbar},\boldsymbol{j}}}^{\text{specified by } \mu}, \boldsymbol{\hbar}, \boldsymbol{j}) \, \mathrm{d}F(\boldsymbol{\hbar}, \boldsymbol{j}). \tag{5.2.23}$$

An interesting fact is that the steady-state probability $\pi_{\mu^{\text{tot}}}(b)$ depends upon the sub-policy μ^{tot} only. This is because $\pi_{\mu^{\text{tot}}}(b)$ describes the battery energy evolution, that depends only upon the total energy consumption in a slot, not upon the particular power splitting scheme. This result leads to the following theorem.

Theorem 5.2.6. *The maximization of C_μ can be decomposed into two steps:*

1. *Fix a value x and the channel gain vectors $\boldsymbol{\hbar}$, $\boldsymbol{j}$ and find the optimal power splitting choice*

$$\boldsymbol{P}^\star = \arg\max_{\boldsymbol{P}} c(\boldsymbol{P}, \boldsymbol{\hbar}, \boldsymbol{j}), \tag{5.2.24a}$$

s.t.:

$$\boldsymbol{P} \in \mathscr{P}_=(x) \triangleq \left\{ \boldsymbol{P} \ : \ \boldsymbol{P} \succeq 0, \ x = \mathbf{1}_M^T \boldsymbol{P} \right\}. \tag{5.2.24b}$$

2. Maximize C_μ by considering only μ^{tot}

$$\mu^{\text{tot}^\star} = \arg\max_{\mu^{\text{tot}}} C_\mu, \tag{5.2.25a}$$

s.t.:

$$\mu^{\text{tot}} \text{ and } \mu \text{ are consistent,} \tag{5.2.25b}$$

$$\boldsymbol{P}_{\mathbf{s}} \text{ solves (5.2.24) with } x = P_{\mathbf{s}}^{\text{tot}}, \forall \mathbf{s} \in \mathcal{B} \times \mathcal{H} \times \mathcal{G}. \tag{5.2.25c}$$

The optimal $\mu^\star$ can be found by fixing $P^{\text{tot}^\star}$ according to point 2) and choosing $\boldsymbol{P}$ with the optimal power splitting choice of point 1).

Proof. See Appendix 5.E. ■

The *optimal power splitting choice* $\boldsymbol{P}^\star$ that solves (5.2.24) can be found with a Lagrangian approach (for further details, see Theorem 1 and Equation (7) in [42]):

$$P^{(r)\star} = \left[\sqrt{\frac{1}{4}\left(\frac{1}{\jmath^{(r)}} - \frac{1}{\hbar^{(r)}}\right)^2 + \frac{1}{\eta}\left(\frac{1}{\jmath^{(r)}} - \frac{1}{\hbar^{(r)}}\right)} - \frac{1}{2}\left(\frac{1}{\jmath^{(r)}} + \frac{1}{\hbar^{(r)}}\right) \right]^+, \tag{5.2.26}$$

where η is a parameter used to satisfy $x = \sum_{r=1}^{M} P^{(r)\star}$. In the remainder of the section we assume that this optimal power splitting choice is used, unless otherwise stated. We highlight that OSP yields $P^{(r)\star} = 0$ if $\hbar^{(r)} \leq \jmath^{(r)}$, which implies that the achievable secrecy rate with complete CSI is independent of the coding scheme (the two expressions in Equation (5.2.7) coincide).

To solve Step 2) instead, the Optimal Secrecy Policy can be found numerically via dynamic programming techniques, e.g., using PIA or VIA as described in Section 2.2.2.[6]

Note that Theorem 5.2.6 with Equation (5.2.26) decomposes the optimization into two steps. Therefore, the numerical evaluation only requires to study the two points separately instead of performing a (more computationally intensive) bidimensional optimization.

We also remark the following.

Lemma 5.2.7. *By restricting the study to the unichain policies constructed as in Appendix 5.D, OSP is uniquely determined.*

Proof. In all the transient states, by construction (Appendix 5.D), we have $P_{\mathbf{s}}^{\text{tot}^\star} = 0$. For the recurrent states, thanks to [14, Vol. II, Sec. 4], we know that $P_{\mathbf{s}}^{\text{tot}^\star}$ is uniquely determined. ■

5.2.4.3 Properties

We now derive a property that is useful to understand when the transmission power increases or decreases.

[6] A key assumption of PIA is that, at every algorithm step, a *unichain* policy is produced. In order to satisfy this condition, we apply the technique of Appendix 5.D.

Proposition 5.2.8. *Consider two channel states $\boldsymbol{\hbar}'$, $\boldsymbol{\jmath}'$ and $\boldsymbol{\hbar}''$, $\boldsymbol{\jmath}''$ and define*

$$D(P^{\mathrm{tot}};\boldsymbol{\hbar}',\boldsymbol{\jmath}';\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') \triangleq \frac{\partial}{\partial P^{\mathrm{tot}}}\Big(c(\boldsymbol{P}^{\star}_{b,\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - c(\boldsymbol{P}^{\star}_{b,\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}')\Big), \tag{5.2.27}$$

where $\boldsymbol{P}^{\star}_{b,\boldsymbol{\hbar}'',\boldsymbol{\jmath}''}$ and $\boldsymbol{P}^{\star}_{b,\boldsymbol{\hbar}',\boldsymbol{\jmath}'}$ are defined as the solutions[7] *of Problem* (5.2.24) *with $x = P^{\mathrm{tot}}$.*
OSP has the following trend:

- *if $D(P^{\mathrm{tot}};\boldsymbol{\hbar}',\boldsymbol{\jmath}';\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') \geq 0,\ \forall P^{\mathrm{tot}}$, then $P^{\mathrm{tot}^{\star}}_{b,\boldsymbol{\hbar}'',\boldsymbol{\jmath}''} \geq P^{\mathrm{tot}^{\star}}_{b,\boldsymbol{\hbar}',\boldsymbol{\jmath}'}$;*
- *if $D(P^{\mathrm{tot}};\boldsymbol{\hbar}',\boldsymbol{\jmath}';\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') \leq 0,\ \forall P^{\mathrm{tot}}$, then $P^{\mathrm{tot}^{\star}}_{b,\boldsymbol{\hbar}'',\boldsymbol{\jmath}''} \leq P^{\mathrm{tot}^{\star}}_{b,\boldsymbol{\hbar}',\boldsymbol{\jmath}'}$.*

Proof. See Appendix 5.F. ■

In practice, it is better to use more energy in the directions where the function $c(\cdot,\cdot,\cdot)$ increases. A consequence of the previous proposition is derived in the following theorem.

Theorem 5.2.9. *Consider $M = 1$. The transmission power of OSP is non-decreasing with $\hbar$ and non-increasing with $\jmath$ (we omit the "1" subscripts). Formally*

- *if $\hbar'' \geq \hbar'$, then $P^{\mathrm{tot}^{\star}}_{b,\hbar'',\jmath} \geq P^{\mathrm{tot}^{\star}}_{b,\hbar',\jmath}$;*
- *if $\jmath'' \geq \jmath'$, then $P^{\mathrm{tot}^{\star}}_{b,\hbar,\jmath''} \leq P^{\mathrm{tot}^{\star}}_{b,\hbar,\jmath'}$.*

Proof. See Appendix 5.G. ■

This is an expected result, i.e., when the legitimate channel improves, then it is reasonable to use more energy in order to get a higher rate. Conversely, when the eavesdropper's channel improves, it is better not to use a lot of energy because only low rates can be obtained. In this case, it is better to conserve energy and wait for a better slot. The previous theorem is useful to prune the action space in the numerical computation: if we found the optimal transmission power for a given channel state, we could exploit it as lower [upper] bound for better [worse] channel states.

We expect that a result similar to Theorem 5.2.9 holds for a generic $M > 1$. A formal proof would require to explicitly compute $D(P^{\mathrm{tot}};\boldsymbol{\hbar}',\boldsymbol{\jmath}';\boldsymbol{\hbar}'',\boldsymbol{\jmath}'')$ and show that it is non-negative or non-positive (see Appendix 5.G). However, this would require the computation of an analytical expression for η in Equation (5.2.26). Even though this is in principle possible for any fixed M, the corresponding expression is very complicated and, in practice, the resulting $D(P^{\mathrm{tot}};\boldsymbol{\hbar}',\boldsymbol{\jmath}';\boldsymbol{\hbar}'',\boldsymbol{\jmath}'')$ is too long to be analytically tractable.

5.2.5 Optimal Secrecy Policy with Partial CSI

In the previous sections we assumed that the realizations of $\boldsymbol{H}$ and $\boldsymbol{J}$, namely $\boldsymbol{\hbar}$ and $\boldsymbol{\jmath}$, are known at the transmitter. This may not be true in practice. In particular, it is likely that, since the eavesdropper does not cooperate with the transmitter, its channel gain is unknown. In this section we gradually remove these assumptions and discuss how the achievable secrecy rate changes as a result.

We assume that $\boldsymbol{H} = [H^{(1)},\dots,H^{(M)}]$ and $\boldsymbol{J} = [J^{(1)},\dots,J^{(M)}]$ have independent components and are independent of each other. In this section we assume that all links

[7] Note that $\boldsymbol{P}^{\star}_{b,\boldsymbol{\hbar}'',\boldsymbol{\jmath}''}$ and $\boldsymbol{P}^{\star}_{b,\boldsymbol{\hbar}',\boldsymbol{\jmath}'}$ depend upon P^{tot}.

are affected by i.i.d. Nakagami fading. This means that the amplitude of a received signal has a Nakagami pdf with parameters m and κ, i.e.,

$$f(x;m,\kappa) = 2\left(\frac{m}{\kappa}\right)^M \frac{1}{\Gamma(m)} x^{2m-1} e^{-\frac{m}{\kappa}x^2}, \quad x \geq 0, \tag{5.2.28}$$

$$\Gamma(m) \triangleq \int_0^\infty e^{-t} t^{m-1}\, \mathrm{d}t, \tag{5.2.29}$$

where e is the Napier's constant. Therefore, $H^{(r)}$ and $J^{(r)}$ exhibit a Gamma distribution. The pdf of $H^{(r)}$ (with mean $\bar{h}^{(r)}$) is

$$f_{H^{(r)}}(\hbar;m) = \left(\frac{m}{\bar{h}^{(r)}}\right)^M \frac{1}{\Gamma(m)} \hbar^{m-1} e^{-\frac{m}{\bar{h}^{(r)}}\hbar}, \quad \hbar \in \mathbb{R}_+, \quad m \geq 1 \tag{5.2.30}$$

and similarly for $J^{(r)}$ (for presentation simplicity, we assume that the legitimate receiver and the eavesdropper have the same index m, but the analysis can be extended to a more general case). Note that $m = 1$ corresponds to Rayleigh fading and $f_{H^{(r)}}(\hbar;1) = \frac{1}{\bar{h}^{(r)}} e^{-\hbar/\bar{h}^{(r)}}$ is an exponential distribution. As m increases, the strength of the line of sight component increases. For ease of notation, in the remainder of the section we drop the dependence on m and implicitly assume $f_{H^{(r)}}(\hbar) = f_{H^{(r)}}(\hbar;m)$.

5.2.5.1 Unknown Eavesdropper's Channel

In this section, we assume that both the legitimate and the eavesdropper's channels are affected by fading but CSI is available only for $\boldsymbol{H}$. In this case, due to this lack of information, it may happen that EHD transmits even when the eavesdropper's channel gain is higher than the legitimate one.

Similarly to Expression (5.2.21) in the previous section, a policy μ can be defined as

$$\mu = \{\boldsymbol{P}_{b,\boldsymbol{\hbar}} \triangleq [P_{b,\boldsymbol{\hbar}}^{(1)}, \ldots, P_{b,\boldsymbol{\hbar}}^{(M)}] \in \mathscr{P}_{\leq}(b),\ \forall b \in \mathcal{B},\ \forall \boldsymbol{\hbar} \in \mathcal{H}\}, \tag{5.2.31}$$

and similarly for μ^{tot}. $\boldsymbol{P}_{b,\boldsymbol{\hbar}}$ represents the transmission power used in state $(b,\boldsymbol{\hbar})$ (since $\boldsymbol{\jmath}$ is unknown, it cannot be included in the state of the system). We remark that μ performs a *power control* mechanism, i.e., a policy specifies only the transmission power $\boldsymbol{P}_{b,\boldsymbol{\hbar}}$. However, in addition to power control, in every slot also the code rate can be changed (see Section 5.2.2.2). In particular, variable rate coding provides higher secrecy rates than constant rate coding, but is more difficult to implement. In the following we analyze both these approaches.[8]

Constant Rate Coding. The simplest assumption is that the coding scheme has constant rate and its choice only depends on the overall channel statistics. Using constant rate coding, the eavesdropper is able to gather more information than the legitimate receiver when its channel is better. Because of this, for some r, we may have (see Equation (5.2.7))

$$R_{\hbar^{(r)},\jmath^{(r)}}(P_{b,\boldsymbol{\hbar}}^{(r)}) < 0. \tag{5.2.32}$$

[8]Differently from the complete CSI case of Section 5.2.4, $P^{(r)}$ cannot be set to 0 if $\hbar^{(r)} \leq \jmath^{(r)}$ (see Equation (5.2.26)), thus using constant rate or variable rate coding leads to different results.

The secrecy rate expression becomes

$$C_\mu = \sum_{b=0}^{b_{\max}} \pi_{\mu^{\text{tot}}}(b) \int_{\mathbb{R}_+^M} \int_{\mathbb{R}_+^M} \sum_{r=1}^{M} \log_2 \left(\frac{1 + \hbar^{(r)} P_{b,\boldsymbol{\hbar}}^{(r)}}{1 + j^{(r)} P_{b,\boldsymbol{\hbar}}^{(r)}} \right) \prod_{r=1}^{M} \left(f_{G^{(r)}}(\hbar^{(r)}) f_{H^{(r)}}(j^{(r)}) \right) \mathrm{d}\boldsymbol{\hbar} \, \mathrm{d}\boldsymbol{j}, \tag{5.2.33}$$

Note that in (5.2.33) we integrate both positive and negative terms. The negative terms are due to the fact that the eavesdropper's channel may be better than the legitimate one ($j^{(r)} > \hbar^{(r)}$).

We now want to extract some properties of the optimal secrecy policy in this context. We start by performing the following computations, which will be used to extend the first point of Theorem 5.2.9.

The channel memoryless property can be used to simplify (5.2.33) and recast the problem using an MDP. By integrating over $\boldsymbol{j}$, we obtain

$$C_\mu = \sum_{b=0}^{b_{\max}} \pi_{\mu^{\text{tot}}}(b) \int_{\mathbb{R}_+^M} \sum_{r=1}^{M} T_{\text{con}}^{(r)}(\hbar^{(r)}, P_{b,\boldsymbol{\hbar}}^{(r)}) \prod_{r=1}^{M} f_{G^{(r)}}(\hbar^{(r)}) \, \mathrm{d}\boldsymbol{\hbar}. \tag{5.2.34}$$

$$T_{\text{con}}^{(r)}(\hbar, P) \triangleq \int_{\mathbb{R}_+} \log_2 \left(\frac{1 + \hbar P}{1 + jP} \right) f_{H^{(r)}}(j) \, \mathrm{d}j. \tag{5.2.35}$$

The function $T_{\text{con}}^{(r)}(\hbar, P)$ is presented in Equation (5.2.36), where $\mathrm{Ei}(z) = -\int_{-z}^{\infty} \frac{e^{-t}}{t} \, \mathrm{d}t$ is the exponential integral function and s_i, t_i are constants.[9]

$$\begin{aligned} T_{\text{con}}^{(r)}(\hbar, P) = \; & \log_2(1 + \hbar \, P) + \frac{1}{\log 2} \Bigg(\sum_{i=2}^{M} s^{(i)} \times \left(P \, \bar{h}^{(r)} \right)^{i-m} \\ & + e^{\frac{m}{P \, \bar{h}^{(r)}}} \mathrm{Ei}\left(-\frac{m}{P \, \bar{h}^{(r)}} \right) \sum_{i=1}^{M} t^{(i)} \times \left(P \, \bar{h}^{(r)} \right)^{i-m} \Bigg). \end{aligned} \tag{5.2.36}$$

A secure transmission can be performed only if $C_\mu > 0$. The maximum of (5.2.34) can be found with an MDP approach, where the MC state is given by the pair $(b, \boldsymbol{\hbar})$.

A property, that directly follows from the definitions of $T_{\text{con}}^{(r)}(\hbar, P)$, is the following.

Proposition 5.2.10. *If for $P > 0$ we obtain $T_{\text{con}}^{(r)}(\hbar, P) < 0$, then allocating a power P over sub-carrier r is strictly suboptimal.*

This result is intuitive. Indeed, if $T_{\text{con}}^{(r)}(\hbar, P) < 0$ and $P > 0$, then in (5.2.34) we are adding negative terms. This is clearly suboptimal because it lowers the secrecy rate and wastes energy at the same time.

Even if $T_{\text{con}}^{(r)}(\hbar, P)$ has a complicated expression, as we will see, we are interested in its double derivative with respect to $\hbar$ and P:

$$\frac{\partial^2}{\partial P \partial \hbar} T_{\text{con}}^{(r)}(\hbar, P) = \frac{1}{\log 2} \frac{1}{(1 + \hbar P)^2}. \tag{5.2.37}$$

[9] Closed form expressions for $s^{(i)}$ and $t^{(i)}$ can be derived but are quite complicated. Moreover, we will see that they do not contribute to our next results.

We now show that even with partial CSI the optimal secrecy policy increases with the legitimate channel gain. As for Theorem 5.2.9, the following result can be used to prune the action space.[10]

Theorem 5.2.11. *Consider $M = 1$. With partial CSI, the transmission power of OSP is non-decreasing with $\hbar$ (we omit the "1" subscripts). Formally, if $\hbar'' \geq \hbar'$, then $P^{\text{tot}^\star}_{b,\hbar''} \geq P^{\text{tot}^\star}_{b,\hbar'}$.*

Proof. The proof follows the same steps presented in Appendices 5.B, 5.F, 5.G. To prove the theorem the key point is that

$$\frac{\partial^2}{\partial P \partial \hbar} T^{(r)}_{\text{con}}(\hbar, P) \geq 0. \tag{5.2.38}$$

Note that, considering the derivative with respect to P, it follows from (5.2.38) that $\frac{\partial}{\partial \hbar} T^{(r)}_{\text{con}}(\hbar, P_B) - \frac{\partial}{\partial \hbar} T^{(r)}_{\text{con}}(\hbar, P_A) \geq 0$, for $P_A \leq P_B$. We can rewrite the inequality as $\frac{\partial}{\partial \hbar}\left(T^{(r)}_{\text{con}}(\hbar, P_B) - T^{(r)}_{\text{con}}(\hbar, P_A)\right) \geq 0$ and obtain

$$T^{(r)}_{\text{con}}(\hbar + \Delta, P_A) - T^{(r)}_{\text{con}}(\hbar, P_A) \leq T^{(r)}_{\text{con}}(\hbar + \Delta, P_B) - T^{(r)}_{\text{con}}(\hbar, P_B), \tag{5.2.39}$$

$\forall \Delta \geq 0$ and $P_A \leq P_B$. This condition can be replaced with Equation (5.F.5) in Appendix 5.F to prove the theorem. ■

Variable Rate Coding. Better performance can be obtained with variable rate coding (see Equations (5.2.5) and (5.2.6)). In this case, in every slot, the code rate is matched to the legitimate channel rate. Thus, even if $\hbar^{(r)} \leq \jmath^{(r)}$ (eavesdropper's channel is better), the eavesdropper can gather at most $R_{\hbar^{(r)}}$ bits (legitimate transmission rate) and not $R_{\jmath^{(r)}}$ (eavesdropper's transmission rate). The secrecy rate expression is

$$C_\mu = \sum_{b=0}^{b_{\max}} \pi_{\mu^{\text{tot}}}(b) \int_{\mathbb{R}_+^M} \int_{\mathbb{R}_+^M} \sum_{r=1}^{M} \left[\log_2\left(\frac{1+\hbar^{(r)} P^{(r)}_{b,\boldsymbol{\hbar}}}{1+\jmath^{(r)} P^{(r)}_{b,\boldsymbol{\hbar}}}\right)\right]^+ \prod_{r=1}^{M} \left(f_{G^{(r)}}(\hbar^{(r)}) f_{H^{(r)}}(\jmath^{(r)})\right) \mathrm{d}\boldsymbol{\hbar}\, \mathrm{d}\boldsymbol{\jmath}, \tag{5.2.40}$$

As before, we introduce a function $T^{(r)}_{\text{var}}(\hbar, P^{(r)}_{b,\boldsymbol{\hbar}})$ such that

$$C_\mu = \sum_{b=0}^{b_{\max}} \pi_{\mu^{\text{tot}}}(b) \int_{\mathbb{R}_+^M} \sum_{r=1}^{M} T^{(r)}_{\text{var}}(\hbar^{(r)}, P^{(r)}_{b,\boldsymbol{\hbar}}) \prod_{r=1}^{M} f_{G^{(r)}}(\hbar^{(r)})\, \mathrm{d}\boldsymbol{\hbar}. \tag{5.2.41}$$

$$T^{(r)}_{\text{var}}(\hbar, P) \triangleq \int_{\mathbb{R}_+} \left[\log_2\left(\frac{1+\hbar P}{1+\jmath P}\right)\right]^+ f_{H^{(r)}}(\jmath)\, \mathrm{d}\jmath \tag{5.2.42}$$

$$= \int_0^{\hbar} \log_2\left(\frac{1+\hbar P}{1+\jmath P}\right) f_{H^{(r)}}(\jmath)\, \mathrm{d}\jmath. \tag{5.2.43}$$

In Equation (5.2.43) we integrate from zero to $\hbar$, thus we remove the $[\cdot]^+$ notation (see the structure of Equation (5.2.7) with variable rate coding).

[10] We provide a formal proof only for the case $M = 1$ because, even if theoretically possible, the proof for a generic $M > 1$ is not analytically tractable (see the related discussion just after Theorem 5.2.9).

Note that $T_{\text{var}}^{(r)}(\hbar, P) \geq T_{\text{con}}^{(r)}(\hbar, P)$, which justifies the fact that the achievable secrecy rate with variable rate coding is higher than with constant rate coding.

The analogous of Theorem 5.2.11 holds in this case, as can be proved by exploiting the structure of the double derivative of $T_{\text{var}}^{(r)}(\hbar, P)$:

$$\frac{\partial^2}{\partial P \partial \hbar} T_{\text{var}}^{(r)}(\hbar, P) = \frac{1}{\log 2} \frac{\Gamma(m) - \Gamma\left(m, \frac{m\hbar}{\bar{h}^{(r)}}\right)}{(1+\hbar P)^2 \Gamma(m)}, \tag{5.2.44}$$

where $\Gamma(m, z) \triangleq \int_z^\infty e^{-t} t^{m-1} \, \mathrm{d}t$ is the incomplete gamma function.

5.2.5.2 No Channel State Information

Lower secrecy rates are obtained when also the legitimate receiver's channel is unknown. In particular, the transmission power cannot be adapted to the current channel state. It is easy to show that C_μ can be greater than zero only if $\bar{g}^{(r)} > \bar{h}^{(r)}$ for some r. However, the mean values of the channel gains are not controlled by the transmitter (they are physical quantities), thus if the legitimate channel is (statistically) worse, no secrecy can be achieved.

5.2.6 Numerical Evaluation

In this section we discuss how the secrecy rate changes as a function of the different system parameters.

We compare the following scenarios: OSP with Full CSI (OSP-FULL), OSP with only legitimate channel knowledge and constant rate coding (OSP-PAR-CON) or variable rate coding (OSP-PAR-VAR) and OSP with only statistical channel knowledge (OSP-STAT).

If not otherwise stated, the simulation parameters are: $b_{\max} = 30$, truncated geometric energy arrivals with $e_{\max} = 6$ and $\bar{e} = 1$, $n = 15$ quantization intervals (see Section 5.2.3.2), $M = 1$ (single sub-carrier), $\bar{g} = \bar{h} = 1$ (symmetric scenario), $\mathcal{H} = \mathcal{G} = \mathbb{R}_+$ with $m = 1$ (Rayleigh fading). Since the slot length is fixed, we will deal with *powers* instead of *energies* and express them in energy quanta without loss of generality. After showing results for this choice of parameters, we study the sensitivity of the system performance by changing one or more parameters while keeping the others fixed.

5.2.6.1 Fixed Parameters

Figure 5.6 shows the optimal transmission power $P_{\mathbf{s}}^{\text{tot}^\star}$ as a function of the battery level e when $\hbar \in [0.41, 0.51)$ and $j \in \mathbb{R}_+$. We recall that, when $\mathcal{H} = \mathcal{G} = \mathbb{R}_+$, we use the technique explained in Section 5.2.3.2, i.e., we have a finite number of points where the transmission power is computed ($n = 15$). When $j \geq 0.51$, the transmission power is identically zero because the eavesdropper is always advantaged. Also when $j \in [0.41, 0.51)$ the transmission power is zero. This is not obvious a priori and strongly depends upon the considered interval of $\hbar$. It can be seen that Theorem 5.2.9 holds, i.e., $P_{\mathbf{s}}^{\text{tot}^\star}$ does not increase with j. Finally, we note that the behavior of the transmission power is not obvious a priori, e.g., it is significantly different from a simple greedy policy ($P_{\mathbf{s}}^{\text{tot}^\star} = b$) even when j is low.

Figure 5.7, instead, shows the steady-state probabilities as a function of the energy level b, for fixed $b_{\max}$ and in the different scenarios. In all cases, the curves are similar. This is because the device tends to operate in an efficient region, i.e., approximately at $b_{\max}/2$. This is in order to avoid energy outage and overflow, that degrade the performance of the

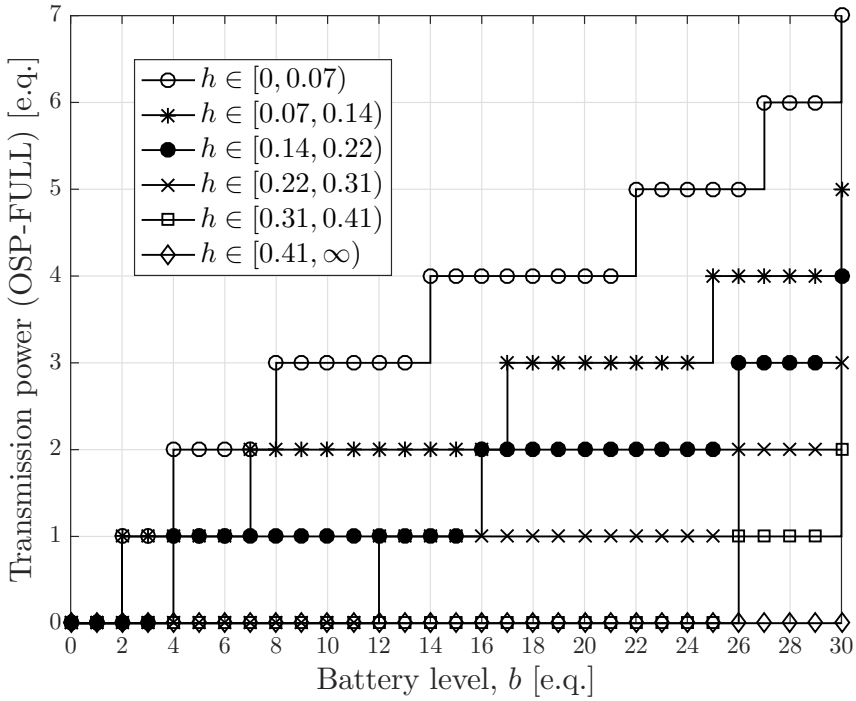

Figure 5.6. Transmission power $P_{\mathbf{s}}^{\mathrm{tot}^\star}$ as a function of the battery level b for several values of j and $\hbar \in [0.41, 0.51)$.

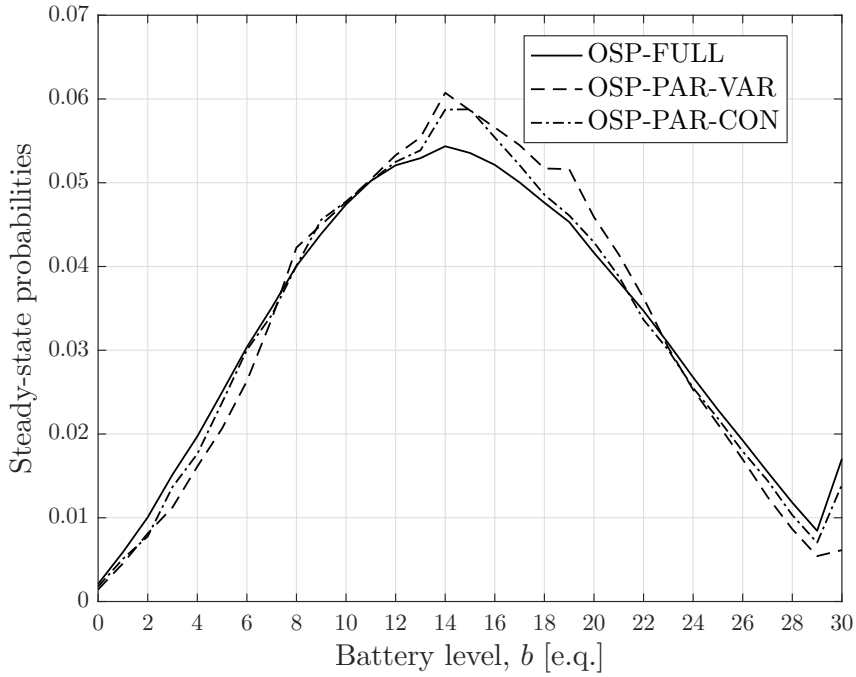

Figure 5.7. Steady-state probabilities $\pi_{\mu^{\mathrm{tot}}}(b)$ as a function of the battery level b.

system. When b approaches $b_{\max}$, the steady-state tails increase because of the overflow (when the battery is almost full, all harvesting events leading to overflow contribute to increasing the steady-state probability of state $b_{\max}$, which is then higher than those of the immediately lower states).

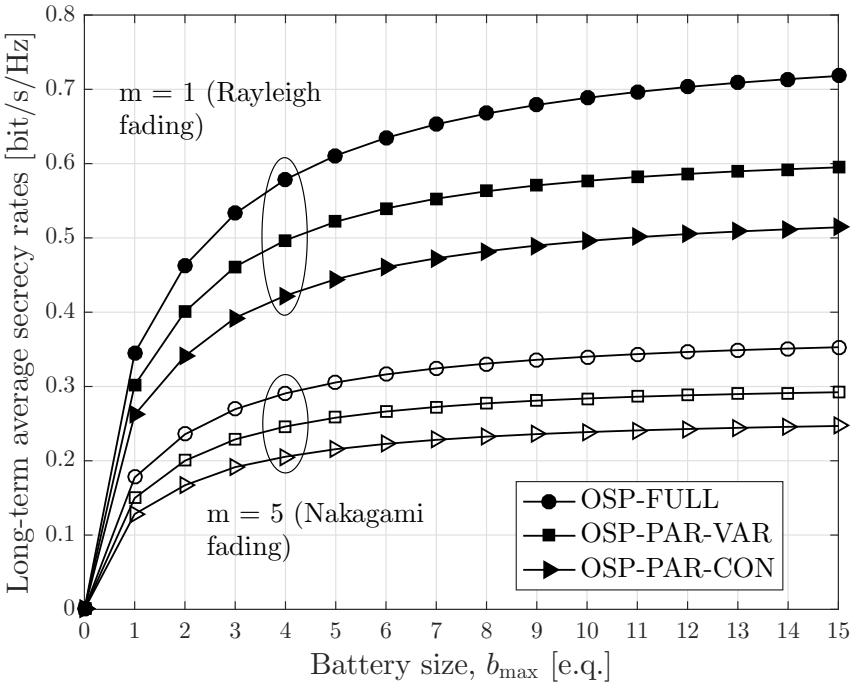

Figure 5.8. Secrecy rate C_μ as a function of the battery size $b_{\max}$ in the case of symmetric channel conditions.

5.2.6.2 Battery Size

In Figure 5.8 we show the rate achieved by the various policies as a function of the battery size $b_{\max}$. We use Rayleigh ($m = 1$) and a general Nakagami fading with a strong Line of Sight (LoS) component ($m = 5$). The curves of OSP-STAT are identically zero because $\bar{h} = \bar{j}$. As expected, OSP-FULL has the highest secrecy rate for every value. It can be seen that the curves saturate after a certain value. This is due to the combination of two effects: 1) the harvesting rate of the EHD is limited (it can be shown that the performance of an EH system is bounded) and 2) the achievable secrecy rate always saturates in the high power regime (because of the structure of Equation (5.2.7)). Note that the curves saturate already for small $b_{\max}$, therefore, in practice, it may be sufficient to use small batteries to obtain high secrecy rates.

In [42, Section IV-B] the authors showed that, when the transmission is subject to an average power constraint, the performance of the optimal transmission scheme with variable rate coding and partial CSI knowledge approaches the performance of the full CSI case when the transmission power is sufficiently high. In our previous example, OSP-PAR-VAR does not achieve OSP-FULL when $b_{\max}$ increases because an energy harvesting system imposes an average power constraint $\bar{e}$.[11] It can be verified that, when $\bar{e}$ increases,

[11] This can be easily derived starting from the causality constraint

$$\sum_{k=0}^{K-1}\sum_{r=1}^{M}\Sigma_k^{(r)} \leq b_0 + \sum_{k=0}^{K-2} e_k, \qquad \forall K = 0, 1, \ldots \tag{5.2.45}$$

where, according to Equation (5.2.11), $\Sigma_k^{(r)}$ is the transmission power over sub-carrier r in time slot k, e_k is the amount of energy harvested in slot k and b_0 is the amount of energy initially available in the battery. In the long run, the right-hand side becomes the power constraint of our system.

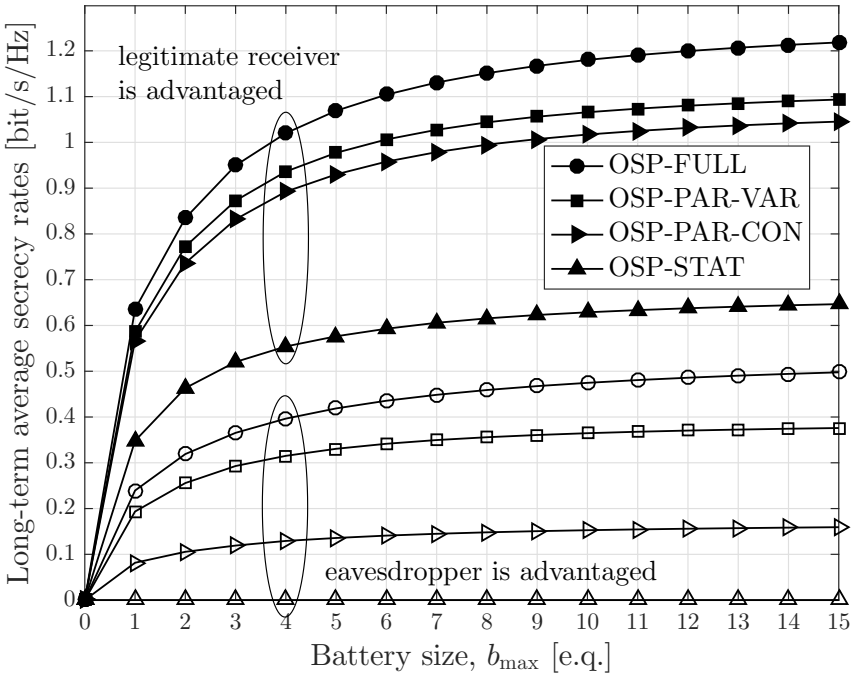

Figure 5.9. Secrecy rate C_μ as a function of the battery size $b_{\max}$ in the case of asymmetric channel conditions and Rayleigh fading.

if the battery size is sufficiently large, the gap between OSP-PAR-VAR and OSP-FULL is smaller.

Note that the achievable secrecy rates strongly depend upon the fading statistics. With $m = 5$, we have strong LoS components, i.e., the channel pdfs tend to be narrow around their means ($\bar{h} = \bar{j}$). It follows that the legitimate and eavesdropper's channel gains are close to each other most of the time. This corresponds to low values of $R_{\hbar^{(r)},P^{(r)}}(P^{(r)})$, thus a low secrecy rate. With Rayleigh fading, instead, exploiting channel diversity allows to obtain higher rewards. This is also the reason why, with Rayleigh fading, full channel state information (OSP-FULL) provides a great improvement with respect to the partial knowledge cases.

Figure 5.9 is similar to the previous one but with asymmetric channel gains. When the eavesdropper is advantaged ($\bar{h} = 1$, $\bar{j} = 2$), even if low performance can be achieved, secret transmission is still possible. When OSP-PAR-CON is used, it is likely that EHD transmits even when the eavesdropper's channel is better and in this case, from Equation (5.2.34), the secrecy rate is lower. This effect is emphasized if the eavesdropper's channel is advantaged, because it is more likely that the legitimate channel is the worse of the two.

On the other hand, if the legitimate channel is better ($\bar{h} = 2$, $\bar{j} = 1$), the secrecy rate can reach high values. In this case, OSP-STAT is also considered and, as expected, is the worst among the optimal policies.

5.2.6.3 Number of Sub-Carriers

When $M = 1$, finding the optimal policies for high values of n (fine quantization of the channel gains) is feasible. We recall that the number of states of the MC is directly proportional to the number of possible combinations of channel gains. Thus, with $M = 1$, the possible combinations are $n \times n$ (legitimate channel $\times$ eavesdropper's channel). With a generic M, the combinations become $n^M \times n^M$. Thus, the number of states grows

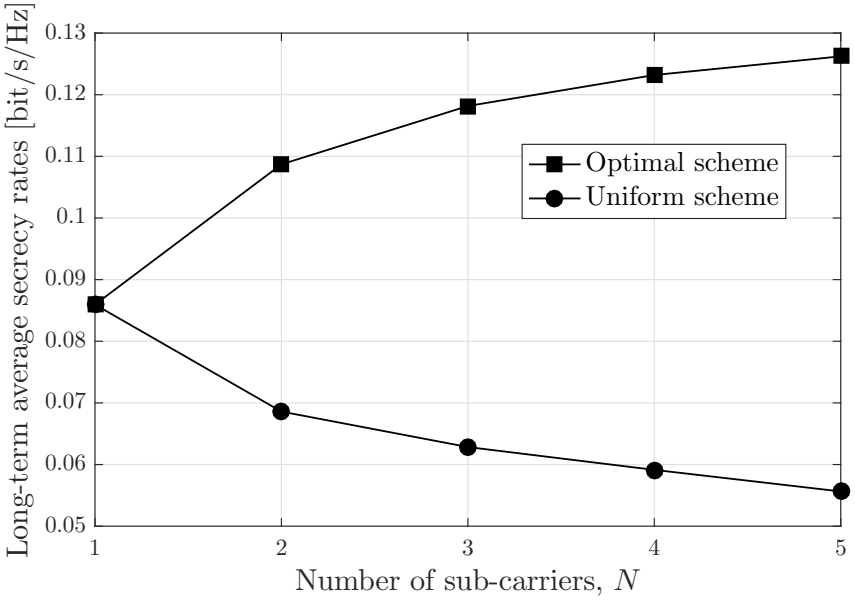

Figure 5.10. Secrecy rate C_μ as a function of the number of sub-carriers M.

exponentially with the number of sub-carriers, making the optimization process for high M infeasible in practice (curse-of-dimensionality). Even when the problem symmetry can be exploited (when $H^{(r)}$ and $J^{(r)}$ are i.i.d.), the computational effort still remains heavy. In pratice, this approach can be applied to multi-carrier scenarios if the number of carriers, M, and the number of quantization levels for the channel, n, are not too large. Note however that our solution suffers from a dimensionality problem because it is the *optimal* solution. Part of our future work agenda may include the design of suboptimal schemes and the study of trade-offs between computational times and performance.

In the following, as an example, we consider a discrete GOOD-BAD channel and discuss the importance of the power splitting scheme. We define $\mathcal{H} = \mathcal{G} = \{\mathrm{B}, \mathrm{G}\} = \{1/30, 3/30\} = \{-15 \text{ dB}, -10 \text{ dB}\}$ with probabilities 0.7 and 0.3, respectively. We also set $b_{\max} = 10$ because, generally, the saturation region is almost reached for this battery size (see Figures 5.8 and 5.9). In Figure 5.10, we plot OSP-FULL as a function of the number of sub-carriers M when the optimal (Equation (5.2.26)) or a uniform power splitting is used. In the optimal case, as M increases, the reward also increases. This is expected because, when one user experiences a bad channel condition, then the power can be directed to other good sub-carriers. Instead, with uniform power splitting, the secrecy rate decreases with M. In practice, this happens because, instead of sending all the transmission power in the "good" sub-carriers, a fraction of this is wasted in the "bad" sub-carriers. For example, with $M = 2$, it may happen that over sub-carrier 1 the pair legitimate-eavesdropper's channel gain is (G, B) whereas, for sub-carrier 2, the pair is (B, B), i.e., sub-carrier 1 is a "good" sub-carrier while sub-carrier 2 is not. In this case, if a positive transmission power were used, the corresponding reward would be greater than zero but the power sent over sub-carrier 2 would be wasted (only when the two pairs are (G, B) and (G, B), is no power wasted during the transmission). This explains why the performance degrades as the number of sub-carriers increases. Moreover, the effect is emphasized with larger M because there are more cases where the transmission power cannot be fully exploited.

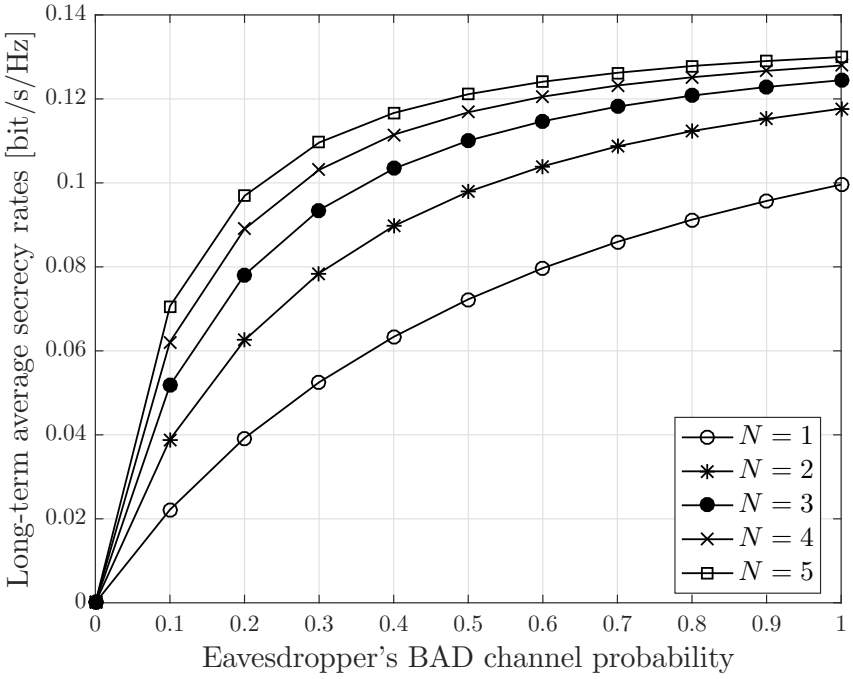

Figure 5.11. Secrecy rate C_μ of OSP-FULL as a function of the eavesdropper's BAD channel probability in a binary channel system.

When the legitimate and the eavesdropper's channel gains are known in every slot, using a smart power splitting scheme is convenient because it can significantly improve the network performance. If this is not possible (e.g., because this information is not available or not reliable), a suboptimal strategy needs to be adopted, e.g., uniform power splitting, which is simpler to implement but yields lower performance in general. The study of the information/performance trade-off for power splitting strategies is left for future work.

Finally, Figure 5.11 shows how the optimal secrecy rate changes as a function of $\mathbb{P}(\hbar_1 = \mathrm{B}) = \mathbb{P}(\hbar_2 = \mathrm{B}) \in [0, 1]$ for different numbers of sub-carriers. It can be noticed that the case with five sub-carriers and $\mathbb{P}(\hbar_1 = \mathrm{B}) = 0.2$ achieves the same performance as the system with only one sub-carrier but $\mathbb{P}(\hbar_1 = \mathrm{B}) = 1$. In practice, the diversity offered by a greater number of sub-carriers can be efficiently exploited to obtain higher secrecy rates. An interesting point is that, as M increases, the improvement obtained from M to $M + 1$ decreases. This is due to the concavity properties of Equation (5.2.8). Therefore, it may not be necessary to use a large number of sub-carriers to obtain high secrecy rates.

5.2.7 Conclusions of Section 5.2

In the second part of this chapter we analyzed an energy harvesting device that has a finite energy storage and transmits secret data to a receiver over M parallel channels exploiting physical layer characteristics. We found the best power allocation technique, namely the Optimal Secrecy Policy (OSP), in several contexts depending on the degree of channel knowledge the device has. We proved several properties of OSP and in particular that it is deterministic and monotonic. We also described a technique to compute OSP by decomposing the problem in two steps and using a dynamic programming approach. When only partial channel state information is available, we described how the maximum secrecy rate varies with constant and variable rate coding, explaining and numerically evaluating the advantages of variable rate coding. We numerically showed that, because

of the limited harvesting rate that is inherently provided by the renewable energy source, OSP-PAR-VAR does not achieve the same performance of OSP-FULL as the battery size increases, and noted that it is not necessary to use very large batteries to achieve close to optimal performance. We also set up the problem when more than one sub-carrier is considered, and discussed the scalability problems related to such scenario. Also, we found that using the optimal power splitting scheme provides a significant advantage with respect to the simpler uniform splitting approach.

Future work may include the study of suboptimal strategies for the case with N sub-carriers in order to avoid the curse-of-dimensionality problem. Also, other optimization techniques can be investigated, e.g., offline approach, Lyapunov optimization or reinforcement learning approach. Finally, it would be interesting to set up a simulation experiment with real data measurements (e.g., for the harvesting process) in order to validate our results in a realistic scenario.

Appendix 5.A Proof of Proposition 5.1.1

The energy causality constraint imposes

$$\sum_{i=0}^{k} P_i \leq \frac{1}{\tau_{\mathrm{TX}}} \sum_{i=0}^{k-1} e_i, \qquad \forall k = 0, 1, 2, \dots \tag{5.A.1}$$

that immediately leads to the upper bound $g(\bar{e}/\tau_{\mathrm{TX}})$ as k grows to infinity [90]. However, in the case of storage losses, the previous inequality can be changed as follows in order to provide a better upper bound. The slot evolution is as follows

$$\begin{aligned}
&b_0 = 0 \implies P_0 = 0,\\
&b_1 = y_T(e_0, b_0) \implies P_1 \leq y_T(e_0, b_0)/\tau_{\mathrm{TX}},\\
&b_2 = y_T(e_0, b_0) + y_T(e_1, b_1) - \tau_{\mathrm{TX}} P_1 \implies P_1 + P_2 \leq (y_T(e_0, b_0) + y_T(e_1, b_1))/\tau_{\mathrm{TX}},\\
&\vdots\\
&b_k = \sum_{i=0}^{k-1} y_T(e_i, b_i) - \tau_{\mathrm{TX}} \sum_{i=0}^{k-1} P_i \implies \sum_{i=0}^{k} P_i \leq \frac{1}{\tau_{\mathrm{TX}}} \sum_{i=0}^{k-1} y_T(e_i, b_i), \quad k > 0.
\end{aligned}$$

In order to obtain a simple upper bound, we remove the terms E_i from the previous equation, assuming that they are always chosen optimally. Obviously, in the general case this is not possible and this is why, with this assumption, we only obtain an upper bound

$$\sum_{i=1}^{k} P_i \leq \frac{1}{\tau_{\mathrm{TX}}} \sum_{i=1}^{k-1} y_T^{\star}(B_i), \qquad \forall k = 1, 2, \dots \tag{5.A.2}$$

With the previous equation, by exploiting the concavity of g, we obtain the upper bound (5.1.9).

Appendix 5.B Proof of Theorem 5.2.1

We want to show that OSP is a deterministic policy, i.e., given the state of the system, $\mu(\boldsymbol{P}; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) = \delta_{\boldsymbol{P}, \boldsymbol{P}^{\star}_{b,\boldsymbol{\hbar},\boldsymbol{\jmath}}}$, where $\delta_{\cdot,\cdot}$ is the Kronecker delta function.[12]

Note that the study can be split into two parts according to Equation (5.2.17). Thus, we only need to prove that both $\gamma_\mu(P^{\text{tot}}; b, \boldsymbol{\hbar}, \boldsymbol{\jmath})$ (transmit power policy) and $\phi_\mu(\boldsymbol{P}; P^{\text{tot}}, b, \boldsymbol{\hbar}, \boldsymbol{\jmath})$ are deterministic. In the following we prove the first part. The latter is derived in [42].

5.B.1 Deterministic Transmit Power Policy

As a preliminary result, we need the following proposition (in this subsubsection, the expectation is always taken with respect to $\boldsymbol{H}$ and $\boldsymbol{J}$).

Proposition 5.B.1. *$\mathbb{P}(b_k = b|b_0)$ depends upon the policy only through $\mathbb{E}[\gamma_\mu(P^{\text{tot}}; b, \boldsymbol{H}, \boldsymbol{J})]$, $\forall P^{\text{tot}} \in \{0, \dots, b\}$, $\forall b \in \mathcal{B}$.*

Proof. The proof is by induction on k. At $k = 0$, $\mathbb{P}(b_0 = b|b_0 = b_0)$ is equal to 1 if $b = b_0$ and to 0 otherwise. In this case there is no dependence upon the policy.

Assume that the thesis is true for k (inductive hypothesis). Using the chain rule, the probability that $b_{k+1} = b'$ given the initial state is

$$\mathbb{P}(b_{k+1} = b'|b_0) = \sum_{b=0}^{b_{\max}} \mathbb{P}(b_{k+1} = b'|b_k = b) \times \mathbb{P}(b_k = b|b_0). \tag{5.B.1}$$

Thus, to prove the thesis, we focus on $\mathbb{P}(b_{k+1} = b'|b_k = b)$, whereas for $\mathbb{P}(b_k = b|b_0)$ we use the inductive hypothesis. Assume $b' < b_{\max}$ ($\mathbb{P}(e)$ is the probability of gathering e energy quanta via the EEH interface)

$$\mathbb{P}(b_{k+1} = b'|b_k = b) = \sum_{e=\max\{0,b'-b\}}^{\min\{b', e_{\max}\}} \mathbb{P}(e)\mathbb{E}[\gamma_\mu(b - b' + e; b, \boldsymbol{H}, \boldsymbol{J})], \tag{5.B.2}$$

whereas, if $b' = b_{\max}$

$$\mathbb{P}(b_{k+1} = b_{\max}|b_k = b) = \sum_{e=\max\{0,b_{\max}-b\}}^{e_{\max}} \mathbb{P}(e) \sum_{d=0}^{b-b_{\max}+e} \mathbb{E}[\gamma_\mu(d; b, \boldsymbol{H}, \boldsymbol{J})]. \tag{5.B.3}$$

Note that we used the *transmit power policy* $\gamma_\mu(\cdot)$ and not the *power allocation policy* $\mu(\cdot)$. Indeed, the battery evolution does not depend upon the particular power splitting scheme but only on the total energy consumed. Thus, $\mathbb{P}(b_{k+1} = b'|b_0)$ depends upon the policy only through the expectations $\mathbb{E}[\gamma_\mu(P^{\text{tot}}; E_k, \boldsymbol{H}, \boldsymbol{J})]$. ■

Define now the long-term probabilities of being in the energy level e given the initial level b_0 as $\pi(b|b_0) = \liminf_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} \mathbb{P}(b_k = b|b_0)$. Thanks to the above proposition, we know that $\pi(b|b_0)$ depends upon the policy only through $\mathbb{E}[\gamma_\mu(P^{\text{tot}}; b, \boldsymbol{H}, \boldsymbol{J})]$, $\forall P^{\text{tot}} \in \{0, \dots, b\}$, $\forall b \in \mathcal{B}$.

Fix a value $\alpha(P^{\text{tot}}; b)$ for every pair P^{tot} and b, and consider the set of policies Ξ that induce $\mathbb{E}[\gamma_\mu(P^{\text{tot}}; b, \boldsymbol{H}, \boldsymbol{J})] = \alpha(P^{\text{tot}}; b)$ for every pair. For every policy in Ξ, the long-term

[12] A proof of this result in the discounted horizon case can be found in [122, Theorems 6.2.9 and 6.2.10]. In our discussion we follow a different approach which will also be useful to prove Proposition 5.2.8.

probabilities are the same. The long-term average secrecy rate given an initial state b_0 can be expressed as in Equation (5.2.16)

$$C_\mu(b_0) = \sum_{b\in\mathcal{B}} \pi(b|b_0) \times \mathbb{E}\Bigg[\sum_{\boldsymbol{P}\in\mathscr{P}_{\leq}(b)} \mu(\boldsymbol{P}; b, \boldsymbol{H}, \boldsymbol{J}) c(\boldsymbol{P}, \boldsymbol{H}, \boldsymbol{J}) \Bigg]. \tag{5.B.4}$$

For every policy in Ξ, the terms $\pi(b|b_0)$ of the previous expression are the same. Therefore, in order to maximize $C_\mu(b_0)$, we focus on the terms $\mathbb{E}[\cdot]$ for each value of b. In particular, the problem can be decomposed in $b_{\max}+1$ simpler optimization problems (according to (5.2.13), define $\mu(b) \triangleq \{\mu(\cdot; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}),\ \forall \boldsymbol{\hbar}\in\mathscr{H},\ \boldsymbol{\jmath}\in\mathscr{J}\}$)

$$\max_{\mu(b)}\ \mathbb{E}\Bigg[\sum_{\boldsymbol{P}\in\mathscr{P}_{\leq}(b)} \mu(\boldsymbol{P}; b, \boldsymbol{H}, \boldsymbol{J}) c(\boldsymbol{P}, \boldsymbol{H}, \boldsymbol{J}) \Bigg], \tag{5.B.5a}$$

s.t.:

$$\text{Constraints in (5.2.15);} \tag{5.B.5b}$$

$$\mathbb{E}[\gamma_\mu(P^{\text{tot}}; b, \boldsymbol{H}, \boldsymbol{J})] = \alpha(P^{\text{tot}}; b),\ \forall P^{\text{tot}} \in \{0,\dots,b\}. \tag{5.B.5c}$$

We rewrite the first expression as follows

$$\max_{\mu(b)}\ \mathbb{E}\Bigg[\sum_{P^{\text{tot}}\in\{0,\dots,b\}} \gamma_\mu(P^{\text{tot}}; b, \boldsymbol{H}, \boldsymbol{J}) \times \sum_{\boldsymbol{P}\in\mathscr{P}_{=}(P^{\text{tot}})} \phi_\mu(\boldsymbol{P}; P^{\text{tot}}, b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) c(\boldsymbol{P}, \boldsymbol{H}, \boldsymbol{J}) \Bigg]. \tag{5.B.6}$$

where $\mathscr{P}_{=}(P^{\text{tot}}) \triangleq \{\boldsymbol{P} \ :\ \boldsymbol{P}\succeq 0,\ P^{\text{tot}} = \sum_{r=1}^{M} P^{(r)}\}$. As derived in [42, Eq. 7] with a Lagrangian approach, $\phi_\mu(\boldsymbol{P}; P^{\text{tot}}, b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) = \delta_{\boldsymbol{P}, \boldsymbol{\tau}^\star_{P^{\text{tot}}, \boldsymbol{\hbar}, j}}$ ($\phi_\mu(\cdot)$ is deterministic and there is no dependence upon b when P^{tot} is fixed). $\boldsymbol{\tau}^\star_{P^{\text{tot}}, \boldsymbol{\hbar}, \boldsymbol{\jmath}}$ is the optimal transmit power splitting given the total transmission power P^{tot} and the channel gains (we use $\boldsymbol{\tau}$ instead of $\boldsymbol{P}$ for notation clarity). Therefore, we can rewrite (5.B.6) as

$$\max_{\gamma_\mu(b)}\ \mathbb{E}\Bigg[\sum_{P^{\text{tot}}\in\{0,\dots,b\}} \gamma_\mu(P^{\text{tot}}; b, \boldsymbol{H}, \boldsymbol{J}) c(\boldsymbol{\tau}^\star_{P^{\text{tot}}, \boldsymbol{H}, \boldsymbol{J}}, \boldsymbol{H}, \boldsymbol{J}) \Bigg]. \tag{5.B.7}$$

For every fixed b, we want to define $\gamma_\mu(b) \triangleq \{\gamma_\mu(\cdot; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}),\ \forall \boldsymbol{\hbar}\in\mathscr{H},\ \boldsymbol{\jmath}\in\mathscr{J}\}$. Note that the problem is concave, thus a Lagrangian approach can be used. The Lagrangian function is

$$\mathcal{L}(b) = \mathbb{E}\Bigg[\sum_{P^{\text{tot}}\in\{0,\dots,b\}} \gamma_\mu(P^{\text{tot}}; b, \boldsymbol{H}, \boldsymbol{J}) \times \Big(c(\boldsymbol{\tau}^\star_{P^{\text{tot}}, \boldsymbol{H}, \boldsymbol{J}}, \boldsymbol{H}, \boldsymbol{J}) - \lambda(P^{\text{tot}}; b) \Big) \Bigg], \tag{5.B.8}$$

where $\lambda(P^{\text{tot}}; b)$ is the Lagrange multiplier associated with constraint $\mathbb{E}[\gamma_\mu(P^{\text{tot}}; b, \boldsymbol{H}, \boldsymbol{J})] = \alpha(P^{\text{tot}}; b)$.

We now show that an optimal policy is $\gamma_\mu(P^{\text{tot}}; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) = 1$ if $P^{\text{tot}} = P^{\text{tot}\,\star}_{b,\boldsymbol{\hbar},\boldsymbol{\jmath}}$ and zero otherwise, with

$$P^{\text{tot}\,\star}_{b,\boldsymbol{\hbar},\boldsymbol{\jmath}} = \underset{P^{\text{tot}}\in\{0,\dots,b\}}{\arg\max} \left\{ c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}},\boldsymbol{\hbar},\boldsymbol{\jmath}}, \boldsymbol{\hbar}, \boldsymbol{\jmath}) - \lambda(P^{\text{tot}}; b) \right\}. \tag{5.B.9}$$

In order to maximize Equation (5.B.8), we can focus on each argument of the expectation

$$\max_{\substack{\gamma_\mu(P^{\text{tot}};b,\boldsymbol{\hbar},\boldsymbol{\jmath}),\\ \forall P^{\text{tot}}\in\{0,\dots,b\}}} \sum_{P^{\text{tot}}\in\{0,\dots,b\}} \gamma_\mu(P^{\text{tot}}; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) \times \underbrace{\left(c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}},\boldsymbol{\hbar},\boldsymbol{\jmath}}, \boldsymbol{\hbar}, \boldsymbol{\jmath}) - \lambda(P^{\text{tot}}; b) \right)}_{u(P^{\text{tot}},b,\boldsymbol{\hbar},\boldsymbol{\jmath})}. \tag{5.B.10}$$

We recall that $\sum_{P^{\text{tot}}\in\{0,\dots,b\}} \gamma_\mu(P^{\text{tot}}; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) = 1$. (5.B.10) is a weighted sum that is maximized when $\gamma_\mu(P^{\text{tot}}; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) = 1$ if $P^{\text{tot}} = P^{\text{tot}\,\star}_{b,\boldsymbol{\hbar},\boldsymbol{\jmath}}$ and zero otherwise. Indeed, suppose by contradiction that there exist $P^{\text{tot}\prime}$ and $P^{\text{tot}\prime\prime}$ (the argument can be generalized to more than two) such that $\gamma_\mu(P^{\text{tot}\prime}; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) > 0$, $\gamma_\mu(P^{\text{tot}\prime\prime}; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) > 0$ and $\gamma_\mu(P^{\text{tot}\prime}; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) + \gamma_\mu(P^{\text{tot}}_B; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) = 1$. The max argument in (5.B.10) would be $\gamma_\mu(P^{\text{tot}\prime}; b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) u(P^{\text{tot}\prime}, b, \boldsymbol{\hbar}, \boldsymbol{\jmath}) + (1 - \gamma_\mu(P^{\text{tot}\prime}; b, \boldsymbol{\hbar}, \boldsymbol{\jmath})) u(P^{\text{tot}\prime\prime}, b, \boldsymbol{\hbar}, \boldsymbol{\jmath})$, which is smaller than or equal to $u(P^{\text{tot}\,\star}_{b,\boldsymbol{\hbar},\boldsymbol{\jmath}}, b, \boldsymbol{\hbar}, \boldsymbol{\jmath})$.

Appendix 5.C Proof of Proposition 5.2.3

The MC has three dimensions: the battery, the legitimate channel and the eavesdropper's channel. Since the fading is not controlled by the EHD, the MC is always free to move along the last two dimensions (we assume that the channel evolution is i.i.d. over time). Thus, the only potential problem is related to the battery dimension, i.e., if the policy is not unichain, the device energy level may be stuck in different subsets of $\mathcal{E}$.

Also, we recall that we consider only discrete channel conditions with non-zero probability (Remark 5.2.3.2). We now discuss Point 1). We want to show that the recurrent class is composed by the states with high energy levels, i.e., for every $b < b_{\max}$, there exists a positive probability of increasing the energy level. This is true by hypothesis because the maximum transmit power in state b is lower than the maximum number of energy arrivals $e_{\max}$ ($P^{\text{tot}}_{b,\boldsymbol{\hbar}',\boldsymbol{\jmath}'} < e_{\max}$). Therefore, since it is possible to reach the energy level $b_{\max}$ (fully charged battery) within a certain number of steps from every state, the policy is unichain. To prove Point 2), a symmetric reasoning can be followed.

If both conditions hold, it is possible to reach every $b \in \mathcal{B}$ from any element of $\mathcal{B}$, thus the policy induces an irreducible MC. Since the number of states is finite, the MC is positive recurrent.

Appendix 5.D Deriving a Unichain Policy

As in Appendix 5.C, it is always possible to move along the channel dimensions. Therefore, we focus on the battery dimension, which represents the only limitation for obtaining a unichain policy.

Consider a policy μ_A that has two recurrent classes, namely Π'_A and Π''_A (this approach can be generalized to more than two classes) and assume, without loss of generality, that if $b_0 \in \Pi''_A$ the greatest long-term reward is reached. We now propose a technique to derive a new policy that, regardless of the initial state, achieves the same maximum reward of μ_A.

Consider a second policy, namely μ_B, obtained from μ_A as follows. For every $b_A = 0, \ldots, \max\{\Pi''_A\}$, set $\boldsymbol{P}^{\mu_B}_{b_B,\boldsymbol{\hbar},\boldsymbol{j}} = \boldsymbol{P}^{\mu_A}_{b_A,\boldsymbol{\hbar},\boldsymbol{j}}$, with $b_B = b_A + b_{\max} - \max_e\{\Pi''_A\}$, i.e., we shift the recurrent class Π''_A toward higher energy levels (we name Π''_B the new recurrent class). For $b_B \in \{0, \ldots, b_{\max} - \max\{\Pi''_A\} - 1\}$, set $\boldsymbol{P}^{\mu_B}_{b_B,\boldsymbol{\hbar},\boldsymbol{j}} = 0$. In this way, the device cannot be stuck in energy levels lower than $b_{\max} - |\Pi''_B| + 1$ (the harvested energy increases the battery level) and, after a certain number of transitions, it reaches the recurrent class Π''_B. Finally, since the power splitting vectors in the recurrent classes Π''_A and Π''_B coincide, μ_B achieves the same maximum reward of μ_A, regardless of the initial b_0.

This proves that it is always possible to obtain a unichain policy with the same maximum long-term secrecy rate as the initial one and shows how to derive it.

Appendix 5.E Proof of Theorem 5.2.6

Problem (5.2.12) can be rewritten using (5.2.22) in the following form:

$$\max_{\mu} C_\mu = \max_{\mu^{\text{tot}}} \max_{\mu \in \mathcal{X}(\mu^{\text{tot}})} C_\mu \tag{5.E.1}$$

$$\mathcal{X}(\mu^{\text{tot}}) \triangleq \{\mu \ : \ \mu^{\text{tot}} \text{ and } \mu \text{ are consistent}\}, \tag{5.E.2}$$

i.e., we fix the transmission powers (outer "max") and focus on all the policies which are consistent with such choice (inner "max"). This is equivalent to searching through all the possible feasible policies (as in (5.2.12)).

Consider the expression of C_μ in Equation (5.2.23) and note that $\pi_{\mu^{\text{tot}}}(b)$ does not depend upon the particular power splitting scheme, but only upon μ^{tot}. Thus, the inner "max" can be moved inside the integral

$$\max_{\mu^{\text{tot}}} \left(\sum_{b=0}^{b_{\max}} \pi_{\mu^{\text{tot}}}(b) \times \int_{\mathcal{H}\times\mathcal{J}} \max_{\mu \in \mathcal{X}(\mu^{\text{tot}})} \Big(c(\boldsymbol{P}_{b,\boldsymbol{\hbar},\boldsymbol{j}}, \boldsymbol{\hbar}, \boldsymbol{j}) \Big) \, \mathrm{d}F(\boldsymbol{\hbar}, \boldsymbol{j}) \right). \tag{5.E.3}$$

Note that inside the integral b, $\boldsymbol{\hbar}$ and $\boldsymbol{j}$ are fixed. Therefore, the only degree of freedom in the inner "max" operation is given by the power splitting choice $\boldsymbol{P}_{b,\boldsymbol{\hbar},\boldsymbol{j}}$.

Since μ^{tot} and μ are consistent, in the inner "max" we have $\boldsymbol{P}_{b,\boldsymbol{\hbar},\boldsymbol{j}} \in \mathscr{P}_{=}(P^{\text{tot}}_{b,\boldsymbol{\hbar},\boldsymbol{j}})$ (specified in (5.2.24)). Therefore,

$$\max_{\mu \in \mathcal{X}(\mu^{\text{tot}})} \Big(c(\boldsymbol{P}_{b,\boldsymbol{\hbar},\boldsymbol{j}}, \boldsymbol{\hbar}, \boldsymbol{j}) \Big) \equiv \text{Problem (5.2.24) with } x = P^{\text{tot}}_{b,\boldsymbol{\hbar},\boldsymbol{j}} \tag{5.E.4}$$

Thus, Points 1) and 2) of the theorem solve the internal and external "max" operations, respectively.

Appendix 5.F Proof of Proposition 5.2.8

The proof exploits the results of Appendix 5.B, and in particular Equation (5.B.9). Also, we focus on the energy levels in the unique recurrent class (for the transient states the proposition is trivial to prove since $P^{\text{tot}^\star}_{b,\boldsymbol{\hbar}',\boldsymbol{j}'}$ is always zero).

Assume that $P^{\text{tot}'} \triangleq P^{\text{tot}^\star}_{b,\boldsymbol{\hbar}',\boldsymbol{j}'}$ is the optimal transmission power given the state of the system $(b, \boldsymbol{\hbar}', \boldsymbol{j}')$, i.e., $P^{\text{tot}^\star}_{b,\boldsymbol{\hbar}',\boldsymbol{j}'} = \arg\max_{P^{\text{tot}} \in \{0,\ldots,b\}} \{c(\boldsymbol{\tau}^\star_{P^{\text{tot}},\boldsymbol{\hbar}',\boldsymbol{j}'}, \boldsymbol{\hbar}', \boldsymbol{j}') - \lambda(P^{\text{tot}}; b)\}$ (we

remark that $\boldsymbol{\tau}^{\star}_{P^{\text{tot}},\boldsymbol{\hbar}',\boldsymbol{\jmath}'}$ is the *optimal* power splitting vector given P^{tot} and the channel gains). Similarly, $P^{\text{tot}''} \triangleq P^{\text{tot}^{\star}}_{b,\boldsymbol{\hbar}'',\boldsymbol{\jmath}''}$ is the optimal power for state $(b,\boldsymbol{\hbar}'',\boldsymbol{\jmath}'')$.

We first show by contradiction that if $D(P^{\text{tot}};\boldsymbol{\hbar}',\boldsymbol{\jmath}';\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') \geq 0, \forall P^{\text{tot}}$, then $P^{\text{tot}''} \geq P^{\text{tot}'}$. Assume $P^{\text{tot}'} > P^{\text{tot}''}$. We now derive some properties of $P^{\text{tot}'}$ and $P^{\text{tot}''}$ and combine these with the hypothesis to obtain the contradiction. From the definitions of $P^{\text{tot}'}$ and $P^{\text{tot}''}$, we have

$$c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}') - \lambda(P^{\text{tot}'};b) \geq c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}''},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}') - \lambda(P^{\text{tot}''};b), \tag{5.F.1}$$

$$c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - \lambda(P^{\text{tot}''};b) \geq c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}'},\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - \lambda(P^{\text{tot}'};b). \tag{5.F.2}$$

By hypothesis, we have, for every P^{tot},

$$\frac{\partial}{\partial P^{\text{tot}}}\Big(c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}},\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}')\Big) \geq 0. \tag{5.F.3}$$

Assume that the inequality is strict. This implies, for every $P_A < P_B$

$$c(\boldsymbol{\tau}^{\star}_{P_A,\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - c(\boldsymbol{\tau}^{\star}_{P_A,\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}') < c(\boldsymbol{\tau}^{\star}_{P_B,\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - c(\boldsymbol{\tau}^{\star}_{P_B,\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}'). \tag{5.F.4}$$

In particular, since $P^{\text{tot}'} > P^{\text{tot}''}$, choose $P_A = P^{\text{tot}''}$ and $P_B = P^{\text{tot}'}$ and obtain

$$c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}''},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}') < c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}'},\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}'). \tag{5.F.5}$$

Finally, by combining (5.F.2) with (5.F.5), we obtain

$$\begin{aligned} &c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}'},\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - \lambda(P^{\text{tot}'};b) + \lambda(P^{\text{tot}''};b) \\ &\leq c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') \\ &< c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}'},\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}') + c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}''},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}'), \end{aligned} \tag{5.F.6}$$

which is equivalent to

$$c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}') - \lambda(P^{\text{tot}'};b) < c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}''},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}') - \lambda(P^{\text{tot}''};b), \tag{5.F.7}$$

and violates Equation (5.F.1), leading to a contradiction.

Assume now that (5.F.3) holds with equality. Following the previous reasoning, we obtain

$$c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}''},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}') = c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}'},\boldsymbol{\hbar}'',\boldsymbol{\jmath}''},\boldsymbol{\hbar}'',\boldsymbol{\jmath}'') - c(\boldsymbol{\tau}^{\star}_{P^{\text{tot}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}') \tag{5.F.8}$$

and, instead of (5.F.7),

$$c(\boldsymbol{\tau}^{\star}_{P^{\mathrm{tot}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}')-\lambda(P^{\mathrm{tot}'};b)\leq c(\boldsymbol{\tau}^{\star}_{P^{\mathrm{tot}''},\boldsymbol{\hbar}',\boldsymbol{\jmath}'},\boldsymbol{\hbar}',\boldsymbol{\jmath}')-\lambda(P^{\mathrm{tot}''};b), \tag{5.F.9}$$

Inequality (5.F.9) must be satisfied with equality, otherwise it would violate (5.F.1). This means that, for the same state $(b,\boldsymbol{\hbar}',\boldsymbol{\jmath}')$, there exist two distinct values of P^{tot} (i.e., $P^{\mathrm{tot}'}$ and $P^{\mathrm{tot}''}$) that maximize (5.B.9). This is not possible because in the recurrent states the optimal solution is unique [14, Vol. II, Sec. 4].

The first point of Proposition 5.2.8 is thus proved. The proof of the second point is symmetric.

Appendix 5.G Proof of Theorem 5.2.9

We want to prove that, for OSP and $M=1$, $P^{\mathrm{tot}\star}_{b,\hbar,j}$ does not decrease with $\hbar$ and does not increase with j.

$D(P^{\mathrm{tot}};\hbar',\jmath';\hbar'',\jmath'')$ can be written as

$$D(P^{\mathrm{tot}};\hbar',j;\hbar'',j)=\frac{\partial}{\partial P^{\mathrm{tot}}}\left(\left[\log_2\left(\frac{1+\hbar''P^{\mathrm{tot}}}{1+jP^{\mathrm{tot}}}\right)\right]^{+}-\left[\log_2\left(\frac{1+\hbar'P^{\mathrm{tot}}}{1+jP^{\mathrm{tot}}}\right)\right]^{+}\right). \tag{5.G.1}$$

Assume $\hbar''\geq\hbar'$. If $\hbar''\leq j$, then both terms are zero because $\hbar'\leq\hbar''\leq j$. If $\hbar'\leq j<\hbar''$, then only the right term is zero. In this case, $D(P^{\mathrm{tot}};\hbar',j;\hbar'',j)\propto\hbar''-j>0$. If $j<\hbar'\leq\hbar''$, then $D(P^{\mathrm{tot}};\hbar',j;\hbar'',j)\propto\hbar''-\hbar'\geq 0$.

The proof of the second part is similar.

CHAPTER 6

Conclusions

In this thesis, we investigated how Environmental Energy Harvesting (EEH) and Wireless Energy Transfer (WET) can be used to improve the lifetime of the devices in wireless networks. We studied the problem from a communications point of view, introducing optimal and suboptimal policies that define when the nodes should transmit and their transmission parameters. Our results are always based on a dynamic programming framework, in which a Markov Chain (MC) models the battery evolution of the devices.

We focused on several different scenarios with EH capabilities. In particular, we studied and analyzed the centralized and decentralized networks with more than one energy harvesting device and their interrelations, the energy cooperation paradigm in which EEH and WET are used jointly, the wireless powered communication networks, the impact of data storage inefficiencies on the optimal policies, and finally we investigated how to maximize the secrecy rate of a communication link using optimal secrecy policies.

Although the model we presented in every chapter is different, they all share some common features. For example, the nodes are always *battery-equipped* and batteries evolve according to a random process related to EEH or WET; the performance of the networks is generally evaluated in the long run, so that we do not focus on greedy approaches that only take into consideration the current slot; the policies we introduced are *online* and thus they are *causal*, i.e., they do not require precise knowledge about the future and therefore can be used in practice.

Although different conclusions apply to each chapter, there are also some common considerations that can be made. For example, the performance of the networks saturates as the battery sizes increase; this implies that it is not necessary to equip the devices with very large batteries to achieve high performance. Another interesting aspect is that, in many cases, the optimal policy can be replaced by simpler ones that are easier to compute but still achieve high rewards (e.g., the balanced policy, the low-complexity policy, the policy derived via the approximate value iteration algorithm, the WCSP-based policy, etc.); this feature is fundamental in practice and can be considered as an important starting point for future work.

Acknowledgments

The work presented in this thesis has been supported in part by the European Commission through the FP7 EU project "Symbiotic Wireless Autonomous Powered system (SWAP)" (G.A. no. 251557, http://www.fp7-swap.eu/), Fondazione Ing. Aldo Gini, and Intel's Corporate Research Council.

List of Publications

The work presented in this thesis has in part appeared in the articles reported below.

Journal Papers.

[**J8**] **A. Biason**, Mitra, U., and Zorzi, M. (2017). Exploiting channel state information for distributed active state tracking. To be submitted to *IEEE Trans. on Wireless Communications*

[**J7**] **A. Biason** and Zorzi, M. (2017). Multicast via point to multipoint transmissions in directional 5G mmWave communications. Submitted to *IEEE Communications Magazine*

[**J6**] **A. Biason**, Dey, S., and Zorzi, M. (2016). A decentralized optimization framework for energy harvesting devices. Submitted to *IEEE Trans. on Mobile Computing, arXiv:1701.02081*

[**J5**] **A. Biason**, Pielli, C., Zanella, A., and Zorzi, M. (2016). Optimized access control for iot nodes with energy and fidelity constraints. Submitted to *IEEE Trans. on Wireless Communications, arXiv:1702.03695*

[**J4**] **A. Biason**, Pielli, C., Rossi, M., Zanella, A., Zordan, D., Kelly, M., and Zorzi, M. (2017). An energy- and context-centric perspective on IoT architecture and protocol design. *IEEE Access*, 5:6894–6908

[**J3**] **A. Biason** and Zorzi, M. (2016). Battery-powered devices in WPCNs. *IEEE Trans. Communications*, 65(1):216–229

[**J2**] **A. Biason**, Laurenti, N., and Zorzi, M. (2016). Achievable secrecy rates of an energy harvesting device. *IEEE J. Sel. Areas in Commun.*, 34(5):1502–1517

[**J1**] **A. Biason** and Zorzi, M. (2015). Joint transmission and energy transfer policies for energy harvesting devices with finite batteries. *IEEE J. Sel. Areas in Commun.*, 33(12):2626–2640

Conference and Workshop Papers.

[**C15**] **A. Biason** and Zorzi, M. (2017). Multicast in directional mmWave communications. In *Proc. 23rd European Wireless Conf. (EW)*, pages 375–381

[**C14**] Abd-Elmagid, M. A., **A. Biason**, ElBatt, T., Seddik, K. G., and Zorzi, M. (2017). Non-orthogonal multiple access schemes in wireless powered communication networks. In *Proc. IEEE Conf. on Commun. (ICC)*, pages 192–197

[**C13**] **A. Biason**, Dey, S., and Zorzi, M. (2017). Decentralized transmission policies for energy harvesting devices. In *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*

[**C12**] Pielli, C., **A. Biason**, Zanella, A., and Zorzi, M. (2016). Joint optimization of energy efficiency and data compression in TDMA-based medium access control for the IoT. In *Proc. IEEE Global Communications Conf. (GLOBECOM), IoT-LINK Workshop*

[**C11**] **A. Biason** and Zorzi, M. (2016). Transmission policies in wireless powered communication networks with energy cooperation. In *Proc. European Signal Processing Conference (Eusipco)*, pages 592–596

[**C10**] **A. Biason**, Mitra, U., and Zorzi, M. (2016). Improved active sensing performance in wireless sensor networks via channel state information. In *Proc. IEEE Symp. on Information Theory (ISIT)*, pages 2469–2473

[**C9**] **A. Biason**, Dittadi, A., and Zorzi, M. (2016). Spreading and repetitions in satellite MAC protocols. In *Proc. IEEE Conf. on Commun. (ICC)*, pages 3697–3702

[**C8**] Abd-Elmagid, M. A., **A. Biason**, ElBatt, T., Seddik, K. G., and Zorzi, M. (2016b). On optimal policies in full-duplex wireless powered communication networks. In *Proc. Symp. Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pages 243–249

[**C7**] **A. Biason** and Zorzi, M. (2016). Long-term throughput optimization in WPCN with battery-powered devices. In *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, pages 391–397

[**C6**] **A. Biason** and Zorzi, M. (2016). On the effects of battery imperfections in an energy harvesting device. In *Proc. IEEE Conf. on Computing, Networking and Communications (ICNC)*, pages 942–948

[**C5**] **A. Biason**, Khamesi, A. R., Laurenti, N., and Zorzi, M. (2015). Achievable secrecy rates of an energy harvesting device with a finite battery. In *Proc. IEEE Global Communications Conference (GLOBECOM)*

[**C4**] **A. Biason** and Zorzi, M. (2015). Energy harvesting communication system with SOC-dependent energy storage losses. In *Proc. 12th IEEE Symp. on Wireless Communication Systems (ISWCS)*, pages 406–410

[**C3**] **A. Biason** and Zorzi, M. (2015). Joint online transmission and energy transfer policies for energy harvesting devices with finite batteries. In *Proc. 21st European Wireless Conf. (EW)*, pages 318–324

[**C2**] **A. Biason** and Zorzi, M. (2015). Transmission policies for an energy harvesting device with a data queue. In *Proc. IEEE Conf. on Computing, Networking and Communications (ICNC)*, pages 189–195

[**C1**] **A. Biason**, Del Testa, D., and Zorzi, M. (2014). Low-complexity policies for wireless sensor networks with two energy harvesting devices. In *Proc. 13th IEEE Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET)*, pages 180–187

Other Publications

In addition to the main topics presented in Chapters 3-5, our research activity also covered several other themes which are briefly summarized below.

Exploiting Channel State Information for Distributed Active State Tracking

Distributed estimation is a common application in Wireless Sensor Networks (WSNs) and in the Internet of Things (IoT) in which different devices collaborate to detect a common underlying state of a system. Indeed, due to the dense nature of these networks, limited individual computational capabilities, and energy availability of the devices, data from multiple sensors can be combined to improve tracking accuracy [9, 153]. Among all the physical quantities that can be exploited for state estimation and tracking, the use of Channel State Information (CSI) has been studied only marginally to date. Indeed, most of the previous works only focused on certain contexts (e.g., target localization [118]), or channel-aware applications [23], but did not investigate how the channel is influenced by the underlying state of the system. Nevertheless, in many scenarios, the channel does depend on the state of the system, therefore CSI can be *directly* used for tracking. Our goal in [**J8**] and [**C10**] is to exploit this behavior, in addition to standard sensor measurements, to improve the tracking performance of the system.

In particular, our papers extend the framework of [173] to consider CSI; however, this model extension necessitates new methods to analyze the resulting POMDP and to develop complexity reduction techniques. The main contributions of the papers can be summarized as follows. We set up a *distributed estimation* model in which, at every time step, a set of sensors is chosen to track the underlying state of the system and report the measurements to a fusion center through a wireless communication channel. The optimal performance in terms of tracking quality and energy consumption is derived exploiting a Partially Observable Markov Decision Process (POMDP) framework for the infinite horizon setup, so that only a stationary scheduler needs to be stored in the nodes. We assume that the sensors are *passive* (i.e., they do not influence the underlying state) and heterogeneous in terms of: sensing cost, quality of the measurements and communication channel quality. As in [141, 173], and differently from most previous works [11, 25, 68–70] we consider an energy constrained fusion center. Since using the channel as an additional source of information adds a layer of complexity, we simplify the problem in different steps. First, with the goal of reducing the size of the belief space [130], we prove and then exploit

the concavity properties of the cost-to-go function to characterize a lower bound [137]. This, in conjunction with an upper bound can be used to derive the performance of the optimal sensing strategy. These bounds are instrumental in determining lower complexity, near optimal strategies. Then, we decompose the tracking procedure into a simpler set of operations and cast a multi-dimensional problem as a set of simpler lower-dimensional sub-problems. Finally, we propose a suboptimal greedy technique which further simplifies the optimization process. Numerical results support the importance of considering channel and measurements jointly and validate our theoretical results. In particular, we compare our new method with another without CSI (this is similar to the one proposed in [173]) and show that the channel information can improve the estimation capabilities of the system by a factor of two. While we use the Wireless Body Area Network (WBAN) scenario as an exemplar for the numerical evaluations, the proposed model adopts very general assumptions and the theoretical framework can be applied to a large variety of applications (e.g., object tracking, indoor environmental monitoring, etc.).

Acknowledgment. This activity is the result of a collaboration with Prof. Urbashi Mitra, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles.

On the Energy/Distortion Trade-off in the IoT

The presence of many battery-powered sensors in the Internet of things paradigm appeals for the design of energy-aware protocols. Source coding techniques make it possible to save some energy by compressing the packets sent over the network, but at the cost of a poorer accuracy in the representation of the data. Our paper **[J5]** addresses the problem of designing efficient policies to jointly perform processing and transmission tasks. In particular, we aim at defining an optimal scheduling strategy with the twofold goal of extending the network lifetime and guaranteeing a low overall distortion of the transmitted data. We propose a Time Division Multiple Access (TDMA)-based access scheme that optimally allocates resources to heterogeneous nodes. We use realistic rate-distortion curves to quantify how compression impacts on the data quality and propose a complete energy model that includes the energy spent for processing and transmitting the data, as well as the circuitry energy costs. We consider both full and statistical knowledge of the wireless channels, and derive optimal policies for the two cases. The overall problem is structured in modules and solved through convex and alternate programming techniques. Finally, we thoroughly evaluate the proposed algorithms and the influence of the design variables on the system performance adopting parameters of real sensors.

Acknowledgment. This activity was developed as part of a collaboration with Intel Labs Europe.

Multicast via Point to Multipoint Transmissions in Directional 5G mmWave Communications

Sustaining the increasing demand of new communication services, especially those driven by multimedia applications (e.g., mobile video), is becoming very difficult with traditional microwave wireless transmissions. On the other hand, in the past few years, new studies on the propagation characteristics of millimeter wave (mmWave) frequencies as well as advances in the hardware design of RF components have shown that using mmWave bands for the new fifth generation of cellular systems is actually feasible [125]. The new

communication frequencies in the mmWave spectrum go from tens to hundreds of GHz (e.g., $30 - 300$ GHz). The consequences of such changes are huge and involve both the available bandwidth, which can be impressively increased, and the design of new devices in which it may be possible to install large arrays of antennas thanks to their reduced dimension. However, to fully exploit the benefits of this new technology, the communication layer has to be redefined, as mmWave links are substantially different from traditional ones.

On the other hand, a well known technique to increase the bandwidth efficiency in traditional communication is the use of *multicast*. Multicast links, by exploiting the properties of point-to-multipoint wireless communications, allow one device (e.g., the access point) to transmit the same information to many devices (e.g., smart phones, tablets, smart TVs, etc.) simultaneously using the same band, modulation and coding scheme. This technique can be naturally exploited in the microwave spectrum, as a transmission is generally omnidirectional. However, its application to directional communication links, such as those at mmWave, poses new challenges, that have been studied only marginally to date.

The goal of our papers **[J7]** and **[C15]** is to describe how multicast can be realized in a mmWave system, with particular focus on the design of the transmission beams and retransmissions. We focus on an analytical model that captures the trade-off between transmitting more redundancy and consuming more resources at the access point using directional communications. Our results show that, despite the difficulties due to directionality, it is possible to use multicast with mmWaves and that by doing so the performance of the system is significantly improved with respect to sequential unicast schemes.

Spreading and Repetitions in Satellite MAC Protocols

Supporting networks with a large number of devices which generate sporadic data (e.g., wireless sensor networks, monitoring applications, position reporting, signaling, etc.) is a key issue in satellite communications. In this kind of scenario, contention free protocols, e.g., Demand Assignment Multiple Access (DAMA) [73], do not provide good performance because of their high overhead. Instead, random access schemes, in which users are not coordinated and collisions may occur, are well suited for uploading packets from ground nodes to a common satellite receiver and are recognized as one of the most appealing solutions. In the past few years, the well-known Aloha protocol has been improved and several new schemes have been proposed. The main focus was on diversity slotted Aloha-like protocols [21, 22, 28, 81, 115] and spread spectrum techniques [2, 32, 117]. The goal of our paper **[C9]** is to study and combine these two approaches. In particular, we discuss the trade-offs between replicas and spreading by combining the two paradigms in a hybrid Asynchronous Contention Resolution Diversity Aloha (ACRDA) [28] protocol. Practically, our algorithm uses the ACRDA scheme and spreads the replicas in order to further reduce the packet loss rate. The result is a highly tunable protocol which can be dynamically adapted to the requirements of different devices (e.g., in terms of power) as well as to the current traffic conditions. The new technique is shown to perform better than the traditional ACRDA or Enhanced Spread Spectrum Aloha (E-SSA) [32] protocols in terms of packet loss rates when the offered traffic is low (stable region).

Acknowledgment. This activity has been carried out under a programme funded by the European Space Agency.
(Satnex IV Project - ESA/ESTEC Contr. No.: 4000113177/15/NL/CLP)

Non-Orthogonal Multiple Access Schemes in Wireless Powered Communication Networks

In our paper [**C14**] we characterize time and power allocations to optimize the sum-throughput of a Wireless Powered Communication Network (WPCN) with Non-Orthogonal Multiple Access (NOMA). In our setup, an Energy Rich Node (ERN) broadcasts wireless energy to several devices, which use it to simultaneously transmit data to an Access Point (AP) in uplink. Differently from most prior works, in [**C14**] we consider a generic scenario, in which the ERN and AP do not coincide, i.e., are two separate entities. We study two NOMA decoding schemes, namely Low Complexity Decoding (LCD) and Successive Interference Cancellation Decoding (SICD). For each scheme, we formulate a sum-throughput optimization problem over a finite horizon. Despite the complexity of the LCD optimization problem, due to its non-convexity, we recast it into a series of geometric programs. On the other hand, we establish the convexity of the SICD optimization problem and propose an algorithm to find its optimal solution. Our numerical results demonstrate the importance of using successive interference cancellation in WPCNs with NOMA, and show how the energy should be distributed as a function of the system parameters.

Acknowledgment. This paper was produced in collaboration with the Department of Electronics and Electrical Communications Engineering, Cairo University, Giza, Egypt.

References

[1] Abd-Elmagid, M. A., ElBatt, T., and Seddik, K. G. (2016). A generalized optimization framework for wireless powered communication networks. *arXiv:1603.01115*.

[2] Abramson, N. (1994). Multiple access in wireless digital networks. *Proceedings of the IEEE*, 82(9):1360–1370.

[3] Agrawal, V. (2015). Security and privacy issues in wireless sensor networks for healthcare. In *Internet of Things. User-Centric IoT*, pages 223–228. Springer.

[4] Al Ameen, M., Liu, J., and Kwak, K. (2012). Security and privacy issues in wireless sensor networks for healthcare applications. *J. of Medical Systems*, 36(1):93–101.

[5] Alrajeh, N. A., Khan, S., Lloret, J., and Loo, J. (2013). Secure routing protocol using cross-layer design and energy harvesting in wireless sensor networks. *Int. J. Distributed Sensor Networks*.

[6] Alsheikh, M. A., Hoang, D. T., Niyato, D., Tan, H.-P., and Lin, S. (2015). Markov decision processes with applications in wireless sensor networks: A survey. *IEEE Commun. Surveys & Tutorials*, 17(3):1239–1267.

[7] Amato, C., Chowdhary, G., Geramifard, A., Üre, N. K., and Kochenderfer, M. J. (2013). Decentralized control of partially observable Markov decision processes. In *Proc. 52nd IEEE Conf. on Decision and Control*, pages 2398–2405.

[8] Arafa, A. and Ulukus, S. (2014). Single-user and multiple access channels with energy harvesting transmitters and receivers. In *Global Conference on Signal and Information Processing (GlobalSIP)*, pages 213–217.

[9] Aráuz, J., Krishnamurthy, P., and Labrador, M. A. (2004). Discrete Rayleigh fading channel modeling. *Wireless Communications and Mobile Computing*, 4(4):413–425.

[10] Arroyo-Valles, R., Marques, A., and Cid-Sueiro, J. (2011). Optimal selective forwarding for energy saving in wireless sensor networks. *IEEE Trans. Wireless Commun.*, 10(1):164–175.

[11] Au, L. K., Bui, A. A., Batalin, M. A., and Kaiser, W. J. (2012). Energy-efficient context classification with dynamic sensor control. *IEEE Trans. Biomedical Circuits and Systems*, 6(2):167–178.

[12] Baldi, M., Chiaraluce, F., Laurenti, N., Tomasin, S., and Renna, F. (2014). Secrecy transmission on parallel channels: Theoretical limits and performance of practical codes. *IEEE Trans. Inf. Forens. and Sec.*, 9(11):1765–1779.

[13] Beeby, S. P., Tudor, M. J., and White, N. M. (2006). Energy harvesting vibration sources for microsystems applications. *Measurement Science and Technology*, 17(12):R175–R195.

[14] Bertsekas, D. (2005). *Dynamic programming and optimal control*. Athena Scientific, Belmont, Massachusetts.

[15] Bertsekas, D. and Tsitsiklis, J. N. (1996). *Neuro-Dynamic Programming*. Athena Scientific, Belmont, Massachusetts.

[16] Bi, S., Ho, C. K., and Zhang, R. (2015). Wireless powered communication: opportunities and challenges. *IEEE Commun. Magazine*, 53(4):117–125.

[17] Blasco, P., Gunduz, D., and Dohler, M. (2013). A learning theoretic approach to energy harvesting communication system optimization. *IEEE Trans. Wireless Commun.*, 12(4):1872–1882.

[18] Bloch, M. and Barros, J. (2011). *Physical-layer security: from information theory to security engineering*. Cambridge University Press.

[19] Brown, W. C. (1984). The history of power transmission by radio waves. *IEEE Trans. Microwave Theory and Techniques*, 32(9):1230–1242.

[20] Bruce, N., Kang, Y., Kim, H. R., Park, S., and Lee, H.-J. (2015). A security protocol based on mutual authentication application toward wireless sensor network. In *Information Science and Applications*, pages 27–34. Springer.

[21] Bui, H.-C., Lacan, J., and Boucheret, M.-L. (2012). An enhanced multiple random access scheme for satellite communications. In *Proc. IEEE Wireless Telecommunications Symposium (WTS)*.

[22] Casini, E., De Gaudenzi, R., and Herrero, O. R. (2007). Contention resolution diversity slotted ALOHA (CRDSA): An enhanced random access scheme for satellite access packet networks. *IEEE Trans. Wireless Commun.*, 6(4):1408–1419.

[23] Chen, B., Tong, L., and Varshney, P. K. (2006). Channel aware distributed detection in wireless sensor networks. *IEEE Signal Processing Magazine*, 23(4):16–26.

[24] Chen, H., Li, Y., Luiz Rebelatto, J., Uchoa-Filho, B. F., and Vucetic, B. (2015). Harvest-then-cooperate: Wireless-powered cooperative communications. *IEEE Trans. Signal Processing*, 63(7):1700–1711.

[25] Chen, Y., Zhao, Q., Krishnamurthy, V., and Djonin, D. (2007). Transmission scheduling for optimizing sensor network lifetime: A stochastic shortest path approach. *IEEE Trans. Signal Processing*, 55(5):2294–2309.

[26] Coarasa, A., Nintanavongsa, P., Sanyal, S., and Chowdhury, K. (2013). Impact of mobile transmitter sources on radio frequency wireless energy harvesting. In *Conf. on Computing, Networking and Communications (ICNC)*, pages 573–577.

[27] Csiszár, I. and Korner, J. (1978). Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, 24(3):339–348.

[28] De Gaudenzi, R., del Rio Herrero, O., Acar, G., and Garrido Barrabes, E. (2014). Asynchronous contention resolution diversity ALOHA: Making CRDSA truly asynchronous. *IEEE Trans. Wireless Commun.*, 13(11):6193–6206.

[29] De Rossi, F., Pontecorvo, T., and Brown, T. M. (2015). Characterization of photovoltaic devices for indoor light harvesting and customization of flexible dye solar cells to deliver superior efficiency under artificial lighting. *Applied Energy*, 156:413–422.

[30] Dechter, R. (1997). Bucket elimination: a unifying framework for processing hard and soft constraints. *Constraints*, 2(1):51–55.

[31] Dechter, R. (2003). *Constraint processing*. Morgan Kaufmann.

[32] del Rio Herrero, O. and De Gaudenzi, R. (2012). High efficiency satellite multiple access scheme for machine-to-machine communications. *IEEE Trans. Aero. and Elect. Syst.*, 48(4):2961–2989.

[33] Del Testa, D., Michelusi, N., and Zorzi, M. (2016). Optimal transmission policies for two-user energy harvesting device networks with limited state-of-charge knowledge. *IEEE Trans. Wireless Commun.*, 15:1393–1405.

[34] Devillers, B. and Gunduz, D. (2012). A general framework for the optimization of energy harvesting communication systems with battery imperfections. *IEEE J. of Commun. and Netw.*, 14(2):130–139.

[35] Dibangoye, J. S., Amato, C., Buffet, O., and Charpillet, F. (2014). Optimally solving Dec-POMDPs as continuous-state MDPs: theory and algorithms. *INRIA*.

[36] Dibangoye, J. S., Amato, C., and Doniec, A. (2012). Scaling up decentralized MDPs through heuristic search. In *Proc. Uncertainty in Artificial Intelligence (UAI)*.

[37] Dibangoye, J. S., Amato, C., Doniec, A., and Charpillet, F. (2013). Producing efficient error-bounded solutions for transition independent decentralized MDPs. In *Proc. ACM Conf. Autonomous Agents and Multi-agent Systems (AAMAS)*.

[38] Dolgov, A., Zane, R., and Popovic, Z. (2010). Power management system for online low power RF energy harvesting optimization. *IEEE Trans. Circuits and Systems I: Regular Papers*, 57(7):1802–1811.

[39] Erturk, A. and Inman, D. J. (2011). *Piezoelectric energy harvesting*. John Wiley & Sons.

[40] Eu, Z. A., Tan, H.-P., and Seah, W. K. (2011). Design and performance analysis of MAC schemes for wireless sensor networks powered by ambient energy harvesting. *Elsevier Ad Hoc Networks*, 9(3):300–323.

[41] Gakkestad, J. and Hanssen, L. (2011). Powering wireless sensor networks nodes in northern Europe using solar cell panel for energy harvesting. In *Proc. 4th IEEE IFIP Conf. on New Technologies, Mobility and Security (NTMS)*.

[42] Gopala, P. K., Lai, L., and El Gamal, H. (2008). On the secrecy capacity of fading channels. *IEEE Trans. Inf. Theory*, 54(10):4687–4698.

[43] Gorlatova, M., Wallwater, A., and Zussman, G. (2013). Networking low-power energy harvesting devices: Measurements and algorithms. *IEEE Trans. Mobile Computing*, 12(9):1853–1865.

[44] Grover, P. and Sahai, A. (2010). Shannon meets Tesla: wireless information and power transfer. In *IEEE Symp. on Information Theory Proceedings (ISIT)*, pages 2363–2367.

[45] Gunduz, D., Stamatiou, K., Michelusi, N., and Zorzi, M. (2014). Designing intelligent energy harvesting communication systems. *IEEE Commun. Magazine*, 52(1):210–216.

[46] Gurakan, B., Ozel, O., Yang, J., and Ulukus, S. (2012). Two-way and multiple-access energy harvesting systems with energy cooperation. In *Proc. 46th IEEE Asilomar Conf. on Signals, Systems and Computers (ASILOMAR)*, pages 58–62.

[47] Gurakan, B., Ozel, O., Yang, J., and Ulukus, S. (2013). Energy cooperation in energy harvesting communications. *IEEE Trans. Communications*, 61(12):4884–4898.

[48] Gurakan, B. and Ulukus, S. (2014). Energy harvesting diamond channel with energy cooperation. In *IEEE Symp. on Information Theory (ISIT)*, pages 986–990.

[49] Hagerty, J., Helmbrecht, F. B., McCalpin, W. H., Zane, R., Popović, Z. B., et al. (2004). Recycling ambient microwave energy with broad-band rectenna arrays. *IEEE Trans. Microwave Theory and Techniques*, 52(3):1014–1024.

[50] Hauskrecht, M. (2000). Value-function approximations for partially observable Markov decision processes. *Journal of Artificial Intelligence Research*, 13:33–94.

[51] Ho, C. K., Khoa, P. D., and Ming, P. C. (2010). Markovian models for harvested energy in wireless communications. In *IEEE Conf. on Communication Systems (ICCS)*, pages 311–315.

[52] Hoang, D. T., Niyato, D., Wang, P., and Kim, D. I. (2014). Optimal decentralized control policy for wireless communication systems with wireless energy transfer capability. In *Proc. IEEE Conf. on Commun. (ICC)*, pages 2835–2840.

[53] Hui, S., Zhong, W., and Lee, C. (2014). A critical review of recent progress in mid-range wireless power transfer. *IEEE Trans. Power Electronics*, 29(9):4500–4511.

[Infinite Power Solutions] Infinite Power Solutions. THINERGY MEC200 Series.

[55] Jaggi, N., Kar, K., and Krishnamurthy, A. (2007). Rechargeable sensor activation under temporally correlated events. In *5th Symp. on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops (WiOpt)*.

[56] Jakobsen, M. K., Madsen, J., and Hansen, M. R. (2010). DEHAR: a distributed energy harvesting aware routing algorithm for ad-hoc multi-hop wireless sensor networks. In *IEEE Symp. on a World of Wireless Mobile and Multimedia Networks (WoWMoM)*, pages 1–9.

[57] Ju, H. and Zhang, R. (2014a). Optimal resource allocation in full-duplex wireless-powered communication network. *IEEE Trans. Communications*, 62(10):3528–3540.

[58] Ju, H. and Zhang, R. (2014b). Throughput maximization in wireless powered communication networks. *IEEE Trans. Wireless Commun.*, 13(1):418–428.

[59] Ju, H. and Zhang, R. (2014c). User cooperation in wireless powered communication networks. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, pages 1430–1435.

[60] Kang, X., Ho, C. K., and Sun, S. (2015). Full-duplex wireless-powered communication network with energy causality. *IEEE Trans. Wireless Commun.*, 14(10):5539–5551.

[61] Kaushik, K., Mishra, D., De, S., Basagni, S., Heinzelman, W., Chowdhury, K., and Jana, S. (2013). Experimental demonstration of multi-hop RF energy transfer. In *24th Symp. on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 538–542.

[62] Khan, M., Misic, J., and Misic, V. (2015). Impact of network load on the performance of a polling MAC with wireless recharging of nodes. *IEEE Trans. Emerging Topics in Computing*, 3(3):307–316.

[63] Khandaker, M. R. and Wong, K.-K. (2014). SWIPT in MISO multicasting systems. *IEEE Wireless Communications Letters*, 3(3):277–280.

[64] Kim, D. I. and Choi, K. W. (2016). Stochastic optimal control for wireless powered communication networks. *IEEE Trans. Wireless Commun.*, 15(1):686–698.

[65] Kim, J. and Lee, J.-W. (2011). Energy adaptive MAC protocol for wireless sensor networks with RF energy transfer. In *Third Conf. on Ubiquitous and Future Networks (ICUFN)*, pages 89–94.

[66] Korf, R. E. (1990). Real-time heuristic search. *Artificial intelligence*, 42(2-3):189–211.

[67] Krikidis, I., Timotheou, S., and Sasaki, S. (2012). RF energy transfer for cooperative networks: Data relaying or energy harvesting? *IEEE Commun. Letters*, 16(11):1772–1775.

[68] Krishnamurthy, V. (2002). Algorithms for optimal scheduling and management of hidden Markov model sensors. *IEEE Trans. Signal Processing*, 50(6):1382–1397.

[69] Krishnamurthy, V. (2013). How to schedule measurements of a noisy Markov chain in decision making? *IEEE Trans. Information Theory*, 59(7):4440–4461.

[70] Krishnamurthy, V. and Djonin, D. V. (2007). Structured threshold policies for dynamic sensor scheduling–A partially observed Markov decision process approach. *IEEE Trans. Signal Processing*, 55(10):4938–4957.

[71] Kurs, A., Karalis, A., Moffatt, R., Joannopoulos, J. D., Fisher, P., and Soljačić, M. (2007). Wireless power transfer via strongly coupled magnetic resonances. *Science*, 317(5834):83–86.

[72] Kurs, A., Moffatt, R., and Soljacic, M. (2010). Simultaneous mid-range power transfer to multiple devices. *Applied Physics Letters*, 96(4).

[73] Le-Ngoc, T., Mohammed, J., et al. (1993). Combined free/demand assignment multiple access (CFDAMA) protocols for packet satellite communications. In *Proc. IEEE 2nd Int. Conference on Personal Communications: Gateway to the 21st Century*, volume 2, pages 824–828.

[74] Lei, J., Yates, R., and Greenstein, L. (2009). A generic model for optimizing single-hop transmission policy of replenishable sensors. *IEEE Trans. Wireless Commun.*, 8(2):547–551.

[75] Levin, D. A., Peres, Y., and Wilmer, E. L. (2009). *Markov chains and mixing times*. American Mathematical Soc.

[76] Li, H., Jaggi, N., and Sikdar, B. (2011). Relay scheduling for cooperative communications in sensor networks with energy harvesting. *IEEE Trans. Wireless Commun.*, 10(9):2918–2928.

[77] Li, Q., Ma, W.-K., and So, A. M.-C. (2014). Robust artificial noise-aided transmit optimization for achieving secrecy and energy harvesting. In *Proc. IEEE Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1596–1600.

[78] Liang, Y., Poor, H. V., and Shamai (Shitz), S. (2008). Secure communication over fading channels. *IEEE Trans. Inf. Theory*, 54(6):2470–2492.

[79] Liu, L., Zhang, R., and Chua, K.-C. (2013). Wireless information transfer with opportunistic energy harvesting. *IEEE Trans. Wireless Commun.*, 12(1):288–300.

[80] Liu, L., Zhang, R., and Chua, K.-C. (2014). Multi-antenna wireless powered communication with energy beamforming. *IEEE Trans. Communications*, 62(12):4349–4361.

[81] Liva, G. (2011). Graph-based analysis and optimization of contention resolution diversity slotted ALOHA. *IEEE Trans. Commun.*, 59(2):477–487.

[82] López, J. and Zhou, J. (2008). *Wireless sensor network security*, volume 1. Ios Press.

[83] Lovejoy, W. S. (1991). Computationally feasible bounds for partially observed Markov decision processes. *Operations research*, 39(1):162–175.

[84] Lu, X., Wang, P., Niyato, D., Kim, D. I., and Han, Z. (2015). Wireless networks with RF energy harvesting: A contemporary survey. *IEEE Commun. Surveys Tutorials*, 17(2):757–789.

[85] Mahdavi-Doost, H. and Yates, R. (2013). Energy harvesting receivers: Finite battery capacity. In *IEEE Symp. on Information Theory Proceedings (ISIT)*, pages 1799–1803.

[86] Mahdavi-Doost, H. and Yates, R. D. (2014). Fading channels in energy-harvesting receivers. In *48th Annual Conference on Information Sciences and Systems (CISS)*.

[87] Martinez, G., Li, S., and Zhou, C. (2014). Wastage-aware routing in energy-harvesting wireless sensor networks. *IEEE Sensors Journal*, 14(9):2967–2974.

[88] Maurer, U. M. (1993). Secret key agreement by public discussion from common information. *IEEE Trans. Inf. Theory*, 39(3):733–742.

[89] Michelusi, N., Badia, L., Carli, R., Corradini, L., and Zorzi, M. (2013a). Energy management policies for harvesting-based wireless sensor devices with battery degradation. *IEEE Trans. Communications*, 61(12):4934–4947.

[90] Michelusi, N., Badia, L., and Zorzi, M. (2014). Optimal transmission policies for energy harvesting devices with limited state-of-charge knowledge. *IEEE Trans. Communications*, 62(11):3969–3982.

[91] Michelusi, N. and Levorato, M. (2017). Energy-based adaptive multiple access in LPWAN IoT systems with energy harvesting. In *Proc. IEEE Symp. on Information Theory (ISIT)*, pages 1112–1116.

[92] Michelusi, N., Stamatiou, K., and Zorzi, M. (2012). On optimal transmission policies for energy harvesting devices. In *Proc. IEEE Information Theory and Applications Workshop (ITA)*, pages 249–254.

[93] Michelusi, N., Stamatiou, K., and Zorzi, M. (2013b). Transmission policies for energy harvesting sensors with time-correlated energy supply. *IEEE Trans. Communications*, 61(7):2988–3001.

[94] Michelusi, N. and Zorzi, M. (2015). Optimal adaptive random multiaccess in energy harvesting wireless sensor networks. *IEEE Trans. Communications*, 63(4):1355–1372.

[95] Mohrehkesh, S. and Weigle, M. C. (2014). Optimizing energy consumption in Tera-Hertz band nanonetworks. *IEEE J. Sel. Areas in Commun.*, 32(12):2432–2441.

[96] Morsi, R., Michalopoulos, D. S., and Schober, R. (2015). Performance analysis of wireless powered communication with finite/infinite energy storage. In *Proc. IEEE Conf. on Commun. (ICC)*, pages 2469–2475.

[97] Mukherjee, A., Fakoorian, S. A., Huang, J., Swindlehurst, A. L., et al. (2014). Principles of physical layer security in multiuser wireless networks: A survey. *IEEE Commun. Surveys & Tutorials*, 16(3):1550–1573.

[98] Mukherjee, A. and Huang, J. (2012). Deploying multi-antenna energy-harvesting cooperative jammers in the MIMO wiretap channel. In *Proc. 46th IEEE Asilomar Conf. on Signals, Systems and Computers (ASILOMAR)*, pages 1886–1890.

[99] Naderi, M., Nintanavongsa, P., and Chowdhury, K. (2014). RF-MAC: A medium access control protocol for re-chargeable sensor networks powered by wireless energy harvesting. *IEEE Trans. Wireless Commun.*, 13(7):3926–3937.

[100] Nasir, A., Zhou, X., Durrani, S., Kennedy, R., et al. (2013). Relaying protocols for wireless energy harvesting and information processing. *IEEE Trans. Communications*, 12(7):3622–3636.

[101] National Renewable Energy Laboratory (NREL), US DOE (2014). Measurement and Instrumentation Data Center. *http://midcdmz.nrel.gov/ecsu/*.

[102] Ng, D. W. K., Lo, E. S., and Schober, R. (2014). Robust beamforming for secure communication in systems with wireless information and power transfer. *IEEE Trans. Wireless Commun.*, 13(8):4599–4615.

[103] Ni, W. and Dong, X. (2015). Energy harvesting wireless communications with energy cooperation between transmitter and receiver. *IEEE Trans. Communications*, 63(4):1457–1469.

[104] Nintanavongsa, P., Muncuk, U., Lewis, D. R., and Chowdhury, K. R. (2012). Design optimization and implementation for RF energy harvesting circuits. *IEEE J. Emerging and Selected Topics in Circuits and Systems*, 2(1):24–33.

[105] Nintanavongsa, P., Naderi, M., and Chowdhury, K. (2013). Medium access control protocol design for sensors powered by wireless energy transfer. In *Proc. IEEE Int. Conf. on Computer Communications (INFOCOM)*, pages 150–154.

[106] Niyato, D., Hossain, E., and Fallahi, A. (2007). Sleep and wakeup strategies in solar-powered wireless sensor/mesh networks: Performance analysis and optimization. *IEEE Transactions on Mobile Computing*, 6(2):221–236.

[107] Niyato, D., Wang, P., and Kim, D. I. (2014). Admission control policy for wireless networks with RF energy transfer. In *Proc. IEEE Conf. on Commun. (ICC)*, pages 1118–1123.

[108] Oggier, F. and Hassibi, B. (2011). The secrecy capacity of the MIMO wiretap channel. *IEEE Trans. Inf. Theory*, 57(8):4961–4972.

[109] Ozel, O., Ekrem, E., and Ulukus, S. (2012a). Gaussian wiretap channel with a batteryless energy harvesting transmitter. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 89–93.

[110] Ozel, O., Ekrem, E., and Ulukus, S. (2012b). Gaussian wiretap channel with an amplitude constraint. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 139–143.

[111] Ozel, O., Tutuncuoglu, K., Yang, J., Ulukus, S., and Yener, A. (2011). Transmission with energy harvesting nodes in fading wireless channels: Optimal policies. *IEEE J. Sel. Areas in Commun.*, 29(8):1732–1743.

[112] Ozel, O. and Ulukus, S. (2011). AWGN channel under time-varying amplitude constraints with causal information at the transmitter. In *Proc. 45th IEEE Asilomar Conf. on Signals, Systems and Computers (ASILOMAR)*, pages 373–377.

[113] Ozel, O. and Ulukus, S. (2012). Achieving AWGN capacity under stochastic energy harvesting. *IEEE Trans. Inf. Theory*, 58(10):6471–6483.

[114] Pandey, A. and Tripathi, R. (2010). A survey on wireless sensor networks security. *Int. J. Computer Applications*, 3(2):43–49.

[115] Paolini, E., Liva, G., and Chiani, M. (2011). High throughput random access via codes on graphs: Coded slotted ALOHA. In *Proc. IEEE International Conference on Communications (ICC)*.

[116] Park, J. and Clerckx, B. (2013). Joint wireless information and energy transfer in a two-user MIMO interference channel. *IEEE Trans. Wireless Commun.*, 12(8):4210–4221.

[117] Pateros, C. (2000). Novel direct sequence spread spectrum multiple access technique. In *Proc. IEEE 21st Century Military Communications Conference (MILCOM)*, volume 2, pages 564–568.

[118] Paul, A. S. and Wan, E. (2009). RSSI-based indoor localization and tracking using sigma-point Kalman smoothers. *IEEE J. Sel. Topics in Signal Processing*, 3(5):860–873.

[119] Perrig, A., Stankovic, J., and Wagner, D. (2004). Security in wireless sensor networks. *Communications of the ACM*, 47(6):53–57.

[120] Piñuela, M., Mitcheson, P. D., and Lucyszyn, S. (2013). Ambient RF energy harvesting in urban and semi-urban environments. *IEEE Trans. Microwave Theory and Techniques*, 61(7):2715–2726.

[121] Powercast Corporation (2016). TX91501 User's Manual. *http://www.powercastco.com/products/powercaster-transmitter/*.

[122] Puterman, M. L. (1995). *Markov decision processes: Discrete stochastic dynamic programming*. John Wilson and Sons Ed.

[123] Raghunathan, V., Kansal, A., Hsu, J., Friedman, J., and Srivastava, M. (2005). Design considerations for solar energy harvesting wireless embedded systems. In *Proc. 4th IEEE Symp. on Information Processing in Sensor Networks (IPSN)*, pages 457–462.

[124] RamRakhyani, A., Mirabbasi, S., and Chiao, M. (2011). Design and optimization of resonance-based efficient wireless power delivery systems for biomedical implants. *IEEE Trans. Biomedical Circuits and Systems*, 5(1):48–63.

[125] Rappaport, T. S., Sun, S., Mayzus, R., Zhao, H., Azar, Y., Wang, K., Wong, G. N., Schulz, J. K., Samimi, M., and Gutierrez, F. (2013). Millimeter wave mobile communications for 5G cellular: It will work! *IEEE Access*, 1:335–349.

[126] Ross, K. W. (1989). Randomized and past-dependent policies for Markov decision processes with multiple constraints. *Operations Research*, 37(3):474–477.

[127] Roundy, S. J. (2003). *Energy scavenging for wireless sensor nodes with a focus on vibration to electricity conversion*. PhD thesis, Department of Mechanical Engineering, University of California, Berkeley.

[128] Sample, A., Meyer, D., and Smith, J. (2011). Analysis, experimental results, and range adaptation of magnetically coupled resonators for wireless power transfer. *IEEE Trans. Industrial Electronics*, 58(2):544–554.

[129] Seyedi, A. and Sikdar, B. (2010). Energy efficient transmission strategies for body sensor networks with energy harvesting. *IEEE Trans. Communications*, 58(7):2116–2126.

[130] Shani, G., Pineau, J., and Kaplow, R. (2013). A survey of point-based POMDP solvers. *Autonomous Agents and Multi-Agent Systems*, 27(1):1–51.

[131] Shannon, C. (1949). Communication theory of secrecy systems. *Bell System Tech. Journ.*, 28(4):656–715.

[132] Sharma, V., Mukherji, U., Joseph, V., and Gupta, S. (2010). Optimal energy management policies for energy harvesting sensor nodes. *IEEE Trans. Wireless Commun.*, 9(4):1326–1336.

[133] Sharma, V. and Rajesh, R. (2011). Queuing theoretic and information theoretic capacity of energy harvesting sensor nodes. In *Proc. 45th IEEE Asilomar Conf. on Signals, Systems and Computers (ASILOMAR)*, pages 383–388.

[134] Shaviv, D. and Özgür, A. (2016). Universally near-optimal online power control for energy harvesting nodes. *IEEE J. Sel. Areas in Commun.*, 34(12):3620–3631.

[135] Shi, Q., Liu, L., Xu, W., and Zhang, R. (2014). Joint transmit beamforming and receive power splitting for MISO SWIPT systems. *IEEE Trans. Wireless Commun.*, 13(6):3269–3280.

[136] Shi, Y., Xie, L., Hou, Y., and Sherali, H. (2011). On renewable sensor networks with wireless energy transfer. In *Proc. IEEE Int. Conf. on Computer Communications (INFOCOM)*, pages 1350–1358.

[137] Smallwood, R. D. and Sondik, E. J. (1973). The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 21(5):1071–1088.

[138] Smith, J. R. (2013). *Wirelessly Powered Sensor Networks and Computational RFID*. Springer Science & Business Media.

[139] Tan, Y. K. and Panda, S. K. (2011). Energy harvesting from hybrid indoor ambient light and thermal energy sources for enhanced performance of wireless sensor nodes. *IEEE Trans. Industrial Electronics*, 58(9):4424–4435.

[140] Texas Instruments (2013). MSP430 SoC with RF Core. *http://www.ti.com/product/cc430f6137*.

[141] Thatte, G., Li, M., Lee, S., Emken, A., Annavaram, M., Narayanan, S., Spruijt-Metz, D., and Mitra, U. (2011). Optimal time-resource allocation for energy-efficient physical activity detection. *IEEE Trans. Signal Processing*, 59(4):1843–1857.

[142] Timotheou, S., Krikidis, I., Zheng, G., and Ottersten, B. (2014). Beamforming for MISO interference channels with QoS and RF energy transfer. *IEEE Trans. Wireless Commun.*, 13(5):2646–2658.

[143] ToulBar2 (2016). an open source WCSP solver. *https://mulcyber.toulouse.inra.fr/projects/toulbar2/*.

[144] Tutuncuoglu, K. and Yener, A. (2012a). Communicating with energy harvesting transmitters and receivers. In *Proc. IEEE Information Theory and Applications Workshop (ITA)*, pages 240–245.

[145] Tutuncuoglu, K. and Yener, A. (2012b). Optimum transmission policies for battery limited energy harvesting nodes. *IEEE Trans. Wireless Commun.*, 11(3):1180–1189.

[146] Tutuncuoglu, K. and Yener, A. (2013a). Cooperative energy harvesting communications with relaying and energy sharing. In *IEEE Information Theory Workshop (ITW)*.

[147] Tutuncuoglu, K. and Yener, A. (2013b). Multiple access and two-way channels with energy harvesting and bi-directional energy cooperation. In *Information Theory and Applications Workshop (ITA)*.

[148] Tutuncuoglu, K. and Yener, A. (2014). The energy harvesting and energy cooperating two-way channel with finite-sized batteries. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, pages 1424–1429.

[149] Tutuncuoglu, K., Yener, A., and Ulukus, S. (2015). Optimum policies for an energy harvesting transmitter under energy storage losses. *IEEE J. Sel. Areas in Commun.*, 33(3):467–481.

[150] Ulukus, S., Yener, A., Erkip, E., Simeone, O., Zorzi, M., Grover, P., and Huang, K. (2015). Energy harvesting wireless communications: A review of recent advances. *IEEE J. Sel. Areas in Commun.*, 33(3):360–381.

[151] Đurišić, M. P., Tafa, Z., Dimić, G., and Milutinović, V. (2012). A survey of military applications of wireless sensor networks. In *Proc. IEEE Mediterranean Conf. on Embedded Computing (MECO)*, pages 196–199.

[152] van der Pijl, F., Bauer, P., and Castilla, M. (2013). Control method for wireless inductive energy transfer systems with relatively large air gap. *IEEE Trans. Industrial Electronics*, 60(1):382–390.

[153] Wang, H. S. and Moayeri, N. (1995). Finite-state Markov channel–a useful model for radio communication channels. *IEEE Trans. Vehicular Technology*, 44(1):163–171.

[154] Wang, P., Yu, G., and Zhang, Z. (2007). On the secrecy capacity of fading wireless channel with multiple eavesdroppers. In *Proc. IEEE Symp. on Information Theory (ISIT)*, pages 1301–1305.

[155] Wang, W. S., O'Donnell, T., Wang, N., Hayes, M., O'Flynn, B., and O'Mathuna, C. (2010). Design considerations of sub-mW indoor light energy harvesting for wireless sensor systems. *ACM J. Emerging Technologies in Computing Systems (JETC)*, 6(2):6.

[156] Wang, X., Gong, J., Hu, C., Zhou, S., and Niu, Z. (2015). Optimal power allocation on discrete energy harvesting model. *EURASIP J. Wireless Commun. and Networking*, 2015(1):1–14.

[157] Winkler, M., Tuchs, K.-D., Hughes, K., and Barclay, G. (2008). Theoretical and practical aspects of military wireless sensor networks. *J. Telecommunications and Information Technology*, (2):37–45.

[158] Wyner, A. (1975). The wire-tap channel. *Bell System Technical Journal*, 54(8):1355–1387.

[159] Xiao, L., Wang, P., Niyato, D., Kim, D., and Han, Z. (2015). Wireless networks with RF energy harvesting: A contemporary survey. *IEEE Commun. Surveys & Tutorials*, 17(2):757–789.

[160] Xie, L., Shi, Y., Hou, Y. T., Lou, W., Sherali, H., and Midkiff, S. F. (2015). Multi-node wireless energy charging in sensor networks. *IEEE/ACM Trans. Networking*, 23(2):437–450.

[161] Xu, J. and Zhang, R. (2014). Energy beamforming with one-bit feedback. *IEEE Trans. Signal Processing*, 62(20):5370–5381.

[162] Yang, G.-M., Ho, C.-C., Zhang, R., and Guan, Y. (2015). Throughput optimization for massive MIMO systems powered by wireless energy transfer. *IEEE J. Sel. Areas in Commun.*, 33(8):1640–1650.

[163] Yang, J. and Ulukus, S. (2012). Optimal packet scheduling in an energy harvesting communication system. *IEEE Trans. Communications*, 60(1):220–230.

[164] Yao, Y., Zhang, H., and Geng, Z. (2011). Wireless charger prototype based on strong coupled magnetic resonance. In *Conf. on Electronic and Mechanical Engineering and Inf. Technology (EMEIT)*, volume 5, pages 2252–2254.

[165] Yates, R. D. and Mahdavi-Doost, H. (2015). Energy harvesting receivers: Packet sampling and decoding policies. *IEEE J. Sel. Areas in Commun.*, 33(3):558–570.

[166] Zhang, M., Liu, Y., and Feng, S. (2014). Energy harvesting for secure OFDMA systems. In *Proc. 6th IEEE Conf. on Wireless Communications and Signal Processing (WCSP)*.

[167] Zhang, R. and Ho, C. K. (2013). MIMO broadcasting for simultaneous wireless information and power transfer. *IEEE Trans. Wireless Commun.*, 12(5):1989–2001.

[168] Zhao, M., Li, J., and Yang, Y. (2014). A framework of joint mobile energy replenishment and data gathering in wireless rechargeable sensor networks. *IEEE Trans. Mobile Computing*, 13(12):2689–2705.

[169] Zhou, G., Huang, L., Li, W., and Zhu, Z. (2014). Harvesting ambient environmental energy for wireless sensor networks: A survey. *J. Sensors*.

[170] Zhou, R. and Hansen, E. A. (2001). An improved grid-based approximation algorithm for POMDPs. In *Proc. 17th Int. Joint Conf. on Artificial Intelligence*, volume 1, pages 707–716.

[171] Zhou, X., Ho, C. K., and Zhang, R. (2016). Wireless power meets energy harvesting: a joint energy allocation approach in OFDM-based system. *IEEE Trans. Wireless Commun.*, 15(5):3481–3491.

[172] Zin, S. M., Anuar, N. B., Kiah, M. L. M., and Pathan, A.-S. K. (2014). Routing protocol design for secure WSN: review and open research issues. *J. Network and Computer Applications*, 41:517–530.

[173] Zois, D.-S., Levorato, M., and Mitra, U. (2013). Energy-efficient, heterogeneous sensor selection for physical activity detection in wireless body area networks. *IEEE Trans. Signal Processing*, 61(7):1581–1594.