MACQUARIE
University

Macquarie University
Department of Engineering

RELIABLE COMMUNICATIONS IN WIRELESS BODY AREA NETWORKS

By
Mohammad Sadegh MOHAMMADI

A thesis submitted in fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in
ELECTRICAL ENGINEERING

August 2017

STATEMENT OF CANDIDATE

As a cotutelle student, I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University and Aarhus University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis.

M. S. Mohammadi
August 2017

Abstract

This thesis investigates the problem of reliability in wireless body area networks (WBANs). WBANs constitute a class of wireless networks that are composed of miniaturized wearable or implantable nodes inside or in the vicinity of human body, with diverse applications in medicine, personal care, and entertainment, or more broadly Internet of Things (IoT).

Since WBANs can potentially convey sensitive health-related signals, high reliability of the wireless links is a necessity. Typically, reliability of communications can be increased by leveraging node resources such as energy or computational complexity. Nonetheless strict limitation of such resources in small battery-operated WBAN nodes to overcome severe channel conditions unique to WBANs pose a significant challenge that demand high transmission efficiency with low complexity. A key idea in this thesis is that achieving an optimal transmission efficiency can relax the constraint on node resources such that the node's expected lifetime is increased and it can allocate more energy and computational power to countering harsh noise and fading conditions.

The thesis studies different types of WBANs and propose novel techniques across different layers of the communication protocol to achieve high transmission efficiency. The proposed methods rely on statistical signal processing, adaptation, error control coding, and optimization. However, maintaining a low computational demand at the transmitter node is a key requirement that has been considered in all methods.

The main problem is addressed in two different parts of this thesis. The

first part deals with the problem of optimization of WBANs based on the IEEE 802.15.6 recommendations, which is the state-of-the-art communication protocol for WBANs. Novel MAC-level adaptation and optimization schemes are proposed for impulse-radio ultra-wideband systems and a theoretical framework to achieve the optimal energy-delay tradeoff for reliable communications is provided. Simulations confirm that the transmission efficiency can be improved by up to a factor of two by link adaptation. Also the energy efficiency is maximized with respect to the frame length and a closed formula is derived.

In the second part, a more general system is considered and the problem of optimal transmission efficiency in a typical WBAN/IoT device is addressed. The thesis proposes novel coding techniques based on random linear coding (RLC) as well as capacity-achieving low-complexity polar codes to outperform the state-of-the-art error control techniques. Specifically, novel hybrid-ARQ schemes based on systematic polar codes are proposed that are designed for low complexity and short code-length implementations that can achieve about 4 dB gain at low SNR. It is also proposed to leverage the available receiver-side computational power to fix partially corrupted packets without asking for any further redundancy from the transmitter which relies on transmitter-side RLC, combined with receiver-side sparse recovery.This technique can lead to performance gain in point-to-point RLC coded systems and improves transmission efficiency typically by 50% in multicast.

This thesis also proposes a novel joint sampling-quantization architecture with nonuniform sampling time and numbers of bits per sample that can represent a segment of a band-limited signal with the smallest number of bits, compared to the state-of-the-art previously known techniques, without performing any transform coding compression. In this way, it is shown that node resources are utilized efficiently and the resources required for quantization, compression, and transmission are saved since fewer bits are produced and transmitted assuming a given segment of signal acquired by the sensor. Simulations show that the proposed sub-Nyquist sampling scheme leads to only 12% of total number of bits compared to the conventional uniform Nyquist sampling.

Acknowledgment

I was greatly honored to pursue my PhD under a joint program between Aarhus University, Denmark, and Macquarie University, Australia. This PhD thesis will be endorsed by both universities separately and independently, based on the specific regulations in each university.

I am truly grateful to my supervisors Prof. Iain B. Collings and Prof. Qi Zhang, and my former supervisor Prof. Eryk Dutkiewicz. I would like to extend my special gratitude to my co-supervisors Prof. Rein Vesilo and Prof. Xiaojing Huang. I am very thankful to Prof. Rune H. Jacobsen and other researchers, administrative staff, and fellow PhD candidates at Aarhus University and Macquarie University for their friendship, support, and the great time we had together.

Parts of this thesis have been published in form of several conference and journal papers. It is acknowledged that the published papers are mainly based on the contribution of the first author and no help is received from the co-authors regarding the ideas, algorithms, implementation of the algorithms, or writing the papers, except as a few minor corrections or proofreading. Warm thanks to Keith Imrie for proofreading the first draft of this thesis and his help in fixing minor grammatical and structural errors.

I dedicate this thesis with love to my family. To my parents whose dedication, wisdom and love inspired me the most, and to my wife for her infinite love, motivation, and support.

Contents

Abbreviations

AC	Auto-Correlation
AMC	Adaptive Modulation and Coding
ARQ	Automatic Repeat Request
BFT	Binary Field Transform
BPS	Bits Per Sample
CM	Channel Model
CRC	Cyclic Redundancy Check
CS	Compressive Sensing
CSI	Channel State Information
ED	Energy Detection
FCS	Frame Check Sequence
FEC	Forward Error Correction
GNS	Generalized Nonuniform Sampling
IR	Impulse Radio
IoT	Internet of Things
JSC	Joint Source-Channel
LA	Link Adaptation
MAC	Medium Access Control
MICS	Medical Implant Communications Service
PC	Power Control
PER	Packet Error Rate
PHR	Physical-layer Header
PHY	Physical [layer]
PPDU	Physical-layer Protocol Data Unit
PPR	Partial Packet Recovery
PSD	Power Spectral Density
PSR	Packet Success Ratio
QLI	Queue Length Information
QoS	Quality of Service
RLC	Random Linear Coding
SC	Successive Cancellation
SFD	Start Frame Delimiter
SHR	Synchronization Header
UWB	Ultra Wideband
VLC	Variable Length Coding
WBAN	Wireless Body Area Network

Chapter 1

Introduction

The focus of this thesis is reliable communications in low-complexity, low-power, mostly wearable or implantable wireless devices in a wireless body area network (WBAN). A WBAN consists of a set of wireless implantable or wearable sensor nodes that communicate with each other or to a central *hub* node in the proximity of a human body. A hub node is able to handle more complex tasks and is less constrained in terms of energy resources. The sensor nodes are miniaturized electronic devices that are usually powered by a small battery and consist of a sensing device as well as communication and control circuits. They can collect various important physiological data for diagnosis or fast emergency response, or can deliver various personalized therapeutic-treatment-related applications and services, or can be a set of body-centric wireless devices that enable diverse applications in the realm of the Internet of Things (IoT).

1.1 Motivations and Challenges

There are several motivations to exploit WBANs, from health monitoring to sport and entertainment. The growth of the elderly population, and limitations in the numbers of medical staff and clinical facilities, demand an affordable and efficient health-care system. Considering the promotion of technology in the society and the recent advances in miniaturizing the elec-

Table 1.1: Characteristics of In-Body and On-Body Sensor Applications [1]

Application Type	Sensor Node	Data Rate	Duty Cycle	Power Consumption	Latency Sensitivity	Privacy
In-body Applications	Glucose Sensor	few Kbps	<1%	Extremely Low	Yes	High
	Pacemaker	few Kbps	<1%	Low	Yes	High
	Endoscope Capsule	>2Mbps	<50%	Low	Yes	Medium
On-body Medical Applications	ECG	3Kbps	<10%	Low	Yes	High
	SpO2	32 bps	<1%	Low	Yes	High
	Blood Pressure	<10bps	<1%	High	Yes	High
On-body Non-Medical Applications	Music for Headsets	1.4 Mbps	High	Relatively High	Yes	Low
	Forgotten Things Monitor	256 Kbps	Medium	Low	No	Low
	Social Networking	<200 Kbps	<1%	Low	Low	High

tronic devices, WBAN is a promising technology that can enable ubiquitous IT solutions for mobile health monitoring and treatment and similar applications with minimum cost. The medical applications of WBANs can be categorized into clinical and in-house deployments. Wireless nodes permit live monitoring with more flexibility, mobility, and durability than wired sensors. Furthermore, they can enable novel treatments when implanted in body. Table 1.1 represents some in-body and on-body sensor applications. A wide range of data rates from a dozen bits per second to several Mbps, a demand for different levels of QoS and privacy, and heterogeneity in duty cycle, can be observed.

Non-medical applications of WBANs vary from fitness monitoring, sport science and research, games, and emergency response for hazardous situations and professions. However, as will be explained in this thesis, the reliability of wireless links poses a challenging problem, and novel techniques and methods are required to be able to deploy WBANs for such applications. An extensive list of applications of WBANs and their advantages can be found in [2].

Besides all the motivations, there are several challenges that bring about a theoretical gap between the capabilities of the current wireless networks and those required by a reliable WBAN. It is now evident that the link quality in WBANs can dramatically change based on different factors such as body movements and occasional signal blockages [3], body shape and gender [4], posture [5], clothing and ornaments [6], existence of in-body medical implants with unknown reflective and dispersive properties [7], transmitter-receiver orientation [8], and finally different propagation characteristics of

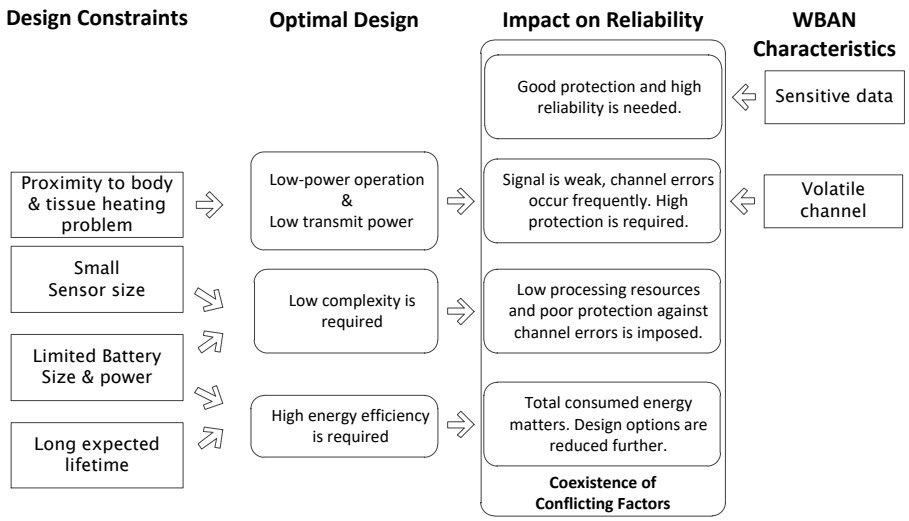


Figure 1.1: WBAN characteristics, design factors and their interactions. Reliability is a challenging design factor because a high level of reliability is needed with minimum possible computational and energy resources.

tissue among different individuals and different body parts [9]. The small form-factor of the sensor nodes as well as their limited energy resources (mostly from a battery) poses several design constraints in terms of energy efficiency and computational complexity. The reason why reliability is a challenging issue in the design of WBANs is imparted in Fig. 1.1. The nature of data in such systems is typically health-related and can be vitally important. Hence, a high level of reliability is needed. Nevertheless, the fact that the computational and energy resources of a typical implantable or sensor node are scarce impose a set of constraints that complicate the reliability challenge even more. Due to this complication, reliability in WBANs poses a research challenge to achieving an optimal design in its most strict sense.

1.2 Previous Work

Reliable communications in WBANs have been the subject of several papers as well as the recent IEEE 802.15.6 standard [10]. Indeed, [10] is a collection of state-of-the-art techniques ideal for low-power in/on-body sen-

sor networks, realized as a sophisticated MAC protocol that can be paired with three different PHY technologies. The default PHY in this standard is IR-UWB which is briefly described in Chapter 2. Nonetheless, as it will be elaborated with further detail in the next section, there are still several problems in this standard that should be appropriately addressed in order to achieve the optimal performance that is expected from WBANs.

Prior to this, reliability in WBANs has been addressed using different techniques such as cooperative diversity [11], wireless network coding [12], adaptive modulation and power control [11, 13], or optimal design of the MAC protocol [14]. Although the mentioned works are promising, they mostly extend the classical approaches for energy-efficient and reliable design of wireless systems to WBANs and are not exclusively based on the requirements and characteristics of WBANs. The problem of energy-efficient compression and forwarding of physiological data considering the heterogeneity of measurements is studied in [15], where compressive sensing is used to reduce data size, and then the importance of each data stream is considered for sensor selection and data transmission to obtain balanced energy consumption across different sensor nodes. The fact that implantable sensor nodes are under strict energy and complexity constraints motivates leveraging signal processing tasks on data-collecting (hub) nodes. This idea is considered in [16] and the transmission power is minimized subject to a target QoS and considering channel properties. A major problem with approaches that rely on cooperative communications is that they require a considerable amount of overhead and complexity for node synchronization which is not practically possible in typical WBANs. For this reason, a maximum of one hub is usually considered to enable cooperation diversity gain, as is also recommended in the default mode of [10].

The ideas of positioning, selecting and combining of the sensor nodes to improve reliability and extend the lifetime of network has been proposed in [17] and [11]. However, these studies are performed assuming a very specific application, bandwidth, and data rate and cannot be generalized easily. In addition, they only find a static state in which the observable performance metric is optimized and don't optimize the protocol in an effective way.

The classic definition of reliability suggests an idealized solution where either all information is delivered error-free or nothing can be interpreted. Alternatively, when the total received distortion becomes proportional to channel conditions, a gradual change of distortion can be achieved. In such cases, if the end-user can tolerate some level of distortion, then the channel can still be reliable, even in medium or harsh conditions. This idea is interesting for WBAN as well since it can overcome the so called "threshold effect". With this objective, some joint source and channel coding schemes have been proposed. However, very few papers have studied the problem of merging the source and channel codes to achive graceful degradation. The authors in [18] reveal the necessary and sufficient conditions of the basis functions over a finite field in order to construct an orthogonal wavelet transform that can be deployed by an error correcting code. Also in [19] they propose a code construction algorithm based on finite field wavelets for arbitrary-rate maximum separable block codes using spectral constraints.

The transmission protocol typically consists of different layers and an optimization framework can be applied to all these layers. For instance, it has been shown that packet size can be optimized to maximize throughput or energy efficiency [20]. Generally speaking, packet size optimization is a classic approach to achieve energy efficiency in wireless sensor networks (WSNs). This problem is investigated for underwater and underground WSNs in [21]. Packet size optimization is proposed to maximize energy efficiency in cognitive radio WSNs in [22]. A similar approach is pursued in [23], and also in [24] for WiFi-WBAN coexistence. Notwithstanding, this technique is not exclusively studied for WBANs. As also pointed out earlier, WBAN channels have significant and unique differences with radio channels in other systems that should be taken into account.

An important problem in optimization of the protocol is how to optimally relate different layers of the transmission scheme to each other. This is especially important to know the relation between delay at the MAC layer and the amount of energy consumed at the PHY layer. In the study of delay-sensitive communications over wireless networks, cross-layer approaches combine physical layer models with queue theory to optimize the

spectrum and energy resource allocation based on the tradeoff between radio resource utilization and the average delay [25]-[26]. Moreover, another goal is to adjust the scheduling policy to achieve the desired energy-delay tradeoff with an energy expenditure arbitrarily close to the optimal. Surprisingly, this important problem has not been addressed for WBANs previously.

An elegant technique that leverages processing for reliability is recovery of partially corrupted packets at the receiver side. It can be implemented using various techniques. [27] and [28] propose sharing physical layer soft information with higher layers for post-processing for recovery of lost data due to residual channel errors. In [29] a recovery strategy based on negative acknowledgment messages is presented that is compatible with the IEEE 802.11 standard and attempts to recover the errors by receiving further redundancy about the corrupted parts of the corrupted packets. Later on, some papers such as [30] propose methods to optimize this framework, for example, by adjusting the rate of the repair packets or by applying a more selective strategy in communicating the checksums [31]. [32] introduce an inter-frame acknowledgment, Micro-Ack, where they improved the efficiency of re-transmissions by adaptively changing the frame length. There are also approaches that recover partially corrupted packets based on some specific a priori knowledge on the content of the packets, or a form of the structure in the packets. For instance, known pilot bits are used in [33]. A similar scheme is suggested in [34] that exploits the algebraic consistency of the packets to recover the message blocks without any further retransmission. Another recent work [35] have suggested recovery of errors by estimating the errors. Error estimation is a process to acquire the approximate number of errors and their location by evaluating probabilistic models and a feedback of checksums from the receiver.

A promising approach to improve efficiency is adaptation. Using adaptive schemes, one can leverage processing power efficiently to deal with the difficulties inherent in WBANs. Link adaptation (LA) is an example which has been extensively studied. LA algorithms can be categorized based on the adaptation method and objective. Most of the existing work in this area is related to cellular [36], and WLANs [37, 38], where adaptation is

exploited to improve data throughput. In wireless sensor networks, however, adaptive techniques are usually proposed to reduce the total energy consumption and improve the energy efficiency [39, 40]. However, there has not been much work on LA in UWB systems that are regarded as the most important transmission systems for low-power low-cost applications. The concept of a PHY-aware medium access control (MAC) protocol in UWB systems is presented in [41]. An LA method based on the estimated frame error rate to improve data throughput under QoS constraints is proposed in [42] for multiband OFDM UWB systems. Additionally, in [43] LA is used to improve system throughput in OFDM-based UWB systems. In [44] and [45], SNR-based LA strategies are proposed to reduce the packet error probability in IEEE 802.15.4 networks. Also, in [46] interference mitigation in WBAN applications is performed by means of data rate adaptation. Nonetheless, LA in impulse radio UWB systems have not been studied.

Adaptive schemes have also been used in the sensor nodes, even for collecting the data. Sampling is one of the key subjects in this regard. The reason is that if a signal could be represented with fewer bits, then the transmission of that signal is a smaller burden for the transmission scheme (for example in terms of energy consumption). For band-limited signals, the Nyquist-Shannon sampling theorem addresses the minimum sampling rate as well as the unique reconstruction method when the samples are equally spaced. On the other hand, nonuniform sampling suggests that perfect reconstruction of the signal is still possible when the samples are not equally spaced, in exchange for further reconstruction complexity [47]. This problem can arise for several reasons. For example, to reduce the total number of bits required for reconstruction given a maximum acceptable distortion. This can lead to a more energy- and bandwidth-efficient transmission.

Although data size can be reduced through compression, it incurs extra cost due to further processing, complexity, and energy consumption. This justification has also led to the development of compressive sensing (CS) and related techniques. Some examples are recent papers such as [48] and [49] that apply CS to analog-to-digital converter design for sparse band-limited signals. Note that sparsity is a necessary condition to apply

CS methodologies. In addition, the CS reconstruction algorithms are nonlinear and usually include solving a convex optimization problem. These new concepts in CS theory have resulted in a reconsideration of quantization schemes that exhibit optimal performance when deployed in CS frameworks [50, 51]. The tradeoffs between the number of quantization levels (i.e. bit depth) and the number of samples required by the CS-based reconstruction schemes have been investigated in [52]. Historically, it is known that if the sampling rate satisfies the Nyquist rate on average, then nonuniform samples of a band-limited signal can uniquely represent the original signal [47]. The reconstruction algorithm usually is based on a form of interpolation. Indeed, despite the uniform sampling where the only prior (and necessary) information about the signal is its bandwidth, one can decrease the sampling rate to even below the Nyquist rate by assuming more prior knowledge about the signal. This knowledge can eventually narrow the reconstruction algorithm down to the original signal. Only a few papers such as [53] have studied this framework.

1.3 Solutions Proposed in This Thesis

The characteristics of WBANs necessitate *optimal design* of sensor nodes. It should satisfy the design requirements with minimum energy consumption and computational complexity. This principle lays out the basis for the ideas presented in this thesis. Albeit, to achieve this objective various strategies that suit different layers of the protocol stack are pursued, as depicted in Fig. 1.2. It is also tried to explore approaches that are very different from classic design principles. This is because most classic methods are based on presumptions, such as unconstrained complexity or energy resources, that don't hold in WBANs. As an example, separation of source and channel coding results in a simpler and more tractable design but can also impose a degraded performance. While such degradation may not be significant in general, WBANs demand strict optimality.

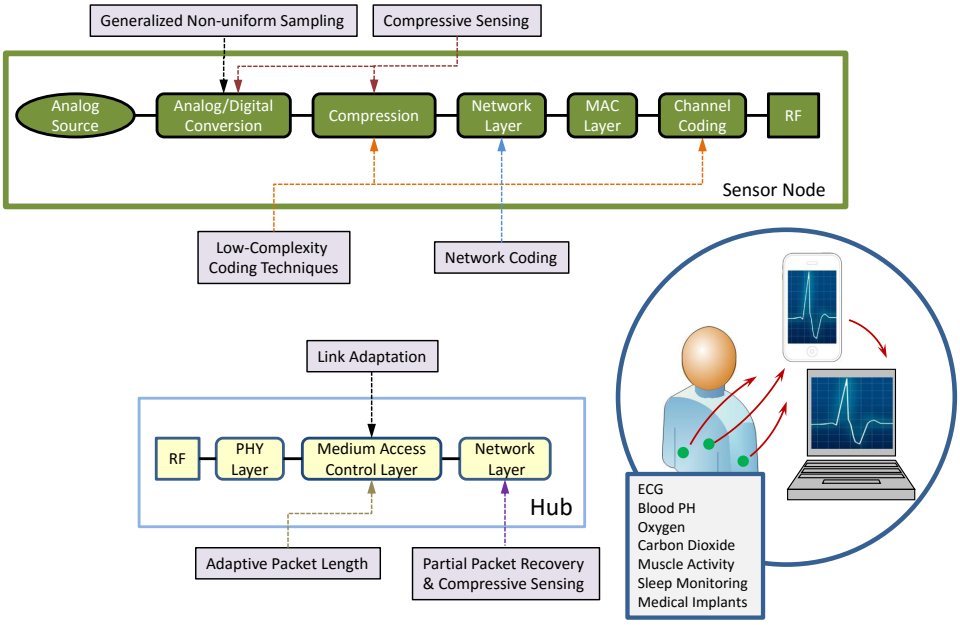


Figure 1.2: Different techniques in different layers of the protocol stack as proposed in this thesis.

1.3.1 Optimization of the Noncoherent UWB PHY

We can think of the optimal design as a distinct point in the energy-complexity-performance tradeoff. Hence, different techniques may be used to achieve a certain level of optimality. One approach is to adopt a simple transmitter structure and improve its performance through adaptation. Nevertheless, the modifications should be strictly minimal since complexity increases with adaptive capabilities. Ultra-wideband (UWB) transmitters are very promising for WBANs due to their simple structure, localization features and fading immunity. For this reason impulse-radio UWB is recommended in the default mode of the IEEE 802.15.6 standard. Addressing UWB systems, techniques such as link adaptation and packet length optimization with an optimal energy-efficiency constraint are proposed. Additionally, the theoretical amount of energy as well as the optimal scheduling policy to stabilize a transmission scheme based on UWB PHY are derive.

1.3.2 Partial Packet Recovery

One important observation in WBANs is that the amount of available resources in the sensor nodes and the hub is asymmetric. This is particularity the case with implantable sensor nodes, when the battery of sensor nodes should maintain the available power for a reasonably long time. The hub, however, can be much more computationally powerful with a less constrained energy resource. Thus, it is capable of more complex processing. This fact is used to design coding schemes and algorithms that are implemented at the hub and improve the link performance without imposing further complexity on the sensor node. Therefore, the processing power at the hub is leveraged to enhance reliability in the sensor-to-hub link. Partial packet recovery (PPR) is one such method that is investigated in this thesis. Compressive sensing and random linear coding are applied to devise a PPR scheme that repairs a partially corrupted packet without requiring any redundancy from the transmitter.

1.3.3 Nonuniform Sampling

Focusing on the sensor nodes alone, we can also improve the total energy efficiency by adopting new coding algorithms and signal processing techniques. Historically, the typical chain of operations in these nodes includes uniform sampling, compression, channel coding, and transmission. These operations are usually performed independently to simplify the design at both transmitter and receiver. However, it may not always lead to the optimal performance. Also it should be noted that the above chain is intrinsically inefficient since a huge number of samples are produced, then a significant part of that effort is thrown away by compression. Sampling, quantization, and compression are the most energy-hungry procedures at the sensor node. Hence, we can improve the energy efficiency by using an approach that takes fewer samples and omits compression. In this regard, nonuniform sampling schemes are investigated that result in fewer total bits to reconstruct a band-limited signal, compared to a uniform approach with Nyquist-rate sampling. This strategy follows the idea of data reduc-

tion presented in the previous section. However, except collecting a large amount of data and then reducing its volume (which also requires energy and consumes node resources), less data is produced in the first place. This will save resources in the following ways:

- The required resources for sampling and quantization are reduced since fewer samples and fewer bits are produced.
- The required resources for compression can be saved.
- By reducing the total number of bits, the required resources for data transmission are reduced.

1.3.4 Coding Techniques Based on Low-Complexity Capacity-Achieving Polar Codes

We can also improve transmission efficiency by providing a different interpretation of channel reliability. This thesis combines finite-field wavelets in Galois field of order two with polar codes to generate a new class of lossy/lossless joint source-channel codes. This approach can be used to construct joint source-channel codes of rate 1 (i.e. the code length is equal to the message length) that can provide an approximation of the message at low SNRs.

A very promising approach to enable high reliability and transmission efficiency at the same time is to rely on efficient coding techniques. This strategy is pursued in this thesis considering polar codes, due to their capacity achieving feature as well as their low-complexity nature. Hybrid automatic-repeat-request (ARQ) schemes based on polar codes that are strictly designed under the assumption of low complexity and short code length are discussed. The thesis proposes hybrid ARQ schemes based on systematic polar codes to achieve optimal transmission efficiencies that can approach the Shannon limit of binary additive white Gaussian noise channels.

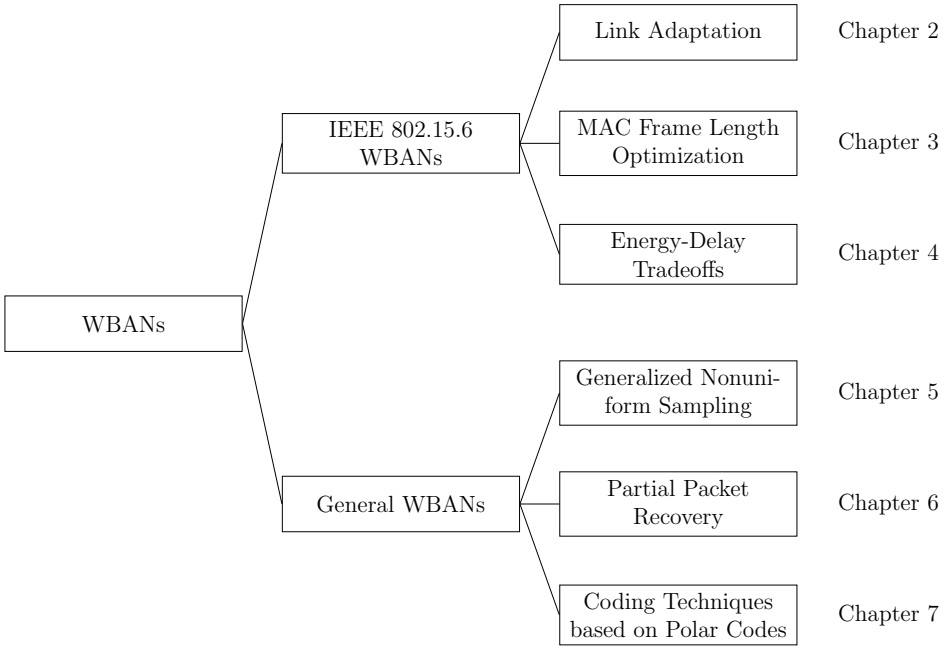


Figure 1.3: Dependency diagram of the subjects presented in this thesis, and the proposed techniques to analyze and improve transmission reliability and efficiency.

1.4 Organization of the Thesis

The thesis is organized as illustrated in Fig. 1.3. Chapters 2-4 deal with the problem of optimization of WBANs based on the IEEE 802.15.6 standard, while Chapters 5-7 pursue cross-layer techniques in general WBANs to improve transmission efficiency and reliability with low transmission complexity.

Chapter 2 briefly introduces the system model, the recommendations given by the standard, receiver structures, decoding algorithms, and the physical-layer characteristics of IEEE 802.15.6 WBANs. A novel link adaptation scheme is also proposed in the last section of this chapter. Chapter 3 presents frame-length optimization in IEEE 802.15.6 WBANs, as well as some more details on possible techniques to deploy the proposed scheme adaptively. The energy-delay tradeoffs in IEEE 802.15.6 IR-UWB systems are analyzed in Chapter 4. Using this framework, one can analytically find the minimum required amount of energy that is required to stabilize the

transmission queues associated with the sensors. In addition, optimal rate scheduling policies to achieve the tradeoffs are derived.

In Chapter 5, a new sampling architecture is proposed to represent a chunk of band-limited signal with the smallest number of bits, without using compression. The idea is to sample the signal nonuniformly in time and nonuniformly in the number of bits per sample, such that the bits per sample per time are allocated adaptively.

Chapter 6 proposes a PPR framework based on compressive sensing assuming random linear coded transmission. The PPR problem is formulated as a sparse recovery problem that has been solved using l_0 and l_1 norm minimization, as well as a Kronecker-based approach.

Finally, Chapter 7 provides different coding techniques based on polar codes. These techniques are designed with the objective of achieving smooth distortion, transmission efficiency and high reliability. A joint source-channel technique is presented in the first section that benefits from binary wavelets to reduce the total number of bits and at the same time increase the information transmission rate at the sensor nodes. In the next section, hybrid ARQ schemes based on systematic polar codes are designed under strict assumptions of low complexity and short code length.

1.5 Contributions of the Thesis

The main contributions of this thesis can be enumerated as follows:

1. A novel link-adaptation scheme for impulse-radio noncoherent UWB systems (ideal for low-complexity implementations). This scheme can adapt the number of pulses per symbol to the channel conditions to maximize energy efficiency without requiring channel knowledge. Furthermore, unbiased estimators are given in (2.28) and (2.29) for energy-detection and autocorrelation-based receivers, respectively, to estimate the *energy capture index* given in (2.6). There are two important diagrams in Chapter 2 (Fig. 2.6 and Fig. 2.7) that demonstrate the energy efficiency of different PHY modes of the IEEE 802.15.6

standard as well as the optimal adaptation maps, as a function of SNR and energy capture index.

2. An Optimization framework for maximizing the energy efficiency in UWB systems by adjusting the MAC frame length. An analytical expression is given in (3.12), and Fig. 3.2 presents the optimal lengths for default and High QoS modes in IEEE 802.15.6 systems.

3. Derivation of optimal energy-delay tradeoffs in IEEE 802.15.6 systems with noncoherent receivers, as well as the *minimum energy function* which is defined as the minimum average energy per symbol to stabilize a transmission system with a given data arrival rate. This has been done under channel-state and queue-length information scenarios. Fig. 4.1 presents the minimum energy function for the default mode of the IEEE 802.15.6 standard. Furthermore, (4.19) presents the optimal rate scheduling policy to achieve the optimal tradeoff while stabilizing the transmission queues.

4. A novel sampling architecture to sample band-limited signals at nonuniform times, and with nonuniform number of bits per sample. The generalized architecture is presented in Fig. 5.1 with one simple realization depicted in Fig. 5.3. Based on the results presented in Table 5.1, this scheme can represent a segment of band-limited ECG signal with the least possible number of bits compared to other state-of-the-art schemes *without performing any compression.*

5. A novel scheme based on compressive sensing and sparse recovery to boost the performance of cross-packet random linear coding (RLC) by incorporating the partial packets in the decoding algorithm, without using any form of pre-coding or any type of cross-layer soft information. The most important equations that mathematically describe this scheme are (6.12) to (6.16) as well as (6.20) and (6.21). In addition, the throughput performance of RLC systems without, and with, PPR is analytically expressed in (6.30) and (6.38) in conjunction with a Markov chain model which is given in (6.32). There are two im-

portant figures (Fig. 6.6 and Fig. 6.7) in Chapter 6 that illustrate the performance advantages of the proposed PPR scheme in terms of throughput, compared to an RLC system without PPR. Numerical results are also tabulated in Table 6.1 and Table 6.2.

6. Proposal of two novel Hybrid ARQ schemes based on systematic polar codes. In addition, the performance of polar decoding with incomplete codewords is analytically derived in (7.39).

In addition to the main contributions, the following minor contributions can be listed:

1. A new energy efficiency model assuming different energy costs for receiver and transmitter, presented in Appendix A. This model takes into consideration the packet overhead bits, as well as bits corresponding to synchronization and error control coding.

2. Analytical derivation of energy efficiency in noncoherent UWB systems with noncoherent (energy-detection and autocorrelation) receivers in IEEE 802.15.6 systems, presented in (3.7) and (3.10).

3. Improvement the existing analytical BER performance equations for noncoherent UWB systems by incorporating the impact of intra-symbol interference, described in (2.17) and (2.23).

4. Two different approaches to incorporate the impact of the time-varying channel gain in the transmission policy, aiming to improve the overall energy efficiency of the system. The flowchart of a simple MAC-layer technique is presented in Fig. 3.5.

5. Proposal of code inversion in polar codes as in (7.48).

6. Proposal of a novel joint source-channel code to achieve smooth distortion. Fig. 7.4 represents the performance of the proposed technique compared to separate source-channel coding, assuming a comparable overall code rate.

7. Proposal of FEC-assisted parallel decoding of short polar codes to reduce the overall decoding computational complexity of polar coded frames.

1.6 List of Published Papers

The contributions of this thesis are published in top-tier conferences and journals. A list of publications sorted by subject follows.

Optimization of IEEE 802.15.6.2012 WBANs

1. M. S. Mohammadi, E. Dutkiewicz, Q. Zhang, and X. Huang, "Optimal energy efficiency link adaptation in IEEE 802.15.6 IR-UWB body area networks," *IEEE Communications Letters*, vol. 18, no. 12, pp. 2193–2196, 2014.

2. M. S. Mohammadi, Q. Zhang, E. Dutkiewicz, and X. Huang, "Optimal frame length to maximize energy efficiency in IEEE 802.15.6 UWB body area networks," *IEEE Wireless Communications Letters*, vol. 3, no. 4, pp. 397–400, 2014.

3. M. S. Mohammadi, Q. Zhang, E. Dutkiewicz, X. Huang, and R. Vesilo, "Energy-delay tradeoffs in impulse-based ultra-wideband body area networks with noncoherent receivers," in *IEEE Global Communications Conference (GLOBECOM)*, 2014, pp. 4014–4019. 2014.

4. M. S. Mohammadi, Q. Zhang, and E. Dutkiewicz, "Channel-adaptive MAC frame length in wireless body area networks," in *International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 584–588, 2014.

Improving Transmission Efficiency using PPR

5. M. S. Mohammadi, Q. Zhang, and E. Dutkiewicz, "Reading damaged scripts: Partial packet recovery based on compressive sensing

for efficient random linear coded transmission," *IEEE Transactions on Communications*, vol. 64, no. 8, pp. 3296–3310, Aug. 2016.

6. M. S. Mohammadi, Q. Zhang, and E. Dutkiewicz, "Exploiting partial packets in random linear codes using sparse error recovery," in *IEEE International Conference on Communications (ICC)*, pp. 2577–2582, 2015.

Improving Transmission Efficiency using PHY Coding Schemes

7. M. S. Mohammadi, I. B. Collings, and Q. Zhang, "Simple hybrid ARQ schemes based on systematic polar codes for IoT applications," *IEEE Communications Letters*, vol. 21, no. 5, pp. 975–978, May. 2017.

8. M. S. Mohammadi, Q. Zhang, and E. Dutkiewicz, "Joint binary field transform and polar coding," in *IEEE International Conference on Communications (ICC)*, pp. 4406–4411, 2015.

9. M. S. Mohammadi, E. Dutkiewicz, and Q. Zhang, "Sampling of band-limited signals with nonuniform sampling-time and bit-depth," in *IEEE Global Communications Conference (GLOBECOM)*, 2015.

10. M. S. Mohammadi, E. Dutkiewicz, and Q. Zhang, "Joint source-channel optimization of vector quantization with polar codes," in *IEEE Vehicular Technology Conference (VTC)*, 2016.

Chapter 2

Link Adaptation in IEEE 802.15.6 WBANs

2.1 Introduction

Most of the recent studies of UWB systems have been conducted to comply with the FCC spectral mask for indoor environments that allows a maximum transmit power of -41.3 dBm, which is less than 100 nW per MHz. Low-power operation is usually highly recommended for medical and implantable devices. Therefore, UWB can be a standard communication medium in the proximity of the human body. Taking into account its other benefits, including support for high data rates due to the vast amount of bandwidth, low complexity and strong ranging capability due to a fine time resolution, UWB has been chosen as a mandatory physical layer technology for on-body applications in wireless body area networks in the recent IEEE 802.15.6 standard [10].

The mandatory physical layer (PHY) scheme in this standard is impulse radio (IR) UWB, which consists of two PHY options: the default PHY and the high quality of service (QoS) PHY. The former is a general-purpose PHY consisting of mandatory on-off signaling, and the latter is for high-priority medical applications and consists of a mandatory differential phase shift keying (DPSK) modulation and a type II hybrid automatic-repeat-

request (H-ARQ) mechanism. Both PHY options allow for a set of different modulation and coding techniques to obtain various data rates. However, the question of how to coordinate the available data rates in an optimal manner, referred to as *rate adaptation* or *link adaptation* (LA), remains unanswered.

In traditional wireless systems (e.g. cellular systems and wireless local area networks), LA techniques can improve data throughput by adapting the physical layer parameters in response to noise, fading, and interference. Normally, a form of channel estimation is performed at the wireless device and is transmitted back to the coordinating node (i.e. base station/access point) using a feedback channel. The coordinating node can then apply a form of adaptation on the physical layer parameters (e.g. constellation size, coding rate, etc.) in order to improve system capacity and reliability. In WBANs, however, the main concerns include maximizing energy efficiency and the overall lifetime of the system, as well as improving the reliability of communications. Due to complexity and energy constraints, channel estimation is infeasible at the transmitting node, and hence traditional methods in automatic modulation and coding (AMC) are not directly applicable. The immensely dynamic and severe channel conditions resulting from body movements, occasional signal blockages, and the different dispersive characteristics of body tissues, along with the strict design limitations unique to WBANs, necessitate a significant level of optimality and flexibility in order to satisfy the QoS requirements in a typical WBAN application. It is believed that LA strategies are of significant importance to attain such levels of performance [40, 54].

In this chapter, we propose a link adaptation strategy for IEEE 802.15.6 IR-UWB body area networks that does not require channel estimation at the sensor nodes. Instead, the link quality is estimated using a novel estimator at the hub node which is typically more potent in in WBANs in terms of energy resources and hardware. Assuming noncoherent reception, we model the bit error probability of different PHY modes of the standard, considering the imperfect energy capture caused by the multipath channel and the intra-symbol interference caused by the multipath channel, and use

Gaussian approximations to approximate the corresponding bit error probabilities. We discuss the confidence level of the Gaussian approximation and address the impact of the integration time on the performance of the system. As is shown in this chapter, a significant energy loss will result from the short integration times at higher data rates. We pursue a semi-analytical approach to model this performance degradation and present the so-called energy capture index in two IEEE 802.15.6 UWB body area channel models. In order to obtain an interference model, and targeting wireless local area network (WLAN) devices as the primary interference source, we assume an orthogonal frequency division multiplexing (OFDM) PHY for the interference signal and show that under the assumptions of IEEE 802.15.6 the signal-to-noise expressions in our model can be replaced by a signal-to-noise-plus-interference (SINR) term. The energy efficiency of each PHY mode is calculated based on the corresponding packet success rate (PSR) and energy consumptions of the PHY mode. We also show that the throughput optimal adaptation method is not necessarily identical to the energy efficiency optimal scheme. In order to approximate the amount of captured energy, two unbiased estimators are proposed.

Fixing the transmit pulse energy, we also propose a novel LA scheme that can adapt the number of pulses per symbol to the channel conditions in order to maximize energy efficiency. By incorporating the impact of the instantaneous energy capture index in the adaptation process, the system can respond to dynamic variations in the wireless channel, noise, and interference to preserve a desired level of performance. The key fact is that, to collect a certain amount of energy from the transmitted pulses, the nodes should be able to adapt the number of pulses per symbol to the variations in the channel conditions (e.g. delay spread, decay factor, etc.), and the level of interference and noise.

To the best of our knowledge none of the existing works address LA in IR-UWB systems. Also, there has not been a similar adaptation strategy based on observing the amount of captured energy in UWB systems. The benefits of such a strategy are: *sensitivity to variations in channel conditions*, *sensitivity to interference and noise*, and *sensitivity to signal strength*,

Table 2.1: UWB PHY modes of IEEE 802.15.6 [10]

	Default					High QoS				
PHY mode	T_s (ns)	T_w (ns)	Rate (Mbps)	Mod.	N_p	T_s (ns)	T_w (ns)	Rate (Mbps)	Mod.	N_p
PHY 0	2051.300	64.103	0.3948	OOK	32	2051.300	64.103	0.243	DBPSK	32
PHY 1	1025.600	32.051	0.7897	OOK	16	1025.600	32.051	0.457	DBPSK	16
PHY 2	512.820	16.026	1.579	OOK	8	512.820	16.026	0.975	DBPSK	8
PHY 3	256.410	8.012	3.159	OOK	4	256.410	8.012	1.950	DBPSK	4
PHY 4	128.210	4.006	6.318	OOK	2	128.210	4.006	3.900	DBPSK	2
PHY 5	64.103	2.003	12.636	OOK	1	128.210	4.006	7.800	DQPSK	2
PHY 6	-	-	-	-	-	1794.900	8.012	0.278	DBPSK	4
PHY 7	-	-	-	-	-	1794.900	8.012	0.557	DQPSK	4

which are all crucial in WBAN applications.

This chapter is organized as follows. Section 2.2 describes the system and channel model, and provides a brief introduction to the IR-UWB physical layer in the IEEE 802.15.6 standard, and our interference model for performance evaluation. Section 2.3 presents a discussion on optimal and suboptimal UWB receivers and the impact of the integration time on the performance of noncoherent receivers. It also provides a semi-analytic model of the amount of collected energy in UWB channel models that we use for performance evaluation. In Section 2.4, the bit error probabilities of various PHY modes are derived using a Gaussian approximation, and the impact of narrowband interference is incorporated in the analysis by modeling the interference signal. Section 2.5 presents the proposed LA strategy and energy estimators. In Section 2.6 the simulation results are provided, and finally the conclusions are given in Section 2.7.

2.2 System Model

Assume that a set of N_s *sensor* nodes communicates with a master *hub* node. The hub uses a *downlink* channel for sending its command and control signals, and the sensors communicate their data back on the *uplink* channel. The sensors can transmit two types of data: measurements, and system parameters. The former varies depending on the application and sensor type. The latter is specific information about the state of the sensor device (e.g. battery, location, etc.), its measurements (e.g. data rate, queue length, etc.), and channel. The hub can utilize the feedback from the sensors to adjust a set of communication parameters in order to achieve a desired level

of performance and send it to the sensors on the downlink channel. We assume that both uplink and downlink communications use the mandatory IR-UWB physical layer of the IEEE 802.15.6 standard.

2.2.1 Channel Model

The characteristics of UWB channels are studied in [55, 56, 57] for indoor environments. Correlated log-normal shadowing is the most precise model that has been proposed for representing the small-scale fading in UWB WBANs [58, 59, 60, 5, 7]. In this model, the amplitudes of multipath components are strongly correlated and follow the log-normal distribution. It is believed that correlation mainly results from the frequency-selective properties of body tissues and UWB antennas. The multipath components tend to arrive in separate clusters. This property usually results from ground reflections or the effects of metal pieces in clothing or implants inside the body. The UWB waves tend to disperse around the body rather than pass through the tissue. There is a significant degradation in signal quality when the antenna is attached to the body. Therefore, in most cases the antenna and the body are separated by using a small dielectric spacer. We consider the channel model for 3.1-10.6 GHz in a typical hospital environment [61] as

$$h(t) = \sum_{l=0}^{N_{mp}-1} a_l \delta(t - \tau_l), \tag{2.1}$$

in which N_{mp} is the number of resolvable multipath components and a_l is a complex attenuation coefficient corresponding to the l'th multipath delay τ_l. We consider two channel models CM3 and CM4 [61]. CM3 can be used to model *body-surface to body-surface* channels, e.g. a sensor-sensor or sensor-hub channel when the hub is in the vicinity of the sensors and is less than two cm away from the body. CM4 is actually a *body-surface to external-node* channel model and can be used for distances between a few centimeters to up to five meters.

2.2.2 IEEE 802.15.6 UWB PHY

The IEEE 802.15.6 standard is composed of two different physical layer options. The default PHY exploits on-off signaling which is a combination of M-ary waveform coding with on-off modulation. Such a signaling strategy maps K information bits onto coded-pulse sequences of length $2K$ using a 1/2 rate symbol mapper ($K = 1$ and $M = 2$ in the mandatory symbol mapper). The high QoS PHY benefits from DPSK modulation as well as a type II H-ARQ mechanism. The characteristics of both PHY options are listed in Table 2.1. We refer to each modulation and coding scheme (i.e. data rate) as a *physical layer mode* for simplicity. The PHY modes are represented by an index m, $m \in [0, N_m]$ where N_m is the number of available modes, and are referred to by "PHY m". The only modulation scheme in the default PHY is on-off keying, however, the high QoS PHY can use either DBPSK or DQPSK. As expected, the number of pulses per symbol and the symbol duration are decreased to achieve higher data rates. Since complexity is a serious limitation in most wireless body area applications, we assume that the receiver can only select between PHY modes with the same modulation. Hence, $N_m = 5$ in the high QoS PHY corresponding to PHY 0-PHY 4, and $N_m = 6$ in the default PHY corresponding to PHY 0-PHY 5.

The transmitted IR-UWB signals corresponding to the default and high QoS PHY can be expressed as

$$x_D(t) = \sum_i [d_i w(t - c_i T_w - 2i\frac{T_s}{2}) + (1 - d_i) w(t - c_i T_w - (2i+1)\frac{T_s}{2})], d_i \in \{0, 1\}, \quad (2.2)$$

$$x_H(t) = \sum_i d_i w(t - c_i T_w - iT_s), d_i \in \{+1, -1\}, \quad (2.3)$$

respectively, where d_i is the i'th data symbol transmitted within the symbol time T_s, $w(t) = \sum_{k=0}^{N_p-1} p(t - kT_p)$ is a waveform of duration $T_w = N_p T_p$, and c_i is the time hopping value in the range $0 < c_i < N_h - 1$, in which N_h is the total number of hops. Regardless of the PHY option, there are 32 timeslots in each symbol duration. We assume N_p pulses of form $p(t)$,

duration T_p, and energy E_p are transmitted per symbol so that $w(t) = \sum_{k=0}^{N_p-1} p(t - kT_p)$, $T_w = N_pT_p$, and symbol energy $E_s = E_w = N_pE_p$, which is consistent with the *burst pulse option* in [10]. A special case is when only one pulse is transmitted per symbol such that $w(t) = p(t)$, $T_w = T_p$, and $E_s = E_p$ which is referred to as the *single pulse option*. The data symbols $\{d_i\}$, depending on the modulation scheme, take different values. Explicitly, $d_i \in \{0, 1\}$ in on-off keying, $d_i \in \{-1, +1\}$ in BPSK, and $d_i \in \{e^{j\pi/2}, e^{j\pi}, e^{-j\pi/2}, e^{j0}\}$ in QPSK modulation. Note that QPSK symbols are constructed by a combination of two in-phase and quadrature BPSK modulated symbols given by $\mathrm{Re}\{d_i\}$ and $\mathrm{Im}\{d_i\}$, respectively.

2.2.3 Narrowband Interference Model

The performance of IR-UWB systems in multipath fading environments and in narrowband interference is studied in [62, 63, 64] and [65], respectively. Narrowband interference (NBI) is a big challenge for UWB systems. The major sources of interference for UWB devices operating in indoor environments are the very popular 802.11a WLAN devices. They operate in the 5 GHz frequency band and use an OFDM air interface to achieve data rates of up to 54 Mbps. The multi-carrier interference signal can be modeled by

$$I_{NB}(t) = \sum_{i=0}^{N_{sc}-1} y_i \cos(2\pi f_{NB} t + \phi_i), \tag{2.4}$$

where N_{sc}, $\{y_i\}$, and $\{\phi_i\}$ represent the total number of sub-carriers, and the corresponding amplitudes and phases respectively, and f_{NB} is the center frequency. It is well known that OFDM signals can be approximated by a complex Gaussian random process due to the high number of independently modulated sub-carriers and the central limit theorem [66]. In fact, the Gaussian approximation is valid even for a reduced number of sub-carriers, as is shown in [67]. Due to the small bandwidth of the narrowband interference process in comparison with the UWB bandwidth, its power spectral density (PSD) represents colored characteristics. However, it can be modeled by a PSD which is constant over the frequency band W_{NB} and zero elsewhere. Hence, the autocorrelation function of the narrowband

interference process, which is the inverse Fourier transform of the above PSD, is given by

$$R_{NB}(\tau) = \frac{J_0}{2}\text{sinc}(W_{NB}\tau)\cos(2\pi f_{NB}\tau), \tag{2.5}$$

which is approximately equal to zero for $\tau > 1/W_{NB}$. This implies that the samples of the narrowband interference process separated by a time interval larger than $1/W_{NB}$ are uncorrelated, and form a white Gaussian random process with variance $\frac{J_0}{2}$.

2.3 UWB Receiver Structures

2.3.1 Rake Reception

Rake receivers can capture a significant portion of the received signal by resolving multipath components. When UWB signals pass though the body, the received signal strength is very low since the attenuation caused by the body is relatively high. But since there is usually a ground reflected wave that arrives shortly after the main pulse, a rake receiver could be able to gather this energy and detect the signal more reliably. In order to capture all of the multipath energy, a rake receiver should accurately estimate the number of resolvable multipath components N_{mp}, the corresponding channel coefficients $\{a_l\}$, and delays $\{\tau_l\}$ ($l \in [0, N_{mp} - 1]$). In general, an exponential increase of complexity due to multipath acquisition, tracking and channel estimation is imposed on the receiver by increasing the number of rake fingers. Consequently, selective or partial rake receivers, with a limited number of fingers, are typically used instead of a full rake receiver. Indeed, due to significant complexity concerns, even for partial rake receivers, non-coherent receivers are more suitable in a wide range of WBAN applications.

2.3.2 Suboptimal Receivers

Due to a high demand for low complexity and low power consumption in WBAN applications, suboptimal noncoherent receivers are of more inter-

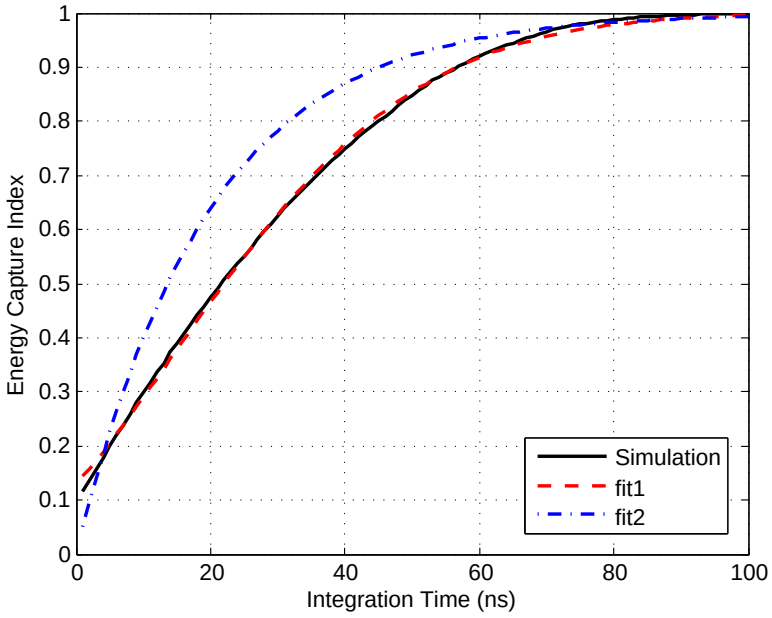


Figure 2.1: Simulated mean energy capture index of CM3 and its exponential decay fitting curves.

Table 2.2: $\mu(T_i)$ Fitting parameters

	Position	$E[\tau_{rms}]$	ω	ν	α
CM3	-	19.6	15.56	46.22	1.843
CM4	0°	34.8	29.13	50.31	1.115
	90°	41.6	23.12	60.60	1.176
	180°	42.0	16.61	57.83	1.160
	270°	49.6	29.53	90.62	1.426

est. In noncoherent and partially coherent receivers the perfect knowledge of channel state information (CSI) is not necessary in detection of data symbols, and this can make the receivers significantly simpler than rake receivers. This is however achieved by a compromise on the receiver's performance. There are two main groups of noncoherent UWB schemes: energy detection (ED), and autocorrelation (AC). In energy detection systems the received energy during each timeslot is measured to detect the existence of data bits, while the AC receivers use a delayed version of a reference pulse from the transmitter as the local template for signal correlation and detection.

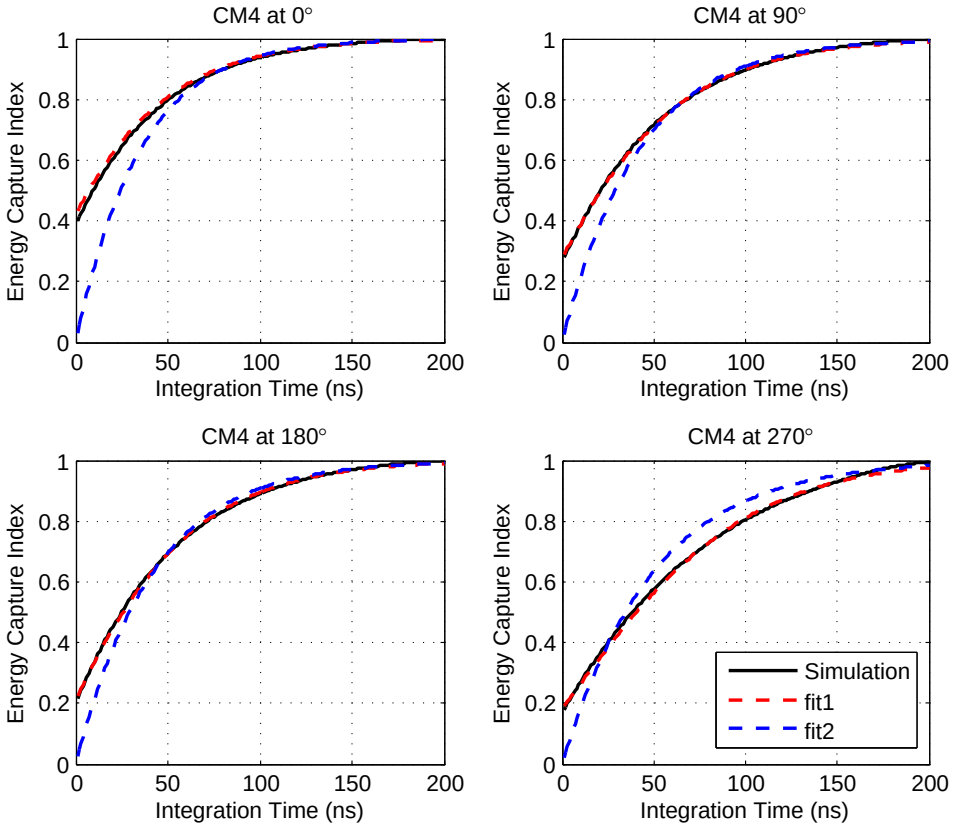


Figure 2.2: Simulated mean energy capture index in CM4 for four different positions and its exponential decay fitting curves.

In a noncoherent receiver (AC or ED), given $g(t)$ the received pulse waveform ($g(t) = p(t) * h(t)$, where $*$ denotes linear convolution), the instantaneous energy of the received pulse is $\int_0^{T_i} g(t)^2 \mathrm{d}t = \mu(T_i|h)E_p$, and the decision variable is formed by combining the sampled energy values received in a symbol duration. The amount of collected energy depends on the integration time T_i as well as the experienced channel $h(\tau)$, and can be modeled by the *energy capture index* [63] defined as

$$\mu(T_i|h) = \int_0^{T_i} |h(\tau)|^2 \mathrm{d}\tau, \tag{2.6}$$

where $\mu(T_i|h) \in [0, 1]$. The mean value of $\mu(T_i|h)$ averaged over 1000 channel realizations of two body area channel models, CM3 and CM4, is depicted in Fig. 2.1 and Fig. 2.2. To obtain an analytical model for $\mu(T_i|h)$, it is also

approximated by the following exponential decay models [63, 68]

$$\mu_1(T_i) = 1 - \exp\left[-\left(\frac{(T_i+\omega)}{\nu}\right)^{\alpha}\right], \tag{2.7}$$

$$\mu_2(T_i) = 1 - \exp\left(-T_i/\tau_{rms}\right), \tag{2.8}$$

where τ_{rms} is the channel's root-mean-squared delay spread, which is the second central moment of the channel's power delay profile.

The fitting parameters are shown in Table 2.2. It can be observed that, while $\mu_1(T_i)$ fits well to the energy capture index, the approximation provided by $\mu_2(T_i)$ is not accurate, specifically for lower values of T_i. We will use μ_1 in our link adaptation strategy.

At the receiver input, first a bandpass filter with bandwidth B removes the unwanted spectra. The filtered signal can be expressed as

$$r(t) = \sum_{i} \sum_{l=0}^{N_{mp}-1} a_l d_i \widetilde{w}(t - c_i T_w - iT_s - \tau_l) + \widetilde{n}(t), \tag{2.9}$$

where $\widetilde{n}(t)$ is zero-mean band-limited Gaussian noise with two-sided power spectral density $\frac{N_0}{2}$ over the frequency band B and zero elsewhere, and $\widetilde{w}(t)$ and $\widetilde{g}(t)$ are the filtered versions of $w(t)$ and $g(t)$ respectively.

2.4 Physical Layer Characteristics

The error performance of the noncoherent UWB systems has already been studied in [63] and [62]. However, since the standard uses a special terminology (e.g. on-off signaling that combines on-off keying with waveform coding), we briefly include the error analysis in this section to prevent confusion. We will also require the equations to estimate the energy capture index based on the decision statistic. Nevertheless, we complete the previous works on the error performance of noncoherent IR-UWB systems by incorporating the impact of intra-symbol interference.

The key parameter in our analysis is the integration window in the receiver front end. The IEEE 802.15.6 standard does not provide specific options for pulse duration and the integration window. From Table 2.1,

since the symbol time linearly increases with N_p, it implies that a constant pulse duration is assumed. Considering $N_w = 32$ timeslots (fixed by the standard), it implies that $T_p = 2.003$ ns. We assume that the integration window covers a complete timeslot such that $T_i = T_w = N_p T_p$. Therefore, intra-symbol interference can occur in a timeslot when the channel's RMS delay spread is large enough compared to T_p. This is actually a constructive property. The overall captured energy per symbol can be increased since the multipath components of the earlier pulses can be captured during the integration time of the proceeding pulses.

2.4.1 Energy Detection with On-Off Signaling

In energy detection, the energy of the received signal is detected by passing through a square-law device. Combined with on-off signaling, the decision statistic is the difference of energy in the first and the second half of the symbol duration. Generally, the symbol decision statistic is formed by a linear combination of N_p independent pulses per symbol. The i'th decision variable corresponding to th i'th symbol can be expressed as

$$Z_D[i] = \int_{iT_s+c_iT_w}^{iT_s+(c_i+1)T_w} \mid r(t) \mid^2 \mathrm{d}t - \int_{(i+1/2)T_s+c_iT_w}^{(i+1/2)T_s+(c_i+1)T_w} \mid r(t) \mid^2 \mathrm{d}t, \tag{2.10}$$

Assuming that $d_i = 1$ the decision variable can be written as

$$Z_D[i] \mid_{d_i=1} = \int_0^{T_w} \left[\sum_{k=0}^{N_p-1} \widetilde{g}(t - kT_p) + \widetilde{n}_1(t) \right]^2 \mathrm{d}t - \int_0^{T_w} \widetilde{n}_1^2(t) \mathrm{d}t, \tag{2.11}$$

Considering the Landau-Pollak theorem [69] which states that there are $2BT$ independent signals of bandwidth B spread in a time interval of length T, we can re-write the decision statistic as

$$Z_D[i] \mid_{d_i=1} = \sum_{k=0}^{N_p-1} \left(\sum_{j=0}^{2B(T_w-kT_p)-1} g_{j,k} + \sum_{j=0}^{2BT_p-1} n_{1,j,k} \right)^2 - \sum_{k=0}^{N_p-1} \left(\sum_{j=0}^{2BT_p-1} (n_{1,j,k})^2 \right). \tag{2.12}$$

which is the difference of a non-central and a central chi-squared random variable with $2N_pBT_p$ degrees of freedom, where $g_{j,k}$, and $n_{j,k}$ are independent samples of the band limited signals $\widetilde{g}(t)$ and $\widetilde{n}(t)$. The conditional mean and variance of Z_D, the decision variable in the default PHY, are

$$m_{Z_D|1} = E_p \sum_{k=0}^{N_p-1} \mu(T_w - kT_p \mid h), \tag{2.13}$$

$$\sigma^2_{Z_D|1} = 2E_pN_0 \sum_{k=0}^{N_p-1} \mu(T_w - kT_p \mid h) + 2N_pBT_pN_0^2. \tag{2.14}$$

When the total number of samples ($2N_pBT_w$) is large, according to the Central Limit Theorem, the sum PDF tends to the Gaussian distribution. Based on the justification in [62] and the K-S statistic represented in Fig. 2.3, this approximation is valid for $N_pBT_p > 20$. This implies that when two 500 MHz channels are used (as is mandatory in the 802.15.6 standard), the approximation is valid for $T_p > 2$ ns. In such a case, the false and miss detection probabilities are given by [70]

$$P_{fd} \approx Q\left(\frac{\xi + E_p\bar{\mu}_t}{\sqrt{2E_pN_0\bar{\mu}_t + 2N_pBT_pN_0^2}}\right), \tag{2.15}$$

$$P_{md} \approx Q\left(\frac{E_p\bar{\mu}_t\xi}{\sqrt{2E_pN_0\bar{\mu}_t + 2N_pBT_pN_0^2}}\right), \tag{2.16}$$

where ξ is the decision threshold which is set to zero due to symmetry, $Q(\cdot)$ is the tail probability of the standard normal distribution and $\bar{\mu}_t = \mathbb{E}\{\mu_t\}$ in which $\mu_t = \sum_{k=0}^{N_p-1} \mu(T_w - kT_p \mid h)/N_p$. The bit error probability can be approximated by

$$P_b \approx Q\left(\frac{\bar{\mu}_tN_pE_p/N_0}{\sqrt{2\bar{\mu}_tN_pE_p/N_0 + 2N_pBT_p}}\right). \tag{2.17}$$

2.4.2 Autocorrelation with Differential Signaling

In autocorrelation receivers, the received signal is multiplied by a delayed reference waveform. In differential signaling, the reference is the received

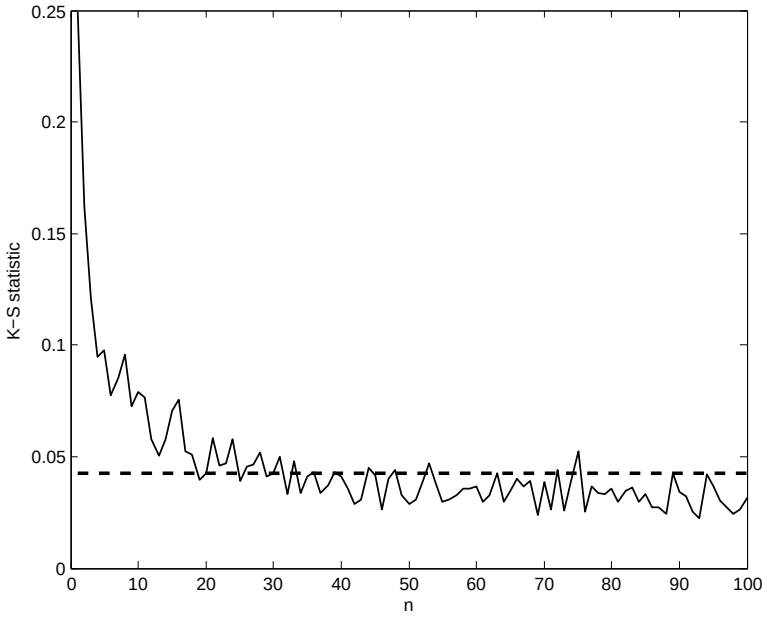


Figure 2.3: Kolmogorov-Smirnov Goodness-of-fit test for Gaussian approximation of Chi-squared variables with n degrees of freedom. The approximation is rejected at the 5% significance level, when the test result is above the critical value (for $n < 40$), and is valid for $n > 40$.

waveform corresponding to the last decoded symbol. Similar to (2.10), the i'th decision variable corresponding to the i'th symbol is expressed as

$$Z_H[i] = \int_{iT_s+c_iT_w}^{iT_s+(c_i+1)T_w} r(t)r(t-T_s)\mathrm{d}t, \tag{2.18}$$

There are two elements of the desired signal: an information-bearing waveform and a reference signal. We assume that the coherence time of the channel is greater than T_s and that these two signals experience the same channel conditions.

$$Z_H[i] = \int_0^{T_w} \Bigg[\left(\sum_{k=0}^{N_p-1} \mathrm{d}_1 \widetilde{g}(t-kT_p) + \widetilde{n}_1(t) \right) \cdot \left(\sum_{k=0}^{N_p-1} d_2 \widetilde{g}(t-kT_p) + \widetilde{n}_2(t) \right) \Bigg] dt, \tag{2.19}$$

The symbol decision statistic, which is the sum of N_p samples, can be

expressed in terms of the independent components $g_{j,k}$, $n_{1,j,k}$, and $n_{2,j,k}$ as

$$Z_H[i] \mid_{d_1,d_2=+1} = \sum_{k=0}^{N_p-1} \left[\sum_{j=0}^{2B(T_w-kT_p)-1} g_{j,k}^2 + \sum_{j=0}^{2BT_p-1} g_{j,k}n_{1,j,k} + g_{j,k}n_{2,j,k} + n_{1,j,k}n_{2,j,k} \right]. \tag{2.20}$$

There are four equiprobable combinations of d_1 and d_2. We assume $d_1 = +1$, and calculate the conditional error probability for $d_2 = +1$, which is equal to the bit error probability due to symmetry. The conditional mean and variance of the decision variable Z_H corresponding to the high QoS PHY are

$$m_{Z_H|+1} = \mu_t N_p E_p + N_p B T_p N_0, \tag{2.21}$$

$$\sigma^2_{Z_H|+1} = \mu_t N_p E_p N_0 + N_p B T_p N_0^2/2. \tag{2.22}$$

The bit error probability of the DBPSK system with the Gaussian approximation is expressed as

$$P_b \approx Q\left(\frac{\bar{\mu}_t N_p E_p/N_0}{\sqrt{\bar{\mu}_t N_p E_p/N_0 + N_p B T_p/2}} \right). \tag{2.23}$$

Note that the decision threshold is $N_p B T_i N_0$ in this case.

2.4.3 Performance in Presence of Narrowband Interference

As it is shown in Section 2.2.3, we can model the narrowband interference, i.e. WLAN OFDM signal, by a Gaussian process which is approximately white in a frequency band W_{NB} and zero elsewhere. In 802.11a WLAN, the total bandwidth W_{NB} is equal to 20 MHz. Taking into account the minimum symbol duration among the PHY modes of the IEEE 802.15.6 standard (64 ns), and the NBI autocorrelation function (2.5), the narrowband interference samples form a white Gaussian process with variance $\frac{J_0}{2}$.

Hence, the signal-to-noise term E_p/N_0 in (2.17) and (2.23) can be replaced by the following signal-to-interference-plus-noise (SINR) term.

$$\mathsf{SINR} = \frac{E_p}{N_0 + J_0}. \tag{2.24}$$

Table 2.3: Simulation results

	N_p	T_w (ns)	$\bar{\mu}_t$	η (Default)	η (High QoS)
LA	-	-	-	**0.58**	**0.43**
PHY 0	32	64.103	0.62	0.22	0.12
PHY 1	16	32.051	0.41	0.34	0.20
PHY 2	8	16.026	0.27	0.44	0.29
PHY 3	4	8.012	0.20	0.42	0.34
PHY 4	2	4.006	0.17	0.35	0.22
PHY 5	1	2.003	0.15	0.24	-

A comparison of the bit error performance of different data rates of the 802.15.6 UWB PHY, assuming energy detection for the default PHY and autocorrelation-differential detection for the high QoS PHY, is summarized in Fig. 2.4. The solid lines correspond to full energy capture and the dashed lines result by taking into account the impact of $\bar{\mu}_t$. The corresponding values of $\bar{\mu}_t$ for each PHY mode can be found in Table 2.3. At higher data rates, where the symbol duration (as well as the integration time) is shorter, the amount of the collected energy is lower. This effect leads to a performance degradation of up to 7 dB at higher data rates.

2.4.4 Comparison of Burst and Single Pulse Options

By increasing the number of pulses per symbol the level of noise power at the receiver's input will increase. From this point of view, and assuming constant symbol energy $E_s = N_p E_p$ in (2.17) and (2.23), the single pulse option outperforms the error performance of the burst pulse option. On the other hand, the burst pulse option is more immune to interference. Another interesting property of the burst option is the fact that, when the pulse energy E_p is constant, the average transmit energy per symbol in different physical layer modes varies depending on the number of pulses per symbol. Based on this property, the transmit energy per symbol is lower in PHY

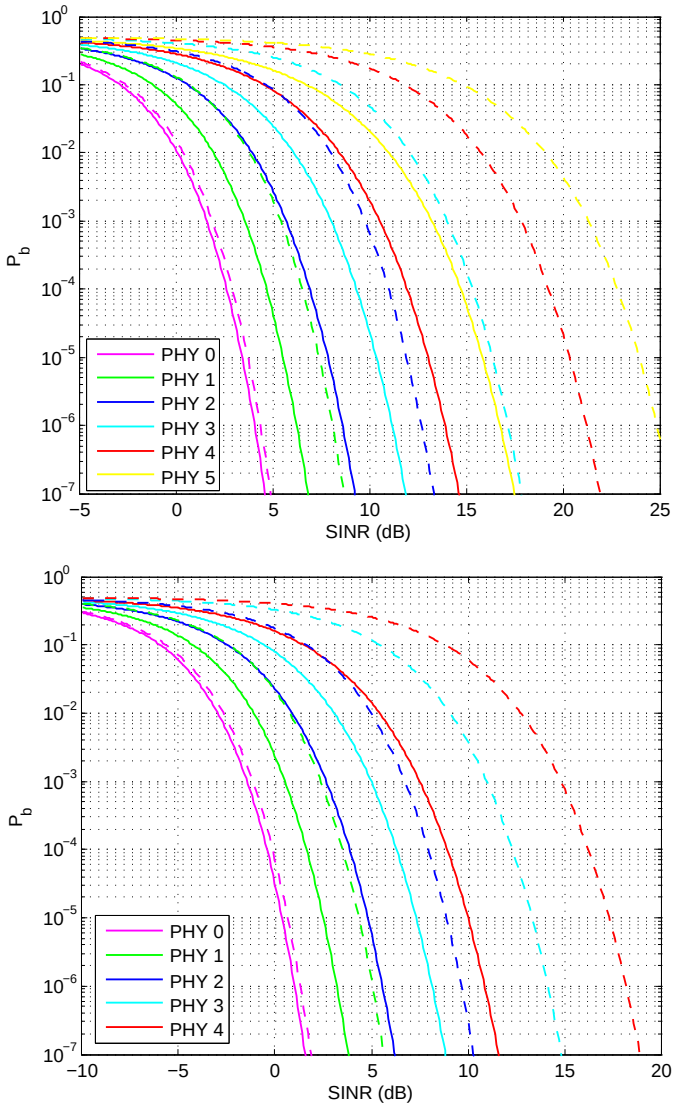


Figure 2.4: Bit error performance of various PHY modes of 802.15.6 considering energy detection with on-off signaling in default PHY (top), and autocorrelation-differential detection in high QoS PHY (bottom). The solid lines correspond to full energy capture ($\mu_t = 1$), and the dashed curves result by taking into account the impact of $\bar{\mu}_t$ in CM3.

modes with higher data rates, which is consistent with traditional adaptive modulation and coding (AMC) schemes. Normally, AMC mechanisms avoid transmission at higher data rates in harsh channel conditions. It is because of the reduced average transmit energy per bit due to the usually larger

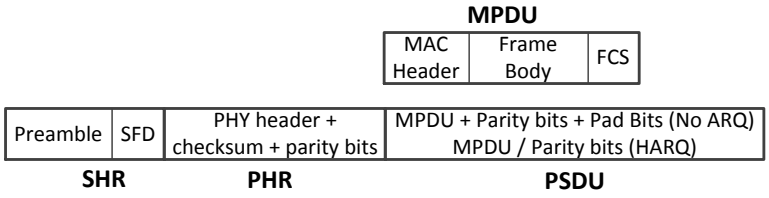


Figure 2.5: IEEE 802.15.6 UWB PHY packet structure [10].

constellation sizes at higher data rates. Indeed, high data rate PHY modes are more prone to channel errors and should be used when errors are less likely.

In the rest of this section, we assume a **constant pulse energy** and **different number of pulses per symbol** for different PHY modes. We also assume that the value of E_p is known at the hub.

2.4.5 Energy Efficiency

As illustrated in Fig. 2.5, the physical layer protocol data unit (PPDU) is composed of three parts: a synchronization header (SHR), a physical layer header (PHR), and the corresponding PHY service data unit (PSDU). The SHR is used for packet detection and synchronization and is composed of a preamble and a start of frame delimiter (SFD). The packet success rate (PSR) is defined as the probability of successful reception of the PPDU.

The calculation of the PSR and the energy efficiency of different PHY modes can be found in Appendix A. We can model the energy efficiency $\eta \in [0, 1]$ of the PHY mode m as

$$\eta = \frac{\varepsilon_1 L_p}{\varepsilon_1(L_p + L_h(R_p/R_h)) + E_0}\mathsf{PSR}(\gamma, \bar{\mu}_t, m) \tag{2.25}$$

where $\gamma = E_p/N_0$, ε_1 represents the useful energy for communication of one bit of information, L_p and L_h are the lengths of data payload and overhead having rates R_p and R_h respectively, and E_0 is the total required extra energy which models the energy required for synchronization, data encoding/decoding, and transmission and reception of packet acknowledgment messages. For further details of the above energy efficiency model see Appendix I. We assume that E_p is constant and is known at the hub.

Comparing (A.9) with the throughput of the system defined by [38]

$$T(\mathsf{PSR}_m) = \frac{L_p}{L_p + L_h(R_h/R_p)} R_p \mathsf{PSR}_m, \tag{2.26}$$

one observation is that a throughput-optimal strategy is not necessarily energy efficiency optimal. However, the two performance measures are closely related and their joint optimality is application specific. We will compare both adaptation strategies in our simulations.

2.5 Link Adaptation

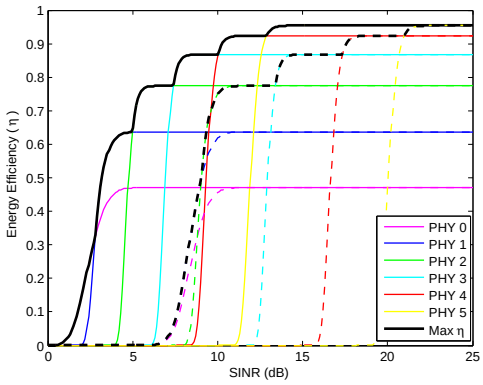
2.5.1 PHY-Aware MAC Protocol

Our LA is based on the (γ, $\bar{\mu}_t$) pair. Assume that a sensor s ($s \in [1, N_s]$) transmits data to the hub. The task of the LA algorithm is to identify for the hub the PHY mode m_s^* a given sensor s should select in order to maximize η in (A.9).

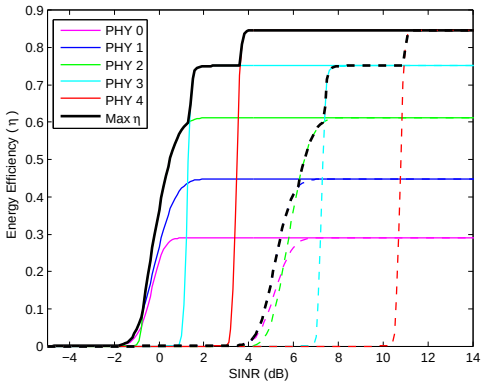
$$m_s^* = \arg\max_m \{\eta \mid (\gamma, \bar{\mu}_t, m)\},\ m \in [0, M-1]. \tag{2.27}$$

Assuming full energy capture ($\bar{\mu}_t = 1$), this mechanism is straightforward and is depicted in Fig. 2.6 for default and high QoS PHY. The solid lines represent full energy capture. The amount of captured energy depends on the instantaneous channel as well as the integration time. Our simulations for imperfect energy capture are carried out assuming channel model CM3 and are demonstrated by dashed lines. The average statistic for $\bar{\mu}_t$ is used, based on the approximation provided in Section 2.3.2. It can be seen that a significant energy loss will result from a short integration time at higher data rates.[1] Therefore, the objective is to utilize the transmitted energy as much as possible. However, knowledge of $\bar{\mu}_t$ is required for selecting the optimal PHY mode.

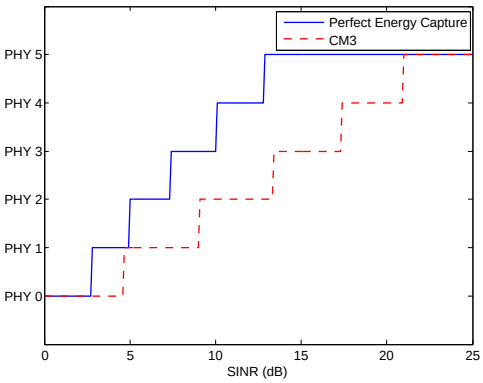
[1]Typically, the RMS delay spread in a hospital environment is between 10 ns and 17 ns [71]. So, according to Chebyshev's inequality, the integration time should be between 32 ns and 41 ns to capture 99% of the pulse energy.



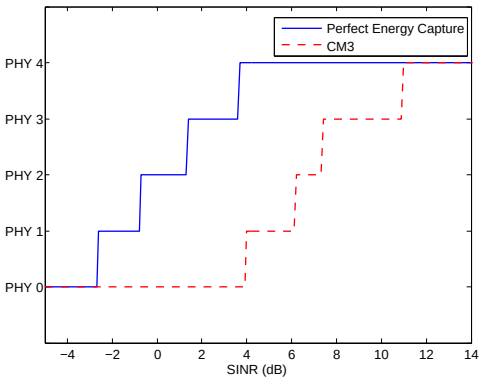
(a) Energy efficiency of different data rates - Default PHY.



(b) Energy efficiency of different data rates - High QoS PHY.



(c) Optimal adaptation map - Default PHY.

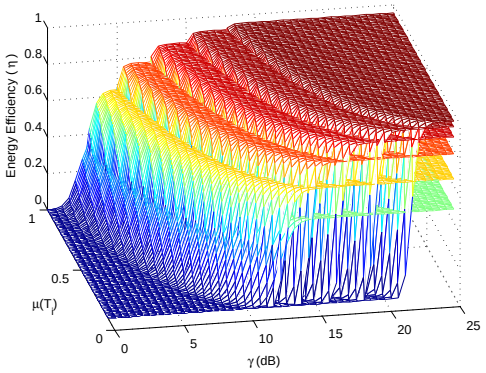


(d) Optimal adaptation map - High QoS PHY.

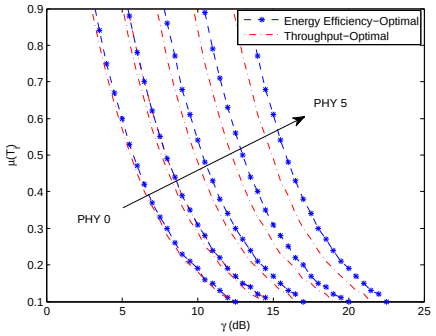
Figure 2.6: Energy efficiency of different data rates and the optimal adaptation maps. The solid lines are for full energy capture and the dashed lines are for partial energy capture in CM3.

2.5.2 The Proposed Adaptation Strategy

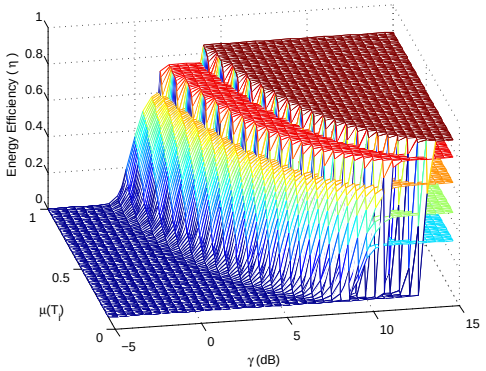
For any given sensor we first define $\gamma = E_p/(N_0 + J_0)$, which is the received SINR at the hub with full energy capture ($\bar{\mu}_t = 1$), and is assumed to be known. Note that E_p is already known at the hub and $N_0 + J_0$ should be known for setting the detection threshold in the high QoS PHY, anyway. The hub can measure the received energy in a null timeslot to estimate $(N_0 + J_0)$.



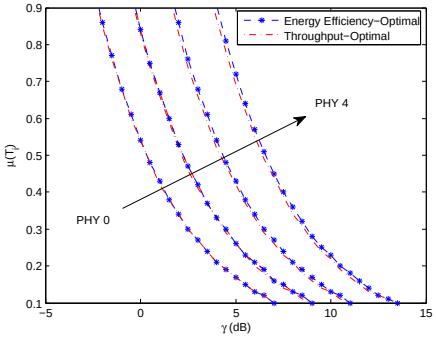
(a) Energy efficiency in different modes - Default PHY.



(b) Optimal adaptation map - Default PHY.



(c) Energy efficiency in different modes - High QoS PHY.



(d) Optimal adaptation map - High QoS PHY

Figure 2.7: Energy efficiency of different PHY modes (a,c), and the optimal adaptation map corresponding to maximum energy efficiency (b,d). The lines represent the border between two neighboring PHY modes.

The LA strategy is based on a local statistic derived from the decision variable as well as γ. Considering (2.13) and (2.21), we can use the first moment of the decision variable to estimate $\bar{\mu}_t$. The following estimators over the interval $A = [0, 1]$ can approximate the energy capture index for a

given pulse index j, in the ED and AC receivers, respectively

$$\hat{\mu}_D[i] = \Psi_A\left(\frac{\sum_{j=0}^{\tau-1}|Z_{ED}[i-j]|}{\tau N_p E_p}\right), \tag{2.28}$$

$$\hat{\mu}_H[i] = \Psi_A\left(\frac{\sum_{j=0}^{\tau-1}|Z_{AC}[i-j] - N_p B T_p N_0|}{\tau N_p E_p}\right), \tag{2.29}$$

The hard-limiting function $\Psi_A(\cdot)$ is to make sure that the estimation is within the interval $A = [a, b]$, and is given by

$$\Psi_A(x) = \begin{cases} b, & x \geq b \\ x, & a < x < b \\ a, & x \leq a. \end{cases} \tag{2.30}$$

After LA finds the optimal mode at the hub, it should send a management/control packet to the node to provide it with the optimal PHY mode.

Fig. 2.7 demonstrates the adaptation maps and the corresponding energy efficiency values. It also shows the throughput-optimal map, which is different from the energy efficiency optimal map. The energy efficiency of each mode is the overall maximum only in a portion of the curve. One observation is that in the mandatory mode of the default PHY (PHY 0), the maximum energy efficiency is less than 30%, which is optimal at low γ. Indeed, with static communication on PHY 0 (no adaptation), the achieved energy efficiency is less than half that of the case with LA capability. For the high QoS PHY, the optimal region corresponding to PHY 0 is in low values of the curve. Consequently, there is always a better option with superior energy efficiency. In other words, communication on the mandatory mode of the high QoS PHY is not optimal when the higher data rates (e.g. PHY 1, or PHY 2) are available. Note that our adaptation algorithm is based on the pair ($\hat{\mu}_t$, γ), and therefore, it is able to respond to variations of the captured energy caused by dynamic evolutions of the channel resulting from body movements and other dynamic behaviors in typical WBAN applications.

2.6 Simulation Results

The proposed scheme has been extensively simulated assuming different γ values and the channel model CM3. The average values of $\bar{\mu}_t$ and η are compared with the static system, i.e. fixed communication on different PHY modes in Table 2.3 (PHY 0 is the mandatory mode of the standard). Since the LA scheme uses all modes in different conditions, only the overall energy efficiencies are written for this scheme. The range of γ (in dB) is uniformly within [-5, 25] and [4, 14] for the default and high QoS PHY simulations, respectively. It can be seen that the proposed scheme can significantly improve the energy efficiency of the system. To quantify the estimation accuracy of the proposed estimators for the energy capture index, we also used (2.13), (2.14) and (2.21), (2.22) to generate random samples of the decision variable for a given pair (T_w, γ), assuming $\tau = 100$ and $E_p = 1$ for simplicity. The achieved mean square error (MSE), for the estimators given in (2.28) and (2.29) are 0.0595 and 0.0523, respectively. We can summarize the simulation results as follows:

Fixed integration time

In a given integration time, as γ increases it is optimal to shift to a higher physical layer mode. The LA scheme can also respond to any increase/decrease in the level of noise and interference in a fixed signal level by shifting to a lower/higher data rate or equivalently by increasing/decreasing the transmit energy per symbol.

Fixed γ

When γ is constant, as the energy capture index increases (e.g. when the channel delay spread slightly changes to a lower value), it is optimal to shift to higher data rates. This can be interpreted by the fact that the transmit energy per symbol is lower in higher PHY modes and an increase in the amount of captured energy indicates an improved channel state which means a smaller bit-level error probability. Therefore, to preserve a certain

level of performance, it is more energy efficient to shift to a higher data rate. On the other hand, as $\bar{\mu}_t$ decreases, a lower physical layer mode (with a higher symbol energy) is necessary to increase the total amount of captured energy per symbol.

2.7 Conclusion

Reliability and energy efficiency are two crucial requirements in WBANs, which necessitate a sophisticated mechanism to overcome the highly dynamic and harsh channel conditions and strict design limitations. Achieving these requirements is more challenging in noncoherent UWB receivers, where no channel state information is available. In this chapter, we address this problem by proposing a novel link adaptation strategy for IEEE 802.15.6 IR-UWB systems. Assuming a constant pulse energy among different PHY modes, and considering the different number of pulses recommended by the standard, the symbol energy decreases in higher PHY modes. However, the higher data rates should only be selected in reliable channel conditions, since they are more prone to channel errors. To solve this problem, the proposed adaptation scheme is based on the estimated SINR, as well as the energy capture index which is a local statistic derived from the decision variable. Since the energy capture index depends on several factors including the instantaneous channel and the integration time, the proposed scheme is sensitive to temporal variations in channel conditions, noise, and interference. Therefore, it can correctly select the most energy efficient PHY mode to maintain the required level of performance.

Chapter 3

Frame Length Optimization in IEEE 802.15.6 WBANs

3.1 Introduction

In this chapter we tackle the frame length optimization to maximize the energy efficiency of WBANs based on the impulse radio (IR) UWB [72] physical layer (PHY), defined in the IEEE 802.15.6 standard [10]. The motivation to exploit adaptive schemes in WBANs has been discussed in [40]. Indeed, to achieve stable quality of service (QoS) performance it is indispensable to design agile adaptive schemes for WBANs. One approach is to optimize the frame body length in the MAC frame, which is addressed here.

The IEEE 802.15.6 standard is composed of two different IR-UWB PHY modes. The default mode is for general purpose WBAN applications and exploits on-off signaling and the BCH(63, 51) code for forward error correction (FEC). The high QoS mode is exclusive to medical and high priority applications and benefits from differential signaling with type II hybrid automatic-repeat-request (H-ARQ) that combines the BCH(126, 63) code with ARQ. In this chapter we also derive the packet success rate of both PHY modes of the standard, which is fundamental in cross-layer analysis and adaptive link layer strategy design. Furthermore, a generic energy effi-

ciency model is proposed incorporating separate energy consumption costs for uplink and downlink channels, as well as for reception and transmission. The chapter is organized as follows. We first describe the system model, followed by the calculation of packet success probabilities. Then we formulate and solve the optimization problem to maximize the energy efficiency in both UWB PHY modes of the standard. At the end of this chapter we propose adaptive frame length optimization schemes applicable to narrowband and UWB WBANs. The proposed adaptive scheme is verified using narrowband channel measurements at 2.4 GHz.

Notation: We use a superscript letter to denote the PHY mode, i.e., "d" for the default mode, "q" for high QoS mode, and "m" for mode in general, and use double parenthesis to prevent confusion with power.

3.2 System Model

We assume a set of wireless sensor nodes that communicate with a master hub node. The hub-to-sensor and sensor-to-hub links are referred to as the downlink and uplink, respectively.

It is well known that rake receivers can capture a significant amount of the transmitted energy by resolving multipath fading. However, due to a demand for low complexity, suboptimal non-coherent receivers are of more interest. Such receivers are based on either energy detection (ED) or autocorrelation (AC) [68], which are assumed in this section.

As illustrated in Fig. 2.5, the physical layer protocol data unit (PPDU) is composed of three parts: a synchronization header (SHR), a physical layer header (PHR), and the corresponding PHY service data unit (PSDU). The SHR is used for packet detection and synchronization and is composed of a preamble and a start of frame delimiter (SFD). The preamble provides four logical channels by using a pool of four Kasami sequences (symbols) of length 63. For coexisting WBANs, it is assumed that the logical channel with minimum received energy is assigned by the hub. The preamble is constructed by four repetitions of the Kasami symbol. At the receiver side, each packet is successfully detected if the cross-correlation of the ex-

pected synchronization sequence with the received sequence is higher than an implementation-dependent sensitivity margin, or their hamming distance is less than a specific threshold ρ (we assume $\rho = 6$ in the simulations). Therefore, the probability of successful packet detection can be given by

$$P_{SHR}^{(m)} = P_{SFD}[1 - (1 - P_{syn}^{(m)})^4], \tag{3.1}$$

where m is the mode index, P_{syn} and P_{SFD} are the probabilities of successful detection of the corresponding Kasami sequence and the SFD respectively. Here $P_{SFD} = P_{syn}$, since the SFD is constructed by inverting the selected synchronization symbol. P_{syn} can be expressed as

$$P_{syn}^{(m)} = \sum_{i=0}^{\rho} \binom{63}{i} [P_b^{(m)}]^i [1 - P_b^{(m)}]^{63-i}, \tag{3.2}$$

where $P_b^{(m)}$ is the bit error probability of PHY mode m.

The PHR which contains the PHY settings is composed of 24 bits of PHY header, 4 bits of check sequence from CRC-4 ITU error detection code, plus 12 and 63 parity bits from a shortened BCH(40, 28) and BCH(91, 28) code for default mode and high QoS mode, respectively. The probability of successful reception of the PHR is

$$P_{PHR}^{(m)} = \sum_{i=0}^{t^{(m)}} \binom{N_{PHR}^{(m)}}{i} [P_b^{(m)}]^i [1 - P_b^{(m)}]^{N_{PHR}^{(m)}-i}, \tag{3.3}$$

where $t^{(d)} = 2$ and $t^{(q)} = 10$ are the correction capabilities of the BCH codes, and $N_{PHR}^{(d)} = 40$ and $N_{PHR}^{(q)} = 91$ are the lengths of coded data for the two modes respectively.

The MAC PDU consists of a header and a frame checksum of fixed lengths of 7 and 2 octets respectively, plus a MAC frame body of a variable length $L_{FB}^{(m)}$. It is mapped in blocks of length $k^{(m)}$ to $N_{CW}^{(m)}$ codewords of length $n^{(m)}$ ($n^{(d)} = 63$, $k^{(d)} = 51$, $n^{(q)} = 126$, $k^{(q)} = 63$, considering different BCH encoders), to form a PHY frame. If $\mathrm{mod}(72 + L_{FB}^{(m)}, k^{(m)}) \neq 0$, then $N_{BS} = k^{(m)} N_{CW}^{(m)} - (72 + L_{FB}^{(m)})$ bits are appended to the last codeword so that

$$N_{CW}^{(m)} = \left\lceil \frac{72 + L_{FB}^{(m)}}{k^{(m)}} \right\rceil, \tag{3.4}$$

where $\lceil . \rceil$ denotes the ceiling operator. Data bits are interleaved prior to modulation to remove the effect of burst errors, and pad bits are added for symbol boundary alignment, if necessary[1]. The probability of successful reception of PSDU can be expressed as

$$P_{PSDU}^{(m)} = [P_{CW}^{(m)}]^{N_{CW}^{(m)}}. \tag{3.5}$$

Note that $P_{PSDU}^{(m)}$ is a function of the frame length presented by the number of codewords $N_{CW}^{(m)}$. Here $P_{CW}^{(m)}$ is the probability of successful reception of a codeword, and can be given by

$$P_{CW}^{(m)} = \sum_{i=0}^{t^{(m)}} \binom{n^{(m)}}{i} (P_b^{(m)})^i (1 - P_b^{(m)})^{n^{(m)}-i}. \tag{3.6}$$

Based on the derived success probabilities for the packet headers and the synchronization bits, the packets are successfully delivered in the default mode with the probability

$$P_{PPDU}^{(d)} = P_{SHR}^{(d)} P_{PHR}^{(d)} P_{PSDU}^{(d)}. \tag{3.7}$$

The high QoS mode benefits from the H-ARQ mechanism combining FEC with ARQ. This scheme will require both the transmitter and receiver to store data packets. At first, a preliminary BCH(126, 63) encoding is performed on the systematic bits D (D = MAC header + MAC frame body) to generate the parity bits P of the same length. Then D is encoded with the CRC-16-CCITT code and transmitted along with the corresponding frame check sequence (FCS). At the receiver, the content of D is not discarded in case of errors detected by the FCS. Instead, it is stored and error correction will be carried out based on the existing erroneous packet and the following retransmissions. On the next trial, P is encoded with the CRC-16-CCITT code and is transmitted along with the corresponding FCS. The systematic bits are recovered by inverting P bits, if they are error free. Otherwise, if the receiver could not recover the errors based on the received P and the stored D, D is discarded and the parity bits are stored. This process will

[1]No pad bits are required at the mandatory data rates.

take place for a maximum of $R_{max} = 4$ times. After reaching the maximum number of retransmissions, the receiver resets its data storage and H-ARQ begins at the initial stage.

If the MAC checksum verifies the integrity of the first packet, the data/parity fragments are received successfully with probability $\delta = P_{SHR}^{(q)} P_{PHR}^{(q)} (1 - P_b^{(q)})^{N_{CW}^{(q)} n^{(q)}/2}$. In case of an error, retransmission is required, and the probability of successful reception after each retransmission, under the assumption that the error occurrence process is stationary during alternating retransmissions and is independent of data/parity fragments, can be given by

$$P_x = \delta + (1 - \delta)[P_{SHR}^{(q)} P_{PHR}^{(q)}]^2 P_{PSDU}^{(q)}. \tag{3.8}$$

The rationale is that either P or D is received correctly with probability δ, or D is recovered by the FEC. Both packets should be decodable by the physical layer in the latter case. So we can summarize the probability of successful reception of the data fragment D after the r'th retransmission as

$$P_D(r) = P_x(1 - P_x)^{r-1}, r \in [1, R_{max}]. \tag{3.9}$$

The H-ARQ mechanism is able to deliver the packets with probability

$$P_{PPDU}^{(q)} = \delta + (1 - \delta) \sum_{r=1}^{R_{max}} P_D(r). \tag{3.10}$$

3.3 Optimal Frame Length

Substituting (3.4) to (3.7) into (A.9) and dropping the mode index for simplicity, the energy efficiency of the default mode is

$$\eta = \frac{\varepsilon_1 L_{FB}}{\varepsilon_1 L_{FB} + c_0} P_{SHR} P_{PHR} [P_{CW}]^{\frac{L_{FB}+72}{63}}, \tag{3.11}$$

where $c_0 = \varepsilon_1(\sum_i \tau_i (R_b/R_{\tau_i})) + \varepsilon_2 + E_0$ and P_{CW} is given in (3.6). The optimal frame body length L_{FB}^* is obtained by setting $\frac{d\eta}{dL_{FB}} = 0$ in (3.11)

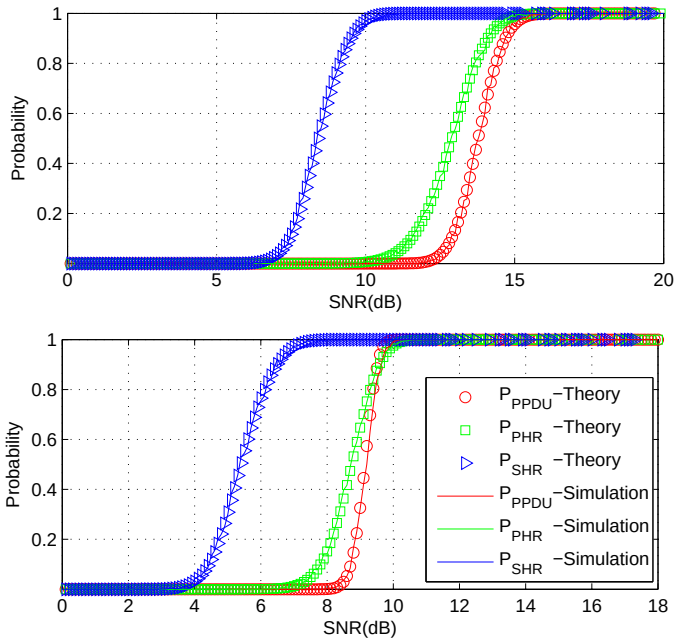


Figure 3.1: Comparison of the theoretical model with the simulation results for the default (top), and high QoS (Bottom) PHY.

and is expressed as[2]

$$L_{FB}^{*} = \left[\sqrt{\left(\frac{c_0}{2\varepsilon_1}\right)^2 - \frac{63c_0}{\varepsilon_1 \ln(P_{CW})}} - \frac{c_0}{2\varepsilon_1} \right]. \tag{3.12}$$

It can be observed that when P_{CW} is low (which represents a high probability of packet errors), the second term under the square root tends to a smaller positive number and hence a smaller frame length is optimal. For the high QoS mode, there is no explicit form. However, (A.9) can be directly calculated and numerically maximized for a given bit error probability using conventional convex optimization techniques.

[2]It can be shown that η has a maximum when $L_{FB} > 0$. The negative solution is not feasible and is discarded. We round-off the answer to the nearest integer which is denoted by the $[\cdot]$ operator.

Table 3.1: Optimal Frame Lengths (octets)

	Bit Error Probability (P_b)	L^*_{FB}
Default	5.2×10^{-3}	300
	8.8×10^{-3}	134
	1.2×10^{-2}	76
High QoS	4.7×10^{-2}	335
	5.4×10^{-2}	190
	6.2×10^{-2}	108

3.4 Numerical Analysis

To verify the statistical model presented in Section 3.2, we have simulated the mandatory data rates of the IEEE 802.15.6 physical layer. The same packet reception procedures described in Section 3.2 are simulated in the Additive White Gaussian Noise (AWGN) channel which is also used for simulations within the standard. We first compare the output of the theoretical model with the simulation results. Fig. 3.1 shows that the experimental success probabilities of PPDU, SHR, and PHR, averaged over 1000 realizations, are accurately followed by their theoretical counterparts. The minimum required SNR for 99% packet success probability (P_{PPDU}) is 15.5 dB for ED with on-off signaling (default mode), and 9.8 dB for AC with differential signaling (high QoS mode). In Fig. 3.2, the energy efficiency is plotted using (A.9) for different frame lengths and bit error probabilities for the two modes. It is assumed that $c_{tu} = c_{rd} = 0.9$, $c_{ru} = c_{td} = 0.1$, $E_{enc} = 4$ pJ, $E_{dec} = 2$ nJ, $\bar{\tau}^{(d)} = 427$, and $\bar{\tau}^{(q)} = 487$. The optimal length values are given in Table 3.1. Note that the optimal length grows quickly as the bit error probability decreases. For extremely low bit error probabilities, the corresponding optimal length is unfeasibly large. Indeed, the realistic frame length values are upper-bounded by other factors such as the maximum allowable delay, the buffer size, and the coherence time of the channel.

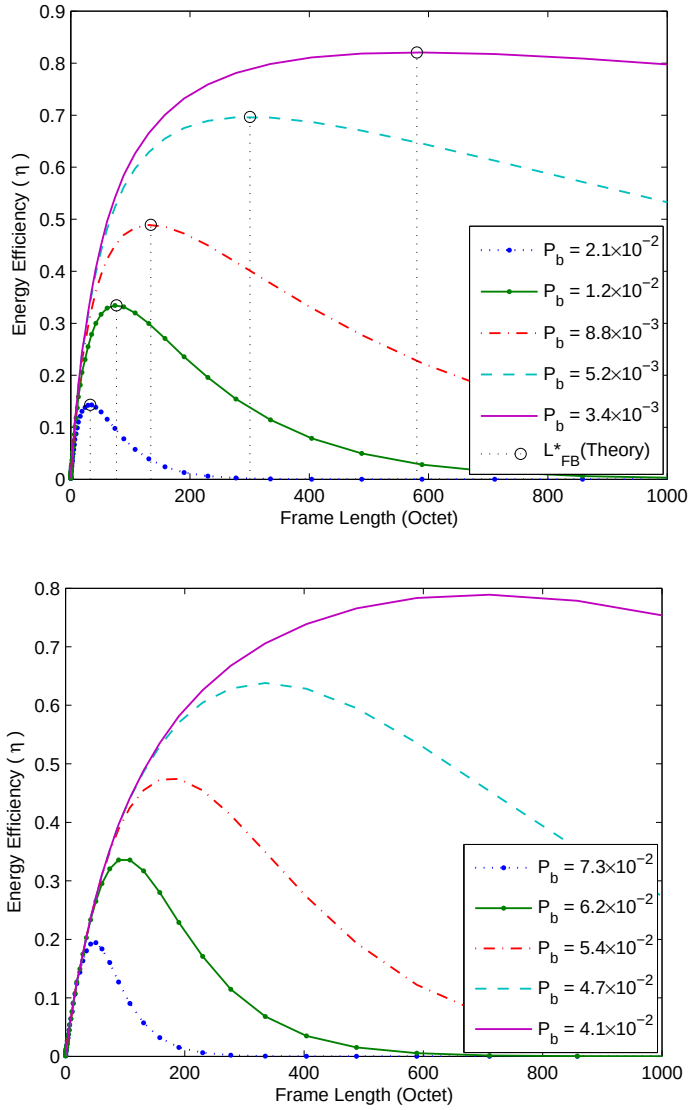


Figure 3.2: Energy efficiency vs frame length, with different bit error probabilities for the default (top), and high QoS (Bottom) PHY. The optimal frame lengths based on the derived closed form in (3.12) are shown for the default mode (top).

3.5 Channel Adaptive Frame Length

The nature of a wireless channel in a wireless body area network is highly dynamic, and hence any static design is trivially sub-optimal. In this sec-

tion we propose a cross-layer channel-adaptive framework to adjust the MAC frame body length of the data packets. We propose two different approaches to incorporate the impact of the time-varying channel gain in the transmission policy aiming to improve the overall energy efficiency of the system. The first approach is based on autoregressive channel prediction and the second scheme is a novel procedure inspired by the well-known slow-start mechanism. Explicitly, in our first scheme the channel gain is predicted and the optimal frame length is selected. In the second method, the MAC frame length is sequentially increased after receiving the acknowledgment message or decreased if no acknowledgment is received. Our simulation results show that the proposed schemes can significantly improve energy efficiency in comparison with the case with a fixed frame length.

3.5.1 Advantage of Adaptive Schemes

Adaptive schemes have brought enormous attention in the design of wireless body area networks (WBANs) [73, 74, 75, 39]. This is due to the strict design constraints on one hand and the unique channel conditions in WBANs on the other hand. Low complexity is the main design constraint and is driven by several factors. The extremely small form factor of the motes, a reasonably long expected lifetime though powered by very limited battery resources, and the need for low-power operation to prevent tissue heating problems, together impose an upper feasibility bound on the maximum possible complexity of the motes.

There is however, an implicit lower bound for the complexity of the motes that is governed by a crucial need for optimality and efficiency in WBANs. It is now evident that the link quality in WBANs can dramatically change based on different factors such as body movements and occasional signal blockages [3], body shape and gender [76, 4], posture [5, 8], clothing and ornaments [6], existence of in-body medical implants with unknown reflective and dispersive properties [7], transmitter-receiver orientation [77, 8], and finally different propagation characteristics of tissue among different individuals and different body parts [78, 79, 9]. Indeed, due to the dynamic

conditions of the wireless channel in typical WBANs a static design fails to satisfy all design requirements since it is inherently sub-optimal and inefficient. Therefore, the minimum imaginable level of complexity for these systems to be efficient is identified by their adaptive capabilities.

Power control (PC) and link adaptation (LA) are two major adaptive schemes that can greatly increase the efficiency of the system. It is shown in [40] that LA can save up to 85% on energy by imposing a compromise of less than 1% on the error probability. Similarly, it can be shown that PC can save nearly 35% on energy without degrading reliability [73]. Nevertheless, cross-layer optimization of the system can still improve the overall efficiency, and since radio is the most energy-consuming part of the motes [40] it can lead to a significant improvement.

It is shown in [80] that the performance of MAC protocols that ignore channel temporal variations in WBANs is dramatically degraded. The length of the medium access control (MAC) frame is one of the parameters that can influence the energy efficiency, and has been addressed in different applications [20]. We propose two different channel-adaptive approaches to dynamically adjust the length of the MAC frame body of the packets in different channel conditions, aiming to improve the energy efficiency. The first scheme exploits autoregressive prediction of the channel based on previous samples. The predicted value is used to estimate the expected packet error probability and then the MAC frame length is optimized to maximize the energy efficiency for the given packet error probability. The optimized packet lengths can be calculated off-line for different channel gains and stored in a lookup table for fast on-line recovery, at the expense of a small amount of memory. Our second scheme is analogous to the slow-start procedure which is a transport-layer strategy for congestion control. It is desirable for situations where channel prediction is not feasible and the transmitter can only adapt to the channel based on the outcome of the communication of the previous packets. Explicitly, we propose an algorithm that increases/decreases the frame length by a predefined value after every successful/failed packet, indicated by the acknowledgment message. There are different parameters that should be considered in the second scheme,

Table 3.2: $\hat{P}_b$ Fitting Parameters

	w	γ_0	ν	σ
$\pi/2$-DBPSK	0.5300	0.0003	0.9779	0.9586
$\pi/4$-DQPSK	1.5199	-0.0086	0.7996	1.5322

including the initial and the maximum frame length and the step size, that are discussed in the sequel. Note that although the WBAN channel is highly dynamic, the speed of variations is not high since the motion is usually not very fast [80, 74].

3.5.2 Formulation

We have simulated the $\pi/2$-DBPSK, and $\pi/4$-DQPSK modulation schemes recommended in the narrow-band physical layer of the IEEE 802.15.6 standard [10] in an additive white Gaussian (AWGN) channel to derive a unified exponential fit for P_b versus different SNR values. The four-parameter fitting model is expressed as

$$P_b(\gamma) \approx \hat{P}_b = w \exp\left(-\frac{(\gamma-\gamma_0)^{\nu}}{\sigma}\right), \tag{3.13}$$

where γ is the received signal-to-noise ratio (SNR). The empirical bit error rate (BER) and their corresponding exponential fits are illustrated in Fig. 3.3 and the fiting parameters are given in Table 3.2. Note that the fitting model is based on the SNR in natural unit (not dB).

For simplicity and without loss of generality, we assume that the total number of bits in each frame is $N_{CW}k$. Since there is a direct relationship between the number of transmitted bits and the consumed energy per bit, to measure the energy efficiency of the system we use the effective throughput which is the ratio of the number of successful data bits to the total number of transmitted bits. The effective throughput of the system is expressed as

$$T(N_{CW},\gamma) = \frac{(N_{CW}k)}{(N_{CW}n)+\tau}\mathsf{PSR}(\gamma) = \frac{(N_{CW}k)}{(N_{CW}n)+\tau}P_{\tau}(P_{CW}(\gamma))^{N_{CW}}. \tag{3.14}$$

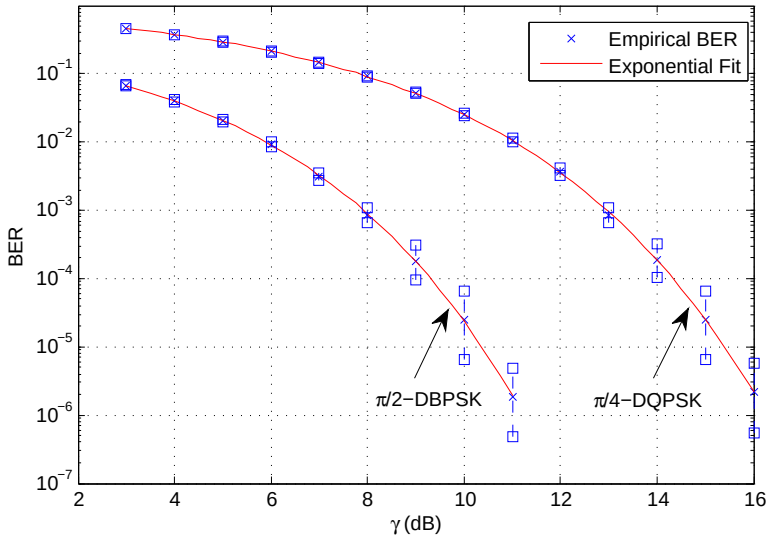


Figure 3.3: The empirical BER of $\pi/2$-DBPSK and $\pi/4$-DQPSK and the corresponding exponential fit.

3.5.3 Channel Prediction

Our first scheme is based on a dynamic prediction of channel gain that takes place ahead of the actual packet transmission time. The prediction task is based on the existing samples of the channel and can be performed both at the hub or motes due to the reciprocity of the channel. The specific temporal distribution of the samples defines the prediction period. A prediction time of up to two seconds has been previously reported for dynamic power control in WBANs [74]. The prediction procedure is completely separate from the adaptation phase and can be implemented according to the specific design requirements, e.g. timing, prediction accuracy, and complexity. We use autoregressive (AR) prediction which is also known as linear prediction (LP) and has already been extensively studied. LP is a simple and accurate method with a performance superior to even more complex methods. In this scheme, the future samples of the channel are modeled by a linear combination of its past p samples expressed as [81]

$$\hat{c}(m) = \sum_{j=1}^{p} d_j(m) c_{m-j}, \tag{3.15}$$

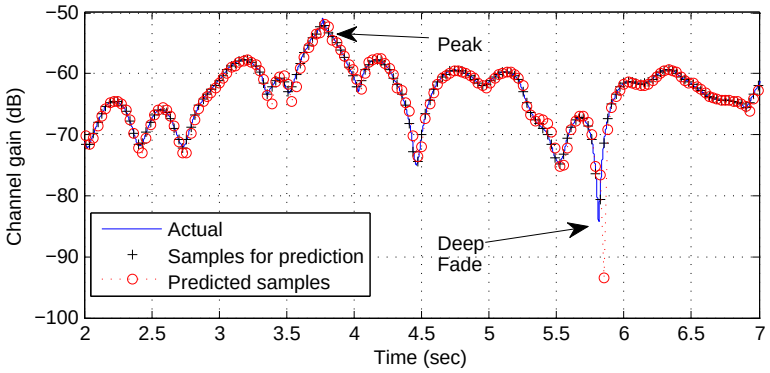


Figure 3.4: Autoregressive prediction of the Tx-Rx channel gain at 820 MHz.

where m is the sample index and p is referred to as the AR model order. The coefficients $\{d_j\}$ are calculated according to the minimum mean square error (MMSE) criterion and the celebrated Yule-Walker equations [70]. The predictor can be implemented in the form of a simple pth order finite impulse response (FIR) filter, which is ideal for WBAN applications with complexity constraints.

An example of the LP algorithm is presented in Fig. 3.4, where the Tx/Rx channel gain in the 820 MHz frequency band are predicted based on the samples taken at 30 ms intervals of the NICTA channel model [61] and with $p = 7$.

For each predicted channel gain $\hat{c}$ the instantaneous SNR is given by

$$\gamma = \frac{\hat{c}P_t}{N_0 B}, \tag{3.16}$$

where P_t is the transmit power, N_0 is the noise power spectral density and B is the bandwidth. We can calculate and adjust the optimal frame length. In fact, the optimal length can be calculated off-line and be stored in a look-up table for different SNR values for fast and efficient on-line adaptation.

3.5.4 Modified Slow-Start Algorithm

The slow-start algorithm was proposed by Jacobson for congestion control [82] and is implemented in the transport layer. Congestion is one of the

factors causing packet drops in the network and happens when the link is overloaded, with too much traffic. Slow-start facilitates reliable communication of data packets by controlling the load of packets that each transmitting node offers to the network. This algorithm works in a window-based approach and has been modified over time. It is mainly composed of two different phases known as additive-increase/multiplicative-decrease (AIMD) which is also adopted in this section.

Let $L(m)$ be the length of the MAC frame at the time index m and assume that its initial value is L_{min}. During the exponential growth phase, which is at the initial stage, the value of $L(m)$ will increase by a constant $a > 0$ after each successful packet indicated by an acknowledgment message from the receiving node. This procedure continues until either no acknowledgment is received, or the size of the MAC frame reaches a predefined threshold ρ_L. This threshold is influenced by several factors e.g. buffer size, maximum permissible delay, etc. (we used ρ_L=500 in the simulations). In the case of a packet drop, the algorithm enters the multiplicative decrease phase, in which the frame length is reduced by a multiplicative factor $0 < b < 1$. Alternatively, if the threshold length is reached, the linear growth phase begins, in which the frame length is increased by v after each successful packet (v is usually one unit). The operation of this scheme is illustrated in the sequential diagram in Fig. 3.5. The $[\cdot]$ operator that is used in the multiplicative reduction phase is to make sure that the frame length is an integer.

3.5.5 Simulation Results and Discussion

Consistent with the narrow-band physical layer of the IEEE 802.15.6 standard, we use the BCH(63, 51) channel code with correction capability $t = 2$. The physical-layer protocol data unit (PPDU) is constructed by a preamble of length 90 bits and a header of length 31 bits, and the data payload. The MAC header and check sequence also increase 72 bits due to the total physical layer convergence protocol (PLCP) overhead bits.

Fig. 3.6 shows the optimal number of codewords in the PLCP for $\pi/2$-

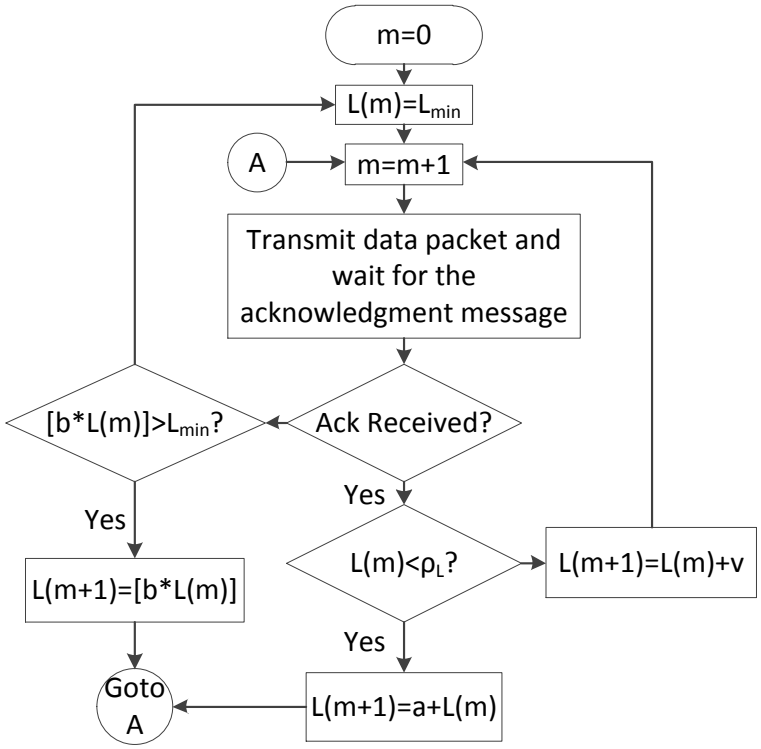


Figure 3.5: Modified slow-start algorithm.

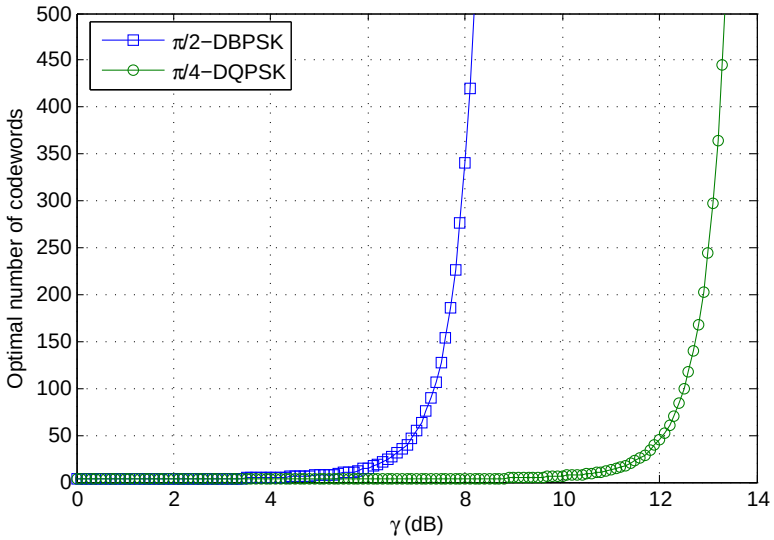


Figure 3.6: The optimal number of codewords for $\pi/2$-DBPSK and $\pi/4$-DQPSK.

Table 3.3: Energy efficiency of the system in different scenarios

		Static	Scheme 1	Scheme 2
Link 1	$\pi/2$-DBPSK	58%	97%	96%
	$\pi/4$-DQPSK	46%	96%	95%
Link 2	$\pi/2$-DBPSK	83%	99%	97%
	$\pi/4$-DQPSK	66%	98%	96%

DBPSK and $\pi/4$-DBPSK modulation schemes. We can observe the corresponding threshold SNRs $\gamma_{th}^{\pi/2\text{-DBPSK}} \approx 8$ dB and $\gamma_{th}^{\pi/4\text{-DQPSK}} \approx 13$ dB for the two schemes, where the optimal packet length grows exponentially. This observation supports the efficiency of burst transmission, e.g., transmitting with maximum frame length (and maximum rate) when channel conditions are suitable and not transmitting otherwise.

To simulate the dynamic behavior of the WBAN channel, we use the NICTA's database which includes channel measurements at 2.4 GHz performed at 200 Hz sampling for up to 12 hours in a typical WBAN [83, 84]. The transmitter is attached on the subject's right hip and two receivers are located at the right wrist (link 1) and right ankle (link 2). The transmit power is 0 dBm and the test scenario consists of different postures and activities, e.g. office work, walking, running, and sleeping. Among all available samples, we use samples with 100 ms time separation.

The energy efficiency of the system for different scenarios (the static scenarios with fixed frame length and our two adaptive schemes) is compared in Table 3.3[3]. It can be seen that the proposed schemes can significantly improve the performance in comparison with the static system. Not surprisingly, the channel prediction scheme outperforms the modified slow-start scheme since it can always select the optimal frame length.

[3]The overall energy efficiency is calculated based on the ratio of the number of successful codewords to the total number of transmitted codewords. Scheme 1 is the prediction-based approach and Scheme 2 is the modified slow-start algorithm.

3.6 Conclusion

The packet success rates in the default and high QoS modes of IR-UWB PHY in the IEEE 802.15.6 standard have been derived and verified by simulations. While these performance measures play a significant role in cross-layer study of the system, we use them to formulate and solve an optimization problem: to maximize the energy efficiency by selecting the optimal length for the MAC frame body. A closed form expression has been derived for the default mode, and the impact of frame length on energy efficiency has been studied. We also propose two channel-adaptive frame optimization schemes that, according to simulations on real WBAN channel measurements, can significantly improve the energy efficiency at the sensor nodes with application in both UWB and narrowband WBANs.

Chapter 4

Energy-Delay Tradeoffs in IR-UWB Systems

4.1 Introduction

For delay-sensitive traffic in WBAN applications, the transmission protocol should take into account the data queuing delay to meet quality of service (QoS) requirements. We consider IR-UWB networks with noncoherent receivers based on energy detection (ED) and autocorrelation (AC), since this setup is an ideal combination of the PHY technologies that can provide unique capabilities such as low complexity and precision localization together with low implementation cost. We assume that the transmission rate (e.g. the transmission PHY mode) can be adapted by an adaptive modulation and coding scheme that controls the number of pulses per symbol. Our study closely follows the theoretical framework in [25]-[85]. Based on the observed instantaneous channel state information (CSI) and queue length information (QLI), the rate scheduling policy is able to select a PHY mode with minimum transmission energy for each sensor node while maintaining the network stability.

The contribution of this chapter is to extend the existing theoretical frameworks in the study of delay-sensitive data transmission to IR-UWB WBANs with noncoherent receivers. We formulate the problem us-

ing our PHY model and apply the results to IEEE 802.15.6 based IR-UWB WBANs. We first calculate the *minimum energy function*, i.e., the minimum energy that is required to achieve network stability for a given traffic rate and provide a mathematical representation for the stability constraint. Then we identify the optimal energy-delay tradeoff in the system, which is fundamental in the design of the scheduling algorithm and expresses the minimum energy required to achieve a specific average delay. Therefore, we pursue a cross-layer approach to integrate the PHY models of the IR-UWB WBANs with the existing higher layer scheduling policies proposed in [25] and [26] to come up with an explicit design model for these networks. Furthermore, we include the total consumed energy required for sensing, processing and communication in the optimization problem for a more comprehensive evaluation, unlike [25]-[85] that only consider the transmission energy in their scheduling algorithm.

4.2 System Model

4.2.1 Physical Layer Characteristics

We assume a WBAN of N_s wireless sensor nodes (motes) and a master hub node that communicate using noncoherent IR-UWB transceivers. To enable various data rates, every node can transmit with a specific modulation and coding scheme which we refer to as a physical layer mode. This can normally be achieved by changing the number of pulses per symbol in IR-UWB systems. Explicitly, considering there are N_m PHY modes and a given physical layer mode $m \in [1, N_m]$, each sensor node uses $N_p^{(m)} \in \mathbb{Z}_+$ pulses of the form $p(t)$ with duration T_p and energy E_p to transmit a symbol. The transmitted waveform is represented as

$$x_k^{(m)}(t) = \sum_i w^{(m)}(t - c_i T_w^{(m)} - i T_s^{(m)}), \tag{4.1}$$

where $k \in [1, ..., N_s]$ and

$$w^{(m)}(t) = \sum_{j=0}^{N_p^{(m)}-1} p(t - jT_p), \tag{4.2}$$

is a waveform of duration $T_w^{(m)} = N_p^{(m)} T_p$, and c_i is the time hopping coefficient selected from the range $[0, N_w - 1]$. N_w indicates the number of possible waveform positions within a symbol time. The symbol time

$$T_s^{(m)} = N_w T_w^{(m)}, \tag{4.3}$$

changes over different PHY modes and hence the amount of transmitted data within a time unit varies depending on the PHY mode. We denote the total energy consumption corresponding to PHY mode m by

$$E_T^{(m)} = N_p^{(m)} E_{TPP}, \tag{4.4}$$

where $E_{TPP} \triangleq \alpha_t E_t + \alpha_r E_r$ is the total weighted energy per pulse consisting of the transmission and reception energies E_t and E_r respectively. α_t and α_r are the energy weights at the transmitter and the receiver, E_t is sum of the pulse energy E_p and the total sensing and processing energies at the transmitter averaged per pulse. Generally, the UWB impulses are subject to regulatory spectral masks and therefore the transmitted pulse energy E_p is usually selected such that $E_p \leq E_{max}$, where E_{max} is the maximum pulse energy permissible by the regulatory. Therefore, to increase the probability of pulse detection, it is useful to assume a fixed $E_p = E_{max}$. The received signal at the hub is given by

$$y_k^{(m)}(t) = \sum_i x^{(m)}(t) * h_k(t) + n(t), \tag{4.5}$$

where $*$ denotes linear convolution, $n(t)$ is a zero-mean additive white Gaussian noise process with two-sided power spectral density $N_0/2$, and $h_k(t)$ is the normalized multipath channel for the kth sensor with $N_{mp,k}$ resolvable multipath components given by

$$h_k(t) = \sum_{l=0}^{N_{mp,k}-1} a_{k,l} \delta(t - \tau_{k,l}), \tag{4.6}$$

in which $\{a_{k,l}\}$ and $\{\tau_{k,l}\}$ represent the sets of corresponding resolvable multipath coefficients and delays, respectively.

We define the average transmission rate corresponding to PHY mode m as

$$r^{(m)} = \frac{1}{N_p^{(m)}},\ m \in [1, .., N_m], \tag{4.7}$$

which is the fractional bits per pulse assuming one bit per symbol. Obviously, the maximum possible rate is achieved by using one pulse per symbol. We define $r^{(0)} = N_p^{(0)} = 0$, which represents the idle mode where the transmission rate and energy are zero. Without loss of generality we assume $0 < r^{(1)} < r^{(2)} < \cdots < r^{(N_m)}$. We define the instantaneously received SNR as $\gamma_k = \mu_k E_p/N_0$. Hence, the average received SNR can be expressed by

$$\bar{\gamma}_k = \bar{\mu}_k E_p/N_0, \tag{4.8}$$

In the sequel, we pursue a cross-layer approach to design a rate scheduling policy that is influenced by the physical layer success probabilities.

4.2.2 Network Model

The medium access of each sensor node is controlled by the hub with full knowledge of the received SNR from all the sensor nodes. Each sensor has a buffer that stores the arrived data from the application layer in a service queue. The network operates in slotted time t with time slot duration Δt. In each beacon period the hub uses management/control packets to inform the sensors about their scheduled physical layer parameters, i.e. the corresponding sensor node and the PHY mode to transmit. For a given time slot t, we denote the scheduled sensor by $K(t)$ and the allocated mode by $M_k(t)$ respectively.

We assume that the channel state remains constant during the transmission of one data packet.

4.3 Minimum Energy with Stability Constraint and CSI

Let $A_k(t)$ represent an i.i.d. ergodic bit-arrival process over time slots corresponding to the kth sensor node with arrival rate $\mathbb{E}[A_k(t)] = \lambda_k$ bits per time slot. We assume that the arrival process is independent of the queue backlog and is bounded, i.e. there exists A_{max} such that $A_k(t) \leq A_{max}$, $\forall t, k$. For each sensor node s_k we denote the state of the queue backlog by $Q_k(t)$ over time. The queue dynamics can be modeled by

$$Q_k(t+1) = \max\{Q_k(t) - F_k(t), 0\} + A_k(t), \tag{4.9}$$

where $F_k(t)$ is the finished work at time t, i.e. the transmitted data from the sensor. We also assume $\lambda_k \in \Lambda, \forall k \in [1, N_s]$, where Λ is the set of all arrival rates for which there exists a randomized scheduling policy that can achieve a finite average delay (Similarly, we denote the complement region of Λ by $\bar{\Lambda}$). Such a policy is defined for any time slot t by the conditional probability distribution

$$P_{K,M|\mathbf{\Gamma}} = \mathbb{P}\left[K = k, M = m | \mathbf{\Gamma} = \boldsymbol{\gamma}\right], \tag{4.10}$$

where K and M are random variables over $\{1, ..., N_s\}$ and $\{1, ..., N_m\}$ indicating the sensor index and the scheduled PHY mode, and $\mathbf{\Gamma}$ is a random vector over the channel state vector $\boldsymbol{\gamma} = (\gamma_1(t), ..., \gamma_{N_s}(t))^{\mathrm{T}}$.

The probability distribution in (4.10) can be written in matrix form $\mathbf{W}(\boldsymbol{\gamma})$ of size $N_s \times (N_m + 1)$ with elements $w_{k,m}(\boldsymbol{\gamma}) \in [0, 1]$ such that

$$w_{k,m}(\boldsymbol{\gamma}) = \mathbb{P}[K = k, M = m | \mathbf{\Gamma} = \boldsymbol{\gamma}]. \tag{4.11}$$

To define the stability of the queue $Q_k(t)$, we define $q_k(U)$ as the largest limiting fraction of time the queue backlog is above U [26], i.e.

$$q_k(U) = \limsup_{t \to \infty} \frac{1}{t} \sum_{t'=0}^{t} \mathbb{P}\{Q_k(t') > U\}, \tag{4.12}$$

then the queue is stable if

$$\lim_{U \to \infty} q_k(U) = 0, \tag{4.13}$$

where lim sup denotes the limit superior. It follows that the necessary and sufficient condition for the stability of the queue for each sensor k is [26]

$$\mathbb{E}[F_k(t)] \geq \lambda_k, \tag{4.14}$$

which necessitate that the average transmission rate for each sensor should be at least equal to the arrival rate[1]. The optimal rate scheduling policy is the solution of the following optimization problem

$$\begin{aligned} \text{Minimize:} \quad & \mathbb{E}\left[\sum_k \sum_m w_{k,m}(\boldsymbol{\gamma}) E_T^{(m)}\right] \\ \text{s.t.} \quad & \mathbb{E}\left[\sum_m w_{k,m}(\boldsymbol{\gamma}) \hat{r}^{(m)} \mathsf{PSR}_k^{(m)}\right] \geq \lambda_k, \ \ \forall k \in [1, N_s] \\ & \mathbf{1}^{\mathrm{T}} \mathbf{W}(\boldsymbol{\gamma}) \mathbf{1} = 1, \ \ \forall \boldsymbol{\gamma} \in \mathbb{R}^{N_s} \\ & w_{k,m}(\boldsymbol{\gamma}) \geq 0, \ \forall \boldsymbol{\gamma} \in \mathbb{R}^{N_s}, k \in [1, N_s], m \in [0, N_m] \end{aligned} \tag{4.15}$$

where $\mathsf{PSR}^{(m)}$ is the packet success rate of the PHY mode m and $\hat{r}^{(m)}$ is the average transmission rate of PHY mode m normalized over a time slot, i.e. bits per time slot, and is given by

$$\hat{r}^{(m)} = r^{(m)} \frac{\Delta t}{T_p}. \tag{4.16}$$

Substituting (4.4), (4.7), and (4.16) it can be written as

$$\begin{aligned} \text{Minimize:} \quad & \mathbb{E}\left[\sum_k \sum_m w_{k,m}(\boldsymbol{\gamma}) N_p^{(m)} E_{TPP}\right] \\ \text{s.t.} \quad & \mathbb{E}\left[\sum_m w_{k,m}(\boldsymbol{\gamma}) \frac{\mathsf{PSR}_k^{(m)}}{N_p^{(m)}} \frac{\Delta t}{T_p}\right] \geq \lambda_k, \ \ \forall k \in [1, N_s] \\ & \mathbf{1}^{\mathrm{T}} \mathbf{W}(\boldsymbol{\gamma}) \mathbf{1} = 1, \ \ \forall \boldsymbol{\gamma} \in \mathbb{R}^{N_s} \\ & w_{k,m}(\boldsymbol{\gamma}) \geq 0, \ \forall \boldsymbol{\gamma} \in \mathbb{R}^{N_s}, k \in [1, N_s], m \in [0, N_m] \end{aligned} \tag{4.17}$$

We denote the term $\mathsf{PSR}_k^{(m)}/N_p^{(m)}$ by $\chi_k^{(m)}$ which is referred to as the "*effective rate*" of PHY mode m and is a function of the channel state γ_k. Let

[1] $\mathbb{E}[F_k(t)] > \lambda_k$ is referred to as strict stability.

$\mathbf{n} = (0, N_p^{(1)}, ..., N_p^{(N_m)})^{\mathrm{T}}$, $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_{N_s})^{\mathrm{T}}$, $\boldsymbol{\chi}_k = (0, \chi_k^{(1)}, ..., \chi_k^{(N_m)})^{\mathrm{T}}$, and $\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{\gamma}) = (\boldsymbol{\chi}_1, ..., \boldsymbol{\chi}_{N_s})$. The stability policy is the solution of the following problem

$$\begin{aligned} \text{Minimize:} \quad & \mathbb{E}\left[\mathbf{1}^{\mathrm{T}}\mathbf{W}(\boldsymbol{\gamma})\mathbf{n}\right] \\ \text{s.t.} \quad & \mathbb{E}\left[\mathbf{W}(\boldsymbol{\gamma}) \circ \boldsymbol{Z}(\boldsymbol{\gamma})\mathbf{1}\right] \geq \frac{T_p}{\Delta t}\boldsymbol{\lambda} \\ & \mathbf{1}^{\mathrm{T}}\mathbf{W}(\boldsymbol{\gamma})\mathbf{1} = 1, \ \ \forall \boldsymbol{\gamma} \in \mathbb{R}^{N_s} \\ & w_{k,m}(\boldsymbol{\gamma}) \geq 0, \ \forall \boldsymbol{\gamma} \in \mathbb{R}^{N_s}, k \in [1, N_s], m \in [0, N_m] \end{aligned} \tag{4.18}$$

where $\circ$ denotes the Hadamard (element by element) product and the total energy is normalized. To solve this problem, the channel state space should be properly discretized to a sufficiently large number of states. In this case, the scheduling policy maps the SNR partitions into a set of PHY mode probabilities for each sensor node. Nevertheless, it follows that the number of optimization variables grows exponentially with the network size. An alternative approach that can simplify the problem would be the case when the stability constraint is satisfied for each channel realization (and hence for each time slot). In this case, the scheduling policy selects with probability one the minimum-energy PHY mode with an instantaneous effective rate greater than or equal to the expected arrival rate of the sensor. Explicitly, $\forall k \in [1, N_s], \forall m \in [1, N_m]$ we have

$$w_{k,m}(\boldsymbol{\gamma}) = \begin{cases} 1, & \text{if } \chi_k^{(m)} \geq \frac{T_p}{\Delta t}\lambda_k \text{ and } N_p^{(m)} \leq N_p^{(j)}, \ \forall j \in [1, N_m] \\ 0, & \text{otherwise.} \end{cases} \tag{4.19}$$

The minimum energy function, which is defined as the minimum average energy per symbol to stabilize the system with a given arrival rate vector $\boldsymbol{\lambda}$ and SNR $\boldsymbol{\gamma}$ can be expressed as

$$\Phi(\boldsymbol{\lambda}, \boldsymbol{\gamma}) = \mathbb{E}\left[\mathbf{1}^{\mathrm{T}}\mathbf{W}_{\boldsymbol{\lambda}}(\boldsymbol{\gamma})\mathbf{n}\right], \tag{4.20}$$

where $\mathbf{W}_{\boldsymbol{\lambda}}(\boldsymbol{\gamma})$ is the scheduling policy corresponding to the given arrival rate vector $\boldsymbol{\lambda}$ and channel state vector $\boldsymbol{\gamma}$. Since $\chi_k^{(m)} \in [0, 1] \, \forall k, m$, we limit the maximum possible arrival rate so that $\lambda_k \leq \lambda_{k,max} = \frac{\Delta t}{T_p}$, $\forall k$ and discard the other scenarios. Clearly, for all k if $\lambda_k > \lambda_{k,max}$, then we have $\lambda_k \in \bar{\Lambda}$.

4.4 Minimum Energy Subject to Stability with QLI

The scheduling policy presented in the previous section is based on the average received SNR from the sensors as well as their average arrival rates. Due to the inherent difficulties in solving (4.18) and its need for the knowledge of average arrival rates, it is useful to pursue approaches based on the QLI of the sensors, instead. Indeed, it is known that scheduling policies that are only based on the channel state and ignore the queue backlog are sub-optimal [25]. The theory of Lyapunov drift seeks scheduling policies based on the queue and channel state information to stabilize the queuing systems. By incorporating a resource optimization part to this theory, it extends to Lyapunov optimization theory, which is extensively studied in the literature [25, 26, 86].

We apply the dynamic scheduling algorithm proposed in [25] and [85] to our IR-UWB network. The main idea is to exploit a specific function, i.e. a Lyapunov function that provides a numerical measure for the current stability level of the system based on its current state. Typically, the function is defined to grow large when the queue backlog of at least one sensor is approaching large values. We refer to the change in the Lyapunov function from one slot to another by the Lyapunov *drift*. It follows that the scheduling policy can stabilize the queuing system by taking control actions (i.e. allocating the PHY modes to the sensors) that make the Lyapunov drift in the negative direction towards zero. In general, the scheduling policy tries to minimize a drift-plus-penalty function which is a sum of the Lyapunov drift and a weighted version of the required spectrum or energy resources that leads to minimum resource utilization while stabilizing the system. We are particularly minimizing the total required energy. Hence, the penalty weight can help the scheduling policy in order to put the system close to any desired point on the optimal energy-delay tradeoff curve.

Given the positive control parameter V, the scheduling policy is based

on the QLI and CSI of all sensor nodes and the following parameters:

$$\epsilon = \frac{1}{\sqrt{(V)}}, \tag{4.21}$$

$$\omega = \frac{\epsilon}{\delta_{max}^2} e^{-\epsilon/\delta_{max}}, \tag{4.22}$$

$$Q_{th} = \frac{6}{\omega} \ln(\frac{1}{\epsilon}), \tag{4.23}$$

where $\delta_{max} = \max\{A_{max}, \max\limits_{(m=0,\ldots,N_m)} \{r^{(m)}\}\}$ is the maximum variation of the queue length, ϵ is the backlog drift, Q_{th} is the buffer partitioning threshold [25], and $\ln(\cdot)$ denotes natural logarithm. At each time slot t, the hub schedules the sensor $K(t)$ to transmit using the PHY mode $M_k(t)$ by solving the following optimization problem

$$(M_k(t), K(t)) = \underset{k\in[0,N_s],m\in[0,N_m]}{\arg\min} \{V N_p^{(m)} - \tilde{W}_k(t)\chi_k^{(m)}\}, \tag{4.24}$$

where $\tilde{W}_k(t) = \max\{W_k(t), 0\}$ and

$$\begin{aligned} W_k(t) =& \mathrm{I}(Q_k(t) \geq Q_{th})\omega e^{\omega(Q_k(t)-Q_{th})} \\ & - \mathrm{I}(Q_k(t) < Q_{th})\omega e^{-\omega(Q_k(t)-Q_{th})} + 2X_k(t). \end{aligned} \tag{4.25}$$

The hub then updates the auxiliary queue $X_k(t)$ such that

$$\begin{aligned} X_k(t+1) = & \max\{X_k(t) - \chi_k^{(M_k(t))}\mathrm{I}(K(t)=k) - \epsilon\mathrm{I}(Q_k(t) < Q_{th}), 0\} \\ & + A_k(t) + \epsilon\mathrm{I}(Q_k(t) \geq Q_{th}), \end{aligned} \tag{4.26}$$

where $X_k(0) = 0$ and $\mathrm{I}(\cdot)$ is the indicator function defined as

$$\mathrm{I}(\textit{condition}) = \begin{cases} 1, & \text{if } \textit{condition} \text{ is true} \\ 0, & \text{otherwise.} \end{cases} \tag{4.27}$$

The average delay of the system is $\frac{1}{N_s}\sum_{k=1}^{N_s} \bar{Q}_k$, where

$$\bar{Q}_k = \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{Q_k(\tau)\} \tag{4.28}$$

is the average delay corresponding to the kth sensor node by Little's theorem [87].

Table 4.1: IEEE 802.15.6 PHY modes

m	Symbol time $T_s^{(m)}$ (ns)	$N_P^{(m)}$	Bit rate (Mbps)
1	2051.300	32	0.487
2	1025.600	16	0.975
3	512.820	8	1.950
4	256.410	4	3.900
5	128.210	2	7.800
6	64.103	1	15.600

4.5 Case Study: IEEE 802.15.6

As an example, we apply our theoretical framework to the IR-UWB physical layer of the IEEE.802.15.6 standard [10]. The mandatory PHY in this standard permits six different data rates in Table 4.1. The success probabilities corresponding to the SHR, PHR, and PSDU packets are derived in the Appendix A and can be calculated based on the bit error probability of the channel.

We use the channel model in [61] (which is explained in section 2.2.1) to simulate the multipath and exploit average channel statistics to quantify the channel quality. The mean μ value for this channel averaged over 1000 realizations is $\bar{\mu}_t = 0.6181$. The sensor subscript k is removed for simplicity. This also applies to λ in the rest of the section. Fig. 4.1 demonstrates $\Phi(\lambda, \bar{\gamma})$ in (4.20) for a single node, assuming different values of the arrival rate λ and average SNR $\bar{\gamma}$ along with the corresponding optimal PHY mode given by (4.19). The infeasible region is labeled by $\bar{\Lambda}$ in which no stability policy can be found. It can be observed that $\Phi(\lambda, \bar{\gamma})$ is a decreasing function of $\bar{\gamma}$ for a fixed λ, since the high-SNR regime can achieve improved effective rates with a lower number of pulses per symbol.

In the sequel we describe the simulations of an IEEE 802.15.6 IR-UWB WBAN consisting of two nodes and a hub with rate scheduling based on CSI and QLI at the hub. To assess the scheduling policy we consider different traffic scenarios and different PHY parameters. The CSI and QLI of all nodes are assumed to be known at the hub. The QLI of a sensor node is informed to the hub in each beacon interval and the CSIs can be blindly estimated by the hub. The scheduling algorithm selects a sensor node and

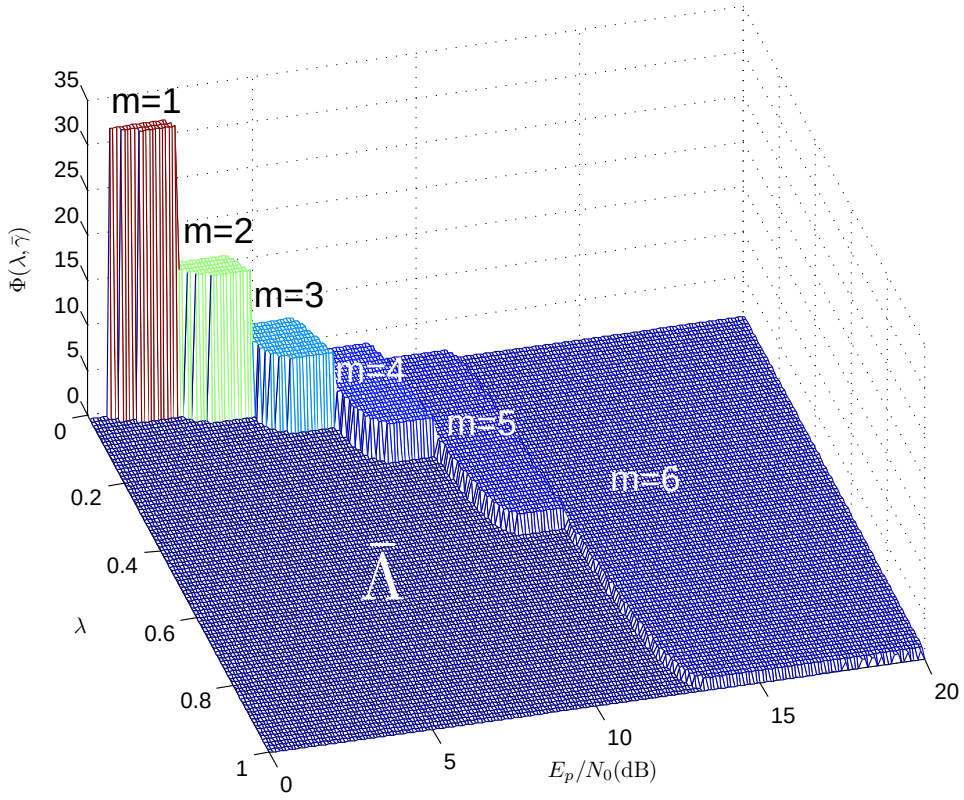


Figure 4.1: Minimum energy function $\Phi(\lambda, \bar{\gamma})$ corresponding to the default IR-UWB PHY option of the IEEE 802.15.6 standard with $\Delta t = T_p$. The required energy and the PHY mode to stabilize a queuing system with arrival rate λ for a single node are given in the figure. The infeasible region is labeled by $\bar{\Lambda}$.

the corresponding optimal PHY mode every beacon period and informs it to the sensors on the downlink channels.

The instantaneous channel at each time slot is simulated based on the model given in [61] and the corresponding multipath factor μ is calculated by integrating the square of the instantaneous channel's impulse response. We assume $B = 1$ GHz, $\Delta t/T_p = 500$, and data arrival is Poisson distributed which means the number of arrived data bits for sensor k the time slot t can be given by

$$\mathbb{P}[A_k(t) = a] = \frac{\lambda^a}{a!} e^{-\lambda}. \tag{4.29}$$

Each sensor node can either transmit using the PHY modes listed in Table 4.1 or wait in the idle mode which is indicated by $m = 0$. The initial

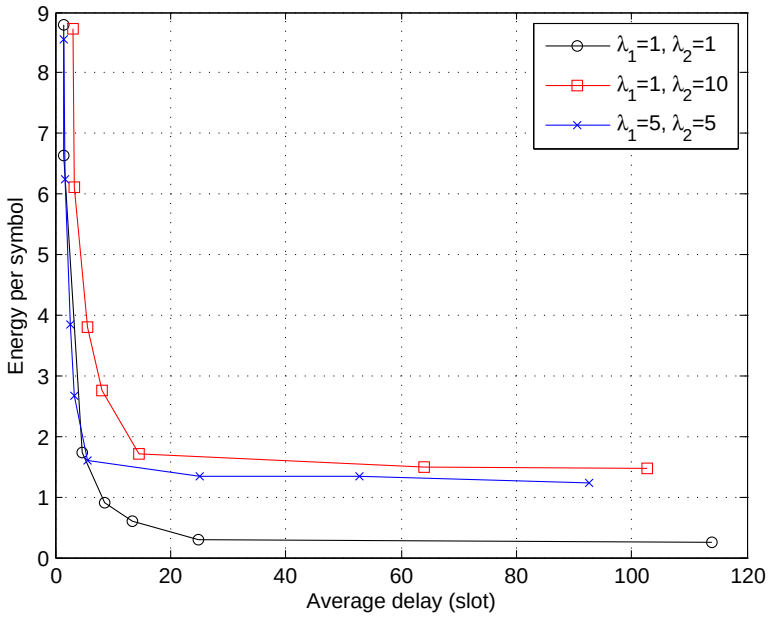


Figure 4.2: The energy-delay tradeoff of the system considering different traffic scenarios.

queue length is assumed zero for all sensors. We are firstly interested in the optimal energy-delay tradeoff in the system which can be achieved by varying the value of V in (4.24). The simulations are performed for three different traffic scenarios and the resulting energy-delay tradeoff curves are depicted in Fig. 4.2. These tradeoffs are of significance importance in the design of the scheduling algorithm since they provide the explicit relationship between the average delay incurred by the queuing system at the sensors and the minimum amount of required energy to achieve it. We assume two scenarios with homogeneous arrivals $\lambda_1 = \lambda_2 = 1$ bit/time slot and $\lambda_1 = \lambda_2 = 5$ bits/time slot and a heterogeneous traffic scenario with $\lambda_1 = 1$ and $\lambda_2 = 10$. The average incurred delay for each node is calculated according to (4.28).

The steady state behavior of the sensor queues is depicted in Fig. 4.3 for various arrival rates. As a rule of thumb, the scheduling algorithm allocates the time slots more often to the sensors with higher arrival rates as long as they are in good channel conditions and the allocated PHY

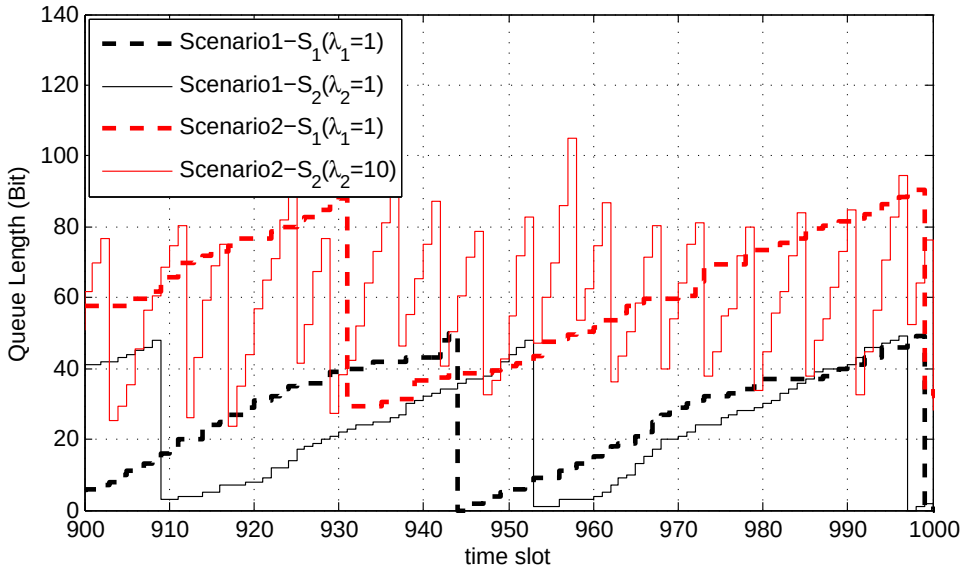
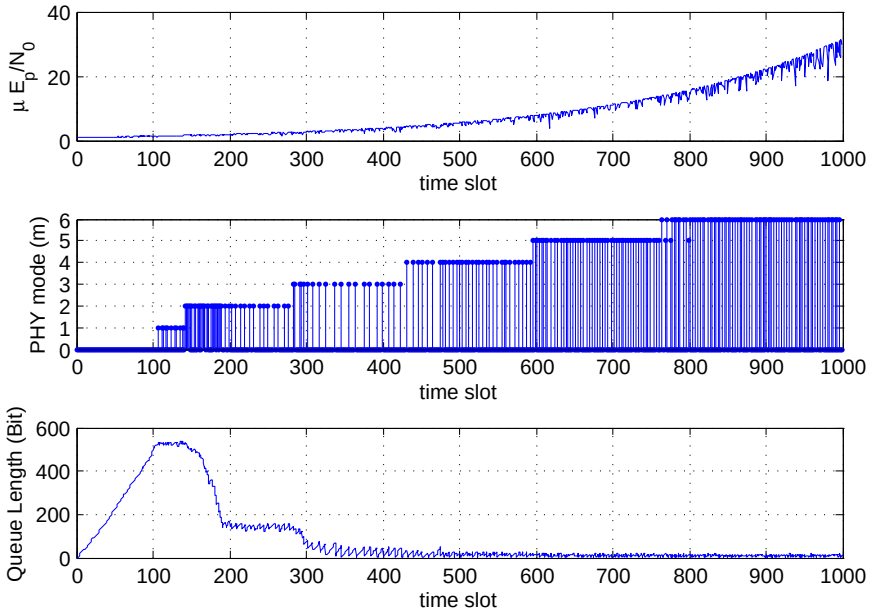
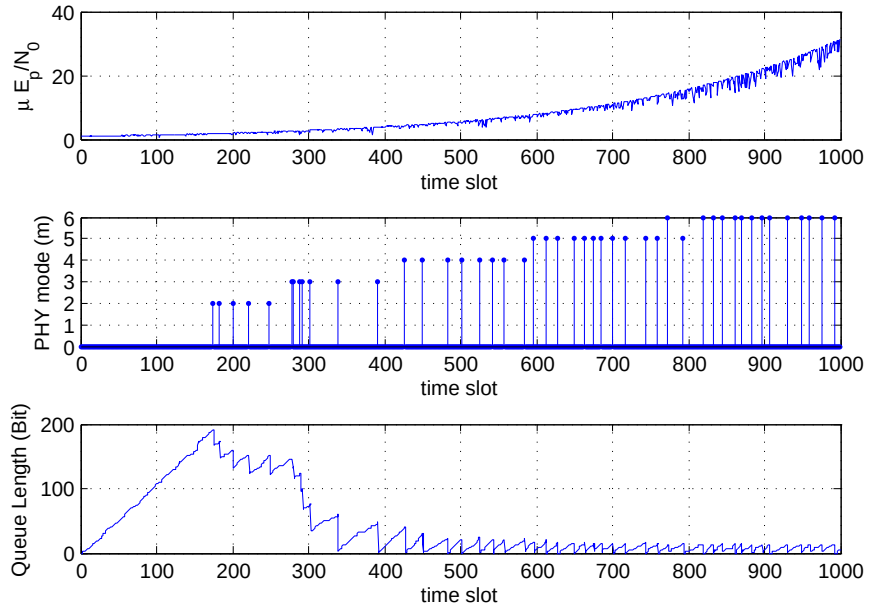


Figure 4.3: Buffer evolution for different traffic arrival rates (equal average delay is assumed for the two scenarios).

mode is more governed by the state of the channel. This idea is validated when we consider the state of the channel and observe the queue length together with the scheduled PHY mode which is illustrated in Fig. 4.4. In these figures we consider two sensors and different arrival rates $\lambda_1 = 5$ and $\lambda_2 = 1$. The value of E_p/N_0 is equal to zero at the first time slot and is increased by 0.015 dB so that we could study its impact on the selected PHY mode and the steady state conditions of the buffers. The state of channel can be assessed by observing the instantaneous value of $\mu E_p/N_0$ at each time slot. It can be seen that when the channel is not in a good condition (low $\mu E_p/N_0$) none of the sensors are selected by the scheduling algorithm and both queue lengths increase linearly (constant arrival and zero service rate). By increasing $\mu E_p/N_0$, the queue backlogs are stabilized after it reaches a specific point and tend to zero as it increases further. It can also be observed that the higher PHY modes are allocated in better channel conditions which is expected a priori. The scheduling algorithm can be tuned by changing the V parameter to approach any desired point in the optimal energy-delay tradeoff curve.



(a) sensor s_1 ($\lambda_1 = 5$)



(b) sensor s_2 ($\lambda_2 = 1$)

Figure 4.4: Buffer evolution and rate scheduling with respect to the instantaneous channel state $\mu E_p/N_0$ (linear scale).

4.6 Conclusion

We studied the energy-delay tradeoffs in IR-UWB WBANs with noncoherent receivers and applied this framework to the IEEE 802.15.6 networks. The stability scheduling policies based on CSI and QLI were addressed and simulated using the standard PDP model of the IEEE 802.15.6 channel model. We also analyzed the dynamic behavior of the rate scheduling policy in different traffic and channel conditions and showed that our rate scheduling policy can respond to these variations in order to preserve the stability with an energy arbitrary close to the minimum possible energy.

Chapter 5

Generalized Nonuniform Sampling of Band-Limited Signals

5.1 Introduction

Optimal design of communication systems has been regarded as a major challenge for many years. This issue still calls for new endeavors with the advent of emerging technologies and applications such as wireless body area networks (WBANs) or, more generally, wireless sensor networks (WSNs). A crucial challenge in the design of these systems is that multiple counteractive constraints are imposed simultaneously. For instance, the sensors in an implant WBANs should have adequately small size and complexity and work with ultra low power to prevent overheating of the proximate tissues, but the channel quality is adverse due to body movements and propagation inconsistencies, and the transmit power should be extremely small to maintain the battery power for a reasonable time. Indeed, a new level of optimal design is necessary in order to meet the constraints. Compression techniques can reduce the total number of bits that the sensor should transmit and save some transmit power as a result. However, a lot of sensor resources due to sampling, quantizing, and processing of the measurements

are actually thrown away in this approach. With the advent of novel techniques such as compressive sensing, one can realize that the current trend in data-gathering applications such as sensor networks is to produce fewer data bits in the first place instead of sampling a big volume of data and then applying data compression which consumes a considerable amount of energy and circuitry [88, 52].

The Sampling theory can provide the analytical means for the equally-spaced discrete-time representation of continuous-time band-limited signals such that the optimal fidelity is achieved after reconstruction. To express such signals in the digital domain, a further level of processing is required that is addressed by the theory of quantization. These two steps are normally dealt with independently, while the same objective is usually sought, i.e. minimizing the reconstruction error. Hence, a theoretical gap can be observed by looking at these two steps together. More explicitly, in today's analog-to-digital-converters (ADCs) the quantization is usually performed when the samples of the continuous-time signal are already taken. The aim and novelty of this work is to seek for more gains in terms of the communications resources by merging these two concepts and introducing more degrees of freedom in the sampling and quantization of the signal. Due to its direct impact on energy and bandwidth utilization, we consider and try to minimize the bit budget, which is the total number of bits required to reconstruct a signal within a given time frame. In the sequel, we explain this generalized nonuniform sampling (GNS) framework. The term "generalized" is used since we assume both the sampling-time and resolutions are nonuniform. This is unlike traditional nonuniform sampling methods that only address the sampling times. We apply the previous uniform and nonuniform sampling times techniques to a test signal for comparison and demonstrate the capabilities of GNS. The proposed sampling architecture reduces the total bit budget of the previous nonuniform sampling time schemes. These schemes can be seen as special cases in the proposed framework where either sampling time or bit depth is fixed and uniform. We propose different schemes under this framework with nonuniform sampling time and bit depth that require a smaller bit budget for

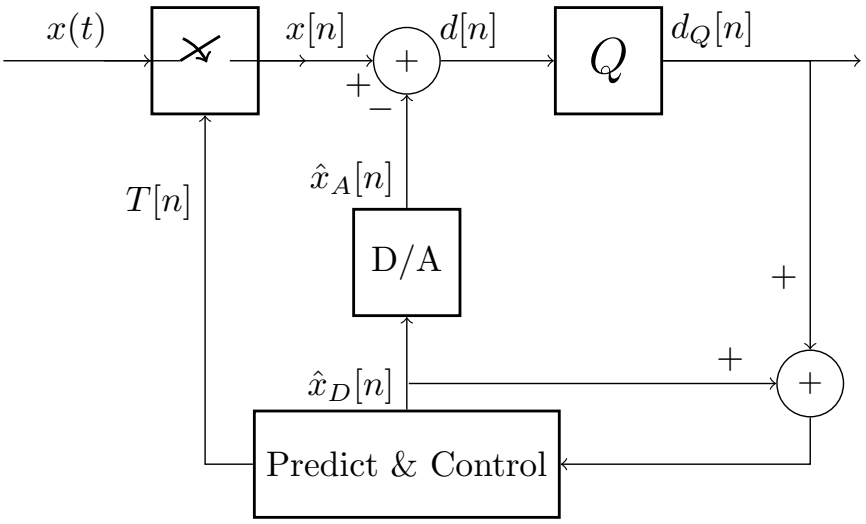


Figure 5.1: Generalized nonuniform sampling architecture. Only $B[n]$ bits are used/generated at the output (depending on the quantizer type).

signal reconstruction. Although our theoretical results apply to arbitrary *compressible*[1] signals, we restrict the evaluation part to ECG signals and leave the other signal models for future studies.

5.2 GNS Framework

In this section we introduce our GNS architecture in which both the sampling times and the bit depths of the sample values are nonuniform. More specifically, just like the conventional nonuniform sampling schemes, we adopt variable sampling times that are controlled by a sampling function f_T. In addition, the number of bits that can represent each sample value (i.e. bit depth) is variable and can be adjusted for each sample according to a function f_B. Consider the analog band-limited signal $x(t)$ having bandwidth W. By definition, the discrete-time samples of this signal are

[1]A compressible signal is one with a form of sparse representation in a particular domain. Arbitrary band-limited signals such as band-limited white noise are not considered. This is motivated by the fact that almost all forms of natural signals of interest in practice belong to this class.

represented by $x[n]$ where the samples are taken at the time stamps $T[n]$. Assume that at time $t = t_i$, the signal's sample is $x[i] = x(t_i)$. Also assume that the number of bits per sample (BPS) for sample $x[i+1]$ is represented by $B[i+1]$ which is given by a function

$$\begin{aligned} B[i+1] =& f_B(t[i-N+1], ..., t[i], \\ & B[i-N+1], ..., B[i], x[i-N+1], ..., x[i]) \end{aligned} \tag{5.1}$$

Generally speaking, the next sample (i.e. $x[i+1]$) should be taken at $t_{i+1} = t_i + T[i]$, where

$$\begin{aligned} T[i] =& f_T(t[i-N+1], ..., t[i], \\ & B[i-N+1], ..., B[i], x[i-N+1], ..., x[i]) \end{aligned} \tag{5.2}$$

is chosen based on a function f_T of the last N sampling times and the corresponding bit depths and sample values as well as the bit depth of the next sample. If evaluated sequentially (rather than jointly), depending on the order, either f_T or f_B can be a function of the other. The sampling functions f_T and f_B will be discussed in more detail in the succeeding sections. We refer to $\sum_i B[i]$ as the required bit budget to reconstruct the signal $x(t)$. Our key objective is to optimize the sampling time and the bit depth of each sample such that the total number of required bits is minimized. We introduce a general sampling architecture illustrated in Fig. 5.1 where a switch is controlled by predictive control logic to sample the signal $x(t)$ at the desired times to produce discrete-time samples $x[n]$. These values should normally be quantized before further processing. In this framework, however, we quantize the difference of the signal and a prediction of the signal that is given by the predictive control logic. The prediction signal is represented by $\hat{x}[n]$ which is fed to a digital-to-analog converter to generate $\hat{x}_A[n]$. The difference signal

$$d[n] = x[n] - \hat{x}_A[n], \tag{5.3}$$

is then quantized and the output signal $d_Q[n]$ is transmitted or stored. Throughout this section, we assume that uniform quantization with a fixed

step size Δ is performed wherever data is quantized. For the sake of simplicity, we first proceed with some special cases of our proposed system. Our main concern here is how to design the block marked "predict and control" where the sampling times $T[n]$ as well as the estimate of the next sample are generated. A tradeoff between the number of samples and the number of bits per sample can be established by adopting different algorithms. By intuition, we expect that a longer time between two consecutive samples makes the expected number of required bits for the next sample larger.

5.2.1 DPCM system

It can be seen that a special case of this framework is when the samples are taken uniformly and a fixed number of bits per sample are allocated, i.e.

$$T[n] = T_s = \frac{1}{2W}, \ \forall n \tag{5.4}$$

$$B[n] = B, \ \forall n \tag{5.5}$$

where W is the signal's bandwidth, T_s is the sampling time period, and B is the bit depth of each sample, is chosen based on the signal's dynamic range. This case corresponds to the well known differential pulse code modulation (DPCM) which is prevalent in voice encoding. The gain in DPCM systems is achieved by reducing the variance and dynamic range of the signal since the difference of the prediction is usually small. Generally, a linear prediction is used in different applications to represent an estimate for the next sample. More specifically,

$$\hat{x}_D[i] = \sum_{j=1}^{m} w_j x[i-j] \tag{5.6}$$

where m is the prediction order and $\{w_j\}, j \in [1, m]$ are the prediction weights. The optimal weights for stationary signals are given by solving the Yule-Walker equations. A simple case is when $m = w_1 = 1$ which is called first-order prediction or delta modulation. In Fig. 5.2, the probability mass function (PMF) of the sample values of an ECG signal[2] and its first-

[2]We use 10 seconds of lead-I ECG of record 3 from T-Wave Alternans Challenge Database which includes 12-lead ECG records sampled at 500 Hz with 16-bit resolution over a ±32 mV range [89]. The signal amplitude is normalized to one.

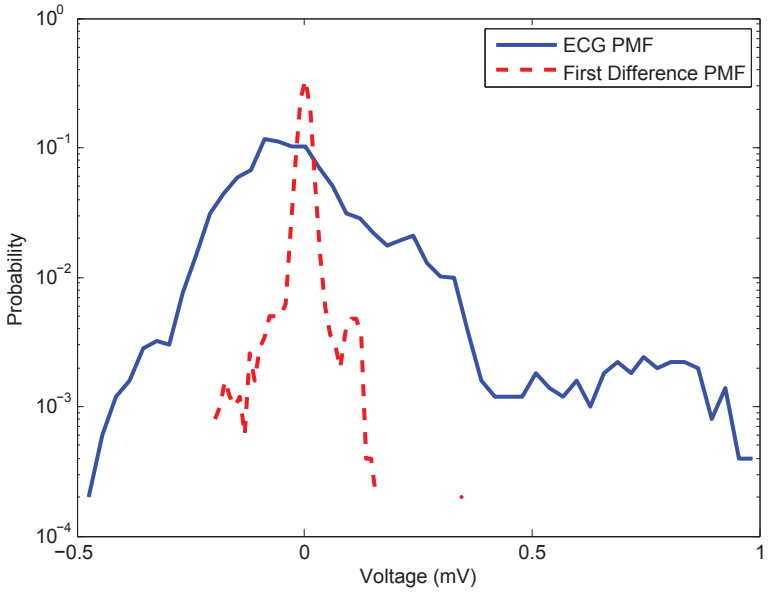


Figure 5.2: The probability mass function (PMF) of the sample values of an ECG signal and its first order prediction (first-difference). Note that the dynamic range and variance of the first difference are much less than the signal itself. The variance of an ECG signal and its first difference are 3.2×10^{-2} and 8.2×10^{-4} respectively.

order prediction are compared. The variance and dynamic range of the first difference are much smaller than the original ECG signal and therefore it can be represented with fewer bits.

The total required number of bits may be reduced further using adaptive schemes. In adaptive differential pulse code modulation (ADPCM), for instance, the quantization step is adaptively adjusted for each sample. For instance, in [90] the step size is multiplied by a time-invariant function of the magnitude of the codeword corresponding to the previous sample.

5.2.2 DPCM with Adaptive Sampling Time

Consider the DPCM system with variable sampling times. A simple case is when the last sample is used to produce an estimate for the next sample. This approach is illustrated in Fig. 5.3. In this figure, the delay element D reproduces $\hat{x}_D[n] = x_D[n-1]$ as the estimated value for the next sample. There are several approaches to control the sampling times.

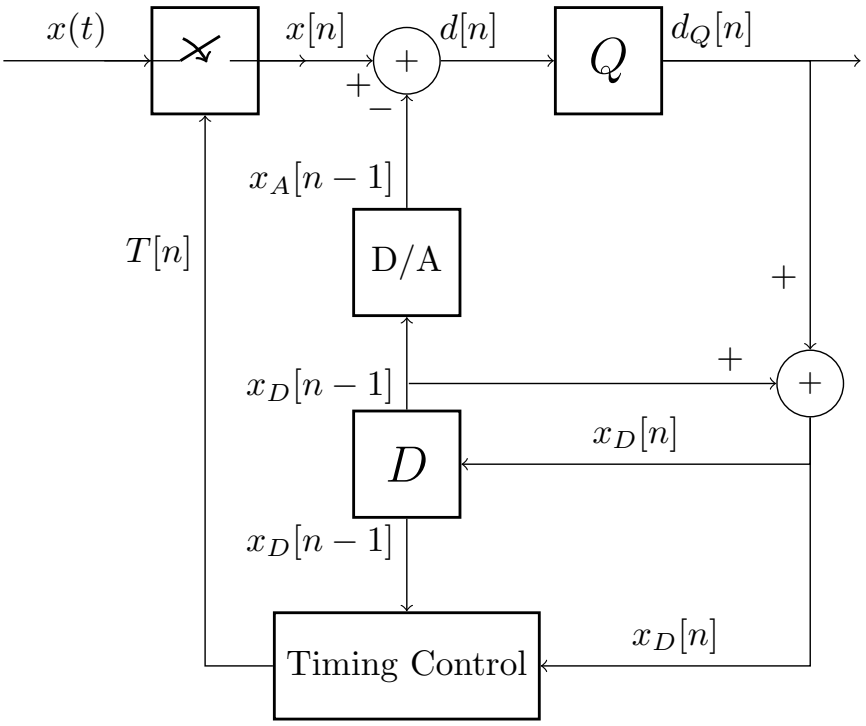


Figure 5.3: Variable sampling time with first-order prediction, also known as delta modulation.

A straightforward method is to observe one of the derivatives of the signal (e.g. an approximate of the first derivative such as $\omega[n] = d_Q[n]/T[n-1]$) and update the sampling times adaptively [53]. More explicitly, we can set

$$T[i] = T[i-1] - \epsilon \Psi_{(\omega_1,\omega_2)}(\omega[i]), \tag{5.7}$$

$$B[n] = B, \ \forall n \tag{5.8}$$

where ϵ is a constant time and

$$\Psi_{(\omega_1,\omega_2)}(\omega) = \begin{cases} +1 & \omega > \omega_2 \\ -1, & \omega < \omega_1 \\ 0, & \text{otherwise} \end{cases} \tag{5.9}$$

is defined for the given thresholds ω_1 and ω_2. The sampling time $T[i]$ is also hard-limited to maintain the result in the interval $[T_{min}, T_{max}]$. The advantage is that the time stamps $T[n]$ can be recovered locally at the

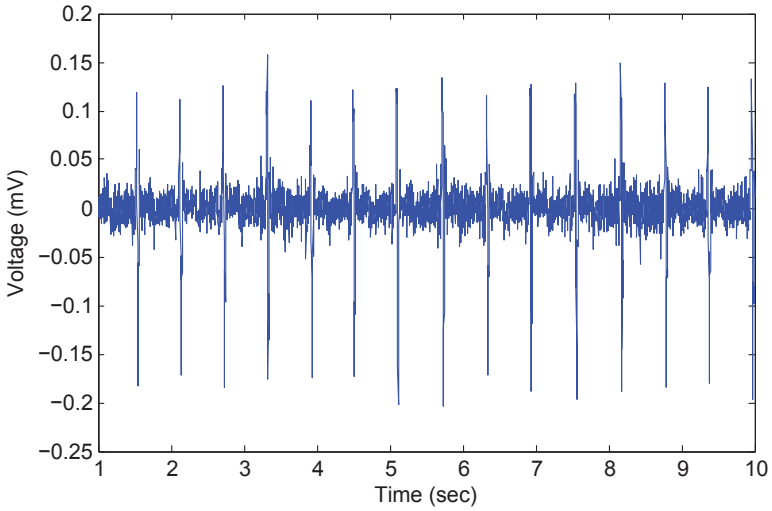


Figure 5.4: First difference of an ECG signal. Both the variance and mean are small, but more significant peaks appear occasionally.

decoder from the sample value, hence they should not be transmitted or preserved. In the sequel we keep using this nonuniform sampling time approach but suggest different schemes to extend the nonuniformity to the bit depth of the samples.

5.2.3 In-Message Signaling of the Bit Depths

Assume that the set of differential symbols $d_Q[n]$ are represented (for storage or communication) in the form of a bit stream. Obviously, if the symbol lengths are nonuniform, then some form of punctuation is required to separate two different symbols. In this regard, one solution is to stuff the length of each symbol before the actual symbol in form of a fixed-size header segment. In this way, every symbol is decoded by first looking at the length segment, and then the required number of bits are accumulated. The actual required number of bits per symbol is given by

$$b[n] = \left\lceil \log_2 \left(\left\lfloor \frac{|d_Q[n]|}{\Delta} \right\rfloor + 1 \right) \right\rceil + 1, \tag{5.10}$$

where $\lceil . \rceil$ and $\lfloor . \rfloor$ are ceiling and floor operators. One bit is allocated to represent the sign and the rest is for the amplitude. Fig. 5.5 illustrates this

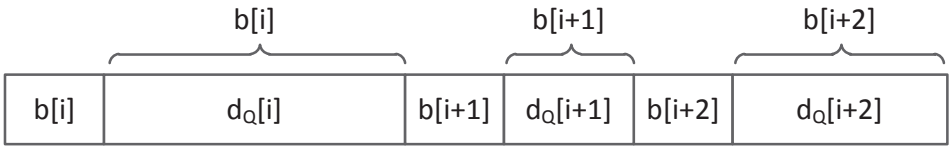


Figure 5.5: In-message signaling of variable bit depths

approach.

5.2.4 GNS for Band-limited Signals

A key point about band-limited signals is that the maximum variation of such signals at each time is upper bounded. This can help to allocate in advance the required number of bits for quantization of the next sample.

A band-limited signal can either be of finite power, which refers to a form of periodic function, or finite energy, that belongs to transient signals. These bounds are investigated in [91] and [92] and are reiterated here for clarity. Assume a periodic band-limited signal with period T_0 whose power spectral density is zero for $|f| > W = L/T_0$. It can be represented by a Fourier series as

$$x(t) = \sum_{k=-L}^{L} c_k e^{j2\pi kt/T_0}, \tag{5.11}$$

where $\{c_k\}$, $k \in [-L, L]$ are the Fourier coefficients. Using the Cauchy-Schwarz inequality and Parseval's Theorem the bound on $x(t)$ and its first derivative $x'(t)$ are given by

$$lCl|x(t)| \leq \sqrt{P(2L+1)}, \tag{5.12}$$

$$|x'(t)| \leq \frac{2\pi}{T_0}\sqrt{\frac{PL(L+1)(2L+1)}{3}}, \tag{5.13}$$

where $P = \sum_{k=-L}^{L} |c_k|^2$ is the signal's average power. We can also calculate the signal's variation within a given time horizon τ using the signal's Fourier series and the shift property as follows

$$g(\tau) = x(t+\tau) - x(t) = \sum_{k=-L}^{L} c_k(e^{j2k\pi\tau/T_0} - 1)e^{j2k\pi t/T_0}. \tag{5.14}$$

Applying the Cauchy-Schwarz inequality and Parseval's theorem yields

$$|g(\tau)|^2 \leq \sum_{k=-L}^{L} |c_k|^2 \cdot 4 \sum_{k=-L}^{L} \sin^2(\frac{k\pi\tau}{T_0}), \tag{5.15}$$

and noting the fact that $\sin^2(\phi) \leq \phi^2$ we conclude that

$$|g(\tau)| \leq \frac{2\pi\tau}{T_0}\sqrt{\frac{PL(L+1)(2L+1)}{3}} = G(\tau), \tag{5.16}$$

where $G(\tau)$ is the peak absolute variation of the signal in time τ. Alternatively, for a band-limited signal with finite energy $E = \int_{-\infty}^{+\infty} |x(t)|^2 \mathrm{d}t$, we can establish the following bounds for the signal, its derivative, and the peak variation for a given time horizon τ [91]

$$|x(t)| \leq \sqrt{2EW}, \tag{5.17}$$

$$|x'(t)| \leq 2\pi W\sqrt{\frac{2EW}{3}}, \tag{5.18}$$

$$|g(\tau)| \leq \sqrt{E\left(4W - \frac{2}{\pi\tau}\sin(2\pi W\tau)\right)} = G(\tau). \tag{5.19}$$

Assume that the next sampling time $T[i]$ is chosen, for example by (5.7). The range of the next sample can therefore be fixed at twice the above bounds to account for both negative and positive variations around the current sample value. After sampling, the difference is divided by a fix quantization step size Δ in order to produce the corresponding codeword. We can fix the bit depth of the next sample by

$$B[i+1] = f_B(T[i]) = \left\lceil \log_2\left(\left\lfloor \frac{G(T[i])}{\Delta} \right\rfloor + 1\right)\right\rceil + 1. \tag{5.20}$$

In this expression, one bit is reserved for the sign of the difference and another bit inside the $\log_2$ function is to take into account zero value. Unfortunately, discussion on whether these bounds are tight enough to work effectively is signal-dependent. Indeed, the spectral properties of the signal and how its energy is distributed throughout the spectrum are the influential factor.

In order to benefit from the above bounds, some form of prior knowledge of the signal such as the period, average power, or energy is necessary depending on the signal's type. For instance, the ECG signal can be approximated by a periodic signal. However, the sampling process might be initiated when no prior information is available. In this case, one can initialize the bit depth from a fixed value large enough to cover the dynamic range of the signal. The required statistics can be collected dynamically. For example, the QRS part of a typical ECG signal that corresponds to a significant peak can be detected by level-crossing detection with appropriate thresholding. The time durations between the peaks can then approximate the signal's period. After that, one can approximate and update the required bit depth of the succeeding samples more accurately by using the statistics harvested from the signal.

Another approach is to bound the signal's sample at a time in the future only by its current first few derivatives. Consider the Taylor series expansion of $x(t+\tau)$ around t

$$x(t+\tau) = x(t) + \frac{x'(t)}{1!}\tau + ... + \frac{x^{(k)}(t)}{k!}\tau^k + ..., \tag{5.21}$$

where $x^{(k)}(t)$ is the kth derivative of $x(t)$. If we use the first k terms for approximation, then the truncation error $R_k(t+\tau)$ is bounded by

$$R_k(t+\tau) \leq \frac{M_{k+1}}{(k+1)!}\tau^{k+1}, \tag{5.22}$$

where

$$M_{k+1} = \max_{\zeta \in [t,t+\tau]} x^{(k+1)}(\zeta), \tag{5.23}$$

is the peak value of the $(k+1)$th derivative in the assumed interval. Therefore, new bounds on maximum variation of the signal can be established by truncating the series up to any term. For instance, the following bound is given when the first two terms are used:

$$|x(t+\tau) - x(t)| \leq |x'(t)\tau + \frac{M_2\tau^2}{2}| = G(\tau). \tag{5.24}$$

The peak value of the derivatives can either be approximated locally similarly to (5.13) and (5.18) for finite power and energy signals respectively, or

Table 5.1: Comparison of the simulation results

	Sampling time	bit depth	Samples	BPS	Total bits	Reconstruction MSE
PCM	Uniform	Uniform	5000	12	60000	3.1886×10^{-7}
Down-sampled PCM	Uniform	Uniform	1000	12	12000	3.0336×10^{-4}
First-order DPCM	Uniform	Uniform	5000	10	50000	4.4984×10^{-7}
Adaptive Sampling Time PCM [53]	Nonuniform	Uniform	966	12	11592	2.1409×10^{-3}
Adaptive Sampling Time First-order DPCM	Nonuniform	Uniform	965	11	10615	4.7972×10^{-3}
In-Message Signaling	Nonuniform	Nonuniform	965	9.10	8784	4.7972×10^{-3}
Generalized Nonuniform Sampling (Band-limited)	Nonuniform	Nonuniform	965	8.89	8576	4.7972×10^{-3}
Generalized Nonuniform Sampling (VLC)	Nonuniform	Nonuniform	965	7.51	7243	4.7972×10^{-3}

assumed known a priori when dealing with signals such as an ECG. In the latter case, global maximum values can be set based on available datasets of realistic signals. Based on the sampling time and using (5.20), the appropriate number of bits can be allocated for the next sample. A heuristic approach that has been applied and will be discussed in our simulations is to choose the value of M_2 such that the minimum total number of bits is achieved.

5.2.5 GNS with Variable Length Coding

One solution to reduce the total number of bits is to allocate shorter codewords to more probable symbols and longer codewords to rare symbols. When differential encoding is used, the mean and variance of the differential signal is usually small and varies depending on the performance of the prediction. For example, as illustrated in Fig. 5.4, the first difference of an ECG signal contains occasional peaks that increase the required dynamic range of the quantizer significantly. Since the peaks are less likely, a variable length coding (VLC) scheme such as Huffman coding can reduce the average number of bits per symbol and hence the total required number of bits. In this case, the sampling times are selected, for example similarly to (5.7) but, instead of using a fixed length, a dictionary is agreed between the encoder and decoder that has symbols with variable lengths. More specifically, the function $B[n] = f_B(x[n])$ has a distinct value for each sample

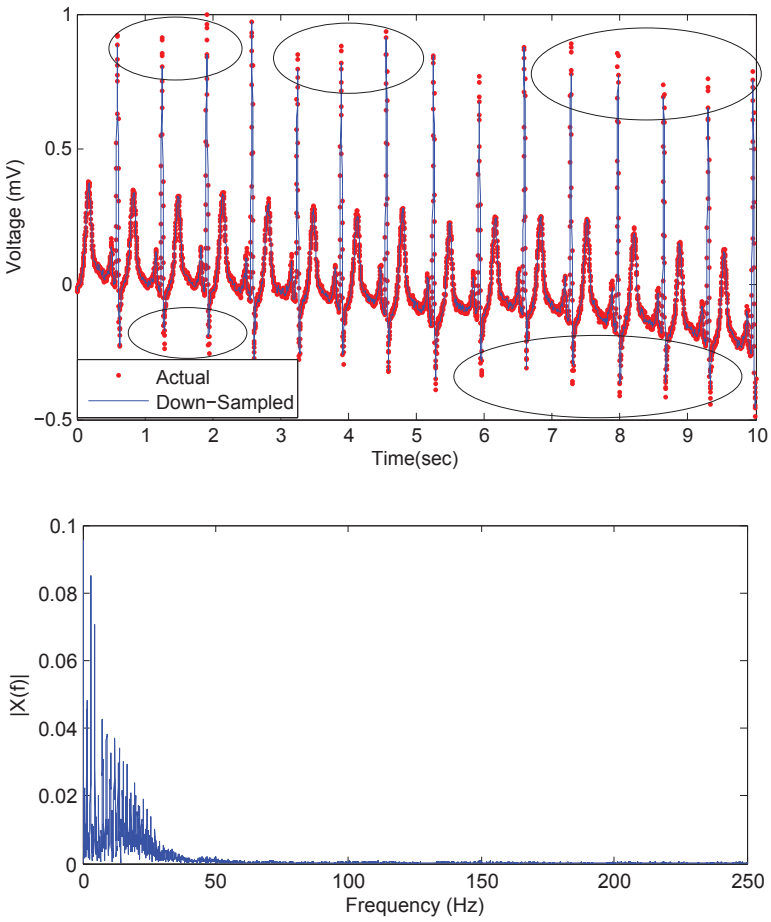


Figure 5.6: The actual and the reconstructed waveforms corresponding to the down-sampled (by a factor of 5) signal (top). The peaks in the reconstructed signal are shrunk as a consequence of aliasing. The Fourier transform indicates that the signal bandwidth is approximately 50 Hz (bottom).

which is equal to the length of the corresponding codeword in a predefined dictionary. This also resolves the punctuation problem when dealing with nonuniform bit depths since the prefix codes are inherently separable.

5.3 Simulation Results and Discussion

In this section we evaluate the performance of the presented schemes using a test ECG signal, originally sampled uniformly at a sampling frequency of

500 Hz with a resolution of 16 bits-per-sample, and compare the results. To reduce the number of samples, the first and most naive approach is down-sampling. Based on the spectrum of the signal presented in Fig. 5.6, the signal is band-limited with $W = 50$ Hz. Therefore, we down-sample the signal by a factor of 5 and reduce the sampling frequency to 100 Hz. As illustrated in the figure, the peaks in the reconstructed signal are shrunk as a consequence of aliasing effects caused by minor frequencies above 50 Hz.

In Table 5.1 the outcome of different simulations on the test signal are presented, where a PCM signal (with uniform sampling time and bit depth) is compared with the signal reconstructed using other approaches. The first column identifies the specific sampling schemes that have been applied. The last four rows, which are distinguished using boldface letters, are the present work's contributions. Assume that the total number of samples is J and the reconstructed signal is x_R. The reconstruction distortion in terms of mean square error (MSE) is

$$\mathsf{MSE} = \frac{1}{J}\sum_{j=1}^{J}\left(x[j] - x_R[j]\right)^2 \tag{5.25}$$

which is given for each scheme. The first row corresponds to a conventional PCM system with uniform sampling time equal to the Nyquist rate and a uniform bit depth proportional to the signal's dynamic range. Note that the reconstruction MSE is due to quantization error in this case. The second row represents the same signal but down-sampled by a factor of 5. While a reasonable MSE is achieved and the total number of bits is reduced by up to 80%, as we saw earlier aliasing effects appear and the peaks shrink significantly. When DPCM is used, the required bit depth is reduced from 12 to 10 due to a reduction in the signal's dynamic range which saves 17% of the total number of bits. Using an adaptive sampling time with uniform bit depth an 81% reduction in the total required number of bits is achieved compared to the PCM system. In order to calculate the MSE for the signal with nonuniform sampling time, a re-sampled version uniformly sampled at $1/T_s$ is generated in which the sequential samples are connected via a line. Explicitly, when $T[i] > T_s$ a linear regression is applied in order to

approximate the sample values between $x[i]$ and $x[i+1]$. This technique had also been applied to reconstruct from the down-sampled signal. The simulation parameters are $T_{min} = 2T_s$, $T_{max} = 6T_s$ and $\epsilon = 2T_s$, with $T_s = 1/500$.

With a nonuniform sampling time and DPCM encoding, the total bit reduction approaches 82%. However, we can observe that the two schemes based on VLC and band-limited prediction can offer an 88% and 84% reduction respectively, compared to the conventional PCM system. While the reconstruction distortion is still negligible, this reduction is equal to 38% and 17%, compared to the case where only the sampling times are nonuniform, which is a significant improvement. A difference in the MSE of the nonuniform sampling time scheme with uniform and nonuniform bit depths can be observed in the table. This is because the quantized version of the symbols is directly used for decoding in the uniform approach. On the other hand, the nonuniform approaches use quantized versions of the differences. Hence, the amount of quantization error is twice on average. Also note that, for a nonuniform bit depth, all MSE values are equal since all use the same quantized values, except the symbols are represented with a different number of bits. Since $b[n] \leq B[n]$, $\forall n$, no distortion caused by nonuniformity of the bit depths is introduced regardless of the bit allocation scheme.

To predict the maximum absolute variations for a given time horizon, we exploited the two inequalities (5.16) and (5.24). However, (5.16) is not a tight bound for the ECG signal and the total number of bits was not small enough. Conversely, using (5.24) and approximating the first derivative of the signal as

$$x'[n] \approx \omega[n] = d_Q[n]/T[n-1], \tag{5.26}$$

the overall reduction in the number of bits per sample is overwhelming. The bit depth of the first sample should be fixed and known at both sides to initiate the algorithm. The main concern in the band-limited prediction-based approach is that the actual number of required bits represented by $b[n]$ should never exceed the allocated bit depths, $B[n]$. In case of such event, not only is the symbol at which this occurs decoded in error, but

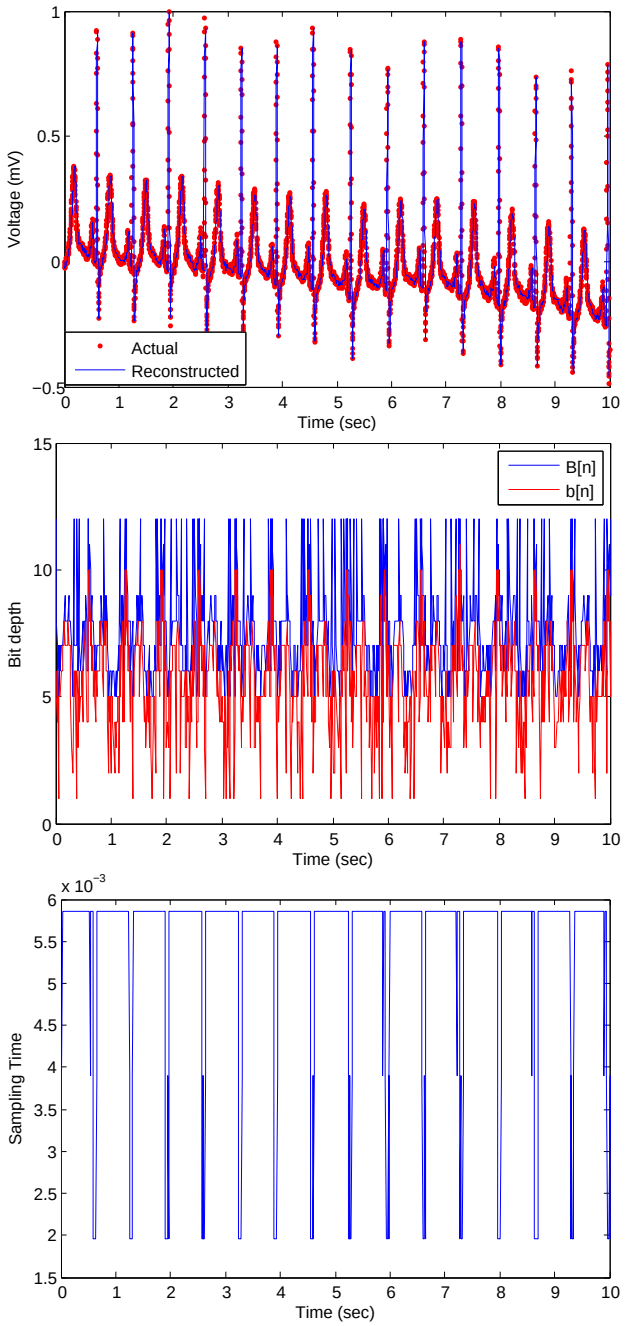


Figure 5.7: The reconstructed signal with nonuniform sampling time and bit depth on top. Variations of bit depth and sampling times are depicted in the middle and bottom figures.

also the rest of the symbols are corrupted consequently, since the wrong bits will be selected from the bit-stream. Assume that a fixed value is selected for M_2. To prevent length overflow, this parameter should be large enough. This however, ends up with all symbols becoming significantly longer, which is not desirable. To solve this problem, we allocate one more bit to all symbols, that maintains the state of each symbol in terms of the length-overflow event. At the encoder side, if a length-overflow occurs at a symbol, the bit depth corresponding to that symbol is overwritten with a fixed value equal to the maximum possible bit depth. At the same time, the length overflow indicator bit is set in order to let the decoder know to correctly overwrite the corresponding bit depth to its maximum. Selecting an appropriate value for M_2 can lead to the minimum total number of bits since very small M_2 values increase the total number of bits just like very large values. A too small M_2 will cause a lot of length overflow events, and for each event the allocated bit depth is overwritten by the maximum bit depth.

A snapshot of the test ECG signal along with the reconstructed signal x_R, the variations of bit depth and the sampling-time functions, $B[n] = f_B$ and $T[n] = f_T$, respectively, are pictured in Fig. 5.7. Note that the sampling times and bit depths are first reconstructed locally at the decoder and then the estimated sample is interpolated. The smaller bit depths are associated with smaller changes between the two given sequential samples which can be represented with fewer bits. It can be verified from Fig. 5.3 that $B[n] \geq b[n]$, $\forall n$, so no loss of information can occur due to an overflow.

5.4 Conclusion

With the objective of reducing the total number of bits to reconstruct a continuous band-limited signal, we investigated and simulated different schemes. Proposing a new sampling architecture, we showed that it can represent a conventional DPCM system when both sampling times and bit depths are fixed and uniform, and the total number of bits per sample is reduced due to a decrease in the signal's dynamic range. In the next steps, we

used nonuniform sampling times with uniform and nonuniform bit depths to further reduce the number of samples, and bits per sample respectively. Only 12% of the total raw bits in a normal PCM system are required in the generalized non-uniform scheme, according to the simulations. Only add and multiply operations, linear with the order of prediction, are required for both sampling and reconstruction. This is much lower than the computational complexity of the existing techniques used in compressive sensing, that rely on solving an optimization problem, or many other known compression schemes.

Chapter 6

Partial Packet Recovery for Efficient Transmission

6.1 Introduction

Packet errors degrade the performance of wireless communication systems in terms of energy efficiency and throughput. In a wireless packet-based system a packet is lost or *erased* due to errors, when the packet checksums are not consistent with the payload. When this happens, normally, the receiver requests a re-transmission by signaling a negative acknowledgment message to the transmitting node. An extra amount of energy and bandwidth is required for re-transmissions. Nonetheless, they are also often corrupted by errors. Therefore, error correction that leverages an additional protective or reconstructive procedure per packet is used.

Obviously, the overall efficiency of the transmission protocol is governed by how and to what extent random errors can be corrected. Traditionally, this issue is addressed either proactively, i.e. by forward error correction (FEC) or reactively by automatic repeat request (ARQ). Besides their advantages, there are some limitations associated with each of these strategies. For example, FEC introduces a constant overhead and has a limited correction capability that results in the threshold effect [93]. Also, ARQ performs poorly in severe channel conditions and cannot achieve high relia-

bility. Hence, FEC and ARQ have been combined to produce more flexible methods such as hybrid ARQ and rateless coding. Another viable technique that has been developed to improve resource utilization in wireless networks is random linear coding (RLC) [94, 95], which is the baseline technique in this chapter. When RLC is performed across packets, blocks of message packets are combined randomly to construct new coded packets to represent a rateless code. A receiver requires enough error-free coded packets to reconstruct the original packets. However, there are usually several packets that are *partially* corrupted by channel errors and are discarded. This can constitute typically up to 20% of total packets depending on the deployment [75]. Normally, all of the packets are encoded by an error detection code such as the cyclic redundancy check (CRC). In this way, upon reception of a packet at the receiver, the integrity of the packet can be verified by looking at the check sequence. Packets with inconsistent checksums contain bit errors and are referred to as *partial packets*, so the receiver can collect all the packets and classify them as either partial or valid. Certainly, discarding partial packets is sub-optimum since some of their segments might actually be error-free. It is pointed out in [96] that up to 95% of the segments in a partially corrupted packet can be correct.

In this chapter we address partial packet recovery (PPR) in random linear coded systems. In contrast to hybrid-ARQ approaches, PPR schemes attempt to repair partial packets based on the existing knowledge at the receiver instead of requesting more redundancy from the transmitter (though there is a form of feedback in acknowledgment messages). This can alleviate computational complexity and offer better throughput and energy efficiency at the transmitter nodes by reducing the total number of transmitted packets per message unit. It is shown [97] that, with one or more sources, linear coding is enough to achieve the upper performance bound in multicast networks. However, linear codes are not sufficient for arbitrary networks [98]. From this point of view, PPR can be interpreted as a nonlinear scheme that complements RLC.

Although RLC is a hybrid coding/ARQ technique it has architectural and functional differences with H-ARQ. As is pointed out in [95], H-ARQ

utilizes the soft information combined in previous transmissions of a packet through different puncturing of a channel code applied to the same packet. Hence, it would face the same challenges of ARQ in broadcast settings. Most notably, retransmission of a packet only benefits the receivers missing that packet [95]. However, RLC exploits coding across multiple packets so a larger number of receivers can benefit by transmission of a RLC packet. It also benefits from the acknowledgment messages more efficiently by including the number of missing/collected packets rather than simply stating successful reception of a packet. The performance of RLC is compared with chase-combining H-ARQ in [99]. In [100], raptor codes are compared with RLC for multimedia broadcast/multicast service. An obvious advantage of RLC over raptor codes is re-coding and the subsequent performance gains. It is shown that RLC offers up to 5.9 times the gain in throughput and 5.5 times the reduction in the overall delay compared to the state-of-the-art H-ARQ mechanism used in WiMAX. A similar comparison is made in [101], assuming single and multihop transmissions, that advocates the advantage of RLC over H-ARQ. We will discuss in the next sections applications where such properties are of interest. Our key assumption for designing this scheme is that channel errors constitute a *sparse process*, i.e, we assume that the number of channel errors is much less than the total size of the packet and the number of partial packets is relatively smaller than the number of correct ones. This also involves burst errors as long as the number of correct packets is higher than the number of partial packets. Using algebraic properties of the coding matrix, the receiver forms a special sensing matrix that can cancel out the message data and provide compressive sensing (CS) of the errors. We use concepts from the Coding theory, Markov chain theory, compressive sensing and optimization to model the system. Since packet elements are selected from a finite field, a special case of a compressive sensing problem is formulated that can be solved by algebraic coding theory with less complexity. Explicitly, as opposed to the normal methodology in the CS frameworks, we show theoretically and by simulation that the complexity of solving the original l_0 optimization problem is less than that of the convex relaxed version. The problem

of compressive sensing over a finite alphabet is addressed in [102]. Using an algebraic framework, it is shown that the complexity of finite-field CS is reduced such that the original l_0 minimization can be solved in polynomial time and with less resources than with the convex relaxation approach. Explicitly, while the compressive sensing of a real-valued m-sparse signal requires $O(m\log(n/m))$ measurements, the counterpart finite alphabet problem with alphabet size q requires $\Theta(m\lceil\log_q n\rceil)$ measurements, where n is the length of the original vector. Consequently, one can solve the exact l_0 minimization problem in polynomial time instead of performing convex relaxation and solving the l_1 minimization. Also, unlike in conventional CS problems where the measurements are in form of a vector, in our PPR problem they are in matrix form. We provide an efficient solution that handles this property in a tractable manner and demonstrate its performance by simulation. We propose to exploit partial packets by solving a set of standard sparse recovery (SR) problems. An error correction scheme is proposed in [103] which is based on sparse recovery of errors by solving a linear program at the receiver. We apply this idea to an RLC framework to recover partial packets and boost the performance of RLC with much fewer transmitted packets. Using concepts from coding theory, we establish a relationship between the RLC encoding matrix and the sensing matrix at the receiver. It is shown that the performance of RLC can be remarkably improved, which can significantly reduce the average required number of encoded packets. Therefore, it can improve energy and bandwidth efficiency at the transmit node. The PPR-enabled receiver is modeled by an absorbing Markov chain and the expected number of packets for successful decoding is analytically derived. Markov chain models are the main tools to model and analyze these systems and have been used in CDMA systems with ARQ [104] and delay [105, 106, 107, 108] and throughput [95, 108] analysis in RLC systems. The advantages of our proposed scheme can be enumerated as follows:

- No cross-layer information or bit-level soft information (i.e. the corresponding reliability measure of the decoded symbols) is required. In

the traditional layered approach exchange of cross-layer information is prohibited. Hence, our scheme can be implemented in systems based on layered design while operating completely transparently from other layers[1].

- No form of checksum, preamble, or postamble is required for PPR. The algorithm is completely implemented at the receiver side, relies on data processing techniques, and performs transparently to the transmitter. Therefore, no loss of bandwidth or extra energy consumption at the transmitter node is imposed. This can be of major interest in several applications as will be covered in this chapter.

- Nevertheless, any desired tradeoff between the computational complexity at the receiver and the energy and bandwidth efficiency at the transmitter can be achieved. This can be done simply by changing a few design parameters such as the FEC coding rate at the transmitter and the worst-case channel error probability under which reliability should be achieved.

The proposed scheme enables point-to-point gain in terms of transmission efficiency in RLC systems, significantly improves the transmission efficiency in multicast, and is implemented at the receiver side only (no change at the transmitter).

6.1.1 Organization and Notations

This chapter is organized as follows. First a preliminary introduction to RLC, CS and SR is presented in Section 6.2 along with the system model. In Section 6.3 we explain our proposed scheme and discuss its performance. The characteristics of the PPR problem and the corresponding error recovery algorithms are considered and elaborated in Section 6.4. The system is modeled by a Markov chain and the number of packets required for decoding of RLC and solving the PPR problem are also discussed. Section 6.5

[1]RLC is traditionally performed at the network layer (so is the proposed PPR scheme).

presents simulation results where the performance of the proposed scheme is compared with conventional RLC and the accuracy of the derived Markov chain model is verified. Finally, concluding remarks are given in Section 6.6.

We use boldface capital letters for matrices and boldface lower case letters for vectors and distinguish column vectors from row vectors by using a prime ($'$) sign.

6.2 Preliminaries and System Model

In this section we present the system model and briefly explain the concept of RLC and the terminology that is used in the CS literature. We assume a packet-based wireless system where the transmit and receive nodes can communicate packets of length L. This framework applies to both point-to-point and multicast channels. We will quantify our scheme in both point-to-point and multicast channels in the simulations. Packet elements are referred to as symbols and are drawn from a Galois Field of order q denoted by $GF(q)$. Hence, each symbol can be represented by $\log_2 q$ bits and all operations are performed in $GF(q)$.

In typical wireless systems, FEC is performed at the physical layer, where the packet is divided into several segments and each segment is encoded separately by the channel code. In a partial packet, there may exist several correct segments that can actually be used to recover the whole packet. Note that our scheme is performed independently above the physical layer (FEC) and exploits all packets (valid or partially corrupted) and all packet segments for recovery.

6.2.1 Random Linear Coding

We first consider a point-to-point system and describe the procedures at both the transmitter and receiver as depicted in Fig. 6.1. Then we assume a multicast system where a transmitter sends coded packets to more than one receiver. In both cases we assume that two-way communications take

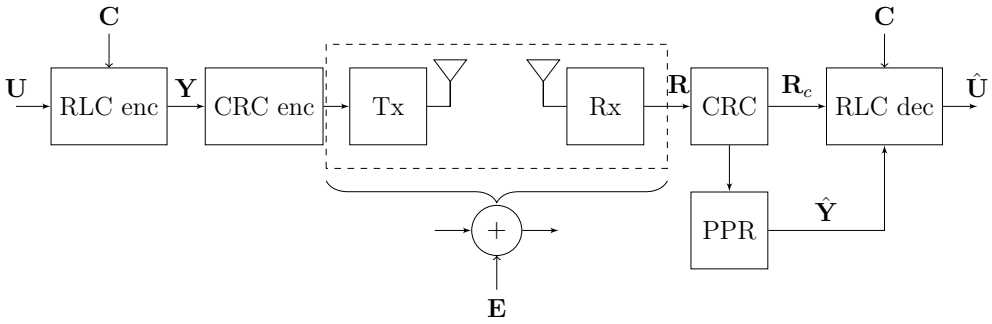


Figure 6.1: Block diagram of the RLC system with PPR.

place based on time division duplexing. Feedback channels are necessary for acknowledgment messages where the receiver(s) can signal the number of coded packets they have collected. The transmitter node is composed of an RLC encoder and a wireless transmitter that caries out FEC encoding, modulation, and other PHY tasks. The receiver node is equipped with a wireless receiver and a conventional RLC decoder plus a PPR block that recovers the partially corrupted RLC packets. The transmitter can perform RLC on the groups of G packets which is commonly referred to as the generation size of the rateless code. At each time instance $t = i$ of this procedure, G packets of length L are combined using a random coding vector $\mathbf{c}_i$ containing G random elements $(c_{i,1}, ..., c_{i,G})$ to form a new packet of the same length, where $c_{i,j} \in GF(q), \forall i, j$. Denoting the coded packet by $\mathbf{y}_i$, the encoding can be represented by

$$\mathbf{y}_{i(1\times L)} = \mathbf{c}_{i(1\times G)}\mathbf{U}_{(G\times L)}, \tag{6.1}$$

where the vector and matrix sizes are given in the parenthesis and the message block is denoted by

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_G \end{bmatrix}, \tag{6.2}$$

in which $\mathbf{u}_i \, \forall i \in [1, G]$ is the row-vector representation of the ith message packet. For convenience, we can denote the process to generate T encoded

packets in the matrix form:

$$\mathbf{Y}_{(T\times L)} = \mathbf{C}_{(T\times G)}\mathbf{U}_{(G\times L)}. \tag{6.3}$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_T \end{bmatrix}. \tag{6.4}$$

For simplicity, we assume $GF(2)$ throughout this chapter, i.e., $q = 2$ and all coding vectors and operations are in the binary field. Ideally, the elements $c_{i,j} \in [0, 1]$ are uniformly independent and identically distributed (iid). They can however be generated by a pseudo random generator using a specific seed value that can be conveyed to the receiver side in the packet headers for local recovery and synchronization [99]. Therefore, we assume that both the transmitter and receiver have perfect knowledge of the encoding vectors $\{\mathbf{c}_i\}$. The encoded packets are then transmitted on a wireless channel. Depending on the channel quality, some parts of the received packets are corrupted by channel errors. This can be represented by

$$\mathbf{R} = \mathbf{Y} + \mathbf{E} \tag{6.5}$$

where $\mathbf{R}$ is the matrix form of the received packets with rows $\mathbf{r}_i = (r_{i,1}, ..., r_{i,L})$ and $\mathbf{E}$ is the error matrix in which the rows $\mathbf{e}_i = (e_{i,1}, ..., e_{i,L})$ represent the bit errors within each packet. Specifically, we have

$$e_{i,j} = \begin{cases} 1, & \text{if the } j\text{th bit of the } i\text{th packet is flipped} \\ 0, & \text{otherwise.} \end{cases} \tag{6.6}$$

The receiver can recover the message block $\mathbf{U}$ only when it receives *sufficient* error-free packets. The decoding problem is a system of linear equations and can be solved only when the equations are linearly independent. Hence, at least G linearly independent and error-free packets are required to recover $\mathbf{U}$. Typically, when the coding vectors are generated randomly as mentioned earlier, it suffices to receive slightly more than G correct packets to make sure with high probability that G linearly independent packets can be found among all the correct packets. When the

number of linearly independent and correct packets is greater than or equal to the generation size, then the receiver selects G packets among them randomly and performs RLC decoding to obtain the original message block $\mathbf{U}$ as follows:

$$\hat{\mathbf{U}} = \mathbf{C}_c^{-1}\mathbf{R}_c, \tag{6.7}$$

where $\mathbf{C}_c$ and $\mathbf{R}_c$ are the submatrices of $\mathbf{C}$ and $\mathbf{R}$ that are constructed by removing the rows corresponding to the partial packets. A packet that is linearly independent of the previous packets is called an innovative packet. An acknowledgment message is sent from the receiver at the end of each generation indicating the number of collected innovative packets. The receiver can calculate this number, for instant, by checking the rank of the coding matrix $\mathbf{C}$. Then the required number of coded packets is sent from the transmitter, and this process continues until the generation block is delivered to the receiver. Similarly, this algorithm can be deployed in a multicast setup where one transmitter communicates with M receivers, except that the transmitter should wait for all acknowledgment messages after sending each generation and then send a number of coded packets equal to the maximum number of required packets missing at different receivers. We refer to this model as the baseline system in the sequel, and design a PPR scheme to improve its performance. For the sake of simplicity we omit errors in the acknowledgment messages and neglect packet erasures due to misdetection and channel errors on the packet header. Nonetheless, it is a straightforward task to extend the results to cover the above situations.

6.2.2 Compressive Sensing and Sparse Recovery

In many practical situations, it is required to infer the actual values of some quantities by observing a set of measurements that are in a reduced subspace. This means that the dimensionality of the measurements is less than that of the original signal. In a linear setting, such cases can be modeled by a set of under-determined equations where the number of variables is higher than the number of equations. Explicitly, given $\mathbf{x}'$ the column vector representation of the original signal and $\mathbf{A}$ the sensing matrix, the

measurements $\mathbf{y}'$ can be obtained from

$$\mathbf{y}' = \mathbf{A}\mathbf{x}' \tag{6.8}$$

where the number of rows of $\mathbf{A}$ is less than the number of columns. Consequently, the size of $\mathbf{y}'$ is less than that of $\mathbf{x}'$. The inverse problem, i.e. restoring the original signal $\mathbf{x}'$ from $\mathbf{y}'$, can generally have an infinite number of solutions. However, when the original signal $\mathbf{x}'$ is sparse, there are efficient algorithms that can reliably recover the signal provided that the number of measurements is enough[2] considering the sparsity of the signal and the dimensionality of the measurements. This is a standard SR problem that typically appears in CS [88].

6.3 The Proposed Scheme

In this section we present our scheme that takes advantage of compressive sensing to recover partial packets. We design a systematic RLC which means that the first G packets are identical to the original packets. There are some advantages in using systematic RLC that can facilitate the procedure as well. For instance, using a systematic code, when the channel quality is desirable the packets can be correctly received with high probability. Hence, the receiver does not need to wait to receive enough independent packets. In addition, PPR can improve performance of the system in two ways. If all of the systematic packets are among the recovered partial packets, then the generation block is successfully received. Otherwise, if less than G linearly independent packets were originally received, the PPR procedure can add new innovative packets to eliminate the need for transmission of new packets.

Assuming that $T > G$ packets are sent, the coding matrix corresponding to systematic RLC is in the form

$$\mathbf{C}_{(T\times G)} = \begin{bmatrix} \mathbf{I}_{(G\times G)} \\ \hline \mathbf{P}_{((T-G)\times G)} \end{bmatrix}, \tag{6.9}$$

[2]The required number of measurements depends on the sensing matrix. It is known for matrices possessing the restricted isometry property or the minimum coherence condition [109].

where the horizontal dashed line represents matrix concatenation along the columns, $\mathbf{I}$ is the identity matrix, submatrix P is in the form

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \\ \mathbf{p}_{T-G} \end{bmatrix}, \tag{6.10}$$

and $\mathbf{p}_i = (p_{i,1}, ..., p_{i,G})$, $\forall i \in [1, (T-G)]$ is a random binary row vector that constitutes the ith row of $\mathbf{P}$. The corresponding coded packets are therefore given by

$$\mathbf{y}_i = \begin{cases} \mathbf{u}_i, & \forall i \in [1, G] \\ \mathbf{p}_{i-G}\mathbf{U} & \forall i \in [G+1, T] \end{cases} \tag{6.11}$$

Suppose that T packets are CRC coded and transmitted on a noisy wireless channel. Upon reception, first the packets are categorized based on their check sequences as correct or partial. After each round of transmission, an acknowledgment message is sent to the transmitter containing the number of collected innovative packets. If G packets are already collected, then no more coded packets are sent. Otherwise, the required number of coded packets missing from G are transmitted until successful reception of the generation block. The transmitter can then proceed to the next message block. Assume that for a given generation $T_c \leq T$ packets are received correctly. Then there are two possible situations:

1. $T_c < G$ or $T_c \geq G$, but less than G packets are linearly independent. RLC decoding is not performed but error recovery can be applied. PPR can recover some of the partial packets and increase the probability of the receiver collecting G independent packets.

2. $T_c \geq G$ and at least G packets are linearly independent. G correct packets are randomly selected and then the generation block $\mathbf{U}$ is obtained by the RLC decoding procedure in (6.7).

It can be seen that, while the RLC approach can only reconstruct the message when $T_c \geq G$ (case 2), the receiver can try error recovery in the first two situations and decode the generation block. Note that, as we mentioned earlier, PPR can either directly lead to the recovery of the original message

or it can facilitate the operation of the RLC decoding by increasing the total number of available valid packets.

Since this procedure is usually performed after FEC, we assume that the partial packets are mostly correct with a few symbols in error. In other words, the errors are sparsely distributed in random positions in the packets. If we assume that the channel is time invariant and errors are identically distributed within the packets, then we can also conclude that the errors can be assumed sparse in a fixed location across different packets (i.e. sparsity in the matrix implies sparsity across the columns too). We therefore assume that the columns of $\mathbf{E}$ denoted by $\mathbf{e'}_1, ..., \mathbf{e'}_L$ are also sparse.

Here we design matrix $\mathbf{A}$ such that

$$\mathbf{AC} = \mathbf{0}, \tag{6.12}$$

where $\mathbf{0}$ is an all-zero matrix. This means that $\mathbf{C}$ is a basis for the null space of $\mathbf{A}$. From the algebraic coding theory, one solution is[3]

$$\mathbf{A}_{((T-G)\times T)} = \left[\begin{array}{c:c} -\mathbf{P} & \mathbf{I}_{((T-G)\times(T-G))} \end{array} \right], \tag{6.13}$$

where the vertical dashed line represents matrix concatenation alongside the rows. Note that matrix $\mathbf{A}$ is perfectly known at the receiver from the knowledge of $\mathbf{C}$. Defining

$$\mathbf{S} = \mathbf{AR}, \tag{6.14}$$

it follows from (6.3) and (6.12) that

$$\mathbf{S} = \mathbf{A}(\mathbf{CU} + \mathbf{E}) = \mathbf{AE}. \tag{6.15}$$

In other words, the product $\mathbf{AE}$ and the known matrix $\mathbf{A}$ are given and the unknown matrix $\mathbf{E}$ is to be determined. The above equation and the procedure to solve it (i.e. error recovery) constitute the core of our PPR scheme. Therefore, we elaborate more on this issue in the sequel to provide a better insight. First, about the unknown matrix $\mathbf{E}$ we notice that

- Its elements are nonnegative integer values.

[3]Note that $-\mathbf{P} = \mathbf{P}$ in $GF(2)$

- The matrix is sparse.

The former observation implies that the problem at hand is naturally an integer program whose complexity is non-polynomial. Such problems are more difficult than the equivalent real-valued linear programs. The conventional approach to solve an integer linear program is to relax the integer constraint and solve it assuming real variables. To transform the results back to a feasible set, they are rounded-off to the nearest integer. However, we may facilitate the error recovery by taking into the account our latter observation which is based on the fact that channel errors do not occur very often and the number of channel errors is normally much smaller than the packet length. Therefore, we can model the problem as a CS of the sparse matrix $\mathbf{E}$ using the sensing matrix $\mathbf{A}$. In contrast to most SR problems tackled in the CS community where a sparse vector is recovered, here we deal with a two-dimensional sparse signal and the variables are in the form of matrices. It can however be seen as a set of L independent standard SR problems to recover different columns of $\mathbf{E}$. Hence, it can be formulated as follows:

$$\begin{aligned} &\hat{\mathbf{e}}'_i = \arg\min \|\mathbf{x}'\|_0 \\ &\text{subject to} \quad \mathbf{A}\mathbf{x}' = \mathbf{s}'_i,\ i = 1, \ldots, L. \end{aligned} \tag{6.16}$$

where $\mathbf{s}'_i$ is the ith column of $\mathbf{S}$ and $\hat{\mathbf{e}}'_i$ is the ith recovered error column. Since the l_0 norm in (6.16) is not convex, the problem is NP-hard in general. Nevertheless, from the CS literature we know that minimization can be performed over the l_1 norm which is a convex function. This approach will be covered in more detail in the next section.

In the rest of this section we focus on the l_0 minimization problem in (6.16), applying concepts from algebraic coding theory. This is due to the close relationship between the sparse recovery of signals in finite fields and the problem of decoding of noisy linear codewords known as *syndrome decoding*. Explicitly, (6.16) can be solved using syndrome decoding in which $\mathbf{A}$ is the parity check matrix and each column of $\mathbf{S}$ is an independent syndrome.

Assume that the column vectors $\{\mathbf{e'}_i\}$ are m-sparse. Since there are T_c error-free packets, the corresponding rows of all column vectors $\{\mathbf{e'}_i\}$ are equal to zero. The total number of m-sparse candidate solutions for each column which is equivalent to the size of the syndrome decoding table is given by $\sum_{k=1}^{m} \binom{T - T_c}{k}$.

We denote the recovered error matrix by $\hat{\mathbf{E}}$ which is the corresponding set of coset leaders[4] required for recovery of the partial packets. After the reconstruction of $\hat{\mathbf{E}}$, the denoised packets can be estimated by performing the modulo-two operation of adding the error matrix to the received packet Matrix $\mathbf{R}$

$$\hat{\mathbf{Y}} = \mathbf{R} + \hat{\mathbf{E}}. \tag{6.17}$$

At this stage, the receiver checks again if the packets are decodable. Explicitly, it first looks at the first G packets. If their errors are resolved, regardless of the status of the other packets it will send an acknowledgment message to the transmitter. Otherwise, it counts the number of correct packets again and compares it with G. If it is larger than G, the receiver looks for independent packets by randomly building the submatrix $\mathbf{C}_c$ which is the corresponding coding matrix of the correct packets and checks if it is invertible. This process can be repeated if the number of correct packets is large enough. The message block $\mathbf{U}$ is recovered if an independent set can be found by solving (6.7). Otherwise the receiver waits for more packets from the transmitter. This process is repeated for the succeeding packets as well, until the successful decoding of the message block.

6.4 Error Recovery

We elaborate more on different error recovery algorithms based on the exact l_0 minimization problem (i.e. the integer program) and the equivalent convex optimization approaches.

[4]In syndrome decoding, a coset is defined as the set of all error patterns that share a common syndrome. A coset leader is a word of minimum Hamming weight among all elements of a coset.

$$\begin{bmatrix} 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & 0 & 1 \\ 0 & \cdots & 0 & 1 & 1 & 0 & \cdots & 0 \end{bmatrix}$$

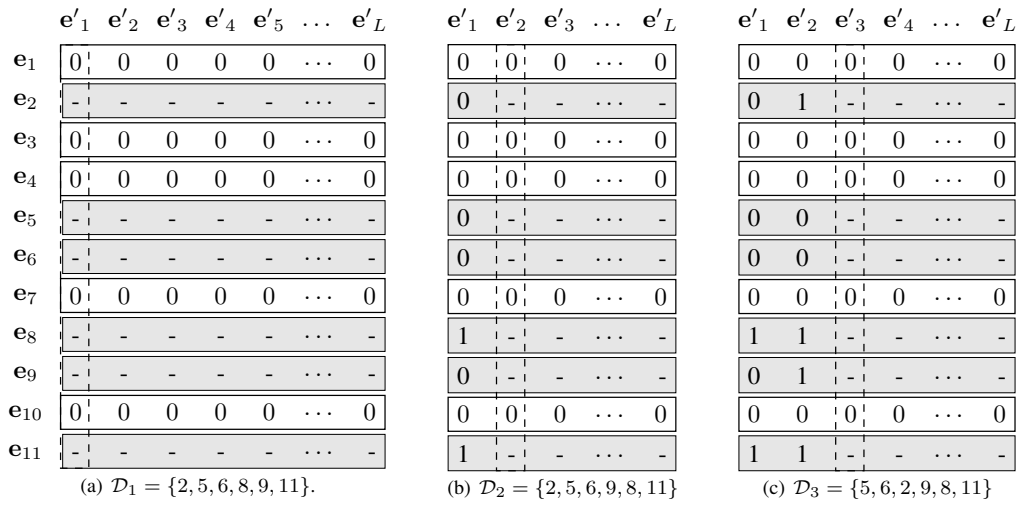
Figure 6.2: Error matrix corresponding to bursty (left) and homogeneous (right) error distributions. Partial packets are distinguished by a light gray shade and bit errors with a dark shade.

6.4.1 Sparse Solutions of the Integer Program and Optimization of the Recovery Algorithm

The row elements of the error matrix $\mathbf{E}$ correspond to channel errors within a packet and tend to be sparse too. The task of PPR is to locate the errors in a partial packet (i.e. a given row), that can be seen as a set of hypothesis test problems. Assume that the problems are dealt with sequentially, beginning from the leftmost element of each row. Initially, all partial packets can *equally* share an error. However, the test statistics for each position in different rows depend on the decision for the previous positions. Channel errors usually depend on factors such as noise, fading, interference and collisions. Hence, they can be distributed within the partial packets homogeneously or in the form of bursts. More clearly, a fixed number of errors can either be homogeneously distributed among *several* partial packets, or can collect in a *few* of them.

These two cases are illustrated in Fig. 6.2. When the columns of the error matrix are processed separately, the order in which different rows are processed is important. For instance, assume a homogeneous distribution of errors (e.g. when the errors are due to AWGN only). Assume that there are two partial packets and a fixed processing order is used (e.g. the first partial packet is processed first). Without re-ordering, the following candidate solution may appear in the search:

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{bmatrix}.$$



(a) $\mathcal{D}_1 = \{2, 5, 6, 8, 9, 11\}$. (b) $\mathcal{D}_2 = \{2, 5, 6, 9, 8, 11\}$ (c) $\mathcal{D}_3 = \{5, 6, 2, 9, 8, 11\}$

Figure 6.3: The error values corresponding to the partial (shaded) and correct (white) packets. Only the error values of the correct packets are known and are set to zero. The search space shrinks to partial packets only in this way. The search algorithm starts at the first column of $\mathbf{E}$, i.e. $\mathbf{e'}_1$ in (a) and proceeds to the next columns sequentially.

This is because the sparse solutions of the previous columns are not considered when a new column is processed. In such cases, all columns are legitimate sparse solutions of the corresponding SR problem. However, when the columns are merged and an error matrix is formed, the candidate solution is less likely or even infeasible (e.g. some partial packets have no errors). Although such candidate solutions are discarded after CRC check and the correct solution is eventually found, this approach is inefficient. Appropriate re-ordering can result in a more efficient search.

We explain this re-ordering method through an example for simplicity. Fig. 6.3 represents the recovery procedure of the error matrix where $T = 11$ packets are transmitted and $T_c = 5$ packets are received correctly. The corresponding rows in the error-free packets are substituted with zeros. However, for the partial packets identified by the index set $\{2, 5, 6, 8, 9, 11\}$ and gray shading, the algorithm should recover the error values, i.e. decide whether they are 0 or 1. Assume that the search domain at the ith step (corresponding to the ith column of the error matrix) is denoted by the

ordered set $\mathcal{D}_k = \{d_1, d_2, ..., d_{T_p}\}$ in which $T_p = T - T_c = 6$ is the number of partial packets and the ordering of the elements is such that

$$\mathbb{P}\{e'_{d_1,k} = 1\} > \mathbb{P}\{e'_{d_2,k} = 1\} > ... > \mathbb{P}\{e'_{d_{T_p},k} = 1\}, \tag{6.18}$$

where $\mathbb{P}$ represents probability. Due to the sparsity of the errors across the rows we use the following ordering rule for $k > 1$:

$$\mathcal{D}_k = \{d_1, d_2, ..., d_{T_p} : \sum_{l=1}^{k-1} e_{d1,l} < \sum_{l=1}^{k-1} e_{d2,l} < ... < \sum_{l=1}^{k-1} e_{d_{T_p},l}\}. \tag{6.19}$$

At the beginning, since there is no prior knowledge we can use a random or equivalently a fixed order such as $\mathcal{D}_1 = \{2, 5, 6, 8, 9, 11\}$. Assume that the recovered error column corresponding to $\mathbf{e'}_1$ is $\hat{\mathbf{e}}'_1 = \{0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1\}$. Then since there are two error elements equal to 1, corresponding to the packet indices 8 and 11, the new ordered domain set is $\mathcal{D}_2 = \{2, 5, 6, 9, 8, 11\}$. Consequently, after recovering the error column $\mathbf{e'}_2$, $\mathcal{D}_3 = \{5, 6, 2, 9, 8, 11\}$ and the process continues until all of the L columns are recovered.

Note that the re-ordering is only to improve the search efficiency for sparse recovery. PHY-related settings do not change the nature of this procedure. Operations such as FEC or modulation can only change the probability of errors. If interleaving is performed at PHY, it can only change the locations of the errors and the cardinality of the support set is unchanged.

6.4.2 Convex Relaxation of the l_0 Minimization Problem

The normal procedure to solve the l_0 minimization problem in CS is a convex relaxation. That is, the equivalent l_1 norm, which is a convex function, is minimized using linear programming [88]. As long as the sensing matrix satisfies the incoherence property, close to perfect reconstruction is guaranteed using $O(m \log(T/m))$ measurements. Explicitly, we can solve

$$\begin{aligned} &\hat{\mathbf{e}}'_i = \arg\min \|\mathbf{x}'\|_1 \\ &\text{subject to} \quad \mathbf{A}\mathbf{x}' = \mathbf{s}'_i, \ i = 1, \ldots, L. \end{aligned} \tag{6.20}$$

The recovered values should be rounded off to the closest feasible values (e.g. 0 and 1 in $GF(2)$). We refer to this approach as *per-column CS*

since a single CS problem is solved for every column. Alternatively, the error matrix can be converted to a long vector. To solve (6.16) using this approach, one can re-arrange all the columns of $\mathbf{S}$ in a new vector and solve the vectorized problem. In this technique, known as *Kronecker-CS*, only one SR problem needs to be solved, which is in the form

$$\begin{aligned} &\hat{\mathbf{E}} = \text{reshape}(\arg\min \|\mathbf{x}'\|_1, T, L) \\ &\text{subject to} \quad \bar{\mathbf{A}}\mathbf{x}' = \text{reshape}(\mathbf{S}, L(T-G), 1) \end{aligned} \tag{6.21}$$

where the function "reshape" is used to convert between matrix and vector representations by concatenating or splitting different columns in which the appropriate dimensions of the output are given as the last two arguments, and

$$\bar{\mathbf{A}}_{(L(T-G)\times LT)} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A} \end{bmatrix}, \tag{6.22}$$

in which matrix $\mathbf{A}$ is repeated L times across the diagonal. The joint measurement matrix $\bar{\mathbf{A}}$ can also be expressed as $\bar{\mathbf{A}} = \mathbf{I}_L \otimes \mathbf{A}$ where $\mathbf{I}_L$ is the identity matrix of size L and $\otimes$ represents the Kronecker product [110]. The optimization problem can then be solved by conventional CS techniques such as basic pursuit [88].

6.4.3 The Minimum Number of Packets Required for Decoding

The minimum number of packets that the transmitter should send for successful decoding at the receiver(s) depends on the approach selected for error recovery. We discuss this issue and derive the corresponding expressions.

Assume that the columns of the error matrix $\mathbf{E}$ are m-sparse, i.e., there is a maximum of m nonzero elements in each of its columns. In the standard sparse recovery problem (6.16), the measurement matrix $\mathbf{A}$ transforms a set of column vectors of length T to vectors of length $T-G$. Hence, we have $T-G$ measurements to recover the error vectors of size T. For a measurement matrix $\mathbf{A}$ with near-minimal coherence of the order $c/\sqrt{T-G}$,

m-sparse recovery is achievable and the number of measurements scales quadratically in the sparsity order [111]. Conversely, with $\mathbf{A}$ possessing the restricted isometry property, only a linear number of measurements with a small logarithmic factor of the data size is required. Hence, using l_1 minimization we have

$$T_{\mathsf{min}}^{l_1}(m) \approx \alpha m + G, \tag{6.23}$$

where $T_{\mathsf{min}}^{l_1}$ is the minimum number of coded packets that should be sent for successful recovery with the l_1 approach, α is a scalar, typically taking values between two and five and independent of m. Equivalently, the maximum number of errors that can be recovered after receiving T packets is $(T - G)/\alpha$.

The probability of occurrence of at most m bits of error in all the columns given T packets can easily be given by

$$P(m, T) = \Big[\sum_{z=0}^{m} \binom{T}{z} \varepsilon^z (1-\varepsilon)^{(T-z)}\Big]^L. \tag{6.24}$$

So we can calculate the maximum possible sparsity order for a given T. As an example, the maximum sparsity order with 99% confidence versus the number of transmitted packets is depicted in Fig. 6.4 for three different bit error probabilities[5].

If the original l_0 minimization is used, then we can establish a new bound on the number of packets required for successful error recovery. The minimum number of measurements required to solve the exact l_0 minimization is addressed in [112, Lemma 3.1]. To be able to uniquely recover each error column of size T the kernel of the measurement matrix $\mathbf{A}$ should not contain any $2m$-sparse vectors other than the zero vector. Equivalently, any set of $2m$ columns of $\mathbf{A}$ should be linearly independent. Since the rank of $\mathbf{A}$ is at most $T - G$, it follows that

$$T_{\mathsf{min}}^{l_0}(m) = 2m + G, \tag{6.25}$$

which is the optimal bound for the l_1 minimization problem.

[5]The selected bit error probabilities are pessimistic and account for the worst case scenario. Such conditions can occur frequently in wireless sensor networks.

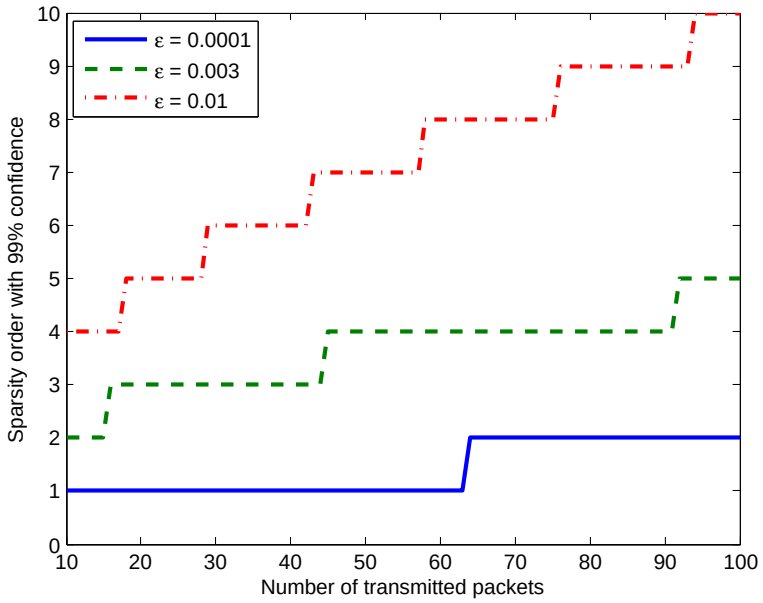


Figure 6.4: Maximum possible sparsity order with 99% confidence versus the number of transmitted packets.

Regardless of how we solve the error recovery problem, the RLC decoding algorithm requires at least G linearly independent packets, which puts a separate constraint on the required number of packets. The procedure is straightforward and is given in [113]. More clearly, the problem can be formulated in the form of an absorbing Markov chain with $G+1$ states (0 to G) as depicted in Fig. 6.5. Each state represents the number of collected innovative packets. The probability of transition from state $i \geq 0$ to state $j \geq 0$ is as follows

$$p_{ij} = \begin{cases} \zeta_i, & i = j, \forall i \in \{0, 1, ..., G\} \\ 1 - \zeta_i, & j = i + 1, \forall i \in \{1, ..., G\} \\ 0, & \text{otherwise.} \end{cases} \tag{6.26}$$

where $\zeta_i = q^{-(G-i)}$ is the probability of staying at the ith state after receiving a new packet (i.e. the probability that the received packet is not innovative). We refer to this system as "baseline" and denote the corresponding transition matrix by $\mathbf{H}^{\mathsf{BL}} = \{p_{ij}\}$. The expected number of steps to complete the process at state G beginning from state 0 can be calculated

by adding up the expected number of transitions from state $i \in \{0, ..., G-1\}$ to $j \in \{1, ..., G\}$. Using the linearity property of the expectation and assuming $E_{ij} = \mathbb{E}\{\text{the number of packets required to change state from } i \text{ to } j\}$, and the fact that a new packet is required only if a transition does not change the state we have[6]

$$E_{ij} = p_{ii}(E_{ij} + 1) + p_{ij}, \tag{6.27}$$

that yields $E_{ij} = \frac{1}{p_{ij}}$. In practice, packets with a failed checksum (i.e. packet erasures) should also be taken into account. Hence, the transition matrix $\mathbf{H}^{\mathsf{ER}}$ corresponding to the Markov chain model of the RLC system with packet erasures can be expressed by ζ' such that

$$\zeta'_i = \zeta_i + \Gamma(1 - \zeta_i), \, \forall i \in \{0, ..., G-1\}, \tag{6.28}$$

where

$$\Gamma = 1 - (1 - \varepsilon)^L \tag{6.29}$$

represents the packet erasure rate and ε denotes the channel's bit error probability. It follows that the average number of packets that should be sent for successful decoding of a generation block of size G (i.e. the expected number of steps required for absorption) is

$$\bar{T}^{\mathsf{RLC}}_{\mathsf{min}} = \sum_{i=0}^{G-1} \frac{1}{(1-\Gamma)(1-\zeta_i)} = \sum_{i=0}^{G-1} \frac{1}{(1-\Gamma)(1-q^{-(G-i)})} \tag{6.30}$$

Since we use a systematic code, the expected number of required packets is slightly less than $\bar{T}^{\mathsf{RLC}}_{\mathsf{min}}$. Hence it can be treated as an upper bound. This is because with a systematic code the first G packets are enough for decoding if they are received correctly. In addition, since during the systematic phase all the transmitted packets are inherently innovative, the total number of collected innovative packets at the end of the transmission of the first G packets is higher, than for a non-systematic code. The exact modeling of systematic RLC with PPR complicates the analysis but does not add much

[6]With probability p_{ij} the state is changed using one innovative packet, and no more packets are required. With probability p_{ii} one new packet is required.

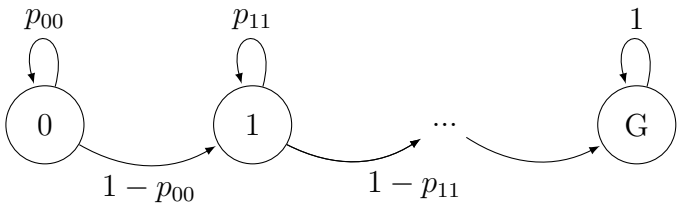


Figure 6.5: The Markov chain of RLC decoding (baseline system). The states represent the number of collected innovative packets for successful decoding. A generation of size G can be decoded at the state G.

practical insight. Hence, we use the general model for non-systematic RLC and treat the result as a plausible approximation. As shown in [114], this approximation improves by increasing the field size.

When partial packet recovery is applied, the Markov chain in Fig. 6.5 should be modified accordingly. Since the chance of the recovery algorithm succeeding depends on the number of collected packets as well as the sparsity of the partial packets, we have a time-varying Markov chain. Hence, we need to assume a fixed T and calculate the probabilities conditioned on T. We recall the three decoding situations presented in Section 6.3 and notice that RLC can only be decoded when $T_c \geq G$ and at least G packets are linearly independent. However, using PPR the first two situations can also lead to successful decoding.

Assume that after the transmission of T packets the receiver can recover the partial packets with probability $\Lambda(T)$. Without loss of generality, we assume that l_0 minimization is used for PPR. We can then generalize the results for the l_1 minimization approach. For simplicity we assume that

$$\Lambda(T) = \begin{cases} 1, & T \geq T_{\mathsf{min}}(m) \\ 0, & \text{otherwise} \end{cases} \tag{6.31}$$

where $T_{\mathsf{min}}(m)$ is substituted from (6.23) or (6.25) depending on the recovery scheme. The exact calculation of $\Lambda(T)$ is complicated as it requires averaging over all possible combinations of the error distribution in the partial packets. Since we look at the columns of the error matrix, the combinations of the column errors in different rows should also be considered.

The relationship between the sparsity order, m, and the number of partial packets T_p is not definitive. The best situation is when all the partial packets share no more than one bit error in each column, which corresponds to $m = 1$. On the other hand, in the worst case all T_p packets are corrupted in at least one column, that leads to $m = T_p$. When PPR is enabled, any k partial packets that are successfully recovered can contribute to complete the Markov chain and reach to at most k higher states. The PPR algorithm can recover the packets only when $k > 0$ partial packets are available and enough packets are collected (i.e. $\Lambda(T) = 1$). Hence, we can derive the transition matrix $\mathbf{H}^{\mathsf{PPR}} = \{p_{ij}^{\mathsf{PPR}}\}$ for the Markov chain model of the PPR system as follows:

$$\mathbf{H}^{\mathsf{PPR}}(T) = \begin{cases} \left(P_K(0)\mathbf{H}^{\mathsf{ER}} + \sum_{k=1}^{T} P_K(k) \left[\mathbf{H}^{\mathsf{BL}}\right]^k\right), & \Lambda(T) = 1 \\ \mathbf{H}^{\mathsf{ER}}, & \text{otherwise,} \end{cases} \tag{6.32}$$

where P_K is the probability that k out of T packets are in error, which is given by

$$P_K(k) = \binom{T}{k} \Gamma^k (1-\Gamma)^{(T-k)}. \tag{6.33}$$

One can show that all the above transition matrices are in the form

$$\mathbf{H} = \begin{bmatrix} \mathbf{Q} & \mathbf{B} \\ \mathbf{0} & 1 \end{bmatrix}, \tag{6.34}$$

since they represent an absorbing Markov chain. In addition any power λ of such matrices can be written in the form

$$\mathbf{H}^{\lambda} = \begin{bmatrix} \mathbf{Q}^{\lambda} & * \\ \mathbf{0} & 1 \end{bmatrix}. \tag{6.35}$$

If we denote the expected number of transitions required for absorption (i.e. successful completion of the Markov chain at the final state) initiating from state i by n_i, then it can be shown that

$$\mathbf{n}' = (\mathbf{I}_{(G-1)\times(G-1)} - \mathbf{Q})^{-1}\mathbf{1}_{(G-1)\times 1} \tag{6.36}$$

is a vector whose elements are n_i, $\forall i \in [0, G-1]$. We can consider the transition matrix given in (6.32) in two different cases. First assume that

$\Lambda(T) = 0$, and the initial distribution of the states in the Markov chain is $\boldsymbol{\mu}(0) = (1, 0, ..., 0)$; this corresponds to beginning from state 0 with probability one. The probability distribution of the states after $T_{\mathsf{min}}^{l_0}$ transmissions becomes

$$\boldsymbol{\mu}(T_{\mathsf{min}}^{l_0}) = \boldsymbol{\mu}(0) \left[\mathbf{H}^{\mathsf{ER}}\right]^{T_{\mathsf{min}}^{l_0}}. \tag{6.37}$$

When Λ changes value from zero to one at $T = T_{\mathsf{min}}^{l_0}$, the transition matrix is updated and will then change with T. However, the average number of packets for successful delivery of each generation block using the PPR approach can be approximated by

$$\bar{T}_{\mathsf{min}}^{\mathsf{PPR}} \approx \begin{cases} \tilde{\boldsymbol{\mu}}(T_{\mathsf{min}}^{l_0})\tilde{\mathbf{n}}' + T_{\mathsf{min}}^{l_0}, & \text{if } T_{\mathsf{min}}^{l_0} > \bar{T}_{\mathsf{min}}^{\mathsf{RLC}} \\ \bar{T}_{\mathsf{min}}^{\mathsf{RLC}}, & \text{otherwise} \end{cases} \tag{6.38}$$

where $\tilde{\boldsymbol{\mu}}(T_{\mathsf{min}}^{l_0})$ is the state distribution vector after $T_{\mathsf{min}}^{l_0}$ transitions corresponding to the $G - 1$ states (excluding the final state), and $\tilde{\mathbf{n}}'$ should be calculated using (6.36) for the transition matrix $\mathbf{H}^{\mathsf{PPR}}(T_{\mathsf{min}}^{l_0})$ identified in (6.32).

6.4.4 Application in Wireless Body Area Networks

In some applications such as wireless body area networks (WBANs), optimal energy efficiency is a feasibility constraint. WBANs are wireless networks consisting of tiny implantable or wearable sensor nodes that operate inside or around the human body and communicate body-related data, typically to a *hub* node. In a WBAN where sensors are implanted in the body, sensor nodes should transmit their data with minimum energy consumption since they only rely on battery power that is unreachable for re-charging. The hub node is less constrained in terms of energy consumption and computational complexity since the re-charging can conveniently be performed often and the device can be much larger than the sensor nodes. For these reasons, an asymmetric design of the communication protocol is needed that imposes minimum computational complexity and energy consumption on the sensor nodes and compensates for this by more sophisticated procedures at the hub. In this way, the computational complexity that is outsourced

to the hub node is leveraged to improve the transmission efficiency at the sensor nodes. By exploiting PPR, the performance of RLC can offer a considerable gain in point-to-point transmissions (sensor to hub), where it usually does not have the advantage.

6.4.5 Discussion on the Generation Size and the Packet Length

Normally, in RLC, a higher generation size leads to a better throughput, but an increase in latency. A small generation size cannot offer much coding gain. Using PPR, however, the gain with a generation size of $G = 4$ can be considerable. The generation size does not significantly influence the complexity of the PPR problem and the required solution time (more details are presented in the simulation results). However, since a larger generation requires more packets, it can consequently increase the sparsity order of the problem. Note that the channel's bit error probability, ε also affects the sparsity order and should be taken into the consideration. Recall that (6.24) implies that, for a channel with a higher bit error probability, a smaller generation size should be selected to decrease the sparsity order.

The impact of the packet length, L, on the complexity of PPR depends on how the problem is solved. Except for the Kronecker CS, the complexity is linearly increased, namely, L problems should be solved sequentially. For large and even moderate L values, Kronecker CS, which is a parallel approach, is infeasible due to the need for a significant amount of memory.

6.5 Simulation Results

In this section, the performance of the proposed system is evaluated by simulation and is compared with that of the conventional RLC and uncoded transmission. We perform three sets of simulations, for uncoded, RLC, and RLC with PPR, with the same parameters assuming point-to-point and multicast channels. In the PPR scheme we compare SR solutions based on l_0 minimization (syndrome decoding) as well as the two l_1 minimiza-

tion solutions (Kronecker CS and per-column CS) discussed earlier. 1000 generations are transmitted in each simulation and the number of required packets that should be transmitted for successful recovery of the message block are plotted. In Fig. 6.6, given $\varepsilon = 0.004$, $L = 256$, and $G = 8$ the empirical cumulative distribution (CDF) of the number of transmitted packets is demonstrated for point-to-point and multicast channels with 5 receivers[7]. As expected, RLC does not make much difference in the point-to-point case in comparison with the uncoded transmission. Nevertheless, combined RLC and sparse error recovery can considerably reduce the required number of transmit packets. For instance, while the uncoded transmission requires 35 packets to achieve a probability of successful decoding of 95%, the combined RLC and SR method requires 15 packets to guarantee the same success probability, which saves about 57% of the transmit packets. It can also be observed that RLC can reduce the total required number of packets in the multicast scenario. In the given case study, 58, 41, and 19 packets are required to attain a success probability of 95% in the uncoded, RLC, and combined RLC and PPR schemes, respectively. The performance of syndrome decoding is determined by the parameter m. A higher m value leads to a smaller number of transmit packets (because the PPR can recover columns with more errors) but will increase the size of the decoding table and the computational complexity. We have set $m = 5$ in this figure.

In the next simulation we repeat this process for a range of different channel bit error probabilities. Explicitly, ε is varied between 0.001 and 0.007, which corresponds to a signal-to-noise ratio from 6.8 dB to 4.8 dB[8] assuming antipodal signaling in an additive white Gaussian Noise (AWGN) channel [9]. The average number of transmit packets along with the corresponding standard deviations are illustrated in Fig. 6.7. The analytical average number of required packets is also calculated using the Markov

[7]Throughout this chapter the packet length L is in bit units.

[8]These values are rather pessimistic for a typical WBAN. However, they can occur in fading conditions [115].

[9]If we assume FEC at the physical layer, depending on the correction capability of the channel code these values can change. Here, no FEC is assumed.

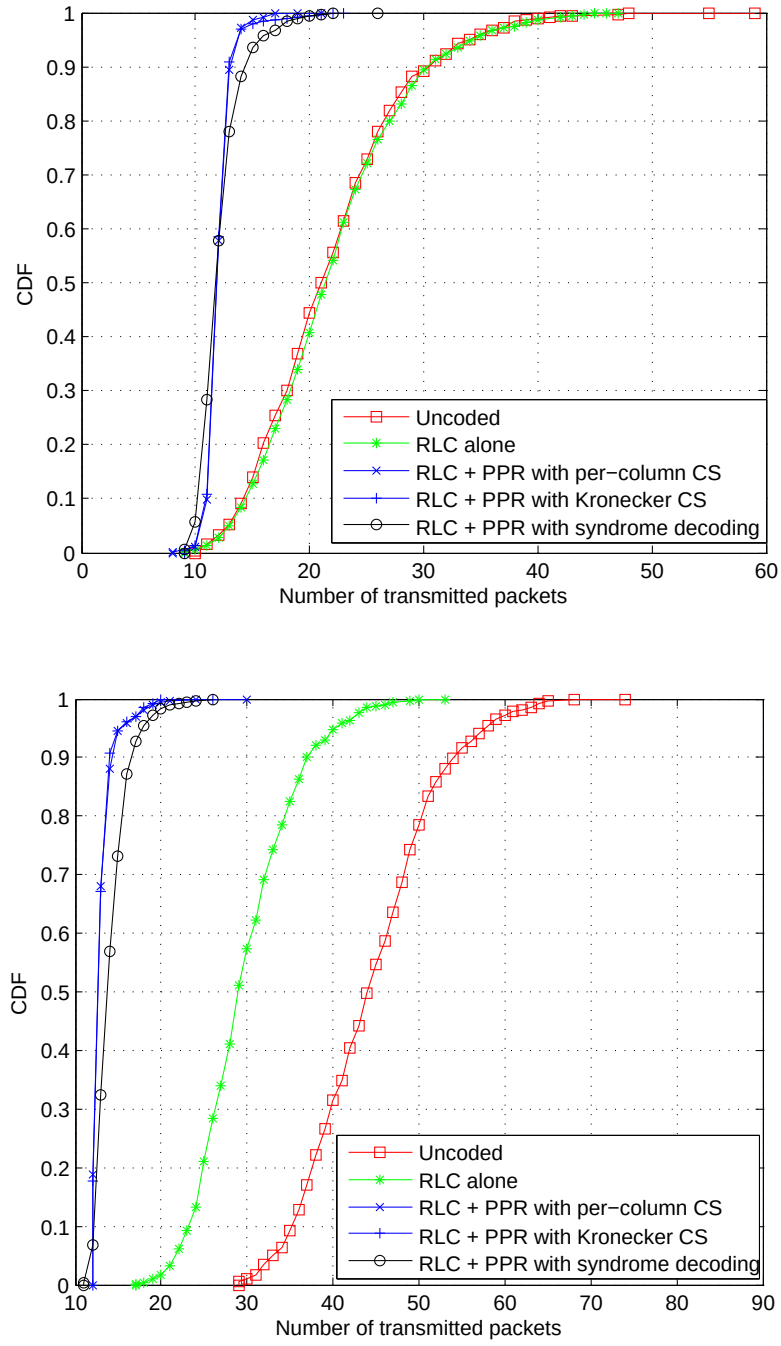


Figure 6.6: The empirical CDF of the number of transmitted packets per generation corresponding to RLC alone and RLC with PPR with $L = 256$, $\varepsilon = 0.004$, $G = 8$ for point-to-point (top) and multicast channel with 5 receivers (bottom).

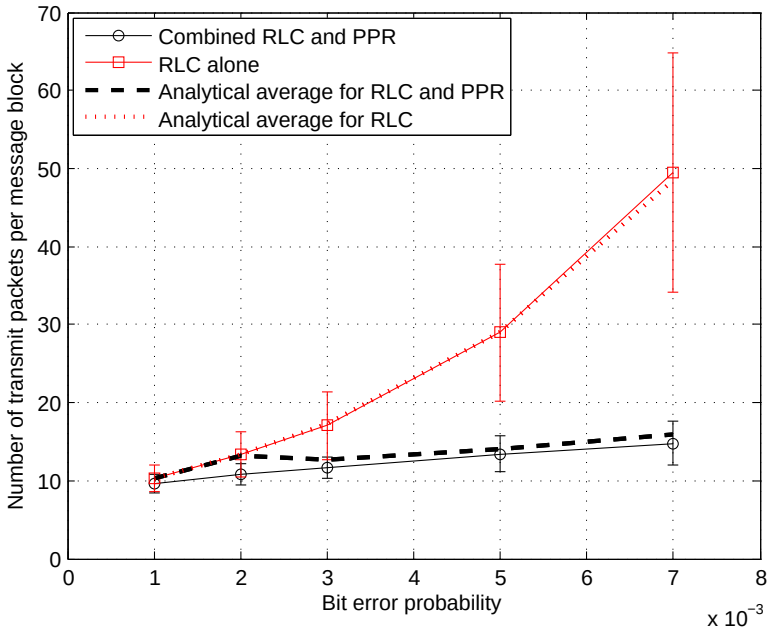


Figure 6.7: The number of transmitted packets per generation as a function of bit error probability. The average values (based on simulation and Markov chain modeling) are illustrated along with the standard deviations, assuming $L = 256$ and $G = 8$. Syndrome decoding is used for PPR.

models presented in the previous section. As mentioned before, for the sake of simplicity we approximated the mean value using the Markov model of a general non-systematic RLC and selected a large field size to approximate the systematic code [114].

The impact of the generation size on the achievable transmission efficiency is also studied. Assuming the same multicast setting with 5 receivers and $\varepsilon = 1 \times 10^{-3}$, the simulation is performed for different G and L values and the transmission efficiency obtained by uncoded transmission, RLC alone, and combined RLC and PPR is reported in Table 6.1. We compare the transmission efficiency η defined as

$$\eta = \frac{G}{\mathbb{E}\{T\}(1 + \tau/L)}, \tag{6.39}$$

where L is the length of the payload in each packet and τ is the total number of overhead bits such as headers and checksums[10]. η is indeed the ratio of

[10]In the simulations we use $\tau = 128$ bits.

Table 6.1: Transmission Efficiency in a Multicast Channel with 5 Receivers.

		$G = 4$	$G = 8$	$G = 16$
RLC and PPR	$L = 2048$	0.39	0.41	0.49
	$L = 1024$	0.40	0.49	0.52
	$L = 512$	0.45	0.52	0.56
	$L = 256$	0.44	0.49	0.53
RLC	$L = 2048$	0.08	0.08	0.09
	$L = 1024$	0.21	0.24	0.26
	$L = 512$	0.33	0.37	0.40
	$L = 256$	0.39	0.43	0.45
Uncoded	$L = 2048$	0.06	0.05	0.06
	$L = 1024$	0.16	0.16	0.16
	$L = 512$	0.27	0.27	0.27
	$L = 256$	0.32	0.31	0.32

the *useful* transmitted information to the total transmitted bits including overhead and preamble data. A similar measure has been widely used, for instance in [116] and the references therein, to quantify the energy efficiency of the system. Since the transmission energy is directly proportional to the number of transmitted bits, this definition of transmission efficiency is particularly appropriate to measure energy efficiency too. As can be observed from Table 6.1, increasing the generation size can lead to a better result. However, a larger generation size results in more computational complexity at both the transmitter and receiver, which should be taken into consideration for a fair complexity-performance tradeoff. In addition, there is an optimal packet length to achieve maximum efficiency that depends on the generation size, the number of overhead bits and the bit error probability [116]. For example, when PPR is used the maximum transmission efficiency is achieved for $L = 512$, while the optimal packet length without PPR is at $L = 256$. Generally speaking, PPR can improve transmission efficiency for all L and G values. This improvement is more significant for longer packet lengths. This is particularly due to an increase in the packet error rate for longer packets. However, since the performance of the PPR is only affected by the bit error probability, it can overcome the destructive impact of long packets. It should be noted that, although the transmission efficiency is improved, the number of SR problems (as well as the complexity of the PPR) increases linearly with the packet length.

Table 6.2: Transmission Efficiency in a Multicast Channel with 5 Receivers ($L = 1024$).

PER		87%	64%	56%	40%	19%	$< 1\%$
RLC and PPR	$G = 16$	0.44	0.52	0.54	0.62	0.72	0.86
	$G = 8$	0.38	0.49	0.51	0.58	0.65	0.87
	$G = 4$	0.37	0.40	0.43	0.48	0.59	0.86
RLC	$G = 16$	0.08	0.26	0.32	0.44	0.64	0.87
	$G = 8$	0.08	0.24	0.29	0.42	0.61	0.86
	$G = 4$	0.07	0.21	0.27	0.38	0.57	0.87
Uncoded	$G = 16$	0.05	0.16	0.20	0.29	0.48	0.87
	$G = 8$	0.05	0.16	0.20	0.28	0.48	0.87
	$G = 4$	0.05	0.16	0.20	0.29	0.48	0.87

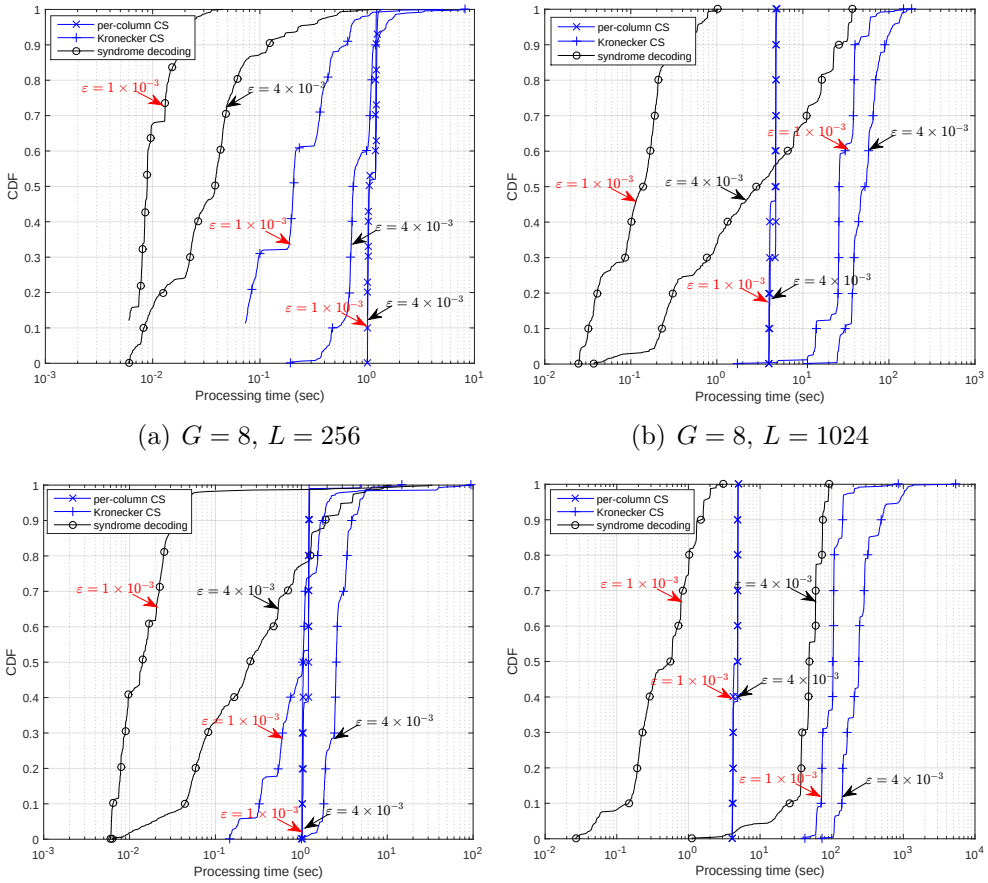
In order to quantify the impact of packet quality on the PPR and RLC, the transmission efficiency is presented in Table 6.2 for different generation sizes and packet error rates (PER), i.e. the ratio of partial packets to the total received packets. For a very small PER[11], RLC alone or RLC with PPR cannot offer any gain. Since the packets are successfully delivered with high probability, $\mathbb{E}\{T\} \approx G$, hence $\eta \approx L/(L+\tau)$. However, with increasing PER, the advantage of RLC over the uncoded scenario as well as the gain from PPR over RLC alone becomes more visible. This trend applies to all PER values, while the gap between RLC alone and RLC with PPR increases with increasing PER. Nonetheless, the transmission efficiency degrades even with PPR. The reason is that a fixed sparsity order, $m = 5$, is used that limits the maximum number of errors in each error column to 5 bits.

6.5.1 Processing Time

The complexity of RLC compared to the uncoded system has been addressed in [117, 118, 94] and the references therein. In Fig. 6.8, we have compared the complexity of different SR solutions in the PPR approach. The CDF of the total elapsed time corresponding to syndrome decoding, Kronecker CS, and per-columns CS are shown[12]. To solve the l_1 minimiza-

[11]PER is lower for a smaller packet length L, or a lower bit error probability ε. We fix L to 1024 bits in the simulation.

[12]It should be noted that we use a non-optimized Matlab software which is much slower than the practical implementations of the network coding tools. Hence, the cor-



(a) $G = 8$, $L = 256$ (b) $G = 8$, $L = 1024$

(c) $G = 16$, $L = 256$ (d) $G = 16$, $L = 1024$

Figure 6.8: A comparison of the total elapsed time for solving the SR problem using different schemes. The absolute increased time (independent of RLC) required by PPR is reported. The SR based on l_0 minimization (syndrome decoding) performs faster than the other two schemes based on l_1 minimization (Kronecker CS and per-column CS).

tion we have used the SparseLab software package and the basic pursuit algorithm [119]. In each figure, two different bit error probabilities are considered ($\varepsilon = 1 \times 10^{-3}$ and $\varepsilon = 4 \times 10^{-3}$). As can be observed, syndrome decoding is much faster than both the CS approaches based on l_1 minimization in all the scenarios except when a long packet and generation size is used in a severe channel (Fig. 6.8(d)). As a rule of thumb, the aver-

responding values of the total processing time are large. Notwithstanding, the relative values are important and we compare the values with each other.

age processing time corresponding to syndrome decoding and Kronecker CS increases with the bit error probability of the channel. Therefore, the processing time reduces when the channel quality improves. This is in contrast with per-column CS since it is independent of the number of errors and is only based on the sparsity order across each error column. Fixing G and comparing the results reveals that increasing the packet length, L, will increase the computational complexity in all the scenarios, which is a natural consequence due to the increased size of the search. The amount of the increase across different approaches is however different. For per-column CS, a linear increase can be observed. More explicitly, the processing time is increased by a factor of four, which is proportional to the increase of the packet length from $L = 256$ to $L = 1024$. However, for Kronecker CS and Syndrome decoding, the increase rate is more than linear. Next we fix L and compare the results to quantify the impact of the generation size on the processing time. Note that the processing time for per-column CS is almost unchanged while it is significantly affected in Kronecker CS and is moderately changed in syndrome decoding. Again, the amount of the increase in the processing time is more than linear in both cases. In summary, for moderate and desirable channel conditions, the syndrome decoding approach outperforms the other two schemes, independent of the generation size and the packet length. Otherwise, per-column CS is the best approach since it is not sensitive to the generation size and the channel error rate. The required computational complexity for the proposed PPR to achieve a high gain can be considerable. Indeed, computational complexity is our only means to make the gains achievable in the highly constrained setup we consider. The complexity is asymmetric (mostly at the receiver side). For applications such as WBAN where transmit node complexity (size) and energy are very limited and a large number of packets are partially corrupted, it is very promising. Note that the computational complexity can be adjusted with respect to the available resources through the assumed sparsity order.

6.6 Conclusions

A novel scheme based on compressive sensing and sparse recovery has been developed to tackle the partial packet recovery problem in a random linear coding setup. This scheme does not require any cross-layer information, works transparently to the other layers, does not require extra overhead and generally does not impose any new constraint on the transmitter node. The receiver first decouples the message from the errors by multiplying the received packet by a matrix and then tries to recover the partial packets using integer programming and sparse recovery techniques. The probabilities of successful recovery have been presented and the performance gains in comparison with the conventional approach have been illustrated by simulation. Based on the simulation results, the average required number of transmit packets for successful reception of the message blocks can be substantially reduced by using this approach, which can yield a significant saving in energy and throughput at the transmitter node.

Chapter 7

Coding Techniques based on Polar Codes

7.1 Introduction

Due to the limited resources in the Internet of things (IoT) devices or nodes in a WBAN, the coding schemes should impose low complexity and have high reliability and high efficiency. Polar codes [120] can provably achieve the capacity of the binary symmetric channel with complexity $\mathcal{O}(n \log n)$. These features cannot be found in other codes at the same time. For example, LDPC codes can only approach the channel capacity with polynomial complexity. Hence there is a high motivation to deploy polar codes in IoT applications where low complexity and high transmission efficiency are of utmost importance.

A binary polar code is identified by the two-tuple $(\mathcal{I}, \mathbf{u}_{\mathcal{F}})$, and the encoding

$$\mathbf{x} = \mathbf{u}\mathbf{G}, \tag{7.1}$$

where $\mathbf{u}$ and $\mathbf{x}$ are the message and code vectors, $\mathbf{G}$ is the encoding matrix of size $N \times N$, and $\mathbf{G} = \mathbf{\Pi}\mathbf{F}^{\otimes n}$, in which $\mathbf{F} \triangleq \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, $n = \log_2(N)$, $\mathbf{F}^{\otimes n}$ is the nth Kronecker power of $\mathbf{F}$ [120], and $\mathbf{\Pi}$ is the bit-reversal permutation

mapping matrix possessing the commutative property, i.e. $\mathbf{G} = \mathbf{F}^{\otimes n}\mathbf{\Pi}$[1]. Only K elements in $\mathbf{u}$ can be substituted with the message bits whose indices are identified in the *information* set $\mathcal{I}$, where $1 \leq K < N$ (i.e. $|\mathcal{I}| = K$). The remaining indices in $\mathbf{u}$ are denoted by $\mathbf{u}_{\mathcal{F}}$ and have fixed or *frozen* bit values (0 or 1), known by both encoder and decoder, where $\mathcal{F}$ denotes the complementary set of $\mathcal{I}$, known as the *frozen* set.

7.2 Joint Source and Channel Coding for Smooth Distortion

The classical approach in the design of communication systems addresses the source and channel coding problems independently. In this way, the codes are constructed very simpler and can be optimized separately. In addition, exclusive codes that match well with different source and channel models can be used together without interfering with each other.

In this approach, normally, source-encoded message blocks at the transmitter side are first encoded by an error detection code such as a cyclic redundancy check (CRC) at higher layers and then are equally protected against channel errors using a forward error correction (FEC) code in the physical layer. At the receiver side, the message blocks are first recovered by the FEC decoder and then the integrity of the received frame is verified using the error detection checksums. If the check sequence verifies the frame, it is error free and can be forwarded to continue the process to the application layer. Otherwise, the frame is either discarded or is recovered using further redundancy from the transmitter in the succeeding transmissions. In this way, however, the system tends to break down completely when the channel quality falls below a certain threshold, since the channel code can only resolve a limited number of errors, known as correction capability. This property which is called the "threshold effect" [93] is detrimental in real-time applications, e.g. communication of voice and video, where re-

[1]Another representation of polar encoding has also been adopted that does not include the bit reversal permutation matrix. For simplicity, we will use the former representation when dealing with non-systematic codes and the latter with systematic codes.

transmission of the corrupted data is practically ineffective, or when the channel is highly dynamic and volatile over time. Two common examples of the latter are the channels in mobile networks or in wireless body area networks (WBANs).

In several applications it is desirable to reliably convey even a slightly distorted version of the signal that can represent the essential characteristics of the original signal, rather than an "all-or-nothing" situation. For instance, consider a wireless sensor node that is attached to body for movement detection in a gaming application, or even a monitoring system that collects sensor measurements for anomaly (spike) detection. Indeed, the general *pattern* of the signal along with a good *approximation* of the original signal is adequate here. This *acceptable* level of distortion is usually influenced by how the message is being interpreted at the endpoint (e.g. when the existence of distortion is not perceptible by the human visual or hearing system).

For this class of applications, usually, energy and bandwidth efficiency are of crucial importance as well. For stationary sources, entropy coding can achieve optimal compression rates, which can eventually reduce the total required transmission energy when the channel quality is good. However, in the presence of unresolved channel errors any loss of the entropy-coded data propagates the error into bursts of the original data. The amount of consumed energy in the transmit node is directly proportional to the number of transmitted bits, and in fact packet loss can be interpreted as loss of energy. Therefore, energy efficiency is only achievable when the channel quality is desirable. However, it is not often the case in such applications due to factors such as low transmit power, interference, and signal propagation inconsistencies related to body randomness. In addition, a significant performance degradation can occur for sources with non-stationary and non-ergodic traits. To prevent these "catastrophic" effects, it is proposed to remove the entropy coding in such scenarios [121].

Although numerous studies have been performed separately in both directions of wavelet transforms for source coding and error correcting codes, only a few contributions can be found on the connections of these two

tracks of research. The authors in [18] reveal the necessary and sufficient conditions of the basis functions over a finite field in order to construct an orthogonal wavelet transform that can be deployed by an error correcting code. Later in [19] they proposed a code construction algorithm based on the finite field wavelets for arbitrary rate maximum separable block codes using spectral constraints. In the sequel, we combine finite field wavelets in a Galois field of order 2 denoted by $GF(2)$ with polar codes to generate a new class of lossy/lossless joint source-channel codes. This approach can be used to construct joint source-channel codes of rate 1 (i.e. the code length is equal to the message length) that can provide an approximation of the message at low SNRs.

7.2.1 Binary Field Transform

When data is in the binary format, for several reasons (such as reducing computational complexities and avoiding an extension in the range of the variables), it is more convenient to apply transforms in the binary domain using binary field (modular) arithmetic. This issue has been addressed in the study of finite field Fourier transforms in [122]. There are a few limitations that introduce difficulties in dealing with finite field transforms, specially in $GF(2)$ [123]. For instance, an element of order N is necessary for the existence of discrete Fourier transform (DFT) of a length N signal [124]. Since $GF(2)$ has only one element of order 1, DFT is only defined for a length 1 sequence in $GF(2)$. Indeed, a DFT of even length does not exist over $GF(2^m)$ for any m. These difficulties hold also for other Fourier-related transforms such as the discrete cosine transform (DCT), discrete Hartley transform (DHT), etc.

For these reasons, non-Fourier based transforms have been considered in the design of binary field transforms [125], [126]. Specifically, a linear binary field transform (BFT) based on two-band perfect reconstruction filter banks in $GF(2)$ is proposed in [127] that is analogous to the Walsh-Hadamard transform (WHT) for real-value signals. Similarly to the Fourier-based transforms that decompose the spectral characteristics of the

input signal to different "frequencies", the BFT can obtain the "sequency" information of the input waveform. For length N signals, the basis vectors of this transformation constitute the columns of the transform matrix B_N, which are rectangular waveforms taking binary values {0,1} with varying sequencies. Note that, since the determinant of a WHT matrix is even, a naive replacement of -1's by 0 in the transform matrix of WHT leads to a singular matrix in $GF(2)$ and a non-invertible transform. Instead, the $N \times N$ matrix $\mathbf{B}_N$ is defined recursively as follows:

$$\mathbf{B}_2 \triangleq \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \tag{7.2}$$

$$\mathbf{B}_4 \triangleq \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}. \tag{7.3}$$

For $N \geq 6$ and even, the construction is based on four submatrices

$$\mathbf{B}_N = \begin{bmatrix} \mathbf{B}_N^{ul} & \mathbf{B}_N^{ur} \\ \mathbf{B}_N^{ll} & \mathbf{B}_N^{lr} \end{bmatrix}, \tag{7.4}$$

where $\mathbf{B}_N^{ul}$ is the upper-left submatrix of size $(N-2) \times (N-2)$ defined as

$$\mathbf{B}_N^{ul} = \begin{bmatrix} \mathbf{1}_{2\times 2} & \mathbf{1}_{2\times(N-4)} \\ \mathbf{1}_{(N-4)\times 2} & \overline{\mathbf{B}}_{(N-4)} \end{bmatrix}, \tag{7.5}$$

in which the subscripts are the size of the matrices, $\mathbf{1}$ denotes a matrix with all elements equal to one, and $\overline{\mathbf{B}}_{(N-4)}$ represents the complementary matrix with all elements of $\mathbf{B}_{(N-4)}$ logically flipped. The lower-left and the upper-right submatrices of sizes $2 \times (N-2)$ and $(N-2) \times 2$ are defined as

$$\mathbf{B}_N^{ll} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{bmatrix}, \tag{7.6}$$

and $\mathbf{B}_N^{ur} = \mathbf{B}_N^{ll\,T}$, respectively. Finally, the lower right submatrix $\mathbf{B}_N^{lr} = \mathbf{B}_2$ is of fixed size 2×2. It can be shown that $\mathbf{B}_N$ is invertible over $GF(2)$ and its inverse can be found recursively [127].

7.2.2 Binary Field Filtering

Binary field filtering is performed using circular convolutions. A simple representational way for this approach is to use circulant matrices. By definition, a one-circulant matrix $\mathbf{X}_1 = 1\text{-circ}(\mathbf{x})$ is a square matrix such that its first row is equal to $\mathbf{x}$ and the next rows are created by a right circular shift of the previous row. This way, the circular convolution of two vectors can be rewritten as the multiplication of a vector and a circulant matrix. Similarly, a 2-circulant matrix for a vector $\mathbf{x}$ of even length N is a matrix $\mathbf{X}_2 = 2\text{-circ}(\mathbf{x})$ of size $N/2 \times N$, where each row is a two circular shifted version of the previous row. The 2-circulant form is a simple representation for a combined filtering and down sampling operation in a filter bank.

Generally, to apply a specific filter $\mathbf{h}$ (high-pass or low-pass) of length N, one needs to first evaluate a circular convolution of the filter on all basis vectors of a BFT. This is denoted by

$$\tilde{\mathbf{H}} = \mathbf{H}\mathbf{B}_N, \tag{7.7}$$

where $\mathbf{H} = 1\text{-circ}(\mathbf{h})$. A filter of higher length can be constructed by zero-padding of $\mathbf{h}$.

7.2.3 Encoding Scheme

The binary wavelet transform represented by the perfect reconstruction low-pass and high-pass filters $\mathbf{l}$ and $\mathbf{h}$ can be given by

$$\hat{\mathbf{x}} = \mathbf{T}\mathbf{x}, \tag{7.8}$$

in which

$$\mathbf{T} = \begin{bmatrix} \tilde{\mathbf{L}}_2 \\ \tilde{\mathbf{H}}_2 \end{bmatrix}, \tag{7.9}$$

and

$$\tilde{\mathbf{L}}_2 = \mathbf{L}_2\mathbf{B}_N, \tag{7.10}$$

$$\tilde{\mathbf{H}}_2 = \mathbf{H}_2\mathbf{B}_N, \tag{7.11}$$

are the BFTs of the analysis filters.

7.2.4 Codes of Length 8

We present the proposed scheme through an intuitive example for length $N = 8$ and then present the general case. Assume an input binary vector $\mathbf{x} = [x_1, \cdots, x_8]^T$ and the BFT matrix

$$\mathbf{B}_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}. \tag{7.12}$$

The columns of this matrix are the basis vectors and each column represents a component referred to as 'sequency' which is analogous to 'frequency' in Fourier-related transforms. For instance, the first column can be treated as zero or DC-sequency. The filters $\mathbf{l}$ and $\mathbf{h}$ can be designed according to three constraints: bandwidth, vanishing moment, and perfect reconstruction. Let's select $\mathbf{l} = [1\,1\,1\,0\,1\,0\,1\,0]$ and $\mathbf{h} = [1\,1\,1\,1\,1\,1\,0\,0]$ [127]. The transform matrix for this choice of filters is

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}, \tag{7.13}$$

and the inverse transform matrix is $\mathbf{T}^{-1}$. More details about the calculation of the inverse transform will be presented in the next section.

The bit distribution in the transform domain is a function of the sequency distribution of the input waveform, as well as the specific filter bank that has been applied. Nevertheless, due to the compactness property, the number of non-zero bits is usually less than the number of zeros. In Fig. 7.1, the bit distribution of 8-bit transform domain strings $\hat{\mathbf{x}} = [\hat{x}_1, \cdots, \hat{x}_8]^T$

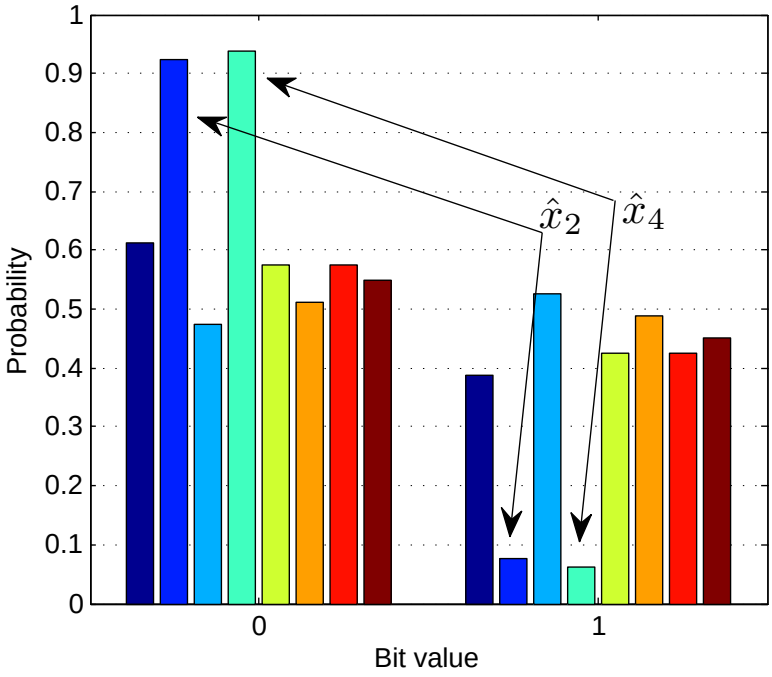


Figure 7.1: Probability distribution of the transformed domain sequences. The bars represent the distribution of bit values (0 and 1) of the transformed symbols. Uneven distribution of $\hat{x}_2$ and $\hat{x}_4$ bits is suitable for lossy compression.

corresponding to a test signal is illustrated. It can be seen that the signal contains minimum spectral strength in terms of the sequencies [1 1 1 1 1 1 0 0] and [1 1 0 0 1 1 0 0] (i.e. the second and fourth columns of $\mathbf{B}_8$) which leads to a very low probability of 1's for $\hat{x}_2$ and $\hat{x}_4$. Therefore, they can be approximated by zero. We denote the set of transform-domain bits for which this property holds by $\mathcal{K}^c$ and the set of remaining bits by $\mathcal{K}$. Particularly, in this case $\mathcal{K}^c = \{\hat{x}_2, \hat{x}_4\}$, $\mathcal{K} = \{\hat{x}_1, \hat{x}_3, \hat{x}_5, \hat{x}_6, \hat{x}_7, \hat{x}_8\}$, and the transform-domain representation of the vector $\mathbf{x}$ can be approximated by

$$\hat{\mathbf{x}}_q = \mathbf{Q}\hat{\mathbf{x}} = \begin{bmatrix} x_1 + x_3 + x_4 + x_5 \\ 0 \\ x_1 + x_4 \\ 0 \\ x_5 + x_7 + x_8 \\ x_5 + x_8 \\ x_6 + x_8 \\ x_5 + x_6 + x_8 \end{bmatrix}. \tag{7.14}$$

where $\mathbf{Q}_{(8\times 8)} = \{q_{i,j}\}$ is the quantization and mapping matrix. So when no quantization is performed (i.e. lossless compression is performed), this matrix is only required for index mapping. Note that in this example the input bit x_2 is automatically discarded in the approximated transform. Our further analysis reveals that this input bit is highly correlated with x_6. We denote $|\mathcal{K}|$ by K, which is the number of elements in $\mathcal{K}$.

Assuming an $(N, K, \mathcal{F})$ polar code, the joint source-channel code is constructed according to the following set assignments:

$$\mathcal{F} = \mathcal{K}^c, \tag{7.15}$$

$$\mathcal{I} = \mathcal{K}. \tag{7.16}$$

This can be done by using a function that maps indices in $\mathcal{I}$ and $\mathcal{K}^c$ to indices in $\mathcal{K}$ and $\mathcal{K}^c$. There are $K!$ combinations (feasible answers for $\mathbf{Q}$) such that $\mathcal{I} = \mathcal{K}$. One simple approach to find $\mathbf{Q}$ can be

$$q_{i,j} = \begin{cases} 0, & \text{if } \ u_i \in \mathcal{F}, \ \forall i,j \\ 1, & \text{else if } \hat{x}_i \text{ has an } x_j \text{ factor}, \ \forall i,j \\ 0, & \text{otherwise} \end{cases} \tag{7.17}$$

In summary, the joint source-channel code is a polar code with

$$\mathbf{u} = \hat{\mathbf{x}}_q. \tag{7.18}$$

Therefore, we can calculate the unified joint source-channel code using (7.42), (7.8), (7.14) and (7.18) in the matrix form as follows:

$$\begin{aligned} \mathbf{c} &= \mathbf{G}_N \hat{\mathbf{x}}_q \\ &= \mathbf{G}_N \mathbf{Q} \hat{\mathbf{x}} \\ &= \underbrace{\mathbf{G}_N \mathbf{Q} \mathbf{T}}_{\mathbf{W}_N^{jsc}} \mathbf{x}, \end{aligned} \tag{7.19}$$

in which $\mathbf{W}_N^{jsc}$ is the joint source-channel coding matrix for a signal of length N. The factor graph representation of this transform corresponding to our 8-bit code is depicted in Fig. 7.2.

.

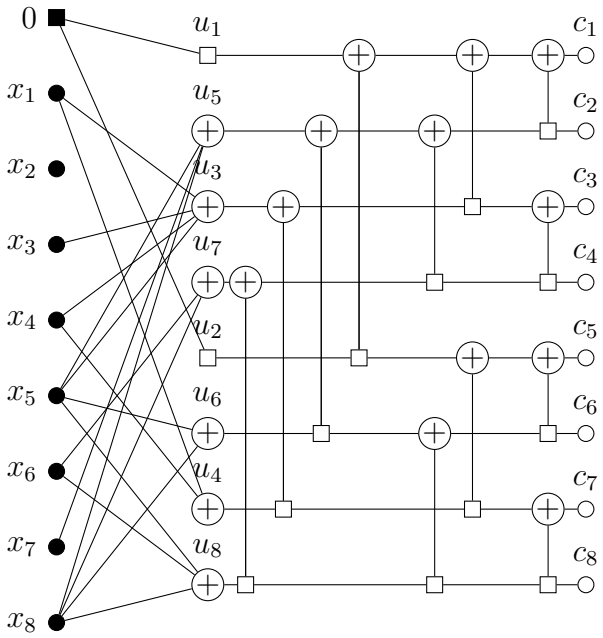


Figure 7.2: The factor graph representation of the joint source-channel code.

7.2.5 Length-N Codes

We mainly use the same theoretical results in [127] to formulate and solve the problem. As was mentioned earlier, $\mathbf{l}$ and $\mathbf{h}$ filters of higher lengths can be designed by padding zeros to the filters of lower lengths. This approach can significantly reduce the search space. However, when the base filter is too short in comparison with N, the filters designed by this approach might not necessarily perform well in terms of the ability to generate an uneven distribution in the transform domain. Therefore, it is necessary to first select a suitable length for the base filters. Nevertheless, derivation of the inverse transform is not straightforward and we present a more practical alternative here. The design procedure is based on a set of constraints on the bandwidth and the vanishing moments of the filters as well as the determinant of the binary transform matrix $\mathbf{T}$ (perfect recovery). Generally, the filter $\mathbf{h}$ of length N can have either one or $N/2$ vanishing moments. By

setting the vanishing moments of $\mathbf{h}$ to a maximum, the filter can suppress more low sequency components and hence better high-pass characteristics can be achieved. This means that the binary field transform of $\mathbf{h}$, decimated by two (due to down-sampling) should produce zero strength at the first $N/2$ sequencies. Therefore, the first $N/2$ columns of the matrix $\tilde{\mathbf{H}}_2 = \mathbf{H}_2\mathbf{B}_N$ are set to zero. Identically, it suffices that [128]

$$h_{2j} = h_{2j-1},\ \forall j \in [1, ..., N/2 - 1]. \tag{7.20}$$

The inverse operation can be initiated after polar decoding of $\mathbf{c}$. Assuming that the output of the decoder is represented by the vector $\mathbf{r}$, then the estimate of the original message denoted by $\mathbf{x}'$ can be recovered by

$$\mathbf{x}' = \mathbf{T}^{-1}\mathbf{r}. \tag{7.21}$$

The synthesis filters $\mathbf{l}^*$ and $\mathbf{h}^*$ can be directly found based on the following duality:

$$\mathbf{l}^* = \mathbf{h}^- \tag{7.22}$$

$$\mathbf{h}^* = \mathbf{l}^-, \tag{7.23}$$

where $\mathbf{v}^-$ represents reverse ordering of the elements of vector $\mathbf{v} = [v_1, \cdots, v_N]$, such that $\mathbf{v}^- = [v_N, \cdots, v_1]$. This approach can significantly simplify the search algorithm even in comparison with [128]. The output of the synthesis filters are up-sampled and summed to recover the message. Assuming $\mathbf{L}_2^* = [2\text{-circ}(\mathbf{l}^*)]^T$ and $\mathbf{H}_2^* = [2\text{-circ}(\mathbf{h}^*)]^T$, it follows that the inverse transform $\mathbf{T}^*$ can be given by

$$\mathbf{T}^{-1} = [\check{\mathbf{L}}_2^* \,|\, \check{\mathbf{H}}_2^*] \tag{7.24}$$

in which

$$\check{\mathbf{L}}_2^* = \mathbf{B}_N^{-1}\mathbf{L}_2^* \tag{7.25}$$

$$\check{\mathbf{H}}_2^* = \mathbf{B}_N^{-1}\mathbf{H}_2^*. \tag{7.26}$$

By $\mathbf{T}\mathbf{T}^{-1} = \mathbf{I}_{N\times N}$, where $\mathbf{I}$ is the identity matrix, we have

$$\mathbf{L}_2\mathbf{H}_2^* = \mathbf{0}_{N/2\times N/2} \tag{7.27}$$

$$\mathbf{L}_2\mathbf{L}_2^* = \mathbf{I}_{N/2\times N/2} \tag{7.28}$$

$$\mathbf{H}_2\mathbf{L}_2^* = \mathbf{0}_{N/2\times N/2} \tag{7.29}$$

$$\mathbf{H}_2\mathbf{H}_2^* = \mathbf{I}_{N/2\times N/2}, \tag{7.30}$$

where $\mathbf{0}$ is an all-zero matrix. Equations (7.27) and (7.30) hold given (7.22) and (7.23). To satisfy (7.28) and (7.30), it suffices that

$$\mathbf{l}(\mathbf{h}^-)^T = \mathbf{h}(\mathbf{l}^-)^T = 1 \tag{7.31}$$

$$\mathbf{l}[\mathrm{cs}(\mathbf{h}^-, j)]^T = \mathbf{h}[\mathrm{cs}(\mathbf{l}^-, j)]^T = 0, \forall j \in [2, ..., N-2] \tag{7.32}$$

where $\mathrm{cs}(\mathbf{h}, j)$ is the j-bit circular shifted vector of $\mathbf{h}$. In addition, the following constraints should be satisfied [127]:

$$\begin{aligned} \sum_{i=2,\text{even}}^{N} l_i = 1, \quad \sum_{i=1,\text{odd}}^{N-1} l_i = 0 \\ \sum_{i=1,\text{odd}}^{N-1} h_i = 1, \quad \sum_{i=2,\text{even}}^{N} h_i = 1. \end{aligned} \tag{7.33}$$

Since more than one filter bank can be designed according to the above constraints, one can also incorporate new application-specific constraints. More specifically, some designs might require less computational complexity [128].

7.2.6 Codes with Adjustable Frozen Bits

For the polar code, the actual value of the frozen bits does not make any difference, and one can set them equal to either zero or one. However, the receiver should know these values to use them in the decoding algorithm. In some cases, the uneven distribution of the transformed data shows more dominance in favor of ones rather than zeros. In other words, the signal has almost full spectral strength with respect to some sequencies. In such cases, it is suitable to quantize the corresponding component to one, rather than

zero. Consistently, the frozen bits of the concatenated polar code should be selected equal to one. The benefit of this approach is that a better compression ratio as well as a stronger polar code (with a lower rate) can be used, given a maximum acceptable level of distortion.

Therefore, we should split the set $\mathcal{F}$ into two subsets, $\mathcal{F}^0$ and $\mathcal{F}^1$, that represent the set of frozen bits equal to zero and one, respectively. We can then modify (7.17) as follows:

$$q_{i,j} = \begin{cases} 0, & \text{if } \ u_i \in \mathcal{F}^0, \ \forall i,j \\ 1, & \text{else if } \ u_i \in \mathcal{F}^1, \ \forall i,j \\ 1, & \text{else if } \hat{x}_i \text{ has } x_j \text{ factor}, \ \forall i,j \\ 0, & \text{otherwise.} \end{cases} \tag{7.34}$$

Let's denote the subset of the transform domain components with zero spectral strength with $\mathcal{K}^{c,0}$ and the components with full spectral strength with $\mathcal{K}^{c,1}$. Therefore we can set $\mathcal{F}^0 = \mathcal{K}^{c,0}$, and $\mathcal{F}^1 = \mathcal{K}^{c,1}$.

7.2.7 Discussion and Simulation Results

In this section we discuss the performance of the proposed scheme in a realistic application. Assume that

$$D \triangleq \mathbb{E}\{(\mathbf{x} - \mathbf{x}')^2\}, \tag{7.35}$$

where $\mathbb{E}\{.\}$ is expectation and D is the total message distortion at the output of the receiver, which is jointly caused by the quantization of the BFT domain signal and the channel. Note that the equivalent decimal values of the samples are required here, not their binary form. In the classic design, there are always two possible outcomes. Either $\mathbf{x} = \mathbf{x}'$ or equivalently $D = 0$, or $\mathbf{x}'$ is not available (which corresponds to when the errors are more than the correction capability of the code) and D cannot be calculated. For ease of notation and illustration we assume that distortion is infinite in this case. Note that we use this notation only to illustrate the threshold effect, and any non-zero value we assume here does not make any difference in the analysis. First, we consider the source transform (BFT) alone. Due to the energy compactness property, as the length of the BFT

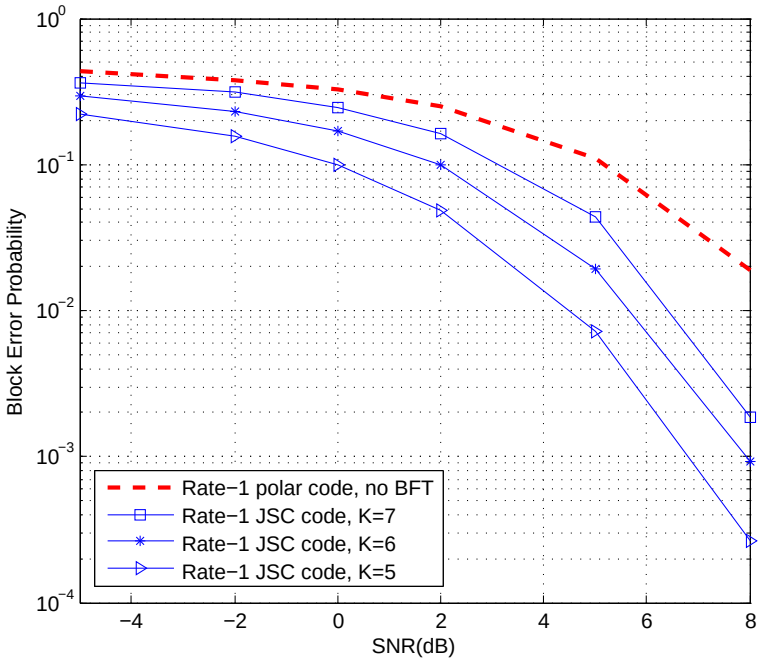


Figure 7.3: The comparison of block error probabilities for block length $N = 8$.

increases, several sequency components appear with zero (or full) spectral strength. In other words, some specific components are always equal to zero (or one) in the transform domain. This is the ideal outcome since it can eliminate the need for quantization, and the signal can be reconstructed perfectly. Here we assume that no quantization is performed (i.e. the compression is lossless). We first simulate the error performance of the system in an additive white Gaussian noise (AWGN) channel for a successive cancellation (SC) decoder. Although the performance of the SC decoder is suboptimal, we can still compare the error probabilities with and without BFT. Indeed, one can exploit more powerful decoders such as maximum likelihood (ML) or successive cancellation list (SCL) decoding [129], since the decoding of the polar code is performed independently. In the proposed scheme, we use a rate-1 code which means that the code length is equal to the message length. Assuming $N = 8$, we have simulated the block error probability of four different codes, namely a rate-1 polar code with no BFT and three JSC codes of rate 1 with K equal to 5, 6, and 7. The results are

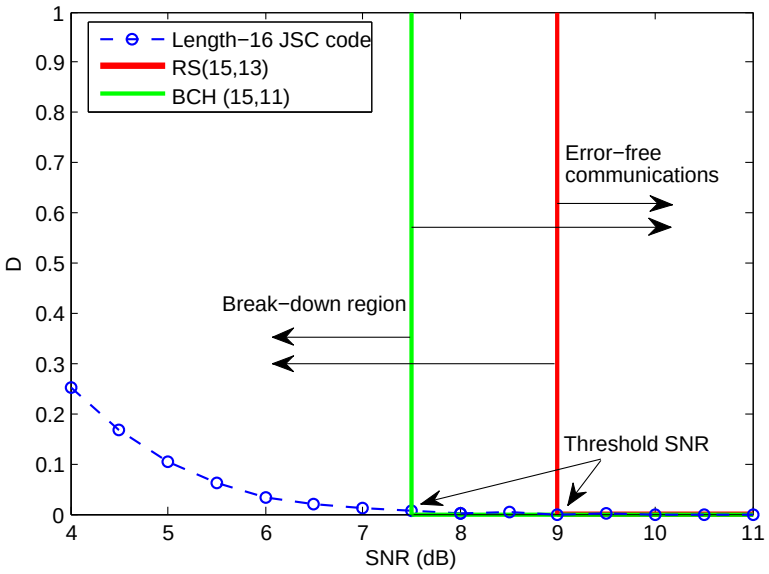


Figure 7.4: Comparison of distortion with and without JSC coding for $N = 16$. Although the actual values of distortion are signal-dependent, here we illustrate the gradual change in distortion in contrast to the idealized situation in the classic system.

illustrated in Fig. 7.3 for comparison. It can be observed that maintaining the same throughput, the performance gain of the proposed JSC code is about 2 dB at a low signal-to-noise ratio (SNR) and 1 dB at higher SNRs, for each of the $N - K$ bits. Note that the performance of the BFT depends on the statistical characteristics of the original message signal, and finding an optimal filter is an independent application-specific problem.

In the next simulation we compare the proposed scheme with the classic design in terms of the overall message distortion. As we mentioned earlier, the classic system experiences a break down at low SNRs due to the threshold effect.

The amount of distortion can be given by (7.35). However, it is difficult to represent a unified tool for comparing these two systems, and to illustrate the threshold effect we require a stochastic model. For convenience, assuming an (N, K, t) channel code in which N is the code length, K is the message length, and t is the correction capability of the code, we adopt an outage model for the distortion. More specifically, for the classic system

the threshold SNR is set at the SNR at which the distortion is equal to zero with 99% confidence (or equivalently, the SNR at which in less than 1% of the outcomes the FEC code cannot correct the errors). The JSC code is of rate 1, corresponding to $K = N$. But since there is no channel code of this rate, we use the closest feasible code for comparison here. For a code with a minimum distance d, the Singleton bound can give the maximum message length

$$K \leq N - d + 1, \tag{7.36}$$

in which $t = \left\lfloor \frac{d-1}{2} \right\rfloor$. For $N = 15$, the maximum message length is $K = 13$, corresponding to the Reed-Solomon (RS) code (15,13). Also, the BCH (15,11) code has the same correction capability but with a lower rate. These codes are extensively used in classic systems. For instance, the FEC scheme in the recent IEEE 802.15.6 standard for wireless body-area networks [10] is based on the BCH codes of lengths 63 and 127. Nevertheless, we actually compare the JSC code of rate 1 with two block codes of lower rates. Hence, the JSC code already outperforms the two block codes in terms of bandwidth. The following filters of length 16 have been used in the BFT part of the simulation:

$$\mathbf{l} = (\,0,\,0,\,0,\,0,\,0,\,0,\,1,\,0,\,1,\,0,\,1,\,1,\,0,\,0,\,1,\,0\,)^T$$
$$\mathbf{h} = (\,0,\,0,\,0,\,0,\,1,\,1,\,1,\,1,\,0,\,0,\,0,\,0,\,1,\,1,\,0,\,0\,)^T$$

The BFT of the message signal has 4 zero components, i.e. $|\mathcal{K}^{c,0}| = 4$. The end-to-end distortion (normalized to the signal energy) corresponding to the two systems is plotted versus different SNRs in Fig. 7.4. In the JSC code, the amount of distortion grows at low SNR values due to the unresolvable channel errors. Nevertheless, it is still possible to acquire an approximation of the original message. On the other hand, when the number of channel errors is more than t with high probability, the classic system is in the break down region where the decoder fails to correctly decode most of the received codewords. The threshold SNR values are 7.5 dB and 9 dB for the BCH(15,11) and RS(15,13) codes.

7.3 Simple Hybrid ARQ Schemes based on Polar Codes

Contrary to forward error correction (FEC) schemes, H-ARQ techniques can adapt the rate of transmission to the channel quality and achieve a higher efficiency. With an H-ARQ scheme based on systematic codes, an even higher level of energy efficiency can be achieved, especially at high signal-to-noise ratio (SNR), since FEC encoding and decoding is only performed when the initial systematic transmission fails. However, for systematic polar codes it is not known how to incrementally correct the errors in H-ARQ deployments. This arises from the special soft-decoding of the polar codes known as successive cancellation (SC) which is performed on a decoding graph. In this chapter, we show how a systematic polar code can be used in a type II incremental redundancy (IR) H-ARQ such that the receiver can improve the frame error rate (FER). Note that since typical sensor-based IoT devices usually demand low to medium data rates and have complexity constraints, short code lengths are considered in this chapter.

There are different Hybrid ARQ mechanisms, mostly based on non-systematic rate-compatible and punctured polar codes, that have been proposed recently. Quasi-uniform puncturing combined with repetition of information bits for H-ARQ was proposed in [130]. The performance of H-ARQ schemes based on puncturing significantly depends on the puncturing pattern. Finding optimal patterns can also impose high complexity, since they mostly rely on exhaustive search. A different strategy based on systematic codes is pursued in this chapter in which no puncturing is required. The transmitter first sends the systematic bits and then the parity segments of a polar codeword sequentially until the message is correctly decoded.

We propose and develop the theoretical framework for type-II H-ARQ based on systematic polar codes. Our first H-ARQ scheme is based on half-rate polar codes and code inversion. We also present a more general H-ARQ scheme assuming an arbitrary code rate in conjunction with calculations of

the frame error rate in polar codes with incomplete codewords.

7.3.1 Error Performance of Polar Codes with an Incomplete Codeword

Assume that an all-zero codeword is sent using antipodal signaling with unit energy per bit over a discrete memoryless zero-mean additive white Gaussian noise (AWGN) channel with variance σ^2, and that the received vector is $\mathbf{y} = (y_1, ..., y_N)$. The channel log-likelihood ratio (LLR) corresponding to a code bit y_i is $\frac{2}{\sigma^2} y_i$. To model decoding with an incomplete codeword, we assume that the set of code bits available at the receiver after a transmission t (denoted by $\mathcal{T}_t$) are transmitted through a channel with noise variance σ^2, and the rest (denoted by $\mathcal{T}_t^c$) is sent over a channel with infinite noise. In the latter case, the channel LLR values are equal to zero. This is equivalent to a likelihood ratio of one, or a bit error probability of 1/2. Therefore, choosing $\mathbf{E}^{[t]} = \{E_{i,j}^{[t]}\}$ as the average LLR matrix corresponding to the decoding graph, the mean LLR values at the first (channel) level are

$$E_{i,1}^{[t]} = \begin{cases} \frac{2}{\sigma^2} y_i & \forall i \in \mathcal{T}_t \\ 0, & \forall i \in \mathcal{T}_t^c \end{cases} \tag{7.37}$$

Note that $\mathcal{T}_t = \{1, .., N\}$ when the codeword is fully available for decoding. Since the FERs of systematic and non-systematic polar codes are equal [131], we assume non-systematic codes for FER analysis. Accurate calculation of FER requires assuming the initial probability density function of each received bit and evolving the densities in the succeeding levels of the decoding graph considering the variable and check nodes. This approach may not be tractable for arbitrary densities. However, it is straightforward for Gaussian random variables since their sum also follows a Gaussian distribution. Another useful observation here is that the channel LLR can also be approximated by a Gaussian random variable whose mean value is half of its variance. Hence, it can be recognized only with one parameter (i.e. its mean value)[132]. Different nodes on the decoding graph are divided into check nodes and variable nodes and can be distinguished tusing the

following rule

$$\text{node type} = \begin{cases} \text{check}, & \text{if } B_{i,j-1} = 1 \\ \text{variable}, & \text{if } B_{i,j-1} = 0, \end{cases} \tag{7.38}$$

where $\mathbf{B} = \{B_{i,j}\}$ is the right-MSB binary representation of the vector $(0, ..., N-1)^\intercal$, $\forall i \in [1, N]$, $j \in [2, n+1]$. It can be shown that the mean LLR values at each level using the Gaussian approximation are given by

$$E_{i,j}^{[t]} = \begin{cases} a + b & \text{check} \\ \phi^{-1}(1 - (1 - \phi(a))(1 - \phi(b))) & \text{variable}, \end{cases} \tag{7.39}$$

where $a = E_{i,j-1}^{[t]}$, $b = E_{i+2^{j-2},j-1}^{[t]}$, and $\phi(x)$ can be modeled by the exponential fit [132]

$$\phi(x) = \begin{cases} \exp\left(0.4527x^{0.86} + 0.0218\right), & \text{if } 0 < x \leq 10 \\ \frac{\sqrt{\pi}}{2} \exp\left(-\frac{x}{4}\right)\left(1 - \frac{10}{7x}\right), & \text{if } x > 10. \end{cases} \tag{7.40}$$

On the assumption that the previous bits are correctly decoded, the approximated error probability of the ith bit is given by

$$\varepsilon_i^{[t]} = \mathcal{Q}(\sqrt{0.5\tilde{e}_i^{[t]}}), \tag{7.41}$$

where $\tilde{\mathbf{e}}^{[t]} = (\tilde{e}_1^{[t]}, ..., \tilde{e}_N^{[t]})^\intercal = \pi(E_{:,n+1}^{[t]})$ is the mapped version of the $(n+1)$th column of $\mathbf{E}^{[t]}$ under bit-reversal permutation mapping π, and $\mathcal{Q}$ is the tail probability of the standard Gaussian density. The FER can be given by

$$\mathsf{FER}^{[t]} = 1 - \prod_{\forall i \in \mathcal{I}} (1 - \varepsilon_i^{[t]}). \tag{7.42}$$

Similarly to [131], it would be useful to split the message vector $\mathbf{u} = (\mathbf{u}_\mathcal{I}, \mathbf{u}_\mathcal{F})$ to consider separately the information and the frozen bits. Note that the frozen bits are not necessarily in a separate part of the message vector and can be distributed among different indices. The polar encoding procedure (7.1) can be written as

$$\mathbf{x} = \mathbf{u}_\mathcal{I}\mathbf{G}_\mathcal{I} + \mathbf{u}_\mathcal{F}\mathbf{G}_\mathcal{F}, \tag{7.43}$$

in which $\mathbf{G}_{\mathcal{I}}$ and $\mathbf{G}_{\mathcal{F}}$ are submatrices of $\mathbf{G}$, with rows specified by $\mathbf{u}_{\mathcal{I}}$ and $\mathbf{u}_{\mathcal{F}}$. Similarly, assuming that the systematic and parity bits of the codeword $\mathbf{x}$ belong to the sets $\mathcal{S}$ and $\mathcal{P}$ respectively, we can split the codeword $\mathbf{x}$ into systematic and parity bits $\mathbf{x}_{\mathcal{S}}$ and $\mathbf{x}_{\mathcal{P}}$, and rewrite (7.43) as follows [131]:

$$\mathbf{x}_{\mathcal{S}} = \mathbf{u}_{\mathcal{I}}\mathbf{G}_{\mathcal{I}\mathcal{S}} + \mathbf{u}_{\mathcal{F}}\mathbf{G}_{\mathcal{F}\mathcal{S}}, \tag{7.44}$$

$$\mathbf{x}_{\mathcal{P}} = \mathbf{u}_{\mathcal{I}}\mathbf{G}_{\mathcal{I}\mathcal{P}} + \mathbf{u}_{\mathcal{F}}\mathbf{G}_{\mathcal{F}\mathcal{P}}, \tag{7.45}$$

where $\mathbf{G}_{\mathcal{I}\mathcal{S}}$ is a sub-matrix of $\mathbf{G}$ with rows and columns specified by $\mathcal{I}$ and $\mathcal{S}$, respectively, and the other three sub-matrices of $\mathbf{G}$ are defined similarly. According to [131], the necessary and sufficient conditions for the existence of a systematic encoder is a one-to-one correspondence between the values of $\mathbf{u}_{\mathcal{I}}$ and $\mathbf{x}_{\mathcal{S}}$, which is achieved if and only if the two sets $\mathcal{I}$ and $\mathcal{S}$ have the same number of elements and $\mathbf{G}_{\mathcal{I}\mathcal{S}}$ is invertible. A straightforward choice that satisfies the above condition and is assumed in the rest of this chapter is to select

$$\mathcal{S} = \boldsymbol{\mu}_{\mathcal{I}}, \tag{7.46}$$

where $\boldsymbol{\mu} = (1, 2, ..., N)\boldsymbol{\Pi}$ is a vector of the image of the input indices under the mapping $\boldsymbol{\Pi}$.

The conventional encoding (7.1) leads to non-systematic codewords, in the sense that the message bits do not necessarily appear in the polar codeword. Hence, a reverse approach is used to construct systematic codewords in which the beginning point is the codeword $\mathbf{x}$, or more precisely, the systematic bits $\mathbf{x}_{\mathcal{S}}$. For a given message, $\mathbf{x}_{\mathcal{S}}$ would be considered known and the vector $\mathbf{u}_{\mathcal{I}}$ calculated using (7.44) as

$$\mathbf{u}_{\mathcal{I}} = \left(\mathbf{x}_{\mathcal{S}} - \mathbf{u}_{\mathcal{F}}\mathbf{G}_{\mathcal{F}\mathcal{S}}\right)\left(\mathbf{G}_{\mathcal{I}\mathcal{S}}\right)^{-1}, \tag{7.47}$$

and the parity bits calculated by substituting $\mathbf{u}_{\mathcal{I}}$ in (7.45).

7.3.2 Simple H-ARQ Schemes Based on Polar Codes

The inversion process in an FEC code refers to the recovery of message bits from redundant bits only [133]. For polar codes, it can be performed

similarly to what presented in Section 7.3.1 for constructing systematic polar codes, however, for a very different purpose and in a different place. Contrary to a systematic encoder in which equations (7.44) and (7.45) are used at the transmitter, we use these equations at the receiver side. The other slight difference is that, instead of beginning with (7.44), here we first use (7.45) to calculate

$$\mathbf{u}_{\mathcal{I}} = (\mathbf{x}_{\mathcal{P}} - \mathbf{u}_{\mathcal{F}}\mathbf{G}_{\mathcal{F}\mathcal{P}})\,(\mathbf{G}_{\mathcal{I}\mathcal{P}})^{-1}, \tag{7.48}$$

and then substitute the result in (7.44) to acquire the systematic bits.

H-ARQ with Half-rate Invertible Polar Codes

We propose a half-rate $(N, N/2)$ systematic polar code as our base FEC scheme. This choice leads to an equal length for the message and the redundancy bits. Using an error detection code $\mathbb{C}_0$ such as a cyclic redundancy check (CRC) C parity bits are appended to the codeword to verify the integrity of each of the received words.

Let $\mathbf{x}_{\mathcal{S}}$ be a message of size $N/2 - C$. At the initial stage, the transmitter temporarily saves the data bits, encodes them using $\mathbb{C}_0$ and sends the codeword of size $N/2$ through the channel. At the receiver side and by decoding $\mathbb{C}_0$, an acknowledgment (Ack) message is sent from the receiver if no error is detected. No further encoding is performed at the transmitter and ithe message proceeds to the next data block. If there was any error, the receiver temporarily stores the received vector and sends a negative Ack message. The transmitter encodes the data bits using an $(N, N/2)$ systematic polar code and then sends the redundant bits only. This time, the receiver finds the decoded message by inversion. If no error is detected after looking at the checksum, then the receiver accepts it. Otherwise, the receiver applies SC decoding using the stored received vector of the systematic bits and the available received vector of the redundant bits. If the decoding is unsuccessful again, the previously stored vector of the received systematic bits is erased and the received vector of the redundant bits is stored, so that the next potential transmission of the systematic bits could

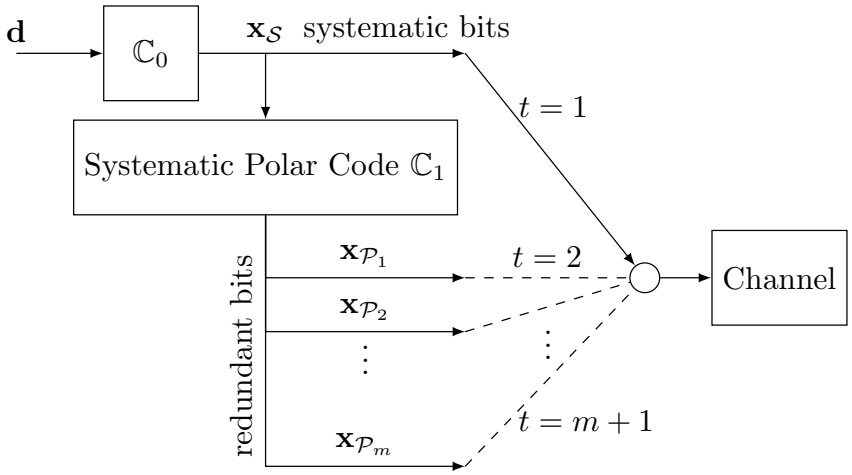


Figure 7.5: Block diagram of the proposed incremental redundancy H-ARQ.

be combined with the redundant bits for SC decoding. This procedure can continue for a pre-defined number of retransmissions which we denote by ρ.

Incremental Redundancy H-ARQ with Polar Codes

The H-ARQ scheme presented in the previous section is simple. However, it cannot use arbitrary-rate codes. In addition, the redundant bits should always be sent at once. An alternative approach is to deploy a systematic polar code of arbitrary rate $R = K/N$ (which is selected for the lowest SNR at which reliability should be guaranteed), divide the redundant bits to a number of disjoint sets, and transmit the systematic bits and then redundant segments, sequentially, until the message is successfully delivered. Our second proposed scheme starts with a message of length $K - C$ which is encoded by the error detection code $\mathbb{C}_0$. The output of length K denoted by $\mathbf{x}_{\mathcal{S}}$ is sent over the channel. The transmission of the systematic bits is denoted by the set $\mathcal{T}_1$. If the receiver detects the message in error, then the transmitter encodes $\mathbf{x}_{\mathcal{S}}$ by the systematic polar code $\mathbb{C}_1$ of rate R and sends a subset $\mathcal{P}_1$ of the redundant (parity) bits, where $r_1 = |\mathcal{P}_1|$ is the total number of bits in $\mathcal{P}_1$. We then decode the polar code using the incomplete set of LLRs which includes all systematic bits plus $\mathcal{P}_1$ such that $\mathcal{T}_2 = \mathcal{T}_1 + \mathcal{P}_1$. In case of decoding failure identified by $\mathbb{C}_0$, the transmitter sends another

disjoint subset of size r_2 of the redundant bits denoted by $\mathcal{P}_2$. This procedure ends when the $\mathbb{C}_0$ checksum is consistent, but it potentially continues until all redundant bits are sent in the final transmission m. Hence,

$$r_1 + r_2 + \cdots + r_m = N - K. \tag{7.49}$$

7.3.3 Performance Evaluation

Denote the number of successfully delivered bits in the tth transmission by d_t and the number of sent bits by s_t. The throughput efficiency is defined as $\eta \triangleq \frac{\sum_t^T d_t}{\sum_t^T s_t}$, for a large number T. d_t is a random variable taking two values $K - C$ or 0, and $s_t \in \{K, K + r_1, K + r_1 + r_2, ..., N\}$. For theretical derivation of the throughput efficiency one needs to calculate

$$\eta = \frac{\mathbb{E}\{d\}}{\mathbb{E}\{s\}}. \tag{7.50}$$

However, the exact calculation requires the conditional frame probabilities, given failure of the previous transmissions, and is not tractable. Hence, we approximate the conditional frame probabilities similarly to the approach pursued in [130]. Assume that A_t represents the event of failed reception in the tth transmission where $t \in [1, m + 1]$ and $t = 1$ corresponds to the initial (systematic) transmission. Let $\Pr(A_t)$ and $\overline{A}_t$ be the probability and the complementary event of A_t, respectively. Considering [130]

$$\Pr(\overline{A}_t \cap A_{t-1} \cap \cdots \cap A_0) \approx \Pr(A_{t-1}) - \Pr(A_t), \tag{7.51}$$

$$\Pr(A_t \cap A_{t-1} \cap \cdots \cap A_0) \approx \Pr(A_t), \tag{7.52}$$

the expected numbers of received and sent bits are as follows:

$$\begin{aligned}
\mathbb{E}\{d\} &= (K - C)(1 - \Pr(A_{m+1} \cap A_m \cap \cdots \cap A_0)) \\
&\approx (K - C)(1 - \Pr(A_{m+1})), \qquad (7.53) \\
\mathbb{E}\{s\} &= \sum_{t=1}^{m+1} (K + \sum_{j=0}^{t-1} r_j) \Pr(\overline{A}_t \cap A_{t-1} \cap \cdots \cap A_0) \\
&\quad + N \Pr(A_{m+1} \cap A_m \cap \cdots \cap A_0) \\
&\approx \sum_{t=1}^{m+1} (K + \sum_{j=0}^{t-1} r_j) [\Pr(A_{t-1}) - \Pr(A_t)] + N \Pr(A_{m+1}), \qquad (7.54)
\end{aligned}$$

where $r_0 \triangleq 0$, $\Pr(A_0) \triangleq 1$, $\Pr(A_1) = 1 - (1 - \mathcal{Q}(1/\sigma))^K$, and $\Pr(A_t) = \mathsf{FER}^{[t]}$ is calculated using (7.42) for $t \in [2, m+1]$. It should be noted that depending on the rate of the base polar code and the channel SNR, the first transmission of redundant bits does not necessarily reduce the FER for arbitrary r_1. To prevent this, r_1 should be selected such that $\Pr(A_2) < \Pr(A_1)$. Note that for the remaining transmissions, $\Pr(A_{t+1}) < \Pr(A_t)$, $\forall t \in [2, m+1]$ due to the polarization effect.

Fig. 7.6 illustrates a comparison between the Monte Carlo FER of two different polar codes of rate 1/2 and their corresponding analytical approximation assuming three different SNRs. The analytical FER matches well with the simulation. A difference can be observed at low SNR due to the inaccuracy of Gaussian approximation for LLR density at this SNR range [132]. The non-smooth appearance of the curves is due to the different levels of polarization for various bit channels.

Fig. 7.7 compares the throughput efficiency of the proposed H-ARQ schemes with that of the H-ARQ scheme in the high-quality-of-service mode of the IEEE 802.15.6-2012 standard for wireless body area networks [10], which is based on the invertible shortened BCH (126,63) code. The CRC-16-CCITT error detection code is assumed for $\mathbb{C}_0$ which is also used in [10] and appends 16 parity bits to each codeword. This will result in a probability equal to 2^{-16} for erroneous detection of errors. Note that although traditionally adopted in previous work, we do not neglect the rate loss corresponding to $\mathbb{C}_0$ (hence, the maximum achievable throughput is less than 1). The bumpy curves are typical in many H-ARQ schemes and are mainly due to different delivery mechanisms in the H-ARQ scheme, each being dominant in a particular SNR interval. A code of length 128 is assumed for a fair comparison in terms of computational complexity. The advantage of the H-ARQ scheme based on an invertible polar code at low SNR is roughly 4 dB compared to [10]. However, at high SNR it achieves a similar performance. The IR H-ARQ scheme can also achieve the 4-dB advantage at low SNR while it can outperform [10] at higher SNR range. Note that the maximum achievable throughput is lower than [10] since no error detection is performed by the polar code, hence, the H-ARQ always relies on $\mathbb{C}_0$ for

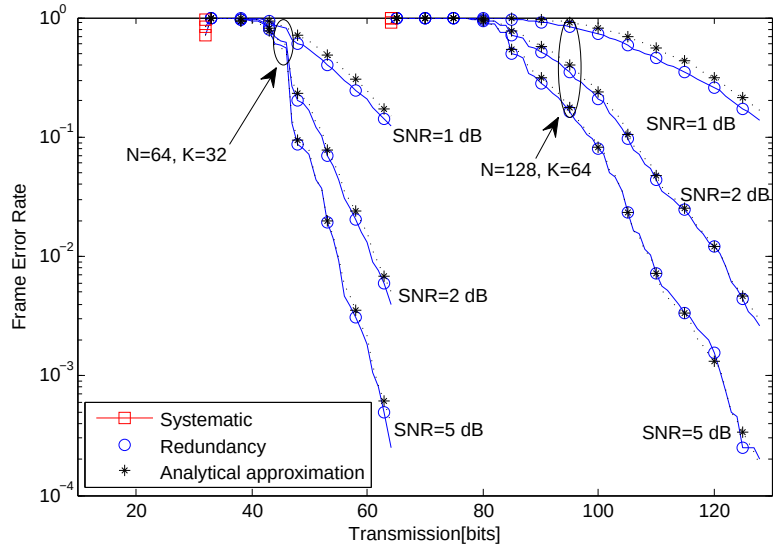


Figure 7.6: Simulated and analytical approximation of the FER of two different polar codes of rate 1/2 versus incremental transmission of the redundant bits.

this task. This can be improved by increasing the code length. Fig. 7.7 can also verify our analytical approximation for η. The capacity bound for a binary AWGN channel (BAWGNC) is also depicted for comparison.

To understand the effectiveness of inversion in polar codes, a separate simulation is performed assuming a Rayleigh fading channel, and the result is presented in Fig. 7.8. Based on this result, the throughput efficiency can be considerably improved in fading conditions when code inversion is enabled.

7.3.4 Conclusions

We proposed a novel joint source and channel coding scheme that combines BFT and polar coding. The BFT, which is a source transform, can reduce the rate of the message in order to compensate for the redundancy imposed by the polar channel code. The theoretical background regarding both BFT and polar codes is presented and then the proposed JSC framework is formulated for length-N codes. Using simulations for the AWGN channel, we show that a noticeable performance gain is achieved considering each single

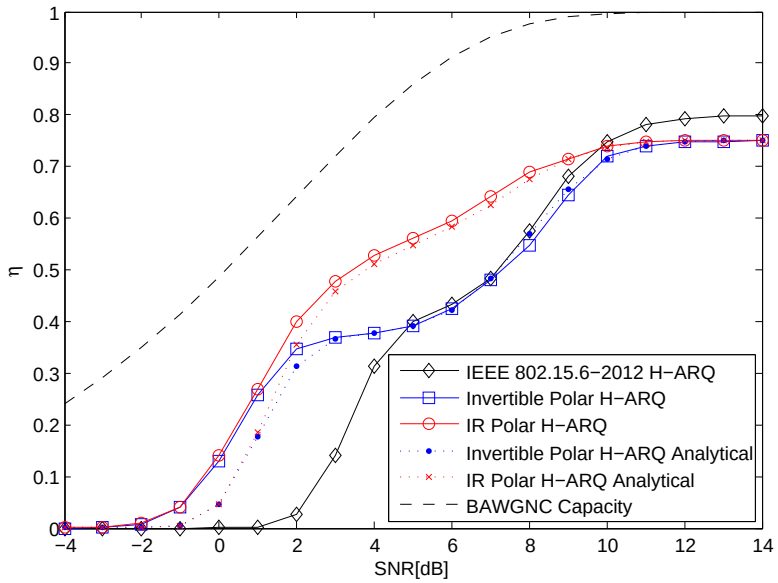


Figure 7.7: Throughput performance of the proposed H-ARQ schemes in comparison with IEEE 802.15.6-2012 H-ARQ [10] for $N = 128$ and $R = 1/2$, and the theoretical approximation.

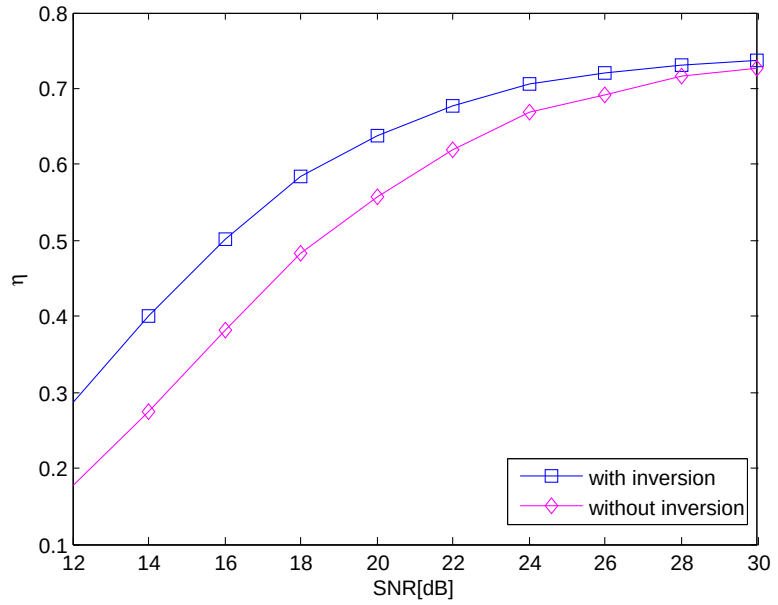


Figure 7.8: Effectiveness of code inversion against fading ($\rho = 4$).

bit compensated by the BFT. We also compare the proposed JSC coding scheme with the classic Reed-Solomon and BCH block codes of a similar rate in terms of the end-to-end distortion and showed that the JSC coding approach can provide an approximation of the original signal, while the block code is in the break down region. Other performance gains of this scheme include a higher throughput as well as a significant complexity reduction. Furthermore, this we introduced type II-IR-H-ARQ schemes for systematic polar codes. The first scheme deploys alternative message/redundancy packets combined with code inversion, which is effective at fading conditions. The second scheme can use arbitrary-rate systematic polar codes. Frame error probabilities with incomplete channel codewords are calculated and the results are verified by simulation.

Chapter 8

Conclusion and Future Work

This thesis addressed the problem of reliable communications in wireless body area networks. Since the available resources in typically tiny low-power low-complexity sensor nodes are extremely scarce, it was proposed to improve the transmission efficiency at the sensor nodes which can ultimately improve the reliability directly (by increasing the amount of successfully delivered data) or indirectly (through allocation of more node resources for error control schemes). In addition, due to the usually more computationally and energy capable hardware at the hub nodes, it was proposed to leverage this feature to achieve a higher efficiency at the sensor nodes. Several methods based on the theory of statistical signal processing were proposed to improve transmission efficiency.

At the sensor level as well as the physical layer of the protocol stack, non-uniform sampling and joint source-channel coding schemes were proposed. It was shown that the proposed joint source-channel code can achieve smooth distortion to improve robustness against the threshold effect in digital systems and the results were compared with the state-of-the-art codes in IEEE 802.15.6 standard. While the classic approach based on FEC codes breaks down below the threshold SNR, the proposed joint source-channel code can enable decoding and represent a distorted version of the original signal, where the level of distortion increases ans the SNR is reduced. Furthermore, non-uniform sampling was proposed to represent a given chunk

of band-limited signal with fewer number of bits. It was confirmed by simulations that the proposed scheme can reduce the total number of bits down to 88% compared to the conventional uniform sampling, without using any compression schemes and with almost imperceptible visual effects on the ECG test signal.

At the MAC/link layer, optimization schemes were deployed for link adaptation as well as frame optimization. It was shown that link adaptation can improve the transmission efficiency even close to twice the efficiency of the static UWB transmission in the IEEE 802.15.6 standard, both in the default and high QoS modes. Novel estimators that are deployed at the receiver side and don't require channel feedback from the transmitter are proposed for SNR estimation which is require for link adaptation. A closed form expression was derived for the optimal frame length to achieve optimal energy efficiency in the default mode of the IEEE 802.15.6 and and the optimal length for the high QoS mode was derived using numerical optimization. Additionally, the optimal energy-delay tradeoffs for queue stability as well as the minimum energy function under UWB physical layer of this standard were derived. The scheduling policy that achieves minimum energy bound for noncoherent UWB communication was also derived.

At the network/data-link layer, novel error control coding schemes were proposed that rely on low complexity at the sensor nodes. More specifically, a partial packet recovery technique was proposed that benefits from the sparsity of errors in partially corrupted packets and can correct residual packet errors by exploiting random linear network coding combined with compressive sensing. Based on the simulation results, it was concluded that the proposed scheme not only improves the transmission efficiency in the multicast scenario (up to 50% in a typical setting), but is can also improve the performance of random linear coding in point-to-point settings where it doesn't have any advantage alone.

Motivated by the capacity-achieving and low-complexity features of the recently discovered polar codes, their application for low-complexity IoT and WBAN devices was considered and novel coding techniques were proposed. In this regard, simple hybrid ARQ schemes based on systematic

polar codes were proposed that can achieve significantly higher efficiencies compared to the state-of-the-art technique in the high QoS transmission mode of the IEEE 802.15.6 standard at low SNR. Simulations confirmed that a gain of up to 4 dB in the SNR can be achieved. Additionally, it was shown that by introducing code inversion the proposed hybrid ARQ scheme is effective against random fading conditions.

In addition to the novel techniques, this thesis presents new topics for future research. There are many aspects in each of the proposed methods that can be investigated and improved further. For nonuniform sampling, the suggested works for future include more advanced nonuniform sampling techniques that can consider the statistical feature of the signal. More clearly, an interesting problem is context modeling as well as optimal bit depth and sampling time allocation based on the specific context. More complex prediction functions may lead to even more reduces set of bits to represent a segment of band-limited signal. There is also room for more investigation for more computationally effective error control schemes that can consider more side information for packet recovery. Indeed, combining the existing knowledge at the receiver side can be leveraged to pinpoint the residual errors. Furthermore, based on the promising results gained in this thesis, advanced coding schemes based on polar codes for IoT applications that consider short code lengths have a high potential to be considered for the next generation of IoT devices. Hence, cross-layer optimization techniques based on these codes remain an open research topic.

Appendix A

Calculation of Energy Efficiency

In this appendix we extend the energy efficiency model proposed in [20] to the cases with different energy costs for the receiver and the transmitter. In fact, the energy resources of the sensors are extremely limited, and they are required to operate for a long time. On the other hand, battery recharging is a convenient task for a hub. Therefore, the hub and sensors have different energy consumption costs. Clearly, the energy consumption cost of transmission on the uplink (sensor $\rightarrow$ hub) is higher, so is the energy consumption cost of reception on the downlink.

Suppose that l encoded bits are transmitted from a sensor to the hub within a data packet with $\bar{\tau} = \sum_i \tau_i$ overhead bits, where τ_i is the length of overhead data of rate R_{τ_i}. This model accounts for different types of overhead data including synchronization bits, header data, or even MAC layer headers and check sequences. The energy required to communicate one bit of information is given by

$$\begin{aligned} E_b &= c_{tu}(\beta E_{tb} + E_{enc}/l) + c_{ru}(E_{rb} + E_{dec}/l) \\ &\quad + c_{td}E_{t-ack}/l + c_{rd}E_{r-ack}/l, \end{aligned} \tag{A.1}$$

in which E_{enc} and E_{dec} are the required energies for data encoding/decoding, and E_{t-ack}, and E_{r-ack} are the energies required for the acknowledgment packets. The constants c_{tu}, c_{ru}, c_{td}, and c_{rd} are energy consumption costs for transmission and reception on the uplink and downlink respectively. β

depends on the type of modulation and is 1/2 for on-off signaling and 1 otherwise. E_{tb} and E_{rb} are the energy consumption for transmission and reception of a bit of data, and can be written as

$$E_{tb} = \frac{1}{l}(N_p \epsilon_{tp} \frac{l}{R_b} + N_p \epsilon_{tp} \sum_i \frac{\tau_i}{R_{\tau_i}} + E_{st}), \tag{A.2}$$

$$E_{rb} = \frac{1}{l}(N_p \epsilon_{rp} \frac{l}{R_b} + N_p \epsilon_{rp} \sum_i \frac{\tau_i}{R_{\tau_i}} + E_{sr}), \tag{A.3}$$

where N_p is the number of pulses per symbol, ϵ_{tp} is the total energy spent on transmission of one pulse, consisting of radiation energy and the processing energy of the electronic circuit, ϵ_{rp} is the processing energy per pulse consumed by the electronic circuits of the receiver, and R_b is the coding rate of the data payload.

Typical values for these parameters can be found in [134]-[135]. We use $\epsilon_{tp} = 20$ pJ and $\epsilon_{rp} = 2.5$ nJ. The required startup energies for transmitter and receiver, E_{st} and E_{sr}, can be approximated using the fact that $k \geq 1$ empty packets (e.g., overheard bits only) are required for synchronization. This will also give an estimate for the startup energies as

$$E_{st} = \epsilon_{tp} N_p \times k\bar{\tau}, \tag{A.4}$$

$$E_{sr} = \epsilon_{rp} N_p \times k\bar{\tau}. \tag{A.5}$$

Equation (A.1) can be re-written in terms of parameters

$$\varepsilon_1 = (c_{tu}\beta\epsilon_{tp} + c_{ru}\epsilon_{rp})N_p/R_b, \tag{A.6}$$

$$\varepsilon_2 = c_{tu}\beta E_{st} + c_{ru}E_{sr} = N_p k\bar{\tau}(c_{tu}\beta\epsilon_{tp} + c_{ru}\epsilon_{rp}), \tag{A.7}$$

and $E_0 = c_{tu}E_{enc} + c_{ru}E_{dec} + c_{td}E_{t-ack} + c_{rd}E_{r-ack}$ as

$$E_b = \varepsilon_1 + \varepsilon_1 \frac{R_b}{l} \sum_i \frac{\tau_i}{R_{\tau_i}} + \frac{\varepsilon_2 + E_0}{l}, \tag{A.8}$$

where ε_1 represents the useful energy for transmission of one bit, and ε_2 is the startup energy consumption for the uplink. From (A.8), energy consumption is proportional to the inverse of packet length l. So it can be reduced by choosing a longer frame length. On the other hand, channel

errors and packet losses are more likely when the frames are longer. Therefore, there is an optimal frame length for a given error probability that can minimize loss of energy. Energy efficiency is defined as the ratio of useful energy for successful communication of l bits to the total consumed energy and is expressed as

$$\eta = \frac{\varepsilon_1 l}{\varepsilon_1(l + \sum_i \tau_i(R_b/R_{\tau_i})) + \varepsilon_2 + E_0} P_{PPDU}^{(m)}, \tag{A.9}$$

where $P_{PPDU}^{(m)}$ is the packet success rate given by (3.7) and (3.10).

References

[1] S. Ullah, P. Khan, N. Ullah, S. Saleem, H. Higgins, and K. S. Kwak, "A review of wireless body area networks for medical applications," *arXiv preprint arXiv:1001.0831*, 2010.

[2] M. Patel and J. Wang, "Applications, challenges, and prospective in emerging body area networking technologies," *IEEE Wireless Communications*, vol. 17, no. 1, pp. 80–88, 2010.

[3] T. Aoyagi, I. Iswandi, M. Kim, J.-i. Takada, K. Hamaguchi, and R. Kohno, "Body motion and channel response of dynamic body area channel," in *Proceedings of the 5th European Conference on Antennas and Propagation (EUCAP)*, Apr. 2011.

[4] M. Gallo, P. Hall, Q. Bai, Y. Nechayev, C. Constantinou, and M. Bozzetti, "Simulation and measurement of dynamic on-body communication channels," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 2, pp. 623–630, Feb. 2011.

[5] Q. Wang, T. Tayamachi, I. Kimura, and J. Wang, "An on-body channel model for UWB body area communications for various postures," *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 4, pp. 991–998, 2009.

[6] A. Eid and J. Wallace, "Accurate modeling of body area network channels using surface-based method of moments," *IEEE Transac-*

tions on Antennas and Propagation, vol. 59, no. 8, pp. 3022–3030, Aug. 2011.

[7] A. Taparugssanagorn, C. Pomalaza-Ráez, A. Isola, R. Tesi, M. Hämäläinen, and J. Iinatti, "UWB channel modeling for wireless body area networks in medical applications," in *Proc. International Symposium on Medical information and Communication Technology (ISMICT)*, 2009.

[8] D. Smith, L. Hanlen, J. Zhang, D. Miniutti, D. Rodda, and B. Gilbert, "Characterization of the dynamic narrowband on-body to off-body area channel," in *IEEE International Conference on Communications (ICC)*, June 2009.

[9] S. Van Roy, F. Quitin, L. Liu, C. Oestges, F. Horlin, J. Dricot, and P. De Doncker, "Dynamic channel modeling for multi-sensor body area networks," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 2200–2208, Apr. 2013.

[10] "IEEE standard for local and metropolitan area networks - part 15.6: Wireless body area networks," *IEEE Std 802.15.6-2012*, 2012.

[11] T. Liang and D. Smith, "Energy-efficient, reliable wireless body area networks: cooperative diversity switched combining with transmit power control," *Electronics Letters*, vol. 50, no. 22, pp. 1641–1643, 2014.

[12] G. Arrobo and R. Gitlin, "New approaches to reliable wireless body area networks," in *IEEE International Conference on Microwaves, Communications, Antennas and Electronics Systems (COMCAS)*, Nov. 2011.

[13] A. Rahim, N. Javaid, M. Aslam, U. Qasim, and Z. Khan, "Adaptive-reliable medium access control protocol for wireless body area networks," in *9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON),*, June 2012.

[14] S. Marinkovic, C. Spagnol, and E. Popovici, "Energy-efficient TDMA-based MAC protocol for wireless body area networks," in *SENSORCOMM '09. Third International Conference on Sensor Technologies and Applications*, June 2009.

[15] D. Wu, B. Yang, H. Wang, D. Wu, and R. Wang, "An energy-efficient data forwarding strategy for heterogeneous wbans," *IEEE Access*, vol. 4, pp. 7251–7261, 2016.

[16] G. D. Ntouni, A. S. Lioumpas, and K. S. Nikita, "Reliable and energy-efficient communications for wireless biomedical implant systems," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1848–1856, 2014.

[17] H. A. Sabti and D. V. Thiel, "Node position effect on link reliability for body centric wireless network running applications," *IEEE Sensors Journal*, vol. 14, no. 8, pp. 2687–2691, 2014.

[18] F. Fekri, R. Mersereau, and R. Schafer, "Two-band wavelets and filterbanks over finite fields with connections to error control coding," *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3143–3151, Dec. 2003.

[19] F. Fekri, S. McLaughlin, R. Mersereau, and R. Schafer, "Block error correcting codes using finite-field wavelet transforms," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 991–1004, Mar. 2006.

[20] Y. Sankarasubramaniam, I. Akyildiz, and S. McLaughlin, "Energy efficiency based packet size optimization in wireless sensor networks," in *IEEE International Workshop on Sensor Network Protocols and Applications*, 2003.

[21] M. C. Vuran and I. F. Akyildiz, "Cross-layer packet size optimization for wireless terrestrial, underwater, and underground sensor networks," in *IEEE Conference on Computer Communications (INFOCOM).*, 2008.

[22] M. C. Oto and O. B. Akan, "Energy-efficient packet size optimization for cognitive radio sensor networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, 2012.

[23] M. C. Domingo, "Packet size optimization for improving the energy efficiency in body sensor networks," *ETRI Journal*, vol. 33, no. 3, pp. 299–309, 2011.

[24] Y. Li, X. Qi, M. Keally, Z. Ren, G. Zhou, D. Xiao, and S. Deng, "Communication energy modeling and optimization through joint packet size analysis of BSN and WiFi networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 9, pp. 1741–1751, 2013.

[25] M. J. Neely, "Optimal energy and delay tradeoffs for multiuser wireless downlinks," *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3095–3113, 2007.

[26] ——, "Energy optimal control for time-varying wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 2915–2934, 2006.

[27] K. Jamieson and H. Balakrishnan, "PPR: Partial packet recovery for wireless networks," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 4, pp. 409–420, 2007.

[28] G. R. Woo, P. Kheradpour, D. Shen, and D. Katabi, "Beyond the bits: cooperative packet recovery using physical layer information," in *Proceedings of the 13th annual ACM International Conference on Mobile Computing and Networking*. ACM, 2007.

[29] B. Han, A. Schulman, F. Gringoli, N. Spring, B. Bhattacharjee, L. Nava, L. Ji, S. Lee, and R. R. Miller, "Maranello: Practical partial packet recovery for 802.11." in *NSDI*, 2010.

[30] H. Lu, S. Zhou, and Z. Zhang, "Retransmission rate selection for block-based partial packet recovery," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2012.

[31] W. Hu, J. Xie, and Z. Zhang, "pmore: Exploiting partial packets in opportunistic routing," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2011.

[32] J. Zhang, H. Shen, K. Tan, R. Chandra, Y. Zhang, and Q. Zhang, "Frame retransmissions considered harmful: improving spectrum efficiency using micro-acks," in *Proceedings of the 18th annual International Conference on Mobile Computing and Networking*. ACM, 2012.

[33] K. C.-J. Lin, N. Kushman, and D. Katabi, "ZipTx: Harnessing partial packets in 802.11 networks," in *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*. ACM, 2008.

[34] G. Angelopoulos, A. Chandrakasan, and M. Medard, "Prac: Exploiting partial packets without cross-layer or feedback information," in *IEEE International Conference on Communications (ICC)*, June 2014.

[35] J. Xie, W. Hu, and Z. Zhang, "Efficient software partial packet recovery in 802.11 wireless LANs," *IEEE Transactions on Computers*, vol. 63, no. 10, pp. 2402–2415, Oct. 2014.

[36] K. Balachandran, S. Kadaba, and S. Nanda, "Channel quality estimation and rate adaptation for cellular mobile radio," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1244–1256, 1999.

[37] D. Qiao, S. Choi, and K. Shin, "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs," *IEEE Transactions on Mobile Computing*, vol. 1, no. 4, pp. 278–292, 2002.

[38] S. Choudhury and J. Gibson, "Payload length and rate adaptation for multimedia communications in wireless LANs," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 4, pp. 796–807, 2007.

[39] S. Lanzisera, A. Mehta, and K. Pister, "Reducing average power in wireless sensor networks through data rate adaptation," in *IEEE International Conference on Communications*, 2009.

[40] Q. Zhang, "Energy saving efficiency comparison of transmit power control and link adaptation in BANs," in *IEEE International Conference on Communications*, 2013.

[41] A. El Fawal, J.-Y. Le Boudec, R. Merz, B. Radunovic, J. Widmer, and G. M. Maggio, "Trade-off analysis of PHY-aware MAC in low-rate low-power UWB networks," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 147–155, 2005.

[42] S. Fu, D. Wang, H. Zhai, and Y. Li, "An enhanced link adaptation for the MB-OFDM UWB system," in *Wireless Communications and Networking Conference (WCNC)*, 2012.

[43] R. Al-Zubi and M. Krunz, "Interference management and rate adaptation in OFDM-based UWB networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 9, pp. 1267–1279, 2010.

[44] F. Martelli, R. Verdone, and C. Buratti, "Link adaptation in wireless body area networks," in *Vehicular Technology Conference (VTC Spring)*, 2011.

[45] ——, "Link adaptation in IEEE 802.15.4-based wireless body area networks," in *International Symposium on Personal, Indoor and Mobile Radio Communications Workshops (PIMRC Workshops)*, 2010.

[46] W.-B. Yang and K. Sayrafian-Pour, "Interference mitigation using adaptive schemes in body area networks," *International Journal of Wireless Information Networks*, vol. 19, no. 3, pp. 193–200, 2012.

[47] F. Marvasti, *Nonuniform sampling: theory and practice*. Springer, 2001.

[48] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals,"

IEEE Transactions on Information Theory, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[49] J. Romberg, "Compressive sensing by random convolution," *SIAM Journal on Imaging Sciences*, vol. 2, no. 4, pp. 1098–1128, 2009.

[50] J. Sun and V. Goyal, "Optimal quantization of random measurements in compressed sensing," in *IEEE International Symposium on Information Theory (ISIT)*, June 2009.

[51] P. Boufounos, "Universal rate-efficient scalar quantization," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1861–1872, Mar. 2012.

[52] J. Laska and R. Baraniuk, "Regime change: Bit-depth versus measurement-rate in compressive sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3496–3505, July 2012.

[53] S. Feizi, G. Angelopoulos, V. Goyal, and M. Medard, "Energy-efficient time-stampless adaptive nonuniform sampling," in *IEEE Sensors*, Oct. 2011.

[54] C. Isheden and G. Fettweis, "Energy-efficient link adaptation on a rayleigh fading channel with receiver CSI," in *IEEE International Conference on Communications*, 2011.

[55] M. Win and R. Scholtz, "Characterization of ultra-wide bandwidth wireless indoor channels: a communication-theoretic view," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 9, pp. 1613–1627, 2002.

[56] A. Molisch, "Ultrawideband propagation channels-theory, measurement, and modeling," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 5, pp. 1528–1545, 2005.

[57] ——, "Ultra-wide-band propagation channels," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 353–371, 2009.

[58] Y. Zhao, Y. Hao, A. Alomainy, and C. Parini, "UWB on-body radio channel modeling using ray theory and subband FDTD method," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 4, pp. 1827–1835, 2006.

[59] A. Fort, J. Ryckaert, C. Desset, P. De Doncker, P. Wambacq, and L. Van Biesen, "Ultra-wideband channel model for communication around the human body," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 4, pp. 927–933, 2006.

[60] A. Fort, C. Desset, P. De Doncker, P. Wambacq, and L. Van Biesen, "An ultra-wideband body area propagation channel model-from statistics to implementation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 4, pp. 1820–1826, 2006.

[61] IEEE 802.15 WPAN task group 6 (TG6) body area networks. [Online]. Available: http://www.ieee802.org/15/pub/TG6.html

[62] P. Humblet and M. Azizoglu, "On the bit error rate of lightwave systems with optical amplifiers," *Journal of Lightwave Technology*, vol. 9, no. 11, pp. 1576–1582, 1991.

[63] S. Dubouloz, B. Denis, S. De Rivaz, and L. Ouvry, "Performance analysis of LDR UWB non-coherent receivers in multipath environments," in *IEEE International Conference on Ultra-Wideband*, 2005.

[64] T. Quek, M. Win, and D. Dardari, "Unified analysis of UWB transmitted-reference schemes in the presence of narrowband interference," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, 2007.

[65] A. Rabbachin, T. Quek, P. Pinto, I. Oppermann, and M. Win, "Non-coherent UWB communication in the presence of multiple narrowband interferers," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3365–3379, 2010.

[66] P. Banelli and S. Cacopardi, "Theoretical analysis and performance of OFDM signals in nonlinear AWGN channels," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 430–441, 2000.

[67] T. Araujo and R. Dinis, "On the accuracy of the Gaussian approximation for the evaluation of nonlinear effects in OFDM signals," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 346–351, 2012.

[68] K. Witrisal, G. Leus, G. Janssen, M. Pausini, F. Troesch, T. Zasowski, and J. Romme, "Noncoherent ultra-wideband systems," *IEEE Signal Processing Magazine*, vol. 26, no. 4, pp. 48–66, 2009.

[69] H. Landau and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-III: The dimension of the space of essentially time- and band-limited signals," *Bell System Technical Journal*, vol. 41, pp. 1295–1336, 1962.

[70] J. Proakis, *Digital Communications*. McGraw-Hill Higher Education, 2001.

[71] L. Hentila, A. Taparungssanagorn, H. Viittala, and M. Hamalainen, "Measurement and modelling of an UWB channel at hospital," in *IEEE International Conference on Ultra-Wideband*, 2005.

[72] M. Z. Win and R. A. Scholtz, "Impulse radio: how it works," *IEEE Communications Letters*, vol. 2, no. 2, pp. 36–38, 1998.

[73] S. Xiao, A. Dhamdhere, V. Sivaraman, and A. Burdett, "Transmission power control in body area sensor networks for healthcare monitoring," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 1, pp. 37–48, Jan. 2009.

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

[75] M. Quwaider, J. Rao, and S. Biswas, "Body-posture-based dynamic link power control in wearable sensor networks," *IEEE Communications Magazine*, vol. 48, no. 7, pp. 134–142, July 2010.

[76] F. Di Franco, C. Tachtatzis, B. Graham, M. Bykowski, D. Tracey, N. Timmons, and J. Morrison, "The effect of body shape and gender on wireless body area network on-body channels," in *IEEE Middle East Conference on Antennas and Propagation (MECAP)*, Oct. 2010.

[77] A. Gupta and T. Abhayapala, "Body area networks: Radio channel modelling and propagation characteristics," in *Australian Communications Theory Workshop (AusCTW)*, Jan. 2008.

[78] A. Fort, F. Keshmiri, G. Crusats, C. Craeye, and C. Oestges, "A body area propagation model derived from fundamental principles: Analytical analysis and comparison with measurements," *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 2, pp. 503–514, Feb. 2010.

[79] S.-H. Pun, Y.-M. Gao, P. Mak, M.-I. Vai, and M. Du, "Quasi-static modeling of human limb for intra-body communications with experiments," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 6, pp. 870–876, Nov. 2011.

[80] A. Boulis, Y. Tselishchev, L. Libman, D. Smith, and L. Hanlen, "Impact of wireless channel temporal variation on mac design for body area networks," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 11, no. S2, p. 51, 2012.

[81] A. Duel-Hallen, "Fading channel prediction for mobile radio adaptive transmission systems," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2299–2313, Dec. 2007.

[82] V. Jacobson, "Congestion avoidance and control," in *ACM SIGCOMM Computer Communication Review*, vol. 18, no. 4. ACM, 1988.

[83] D. Smith, D. Miniutti, L. Hanlen, D. Rodda, and B. Gilbert, "Dynamic narrowband body area communications: Link-margin based performance analysis and second-order temporal statistics," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2010.

[84] L. Hanlen, V. Chaganti, B. Gilbert, D. Rodda, T. Lamahewa, and D. Smith, "Open-source testbed for body area networks: 200 sample/sec, 12 hrs continuous measurement," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications Workshops (PIMRC Workshops)*, Sep. 2010.

[85] E. Rebeiz, G. Caire, and A. F. Molisch, "Energy-delay tradeoff and dynamic sleep switching for bluetooth-like body-area sensor networks," *IEEE Transactions on Communications*, vol. 60, no. 9, pp. 2733–2746, 2012.

[86] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, 1992.

[87] S. Asmussen, *Applied probability and queues*. Springer Science & Business Media, 2008, vol. 51.

[88] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[89] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13).

[90] P. Cummiskey, N. Jayant, and J. Flanagan, "Adaptive quantization in differential PCM coding of speech," *Bell System Technical Journal*, vol. 52, no. 7, pp. 1105–1118, Sept 1973.

[91] A. Papoulis, "Limits on bandlimited signals," *Proceedings of the IEEE*, vol. 55, no. 10, pp. 1677–1686, Oct. 1967.

[92] H. Minn, C. Tellambura, and V. Bhargava, "On the peak factors of sampled and continuous signals," *IEEE Communications Letters*, vol. 5, no. 4, pp. 129–131, Apr. 2001.

[93] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[94] T. Ho, M. Medard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[95] D. E. Lucani, M. Medard, and M. Stojanovic, "On coding for delay-network coding for time-division duplexing," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2330–2348, 2012.

[96] J. Xie, W. Hu, and Z. Zhang, "Revisiting partial packet recovery in 802.11 wireless LANs," in *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*. ACM, 2011.

[97] S. Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371–381, 2003.

[98] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Transactions on Information Theory*, vol. 51, no. 8, 2005.

[99] S. Teerapittayanon, K. Fouli, M. Médard, M.-J. Montpetit, X. Shi, I. Seskar, and A. Gosain, "Network coding as a wimax link reliability mechanism," in *Multiple Access Communications*. Springer, 2012, pp. 1–12.

[100] Q. Zhang, J. Heide, M. V. Pedersen, and F. H. P. Fitzek, "MBMS with user cooperation and network coding," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, 2011.

[101] J. Jin, B. Li, and T. Kong, "Is random network coding helpful in wimax?" in *IEEE Conference on Computer Communications (INFOCOM)*, 2008.

[102] A. Das and S. Vishwanath, "On finite alphabet compressive sensing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013.

[103] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[104] B. Lu, X. Wang, and J. Zhang, "Throughput of CDMA data networks with multiuser detection, ARQ, and packet combining," *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1576–1589, 2004.

[105] M. Nistor, D. E. Lucani, T. T. V. Vinhoza, R. A. Costa, and J. Barros, "On the delay distribution of random linear network coding," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 5, pp. 1084–1093, 2011.

[106] R. Cogill and B. Shrader, "Delay bounds for random linear coding in parallel relay networks," *IEEE Transactions on Mobile Computing*, vol. 14, no. 5, pp. 964–974, 2015.

[107] W. Li, X. Cheng, T. Jing, Y. Cui, K. Xing, and W. Wang, "Spectrum assignment and sharing for delay minimization in multi-hop multi-flow CRNs," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2483–2493, 2013.

[108] X. Shi, M. Medard, and D. E. Lucani, "Whether and where to code in the wireless packet erasure relay channel," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 8, pp. 1379–1389, 2013.

[109] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing.* Springer, 2013.

[110] M. Duarte and R. Baraniuk, "Kronecker compressive sensing," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 494–504, Feb. 2012.

[111] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[112] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 211–231, 2009.

[113] D. Lucani, M. Medard, and M. Stojanovic, "Random linear network coding for time-division duplexing: Field size considerations," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Nov. 2009.

[114] ——, "Systematic network coding for time-division duplexing," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, June 2010.

[115] M. Mohammadi, Q. Zhang, and E. Dutkiewicz, "Channel-adaptive MAC frame length in wireless body area networks," in *International Symposium on Wireless Personal Multimedia Communications (WPMC),*, Sep. 2014.

[116] M. Mohammadi, Q. Zhang, E. Dutkiewicz, and X. Huang, "Optimal frame length to maximize energy efficiency in IEEE 802.15.6 UWB body area networks," *Wireless Communications Letters, IEEE*, vol. 3, no. 4, pp. 397–400, Aug. 2014.

[117] M. Langberg, A. Sprintson, and J. Bruck, "Network coding: A computational perspective," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 147–157, 2009.

[118] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. M. G. M. Tolhuizen, "Polynomial time algorithms for multicast network code construction," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1973–1982, 2005.

[119] Sparselab software package. [Online]. Available: https://sparselab.stanford.edu/

[120] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[121] M. Fresia and G. Caire, "A practical approach to lossy joint source-channel coding," *arXiv preprint cs/0702070*, 2007.

[122] J. M. Pollard, "The fast Fourier transform in a finite field," *Mathematics of computation*, vol. 25, no. 114, pp. 365–374, 1971.

[123] P. Vaidyanathan, "Unitary and paraunitary systems in finite fields," in *IEEE International Symposium on Circuits and Systems*, May 1990.

[124] R. Blahut, *Algebraic methods for signal processing and communications coding*. Springer-Verlag, 1992.

[125] G. Caire, R. Grossman, and H. Poor, "Wavelet transforms associated with finite cyclic groups," in *The Twenty-Sixth Asilomar Conference on Signals, Systems and Computers*, Oct. 1992.

[126] J. Hong and M. Vetterli, "Hartley transforms over finite fields," *IEEE Transactions on Information Theory*, vol. 39, no. 5, pp. 1628–1638, Sep. 1993.

[127] M. Swanson and A. Tewfik, "A binary wavelet decomposition of binary images," *IEEE Transactions on Image Processing*, vol. 5, no. 12, pp. 1637–1650, Dec. 1996.

[128] N.-F. Law and W.-C. Siu, "A filter design strategy for binary field wavelet transform using the perpendicular constraint," *Signal Processing*, vol. 87, no. 11, pp. 2850–2858, 2007.

[129] I. Tal and A. Vardy, "List decoding of polar codes," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, July 2011.

[130] K. Chen, K. Niu, and J. Lin, "A hybrid ARQ scheme based on polar codes," *IEEE Communications Letters*, vol. 17, no. 10, pp. 1996–1999, 2013.

[131] E. Arikan, "Systematic polar coding," *IEEE Communications Letters*, vol. 15, no. 8, pp. 860–862, Aug. 2011.

[132] S.-Y. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a gaussian approximation," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 657–670, 2001.

[133] S. Lin and P. Yu, "A hybrid ARQ scheme with parity retransmission for error control of satellite channels," *IEEE Transactions on Communications*, vol. 30, no. 7, pp. 1701–1719, 1982.

[134] P. P. Mercier, D. C. Daly, and A. P. Chandrakasan, "An energy-efficient all-digital UWB transmitter employing dual capacitively-coupled pulse-shaping drivers," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 6, pp. 1679–1688, 2009.

[135] F. S. Lee and A. P. Chandrakasan, "A 2.5 nj/bit 0.65 v pulsed UWB receiver in 90 nm cmos," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 12, pp. 2851–2859, 2007.