PERFORMANCE LIMITS FOR ENERGY-CONSTRAINED COMMUNICATION SYSTEMS

ANSHOO TANDON
(M.E., Indian Institute of Sciences, Bangalore, India)

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2016

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety.

I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Anshoo Tandon
12 January 2016

Acknowledgements

I am indebted to my supervisor Professor Mehul Motani for his constant support and guidance. His advice and comments have been highly beneficial in shaping my research. I sincerely thank him for mentoring me during different phases of my candidatures and granting me the freedom to pursue my research interests.

I would like to thank my collaborator Dr. Lav R. Varshney for his guidance and valuable discussions on the topic of simultaneous energy and information transfer. I thank the professors at the ECE department in NUS, especially Prof. Marc Andre Armand, Dr. Vincent Y. F. Tan, Dr. Zhang Rui, and Prof. Kam Pooi Yuen, for offering some wonderful modules.

I would like to thank my friends for making my PhD journey memorable and fun. I thank my friends Tram, Hoa, and Quoc, for their lively company during lunch hours. I would like to thank my fellow lab mates, namely Janaka, Shashi, Amna, Aissan, Sanat, Katayuon, Reza, and Ahmed, for all the discussions and help. Special acknowledgment to my table-tennis and badminton friends at NUS: Peichu, Sergey, Parvathy, Bhargav, and Jayshree.

I am grateful to my parents and Prof. D. P. Patil for their support and encouragement. Lastly, special thanks to my best friend and wife, Lavanya, for her patience, understanding, and for being a constant source of strength.

Table of Contents

Summary

The presence of constraints on the availability and usage of energy impacts the performance of communication systems, characterized in terms of metrics such as throughput, reliability, and delay. This impact may be significant when the transmitter and/or receiver does not have a dedicated power source and relies on energy harvesting for its operation. In this dissertation, we formulate and analyze performance limits for three different energy-constrained communication systems, and draw meaningful design insights for each of these three different systems.

First, we consider a polling based multiaccess system where energy-constrained nodes employ automatic repeat request (ARQ) and transmit data over a noisy communication channel. The goal here is to reduce the average packet delay and the average energy usage, relative to an uncoded system, through joint use of forward error correction (FEC) codes and ARQ. We show that reducing the average service time via FEC codes is sufficient to reduce the average packet delay. Further, we show that a reduction in average service time translates into an equal amount of percentage savings in energy usage. Additionally, we characterize the fundamental limit on the percentage reduction in average service time using best possible FEC codes and present useful system design guidelines.

Secondly, we consider an energy harvesting transmitter which uses harvested energy for transmission of data packets. The data and energy arrivals are random, and an exact system analysis is challenging because the data queue dynamics are influenced jointly by the energy arrival process and the data service process. We formulate a two stage virtual queueing system to obtain approximate closed-form expressions for the average packet delay under different energy arrival statistics. The derived results are shown to be robust via comparison with Monte-Carlo simulation results for the physical queueing system.

Thirdly, we consider an energy harvesting receiver which uses the received signal to simultaneously harvest energy as well as to decode the information em-

bedded in the signal. For a given receiver energy requirement, the goal is to obtain the fundamental limit on the information transfer rate. When the receiver energy buffer is small, the energy content in the signal should be sufficiently regular to avoid power outage. Here, we investigate achievable information rates under three different classes of constrained codes for enabling real-time energy transfer. The first class of constrained codes, called the constant subblock-composition codes, require that each subblock within every codeword has the same composition. The second class of constrained codes, called the subblock energy-constrained codes, allows different subblocks to have different composition while ensuring sufficient energy within each subblock. The third class of constrained codes ensures that each codeword carries sufficient energy within a sliding time window. For each of these three classes of constrained codes, we analyze the capacity and highlight the tradeoff between delivery of sufficient energy and achieving high information transfer rates.

List of Figures

List of Tables

List of Acronyms

ARQ	Automatic Repeat Request
AST	Average Service Time
AWGN	Additive White Gaussian Noise
BDD	Bounded Distance Decoding
BEC	Binary Erasure Channel
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BSC	Binary Symmetric Channel
CSCC	Constant Subblock-Composition Code
DMC	Discrete Memoryless Channel
FEC	Forward Error Correction
IID	Independent and Identically Distributed
ML	Maximum Likelihood
NA	Normal Approximation
PEP	Packet Error Probability
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RCB	Random Coding Bound
SECC	Subblock Energy-Constrained Code
SNR	Signal-to-Noise Ratio
SPB	Sphere Packing Bound

Chapter 1

Introduction

Reliable communication over noisy channels was shown possible by Shannon [1] for rates upper bounded by the channel capacity. This could be achieved by using asymptotically long codewords. However, a delay-constrained communication system may not permit the use of long codewords [2] because of high encoding and decoding delays. In this scenario, many insights offered by classical information theory do not apply directly. Towards resolving this issue, researchers have aimed to quantify the capacity for fixed blocklengths [3], [4], and have also explored alternative engineering methods like the automatic repeat request (ARQ) protocol at the data link layer where retransmissions are requested for packets in error [5]. Another practical issue not adequately addressed by classical information theory is that of random and bursty data arrival at the transmitter [2]. In this case, the queueing behavior of the arriving data packets may be analyzed together with characterization of information theoretic limits [6], and investigations made under packet delay constraints [7].

Constraints on the availability and usage of energy may also critically affect the performance of communication systems. For instance, a wireless micro-sensor node is typically battery operated and hence energy is a precious resource due to

finite energy storage capability at the node [8]. When the cost of regularly replacing the batteries is high or when sensors operate in environments that rule out frequent energy replenishment, the energy constraint directly affects the sensor's lifetime and thus entails the implementation of energy-efficient algorithms and protocols [9]. In this regard, the use of energy harvesting [10] holds the promise of virtually perpetual node operation, where energy is harnessed from either natural resources (such as sun or wind) or man-made phenomena (such as wireless energy transfer). However, realizing this potential requires judicious design optimization due to uncertainty in the energy arrival process, both in terms of arrival instances and the amount of energy harvested [11].

In this thesis, we study achievable limits on performance metrics, such as throughput and delay, for energy-constrained communication systems. This introductory chapter provides an outline for the thesis and is arranged as follows. Section 1.1 gives the motivation for the study. Section 1.2 provides a mapping of the different problem scenarios investigated in the thesis to the respective chapters. The thesis contributions are detailed in Section 1.3 while Section 1.4 provides an outline of different chapters in the thesis.

1.1 Motivation

In this section, we motivate the study of performance limits for the three different energy-constrained communication systems analyzed in this thesis.

- First, we consider a system where energy-constrained nodes are polled for transmission by a central server, and the link layer at each node employs automatic repeat request (ARQ) to ensure error free delivery of packets. On noisy communication channels, packets errors may significantly increase the delay due to retransmissions [12]. In this scenario, quantifying the mini-

mum achievable packet delay helps to benchmark practical design approaches towards managing delay sensitive applications. Although the packet error probability (PEP), and hence the number of retransmissions, can be reduced by employing physical layer forward error correction (FEC) codes (also called channel codes) [13], the use of FEC codes results in addition of redundant bits which increase the packet transmission time. A small number of redundant bits may not sufficiently reduce the PEP, while a large number of redundant bits may increase the packet delay due to the high transmission time. This observation naturally leads to the question of finding the appropriate coding scheme for minimizing delay, and quantifying its impact on the average energy usage for packet transmission. We address these questions in this thesis and quantify fundamental performance limits using the best possible FEC codes.

- Secondly, we consider an energy harvesting transmitter where the harvested energy is used for transmission of data packets. The data packets arrive randomly and wait in a queue for accumulation of sufficient energy and for service completion of previously arrived packets. In this scenario, the problem of quantifying the exact average packet delay is hard because the data queue dynamics are influenced jointly by the energy arrival process and the data service process. The difficulty in obtaining an exact delay expression raises questions on selection criteria of system design parameters which will ensure that the average packet delay is less than a given threshold. In this thesis, we present a closed-form expression for the average packet delay which is shown to be exact when the service time is negligible, and robust even for relative high values of the average service time. We derive results for single-source energy harvesting, and extend the results for the

useful case of multi-source energy harvesting.

- Thirdly, we consider the problem of simultaneous energy and information transfer to an energy harvesting receiver. Here, the receiver uses the same received signal both for decoding information and for harvesting energy, which is employed to power its circuitry. The tradeoff between reliable communication and delivery of energy at the receiver was first characterized by Varshney [14] using a general capacity-power function, where transmitted codewords were constrained to have average received energy exceed a threshold. However, in practical applications, imposing only an average power constraint is not sufficient; we also need to regularize the transferred energy content. This is because a codeword satisfying the average power constraint may still cause power outage at the receiver if the energy content in the codeword is bursty, since the receive energy buffer with a relatively small storage capacity may drain during intervals with low signal energy. This scenario raises the following questions: (i) What structure should we impose on transmitted codewords to ensure that the receiver does not suffer from power outage. (ii) How do we characterize achievable information rates for a given set of codeword constraints. (iii) What is the penalty in rate due to codeword constraints providing real-time energy transfer relative to the case where only the average received energy per codeword is constrained to exceed a threshold.

  One approach to providing real-time energy transfer is to divide each codeword into smaller subblocks and then constrain each subblock to carry sufficient energy. Another approach is to ensure that each codeword carries sufficient energy within a moving time window. In this thesis, we investigate achievable information rates under three different classes of constrained

codes for enabling real-time energy transfer. For each of these three classes of constrained codes, we analyze the capacity and highlight the tradeoff between delivery of sufficient energy and achieving high information transfer rates.

1.2 Thesis Overview

In this section, we provide a mapping of the different communication systems analyzed in this thesis to the different chapters in this thesis.

1. First, we consider a multiaccess system where energy-constrained nodes employ automatic repeat request and transmit data over a noisy communication channel. This system is analyzed in detail in Chapter 3, where we investigate the reduction in delay and energy usage via FEC codes. Further, we characterize the fundamental limit on the percentage reduction in average service time using best possible error correcting codes.

2. Secondly, in Chapter 4, we consider an energy harvesting transmitter which uses harvested energy from the environment for transmission of data packets. The data and energy arrivals are random, and an exact system analysis is challenging because the data queue dynamics are influenced jointly by the energy arrival process and the data service process. In Chapter 4, we formulate a two stage virtual queueing system to obtain closed-form approximate expressions for the average packet delay under different energy arrival statistics. The derived results are shown to be robust via comparison with Monte-Carlo simulation results for the physical queueing system.

3. Thirdly, we consider an energy harvesting receiver which uses the received signal to simultaneously harvest energy as well as to decode the information

embedded in the signal. For a given receiver energy requirement, the goal is to obtain the fundamental limit on the information transfer rate. In this scenario, we propose three different codeword constraints for enabling *real-time energy transfer*. The capacity under these three different codeword constraints is analyzed in Chapter 5, Chapter 6, and Chapter 7, respectively.

- In Chapter 5, we propose the use of *constant subblock-composition codes* for providing regular energy content to the receiver. Here, the codewords are partitioned into smaller subblocks, and each subblock is constrained to have the same composition. This composition is chosen to maximize the information rate while ensuring that sufficient energy is carried within every subblock duration.
- In Chapter 6, we propose the use of *subblock energy-constrained codes* which allows different subblocks to have different composition while still carrying sufficient energy within each subblock.
- In Chapter 7, we consider a third class of codeword constraint which requires that each codeword carries *sufficient energy within a sliding time window*.

For each of the above three classes of constrained codes, we provide bounds on capacity and analyze the tradeoff between delivery of sufficient energy and achieving high information transfer rates.

1.3 Thesis Contributions

The research contributions on the different communication systems studied in this thesis are summarized below.

1.3.1 Reducing delay and energy usage via FEC codes

We consider a polling based multiple access scheme where the link layer employs Stop-and-Wait ARQ to ensure error free delivery of data packets and study the impact of FEC codes on the packet delay and energy usage. The main contributions in this thesis, related to use of FEC codes for reducing the packet delay and energy usage in such a polling based system, are summarized as follows:

1. We harness the average waiting time expression for our communication model to prove that a coded system results in lower average packet delay compared to an uncoded system if it has lower average service time (AST) relative to the uncoded system. We also give explicit sufficient conditions comparing any two channel coding schemes, under which the coding scheme with lower AST also achieves lower average packet delay. When the queue switching time is negligible, we show that the percentage reduction in average packet delay is at least as much as the percentage reduction in AST through the use of channel coding.

2. We show that when the energy per transmitted bit is kept constant, the ratio of average transmit energy for the coded and uncoded system is same as the ratio of the AST for the coded and uncoded system. Thus, the percentage reduction in AST due to coding translates into an equal percentage reduction in average transmit energy.

3. We show that the reduction in AST using channel coding for large packet lengths can be bounded as a function of the reduction in AST for smaller packet lengths. This helps to quantify the reduction in AST through channel coding for arbitrarily large packet sizes.

4. We provide several bounds on the reduction in AST using the best possible

FEC codes. Further, we give a necessary condition when FEC will help to reduce AST and a sufficient condition when FEC will not help. The search for best possible codes provides the insight that for Gaussian noise channels with hard-decision decoding, a relatively high PEP ($\sim 10^{-2}$) obtained using as high a coding rate as possible, results in AST which is close to the smallest achievable AST. Additionally, it is observed that the performance of a bounded distance decoder is sufficiently close to that of the optimum maximum likelihood decoder.

5. We show that the packet delay can be further reduced through opportunistic combining and joint encoding of data packets. This is shown by using information theoretic results which assert that the probability of packet error can be reduced by increasing the packet length.

1.3.2 Packet delay analysis at an energy harvesting transmitter

Here, we consider an energy harvesting transmitter which uses the harvested energy for the transmission of data packets. The energy arrival process is modeled as a discrete random process and data is assumed to arrive randomly in the form of fixed sized packets. Since the energy arrival process and the data service processes jointly impact the data queue dynamics, the characterization of the average packet delay is challenging. In this thesis we derive closed-form expressions for the average packet delay and the probability of data loss due to buffer overflow through the use of *virtual queues*. The main contributions are summarized below:

1. We formulate a virtual queueing system which decouples the wait stages for the energy arrival process and the service process. This decoupling enables us to obtain closed-form expressions for the average packet delay and the

probability of data packet loss due to buffer overflow.

2. We derive the average packet delay with single-source energy harvesting in three different settings: (i) *Unit-energy arrivals* where each energy arrival brings energy which is sufficient to transmit one data packet. (ii) *Fractional-energy arrivals* where each energy arrival brings only a fraction of the energy required to transmit one data packet. (iii) *Bulk-energy arrivals* where each energy arrival brings sufficient energy for transmission of multiple data packets.

   The unit-energy arrivals provide a simple base-line setting where the average waiting time for energy arrival is derived by obtaining steady-state probabilities of a one-dimensional Markov chain which models the evolution of data and energy buffer occupancy. On the other hand, the analysis of fractional and bulk energy arrival cases is more challenging, where the buffer occupancy dynamics has to be modeled as a two-dimensional Markov chain. In each of the case of fractional and bulk energy arrivals, a close examination of the relation among the steady-state probabilities reveals a certain structure which we exploit to obtain the exact expression for the average waiting time for energy arrival.

3. We extend the approach used for analyzing fractional and bulk-energy arrivals to the useful case of *multi-source energy harvesting*. Here, the energy is harvested from two independent sources, with potentially different quantity and rate of energy arrival from each source.

4. We use the closed-form expressions for the average packet delay to analyze the maximum throughput which the system can support under an average delay constraint, and discuss the trade-off between throughput and system cost.

5. The derived closed-form expressions for the average packet delay and the probability of data loss are shown to be *exact* when the service time is negligible. The Monte Carlo simulation results for the physical queueing system are compared with the corresponding closed-form expressions for the virtual queueing systems. The derived expressions are shown to be robust even when the service time is increased up to sixty percent of the average packet delay.

1.3.3 Constrained codes for simultaneous energy and information transfer

As a third energy-constrained communication system, we consider an energy-harvesting receiver that uses the same received signal both for decoding information and for harvesting energy, which is employed to power its circuitry. In this thesis, we investigate achievable information rates for three different proposed classes of constrained codes for enabling *real-time* energy transfer. The first class of constrained codes, called *constant subblock-composition codes* (CSCCs), require that each codeword is partitioned into smaller subblocks where each subblock has the same composition and carries sufficient energy. The second class of constrained codes, called *subblock energy-constrained codes* (SECCs), allows different subblocks to have different composition while ensuring that sufficient energy is carried within each subblock. The third class of constrained codes ensures each codeword carries sufficient energy within a *sliding time window*. Our contributions for each of these three classes of constrained codes is summarized below.

- *Constant Subblock-Composition Codes (CSCCs)*

    1. Here, each subblock in any given codeword has the same fixed composition, and this subblock-composition is chosen to maximize the rate

of information transfer while meeting the energy requirement. In this thesis, we consider discrete memoryless channels and characterize the CSCC capacity as a function of the required energy per symbol.

2. We show that CSCC capacity can be computed efficiently by exploiting certain symmetry properties.
3. For a given energy storage capability at the receiver, we derive a necessary and sufficient condition on the subblock length to avoid power outage at the receiver for all possible CSCC sequences.
4. Compared to constant composition codes, we quantify the rate loss incurred due to the additional constraint of restricting all subblocks within codewords to have the same composition.
5. For a given rate of information transfer, we derive a lower bound for the error exponent using CSCC in terms of the error exponent for constant composition codes.

- *Subblock Energy-Constrained Codes (SECCs)*

1. For this class of constrained codes, different subblocks may have different composition while still satisfying the energy constraint per subblock. In this thesis, we characterize the exact SECC capacity as a function of the required energy per symbol.
2. We provide a sufficient condition on the subblock length to avoid power outage at the receiver for all possible SECC sequences.
3. We compare SECC capacity with CSCC capacity, and also provide different bounds on the SECC capacity.
4. We characterize the random coding error exponent for SECCs.

- *Constraining codewords to carry sufficient energy in a sliding time window*

1. Here, we assume that the transmitter uses on-off keying where bit one corresponds to transmission of a high energy signal. The transmitter uses only those codewords which have at least d ones in a sliding window of $W = d+1$ bits. We show that with this constraint, the noiseless code capacity is achieved by sequences generated from a finite state Markov machine.
2. We also characterize achievable rates when such constrained codes are used on the following communication channels:
   - Binary Symmetric Channel
   - Binary Erasure Channel
   - Z-Channel

For each of the three different classes of codeword constraints, we provide numerical examples and analyze the tradeoff between delivery of sufficient energy and achieving high information transfer rates.

1.4 Thesis Outline

In this section, we provide an outline of the different chapters in this thesis.

- *Chapter 2: Background*

  In this chapter, we review some basic concepts related to information theory and queueing theory, and present a concise overview of the Pollaczek-Khinchin formula and the Blahut-Arimoto algorithm. Information theoretic quantities (such as entropy, mutual information and channel capacity) and queueing theoretic terms (such as arrival rate, service rate, waiting time, and packet delay) are defined and discussed briefly. The Pollaczek-Khinchin formula [15] expresses the average waiting time of a packet in a M/G/1 queueing

system, and is applied in Chapter 4 towards obtaining a closed-form expression for the average packet delay at an energy harvesting transmitter. The Blahut-Arimoto algorithm [16], [17] is an iterative algorithm for the capacity of a discrete memoryless channel, and is applied in Chapter 6 to numerically compute the capacity of subblock energy-constrained codes. Additionally, in Section 2.5 of this chapter, we present a detailed literature survey related to different problems addressed in this thesis.

- *Chapter 3: Reducing Delay and Energy Usage Via FEC Codes*

  In this chapter, we consider a polling based multiaccess system where energy-constrained nodes employ automatic repeat request (ARQ) and transmit data over a noisy communication channel. Here, we investigate the reduction in the average packet delay and the average energy usage, relative to an uncoded system, through joint use of forward error correction (FEC) codes and ARQ. We first establish that relative to an uncoded system, it is sufficient to reduce the average service time (AST) using FEC in order to achieve lower average packet delay. We then show that the percentage reduction in AST due to coding translates into an equal percentage reduction in average transmit energy when the energy per coded bit is fixed. Further, we characterize quantify the reduction in AST that can be achieved using the best possible FEC codes. Additionally, we show that the average packet delay can be further reduced in certain cases by opportunistically combining and encoding several packets jointly.

- *Chapter 4: Packet Delay Analysis at an Energy Harvesting Transmitter*

  Consider an energy harvesting transmitter which uses the harvested energy for transmission of data packets. The data packets arrive randomly and wait in a queue for accumulation of sufficient energy and for service com-

pletion of previously arrived packets. Thus, the data queue dynamics are influenced jointly by the energy arrival process, the data arrival process, and the data service process. We formulate a two stage *virtual queueing system* which decouples the wait stages for the energy arrival process and the service process. This decoupling helps us to obtain closed-form expressions for the average packet delay and the probability of data packet loss due to buffer overflow for different energy arrival statistics. These expressions are used to determine the maximum throughput under given quality of service (QoS) constraints. The derived expressions are shown to be exact when the service time is negligible. Even for relative high values of the average service time, we show the robustness of these expressions via Monte Carlo simulations.

- *Chapter 5: Constant Subblock-Composition Codes*

  The study of simultaneous information and energy transfer is relevant for communication from a powered transmitter to a receiver which uses the same received signal both for decoding information and for extracting energy to power its circuitry. When the receiver energy buffer is small, the energy content in the signal should be sufficiently regular to avoid power outage. In this chapter, we introduce the constant subblock-composition codes (CSCCs) where codewords are divided into smaller subblocks and all the subblocks are constrained to have the same composition. This subblock-composition is chosen to maximize the rate of information transfer while meeting the energy requirement. For discrete memoryless channels, we establish the CSCC capacity as a function of the required energy per symbol at the receiver. Compared to constant composition codes, we quantify the rate loss incurred due to the additional constraint of restricting all subblocks within codewords to have the same composition. Further, we derive a lower

bound for the CSCC error exponent in terms of the error exponent for constant composition codes.

- *Chapter 6: Subblock Energy-Constrained Codes*

  For the application of simultaneous energy and information transfer, we analyze the *subblock energy-constrained codes* (SECCs) which by definition are codes which carry sufficient energy in every *subblock* duration. Compared to constant subblock-composition codes, the SECCs allow different subblocks to have different composition and thus achieve higher rates of information transfer while still meeting the real-time energy requirement. We characterize the exact SECC capacity, and also provide different bounds on the SECC capacity. Further, we characterize and bound the random coding error exponent for SECCs. We also provide a sufficient condition on the subblock length to avoid power outage at the receiver for all possible SECC sequences.

- *Chapter 7: Codes Satisfying Sliding Window Energy Constraint*

  We study binary codes in which each codeword is constrained to have at least d ones in a *sliding window* of $W = d + 1$ consecutive bits. This constraint is equivalent to having at least d ones between successive zeros, which in turn defines a Type-1 (d, ∞) run-length limited (RLL) code. We consider on-off keying (bit "1" (resp. bit "0") is represented by the presence (resp. absence) of a carrier), where this sliding window constraint is useful for simultaneous energy and information transfer when the receiver relies on the received signal to fulfill its real-time power requirements. We give a probabilistic proof that the noiseless capacity of a (d, ∞) RLL code can be achieved by using a $d + 1$ state Markovian chain. We also give analytical expressions for achievable rates when these constrained codes are used on the (i) binary symmetric channel, (ii) Z-channel and (iii) binary erasure channel.

Bibliographical Note

Parts of *Chapter 3* appear in the paper:

- A. Tandon, M. Motani, and V. Srivastava, "On the Impact of Channel Coding on Average Packet Delay in a Multiuser Environment," in *Proc. IEEE Wireless Comm. Networking Conf. (WCNC)*, April 2013, pp. 499–504.

and in the journal manuscript:

- A. Tandon and M. Motani, "A Cross-Layer Approach to Reducing Packet Delay in Polling Based Multiuser Systems," accepted, *IEEE Trans. Veh. Tech.*, 2016, DOI:10.1109/TVT.2016.2568578.

Parts of *Chapter 4* appear in the paper:

- A. Tandon and M. Motani, "Has green energy arrived? Delay analysis for energy harvesting communication systems," in *Proc. IEEE Conf. Sensing, Commun. and Netw. (SECON)*, June 2014, pp. 582–590.

and in the journal manuscript:

- A. Tandon and M. Motani, "Diphase: Characterizing Packet Delay in Multi-Source Energy Harvesting Systems," accepted, *IEEE Trans. Commun.*, 2016, DOI:10.1109/TCOMM.2016.2594230.

Parts of *Chapter 5* appear in papers:

- A. Tandon, M. Motani, and L. R. Varshney, "Constant Subblock Composition Codes for Simultaneous Energy and Information Transfer," in *Proc. IEEE SECON Workshop on Energy Harvesting Communications*, June 2014, pp. 45–50.

- A. Tandon, M. Motani, and L. R. Varshney, "Real-time Simultaneous Energy and Information Transfer," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 1124-1128.

and in the journal manuscript:

- A. Tandon, M. Motani, and L. R. Varshney, "Subblock-Constrained Codes for Real-Time Simultaneous Energy and Information," in *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4212–4227, July 2016.

Parts of *Chapter 6* appear in the paper:

- A. Tandon, M. Motani, and L. R. Varshney, "Subblock Energy-Constrained Codes for Simultaneous Energy and Information Transfer," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1969-1973.

and in the journal manuscript:

- A. Tandon, M. Motani, and L. R. Varshney, "Subblock-Constrained Codes for Real-Time Simultaneous Energy and Information," in *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4212–4227, July 2016.

Parts of *Chapter 7* appear in the paper:

- A. Tandon, M. Motani, and L. R. Varshney, "On Code Design for Simultaneous Energy and Information Transfer," in *Proc. 2014 Inf. Theory Appl. Workshop (ITA)*, Feb. 2014.

Chapter 2

Background

In this chapter, we review some basic concepts related to information theory and queueing theory, and present mathematical preliminaries which will be used in the remainder of the thesis. Additionally, in Section 2.5, we present a detailed literature survey related to the different communication systems analyzed in this thesis.

2.1 Communication Systems

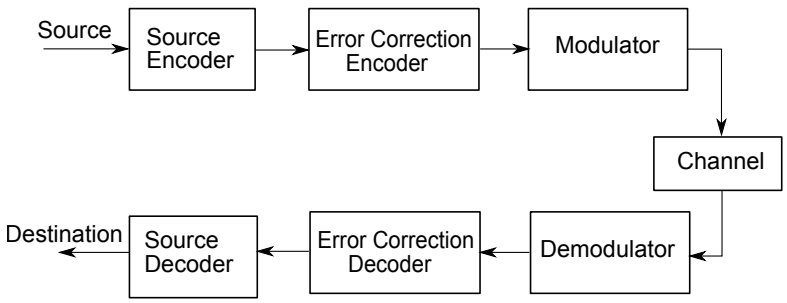

Figure 2.1: Digital Communication System Block Diagram

A general block diagram of a communication system is shown in Fig. 2.1. The source of information is an arbitrary message, e.g. a time varying voltage or a bit stream, which is intended for use by the destination. As pointed out

in [18], information may be intended for transmission from here to there (as in communication links) or from now to then (as in data storage and retrieval). A detailed description of each of the blocks in Fig. 2.1 could be found in standard texts [19], [20]. Here, we provide a quick overview of these basic building blocks.

The source encoder digitizes the data and compresses it to remove redundancy. The error correction encoder, on the other hand, adds redundancy to provide robustness against noise and errors which may potentially be introduced by the communication channel. The modulator has the function of mapping bits to a channel waveform. The demodulator, error correction decoder, and the source decoder, perform corresponding reverse operations at the receiver.

The error correction encoding (also referred to as *channel encoding* [13]) will be used in Chapter 3 to reduce the probability of message error at the receiver. An information theoretic view of the communication system is adopted in Chapters 5, 6, and 7, where we derive bounds on achievable rates for reliable communication to an energy constrained receiver. We next provide a brief overview of information theory.

2.2 Information Theory

In 1948, Claude E. Shannon published a paper titled "A Mathematical Theory of Communication" in the *Bell Labs Technical Journal* [1]. This paper, considered as one of the most remarkable papers in the history of engineering, laid the ground work of the scientific discipline, "information theory". It provides fundamental limits on operations such as compression and fast reliable data communication. Although for the first 25 years of its existence, information theory was only considered a beautiful theory, advancements in device technology in later years enabled system development to follow information theoretic principles.

The concept of *information* is related to the concept of *entropy*, which is a measure of uncertainty of a random variable. Formally, the entropy $H(X)$ of a discrete random variable X, taking values in a finite alphabet $\mathcal{X}$, with probability mass function $P_X(x)$ is defined as

$$H(X) \triangleq \sum_{x \in \mathcal{X}} P_X(x) \log_2 \frac{1}{P_X(x)}, \tag{2.1}$$

where the unit of information is called a *bit*. For a given $x \in \mathcal{X}$, smaller values of $P_X(x)$ result in higher values of $\log_2 \frac{1}{P_X(x)}$, and so, the more unlikely the event $X = x$ is, the more information it contains.

Let $P_{XY}(x, y)$ denote the joint distribution of a pair of discrete random variables (X, Y). Then the *conditional entropy* of random variable X, given random variable Y, is defined as

$$H(X|Y) \triangleq \sum_{x,y} P_{XY}(x, y) \log_2 \frac{P_Y(y)}{P_{XY}(x, y)}, \tag{2.2}$$

where $P_Y(y) = \sum_{x \in \mathcal{X}} P_{XY}(x, y)$ is the distribution of Y.

The *mutual information* $I(X;Y)$ is the relative entropy between the joint distribution $P_{XY}(x, y)$ and the product distribution $P_X(x)P_Y(y)$, and is defined as

$$I(X;Y) \triangleq \sum_{x,y} P_{XY}(x, y) \log_2 \frac{P_{XY}(x, y)}{P_X(x)P_Y(y)} \tag{2.3}$$

$$= H(X) - H(X|Y). \tag{2.4}$$

A *discrete channel* is a system consisting of an input alphabet $\mathcal{X}$, an output alphabet $\mathcal{Y}$, and a probability transition matrix $P_{Y|X}(y|x)$ that expresses the probability of receiving symbol y given that symbol x is transmitted. The channel is *memoryless* if the probability distribution of the output depends only on the input

at that time and is conditionally independent of previous inputs and outputs.

Mutual information is an important quantity in characterizing fundamental limits in channel coding problems. Let $\mathcal{M}$ denote a finite message set $\mathcal{M} = \{1, \ldots, M\}$, with $m \in \mathcal{M}$. Fig. 2.2 shows a transmitter wanting to transmit message m to a receiver through a discrete memoryless channel (DMC). A communication engineer designs an encoder which encodes a message into a codeword X^n, which is then transmitted through the DMC in n channel uses. The receiver needs a decoder which recovers the message based on the received signal Y^n. It can be shown [21] that as the number of channel uses n become sufficiently large, with arbitrary small probability of error, the data rate can be made arbitrarily close to the *channel capacity* defined as

$$C \triangleq \max_{P_X} I(X;Y). \tag{2.5}$$

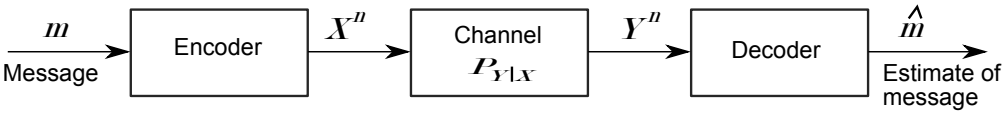

Figure 2.2: A discrete memoryless channel

2.2.1 Binary Block Codes

The encoder in Fig. 2.2 maps message m, belonging to the message set $\mathcal{M}$, to a n-length codeword, X^n. When the input alphabet is binary (i.e. $\mathcal{X} = \{0, 1\}$), the set of codewords form a *binary block code* of length n. If the size of the message set is a power of 2 (i.e. $M = 2^k$), each input message can be represented by k bits. These k input bits to the channel encoder are called *information bits* while the n output bits are called *coded bits*, with $n \geq k$. The *rate* of a binary block code of length n with 2^k codewords is given by $R_c = k/n$, while the *redundancy* is given

by $n - k$.

The *Hamming distance* between two binary n-tuples $\mathbf{x}$ and $\mathbf{y}$, denoted $d(\mathbf{x}, \mathbf{y})$, is the number of places in which they differ. The *minimum distance* of a binary block code $\mathcal{C} = \{\mathbf{c}_i | i = 1, \ldots, 2^k\}$, denoted $d_{\min}$, is the smallest distance between any two distinct codewords of the code, i.e.,

$$d_{\min} \triangleq \min_{i \neq j} d(\mathbf{c}_i, \mathbf{c}_j) \ , \quad \mathbf{c}_i, \mathbf{c}_j \in \mathcal{C}. \tag{2.6}$$

It can be shown [13], [22] that a code with minimum distance $d_{\min}$ can *detect* $d_{\min} - 1$ or fewer bit errors. This code can *correct* t or fewer bit errors as long as $2t < d_{\min}$. In Chapter 3 of this thesis, we use bounds on the performance of best possible binary block codes to reduce the packet delay and energy usage.

Over the years, the design of practical efficient codes has helped us in achieving data transmission rates which are very close to the channel capacity [23]. Next, we provide a brief overview of the Blahut-Arimoto algorithm which helps to numerically compute the capacity for a given discrete memoryless channel.

2.2.2 Blahut-Arimoto Algorithm

An input distribution P_X that maximizes $I(X;Y)$ in (2.5) is called a *capacity-achieving input distribution*. The classical Blahut-Arimoto algorithm [16], [17] solves the problem of computing both the capacity and a capacity-achieving input distribution for a given DMC. If we define

$$Q(x) \triangleq P_X(x) \quad , \quad W(y|x) \triangleq P_{XY}(y|x) \tag{2.7}$$

$$I(Q, W) \triangleq I(X;Y) = \sum_{x,y} Q(x) W(y|x) \log \frac{W(y|x)}{\sum_x Q(x) W(y|x)}, \tag{2.8}$$

then the Blahut-Arimoto algorithm can be described as follows.

Let $Q^{(0)}$ denote an initial chosen input distribution with $Q^{(0)}(x) > 0$ for all $x \in \mathcal{X}$. For iterations $r = 0, 1, 2, \ldots$ perform the following two steps alternatively.

- **Step 1:** For each $x \in \mathcal{X}$, calculate

$$T^{(r)}(x) = \exp \sum_{y \in \mathcal{Y}} W(y|x) \log \frac{Q^{(r)}(x)W(y|x)}{\sum_x Q^{(r)}(x)W(y|x)}. \tag{2.9}$$

- **Step 2:** For each $x \in \mathcal{X}$, the new input probability $Q^{(r+1)}(x)$ is calculated according to

$$Q^{(r+1)}(x) = \frac{T^{(r)}(x)}{\sum_{\tilde{x} \in \mathcal{X}} T^{(r)}(\tilde{x})}. \tag{2.10}$$

For each $r = 0, 1, 2, \ldots$, the sequence $Q^{(r)}$ of input distributions produced by the above two steps fulfills

$$C \geq I(Q^{(r+1)}, W) \geq I(Q^{(r)}, W). \tag{2.11}$$

$Q^{(r)}$ converges to a capacity-achieving input distribution for $r \to \infty$, and

$$\min_x \log \frac{T^{(r)}(x)}{Q^{(r)}(x)} \leq I(Q^{(r)}, W) \leq C(W) \leq \max_x \log \frac{T^{(r)}(x)}{Q^{(r)}(x)}, \tag{2.12}$$

and all the inequalities turn to equalities when $Q^{(r)}$ is a capacity-achieving distribution. Thus, the iterations can be terminated with the assurance that $I(Q^{(r)}, W)$ is within ϵ of capacity when

$$\max_x \log \frac{T^{(r)}(x)}{Q^{(r)}(x)} - I(Q^{(r)}, W) < \epsilon. \tag{2.13}$$

In this thesis, the Blahut-Arimoto algorithm is applied in Chapter 6 to compute the capacity of subblock energy-constrained codes.

2.3 Layering and Network Stack

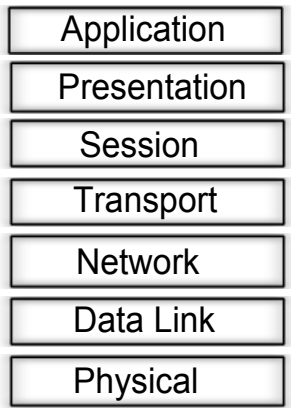

Figure 2.3: The seven layer OSI model

A layered architecture, like the seven-layer open systems interconnect (OSI) model (see Fig. 2.3), divides the overall networking task into layers and defines a hierarchy of services to be provided by the individual layers. The services at the layers are realized by designing protocols for the different layers, and a direct communication between nonadjacent layers is typically forbidden. This modular approach has advantages such as simplicity of design, and availability of standard interchangeable modules. Detailed description of each of the OSI layers can be found in standard texts such as [24] and [15]. In this thesis, we restrict our attention to the *data link layer* and the *physical layer*, and analyze their impact on overall system performance.

The *physical layer* provides a link for transmitting a sequence of bits between any pair of nodes joined by a physical communication channel. Tasks such as error correction coding and modulation are performed at this layer, which aim to provide efficient communication over unreliable channels. The purpose of the *data link layer* is to convert the usually unreliable bit pipe at the physical layer into a higher-level, virtual communication link for sending packets asynchronously but

error-free in both directions over the link. From the standpoint of the link layer, a packet is just a string of bits that comes from the next higher layer. Automatic repeat request (ARQ) is typically implemented at the link layer to detect packet errors at the receiver, and then to request the transmitter to retransmit erroneous packets.

In Chapter 3 of this thesis, we study the joint impact of error correction coding (at the physical layer) and ARQ (at the link layer) on reducing packet delay.

2.4 Queueing Theory

Queueing theory provides a framework for analyzing an important performance measure of data networks: the *average delay* required to deliver a packet from origin to destination. In this thesis, we restrict our attention to the delay at the link layer, where packets are assumed to arrive with a given statistical model. For a given node packet, we define the *packet delay* as the sum of the *queueing delay* (which is the time spent by the packet waiting in the link layer queue) and the *service time* (which is the time taken between the removal of that packet from the link layer buffer for transmission and its error-free delivery to the higher layer at the receiver). We also refer to the queueing delay for a packet as its *waiting time*. We refer the reader to [15], [25] for a detailed review of queueing theory.

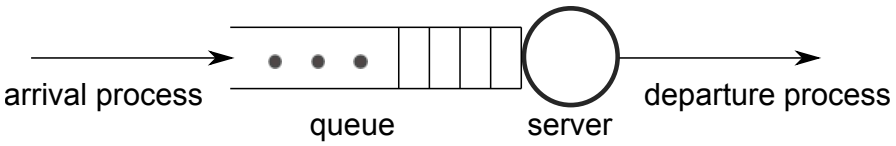

Figure 2.4: A single server queue

A single server queue is illustrated in Fig. 2.4. We are interested in estimating metrics such as the average number of packets in the queue, and the average packet delay. These metrics are estimated in terms of known information such as

the packet *arrival rate* (the average number of packets entering the system per unit time) and the *service rate* (the average number of packets which get served when the server is constantly busy).

If N_{av} denotes the average number of packets waiting in the queue, λ_d denotes the packet arrival rate, and $\overline{T_w}$ denotes the average waiting time, then *Little's Theorem* states that these quantities are related as

$$N_{av} = \lambda_d \overline{T_w}. \tag{2.14}$$

Queueing theory also provides tools to estimate delay in a *multiuser* scenario, when several users share a transmission channel. Such a scenario is considered in Chapter 3 of this thesis, where users are polled in a cyclic order by a central server. A polled user is granted access to the channel for successful transmission of one packet (on top of the queue); once the packet is successfully transferred to the server, the server takes a certain time to switch from one user queue to another during the polling process.

The interesting paradigm of *coupled-queues* is analyzed in detail in Chapter 4 of the thesis, where we consider a communication system where the energy consumed for transmission of data is supplied by an external recharge process. The data packets arrive randomly and wait in a queue for accumulation of sufficient energy and for service completion of previously arrived packets. Here, the data queue dynamics are influenced jointly by the energy arrival process, the data arrival process, and the data service process. In Chapter 4, we formulate a two stage *virtual queueing system* which helps us to obtain closed-form expressions for the average packet delay for different energy arrival statistics.

Next, we present a brief description of the useful *Pollaczek-Khinchin formula* [15] for the average waiting time of a packet in a M/G/1 queueing system,

where M denotes a Poisson packet arrival process, G denotes a general packet service process which is independent of the arrival process, and 1 denotes a single server system. We use the Pollaczek-Khinchin formula in Chapter 4 towards obtaining a closed-form expression for the average packet delay at an energy harvesting transmitter.

2.4.1 Pollaczek-Khinchin Formula

Consider a single-server queueing system where data packets arrive according to a Poisson process with rate λ_d, and the packet service time has a general distribution. Let $\overline{T_s}$ and $\overline{T_s^2}$ denote the mean service time and mean-squared service time, respectively. The *Pollaczek-Khinchin (P-K) formula* states that the mean data packet waiting time, denoted $\overline{T_w}$, can be expressed as [15]

$$\overline{T_w} = \frac{\lambda_d \overline{T_s^2}}{2(1 - \lambda_d \overline{T_s})}, \tag{2.15}$$

where it is assumed that $\lambda_d \overline{T_s} < 1$ to prevent the queue from becoming unbounded in the steady state. Note that when the service time is exponentially distributed (M/M/1 queue), then $\overline{T_s^2} = 2\left(\overline{T_s}\right)^2$, while for a deterministic service process (M/D/1 queue) we have $\overline{T_s^2} = \left(\overline{T_s}\right)^2$. It is interesting to note that the average waiting time for the M/D/1 queue is exactly half their values for the M/M/1 queue with same λ_d and $\overline{T_s}$.

We now give a sketch of the proof of the P-K formula using the concept of mean residual service time. Let T_{s_i} and T_{w_i} denote the service time and the waiting time, respectively, of the i-th data packet. Let T_{r_i} denote the residual service time seen by the i-th packet, i.e., if packet j is already being served when i arrives, then T_{r_i} is the remaining time until packet j service time is complete. If N_i denote the

number of packets found waiting in queue by the i-th packet upon arrival, then

$$T_{w_i} = T_{r_i} + \sum_{j=i-N_i}^{i-1} T_{s_j}. \tag{2.16}$$

By taking the expectation and using the independence of the random variables N_i and $T_{s_{i-1}}, \ldots, T_{s_{i-N_i}}$, we have

$$\overline{T_w} = \mathbb{E}(T_{r_i}) + \overline{T_s}\,\mathbb{E}(N_i) = \overline{T_r} + \overline{T_s} \cdot \overline{N}, \tag{2.17}$$

where $\overline{T_r}$ is the mean residual time and $\overline{N}$ is the expected number of packets in the queue. By Little's law [15], $\overline{N} = \lambda_d \overline{T_w}$, and thus (2.17) can be expressed as

$$\overline{T_w} = \frac{\overline{T_r}}{1 - \lambda_d \overline{T_s}} \tag{2.18}$$

If $M(t)$ denotes the number of service completions within the time interval $[0, t]$, then it follows that [15]

$$\overline{T_r} = \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{M(t)} \frac{\overline{T_{s_i}}}{2} = \lim_{t\to\infty} \frac{1}{2} \frac{M(t)}{t} \sum_{i=1}^{M(t)} \frac{\overline{T_{s_i}}}{M(t)} \tag{2.19}$$

$$= \frac{1}{2} \left(\lim_{t\to\infty} \frac{M(t)}{t} \right) \left(\lim_{t\to\infty} \sum_{i=1}^{M(t)} \frac{\overline{T_{s_i}}}{M(t)} \right) = \frac{1}{2} \lambda_d \overline{T_s^2}, \tag{2.20}$$

and we obtain (2.15) by substituting (2.20) in (2.18).

2.5 Related Work

In this section, we review existing work and present a literature survey related to the different communication systems analyzed in this thesis.

2.5.1 Reducing delay and energy usage via FEC codes

We first consider a system where users are polled for transmission by a central server, and the link layer of each user employs automatic repeat request (ARQ) to ensure error free delivery of packets. In this thesis, we apply bounds on the performance of forward error correction (FEC) codes to quantify their impact on packet delay and energy usage, and show that the knowledge of tradeoffs at the physical layer is vital for optimizing network performance.

Work related to reduction of delay due to FEC codes has been done in different contexts in the past. The joint impact of ARQ and FEC on average service time is analyzed in [26] while its impact on reducing delay in a multi-hop wireless sensor network is investigated in [27]. However, [26] and [27] do not discuss the fundamental limit on delay using the best possible codes.

The use of packet-level FEC (where the message is divided into smaller packets and coding is performed across packets) for reducing the average number of packet transmissions for successful delivery of the message was analyzed for broadcast systems in [28] and for multicast systems in [29].

In [30], the authors analyze the reduction in data retrieval latency due to channel coding in the context of distributed storage. Although the work in [31] focuses on the choice of appropriate modulation and coding scheme (MCS) for improving user throughout in a network, it does not analyze the impact of MCS selection on packet delay.

In [32], a point-to-point wireless packet communication link with truncated ARQ is considered. The transmitter chooses MCS from a fixed set and exhaustive numerical search is employed for finding that MCS which maximizes the throughput subject to a maximum average packet delay and packet loss rate. This exhaustive search does not provide intuitive insight into the trade-off between reducing

packet errors through channel coding and an increase in packet length due to addition of redundancy.

A fixed set of modulation and coding schemes are analyzed in [5] for maximizing spectral efficiency under delay and error performance constraints in wireless fading channels with truncated ARQ. This paper aims at adaptively choosing that modulation and coding scheme which maximizes the spectral efficiency under delay and error performance constraints.

The trade-off between channel coding and ARQ in Rayleigh block-fading channels for maximizing throughput is analyzed in [33]. This work aims at maximizing the long-term average successful throughput by striking a balance between providing sufficient error protection through channel coding and keeping a sufficiently high code rate such that the number of information bits in each coded packet is large. However, [5] and [33] do not address the case where the data to be transmitted arrives in a random fashion and hence ignores the impact of coding on the queueing delay.

The system performance using joint adaptive modulation and coding (AMC) and ARQ is analyzed in [34], [35] where the MCS is chosen from a fixed set, and hence these works do not characterize fundamental performance limits which can be achieved using best possible codes for a given blocklength and coding rate. The use of joint AMC, packet fragmentation, and ARQ is considered in [36] where punctured convolutional codes are employed for FEC. Although [36] analyses the impact of coding and fragmentation on delay and throughput, the issue of characterizing the performance limits using best possible error correction codes is again not addressed.

The queueing delay performance using random linear coding on randomly arriving packets is analyzed in [37] for multicast transmission over packet erasure channels. In [38], the queueing behavior is studied for a discrete-time model where

the packet arrival process is Bernoulli and the communication medium is a bit-erasure channel with memory. It is assumed that random codes with maximum likelihood decoding are used for protection against erasure. However, the analysis is limited to a single user setting.

The delay distribution is characterized in [39] for a point-to-point Markovian modulated binary erasure channel. The use of erasure codes is assumed such that an arbitrarily small error probability can be achieved using any coding rate that is less than the channel capacity. However, such an assumption may not hold when the codeword length is upper bounded [3].

Packet transmission schedules which minimize energy subject to a deadline or a delay constraint are considered in [40]. The work in [40] is motivated by the observation that the energy required to transmit a packet can be reduced by lowering the transmission power and code rate, and therefore transmitting the packet over a longer period of time. In contrast, [41] show that the belief that a longer transmission duration lowers energy consumption may be misleading for short-range applications if the RF circuit power is included in the energy budget.

Tight bounds on the achievable rates for a given communication channel and probability of error were established in [3]. As an application, this work also shows that these bounds could be used to choose the coding rate which maximizes the throughput (or equivalently, minimizes the average service time). However, they did not consider the impact of coding on the packet waiting time.

The impact of link layer scheduling algorithms, other than round-robin polling, on system throughput has also been analyzed in the past. For instance, the greedy maximal scheduling (GMS) [42] determines a schedule by choosing links in decreasing order of queue backlog while conforming to interference constraints. In [43], it was shown that GMS is *throughput-optimal* [44] for networks which satisfy certain conditions. However, GMS may only achieve a fraction of the capacity region on

general network topologies [42], [44].

2.5.2 Packet delay analysis at an energy harvesting transmitter

The ability of energy harvesting communication systems to reduce cost and increase system lifetime has resulted in their use in different practical applications [10]. The research community has aimed to optimize the performance of such systems under varied constraints and assumptions [45]–[57].

Most of the research on energy harvesting communication systems has been on minimizing the data transmission completion time [45]–[47] or maximizing the throughput [48]–[52]. Although the knowledge of future energy and data arrival times is assumed in [45]–[48], such information may not be available in practice owing to the random nature of energy and data arrival processes. In this regard, reasonable assumptions on only the probability distribution of the arrivals, rather than exact knowledge of the time of future arrivals, help to provide a pragmatic approach to system design [49].

The optimal competitive ratio for maximizing the data rate over arbitrary varying energy arrivals and channel fading is derived in [50]. Algorithms for computing the lexicographically maximum data collection rate in energy harvesting sensor networks are presented in [51] under the constraint that no node ever runs out of energy. The work in [52] considers the problem of constructing utility optimal scheduling algorithms in an energy harvesting network where, in every time slot, the network decides on data admission and power allocation over each communication link. Energy management strategies are discussed in [53] for a wireless energy harvesting node with a delay constraint on packet transmission, and it is assumed that a new data packet is generated only after the previous packet is

successfully transmitted or discarded due to delay violation. However, [48]–[53] do not consider the case where the data to be transmitted arrives at random instances and may have to be buffered in a data queue when sufficient energy is not available in the energy buffer.

Random arrival of data and harvested energy is considered in [54] together with the assumption of availability of power from a power grid. The problem of minimizing the long-term average data queue length subject to an average grid power constraint is analyzed. However, the assumption on using grid power in [54] may not hold for low-cost embedded devices. In [58], the energy harvesting process is modeled with a two-state Markov chain for a time-slotted communication system where events occur in each slot with a fixed probability and the probability of event loss due to energy run-out is derived for this model.

In [55], the authors present an energy management scheme where the battery discharge probability falls exponentially with the battery size while the data loss probability falls polynomially with the data buffer size. [56] considers the problem of maximizing the throughput while maintaining the stability of the data queue and proposes online policies for the same. Energy management policies which minimize a linear combination of the mean data queue length and the mean data loss rate are obtained in [57]. Although the buffering of the arriving data is considered in [55]–[57], the problem of quantifying the average delay per data packet is not addressed.

The average packet delay is an important metric for delay sensitive real-time applications. But sufficient attention has not been given to the problem of quantifying the exact average packet delay in an energy harvesting communication system. The fact that the data queue dynamics are influenced jointly by the energy arrival process and the data service process makes this problem hard to solve. In a related work, a closed-form expression for the average delay for event process-

ing is derived in [59] for an energy harvesting system. However, its limitations are that the impact of the event service time on the queueing delay is ignored and the expression for the average event delay is not exact. In [48], an upper bound on the delay is provided in an energy harvesting communication system when the data arrival rate is less than a certain threshold and data buffer size is infinite. This threshold, however, is obtained as a solution to a dynamic programming problem.

In this thesis, we formulate a two stage virtual queueing system which decouples the wait stages for the energy arrival process and the data service process. We assume the knowledge of the distribution of data and energy arrivals which is used to derive closed-form expressions for the average packet delay and the probability of data queue overflow. These expressions are shown to be exact when the service time is negligible, and robust even for relative high values of the average service time. Based on the delay and packet loss requirements for a given application, these expressions can be used to choose design parameters (such as the energy buffer size and the data packet arrival rate) such that the desired performance requirements are satisfied. We also extend the results for the important case of multi-source energy harvesting [60], [61].

2.5.3 Constrained codes for simultaneous energy and information transfer

Although wireless charging of portable electronic devices [62] and implantable biomedical devices [63] has attracted the attention of researchers over the last few years, pioneering work on wireless power transfer was conducted over a century ago by Hertz and Tesla [64]. Similarly, wireless information transfer has a rich history, including works by Popov [65], Bose [66], and Marconi [67]. In fact, Marconi's wireless telegraph device, capable of transatlantic radio communication, helped

save over 700 lives during the tragic accident of the Titanic in 1912 [68]. However, the first work in an information-theoretic setting on analyzing fundamental tradeoffs between *simultaneous* information and energy transfer is relatively recent [14]. The study of simultaneous information and energy transfer is relevant for communication from a powered transmitter to an energy-harvesting receiver which uses the same received signal both for decoding information and for harvesting energy. The energy harvested by the receiver is employed to power its circuitry.

The tradeoff between reliable communication and delivery of energy at the receiver was characterized in [14] using a general capacity-power function, where transmitted codewords were constrained to have average received energy exceed a threshold. This tradeoff between capacity and energy delivery was extended for frequency-selective channels in [69]. Since then, there have been numerous extensions of the capacity-power function in various settings [54], [70]–[72]. Biomedical applications of wireless energy and information transfer have been proposed through the use of implanted brain-machine interfaces that receive data and energy through inductive coupling [63], [73], [74].

Codes with different constraints on the codewords have been suggested in the past, depending on the constraints at the transmitter, the properties of the communication channel, or the properties of the storage medium. For digital information storage on magnetic medium [75], codewords are usually designed to meet the runlength constraint [76] or are optimized for partial response equalization with maximum-likelihood sequence detection (PRML) [77]. The study of information capacity using runlength-limited (RLL) codes on binary symmetric channels (BSC) was carried in [78]–[80]. The use of RLL codes for simultaneous energy and information transfer was proposed and analyzed in [81], [82].

A class of binary block codes called *multiply constant-weight codes* (MCWC), where each codeword of length mn is partitioned into m equal parts and has

weight w in each part, was explored in [83] owing to their potential application in implementation of low-cost authentication methods [84]. Note that MCWC, introduced in [83] as a generalization of *constant weight codes* [85], are themselves a special case of CSCCs with input alphabet size equal to two. When each codeword in an MCWC is arranged as an $m \times n$ array and additional weight constraints are imposed on all the columns, the resulting two-dimensional weight constrained codes have potential application in optical storage systems [86] and in power line communications [87].

Power line communications (PLC) requires the power output to be as constant as possible so that information transfer does not interfere with the primary function of power delivery. One way to achieve this on the PLC channel (which suffers from narrow-band interference, white Gaussian noise, and impulse noise [88]), is to employ *permutation codes* [89] where each codeword of length n is a permutation of n different frequencies, with each frequency viewed as an input symbol. Higher rates of information transfer may be achieved using *constant composition codes* [90] at the cost of local variation in power while ensuring that the power expended is same upon completion of each codeword. When the codeword length is a multiple of the frequency alphabet size, the composition may be chosen such that each frequency occurs equal number times in each codeword [91].

The codewords employed by an energy harvesting transmitter are constrained by the instantaneous energy available for transmission. The capacity of these constrained codes over an additive white Gaussian noise (AWGN) channel has been analyzed when the energy storage capability at the transmitter is zero [92], infinite [93], or some finite quantity [94], [95]. The capacity of an AWGN channel with processing cost at an energy harvesting transmitter was characterized in [96]. The DMC capacity using an energy harvesting transmitter equipped with a finite energy buffer was analyzed in [97]. A comprehensive summary of the recent con-

tributions in the broad area of energy harvesting wireless communications was provided in [11].

One approach to providing real-time energy transfer is to divide each codeword into smaller subblocks, and constrain each subblock to carry sufficient energy. Another approach is to ensure that each codeword carries sufficient energy within a moving time window. In this thesis, we investigate achievable information rates under three different classes of constrained codes for enabling real-time energy transfer. For each of these three classes of constrained codes, we analyze the capacity and highlight the tradeoff between delivery of sufficient energy and achieving high information transfer rates.

Chapter 3

Reducing Delay and Energy Usage Via FEC Codes

We consider a scenario where multiple energy-constrained nodes share a common wireless channel and are polled by a central server for transmission in a round-robin manner. The data to be transmitted by a node arrives randomly at its link layer in the form of fixed length packets. We assume that the link layer of each node uses Stop-and-Wait ARQ [15] to ensure error free transfer of data packets.

For a given node packet, we define the *packet delay* as the sum of the *queueing delay* (which is the time spent by the packet waiting in the link layer queue) and the *service time* (which is the time taken between the removal of that packet from the link layer buffer for transmission and its error-free delivery to the higher layer at the server). The *transmission time* for a packet is the time taken for a single transmission of a packet from a node to the server. We also refer to the queueing delay for a packet as its *waiting time*.

The traditional layered protocol architecture in a communication network has been closely scrutinized by researchers who have used the dependence between protocol layers to propose efficient communication schemes [98]–[100]. The average

packet delay is an important "higher-layer" metric for delay sensitive wireless sensor networks [101]. On noisy communication channels, packets errors may significantly increase the delay due to retransmissions [12]. Although the packet error probability (PEP), and hence the number of retransmissions, can be reduced by employing physical layer forward error correction (FEC) codes (also called channel codes) [13], the use of FEC codes results in addition of redundant bits which increase the packet transmission time. A small number of redundant bits may not sufficiently reduce the PEP, while a large number of redundant bits may increase the packet delay due to the high transmission time.

We remark that although efforts have been made to marry the disciplines of networking and information theory [2], [102], some gaps remain in their respective approaches. For instance, since the probability of packet error (and hence the retransmissions) increases with an increase in length of uncoded packets [12], a typical approach in network protocol design is to limit the packet size. However, from information theory we know that the packet error probability on noisy channels can be *reduced* through channel coding by fixing the coding rate and increasing the packet length [3], [103]. Since by fixing the coding rate we fix the fraction of redundant bits in the coded packet, an increase in the packet length may actually help in increasing the throughput (by reducing the PEP) [3]. This observation gives an intuitive explanation on why our proposed scheme of opportunistic combining and encoding of data packets (Sec. 3.6) helps in reducing delay. Our work in this chapter is a step towards extending the notion that information theoretic results can provide meaningful design guidelines for practical communication networks.

In this chapter, we first establish that relative to an uncoded system, it is sufficient to reduce the average service time (AST) using FEC in order to achieve lower average packet delay. We then study and quantify the reduction in AST

that can be achieved using the best possible FEC codes. Further, we show that when the energy per transmitted bit is kept constant, the ratio of average transmit energy for the coded and uncoded system is same as the ratio of the AST for the coded and uncoded system. Thus, the percentage reduction in AST due to coding translates into an equal percentage reduction in average transmit energy.

The specific findings and contributions from our work are as follows: (i) We provide several bounds on the reduction in AST using the best possible FEC codes. (ii) We give a sufficient condition when no FEC scheme can help reduce the AST. (iii) For Gaussian channels, we find that a relatively high PEP ($\sim 10^{-2}$) obtained using as high a coding rate as possible, typically results in sufficiently small AST. (iv) The performance of optimum maximum likelihood decoding can be approached by a lower complexity bounded distance decoder. (v) Average packet delay can be further reduced in certain cases by opportunistically combining and encoding several packets jointly.

The remainder of this chapter is structured as follows. The system model is described in Sec. 3.1, and sufficient conditions for reducing average packet delay are derived in Sec. 3.2. The metric for quantifying the reduction in average service time is defined in Sec. 3.3 while in Sec. 3.4 we show that the percentage reduction in AST due to coding translates into an equal percentage reduction in average energy for successful packet transfer. FEC codes which minimize average service time are discussed in Sec. 3.5. Reduction in average packet delay via opportunistic packet combining is discussed in Sec. 3.6 and numerical examples are presented in Sec. 3.7. Finally the results are summarized in Sec. 3.8.

3.1 System Model

We consider a scenario where nodes share a common physical channel and are polled by a central server for transmission in a round-robin manner. On being polled, a node transmits if it has data waiting to be sent; else the server polls the next node in the polling list for transmission. The communication model assumes that the link layer of each node uses Stop-and-Wait ARQ, and that the data packets received in error are retransmitted until their successful reception is acknowledged by the receiver. We assume that the ACK/NACK response upon reception of data packet is error-free and instantaneous.

Note that the Stop-and-Wait ARQ protocol is used by the medium access control (MAC) layer of the IEEE 802.11 Wireless LAN standard [104], and the Bluetooth standard [105] for packets protected by the cyclic redundancy check (CRC) bits. The use of ARQ for delay sensitive multimedia applications has been investigated in [106].

In our model, a polled node is granted access to the channel for successful transmission of one packet (on top of the queue); once the packet is successfully transferred to the server, the server takes a certain time to switch from one node queue to another during the polling process. This switching time can be used for exchange of control information between the server and the node before the actual transmission of data packets.

Note that the polling based multiple access scheme has been shown to achieve higher throughput than the contention based approach under heavy traffic conditions [107] and is implemented in the IEEE 802.11 MAC sub-layer as a Point Coordination Function (PCF) [104]. The polling model has also been proposed for use in body area networks [108].

We assume that for all the nodes, the packet arrival process at their respective

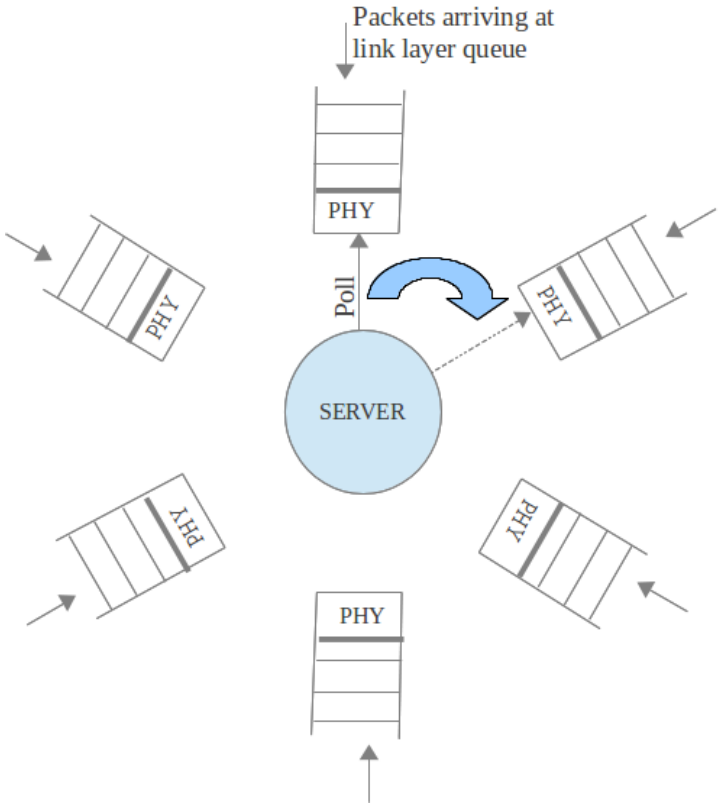

Figure 3.1: A central server polling the nodes in a cyclic order

link layers is Poisson with rate λ_d (measured in terms of packets per second) and that the arrival process for different nodes are independent of each other. The modulation type and the symbol rate for transmission are assumed to be fixed for all the nodes.

The data packet length and the probability of packet error for all nodes are assumed to be same. We assume that each data packet contains necessary header fields for its identification. For error detection, CRC parity check bits are added to each packet, and then FEC code is employed to reduce the probability of packet error. We shall refer to this system model as a Poisson Arrival Polling Model with Packet Retransmission (PAPMPR). A packet with no FEC will be called an *uncoded* packet. The various parameters used in this article are summarized in Table 3.1.

Variables	Description
λ_d	Data packet arrival rate for each node
k	Length of an uncoded packet (including CRC bits)
n	Length of a packet encoded using forward error correction
p	Probability of packet error
T	Time taken to transmit one uncoded packet
R_c	Coding rate ($R_c = k/n$)
$\overline{T_s}$	Average service time for a data packet
$\overline{T_s^2}$	Mean squared service time for a data packet
$\overline{T_w}$	Average waiting time (in queue) for a data packet
$\overline{T_D}$	Average packet delay
$\overline{T_v}$	Average time to switch from one node to another
σ_v^2	Variance of the queue switching time
N	Number of nodes
ρ	Total load offered to the system ($\rho = N\lambda_d\overline{T_s}$)
$\nu^\star$	Metric to quantify reduction in average service time
q	BSC crossover probability

Table 3.1: Description of variables

3.1.1 Average service time of data packets

Let T denote the time taken to transmit an uncoded packet of length k. If a node employs a rate $R_c \leq 1$ code on k bits, then the number of coded bits become $n = k/R_c$. Since the modulation type and the symbol rate for transmission are fixed, an increase in the packet length from k to n increases the packet transmission time from T to T/R_c. Let p denote the probability of packet error and T_s denote the *service time* for error-free transfer of a node packet. Then T_s is a discrete geometric random variable with probability mass function given by

$$Pr\left(T_s = j\frac{T}{R_c}\right) = p^{j-1}(1-p) \quad j = 1, 2, \ldots \tag{3.1}$$

Note that the above expression uses the assumption that the ACK/NACK response upon reception of data packet is error-free and instantaneous.

If $\overline{T_s}$ and $\overline{T_s^2}$ denote the AST and mean squared service time, respectively, for

the transfer of a packet from a node to the server, then

$$\overline{T_s} = \sum_{j=1}^{\infty} \left(j \frac{T}{R_c} \right) \left(p^{j-1}(1-p) \right) = \frac{T}{R_c(1-p)}, \tag{3.2}$$

$$\overline{T_s^2} = \sum_{j=1}^{\infty} \left(j^2 \frac{T^2}{R_c^2} \right) \left(p^{j-1}(1-p) \right) = \frac{T^2(1+p)}{R_c^2(1-p)^2} = \left(\overline{T_s} \right)^2 (1+p). \tag{3.3}$$

The first and second moments of the service time for an uncoded system can be obtained by substituting $R_c = 1$ in the above equations.

3.1.2 Average waiting time of data packets

Let N denote the number of nodes in the system, $\overline{T_v}$ denote the average time taken by the server to switch from one node queue to another, and σ_v^2 denote the variance of the queue switching time. If ρ denotes the total offered load to the system [25], then

$$\rho = N \lambda_d \overline{T_s}. \tag{3.4}$$

Let $\overline{T_w}$ denote the average waiting time for a node packet in the link layer queue. Using the concept of *mean residual service time* and the fact that the order in which the packets are served in a cyclic polling process is independent of the service time, we have [15]

$$\overline{T_w} = \frac{\sigma_v^2}{2\overline{T_v}} + \frac{N\lambda_d \overline{T_s^2} + (N+\rho)\overline{T_v} + \sigma_v^2 N \lambda_d}{2(1-\rho-N\lambda_d \overline{T_v})}. \tag{3.5}$$

Note that the condition for $\overline{T_w}$ to be bounded is $\rho + N\lambda_d \overline{T_v} < 1 \iff N\lambda_d(\overline{T_s} + \overline{T_v}) < 1$. This condition (rather than the condition $\rho < 1$) is required because the server consumes some queue switching time after serving each packet, thereby effectively increasing the average service time from $\overline{T_s}$ to $\overline{T_s} + \overline{T_v}$.

Remark : Note that the average waiting time for the special case of $N = 1$

reduces to the *Pollaczek-Khinchin formula* [15]

$$\overline{T_w} = \frac{\lambda_d \overline{T_s^2}}{2(1-\rho)}, \tag{3.6}$$

since there is no queue switching in a single node case.

3.1.3 Average packet delay

If $\overline{T_D}$ denotes the average packet delay, then we have $\overline{T_D} = \overline{T_s} + \overline{T_w}$, where $\overline{T_s}$ and $\overline{T_w}$ are given by (3.2) and (3.5), respectively.

3.2 Sufficient condition for reducing delay

In this section we compare two coded systems and derive sufficient conditions under which one coded system results in lower average packet delay compared to the other. The following theorem will aid us in expressing the sufficient conditions explicitly in the form of corollaries.

Theorem 1. *For PAPMPR, let* $\overline{\tilde{T}_s}$, $\overline{\tilde{T}_s^2}$, *and* $\overline{\tilde{T}_D}$ *denote the AST, mean squared service time, and average packet delay, respectively, of a coded system. Similarly, let* $\overline{\hat{T}_s}$, $\overline{\hat{T}_s^2}$, *and* $\overline{\hat{T}_D}$ *denote the AST, mean squared service time, and average packet delay, respectively, of another coded system. Then,*

$$\overline{\tilde{T}_s} < \overline{\hat{T}_s}, \quad \overline{\tilde{T}_s^2} < \overline{\hat{T}_s^2} \implies \overline{\tilde{T}_D} < \overline{\hat{T}_D}. \tag{3.7}$$

Proof: Please refer to Appendix A.1. ■

We use Theorem 1 to prove the following corollary comparing the average packet delay of coded and uncoded systems.

Corollary 1. *Let $R_c, \tilde{p}, \overline{\tilde{T}_s}, \overline{\tilde{T}_s^2}$, and $\overline{\tilde{T}_D}$ denote the rate, PEP, AST, mean squared service time, and average packet delay, respectively, for the coded system. Let $p, \overline{T_s}, \overline{T_s^2}$, and $\overline{T_D}$ denote the PEP, AST, mean squared service time, and average packet delay, respectively, for the* uncoded *system. Then,*

$$\overline{\tilde{T}_s} < \overline{T_s} \implies \overline{\tilde{T}_D} < \overline{T_D}. \tag{3.8}$$

Proof: We shall first prove that

$$\overline{\tilde{T}_s} < \overline{T_s} \implies \overline{\tilde{T}_s^2} < \overline{T_s^2}. \tag{3.9}$$

From (3.2), we have

$$\begin{aligned} \overline{\tilde{T}_s} < \overline{T_s} \quad &\iff \quad \frac{T}{R_c(1-\tilde{p})} < \frac{T}{1-p} \\ &\iff \quad R_c(1-\tilde{p}) > 1-p \qquad (3.10) \\ &\implies \quad 1-\tilde{p} > 1-p, \ (\text{since } R_c < 1) \\ &\iff \quad \tilde{p} < p. \qquad (3.11) \end{aligned}$$

Using (3.3) and (3.11), we get (3.9). Now, (3.8) follows from (3.7) and (3.9). ■

We now consider two different coded systems each with coding rate less than one. The following corollaries state sufficient conditions under which the coded system with lower AST also results in lower average delay.

Corollary 2. *Let $\tilde{p}, \overline{\tilde{T}_s}, \overline{\tilde{T}_s^2}$, and $\overline{\tilde{T}_D}$ denote the PEP, AST, mean squared service time, and average packet delay, respectively, for one coded system; and let $\hat{p}, \overline{\hat{T}_s}, \overline{\hat{T}_s^2}$, and $\overline{\hat{T}_D}$ denote the PEP, AST, mean squared service time, and average packet delay, respectively, for another coded system. If $\tilde{p} < \hat{p}$ and $\overline{\tilde{T}_s} < \overline{\hat{T}_s}$, then it implies $\overline{\tilde{T}_D} < \overline{\hat{T}_D}$.*

Proof: Since $\tilde{p} < \hat{p}, \overline{\tilde{T}_s} < \overline{\hat{T}_s}$, we have $\overline{\tilde{T}_s^2} < \overline{\hat{T}_s^2}$ (use (3.3)). Now the claim follows from (3.7). ■

Corollary 3. *Let $\tilde{p}$ and $\hat{p}$ denote the PEP for two different coded system, and let $0 < \epsilon < 1$. If $\tilde{p} < \epsilon$ and $\overline{\tilde{T}_s} < \overline{\hat{T}_s}\left(1 - \frac{\epsilon}{2}\right)$, then $\overline{\tilde{T}_D} < \overline{\hat{T}_D}$.*

Proof:

$$\overline{\tilde{T}_s} < \overline{\hat{T}_s}\left(1 - \frac{\epsilon}{2}\right) \implies \frac{\overline{\hat{T}_s}}{\overline{\tilde{T}_s}} > 1 + \frac{\epsilon}{2}. \tag{3.12}$$

Thus,

$$\left(\frac{\overline{\hat{T}_s}}{\overline{\tilde{T}_s}}\right)^2 > 1 + \epsilon > 1 + \tilde{p} \geq \frac{1 + \tilde{p}}{1 + \hat{p}}. \tag{3.13}$$

Hence $\overline{\hat{T}_s^2} > \overline{\tilde{T}_s^2}$ (from (3.3)) and the claim follows from (3.7). ■

Corollary 4. *Let $\overline{\tilde{T}_s}$ and $\overline{\hat{T}_s}$ be the AST for two different coding schemes having the same coding rate R_c. Then, $\overline{\tilde{T}_s} < \overline{\hat{T}_s}$ implies $\overline{\tilde{T}_D} < \overline{\hat{T}_D}$.*

Proof: Since both coding schemes have the same coding rate, from (3.2) and (3.3) we have

$$\overline{\tilde{T}_s} < \overline{\hat{T}_s} \implies \tilde{p} < \hat{p} \implies \overline{\tilde{T}_s^2} < \overline{\hat{T}_s^2}, \tag{3.14}$$

and the corollary follows from (3.7). ■

The above results imply that for a broad set of cases, the coded system which results in lower AST also results in lower average packet delay. The minimization of average service time results in a simpler optimization problem which is independent of system parameters such as packet arrival rate and the number of nodes in the system.

The following proposition is useful since it quantifies the reduction in average packet delay in terms of the reduction in AST.

Proposition 1. *For PAPMPR, let $\tilde{p}, \overline{\tilde{T}_s}, \overline{\tilde{T}_s^2}, \overline{\tilde{T}_w}$, and $\overline{\tilde{T}_D}$ denote the rate, PEP, AST, mean squared service time, average waiting time and average packet delay,*

respectively, for the coded system. Let $p, \overline{T_s}, \overline{T_s^2}, \overline{T_w}$, *and* $\overline{T_D}$ *denote the PEP, AST, mean squared service time, average waiting time and average packet delay, respectively, for the* uncoded *system. Further, let* k *denote the uncoded packet length. If* $\overline{\tilde{T}_s} < \overline{T_s}$ *and the queue switching time is negligible (that is,* $\overline{T_v} \to 0$ *and* $\sigma_v^2 \to 0$*), then*

$$\frac{\overline{\tilde{T}_D}}{\overline{T_D}} < \frac{\overline{\tilde{T}_s}}{\overline{T_s}}. \tag{3.15}$$

Proof: Please refer to Appendix A.2. ■

Remark: The converse to the above proposition need not always be true: a coded system with lower average packet delay, compared to an uncoded system, may have relatively higher AST. This is because the average packet delay depends on both the first *and* second moments of the service time. The following example illustrates this scenario.

Example: Consider a single-node system where packets, each comprising of 120 bits, arrive at the node link-layer at a rate of $\lambda_d = 500$ packets/sec. Assume that the packet transmission time for the uncoded packet is $T = 1$ ms. When the packets are transmitted over a channel with bit error rate of 0.004, the PEP for the uncoded packet is 0.38, while the corresponding AST is 1.6176 ms. However, using a rate $R_c = 3/5$ code, the PEP can be reduced to less than 10^{-5} (refer Sec. 3.5.1), but the AST increases to 1.6667 ms since the packet transmission is increased due to addition of redundancy. However, in this scenario, the average packet delay for the *uncoded* case is 6.3458 ms while the delay for the *coded* case is 5.8334 ms, since the coded system has lower mean squared service time due to its lower PEP.

3.3 Quantifying the reduction in average service time

From the corollaries in Section 3.2 we know that for a wide variety of cases, the codes which are designed to reduce the AST also achieve lower average packet delay. In this section, we formulate a metric which measures the maximum possible reduction in the AST through channel coding.

Let the probability of error for an uncoded packet of length k be denoted as p. By an (n, k) binary code, we shall mean a code where each codeword length is n and the number of codewords are 2^k. Let $\mathcal{C}_{n,k}$ denote an (n, k) code with PEP $\tilde{p}$. We define

$$\nu(\mathcal{C}_{n,k}) = \frac{n(1-p)}{k(1-\tilde{p})}. \tag{3.16}$$

If $\overline{\tilde{T}_s}$ and $\overline{T_s}$ denote the AST for the coded and uncoded systems, respectively, then from (3.2) and (3.16), we note that

$$\frac{\overline{\tilde{T}_s}}{\overline{T_s}} = \frac{n(1-p)}{k(1-\tilde{p})} = \nu(\mathcal{C}_{n,k}). \tag{3.17}$$

Now, we define

$$\nu^\star(k) = \min_{n,\,\mathcal{C}_{n,k}} \nu(\mathcal{C}_{n,k}), \tag{3.18}$$

where, for a fixed value of k, the minimization above is over all (n, k) block codes with $n \geq k$. Thus $\nu^\star(k)$ is a measure of how much the AST of a coded system be reduced relative to the AST of an uncoded packet of length k.

The following proposition shows that $\nu^\star(k)$ obtained for relatively small values of k can also be used to give an upper bound on $\nu^\star(mk)$ corresponding to larger uncoded packet sizes with length mk where m is a positive integer.

Proposition 2. *Let m be an integer greater than one. Then,*

$$\nu^{\star}(mk) \leq (\nu^{\star}(k))^{m}\,. \tag{3.19}$$

Proof: Please refer to Appendix A.3. ■

3.4 Transmit energy usage

Here, we compare the transmit energy usage for successful transfer of a node packet for the uncoded and coded systems. Let E denote the energy spent in transmitting L *uncoded* bits from a node to the server in a single transmission of the data packet. If the packet transmission is in error, then the packet has to be re-transmitted and every such re-transmission results in an energy spending of E. Let p denote the probability of packet error for the uncoded case and let E_s denote the total energy spent for *successful* transfer (including energy used during re-transmissions) of a packet. Then E_s is a discrete geometric random variable with probability mass function given by

$$P\left(E_s = kE\right) = p^{k-1}(1-p), \qquad k = 1, 2, \ldots \tag{3.20}$$

If $\overline{E_s}$ denote the mean energy spent in successful transfer of packet between the node and the server, then

$$\overline{E_s} = \frac{E}{1-p}. \tag{3.21}$$

Now, if a rate R error correcting code is employed by a node on L information bits, then the effective packet length becomes $\tilde{L} = L/R$. If we assume that the transmit energy per coded bit is equal to the energy per uncoded bit and $\tilde{E}$ denotes the energy spent in transmitting the coded packet, then $\tilde{E} = E/R$. If $\tilde{p}$ denotes

the probability of packet error and $\overline{\tilde{E}_s}$ denotes the average transmit energy for successful packet transfer for the coded case, then

$$\overline{\tilde{E}_s} = \frac{\tilde{E}}{1-\tilde{p}} = \frac{E}{R(1-\tilde{p})} \tag{3.22}$$

Comparing (3.21) and (3.22) we note that the average transmit energy for coded case is less than the uncoded case if and only if

$$R(1-\tilde{p}) > 1-p \tag{3.23}$$

We note that the condition in (3.23) is same as the condition for the average service time of the coded system to be lower than the average service time of the uncoded system (see (3.10)).

Further, using (3.2), (3.21), and (3.22), we have

$$\frac{\overline{\tilde{E}_s}}{\overline{E_s}} = \frac{\overline{\tilde{T}_s}}{\overline{T_s}} \tag{3.24}$$

and hence compared to an uncoded system, the percentage reduction in average service time by employing channel coding also results in an equal reduction in average transmit energy.

3.5 On Codes which Minimize Average Service Time

In this section, we investigate codes which result in smallest possible AST relative to an uncoded system. The following proposition gives a necessary condition for a FEC code to result in lower AST compared to an uncoded system.

Proposition 3. *Let k and p denote the length and the PEP, respectively, for an* uncoded *packet. Then a necessary condition for an (n, k) binary error correcting code to result in lower AST relative to the uncoded system is*

$$n < \frac{k}{1-p}. \tag{3.25}$$

Proof: Let $\tilde{p}$ denote the PEP using the (n, k) code. Let $\overline{T_s}$ and $\overline{\tilde{T}_s}$ denote the AST for the uncoded and coded systems, respectively. Then from (3.17),

$$\overline{\tilde{T}_s} < \overline{T_s} \iff n < \frac{k(1-\tilde{p})}{(1-p)} \implies n < \frac{k}{1-p}, \tag{3.26}$$

where the last inequality follows since $0 < \tilde{p} < 1$. ■

The following proposition shows that when the PEP for the uncoded packet is sufficiently small, then no FEC scheme can reduce the AST.

Proposition 4. *When the PEP, p, of an* uncoded *packet of length k satisfies $p < 1/(k+1)$, then the AST cannot be reduced through any FEC code.*

Proof: When $p < 1/(k+1)$, then we have $\frac{k}{1-p} < k+1$. This inequality together with the statement of Proposition 3 implies that a necessary condition for an (n, k) binary code to reduce the average service time is given by

$$n < \frac{k}{1-p} < k+1. \tag{3.27}$$

The above equation shows that n cannot exceed k and AST is minimized when $n = k$ (corresponding to the uncoded case). ■

Remark: In view of the above proposition, it is interesting to note that the Bluetooth standard for exchanging data over short distances provides the option

of not applying FEC when PEP for the uncoded packet is small [105].

From (3.17) and (3.18) we observe that in order to select the code which minimizes the AST, we need to consider the joint impact of the coding rate $R_c = k/n$ and the associated PEP $\tilde{p}$.

For a fixed k, the choice of n for which an (n, k) code minimizes the AST (and hence yields $\nu^\star(k)$ in (3.18)) is not obvious. A large value of n may offset the reduction in packet retransmissions by the penalty due to an increase in the packet transmission time, while a small value of n may not sufficiently reduce the PEP. However, from Proposition 3 we know that n which minimizes the AST lies in the range $k \leq n < k/(1-p)$ where p is the PEP for the uncoded packet.

For obtaining $\nu^\star(k)$ we require, for a given n, the knowledge of the *smallest* PEP which can be obtained using an (n, k) code. Towards this, we discuss useful bounds on PEP (Sec. 3.5.1) which are applied, in turn, to bound $\nu^\star(k)$.

We consider an additive white Gaussian noise (AWGN) channel with hard-decision decoding at the receiver [19]. We assume that the energy per coded bit is fixed, independent of the coding rate, and binary signaling is employed. Thus, the channel seen from the output of the channel encoder to the input of the corresponding decoder is equivalent to a binary symmetric channel (BSC) [21]. Further, since the energy per coded bit is kept fixed (independent of the coding rate), the crossover probability of the equivalent BSC does not change upon changing the coding rate.

When the energy per coded bit is E_c and the noise power spectral density is N_0, the BSC crossover probability, denoted by q, is given by

$$q = \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\frac{E_c}{N_0}}\right), \tag{3.28}$$

where erfc is the complementary error function:

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-t^2} dt \ . \tag{3.29}$$

We now investigate bounds on the PEP which will be applied to bound the reduction in AST using (3.16) and (3.18).

3.5.1 PEP using the best possible codes

We analyze bounds on $\tilde{p}$, the PEP using FEC, which will be applied to characterize the maximum reduction in AST, and hence the packet delay, achievable using best possible coding schemes. We consider both lower and upper bounds on $\tilde{p}$ in the following.

Upper bound on $\tilde{p}$ using bounded distance decoding

If d denotes the minimum distance of an (n, k) code, then the bounded distance decoder correctly decodes the received word if and only if the error pattern has weight less than or equal to $t = \lfloor \frac{d-1}{2} \rfloor$ [22]. In this case, the PEP on BSC with crossover probability q is given by

$$\tilde{p} = 1 - \sum_{i=0}^{t} \binom{n}{i} q^i (1-q)^{n-i}. \tag{3.30}$$

This bounded distance decoder has the advantage of having lower computational complexity than the Maximum-Likelihood (ML) decoder [109].

For $k \leq n < 256$, we select codes with best known minimum distance [110],[111] and employ bounded distance decoding.

For moderate to large packet sizes, a lower bound on the minimum distance may be obtained using the Gilbert-Varshamov (GV) bound [22]. The GV bound

for binary linear codes states that it is possible to construct an (n, k) code with minimum distance at least d for which the following inequality holds:

$$\sum_{i=0}^{d-2} \binom{n}{i} \geq 2^{n-k}. \tag{3.31}$$

If $\delta = d/n$ denotes the relative distance of the code, then the asymptotic form of GV bound for binary linear codes states that for every $\delta < 0.5$ and $0 < \epsilon \leq 1 - \mathcal{H}(\delta)$, there exists a code with rate $R_c \geq 1 - \mathcal{H}(\delta) - \epsilon$, where $\mathcal{H}(\delta)$ is the binary entropy function $\mathcal{H}(\delta) = -\delta \log_2 \delta - (1-\delta)\log_2(1-\delta)$.

For a binary code with rate R_c, the relative GV distance $\delta_{GV}(R_c)$ is defined as the root $\delta < 0.5$ of the equation $\mathcal{H}(\delta) = 1 - R_c$. If we define $t = \lfloor \frac{n\delta_{GV}(R_c)-1}{2} \rfloor$, then an upper bound on the probability of error may be obtained using (3.30).

Random-coding union bound on $\tilde{p}$

The upper bound on the PEP obtained in (3.30) uses only the knowledge of the minimum distance of the code. A tighter upper bound can be obtained by using the complete *spectral* information of the code. Note that the *spectrum* of a linear code is the set $(A_0, A_1, \ldots, A_n)$ where A_w denotes the number of codewords of weight w. In case we have the knowledge of the spectrum, the following upper bound on the probability of error by Polytrev may be applied [3], [112] to BSC with crossover probability q:

$$\tilde{p} \leq \sum_{l=0}^{n} q^l (1-q)^{n-l} \min \left\{ \binom{n}{l}, \sum_{w=0}^{n} A_w B(l, w, n) \right\}, \tag{3.32}$$

where

$$B(l, w, n) = \sum_{w/2 \leq r \leq \min\{l, w\}} \binom{w}{r} \binom{n-w}{l-r}. \tag{3.33}$$

As mentioned in [3], the upper bound in (3.32) can be extended to a random

binary linear code to obtain (3.35). In the following, we provide a sketch of the steps required to derive (3.35) from (3.32).

If $\mathbf{m}$ is any binary vector of length k, then for a random generator matrix $\mathbf{G}$, $\mathbf{mG}$ is a uniformly distributed over the space of length n binary vectors. Since there are $\binom{n}{w}$ binary vectors of length n and weight w, the probability that $\mathbf{mG}$ has weight w is given by $\binom{n}{w}/2^n$. Further, since there are 2^k codewords, the expected number of codewords of weight w are

$$\mathbb{E}[A_w] = 2^k \times \frac{\binom{n}{w}}{2^n} = 2^{k-n}\binom{n}{w}. \tag{3.34}$$

Note that for a fixed l, $\min\left\{\binom{n}{l}, \sum_{w=0}^{n} A_w B(l,w,n)\right\}$ is a concave function of $\{A_w\}$. This follows because $\sum_{w=0}^{n} A_w B(l,w,n)$ is a linear (and hence concave) function of $\{A_w\}$, and the minimum of two concave functions is again concave [113]. The Jensen's inequality for concave functions states that if $f(x)$ is a concave function of x, then $\mathbb{E}[f(x)] \leq f(\mathbb{E}[x])$. Thus averaging (3.32) over the ensemble of randomly chosen generator matrices, we obtain the expression for the average probability of error as

$$\begin{aligned}
\tilde{p} &\leq \mathbb{E}\left[\sum_{l=0}^{n} q^l (1-q)^{n-l} \min\left\{\binom{n}{l}, \sum_{w=0}^{n} A_w B(l,w,n)\right\}\right] \\
&= \sum_{l=0}^{n} q^l (1-q)^{n-l}\, \mathbb{E}\left[\min\left\{\binom{n}{l}, \sum_{w=0}^{n} A_w B(l,w,n)\right\}\right] \\
&\overset{(a)}{\leq} \sum_{l=0}^{n} q^l (1-q)^{n-l} \min\left\{\binom{n}{l}, \sum_{w=0}^{n} \mathbb{E}[A_w] B(l,w,n)\right\} \\
&\overset{(b)}{=} \sum_{l=0}^{n} q^l (1-q)^{n-l} \min\left\{\binom{n}{l}, \sum_{w=0}^{n} 2^{k-n}\binom{n}{w} B(l,w,n)\right\},
\end{aligned} \tag{3.35}$$

where (a) follows from Jensen's inequality and (b) results from (3.34).

Approximating $\tilde{p}$ using Normal Approximation

If $\tilde{p}$ denotes the target PEP, then the maximum number of information bits which can be carried by the packet using the best possible coding scheme can be approximated using normal approximation [3], [114] as

$$k \approx nC - \sqrt{nV}Q^{-1}(\tilde{p}) + \frac{1}{2}\log_2 n, \tag{3.36}$$

where $C = 1+q\log_2 q+(1-q)\log_2(1-q)$ is the channel capacity for the BSC, $V = \sqrt{q(1-q)}\log_2\frac{1-q}{q}$ is the channel dispersion for the BSC, and Q is the Gaussian Q-function.

The above approximation can be used, in turn, to get a measure of the PEP when the best possible channel coding scheme is used. When the number of uncoded and coded bits are denoted as k and n, respectively, then the PEP using the best possible coding scheme can be approximated as

$$\tilde{p} \approx Q\left(\frac{nC - k + 0.5\log_2 n}{\sqrt{nV}}\right). \tag{3.37}$$

Sphere-Packing lower bound on $\tilde{p}$

The sphere-packing lower bound on the PEP for a (n,k) binary linear code on a BSC with crossover probability q is [22]

$$\tilde{p} \geq \sum_{i=t+1}^{n}\binom{n}{i}q^i(1-q)^{n-i} - \alpha_{t+1}q^{t+1}(1-q)^{n-(t+1)}, \tag{3.38}$$

where t is the greatest integer such that

$$\alpha_{t+1} = 2^{n-k} - 1 - \binom{n}{1} - \binom{n}{2} - \cdots - \binom{n}{t} \geq 0. \tag{3.39}$$

The lower bound given by (3.38) is tight in the sense that equality is achieved for *perfect* (where $\alpha_{t+1} = 0$) and *quasi-perfect* codes [22]. Note that the sphere-packing bound on $\tilde{p}$ is the smallest error probability that can be achieved using any (n, k) binary code. Thus, it acts as a performance benchmark for any practical encoding-decoding scheme on a BSC.

The bounds on the PEP discussed in this section will be employed to obtain bounds on the reduction in AST through channel coding which are illustrated through numerical examples in Sec. 3.7. In the next section, we show that packet delay may be further reduced by opportunistically combining packets and jointly encoding the combined packet.

3.6 Opportunistic Packet Combining for Reducing Delay

Consider the scenario where a node polled for transmission has several packets in its link layer queue which are waiting for their transmission. So far, we have seen that performing channel coding on data packet of fixed size k can potentially lead to significant reduction in the average packet delay. We now enquire if additional reduction in packet delay can be obtained if the node *opportunistically* combines two or more packets in its link layer queue into larger 'super-packet' and then perform error correction coding on such super-packets. We consider only the single user scenario, and thus no switching among queues is required.

Through an appropriate use of channel coding, an increase in packet length, in general, can help in reducing the PEP on noisy channels when the overall coding rate is kept constant, as long as the transmission rate is below channel capacity [3], [103]. This observation can be argued using the *reliability function* of a noisy channel [103], [115]. In fact, the PEP can be made arbitrarily small

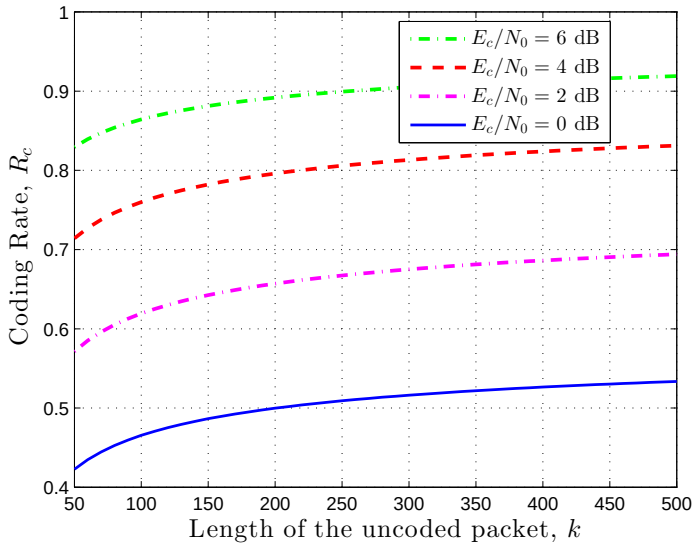

Figure 3.2: Coding rate R_c versus length of the uncoded packet, k, for a fixed packet error probability of $p = 10^{-2}$.

by sufficiently increasing the packet length (provided the transmission rate is less than the channel capacity).

Equivalently, an increase in packet length can help to increase the coding rate (and thus reduce the fraction of redundant bits) while keeping the PEP constant. Fig. 3.2 shows that when the PEP is fixed to 10^{-2}, the required coding rate increases as a function of the uncoded packet size. The curves are obtained for different E_c/N_0 using normal approximation (3.36).

In this section, we study a scheme where a node *opportunistically* combines and jointly encodes m data packets when the data queue has m packets waiting for transmission, where m is an integer greater than one. Thus, in this *opportunistic* transmission scheme, either the data packets are transmitted independently (when the data queue has less than m packets) or m data packets are encoded and transmitted together (when the data queue has at least m packets).

Since each uncoded packet consists of necessary headers (for identification) and CRC bits (for error detection), the formation of super-packets leads to reduction

in overhead since only one set of header and CRC bits are required for each super-packet. Further, by use of appropriate channel coding, the PEP can be reduced while keeping the same coding rate. However, on the other hand, since packet combining increases packet length, the average delay experienced by the first packet in the super-packet may increase due to increase in transmission time. We want to know if, on average, the packet delay may be reduced by formation of larger super-packets.

We now present conditions when *opportunistic* packet combining leads to a reduction in delay for the single node case. Let each data packet comprise of k_I information bits and k_O overhead bits consisting of CRC and header bits. Thus, we have

$$k = k_I + k_O. \tag{3.40}$$

Since the time taken for transmission of k bits is given by T, the time taken for transmitting k_I bits is θT where $\theta = k_I/k$. Similarly, the time taken for transmission of k_O bits is $(1-\theta)T$.

Assume that a node has m data packets waiting for transmission in its data queue, where m is an integer greater than one. When each data packet is encoded and transmitted one at a time, then let R_1 and p_1 denote the coding rate and probability of packet error, respectively, corresponding to the coding scheme which minimizes $\nu^\star$. On the other hand, when the node chooses to form a super-packet by combining the m waiting packets, let R_m and p_m denote the coding rate and probability of packet error, respectively, corresponding to the coding scheme which minimizes $\nu^\star$. The following proposition gives conditions under which the opportunistic combining and joint encoding of packets leads to lower delay in a single use scenario.

Proposition 5. *Consider an opportunistic transmission scheme where the trans-*

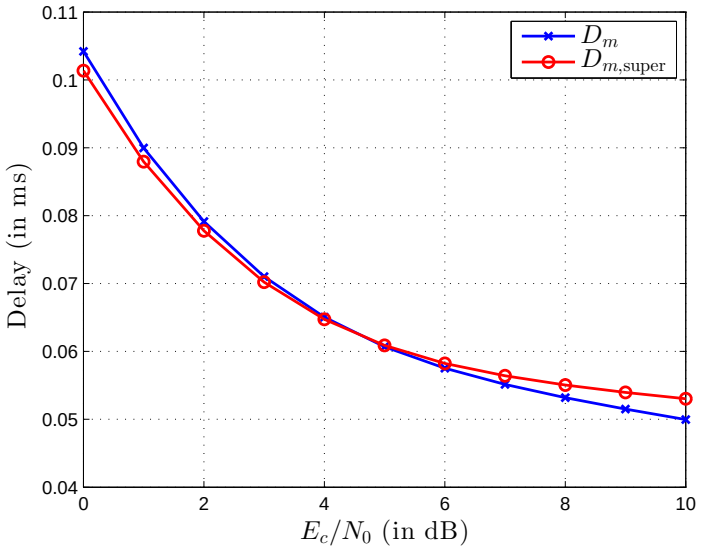

Figure 3.3: Comparison of D_m and $D_{m,\text{super}}$ for $k = 70$, $k_O = 32$ and $m = 6$

mitter jointly encodes m packets whenever there are m data packets waiting in the data queue. Compared to the conventional scheme, where each data packet is always transmitted independently, the opportunistic transmission scheme results in lower average packet delay if

$$\frac{R_m(1-p_m)}{R_1(1-p_1)} > \frac{2(m-1)\theta+2}{m+1}. \tag{3.41}$$

Proof: Please refer to Appendix A.4. ■

Since a larger packet size helps in obtaining a higher coding rate while maintaining the same PEP (see Fig. 3.2), it follows that $R_m(1-p_m) > R_1(1-p_1)$. Thus, the condition given by (3.41) implies that average packet delay can be reduced via packet combining, provided the reduction in redundancy and overhead is sufficient to overcome the increase in transmission time.

Consider a scenario where a node has m packets waiting for transmission at time $t = 0$. Let D_m denote the mean delay for these m waiting packets, relative to $t = 0$, when each packet is encoded and transmitted independently. Let

$D_{m,\text{super}}$ denote the mean delay for these m waiting packets, relative to $t = 0$, when these packets are combined into a super-packet and encoded jointly. As shown in Appendix A.4, the condition given by (3.41) is equivalent to $D_{m,\text{super}} < D_m$.

Fig. 3.3 compares D_m and $D_{m,\text{super}}$ for $k = 70$, $k_O = 32$ and $m = 6$. The BPSK symbol rate is assumed to 5 Mbps, and hence $T = 14\,\mu s$. The values of R_1, p_1, and R_m, p_m (for generating the respective curves) are obtained using the normal approximation (3.37). When E_c/N_0 is low, the probability of error, and hence the average delay, is high. The formation of super-packet and joint encoding of the combined packets with increased length in case $E_c/N_0 < 4$ dB helps to bring the probability of error in the vicinity of 10^{-2} using sufficiently higher rate code (compared to independent transmission of each packet) such that (3.41) is satisfied.

Remark: When m packets are waiting in the data queue at $t = 0$, and these packets are encoded and transmitted independently, the delay relative to $t = 0$ experienced by the first packet (on top of queue) is low while the delay for mth waiting packet is relatively high. However, all the m packets experience the same delay (relative to $t = 0$) in case of joint encoding of the super-packet which also aids in reducing the peak delay.

3.7 Numerical Examples

3.7.1 Using a fixed set of coding schemes

We consider a system where the packet arrival rate per node is $\lambda_d = 1000$ packet/sec, $\sigma_v^2 = 0$, and the number of information bits in each data packet are fixed to $k = 120$ bits. We assume that the transmitter employs binary phase shift keying (BPSK) modulation with a symbol rate of 5×10^6 symbols/sec. Hard-decision

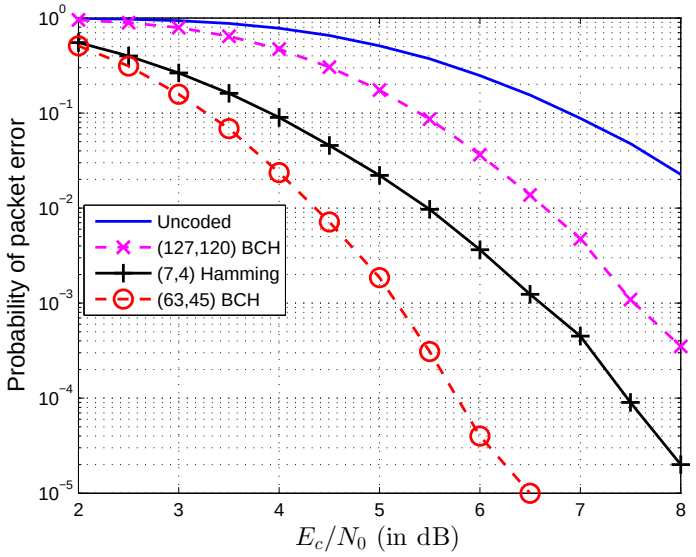

Figure 3.4: Probability of packet error for the uncoded and coded cases. Each data packet comprises of 120 information bits.

decoding is employed at the receiver.

We seek to compare the average packet delay for the uncoded and coded case where we consider the following candidate binary linear block codes: a $(127, 120)$ single-error correcting Bose-Chaudhuri-Hocquenghem (BCH) code, a $(63, 45)$ triple-error correcting BCH code, and a $(7, 4)$ single-error correcting Hamming code [22].

Since each transmitted packet comprises of 120 information bits, the packet consists of 30 codewords using the $(7, 4)$ Hamming code. When $(63, 45)$ BCH code is employed, the information bits are padded to 135 bits prior to encoding which ensures that integer number of code blocks are present in the transmitted packet, and thus each packet in this case comprises of 3 codewords.

Fig. 3.4 plots Monte Carlo simulation results for the probability of packet error for the uncoded and coded cases as a function of E_c/N_0 where E_c is the energy per coded bit and N_0 is the one-sided noise power spectrum density. The results show that the triple-error correcting $(63, 45)$ BCH code yields the lowest error

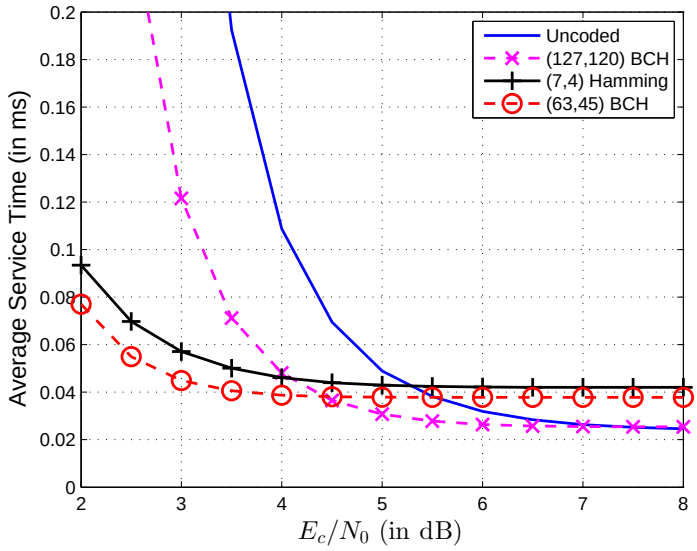

Figure 3.5: Average service time as a function of SNR

probability.

Figs. 3.5 and 3.6 plot the average service time and the average packet delay, respectively, as a function of E_c/N_0 when the number of nodes in the system are $N = 4$ and the queue switching time is $T_v = 0.0024$ ms. These figures indicate that, in general, a coding scheme having lower average service time also results in lower average packet delay, hence corroborating the claims made in Sec. 3.2. When the signal-to-noise ratio (SNR) is low, the probability of packet error for the uncoded case is high and the large number of retransmissions result in high average packet delay. The reduction in error probability from using the triple-error correcting $(63, 45)$ BCH code results in lowest average packet delay when E_c/N_0 is between 2 dB and 4 dB. However, when the SNR is relatively high, the average packet delay is dominated by the packet transmission time of each node (since retransmissions are less likely). When E_c/N_0 varies from 5 dB and 7 dB, the $(127, 120)$ coding scheme with a relatively high coding rate of $120/127$ results in the lowest average packet delay. Moreover, when $E_c/N_0 = 8$ dB the uncoded scheme yields the smallest packet delay since the uncoded error probability is

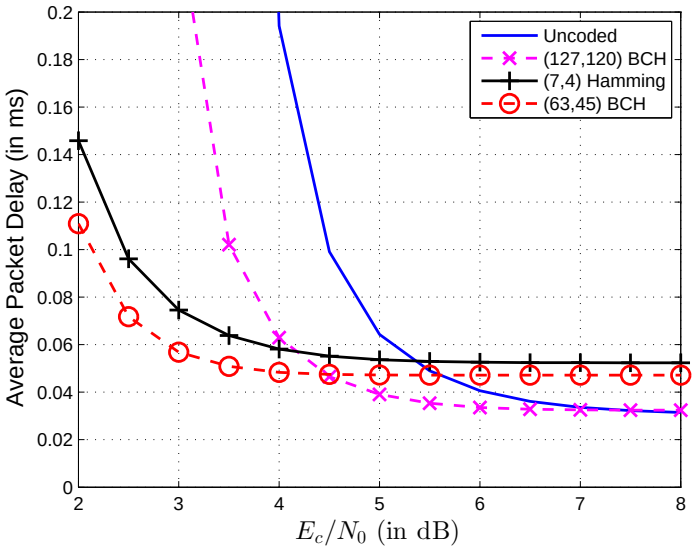

Figure 3.6: Average packet delay as a function of SNR

sufficiently small (close to 2×10^{-2}); any further reduction in error probability through channel coding gets offset by the increase in transmission time due to addition of redundancy.

3.7.2 Performance limits using the best possible coding scheme

In this subsection, we quantify the reduction in average service time through the use of the best possible coding scheme. We first consider the case where the packet lengths are relatively small (less than 200 bits). For bounded distance decoding of these small blocklengths, we select codes with best known minimum distance [110], [111]. Later, we consider the case where the packet lengths are relatively large (between 500 to 5000 bits). For bounded distance decoding of these large packet lengths, we use the asymptotic form of the Gilbert-Varshamov bound to come up with lower bound on the minimum distance of best codes.

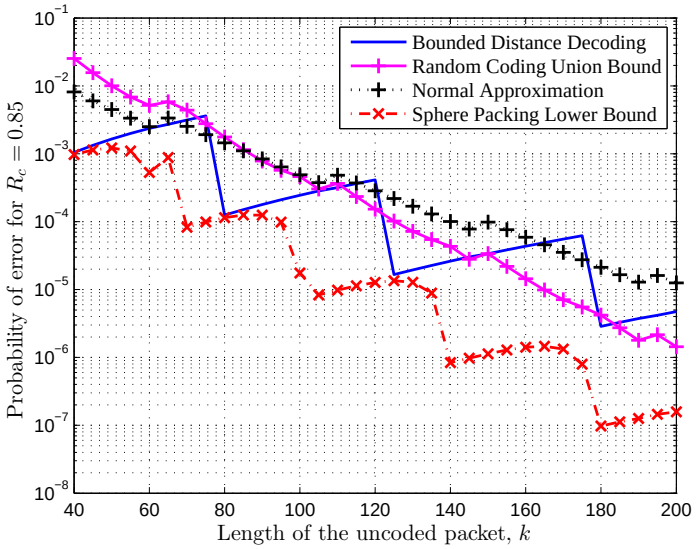

Figure 3.7: Probability of packet error with a fixed coding rate of $R_c = 0.85$ and $E_c/N_0 = 6.8$ dB ($q = 0.001$).

Relatively small packet lengths

Fig. 3.7 plots the different bounds and approximations on the probability of error with coding rate of $R_c = 0.85$ with $E_c/N_0 = 6.8$ dB. This value of E_c/N_0 implies that the equivalent BSC has crossover probability of roughly $q = 0.001$ using (3.28). The length of the coded packet, n, is obtained by rounding k/R_c to the closest integer. The figure shows that for $k = 40$, 80, and 125, the probability of error with bounded distance decoding is very close to the sphere packing bound. This implies that the corresponding codes are close to being 'perfect'. Note that the zigzagging of the curves occurs because the code size is of the form 2^k where k is restricted to be an integer; similar observations were made in [3].

Fig. 3.8 plots $v^\star$ for $E_c/N_0 = 6.8$ dB ($q = 0.001$). For $k \leq 80$, the PEP for the uncoded system (given by $1 - (1-q)^k$) is less than 0.077. For these lengths, from (3.17) we observe that in order for a coded system to achieve lower average service time, we need a coding scheme with rate R_c resulting in PEP $\tilde{p}$ such that $R_c(1-\tilde{p}) > 1 - 0.077 = 0.923$. Fig. 3.8 shows that no such coding scheme exists

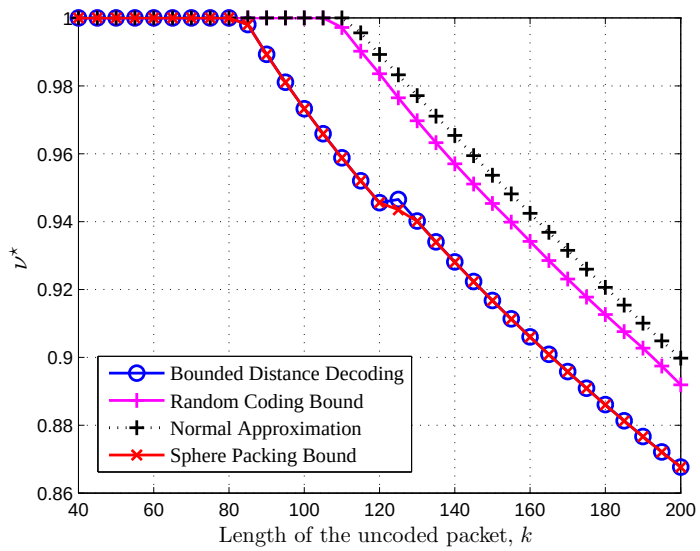

Figure 3.8: $v^\star$ versus uncoded packet length for $E_c/N_0 = 6.8$ dB ($q = 0.001$).

and hence the uncoded system is the most efficient in this case. It can be verified that when $k \leq 80$, the maximization of $R_c(1-\tilde{p})$ using the sphere-packing bound (3.38) occurs at $R_c = 1$, which corresponds to the uncoded case.

One might be tempted to only use short packets in networking, since uncoded is optimal. However, packets in the network stack have header information, which carries no information. So the shorter the packet, the higher the cost of overhead. Hence longer packets are used in practice to mitigate the cost of overhead.

However, the PEP for the uncoded system increases with an increase in the length of the packet. Higher values of PEP increase the number of retransmissions, which in turn lead to an increase in the AST. Thus for relatively large values of k, the reduction in PEP through channel coding at the cost of increase in the packet length has the potential of reducing the overall AST. Indeed, the AST using bounded distance decoding is reduced to less than 87 percent of the AST for an uncoded system for $k = 200$ when a single error correcting $(208, 200)$ code is employed.

Although $\nu^\star$ typically decreases with the uncoded packet length k, Fig. 3.8

shows an interesting phenomenon where $\nu^\star$ increases from $k = 120$ to $k = 125$ for the case of *bounded distance decoding*. We first note that if $R_{c,k}$ and $\tilde{p}_k$ denote the coding rate and PEP, respectively, which achieves $\nu^\star(k)$, then

$$\frac{\nu^\star(125)}{\nu^\star(120)} = (1-q)^5 \times \frac{R_{c,120}(1-\tilde{p}_{120})}{R_{c,125}(1-\tilde{p}_{125})},$$

where $q = 0.001$ is the crossover probability which models the channel in Fig. 3.8. Next, numerical optimization over best possible codes shows that $R_{c,120} = 0.945$, $\tilde{p}_{120} = 0.0074$, while for $k = 125$ we have $R_{c,125} = 0.939$, $\tilde{p}_{125} = 0.0080$. Applying these values in the above equation shows that $\nu^\star(125) > \nu^\star(120)$. We remark that this phenomenon of $\nu^\star$ increasing with k is quite rare because PEP typically decreases with an increase in k and R_c.

Fig. 3.8 also shows that the performance of bounded distance decoding using codes with large minimum distance is very close to the performance limit of any coded system given by the sphere packing bound. This is somewhat surprising since the channel capacity using bounded distance decoding, denoted C_B, on a BSC with crossover probability q is bounded as [116]

$$1 - \mathcal{H}(2q) \leq C_B \leq 1 - \mathcal{H}(0.5 - 0.5\sqrt{1-4q}),\ q < 0.25, \tag{3.42}$$

and is strictly less than the BSC channel capacity $1 - \mathcal{H}(q)$, where $\mathcal{H}$ denotes the binary entropy function.

Some light can be shed on the effectiveness of bounded distance decoding for these packet lengths by noting that the coding rate R_c which minimizes the AST for bounded distance decoding is greater than 0.94 while the corresponding probability of error is around 10^{-2} (refer Table 3.2). Since the AST is minimized when $R_c(1-\tilde{p})$ is maximized, the numerical results outline the fact that instead of

	BDD		RCB		NA		SPB	
k	R_c	$\tilde{p}$	R_c	$\tilde{p}$	R_c	$\tilde{p}$	R_c	$\tilde{p}$
40	1.0000	0.0392	1.0000	0.0392	1.0000	0.0392	1.0000	0.0392
80	1.0000	0.0770	1.0000	0.0770	1.0000	0.0770	1.0000	0.0770
120	0.9449	0.0074	0.9160	0.0078	0.9160	0.0078	0.9449	0.0074
160	0.9524	0.0126	0.9302	0.0131	0.9195	0.0134	0.9524	0.0125
200	0.9615	0.0188	0.9259	0.0014	0.9259	0.0014	0.9615	0.0187

TABLE 3.2: R_c AND $\tilde{p}$ WHICH MINIMIZE $\nu^\star(n,k)$ FOR GIVEN k AND $q = 0.001$

	BDD		RCB		NA		SPB	
k	R_c	$\tilde{p}$	R_c	$\tilde{p}$	R_c	$\tilde{p}$	R_c	$\tilde{p}$
40	0.8696	0.0775	0.7692	0.0154	0.8000	0.0894	0.8696	0.0764
80	0.8511	0.0687	0.8081	0.0775	0.8163	0.0757	0.8696	0.0727
120	0.8451	0.0551	0.8219	0.0598	0.8219	0.0598	0.8633	0.0515
160	0.8333	0.0449	0.8290	0.0457	0.8290	0.0457	0.8377	0.0130
200	0.8333	0.0349	0.8333	0.0349	0.8333	0.0349	0.8584	0.0306

TABLE 3.3: R_c AND $\tilde{p}$ WHICH MINIMIZE $\nu^\star(n,k)$ FOR GIVEN k AND $q = 0.01$

lowering the PEP arbitrarily, it is required that $\tilde{p}$ is brought in the vicinity of 10^{-2} by employing very high code rates. As an example, we note that a coding scheme with rate 0.9 and PEP 10^{-2} achieves lower AST than a coding scheme with rate 0.89 and PEP 10^{-6} since the former results in a higher value of $R_c(1-\tilde{p})$.

Remark: We use the following acronyms in Tables 3.2 and 3.3. (i) BDD: Bounded Distance Decoding, (ii) RCB: Random Coding Bound, (iii) NA: Normal Approximation, (iv) SPB: Sphere Packing Bound.

The value of $v^\star$ when $E_c/N_0 = 4.3$ dB ($q = 0.01$) is plotted in Fig. 3.9 for $k \in [40, 200]$. When $q = 0.01$, the PEP for the uncoded packet for these lengths is relatively high, ranging from 0.33 to 0.87. When channel coding is employed in this case, the coding rate which minimizes the AST is between 0.8 to 0.9, while the corresponding PEP is in the range $[10^{-1}, 10^{-2}]$ (refer Table 3.3). This again shows that instead of trying to make PEP arbitrarily small, it is enough to get it close to 10^{-2} using high rate codes. As shown in Fig. 3.9, the AST using channel

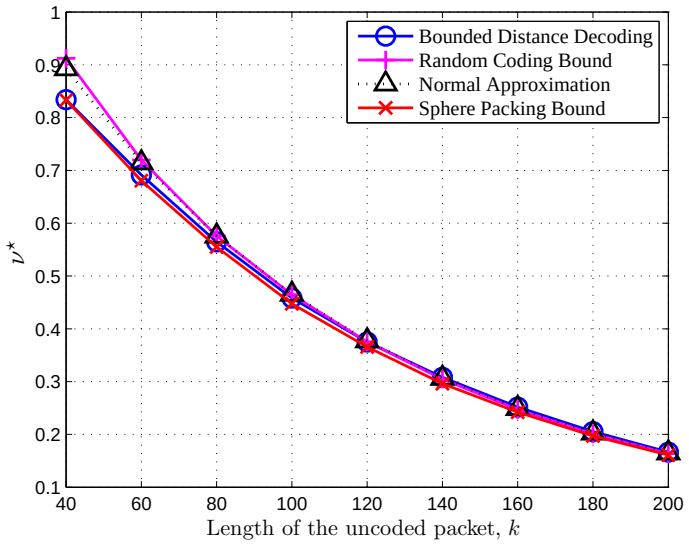

Figure 3.9: $v^\star$ versus uncoded packet length for $E_c/N_0 = 4.3$ dB ($q = 0.01$).

coding is reduced to less than one-fifth of the AST for the uncoded packet when $k = 200$. This shows that the potential for reducing the AST using coding is higher when the PEP for the uncoded system is relatively high. Since the PEP for the uncoded system increases with an increase in the packet length, the reduction in the AST is even more significant for larger packet lengths.

Relatively large packet lengths

Here, we use the asymptotic form of the Gilbert-Varshamov bound for obtaining a lower bound on the minimum distance of good codes. The asymptotic form of the Gilbert-Varshamov bound states that there exists a length n block code of rate R_c for which the minimum distance approaches $n\delta_{GV}(R_c)$ where $\delta_{GV}(R_c)$ is the root of the equation $\mathcal{H}(\delta) = 1 - R_c$. For large packet lengths, the asymptotic version of the sphere-packing bound states that the number of correctable errors is approximately $n\delta_{GV}(R_c)$ [22].

Fig. 3.10 plots $\nu^\star$ for $E_c/N_0 = 6.8$ dB and $k \in [500, 5000]$. It shows that the AST, with an appropriate channel coding scheme, can be reduced to less than

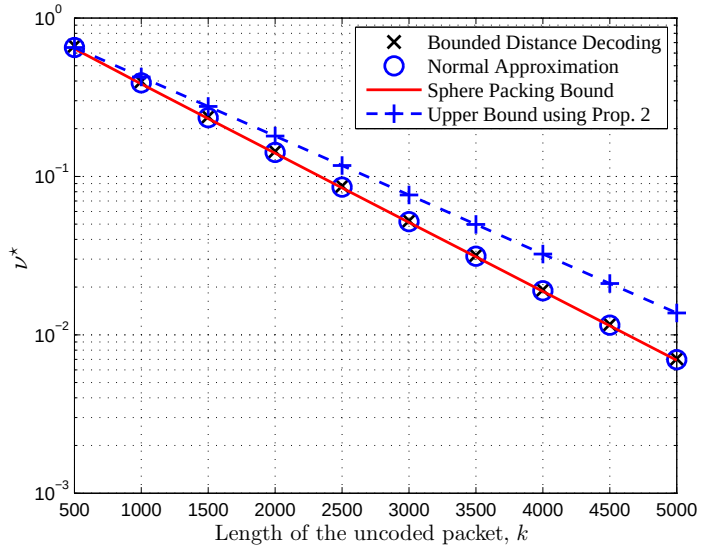

Figure 3.10: $\nu^{\star}$ versus uncoded packet length for $E_c/N_0 = 6.8$ dB ($q = 0.001$).

one-tenth the AST for the uncoded case for $k \geq 2500$. The upper bound on $\nu^{\star}$ is obtained using Proposition 2 where the base value for $k = 500$ is initialized using the normal approximation. The figure shows that the bound is relatively tight for smaller values of m, but gradually becomes weak as m increases. We observe that the performance of bounded distance decoding is again close to the best achievable performance (characterized by the sphere-packing bound). The coding rates which yielded these performance were higher than 0.9 for packet lengths $500 \leq k \leq 5000$ and the corresponding probability of packet error was around 10^{-2}. This again shows that good performance may be achieved when the coding scheme yields a probability of error close to 10^{-2} with a code rate as high as possible.

3.8 Discussion

In this chapter, we applied bounds on the performance of FEC codes to quantify their impact on packet delay, and showed that the knowledge of tradeoffs involved in the choice of physical layer parameters is vital for optimizing network

performance. This work is a step towards extending the notion that information theoretic results can provide meaningful design guidelines for practical communication networks.

We applied results from multi-terminal queueing theory to prove that compared to an uncoded system it is sufficient to reduce the AST using channel coding in order to achieve lower packet delay. The minimization of AST posed a simpler optimization problem, independent of system parameters such as packet arrival rate and the number of nodes in the system. When the queue switching time in the polling system is negligible, we proved that the percentage reduction in average packet delay is at least as much as the percentage reduction in AST obtained using channel coding.

We analyzed the tradeoff between a reduction in packet retransmission and an increase in transmission time due to addition of redundancy using channel coding. For instance, it was shown that a coding scheme with PEP 10^{-2} and coding rate 0.9 results in lower AST compared to a code with PEP 10^{-6} and coding rate 0.89.

The conventional viewpoint in network protocol design is to limit the size of data packets since the probability of error of an uncoded packet increases with packet length. However, we showed that when an appropriate channel coding scheme is employed, the average packet delay can be *reduced* by opportunistically combining several packets and jointly encoding the combined packets.

We showed that network performance can be enhanced by tailoring the physical layer parameters to the specific channel conditions. The key takeaways can be succinctly summarized as follows: (i) instead of trying to make the PEP close to zero via FEC codes, we should aim to bring it close to 10^{-2} using as high a coding rate as possible, (ii) the performance of maximum likelihood decoding can be approached by using a bounded distance decoder, and (iii) delay may be further reduced in certain cases by opportunistically combining and encoding

several packets jointly.

Further, we showed that the percentage reduction in AST due to channel coding translates into an equal reduction in average transmit energy required for successful packet transfer when the energy per coded bit is fixed.

Chapter 4

Packet Delay Analysis at an Energy Harvesting Transmitter

Consider an energy harvesting transmitter which uses the harvested energy for transmission of data packets. The data packets arrive randomly and wait in a queue for accumulation of sufficient energy and for service completion of previously arrived packets. Thus, the data queue dynamics are influenced jointly by the energy arrival process, the data arrival process, and the data service process. We formulate a two stage *virtual queueing system* which decouples the wait stages for the energy arrival process and the service process. This decoupling helps us to obtain closed-form expressions for the average packet delay and the probability of data packet loss due to buffer overflow.

We first derive the results for single-source energy harvesting, and then extend them for the useful case of multi-source energy harvesting. The derived closed-form expressions are shown to be exact when the service time becomes negligible. We provide Monte Carlo simulations and show the robustness of these expressions even for relative high values of the average service time.

The results in this chapter should be useful to both system designers and re-

searchers. Using the closed-form expressions derived in this chapter, an energy harvesting system designer may appropriately choose the system parameters (for example, the energy buffer size, the data packet arrival rate, and the packet transmission rate) such that the desired performance levels are met. The derived expression for the average packet delay is used to numerically quantify the maximum throughput under an average delay constraint.

The rest of the chapter is organized as follows. The system model is presented in Sec. 4.1 and the virtual queueing system is introduced in Sec. 4.2. The *unit-energy arrival* case, where each energy arrival brings energy which is sufficient to transmit one data packet, is analyzed in Sec. 4.3. The *fractional-energy arrival* case, where each energy arrival brings only a fraction of the energy required to transmit one data packet, is presented in Sec. 4.4. The *bulk-energy arrival* scenario, where each energy arrival brings sufficient energy for transmission of multiple data packets, is analyzed in Sec. 4.5. We extend the approach used for analyzing fractional and bulk-energy arrivals to the useful case of *multi-source energy harvesting* in Sec. 4.6, where energy is harvested from two independent sources, with potentially different quantity and rate of energy arrival from each source. In Sec. 4.7, we show that the derived closed-form expressions can be applied to quantify the maximum throughput which the system can support while meeting the required quality of service (QoS) constraints. The Monte Carlo simulation results for a variety of cases are presented in Sec. 4.8 which help us to draw insights regarding the impact of design parameters on packet delay. Concluding remarks and discussions are presented in Sec. 4.9.

4.1 System Model

We consider a communication system (see Fig. 4.1) where the energy consumed for transmission of data is supplied by an external recharge process. The energy arrivals are modeled as a Poisson process and each arrival corresponds to a fixed amount of energy. The Poisson energy arrivals model piezoelectric energy harvesting from mechanical vibrations [117]–[119] and this energy model has been used previously in [120], [121]. We also note that commercial products based on harvesting piezoelectric energy from passing vehicles on roadways have already been built [122]. We assume that the energy storage process is ideal with no energy leakage [59].

The data to be transmitted is assumed to arrive in a fixed packet size, independent of energy arrivals. We assume that each data packet consumes one unit of energy for transmission. This is not too simplifying an assumption since the energy arrivals could correspond to unit-energy arrivals (one energy packet required for transmission of one data packet), or fractional-energy arrivals (multiple energy packets required for one data packet), or bulk-energy arrivals (one energy packet is sufficient for transmission of multiple data packets).

Data packet arrivals are modeled as a Poisson process. This model for real-time arrival of data packets is common and has been used in energy harvesting communication systems [59], [117], [118], [120] and sensor networks [123]–[125]. We assume that receiving or processing of data packets does not consume significant amount of energy and that data transmission is the only energy consuming task. This assumption is justified when the distance between the transmitter and the receiver is large and the energy spent in transmission dominates the total energy consumption. Equivalently, we may assume that the power management at the transmitter uses a dedicated battery for low energy consumption tasks (such as

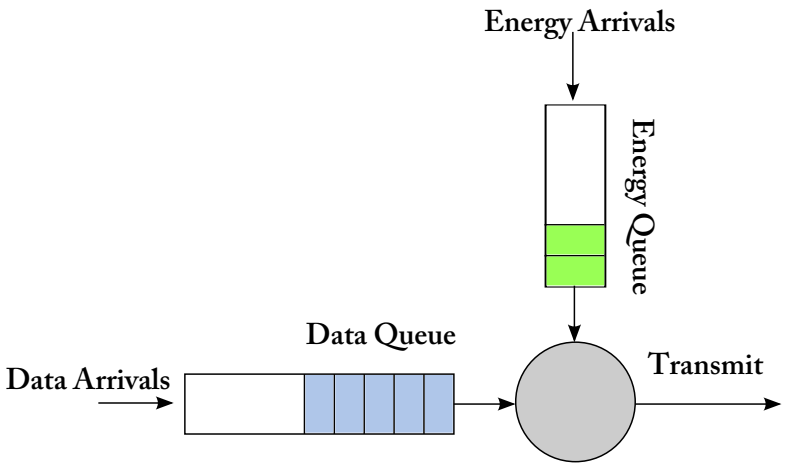

Figure 4.1: Model of the physical queueing system in an energy harvesting transmitter having bounded energy and data storage.

receiving and processing packets) while the harvested energy is used for packet transmission.

The service (transmission) time of each data packet may have a general distribution. The general distribution may be specialized to the deterministic case where the transmission time is fixed. The general distribution of the packet service time can also capture the time variation in the channel access in case of a time-shared channel. We assume that the transmit power is sufficient to meet the required probability of packet error for the given communication channel and hence packet retransmissions are not considered.

The energy buffer can store up to B_e units of energy and the data buffer has a capacity of B_d data packets. The data or energy arrivals which find the corresponding queue to be full are lost. The transmission of an arriving data packet starts as soon as the following two conditions are met: (i) at-least one unit of energy is present in the energy buffer and (ii) the transmission of previously queued data packets has completed. If either of these two conditions are not met, the data packet is either queued (when the data buffer is not full) or else discarded (when the data buffer is full). Note that the energy level in the energy

buffer reduces by one unit after transmission of each data packet.

We refer to the queueing system for this system model as the *Physical Queueing System* (see Fig. 4.1).

We define *data packet delay* to be the time between the arrival of a data packet at the data queue of the energy harvesting transmitter and the completion of its transmission. Thus, data packet delay is the sum of the time spent by the packet waiting in the data queue and the transmission time. We define *data packet loss* to be the event when an arriving data packet is discarded due to the data buffer being full.

We derive closed-form expressions for the average data packet delay and the probability of data packet loss. In order to obtain the above metrics of interest, we formulate a two stage *Virtual Queueing System* which is described next.

4.2 Virtual Queueing System

The virtual queueing system decouples the wait stages for the energy arrival process and the service process as shown in Fig. 4.2. In *stage one*, each data packet waits in the queue until it is paired with one unit of energy in a First-Come-First-Served (FCFS) manner. The buffer sizes for the data and the energy queues in the first stage of the virtual queueing system are B_d and B_e, respectively, which are *equal* to the corresponding buffer sizes in the physical queueing system. Also, the data and energy arrival processes in the first stage are *same* as that in the physical queueing system.

The data packets after being paired with one unit of energy, instantly move to the *stage two* of the virtual queueing system. The second stage is assumed to have infinite storage capacity and the service process for the data packets in the second stage is same as that for the physical queueing system.

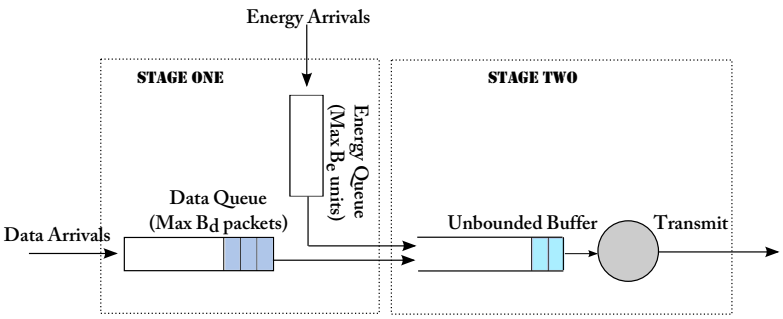

Figure 4.2: Model of the two stage *virtual queueing system*. Arriving data packets which find sufficient energy in stage one energy queue instantly move to the second stage where they wait for their turn to get transmitted.

4.2.1 Comparing the Physical and Virtual Queueing Systems

We now compare the virtual queueing system with the physical queueing system. We will refer to each energy arrival as an *energy packet*. Table 4.1 summarizes the important variables that are used in this chapter. $\overline{T_W}$ denotes the average waiting time for data packets in the physical queueing system. $\overline{T_{W1}}$ and $\overline{T_{W2}}$ denote the average waiting time for data packets in the first and second stages of the virtual queueing system, respectively. These average waiting times correspond to only those data packets which are queued. Data packets which are lost due to buffer overflow are ignored in the delay analysis. The mean transmission time of data packets is denoted by τ. Since the packet delay is the sum of waiting and transmission time, the average data packet delay in the virtual queueing system, denoted by $\overline{T_D}$, is given by

$$\overline{T_D} = \overline{T_{W1}} + \overline{T_{W2}} + \tau. \tag{4.1}$$

The probability of data packet loss in the first stage of the virtual queueing system is denoted as P_{loss}. We will use the closed-form expressions for $\overline{T_D}$ and

Variables	Description
B_d	Buffer size of data queue for stage one of virtual queue
B_e	Buffer size of energy queue for stage one of virtual queue
$N_d(t)$	Number of data packets in stage one of virtual queue at time t
$N_e(t)$	Number of energy packets in stage one of virtual queue at time t
N_{av}	Average number of data packets waiting in stage one of virtual queue
λ_d	Arrival rate of data packets
λ_e	Arrival rate of one unit of energy
τ	Mean transmission time of data packets
β	Work-to-resource ratio (λ_d/λ_e)
$\overline{T_W}$	Avg. waiting time for data packets in physical queue
$\overline{T_{W1}}$	Avg. waiting time for data packets in stage one of virtual queue
$\overline{T_{W2}}$	Avg. waiting time for data packets in stage two of virtual queue
$\overline{T_D}$	Avg. packet delay in virtual queue
P_{loss}	Probability of data packet loss in virtual queue

Table 4.1: Description of variables

P_{loss} to approximate the average data packet delay and the probability of data packet loss, respectively, in the physical queueing system. The motivation for this approximation is as follows: when the transmission time is negligible, the data packets get serviced instantly and $\tau \to 0$, $\overline{T_{W2}} \to 0$. With negligible transmission time, the data packets in the physical queueing system wait only for energy arrival, so $\overline{T_W} \to \overline{T_{W1}}$ and the probability of data loss in the physical queueing system tends to P_{loss}. Thus, the physical queueing system becomes *identical* to the first stage of the virtual queueing system when the transmission time is zero.

When the transmission time is significant, the physical and virtual queueing systems behave identically as long as B_d and B_e are infinite. When B_e is finite and B_d is infinite, the virtual queueing system offers potentially larger energy storage capacity compared to the physical queueing system. This is due to the second stage of the virtual queue having an unbounded buffer and the fact that energy in the first stage energy queue reduces by one unit the moment a data packet moves to the second stage. The additional energy packets in the virtual queueing system reduce the waiting time of the data packets and hence $\overline{T_W} > \overline{T_{W1}} + \overline{T_{W2}}$.

However, when B_d is finite and B_e is infinite, the virtual queueing system

offers potentially larger data storage capacity compared to the physical queueing system. This causes $\overline{T_W} < \overline{T_{W1}} + \overline{T_{W2}}$ because the additionally queued data packets increase the waiting time of the subsequent data packets.

We first analyze the virtual queueing system for the relatively simple case of unit-energy arrivals, where the evolution of the data and energy buffer occupancy in the first stage is modeled as a one-dimensional Markov chain. We deal with fractional and bulk energy arrivals in the subsequent sections.

4.3 Unit-Energy Arrivals

Here, since every energy arrival corresponds to one unit of energy, each data packet needs one energy packet to move to the second stage of the virtual queueing system. We now obtain exact expressions for P_{loss} and $\overline{T_{W1}}$.

4.3.1 Obtaining P_{loss} and $\overline{T_{W1}}$

Let λ_d and λ_e denote the data and energy arrival rates, respectively. Let $N_d(t)$ and $N_e(t)$ denote the number of data and energy packets in the respective queues of the first stage in the virtual system at time t. Since the paired data and energy packets instantaneously move to the second stage, the state of the stage one queueing system is captured by $M(t) = N_d(t) - N_e(t)$. Since data and energy inter-arrival times are exponentially distributed, its memoryless property implies that $M(t)$ is a continuous-time Markov chain. We consider the time instances $0, \delta, 2\delta, \ldots, k\delta, \ldots$ where δ is a small positive number. If we define $M_k := M(k\delta)$, then $\{M_k | k = 0, 1, \ldots\}$ is a discrete-time Markov chain with steady-state occupancy probabilities equal to those of the continuous chain $M(t)$.

When the packet arrival process is Poisson with rate λ, then the probability of exactly one arrival occurs in the time interval $[t, t+\delta]$ is independent of t and is

given by $\lambda\delta\ e^{-\lambda\delta} = \lambda\delta\ (1 - \lambda\delta + (1/2)(\lambda\delta)^2 - \ldots) = \lambda\delta + o(\delta)$, where $o(\delta)$ denotes a function of δ with $\lim_{\delta\to 0}(o(\delta)/\delta) = 0$. Similarly, the probability of two or more arrivals in $[t, t+\delta]$ is easily shown to be $o(\delta)$ while the probability of no packet arrival in $[t, t+\delta]$ is given by $1 - \lambda\delta + o(\delta)$.

Since both data and energy arrivals are Poisson, the transition probabilities $P_{i,j} = P\{M_{k+1} = j | M_k = i\}$ are

$$P_{-B_e,-B_e} = 1 - \lambda_d\delta + o(\delta) \ \ ; \ \ P_{B_d,B_d} = 1 - \lambda_e\delta + o(\delta)$$

$$P_{i,i} = 1 - (\lambda_d + \lambda_e)\delta + o(\delta),\ -B_e < i < B_d \ \ ; \ \ P_{i,i+1} = \lambda_d\delta + o(\delta),\ -B_e \le i < B_d,$$

$$P_{i,i-1} = \lambda_e\delta + o(\delta),\ -B_e < i \le B_d \ \ ; \ \ P_{i,j} = o(\delta),\ j \ne i, i-1, i+1.$$

The discrete-time Markov chain $\{M_k | k = 0, 1, \ldots\}$ has finite number of states with $-B_e \le M_k \le B_d$. It is irreducible because it can transition from any state to any other state in a finite number of steps. It is also aperiodic because $P_{i,i} > 0$ for $-B_e \le i \le B_d$. An irreducible and aperiodic finite Markov chain has a unique steady-state probability vector [15]. Thus, unique steady-state probabilities exist for $\{M_k\}$. Let the steady-state probabilities be denoted as

$$P_n := \lim_{k\to\infty} P\{M_k = n\} = \lim_{t\to\infty} P\{M(t) = n\}. \tag{4.2}$$

Since the frequency of transitions from state n to state $n+1$ is equal to the frequency of transitions from $n+1$ to n in the steady state, we have

$$P_n\left(\lambda_d\delta + o(\delta)\right) = P_{n+1}\left(\lambda_e\delta + o(\delta)\right),\ -B_e \le n < B_d. \tag{4.3}$$

Dividing the above equation by δ and taking the limit $\delta \to 0$,

$$P_n\lambda_d = P_{n+1}\lambda_e, \quad -B_e \le n < B_d. \tag{4.4}$$

We now define the *work-to-resource ratio* as

$$\beta = \lambda_d/\lambda_e. \tag{4.5}$$

Thus β denotes the ratio of the rate at which *work* (data packet) is generated to the rate at which *resource* (energy packet) arrives. From (4.4) and (4.5), it follows that

$$P_n = \beta^{B_e+n} P_{-B_e}, \quad -B_e \leq n \leq B_d. \tag{4.6}$$

When $\beta = 1$, the steady state probabilities for all states are equal. Because the sum of steady state probabilities is equal to 1, we have

$$P_n = 1/(B_e + B_d + 1), \quad -B_e \leq n \leq B_d, \ \beta = 1. \tag{4.7}$$

When $\beta \neq 1$, by summing the probabilities in (4.6) to 1, we get

$$P_{-B_e} = \frac{1-\beta}{1-\beta^{B_e+B_d+1}}. \tag{4.8}$$

Since the data arrival process is Poisson and independent of the energy arrival process, the probability of data packet loss is equal to the unconditional steady-state probability of the data queue being full [15]. Thus, $P_{loss} = P_{B_d}$ and we get

$$P_{loss} = \begin{cases} 1/(B_e + B_d + 1), & \text{if } \beta = 1 \\ \dfrac{\beta^{B_e+B_d} - \beta^{B_e+B_d+1}}{1-\beta^{B_e+B_d+1}}, & \text{if } \beta \neq 1. \end{cases} \tag{4.9}$$

From (4.9) we observe that P_{loss} does not depend on the individual buffer sizes but rather on the sum $B_e + B_d$.

Let N_{av} denote the average number of data packets waiting in stage 1 queueing

sub-system. Then,

$$\begin{aligned} N_{av} &= \sum_{n=1}^{B_d} nP_n = \sum_{n=1}^{B_d} n\beta^{B_e+n} P_{-B_e} \\ &= \begin{cases} \left(\frac{1}{B_e+B_d+1}\right)\left(\frac{B_d(B_d+1)}{2}\right), & \text{if } \beta = 1 \\ \beta^{B_e+1}\left(\frac{\left(1-\beta^{B_d+1}\right)-(1-\beta)(B_d+1)\beta^{B_d}}{(1-\beta)\left(1-\beta^{B_e+B_d+1}\right)}\right), & \text{if } \beta \neq 1. \end{cases} \end{aligned} \tag{4.10}$$

Using Little's Law, we have $N_{av} = (1 - P_{loss})\lambda_d \overline{T_{W1}}$ and hence

$$\overline{T_{W1}} = \frac{N_{av}}{(1 - P_{loss})\lambda_d} \tag{4.11}$$

$$= \begin{cases} \frac{B_d(B_d+1)}{2(B_e+B_d)\lambda_d}, & \beta = 1 \\ \frac{1}{\lambda_d}\left(\frac{\beta^{B_e+1}}{1-\beta^{B_e+B_d}}\right)\left[\left(\frac{1-\beta^{B_d+1}}{1-\beta}\right) - (B_d+1)\beta^{B_d}\right], & \beta \neq 1. \end{cases} \tag{4.12}$$

4.3.2 Asymptotic Analysis of P_{loss} and $\overline{T_{W1}}$

We now present an asymptotic analysis of P_{loss} and $\overline{T_{W1}}$ when the buffer sizes become large. The analysis is divided into three cases depending on the value of β. We will refer to the system as being *adequately-resourced*, *inadequately-resourced*, and *resource-balanced*, for the cases when $\beta < 1$, $\beta > 1$, and $\beta = 1$, respectively.

- When the system is *adequately-resourced* ($\beta < 1$) and $B_e \to \infty$, then from (4.9), (4.12) we note that $P_{loss} \to 0$ and $\overline{T_{W1}} \to 0$. In this case, an arriving data packet is instantly paired with an energy packet. However, when B_e is finite and the data buffer size is large, then the probability of data packet

loss and the average waiting time in stage one can be approximated as

$$P_{loss} \approx (1-\beta)\beta^{B_e+B_d}, \qquad \beta < 1,\ B_d \gg 1 \tag{4.13}$$

$$\overline{T_{W1}} \approx \frac{\beta^{B_e+1}}{\lambda_d\,(1-\beta)}, \qquad \beta < 1,\ B_d \gg 1. \tag{4.14}$$

Here P_{loss} reduces exponentially as the sum $B_e + B_d$ increases. The approximation for $\overline{T_{W1}}$ in (4.14) shows that it becomes independent of the actual value of B_d and reduces exponentially with B_e.

- When the system is *inadequately-resourced* ($\beta > 1$) and B_d is large, then from (4.9) and (4.12) we get

$$P_{loss} \approx \frac{\beta-1}{\beta} = \frac{\lambda_d - \lambda_e}{\lambda_d}, \qquad \beta > 1,\ B_d \gg 1 \tag{4.15}$$

$$\overline{T_{W1}} \approx \frac{\beta}{\lambda_d}\left(B_d - \left(\frac{1}{\beta-1}\right)\right), \qquad \beta > 1,\ B_d \gg 1. \tag{4.16}$$

Here P_{loss} tends to a constant value independent of the actual buffer sizes. From (4.16) we note that $\overline{T_{W1}}$ grows almost linearly with B_d and is independent of the energy buffer size.

Even when B_d is small and $B_e \gg 1$ in an inadequately-resourced system, from (4.9) and (4.12) we observe that both P_{loss} and $\overline{T_{W1}}$ are independent of the actual value of B_e.

- When the system is *resource-balanced* ($\beta = 1$), then from (4.12) we observe that $\overline{T_{W1}} \to \infty$ as $B_d \to \infty$. However, when B_d has a fixed finite value and $B_e \to \infty$, then $\overline{T_{W1}} \to 0$. From (4.9) it follows that P_{loss} falls inversely with the sum $B_d + B_e$ for large buffer sizes.

4.3.3 Obtaining $\overline{T_{W2}}$

The rate at which data packets arrive at the second stage is equal to $\lambda_d (1 - P_{loss})$. If we approximate the packet arrival process at the second stage by a Poisson process, we obtain an approximation for $\overline{T_{W2}}$ as follows:

$$\overline{T_{W2}} \approx \frac{\lambda_d (1 - P_{loss}) \bar{T^2}}{2 (1 - \lambda_d (1 - P_{loss}) \tau)} , \tag{4.17}$$

where the expression on the right is the waiting time in queue for a data packet in an M/G/1 queue with arrival rate $\lambda_d (1 - P_{loss})$, mean service time τ and mean-squared service time $\bar{T^2}$ [15].

The motivation for approximating the packet arrival process at the second stage as Poisson with rate $\lambda_d (1 - P_{loss})$ is as follows. After an arrival of data-energy packet-pair to the second stage, if energy packets are waiting in the first stage energy queue, then the inter-arrival time for the next packet-pair is independent and exponentially distributed with mean $1/\lambda_d$. For the case when $\lambda_e > \lambda_d$ and $B_e \to \infty$, we note that in the steady state, energy packets wait for data packet arrival and the packet-pair arrival process at the second stage tends to Poisson with rate λ_d while $P_{loss} \to 0$. However, when both B_d and B_e are finite, it follows from (4.9) that $P_{loss} > 0$ and the packet arrival rate at the second stage becomes $\lambda_d (1 - P_{loss})$.

The overall average packet delay is obtained using (4.1). Next, we analyze the more challenging case of fractional-energy arrivals where the exact closed-form expression for $\overline{T_{W1}}$ is derived by harnessing the structure in an associated two-dimensional Markov chain.

4.4 Fractional-Energy Arrivals

Here, each energy arrival corresponds to only a fraction of the energy required to transmit a data packet. In this case, we will capture the state of the virtual queueing system by a pair of variables. Then, we exploit the structure of the resulting two-dimensional Markov chain to derive the steady-state system probabilities. These in turn are used to obtain exact closed-form expressions for P_{loss} and $\overline{T_{W1}}$.

We assume that each energy arrival brings $1/K_f$ units of energy where K_f is a positive integer greater than one. We define an *energy-frame* as a contiguous block of K_f energy packets. The inter-arrival time of energy-frames is the sum of K_f independent exponential random variables and hence has an Erlang distribution. Also, one energy-frame corresponds to one unit of energy which is used for transmitting one data packet. We denote data packet and energy-frame arrival rates by λ_d and λ_e, respectively. Note that the energy packet arrival rate in this case is $K_f\lambda_e$.

Let $N_e(t)$ denote the number of energy packets in the energy queue of the first stage at time t. We define $Y(t)$ to be the quotient and $Z(t)$ to be the remainder, respectively, when $N_e(t)$ is divided by K_f. Thus $Y(t)$ is the number of energy-frames present at time t in the energy queue. We define $M(t) = N_d(t) - Y(t)$ where $N_d(t)$ is the number of data packets in the data queue of the first stage at time t. Note that $-B_e \leq M(t) \leq B_d$ and $0 \leq Z(t) \leq K_f - 1$.

The state of the stage one queueing system is captured by the ordered pair $X(t) = (M(t), Z(t))$. Since data and energy inter-arrival times are exponentially distributed, its memoryless property implies that $X(t)$ is a continuous-time Markov chain whose state-space is given by $S = \{(j_1, j_2) \mid -B_e < j_1 \leq B_d, 0 \leq j_2 \leq K_f - 1\} \cup (-B_e, 0)$. We now present the steady-state analysis of the stage one

queueing system which leads to exact closed-form expressions for P_{loss} and $\overline{T_{W1}}$.

4.4.1 Stage One Steady-State Probabilities

The Markov chain $X(t)$ is irreducible because it can transition from any state to any other state in a finite number of steps. We denote the steady-state probability distribution for $X(t)$ as $P_{(j_1,j_2)} := \lim_{t\to\infty} P\{X(t) = (j_1, j_2)\}$, and let $\mathbf{P}_j$ denote the vector $[P_{(j,k-1)} \ \cdots \ P_{(j,1)} \ P_{(j,0)}]^T$, for $-B_e < j \leq B_d$. In the steady-state, since the frequency of transition *out* of a given set of states is equal to the frequency of transition *in* to those set of states, we have the following balance equations

$$A\ \mathbf{P}_{j_1} = B\ \mathbf{P}_{j_1-1},\ -B_e + 1 < j_1 < B_d\ , \tag{4.18}$$

where A and B are $K_f \times K_f$ invertible matrices given by

$$A = \begin{bmatrix} K_f\lambda_e & 0 & 0 & \cdots & 0 \\ \lambda_d + K_f\lambda_e & -K_f\lambda_e & 0 & \ddots & 0 \\ 0 & \lambda_d + K_f\lambda_e & -K_f\lambda_e & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \ddots & \lambda_d + K_f\lambda_e & -K_f\lambda_e \end{bmatrix},$$

$$B = \begin{bmatrix} \lambda_d & \lambda_d & \cdots & \lambda_d & \lambda_d \\ \lambda_d & 0 & \cdots & 0 & 0 \\ 0 & \lambda_d & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \ddots & \lambda_d & 0 \end{bmatrix},$$

with $\det(A) = (-1)^{K_f-1}(K_f\lambda_e)_f^K$ and $\det(B) = (-1)^{K_f-1}(\lambda_d)_f^K$. The probability vector $\mathbf{P}_{B_d}$ is related to $\mathbf{P}_{B_d-1}$ as follows

$$\tilde{A}\ \mathbf{P}_{B_d} = B\ \mathbf{P}_{B_d-1}, \tag{4.19}$$

where $\tilde{A}$ is a $K_f \times K_f$ matrix given as

$$\tilde{A} = \begin{bmatrix} K_f\lambda_e & 0 & 0 & \cdots & 0 \\ K_f\lambda_e & -K_f\lambda_e & 0 & \cdots & 0 \\ 0 & K_f\lambda_e & -K_f\lambda_e & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \ddots & K_f\lambda_e & -K_f\lambda_e \end{bmatrix}. \tag{4.20}$$

Note that $\tilde{A}$ is a lower triangular matrix, like A, with $\det(\tilde{A}) = (-1)^{K_f-1}(K_f\lambda_e)_f^K$. Let $r = \lambda_d/(K_f\lambda_e)$ and $\mathbf{v}$ be defined as $\mathbf{v} = \begin{bmatrix} r & r(1+r) \cdots r(1+r)^{K_f-1} \end{bmatrix}^T$. Then we have $\mathbf{P}_{-B_e+1} = P_{(-B_e,0)}\mathbf{v}$. From (4.18) and (4.19), we get

$$\mathbf{P}_{j_1} = P_{(-B_e,0)}\left(R^{B_e+j_1-1}\ \mathbf{v}\right),\ -B_e+1 \le j_2 < B_d \tag{4.21}$$

$$\mathbf{P}_{B_d} = P_{(-B_e,0)}\left(\tilde{R}\ R^{B_e+B_d-2}\ \mathbf{v}\right), \tag{4.22}$$

where $R = A^{-1}B,\ \tilde{R} = (\tilde{A})^{-1}B$. Equations (4.21) and (4.22) show that the steady state probability of any given state can be expressed in terms of $P_{(-B_e,0)}$. Since the steady-state probability of all the states sum to 1, we have

$$\left(P_{(-B_e,0)}\right)^{-1} = 1 + \mathbf{1}^T\left(\sum_{j=0}^{B_e+B_d-2} R^j + \tilde{R}R^{B_e+B_d-2}\right)\mathbf{v}, \tag{4.23}$$

where $\mathbf{1}$ denotes a column vector of length K_f consisting of all ones, and superscript T denotes the transpose operation.

4.4.2 Obtaining P_{loss} and $\overline{T_{W1}}$

If π_j denotes the probability that there are j data packets in the first stage data queue, then $\pi_j = \mathbf{1}^T \mathbf{P}_j,\ 1 \leq j \leq B_d$. Thus we have

$$P_{loss} = \pi_{B_d} = P_{(-B_e,0)} \mathbf{1}^T \tilde{R} R^{B_e+B_d-2} \mathbf{v}. \tag{4.24}$$

Note that similar to the unit energy arrival case, P_{loss} for fractional-energy arrivals also depends only on the sum of the buffer sizes $B_d + B_e$ rather than on individual data and energy buffer sizes.

The average number of data packets waiting in the first stage data queue, denoted N_{av}, are given by

$$N_{av} = \sum_{j=1}^{B_d} j\pi_j = P_{(-B_e,0)} \Big[\sum_{j=1}^{B_d-1} j\, \mathbf{1}^T R^{B_e+j-1}\, \mathbf{v} + \ B_d\, \mathbf{1}^T \tilde{R} R^{B_e+B_d-2}\, \mathbf{v} \Big]. \tag{4.25}$$

By applying Little's Law, we obtain $\overline{T_{W1}}$ as

$$\overline{T_{W1}} = \frac{N_{av}}{\lambda_d\,(1 - P_{loss})}, \tag{4.26}$$

where P_{loss} and N_{av} are given by (4.24) and (4.25), respectively.

4.4.3 Obtaining $\overline{T_{W2}}$

$\overline{T_{W2}}$ is obtained using (4.17) while the overall average data packet delay is given by (4.1).

4.5 Bulk-Energy Arrivals

Here we analyze the case where each energy arrival corresponds to K_b units of energy, where K_b is an integer greater than 1. Thus, in this case, each energy arrival brings energy which is sufficient for transmission of K_b data packets. Note that each energy arrival in this case can be interpreted as a *bulk arrival* of K_b unit-energy packets. Since λ_e denotes the arrival rate of one unit of energy, the arrival rate of energy packets for bulk-energy arrivals is given by λ_e/K_b.

Let $N_d(t)$ denote the number of data packets in the data queue of the first stage at time t. Let $N_e(t)$ denote the number of unit-energy packets in the energy queue of the first stage at time t. As before, B_d denotes the number of data packets that can be stored in the data buffer, while B_e denotes the number of unit energy packets that be stored in the energy buffer. We assume that both B_d and B_e are multiples of K_b.

We define $Y(t)$ to be the quotient and $Z(t)$ to be the remainder, respectively, when $N_d(t) - N_e(t)$ is divided by K_b. Thus $-B_e/K_b \leq Y(t) \leq B_d/K_b$ and $0 \leq Z(t) \leq K_b - 1$. The state of the stage one queueing system is captured by the ordered pair $X(t) = (Y(t), Z(t))$. Since data and energy inter-arrival times are exponentially distributed, its memoryless property implies that $X(t)$ is a continuous-time Markov chain. The state-space of the stage one queueing system is given by $S = \{(j_1, j_2) \mid -B_e/K_b \leq j_1 \leq B_d/K_b - 1,\ 0 \leq j_2 \leq K_b - 1\} \cup (B_d, 0)$.

4.5.1 Stage One Steady-State Probabilities

The Markov chain $X(t)$ is irreducible because it can transition from any state to any other state in a finite number of steps. We denote the steady-state probability distribution for $X(t)$ as $P_{(j_1,j_2)} := \lim_{t\to\infty} P\{X(t) = (j_1, j_2)\}$, and let $\mathbf{P}_{j_1}$ denote the vector $[P_{(j_1,K_b-1)} \ \cdots \ P_{(j_1,1)} \ P_{(j_1,0)}]^T$, for $-B_e \leq j_1 \leq B_d - 1$. In the steady-

state, since the frequency of transition *out* of a given set of states is equal to the frequency of transition *in* to those set of states, we have the following balance equations

$$A\ \mathbf{P}_{j_1} = B\ \mathbf{P}_{j_1+1},\ -B_e/K_b \leq j_1 < B_d/K_b\ , \tag{4.27}$$

where A and B are $K_b \times K_b$ invertible matrices given by

$$A = \begin{bmatrix} \lambda_d & 0 & 0 & \cdots & 0 \\ \lambda_d + \lambda_e/K_b & -\lambda_d & 0 & \ddots & 0 \\ 0 & \lambda_d + \lambda_e/K_b & -\lambda_d & \ddots & \ddots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_d + \lambda_e/K_b & -\lambda_d \end{bmatrix},$$

$$B = \begin{bmatrix} \lambda_e/K_b & \lambda_e/K_b & \cdots & \lambda_e/K_b & \lambda_e/K_b \\ \lambda_e/K_b & 0 & \cdots & 0 & 0 \\ 0 & \lambda_e/K_b & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \ddots & \lambda_e/K_b & 0 \end{bmatrix},$$

with $\det(A) = (-1)^{K_b-1}(\lambda_d)_b^K$ and $\det(B) = (-1)^{K_b-1}(\lambda_e/K_b)_b^K$.

Let $r = \dfrac{\lambda_e}{K_b\lambda_d}$ and $\mathbf{v}$ be defined as $\mathbf{v} = \begin{bmatrix} r & r(1+r) & \cdots & r(1+r)^{K_b-1} \end{bmatrix}^T$. From the balance equations for the boundary states, we get

$$\mathbf{P}_{(B_d/K_b)-1} = P_{(B_d/K_b,0)}\mathbf{v}. \tag{4.28}$$

From (4.27) and (4.28), we get

$$\mathbf{P}_{j_1} = P_{(B_d/K_b,0)}\left(R^{(B_d/K_b)-j_1-1}\ \mathbf{v}\right),\ -B_e/K_b \leq j_1 \leq (B_d/K_b) - 1, \tag{4.29}$$

where $R = A^{-1}B$. Equation (4.29) shows that the steady state probability of

any given state can be expressed in terms of $P_{(-B_d/K_b,0)}$. Since the steady-state probability of all the states sum to 1, we have

$$\left(P_{(B_d/K_b,0)}\right)^{-1} = 1 + \mathbf{1}^T \left(\sum_{j=0}^{B_e/K_b+B_d/K_b-1} R^j \right) \mathbf{v}. \tag{4.30}$$

4.5.2 Obtaining P_{loss} and $\overline{T_{W1}}$

Since the data arrival process is Poisson and independent of the energy arrival process, the probability of data packet loss is equal to the unconditional steady-state probability of the data queue being full. Thus we have

$$P_{loss} = P_{B_d/K_b,0} = \left(1 + \mathbf{1}^T \left(\sum_{j=0}^{B_e/K_b+B_d/K_b-1} R^j \right) \mathbf{v} \right)^{-1}. \tag{4.31}$$

From (4.31) we note that P_{loss} again depends only on the sum of the buffer sizes $B_d + B_e$ rather than on individual data and energy buffer sizes.

Let β_j be a length K_b vector defined as

$$\beta_j = [(K_b j + K_b - 1) \quad (K_b j + K_b - 2) \quad \cdots \quad (K_b j + 1) \quad K_b j]^T. \tag{4.32}$$

Then the average number of data packets waiting in the first stage data queue is given by

$$N_{av} = \left(\sum_{j=0}^{(B_d/K_b)-1} \beta_j^T \mathbf{P}_j \right) + B_d P_{B_d/K_b,0} = P_{loss} \left(\sum_{j=0}^{(B_d/K_b)-1} \beta_j^T R^{(B_d/K_b)-j_1-1} \mathbf{v} + B_d \right). \tag{4.33}$$

By applying Little's Law, we obtain $\overline{T_{W1}}$ as

$$\overline{T_{W1}} = \frac{N_{av}}{\lambda_d (1 - P_{loss})}, \tag{4.34}$$

where P_{loss} and N_{av} are given by (4.31) and (4.33), respectively.

4.5.3 Obtaining $\overline{T_{W2}}$

$\overline{T_{W2}}$ is obtained using (4.17) while the overall average data packet delay is given by (4.1).

4.6 Multi-Source Energy Arrivals

The topic of multi-source energy harvesting has gained the attention of researchers [60],[61] with the potential to provide a more robust system solution. If the energy arrivals from different sources are (a) independent, (b) Poisson, and (c) energy arrival from each source corresponds to the same amount of energy, then the combined energy arrival process is again Poisson. In this case, the results derived in previous sections can be applied to obtain closed-form expression for average data packet delay by replacing the energy arrival rate with the sum of the rates of individual energy arrivals.

However, when energy arrivals from different sources correspond to *different* amounts of energy, then the combined energy arrival process is no longer Poisson. In this section, we will derive closed-form expression for the average packet delay with multi-source energy arrivals from two independent sources. In particular, we consider the case where each energy arrival corresponds to either one unit of energy or a fraction of one unit of energy. The case of combined unit-energy and bulk-energy arrivals can be handled in a similar fashion.

We assume that the energy harvested from the two independent sources is stored in a single energy buffer which can store up to B_e units of energy. For mathematical convenience, we assume that the data buffer capacity, B_d, and B_e are both positive even integers.

The unit-energy arrival process is assumed to be Poisson and its arrival rate is denoted as λ_{e1}. The fractional-energy arrival process is also Poisson and each fractional arrival brings $1/K_f$ units of energy, where K_f is an integer greater than one. We let the fractional-energy arrival rate be $K_f\lambda_{e2}$ where λ_{e2} is the rate of accumulation of one unit of energy with fractional-energy arrivals. The combined energy arrival rate from the two sources for one unit of energy is given by $\lambda_e = \lambda_{e1} + \lambda_{e2}$.

Let $N_{e1}(t)$ and $N_{e2}(t)$ denote the number of unit-energy and fractional-energy packets, respectively, in the energy queue of the first stage at time t. Let $Y(t)$ be the quotient and $Z(t)$ the remainder, respectively, when $N_{e2}(t)$ is divided by K_f. Define $M(t) = N_d(t) - (N_{e1}(t) + Y(t))$ where $N_d(t)$ is the number of data packets in the data queue of the first stage at time t. Note that $-B_e \leq M(t) \leq B_d$ and $0 \leq Z(t) \leq K_f - 1$.

The state of the stage one queueing system is captured by the ordered pair $X(t) = (M(t), Z(t))$. Since data and energy inter-arrival times are exponentially distributed, its memoryless property implies that $X(t)$ is a continuous-time Markov chain. The state-space of the stage one queueing system is given by

$$S = \{(j_1, j_2) \mid -B_e < j_1 \leq B_d, 0 \leq j_2 \leq K_f - 1\} \cup (-B_e, 0). \tag{4.35}$$

The Markov chain $X(t)$ is irreducible because it can transition from any state to any other state in a finite number of steps. We denote the steady-state probability distribution for $X(t)$ as $P_{(j_1,j_2)} := \lim_{t\to\infty} P\{X(t) = (j_1, j_2)\}$, and let $\mathbf{P}_{j_1}$ denote the vector $[P_{(j_1,K_f-1)} \ \cdots \ P_{(j_1,1)} \ P_{(j_1,0)}]^T$, for $-B_e + 1 \leq j_1 \leq B_d$. In the steady-state, since the frequency of transition *out* of a given set of states is equal to the frequency of transition *in* to those set of states, we have the following balance

equations

$$A\ \mathbf{P}_j = \lambda_d\ \mathbf{P}_{j-1} + B\ \mathbf{P}_{j+1},\ -B_e + 1 < j < B_d\ , \tag{4.36}$$

where A and B are $K_f \times K_f$ matrices given by

$$A = \begin{bmatrix} \lambda_{\text{sum}} & -K_f\lambda_{e2} & 0 & \cdots & 0 \\ 0 & \lambda_{\text{sum}} & -K_f\lambda_{e2} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \ddots & \ddots & \lambda_{\text{sum}} & -K_f\lambda_{e2} \\ 0 & 0 & \ddots & 0 & \lambda_{\text{sum}} \end{bmatrix},$$

$$B = \begin{bmatrix} \lambda_{e1} & 0 & 0 & \cdots & 0 \\ 0 & \lambda_{e1} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \ddots & \ddots & \lambda_{e1} & 0 \\ K_f\lambda_{e2} & 0 & \cdots & 0 & \lambda_{e1} \end{bmatrix}.$$

with $\lambda_{\text{sum}} = \lambda_d + \lambda_{e1} + K_f\lambda_{e2}$.

The balance equations for the boundary states can be expressed as

$$\lambda_d\ P_{-B_e,0} = \mathbf{v}^T\ \mathbf{P}_{-B_e+1}, \tag{4.37}$$

$$\tilde{A}\ \mathbf{P}_{-B_e+1} = B\ \mathbf{P}_{-B_e+2}, \tag{4.38}$$

$$\hat{A}\ \mathbf{P}_{B_d} = \lambda_d\ \mathbf{P}_{B_d-1}, \tag{4.39}$$

where $\mathbf{v}^T$ is a row vector of length K_f while $\tilde{A}$ and $\hat{A}$ are $K_f \times K_f$ matrices given

as follows

$$\mathbf{v}^T = [(\lambda_{e1} + K_f\lambda_{e2}) \;\; \lambda_{e1} \;\; \cdots \;\; \lambda_{e1}], \tag{4.40}$$

$$\tilde{A} = A - \begin{bmatrix} \mathbf{0}_{(K_f-1)\times K_f} \\ \mathbf{v}^T \end{bmatrix}, \tag{4.41}$$

$$\hat{A} = A - \lambda_d\, I, \tag{4.42}$$

with $\mathbf{0}_{(K_f-1)\times K_f}$ denoting a $(K_f-1)\times K_f$ matrix consisting of all zeros, and I denoting the $K_f \times K_f$ identity matrix.

The process of deriving the steady-state probability in case of multi-source-energy arrivals is simplified by pairing adjacent probability vectors and defining

$$\mathbf{Q}_j = \begin{bmatrix} \mathbf{P}_{2j-1} \\ \mathbf{P}_{2j} \end{bmatrix} . \tag{4.43}$$

Now, (4.36) can equivalently be expressed as

$$\mathbf{Q}_{j-1} = R\ \mathbf{Q}_j\ , \quad (-B_e/2) + 2 \le j \le (B_d/2) \tag{4.44}$$

where R is a $2K_f \times 2K_f$ matrix defined as

$$R = \begin{bmatrix} \left(\left(\lambda_d^{-1}A\right)^2 - \lambda_d^{-1}B\right) & -\lambda_d^{-1}AB \\ -\lambda_d^{-1}A & -\lambda_d^{-1}B \end{bmatrix}. \tag{4.45}$$

We can use (4.44) to express $\mathbf{Q}_{j-1}$ in terms of $\mathbf{Q}_{B_d/2}$ as

$$\mathbf{Q}_{j-1} = R^{\frac{B_d}{2}-(j-1)}\ \mathbf{Q}_{B_d/2}\ , \quad (-B_e/2) + 2 \le j \le (B_d/2). \tag{4.46}$$

If we define

$$H_1 = \begin{bmatrix} \lambda_d I & -\hat{A} \end{bmatrix} , \tag{4.47}$$

$$H_2 = \begin{bmatrix} \tilde{A} & -B \end{bmatrix} R^{\frac{B_d+B_e}{2}-1} , \tag{4.48}$$

$$H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} , \tag{4.49}$$

then using (4.46), the relations given by (4.38) and (4.39) can be compactly expressed as

$$H\,\mathbf{Q}_{B_d/2} = \mathbf{0}_{2K_f \times 1} \quad . \tag{4.50}$$

where $\mathbf{0}_{2K_f \times 1}$ is an all-zero vector. Further, if we define

$$\mathbf{w}^T = \begin{bmatrix} \lambda_d^{-1}\mathbf{v}^T & \mathbf{0}_{1\times K_f} \end{bmatrix} R^{\frac{B_d+B_e}{2}-1} , \tag{4.51}$$

then (4.37) can be equivalently expressed as

$$P_{-B_e,0} = \mathbf{w}^T \mathbf{Q}_{B_d/2} \ . \tag{4.52}$$

Since the steady-state probabilities sum to 1, we have

$$P_{-B_e,0} \ + \sum_{j=(-B_e/2)+1}^{B_d/2} \mathbf{1}^T \mathbf{Q}_j = 1 \ , \tag{4.53}$$

where $\mathbf{1}$ denotes a length $2K_f$ column vector consisting of all-ones. Using (4.46) and (4.52), condition (4.53) can be written as

$$\mathbf{u}^T \mathbf{Q}_{B_d/2} = 1 \ , \tag{4.54}$$

where $\mathbf{u}^T$ is a row vector given by

$$\mathbf{u}^T = \mathbf{w}^T + \mathbf{1}^T \left(I + \sum_{j=1}^{\frac{B_d+B_e}{2}-1} R^j \right) . \tag{4.55}$$

Now, $\mathbf{Q}_{B_d/2}$ is the unique vector which satisfies (4.50) and (4.54), and it is given by

$$\mathbf{Q}_{B_d/2} = {H_u}^{-1} \, \mathbf{e}_{2K_f} \ , \tag{4.56}$$

where H_u is a matrix which is obtained by replacing the last row of H matrix (4.49) by $\mathbf{u}^T$, and $\mathbf{e}_{2K_f}$ is a column vector whose last entry is 1 and the other entries are 0.

4.6.1 Obtaining P_{loss} and $\overline{T_{W1}}$

If π_j denotes the probability that there are j data packets in the first stage data queue, then $\pi_j = \mathbf{1}^T \mathbf{P}_j, \ 1 \leq j \leq B_d$. Since the data arrival process is Poisson and independent of the energy arrival process, the probability of data packet loss is equal to the unconditional steady-state probability of the data queue being full. Thus, if we define $\mathbf{g} = [\mathbf{0}_{1\times K_f} \ \mathbf{1}_{1\times K_f}]^T$, then we have

$$P_{loss} = \pi_{B_d} = \mathbf{g}^T \mathbf{Q}_{B_d/2} = \mathbf{g}^T {H_u}^{-1} \, \mathbf{e}_{2K_f}. \tag{4.57}$$

Now, if $\mathbf{h}_j = \begin{bmatrix} (2j-1)\mathbf{1}_{1\times K_f} & (2j)\mathbf{1}_{1\times K_f} \end{bmatrix}^T$, then the average number of data packets waiting in the first stage data queue, denoted N_{av}, can be obtained as

$$N_{av} = \sum_{j=1}^{B_d} j\pi_j = \sum_{j=1}^{B_d/2} \mathbf{h}_j^T \mathbf{Q}_j \overset{(a)}{=} \left(\mathbf{h}_{B_d/2}^T + \sum_{j=1}^{(B_d/2)-1} \mathbf{h}_j^T R^{\frac{B_d}{2}-j} \right) {H_u}^{-1} \, \mathbf{e}_{2K_f} \ , \tag{4.58}$$

where (a) follows from (4.46) and (4.56). Finally, by applying Little's Law we get

$$\overline{T_{W1}} = \frac{N_{av}}{\lambda_d \left(1 - P_{loss}\right)} = \frac{\left(\mathbf{h}_{B_d/2}^T + \sum_{j=1}^{B_d/2} \mathbf{h}_j^T R^{\frac{B_d}{2}-j}\right) {H_u}^{-1} \, \mathbf{e}_{2K_f}}{\lambda_d \left(1 - \mathbf{g}^T {H_u}^{-1} \, \mathbf{e}_{2K_f}\right)}. \tag{4.59}$$

4.6.2 Obtaining $\overline{T_{W2}}$

$\overline{T_{W2}}$ is obtained using (4.17) while the overall average data packet delay is given by (4.1).

4.7 Throughput under QoS constraints

The closed-form expressions derived in the previous section are useful in selecting system design parameters which maximize the throughput while meeting the required QoS constraints. For applications with delay constraints, an energy harvesting communication system designer may wish to quantify the maximum data packet arrival rate while ensuring that the average delay and the packet loss probability are less than the desired thresholds.

The QoS constraints may be specified by restricting the average packet delay and the probability of packet loss due to buffer overflow, whose target value will be denoted by D_{tar} and P_{tar}, respectively. A given data packet arrival rate λ_d is said to be *QoS-feasible* if the corresponding average packet delay and the probability of packet loss do not exceed D_{tar} and P_{tar}, respectively.

For a given set of system parameters such as $\tau, B_d, B_e, \lambda_{e1}$, and λ_{e2}, we let Λ_{QoS} denote the corresponding set of all *QoS-feasible* arrival rates,

$$\Lambda_{QoS} = \{\lambda_d \,|\text{Av. delay} \leq D_{tar}, \text{and Prob. loss} \leq P_{tar}\}. \tag{4.60}$$

Let λ_d^* denote the *maximum* data packet arrival rate for which the average packet

delay does not exceed D_{tar} and the probability of data packet loss due to buffer overflow is less than or equal to P_{tar}. Thus,

$$\lambda_d^* = \max \Lambda_{QoS} \ . \tag{4.61}$$

Note that because a decrease in the data packet arrival rate only reduces the average packet delay and the probability of data packet loss, the set Λ_{QoS} is equal to the closed interval $[0, \lambda_d^*]$ on the real line.

The average number of bits per second arriving at the energy harvesting transmitter is given by the product of the data packet length and the data packet arrival rate. Thus, λ_d^* captures the maximum throughput of the energy harvesting system under QoS constraints on delay and packet loss.

4.7.1 Choice of system design parameters: τ, B_d, and B_e

In this subsection, we highlight the impact of choice of system design parameters on λ_d^*. Note that the quantity and rate of *energy* arrival are typically governed by nature and the environment, and hence may not be directly controllable. On the other hand, parameters such as τ, B_d, and B_e may be *selected* to ensure that λ_d^* is adequate for a given energy harvesting communication application while satisfying the QoS constraints.

A reasonable value of the average service time, τ, may be obtained via an appropriate choice of the transmission bandwidth which determines the rate at which data bits are transmitted. Since the packet delay is the sum of waiting time and service time, the value of τ should be less than D_{tar} for communication to be feasible. Indeed, $\Lambda_{QoS} = \{0\}$ if $D_{tar} \leq \tau$. This follows because a non-zero value of λ_d will result in non-zero average waiting time and thus the average packet delay will become greater than τ.

In scenarios where the average packet delay is dominated by the waiting time for energy arrivals, a further reduction in τ (by an increase in the transmission bandwidth) will not have significant impact on λ_d^* since the waiting time for energy becomes a bottleneck. Thus, a judicious choice of τ is one which balances the cost of bandwidth with the overall average packet delay by ensuring that the average service time is kept modest relative to the average waiting time for arrival of sufficient energy.

The data buffer size, B_d, has a direct impact on the probability of data packet loss due to buffer overflow. An increase in B_d leads to smaller loss probability since the additional storage capacity is used to queue data packets which would otherwise have been lost. However, these additionally queued data packet (held in the data buffer due to increase in B_d) strive for energy and transmission time, and hence increase the average waiting time of future arriving data packets, which leads to an increase in the average packet delay.

In the scenario where the data buffer size is unbounded, the probability of data buffer overflow is zero, and only the delay constraint needs to be satisfied. An increase in the data arrival rate increases the average packet delay since every additional arrival tends to increase the waiting time for future data arrivals. Let the maximum data packet arrival rate (4.61) corresponding to target delay values, D_{tar} and $\tilde{D}_{tar}$, be denoted as λ_d^* and $\tilde{\lambda}_d^*$, respectively. Then, because the average packet delay increases continuously with increase in data arrival rate, it follows that

$$\tau < D_{tar} < \tilde{D}_{tar} \implies \lambda_d^* < \tilde{\lambda}_d^*. \tag{4.62}$$

For a fixed λ_e, whenever $\lambda_d \geq \lambda_e$ and $B_d \to \infty$, the data queue length will grow over time and the average packet delay tends to infinity in the steady state.

Thus with an unbounded data buffer size, we have the inequality

$$\lambda_d^* < \lambda_e, \tag{4.63}$$

for all finite values of the target delay D_{tar}.

A suitable choice of the energy buffer size, B_e, is vital for controlling the average data packet delay. The additional energy storage capacity due to an increase in B_e helps to capture energy packets which might otherwise be discarded due to the energy buffer being full. These additional energy packets aid in providing energy to data packets, leading to a reduction of the average data packet delay. This increase in B_e for reducing delay is especially useful when $\lambda_d < \lambda_e$ where the energy queue is likely to become full over time.

Note that the additional energy packets, available due to an increase in B_e, also reduce the likelihood of the data queue to grow and hence reduce the probability of data packet loss due to buffer overflow. Thus, for given target values of D_{tar} and P_{tar}, an increase in B_e helps to increase in λ_d^* resulting in higher throughput under given QoS constraints. However, an increase in B_e may add to the monetary cost and should be chosen judiciously to trade performance with cost. In the next section, we present simulation results for different scenarios, highlighting the impact of the choice of system design parameters on performance.

4.8 Simulation Results

In this section, we present Monte Carlo simulation results for the probability of data packet loss and the average packet delay in the *physical queueing system*. The simulation results are compared with the corresponding closed-form expressions derived for the *virtual queueing system* in previous sections. The robustness of

the closed-form expressions is shown even for relatively large values of the service time. In the following figures, the numerical values obtained using the closed-form analytical expressions are labeled as 'Analytical' while the Monte Carlo simulation results are labeled as 'Simulation'.

We assume that the data and energy arrivals are independent Poisson processes and the service time is deterministic (as a special case of the general distribution). The energy harvesting process through piezoelectric generators is modeled in [118]. This model is compared with the experimental results in [126] where an electric charge of roughly 3.63 nC gets harvested by each compress-release cycle of nanowires. In our simulations, we allow the electric charge harvested with each compress-release cycle to vary from 1 nC to 4 nC. The arrived energy charges a capacitor which acts as a energy storage device.

We assume that each data packet is comprised of 1000 bits and the energy used in transmitting each bit is 20 pJ. Thus each data packet needs 20 nJ for its transmission and we will refer to that as one *unit* of energy. Data transmission rates of 10 Mbps and 100 kbps will be considered in the simulations which correspond to $\tau = 0.1$ ms and $\tau = 10$ ms, respectively. The capacitor is assumed to have a voltage rating of 10 V with a capacitance that ranges from 2 nF to 24 nF. Hence the energy storage capacity ranges from 100 nJ to 1200 nJ.

Fig. 4.3 depicts the case where each energy arrival brings 4 nC of charge and the capacitor harvests 20 nJ (that is, one unit of energy) with each energy arrival. The value of λ_e, the average number of unit energy recharge cycles per second, is equal to 55. The plot shows the variation in the average packet delay when the capacitance (and hence the energy storage capacity) changes from 2 nF ($B_e = 5$) to 24 nF ($B_e = 60$). The data packets arrive at an average rate of 50 packets per second ($\lambda_d = 50$) and the size of the data buffer is 20 kbits ($B_d = 20$). This is an *adequately-resourced* system with $\beta = \lambda_d/\lambda_e = 50/55 < 1$.

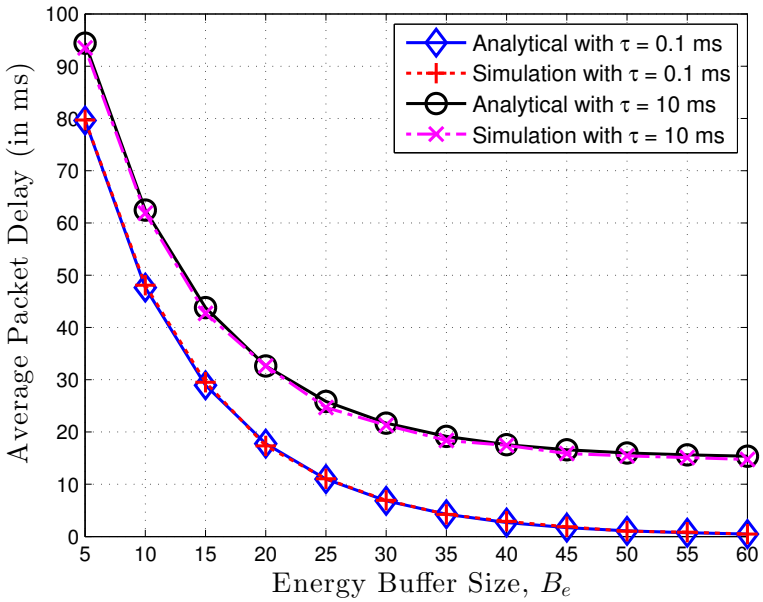

Figure 4.3: Average Packet Delay versus Energy Buffer Size with unit energy arrival, $\lambda_d = 50$, $\lambda_e = 55$, $B_d = 20$.

Fig. 4.3 shows that the service time of $\tau = 0.1$ ms is negligible compared to the average packet delay for $B_e < 35$. The wait for sufficient energy arrival dominates the overall packet delay which falls exponentially with B_e, as predicted by (4.14). When $\tau = 10$ ms and $B_e > 45$, the service time contributes to more than sixty percent of the average packet delay. Even in this case, the numerical values obtained from the closed-form expressions closely match the simulation results.

Fig. 4.4 presents the probability of data packet loss with the same system parameters as for Fig. 4.3. The expression for P_{loss} in the virtual queueing system is independent of τ. The plot shows that the probability of packet loss falls exponentially with B_e (linear in the log scale), as indicated by (4.13). The simulation results for the physical queueing system indicate that the probability of data packet loss is dominated by the event where the data buffer becomes full solely due to packets waiting for sufficient energy and it does not vary much due to a change in transmission time. Note that although the overall buffer size in the virtual queueing system is more than that in the physical queueing system (due to

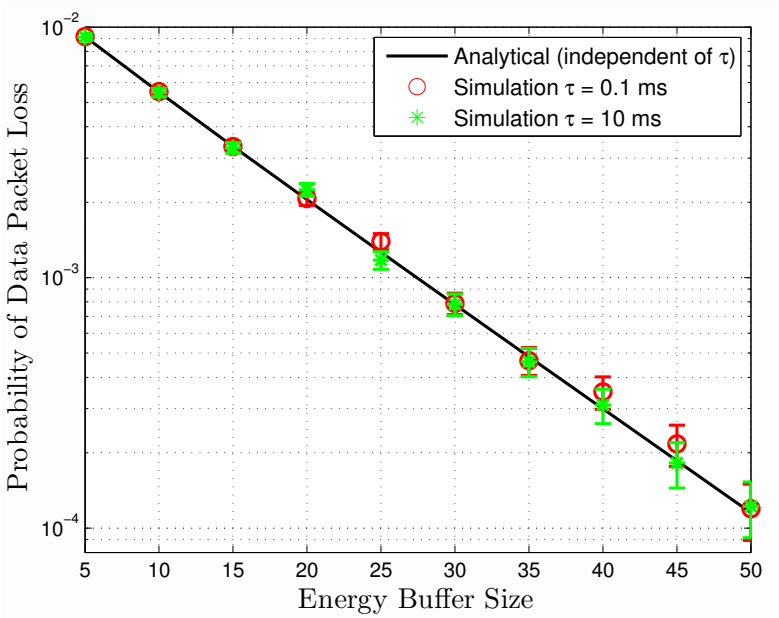

Figure 4.4: Probability of Data Packet Loss versus Energy Buffer Size with unit energy arrival, $\lambda_d = 50$, $\lambda_e = 55$, $B_d = 20$. For simulation results, the 95% confidence interval is also plotted.

the infinite buffer in the *second* stage of the virtual queue), the numerical results for the virtual queueing system match closely with the simulation results for the physical queueing system. This is because in case of virtual queues, a packet loss occurs only due to data packets waiting for sufficient energy to arrive in the *first* stage. Note that Fig. 4.4 also plots the 95% confidence interval for simulation results which highlights the relative accuracy of the results when the number of simulation samples are finite and results may not converge.

Fig. 4.5 depicts the fractional-energy arrival case where each energy arrival corresponds to 1/2 units of energy ($K_f = 2$). The number of 1/2 unit energy recharge cycles are assumed to have an average rate of 110 per second and hence the *energy-frame* arrival rate is $\lambda_e = 55$. The energy buffer size $B_e = 20$ while the data packet arrival rate is $\lambda_d = 50$. The corresponding probability of data packet loss due to data buffer overflow is plotted in Fig. 4.6 which shows that $P_{loss} < 10^{-4}$ for $B_d \geq 40$. The insight from this observation is that when B_d is

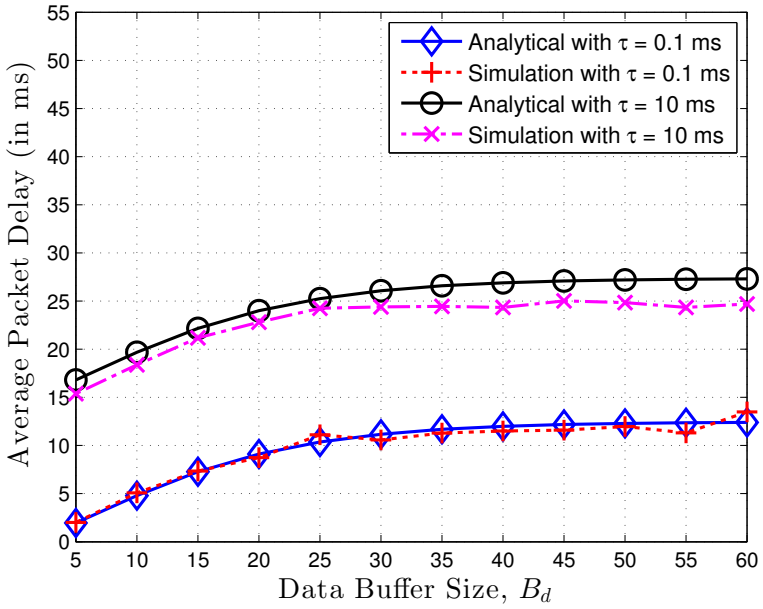

Figure 4.5: Average Packet Delay versus Data Buffer Size with fractional-energy arrival ($K_f = 2$), $\lambda_d = 50$, $\lambda_e = 55$, $B_e = 20$.

large, the packet arrival rate at the second stage of the virtual queueing system is approximately equal to λ_d (since P_{loss} is small) and hence the actual value of B_d does not impact $\overline{T_{W2}}$. It is seen from Fig. 4.5 that the average packet delay saturates to a constant value as B_d exceeds 40. This is not surprising since (4.14) indicates that when $\lambda_d < \lambda_e$ and B_d is large, then $\overline{T_{W1}}$ does not depend on the exact value of B_d. Hence the average packet delay saturates to a constant value for large B_d (since both $\overline{T_{W1}}$ and $\overline{T_{W2}}$ become independent of B_d).

Fig. 4.7 shows the fractional-energy arrival case where each energy arrival corresponds to 1/4 units of energy. The number of 1/4 unit energy recharge cycles are assumed to have an average rate of 220 per second and hence the *energy-frame* arrival rate is $\lambda_e = 55$. The data storage capacity is $B_d = 20$ while $\lambda_d = 50$. For these parameters, the inter-arrival time for arrival of one unit of energy is Erlang distributed with mean $1/\lambda_e$ and variance $1/(K_f\lambda_e^2)$ where $K_f = 4$. In comparison, the parameters chosen for Fig. 4.3 are similar except that the inter-arrival time

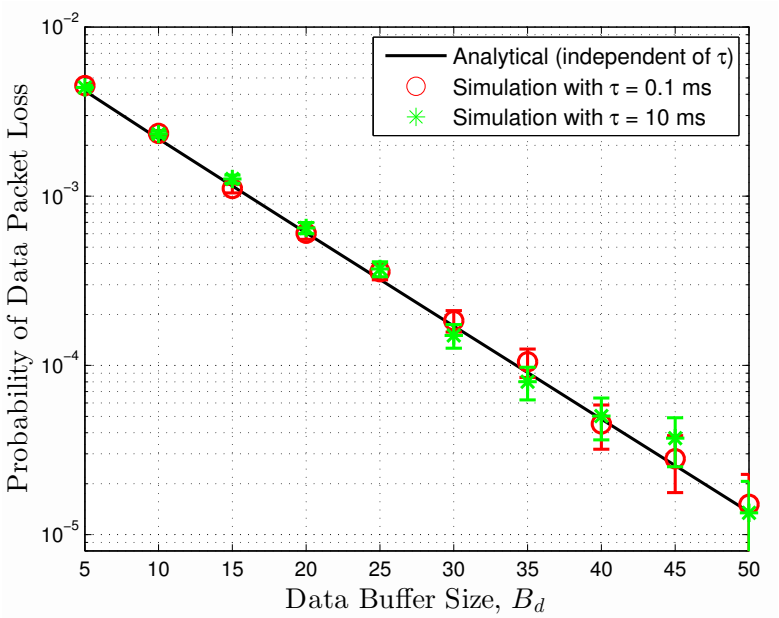

Figure 4.6: Probability of Data Packet Loss versus Data Buffer Size with fractional-energy arrival ($K_f = 2$), $\lambda_d = 50$, $\lambda_e = 55$, $B_e = 20$. For simulation results, the 95% confidence interval is also plotted.

for arrival of one unit of energy is exponentially distributed with mean $1/\lambda_e$ and variance $1/\lambda_e^2$. Thus, the lower variance of energy arrivals causes the average packet delay to be lower in Fig. 4.7 as compared to Fig. 4.3. Note that the overall shape of the curves in Fig. 4.7 is roughly similar to those in Fig. 4.3.

Fig. 4.8 plots the probability of data packet loss with the same parameters as in Fig. 4.7. On comparing Fig. 4.8 with Fig. 4.4, we observe that the probability of data packet loss is lower in Fig. 4.8 due to a reduction in the variance of energy arrivals.

Fig. 4.9 depicts the bulk-energy arrival case where each energy arrival corresponds to 2 units of energy ($K_b = 2$), $\lambda_d = 50$, and $B_d = 20$. As discussed in Section 4.5, each energy arrival in this case can be interpreted as arrival of two unit-energy packets. The average number of energy arrivals (of 2 units each) is assumed to be 55/2 per second and hence the arrival rate of one unit of energy is $\lambda_e = 55$. Since the energy arrival (of 2 units each) is Poisson, it follows

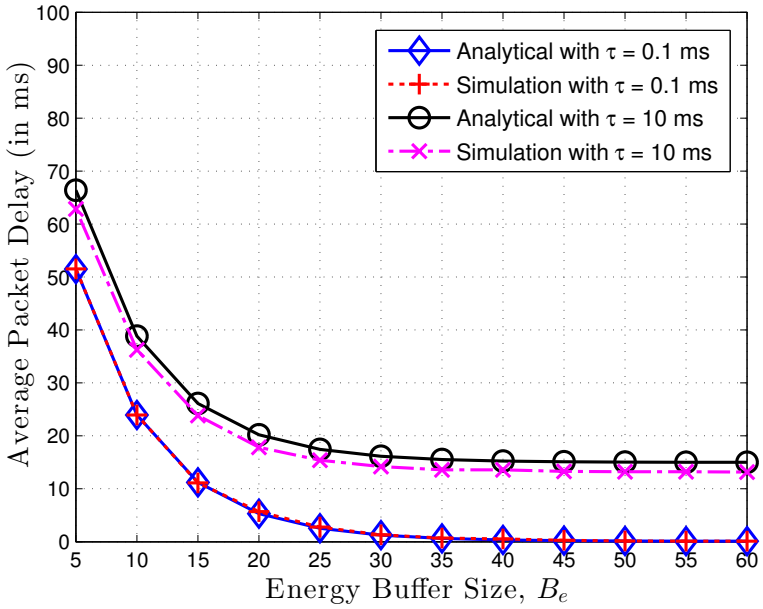

Figure 4.7: Average Packet Delay versus Energy Buffer Size with fractional-energy arrival ($K_f = 4$), $\lambda_d = 50$, $\lambda_e = 55$, $B_d = 20$.

that the inter-arrival time for arrival of *one unit* of energy has variance equal to $(2K_b - 1)/\lambda_e^2$ where $K_b = 2$. In comparison, the parameters chosen for Fig. 4.3 are similar except that the inter-arrival time for arrival of one unit of energy in that case has variance $1/\lambda_e^2$. Thus, the higher variance in the arrival of unit-energy causes the average packet delay to be larger in Fig. 4.9 as compared to Fig. 4.3.

The average packet delay for the *multi-source energy harvesting* comprising of a combination of unit-energy arrivals from one source (with $\lambda_{e1} = 40$) and fractional-energy arrivals from the other source (with $K_f = 4$ and $\lambda_{e2} = 15$) is plotted as a function of B_e in Fig. 4.10 where the data packet arrival rate is assumed to be $\lambda_d = 50$. Note that although the rate of arrival of one unit of energy from each of the two sources is less than λ_d, the combined rate of arrival of one unit of energy, $\lambda_e = \lambda_{e1} + \lambda_{e2} = 55$, exceeds λ_d. We assume $B_d = 20$, and thus the data buffer size as well as the arrival rate of data and energy packets chosen for Fig. 4.10 are same as those chosen for Figs. 4.3 (only unit-energy arrivals)

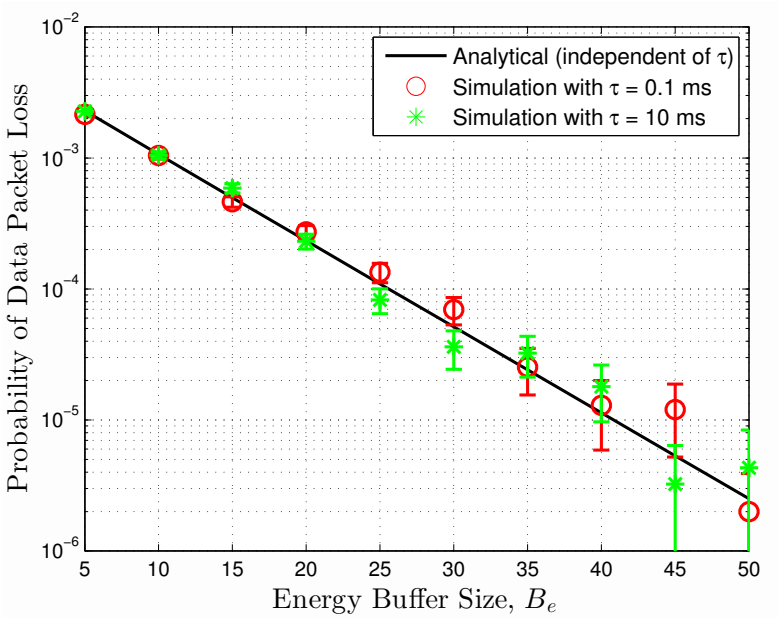

Figure 4.8: Probability of Data Packet Loss versus Energy Buffer Size with fractional-energy arrival ($K_f = 4$), $\lambda_d = 50$, $\lambda_e = 55$, $B_d = 20$. For simulation results, the 95% confidence interval is also plotted.

and 4.7 (only fractional-energy arrivals). Since the energy arrivals in this case are a combination of unit-energy and fractional-energy, we observe that the average packet delay curves in Fig. 4.10 lie in-between the corresponding curves in Figs. 4.3 and 4.7.

Fig. 4.11 plots the maximum possible data packet arrival rate when the average packet delay is constrained not to exceed D_{tar}. Here the data buffer size is unbounded, $\tau = 10$ ms, $\lambda_e = 55$, and it follows from (4.63) that $\lambda_d^* < 55$ for all finite D_{tar}. Fig. 4.11 shows that λ_d^* increases monotonically with D_{tar} as indicated by (4.62). This figure also highlights the impact of increasing the energy buffer size, B_e, on throughput. We observe that when B_e is increased from 5 to 20, the system can support higher throughput (due to increase in λ_d^*) for a given delay constraint. For the case when $B_e = 20$, we find that λ_d^* increases rapidly as D_{tar} is increased from 12 ms to 30 ms. However, the increase in λ_d^* is gradual when D_{tar} is increased beyond 30 ms. Hence in this case, a target average packet delay

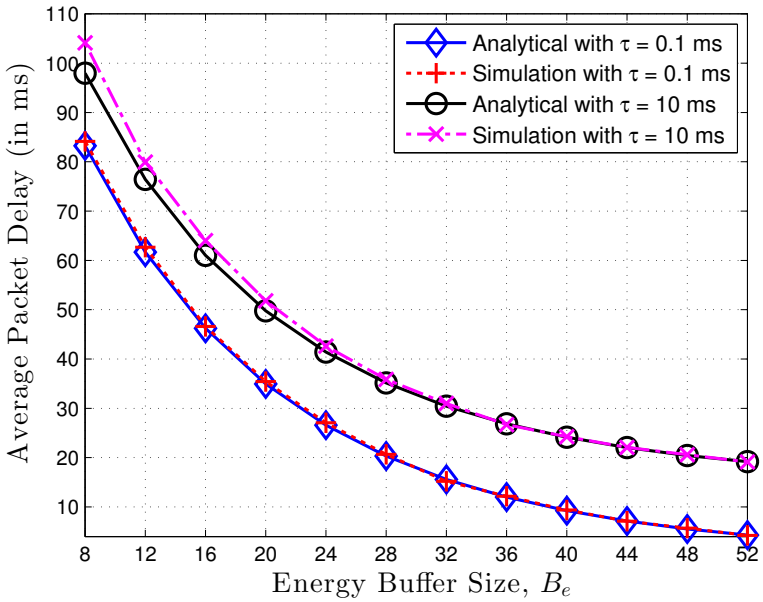

Figure 4.9: Average Packet Delay versus Energy Buffer Size with bulk-energy arrival ($K_b = 2$), $\lambda_d = 50$, $\lambda_e = 55$, $B_d = 20$.

of 30 ms represents a "sweet spot" where a sufficiently high throughput can be achieved with a nominal delay.

In practice, the closed form expressions for the average packet delay obtained in this chapter can help in making a judicious choice of system design parameters by trading resources (such as the energy buffer size and the transmission bandwidth) with performance metrics (such as delay and throughput).

4.9 Discussion

In this chapter, we derived closed-form expressions for the average packet delay and the probability of data packet loss due to buffer overflow for an energy harvesting communication system. After obtaining these expressions for single-source energy harvesting, we extended the results to the important case of multi-source energy harvesting. The derived expressions were shown to be exact when the service time is zero, and robust even when the service time is increased up to sixty percent of

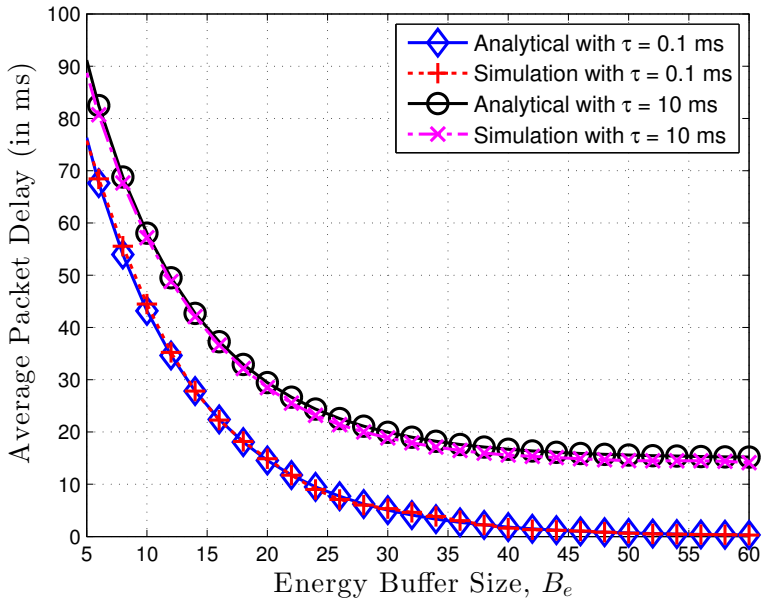

Figure 4.10: Average Packet Delay versus Energy Buffer Size with multi-source energy arrivals, $\lambda_{e1} = 40$, $\lambda_{e2} = 15$, $K_f = 4$, $\lambda_d = 50$, $B_d = 20$.

the average packet delay.

These results should be useful to both system designers and researchers in this field. For instance, an energy harvesting system designer may appropriately choose the system parameters (for example, the energy buffer size, the data packet arrival rate, and the packet transmission rate) based on the derived expressions to meet the desired QoS constraints.

The analytical approach presented in this chapter is also applicable to other related models which are discussed below:

1. The study of coupled data and energy queues is similar to the study of *assembly-like queues* in inventory systems [127]–[130]. The service of *objects* in these systems does not begin until at-least one object from each queue is present; just as a data packet is not transmitted until both data and energy are present. Thus, our work is also useful in analyzing assembly-like queues comprising of two independent queues with bounded capacity. The results

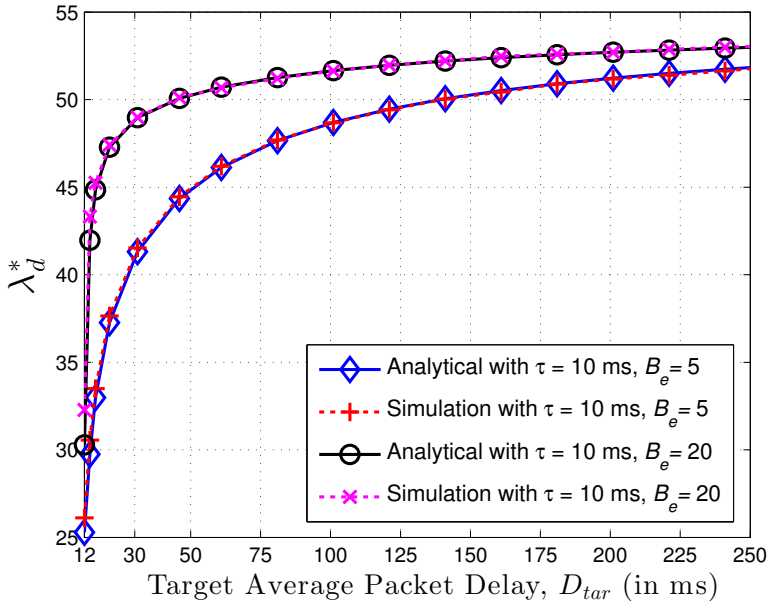

Figure 4.11: Maximum Data Packet Arrival Rate versus Target Average Delay with unit energy arrivals and $\lambda_e = 55$.

on fractional-energy arrival (Section 4.4) could be applied, for instance, on analyzing the delay in assembly of toy cars where the main body of the car arrives in one queue and the wheels arrive in another queue, and each toy gets serviced by fixing 4 wheels to a car body.

2. In multi-source energy harvesting, we assumed that energy is harvested from two independent sources while the energy is consumed for the single task of transmission of data packets. A 'dual' problem scenario is where energy is harvested from a single source but is consumed for the twin tasks of data reception and transmission of the sensed data in a wireless sensor network. This 'dual' problem was considered in [118] where it is mentioned that finding a closed-form expression for the steady-state probabilities of the Markov model is not feasible. However, our analytical approach for multi-source energy harvesting (Section 4.6) can be adapted to obtain exact closed-form expression of the steady-state probabilities for this dual problem.

Chapter 5

Constant Subblock-Composition Codes

The study of simultaneous information and energy transfer is relevant for communication from a powered transmitter to a receiver which uses the same received signal both for decoding information and for extracting energy to power its circuitry. This has applications ranging from wireline [88], [131] to wireless [11], [71] communications. The fundamental tradeoff between reliable communication and delivery of energy at the receiver, in an information-theoretic setting, was first characterized in [14] using a general capacity-power function, where transmitted codewords were constrained to have average received energy exceed a threshold.

For practical application of simultaneous energy and information transfer from a powered transmitter to an energy harvesting receiver, imposing only an average received power constraint may not be sufficient; we may also need to regularize the transferred energy content. This is because a codeword satisfying the average power constraint may still cause power outage at the receiver if the energy content in the codeword is bursty, since a receiver battery with small capacity may drain during periods of low signal energy.

In this chapter, we consider a discrete memoryless channel (DMC) and characterize achievable information rates when codewords are divided into smaller subblocks and each *subblock* is constrained to carry sufficient energy. We assume that corresponding to transmission of each symbol in the input alphabet, the receiver harvests a certain amount of energy as a function of the transmitted symbol. Since different symbols may correspond to different energy levels, the requirement of sufficient energy content within a subblock imposes a constraint on the composition of each subblock. Towards meeting this subblock energy requirement, we introduce the *constant subblock-composition codes* (CSCCs) where all the subblocks in every codeword have the same fixed composition. This subblock-composition is chosen to maximize the rate of information transfer while meeting the energy requirement. Note that if x_1^L denotes a given subblock of length L, then its composition is the distribution $P_{x_1^L}$ on $\mathcal{X}$ defined by $P_{x_1^L}(x) \triangleq \frac{N(x)}{L},\ x \in \mathcal{X}$, where $N(x)$ is the number of occurrences of symbol x in subblock x_1^L.

For meeting the real-time energy requirement at a receiver which uses the received signal to simultaneously harvest energy and decode information, we propose the use of CSCCs (Sec. 5.2.1) and establish their capacity as a function of the required energy per symbol (Sec. 5.2.2). We show that CSCC capacity can be computed efficiently by exploiting symmetry properties (Sec. 5.2.3) and present bounds on subblock length for avoiding receiver energy outage (Sec. 5.2.4).

Compared to constant composition codes, we quantify the rate loss incurred due to the additional constraint of restricting all subblocks within codewords to have the same composition (Sec. 5.3.1). For a given rate of information transfer, we derive a lower bound for the error exponent using CSCC in terms of the error exponent for constant composition codes (Sec. 5.3.2).

For enabling real-time information transfer, we consider local subblock decoding where each subblock is decoded independently (Sec. 5.4), and compare achiev-

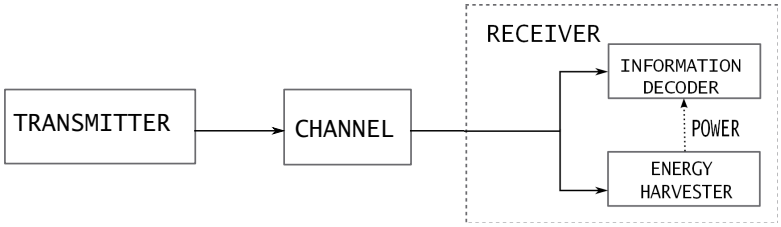

Figure 5.1: Simultaneous information and energy transfer from a transmitter to an energy-harvesting receiver

able rates using local subblock decoding with those when all the subblocks within a codeword are jointly decoded. We also provide numerical results highlighting the tradeoff between delivery of sufficient energy to the receiver and achieving high information rates (Sec. 5.5).

5.1 System Model

Consider communication from a transmitter to a receiver where the receiver uses the received signal both for decoding information as well as for harvesting energy (see Fig. 5.1). We model the effective communication channel from the output of a digital modulator at the transmitter to the input to an information decoder at the receiver as a DMC. Note that a DMC is characterized by input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and a stochastic matrix $W : \mathcal{X} \to \mathcal{Y}$ with $W = \{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ where the matrix entry $W(y|x)$ is the probability that the output is y when the channel input is x.

A DMC is a reasonable communication channel model for simultaneous energy and information transfer. Consider, for instance, the use of a digital modulator at the transmitter which produces symbols from a signal constellation $\mathcal{X} = \{x_1, \ldots, x_r\}$. At the receiver, the signal is split for use by the energy harvesting module and the information processing module, respectively. The input to the information decoder at the receiver comprises of one of s quantized values

$\mathcal{Y} = \{y_1, \ldots, y_s\}$, fed by a quantizer in the information processing path. For each quantized value $y_i, 1 \leq i \leq s$, and each transmitted symbol $x_j, 1 \leq j \leq r$, the likelihood $\Pr(y_i|x_j)$ can be computed based on the effective signal path from the transmit modulator to the quantizer at the receiver. The communication channel is thus a DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and channel transition probabilities $\Pr(y_i|x_j)$.

In practice, the effective channels seen by the information decoder and the energy harvester may be different due to their respective pre-processing stages. In [132], practical architectures for simultaneous information and energy reception were defined: an *integrated* receiver architecture has shared radio frequency chains between the energy harvester and the information decoder, whereas a *separated* architecture has different chains.

In our work, we assume a generic receiver architecture where the received signal is split between the energy harvesting path and the information processing path with a static power splitting ratio. The effective communication channel seen by the decoder in the information processing path is modeled as a DMC. We let $b(x)$ denote the energy harvested by the harvester after the signal split at the receiver, when $x \in \mathcal{X}$ is transmitted. Thus, b is a map from the input alphabet $\mathcal{X}$ to the set of non-negative real numbers, and higher energy is carried by symbols having higher b-value. This map is assumed to be time-invariant, and reflects the scenario where the statistical nature of the effective communication channel is due to the noise in the receiver circuitry, which does not affect the harvested energy. The quantification of b abstracts the implementation of a chosen receiver architecture, which in turn helps to abstract the problem of the code design for simultaneous energy and information transfer from implementation details.

In order to meet the real-time energy requirement at the receiver, we partition the transmitted codeword into equal-sized subblocks (see Fig. 5.2) and require

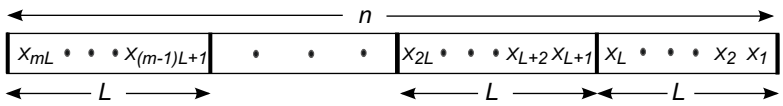

Figure 5.2: Transmitted codeword partitioned into subblocks of length L.

that transmitted symbols be chosen such that the expected harvested energy in each subblock exceeds a given threshold. This threshold is a function of the energy consumption by the receiver circuitry including the information decoder. We will denote the subblock length by L and assume that the codeword length, denoted n, is a multiple of L. If a transmitted codeword is denoted $(X_1, X_2, \ldots, X_n)$, then the constraint on sufficient energy within each subblock can be expressed as

$$\frac{1}{L}\sum_{i=1}^{L} b\left(X_{(j-1)L+i}\right) \geq B, \quad j = 1, 2, \ldots, m \tag{5.1}$$

where j is the subblock index, B denotes the required energy per symbol at the receiver, and m is the number of subblocks in a codeword. The choice of the subblock length L depends on the energy storage capacity at the receiver; a small energy buffer generally requires relatively small value of L to prevent energy outage at the receiver.

The subblock energy constraint given by (5.1) becomes trivial if $b(x)$ is same for all $x \in \mathcal{X}$ (for instance, when the transmitted symbols belong to a phase-shift-keying constellation). However, the constraint is non-trivial when b-values are not constant (for instance, using on-off keying) and threshold B satisfies $b_{\min} < B < b_{\max}$, where $b_{\min} = \min_{x\in\mathcal{X}} b(x)$ and $b_{\max} = \max_{x\in\mathcal{X}} b(x)$.

For a given subblock j within a codeword, if $N(x)$ denotes the number of occurrences of x in the jth subblock, then (5.1) can also be expressed as

$$\sum_{x\in\mathcal{X}} b(x)\frac{N(x)}{L} \geq B. \tag{5.2}$$

Note that $N(x)/L$ denotes the fraction of time when symbol x appears in the subblock. We now introduce constant subblock-composition codes which are a nice way to meet the subblock energy constraint.

5.2 Constant Subblock-Composition Codes

5.2.1 Motivation and Definition

We have seen that for a given subblock, the energy constraint given by (5.1) can equivalently be expressed as (5.2) and this constraint is satisfied provided the fraction of time each symbol appears in the subblock is chosen appropriately. This observation motivates the use of codes where the composition of each subblock in all codewords is constant and is chosen such that (5.2) is satisfied. A *constant subblock-composition code* (CSCC) is one in which all codewords are partitioned into equal-sized subblocks and each subblock (in all codewords) has the same composition P. The subblock composition P in CSCC is chosen to satisfy the subblock energy constraint

$$\mathbb{E}_P\left[b(X)\right] \triangleq \sum_{x\in\mathcal{X}} b(x)P(x) = \sum_{x\in\mathcal{X}} b(x)\frac{N(x)}{L} \geq B. \tag{5.3}$$

5.2.2 Capacity using CSCC

Let $\mathcal{P}_L$ denote the set of all compositions for input sequences of length L. For a given type $P \in \mathcal{P}_L$, the set of sequences in $\mathcal{X}^L$ with composition P is denoted by $\mathcal{T}_P^L$ and is called the *type class* or *composition class* of P. In a CSCC with subblock-composition P, every subblock in a codeword may be viewed as an element of $\mathcal{T}_P^L$.

In order to compute the capacity of a CSCC on a DMC, we may view the L uses of the original channel as a single use of the induced *vector channel* having input

alphabet $\mathcal{T}_P^L$ and output alphabet $\mathcal{Y}^L$. Since the underlying channel is memoryless, the transition probabilities for a pair of input and output vectors is the product of the corresponding transition probabilities of the underlying channel. If we let $x_1^L = x_1 \dots x_L$ and $y_1^L = y_1 \dots y_L$ be given input and output vectors with $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$, respectively, then the transition probabilities for the induced vector channel are:

$$W^L(y_1^L|x_1^L) = \prod_{i=1}^{L} W(y_i|x_i). \tag{5.4}$$

Since each subblock in a codeword may be chosen independently, the capacity using CSCC with subblock-composition P, denoted $C_{CSCC}^L(P)$, is equal to $1/L$ times the capacity of the induced vector channel with input alphabet $\mathcal{T}_P^L$, output alphabet $\mathcal{Y}^L$, and transition probabilities given by (5.4). Thus if we denote $X_1^L = X_1 \dots X_L$ and $Y_1^L = Y_1 \dots Y_L$, then

$$C_{CSCC}^L(P) = \max_{X_1^L \in \mathcal{T}_P^L} \frac{I(X_1^L; Y_1^L)}{L} \tag{5.5}$$

$$= \max_{X_1^L \in \mathcal{T}_P^L} \left(\frac{H(Y_1^L)}{L} - \frac{H(Y_1^L|X_1^L)}{L} \right) \tag{5.6}$$

$$= \max_{X_1^L \in \mathcal{T}_P^L} \left(\frac{H(Y_1^L)}{L} - \frac{\sum_{i=1}^{L} H(Y_i|X_i)}{L} \right) \tag{5.7}$$

where the last equality follows from the memoryless property of the channel. The maximization in (5.5) is over the distribution of input vector in $\mathcal{T}_P^L$. We will show that the maximum is achieved when the input vectors X_1^L are uniformly distributed over $\mathcal{T}_P^L$.

Theorem 2. *The capacity of the induced vector-channel using CSCC with fixed subblock-composition P is obtained via a uniform distribution over $\mathcal{T}_P^L$.*

Proof: Please refer to Appendix B.1. ■

If we define the set of distributions

$$\Gamma_B^L \triangleq \{P \in \mathcal{P}_L : \mathbb{E}_P[b(X)] \geq B\}, \tag{5.8}$$

then the capacity using CSCC with subblock energy constraint (5.1), denoted $C_{CSCC}^L(B)$, is defined as

$$C_{CSCC}^L(B) = \max_{P \in \Gamma_B^L} C_{CSCC}^L(P) \tag{5.9}$$

5.2.3 Computing CSCC Capacity

By Theorem 2, the maximum is achieved in (5.7) when X_1^L is uniformly distributed over $\mathcal{T}_P^L$. The computation of the capacity expression with increasing subblock length L seems challenging since the input and output alphabet size for the induced vector channel grows exponentially with L. However, we will show that the computational complexity of the CSCC capacity expression can be reduced using the following observations.

First note that the probability distribution for the output vector in the induced vector channel is given by

$$P_{Y_1^L}(y_1^L) = \frac{1}{|\mathcal{T}_P^L|} \sum_{x_1^L \in \mathcal{T}_P^L} W^L(y_1^L|x_1^L), \tag{5.10}$$

since the input vectors are uniformly distributed over $\mathcal{T}_P^L$. If $\tilde{y}_1^L$ is another output vector having the same composition as y_1^L, then we have $P_{Y_1^L}(y_1^L) = P_{Y_1^L}(\tilde{y}_1^L)$. This is because the columns $W^L(y_1^L|\cdot)$ and $W^L(\tilde{y}_1^L|\cdot)$ of the vector channel transition matrix are permutations of each other (see Appendix B.1). Thus output vectors having the same composition have equal probability. However, even though the input vectors are uniformly distributed, the output vectors in general are *not*

uniformly distributed. Also, since the symbols within an input vector $x_1^L \in \mathcal{T}_P^L$ are not independent, in general we have $P_{Y_1^L}(y_1^L) \neq \prod_{i=1}^{L} P_Y(y_i)$, where $P_Y(y)$ denotes the probability of output scalar symbol y.

Let $\mathcal{Q}_L$ denote the set of all compositions for output sequences of length L. When X_1^L is uniformly distributed over $\mathcal{T}_P^L$, the $H(Y_1^L)$ term in (5.7) can be expressed as

$$H(Y_1^L) = -\sum_{y_1^L \in \mathcal{Y}^L} P_{Y_1^L}(y_1^L) \log P_{Y_1^L}(y_1^L) \tag{5.11}$$

$$= -\sum_{Q \in \mathcal{Q}_L} \sum_{y_1^L \in \mathcal{T}_Q^L} P_{Y_1^L}(y_1^L) \log P_{Y_1^L}(y_1^L) \tag{5.12}$$

$$= \sum_{Q \in \mathcal{Q}_L} |\mathcal{T}_Q^L| P_{Y_1^L}(y_1^L) \log \frac{1}{P_{Y_1^L}(y_1^L)}, \tag{5.13}$$

where the last equality follows because $P_{Y_1^L}(y_1^L)$ is same for all $y_1^L \in \mathcal{T}_Q^L$. Note that we choose only one representative vector y_1^L from each type class $\mathcal{T}_Q^L$ in the last equality.

Secondly, the following proposition shows that the $H(Y_i|X_i)$ term in (5.7) is same for all $1 \leq i \leq L$, since the corresponding joint probabilities $P_{XY}(X_i = x, Y_i = y)$ are equal.

Proposition 6. *For a random input vector X_1^L uniformly distributed over $\mathcal{T}_P^L$ with corresponding output vector Y_1^L, the pairwise probability $P_{XY}(X_i = x, Y_i = y)$, for $1 \leq i \leq L$, satisfies*

$$P_{XY}(X_i = x, Y_i = y) = \frac{N(x)}{L} W(y|x) = P(x)W(y|x). \tag{5.14}$$

Proof: Please refer to Appendix B.2. ■

Proposition 7. *The CSCC capacity, $C^L_{CSCC}(B)$, is given by*

$$\max_{P\in\Gamma^L_B} \frac{1}{L} \sum_{Q\in\mathcal{Q}_L} |\mathcal{T}^L_Q|\ P_{Y_1^L}(y_1^L) \log \frac{1}{P_{Y_1^L}(y_1^L)} - H(Y|X), \tag{5.15}$$

where only one representative output vector y_1^L is chosen from every type class $\mathcal{T}^L_Q$, $P_{Y_1^L}(y_1^L)$ is given by (5.10)*, and $H(Y|X)$ is evaluated using the joint pairwise probability distribution given by* (5.14)*.*

Proof: Use (5.7) and (5.9) to express $C^L_{CSCC}(B)$. From Thm. 2, a uniform distribution over $\mathcal{T}^L_P$ achieves capacity, and hence the entropy term $H(Y_1^L)$ in (5.7) can be computed using (5.13). The claim in Prop. 7 follows by further noting that the $H(Y_i|X_i)$ term in (5.7) is the same for all $1 \le i \le L$, which can be evaluated using the joint pairwise distribution in (5.14). ■

5.2.4 Choice of Subblock Length L

In this subsection, we derive bounds on subblock length L (as a function of the energy storage capacity at the receiver) which will ensure that the receiver never runs out of energy when the subblock-composition P is chosen to satisfy (5.3). It will be seen that a large energy storage capacity allows for larger values of L and hence results in higher rates of information transfer.

The energy storage capacity at the receiver is denoted E_{max} and we assume that the receiver requires B units of energy per symbol for its processing. Let $E(i)$ denote the level of the energy buffer at the receiver at the completion of $i-1$ uses of the channel. The energy update equation, for $i = 1, 2, \ldots,$ is

$$E(i+1) = \min\left(E_{max}, |E(i) + b(X_i) - B|^+\right), \tag{5.16}$$

where X_i is the symbol transmitted in the ith channel use, and $|z|^+ \triangleq \max(z, 0)$.

We say that an *outage* occurs during ith channel use if $E(i) + b(X_i) < B$, while an *overflow* event occurs if $E(i) + b(X_i) - B > E_{max}$. We partition the input alphabet as $\mathcal{X} = \mathcal{X}_\triangleleft \cup \mathcal{X}_\triangleright$, where

$$\mathcal{X}_\triangleleft = \{x \in \mathcal{X} \,|\, b(x) < B\} \quad , \quad \mathcal{X}_\triangleright = \{x \in \mathcal{X} \,|\, b(x) \geq B\}. \tag{5.17}$$

For CSCC with subblock-composition $P \in \Gamma_B^L$, we define

$$G = \sum_{x \in \mathcal{X}_\triangleleft} LP(x) \left(B - b(x)\right) \ , \tag{5.18}$$

where G will be used to characterize useful properties of the energy update process.

Lemma 1. *The energy update process satisfies the following properties for CSCC with subblock-composition $P \in \Gamma_B^L$:*

(a) If there is no energy outage or overflow during the reception of the first subblock, then $E(L+1) \geq E(1)$.

(b) If $E(1) \geq G$, then there is no energy outage during the reception of the first subblock.

(c) If $E(1) \geq G$ and $E_{max} \geq 2G$, then $E(L+1) \geq G$.

Proof: If there is no energy outage or overflow, then the total energy harvested during the reception of the first subblock is $\sum_{x \in \mathcal{X}} LP(x)b(x)$, while the total energy consumed is LB and claim (a) follows since P satisfies (5.3).

Let X_i denote the transmitted symbol in the ith channel use, $I = \{1, 2, \ldots, L\}$, and $I_< = \{i \in I | X_i \in \mathcal{X}_\triangleleft\}$. For $i \in I$, the level in the energy buffer decreases during the ith channel use if and only if $i \in I_<$, and the corresponding decrease in energy level is $B - b(X_i)$. Since the subblock has composition P, the sum of

energy decrements over the reception of the first subblock is $\sum_{i\in I_<} B - b(X_i) = G$, and claim (b) follows.

For proving claim (c), we note that the condition $E(1) \geq G$ implies that there is no energy outage during the reception of the first subblock (using claim (b)). Further, if there is no overflow then $E(L+1) \geq E(1) \geq G$ (using claim (a)). In case there is energy overflow in the ith channel use for any $i \in I$, we have $E(i+1) = E_{max} \geq 2G$, and thus $E(L+1) \geq E(i+1) - G \geq G$. ■

Lemma 1 is useful in proving the following theorem which gives a necessary and sufficient condition on subblock length in order to avoid outage.

Theorem 3. *A necessary and sufficient condition on L for avoiding energy outage during the reception of CSCC codewords, with subblock-composition P satisfying* (5.3)*, is*

$$L \leq \frac{E_{max}}{\sum_{x\in\mathcal{X}_\triangleleft} 2P(x)\,(B - b(x))}, \tag{5.19}$$

with $E(1) \geq G$.

Proof: Please refer to Appendix B.3. ■

The initial condition on energy level, $E(1) \geq G$, may be ensured by transmitting a preamble, consisting of symbols with high energy content, before the transmission of codewords. This preamble has bounded length and hence does not affect the channel capacity.

5.3 Comparing CSCC with Constant Composition Codes

5.3.1 Rate Comparison

Similar to subblock-composition, a codeword composition represents the fraction of times each input symbol occurs in a codeword and a constant composition code (CCC) is one in which all codewords have the same composition. Note that a CSCC with subblock-composition P may also be viewed as a CCC with codeword composition P, since all the subblocks in CSCC have the same composition. In general for CCC, although all codewords have the same composition, different subblocks within a codeword may have different compositions. Hence CCCs are richer than CSCCs in terms of choice of symbols within each subblock. CCCs were first analyzed by Fano [133] and shown to be sufficient to achieve capacity for any discrete memoryless channel.

Let $C_{CCC}(P)$ denote the maximum achievable rate using CCC with codeword composition P. For $P \in \Gamma_B^L$ (refer (5.8)), a CCC with *codeword* composition P will ensure that the average received energy per symbol in a codeword is at least B. However, it may violate the constraint on providing sufficient energy to the receiver *within every subblock duration*. For a CCC, we have [133]

$$C_{CCC}(P) = I(X;Y) = H(X) - H(X|Y). \tag{5.20}$$

We are interested in quantifying the information rate penalty incurred by using CSCC compared to CCC, given by $C_{CCC}(P) - C_{CSCC}^L(P)$. This information rate penalty is the price we pay for meeting the real-time energy requirement within every subblock duration, compared to the less constrained energy requirement per

codeword. Although the rate penalty can be numerically computed by explicit computation of $C_{CCC}(P)$ and $C^L_{CSCC}(P)$, the numerical approach has the limitation that the computation complexity of $C^L_{CSCC}(P)$ increases with an increase in subblock L.

In CSCC, since a transmitted subblock X_1^L is uniformly distributed over $\mathcal{T}_P^L$, we have [134], p. 26

$$H(X_1^L) = \log|\mathcal{T}_P^L| = LH(P) - L\,r(L,P), \tag{5.21}$$

where $r(L,P)$ denotes a function of L and P given as

$$r(L,P) = \frac{s(P)-1}{2L}\log(2\pi L) + \frac{1}{2L}\sum_{a:P(a)>0}\log P(a) + \frac{\vartheta(L,P)}{12L\ln 2}s(P), \tag{5.22}$$

with $s(P)$ denoting the number of elements $x \in \mathcal{X}$ with $P(x) > 0$, and $\vartheta(L,P)$ is a real number between zero and one which is chosen so that (5.21) is satisfied.

We now present simple analytical bounds for this rate penalty. The following theorem shows that the rate penalty by using CSCC, relative to CCC, is bounded by $r(L,P)$.

Theorem 4. *The rate penalty is bounded as*

$$0 \le C_{CCC}(P) - C^L_{CSCC}(P) \le r(L,P). \tag{5.23}$$

Further, there exist channels for which the rate penalty meets the upper or lower bound in (5.23) *with equality.*

Proof: When X_1^L is uniformly distributed over $\mathcal{T}_P^L$,

$$C_{CSCC}^L(P) = \frac{1}{L}\left[H(X_1^L) - H(X_1^L|Y_1^L)\right] \tag{5.24}$$

$$\begin{aligned} &\stackrel{(a)}{=} H(P) - r(L,P) - \frac{1}{L}\sum_{i=1}^{L} H\left(X_i|Y_1^L, X_1^{i-1}\right) \\ &\stackrel{(b)}{\geq} H(P) - r(L,P) - \frac{1}{L}\sum_{i=1}^{L} H(X_i|Y_i) \\ &\stackrel{(c)}{=} H(P) - r(L,P) - H(X|Y) \\ &\stackrel{(d)}{=} C_{CCC}(P) - r(L,P), \end{aligned} \tag{5.25}$$

where X_1^{i-1} denotes $X_1, \ldots, X_{i-1}$, (a) follows from (5.21) and chain rule for entropy, (b) follows since conditioning only reduces entropy, (c) follows from (5.14), and (d) follows from (5.20). Now, (5.23) follows from (5.25). Explicit channels can be constructed which meet the bounds in (5.23).

- $C_{CCC}(P) = C_{CSCC}^L(P) = 0$ for a binary symmetric channel (BSC) with crossover probability equal to 0.5.
- For a noiseless channel, we have $C_{CCC}(P) - C_{CSCC}^L(P) = r(L,P)$ due to equality in (b) as $\sum_{i=1}^{L} H(X_i|Y_1^L, X_1, \ldots, X_{i-1}) = \sum_{i=1}^{L} H(X_i|Y_i) = 0$.

■

Corollary 5.

$$\lim_{L\to\infty} C_{CSCC}^L(P) = C_{CCC}(P) \tag{5.26}$$

Proof: Note that for a fixed P, the value of $r(L,P)$ as a function of L is non-negative and falls roughly as $\log(L)/L$ and thus tends to zero as $L \to \infty$. Thus (5.26) follows by taking the limit $L \to \infty$ in (5.23). ■

Remark: For a fixed subblock length L, the CSCC capacity can be achieved by making the number of subblocks in a codeword arbitrarily large and performing

joint decoding over all the subblocks. However, when the number of subblocks in a codeword are kept constant and the subblock length is increased without bounds, then achievable rates using CSCC tend to CCC capacity. In particular, when there is only one subblock in a codeword, then the CSCC code is same as a CCC code whose capacity can be achieved by making L arbitrarily large.

Let $X_1^n = (X_1, X_2, \ldots, X_n)$ denote any codeword of length n. If we impose the average energy constraint on codewords,

$$\frac{1}{n}\sum_{i=1}^{n} b(X_i) \geq B, \tag{5.27}$$

then the channel capacity with this constraint is [14], [134]

$$\max_{P_X:\, \mathbb{E}_{P_X}[b(X)] \geq B} I(X;Y). \tag{5.28}$$

Information rates arbitrarily close to this capacity can be achieved by making the codeword length sufficiently large. Moreover, if P_X^* is an input distribution which maximizes (5.28), then this capacity can be achieved by a sequence of CCCs with codeword composition tending to P_X^* [133], [134]. Thus, if $C_{CCC}(B)$ denotes the capacity using CCC when the average energy per symbol is constrained to be at least B, then

$$C_{CCC}(B) = \max_{P:\, \mathbb{E}_P[b(X)] \geq B} C_{CCC}(P) \tag{5.29}$$

$$= \max_{P_X:\, \mathbb{E}_{P_X}[b(X)] \geq B} I(X;Y). \tag{5.30}$$

Thus the capacity with codeword constraints can be achieved by restricting the codewords to have a fixed composition. This is possible because for a given transmission rate, the codebook size increases exponentially with codeword length n

while the number of different types of sequences only increase polynomially with n.

The upper bound (5.22) on the rate penalty given by $r(L, P)$ is independent of the underlying channel. In general, given a communication channel, the bounds on rate penalty can be further improved. Consider, for example, a BSC with crossover probability p_0 where $0 < p_0 < 0.5$. For this channel, the upper bound can be tightened using Thm. 5. We first define a binary operator $\star$ and a function h, respectively, as

$$a \star b \triangleq a(1-b) + (1-a)b. \tag{5.31}$$

$$h(x) \triangleq -x \log x - (1-x)\log(1-x). \tag{5.32}$$

We employ the above definitions to state the following theorem on bounding the rate penalty for a BSC.

Theorem 5. *For a BSC with crossover probability $0 < p_0 < 0.5$, input distribution denoted by $P(0) = \Pr(X = 0)$, $P(1) = \Pr(X = 1)$, and $0 < \gamma = \min(P(0), P(1)) \leq 0.5$ we have,*

$$0 < C_{CCC}(P) - C^{L}_{CSCC}(P) \leq h(p_0 \star \gamma) - h(p_0 \star \alpha) < r(L, P), \tag{5.33}$$

where α is chosen such that

$$h(\alpha) = h(\gamma) - r(L, P), \quad 0 \leq \alpha < 0.5\ . \tag{5.34}$$

Proof: Please refer to Appendix B.4. ■

The proof of Theorem 5 uses Mrs. Gerber's Lemma (MGL) [135]. Using an extension [136] of MGL, the upper bound on the rate penalty can similarly be improved for general memoryless binary-input symmetric-output channels. In

particular, we have the following theorem for the binary erasure channel (BEC).

Theorem 6. *For a BEC with erasure probability $\epsilon > 0$,*

$$C_{CCC}(P) - C^{L}_{CSCC}(P) \leq (1-\epsilon) r(L,P) < r(L,P) \tag{5.35}$$

Proof: Please refer to Appendix B.5. ■

For memoryless asymmetric binary-input, binary-output channels, an alternate upper bound on the rate penalty (other than (5.23)) may be obtained using the equality of the channel *characteristic function* and the *gerbator* [137]. As an example, we have the following theorem for the Z-channel.

Theorem 7. *For a Z-channel with $\gamma = \Pr(X=1)$, and $p_0 = \Pr(1 \to 0)$, we have*

$$C_{CCC}(P) - C^{L}_{CSCC}(P) \leq h\left(\gamma(1-p_0)\right) - h\left(\alpha(1-p_0)\right), \tag{5.36}$$

where $h(\cdot)$ is given by (5.32)*, and α is chosen such that*

$$h(\alpha) = h(\gamma) - r(L,P), \quad 0 \leq \alpha < 0.5\ . \tag{5.37}$$

Proof: Please refer to Appendix B.6. ■

The rate penalty bound given by (5.36) may sometimes be worse than the bound in (5.23), depending on γ and p_0. In general, the rate penalty for the Z-channel can be upper bounded by $\min\left(r(L,P),\, h\left(\gamma(1-p_0)\right) - h\left(\alpha(1-p_0)\right)\right)$.

5.3.2 CSCC Error Exponent

For CCCs, it is well known [134], Thm. 10.2 that for every $R > \delta > 0$ and every type P of sequences in $\mathcal{X}^n$ there exists an n-length block code of rate at least $R - \delta$ such that all codewords are of type P and the maximum probability of

error is upper bounded by $\exp[-n(E_r(R,P,W)-\delta)]$ for every DMC W, whenever n is sufficiently large. Here $E_r(R,P,W)$, characterizing the exponential rate of decay of the probability of error with the blocklength, is called the *random coding exponent function* of channel W with input distribution P, and is defined as [134]

$$E_r(R,P,W) \triangleq \min_V \left(D(V||W|P) + [I(P,V)-R]^+ \right), \tag{5.38}$$

V ranging over all channels $V : \mathcal{X} \to \mathcal{Y}$, $D(V||W|P) = \sum_{x,y} P(x)V(y|x) \log \frac{V(y|x)}{W(y|x)}$, and $I(P,V) = H_P(X) - H_{P\times V}(X|Y)$.

The following theorem shows that the CSCC error exponent using subblock-composition P is related to the CCC error exponent by the same term, $r(L,P)$, used in rate loss bound (5.23).

Theorem 8. *For every $R > \delta > 0$, there exists a CSCC with subblock-composition P, fixed subblock length L, codeword length n, rate at least $R-\delta$, for which the maximum probability of error on DMC W is upper bounded as*

$$P_e \leq \exp[-n\left(E_r(R + r(L,P), P, W) - \delta\right)], \tag{5.39}$$

whenever n is a sufficiently large multiple of L. Thus, the CSCC error exponent using subblock-composition P, with rate R on DMC W is lower bounded by $E_r\left(R + r(L,P), P, W\right)$.

Proof: Please refer to Appendix B.7. ■

5.4 Real-time Information Transfer

So far, we could ensure real-time energy transfer to the receiver by placing constraints on the subblock-composition. For information transfer, although joint

decoding of all the subblocks within a codeword is preferred for reducing the probability of error, it also causes delay in information arrival.

For enabling real-time information transfer, the receiver may decode each subblock independently, and thus avoid waiting for arrival of future subblocks. Here, since the subblock decoding proceeds the instant that subblock has been completely received, the information transfer delay is only due to subblock transfer time and the corresponding decoding delay.

When each subblock within the transmitted sequence is decoded independent of other subblocks, then each subblock may itself be viewed as a codeword. We will refer to the independent decoding of subblocks as *local subblock decoding* (LSD). We remark that this subblock based decoding is distinct from decoding for locally decodable codes that allows any bit of the message to be decoded with high probability by only querying a small number of received bits [138].

5.4.1 Local Subblock Decoding

In case of local subblock decoding, each subblock may be treated as an independent codeword since every subblock is decoded independently. We are interested in estimating achievable rates with bounded error probability when local subblock decoding is employed. We now provide a short review of an existing result on achievable rates for constant composition finite blocklength codes. This result will then be used (in Sec. 5.5) to compare rates between local (independent) subblock decoding and joint subblock decoding.

Let $M^*(n, \epsilon)$ denote the maximum size of length-n constant composition code for a DMC with average error probability no larger than ϵ. When the composition of codewords is equal to an input probability distribution which maximizes the mutual information and the channel satisfies some regularity conditions, then [4],

[139], [140]

$$\log M^*(n,\epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + \frac{1}{2}\log n + O(1) \tag{5.40}$$

where C is the channel capacity, V is the *information variance*, and Q is the Gaussian Q-function [140]. We remark that V is also termed *channel dispersion* in literature [3]. Early results on finite blocklength capacity for memoryless symmetric channels are due to Weiss [141], which were generalized for the DMC and strengthened by Strassen [142].

When each codeword has equal number of ones and zeros, the achievable rate in bits per channel use for BSC with crossover probability p using CCC is approximated as [139]:

$$\frac{\log_2 M^*(n,\epsilon)}{n} \approx C - \sqrt{\frac{p(1-p)}{n}}\log_2\frac{1-p}{p}\,Q^{-1}(\epsilon) + \frac{1}{2n}\log_2 n, \tag{5.41}$$

with $C = 1 + p\log_2 p + (1-p)\log_2(1-p)$.

5.5 Numerical Results and Discussion

In this section, we provide examples highlighting the tradeoff between delivery of sufficient energy to the receiver and achieving high information transfer rates. These results are used to draw meaningful insights into choice of subblock length and subblock composition as a function of required energy per symbol at the receiver.

Fig. 5.3 plots $C^L_{CSCC}(B)$ as a function of B for different values of L for a BSC with crossover probability $p_0 = 0.1$. The b-values are assumed to be $b(0) = 0$ and $b(1) = 1$. These b-values reflect the case of on-off keying where bit-1 (bit-0) is represented by the presence (absence) of a carrier signal. Fig. 5.3 shows that, in general, the value of information rate given by $C^L_{CSCC}(B)$ increases with an

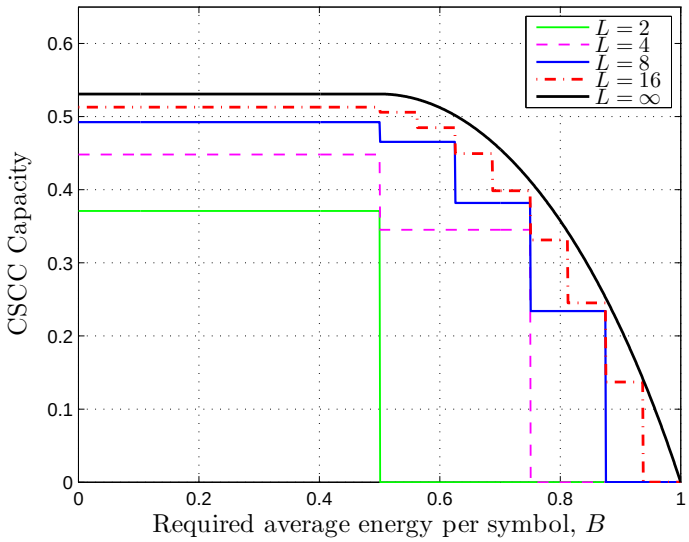

Figure 5.3: Plot of $C^L_{CSCC}(B)$ versus B for BSC with crossover probability $p_0 = 0.1$, $b(0) = 0$, $b(1) = 1$.

increase in the subblock length L, for a given B. This is because an increase in L leads to greater choice for input symbols within a subblock. Note that the smaller the value of L, the greater the uniformity in energy distribution within a codeword. The reduction in capacity due to choice of smaller L is the price we pay for providing smoother energy content.

The plot for $L = \infty$ is evaluated using (5.28); this follows from (5.9), (5.26) and (5.29). Thus the curve corresponding to $L = \infty$ is same as the $C_{CCC}(B)$ curve. This curve is a non-increasing concave function of B for $0 \leq B \leq b_{\max}$. This claim can be proved using the approach in [14]. It is non-increasing since the feasibility set Γ^L_B will only become smaller on increasing B. The concavity of $C_{CCC}(B)$ follows from the concavity of $I(X;Y)$ as a function of probability distribution of X and the fact that for $0 < \alpha < 1$, the conditions $\mathbb{E}_{P_1}[b(X)] \geq B_1$ and $\mathbb{E}_{P_2}[b(X)] \geq B_2$ imply that

$$\mathbb{E}_{\alpha P_1 + (1-\alpha)P_2}[b(X)] \geq \alpha B_1 + (1-\alpha)B_2. \tag{5.42}$$

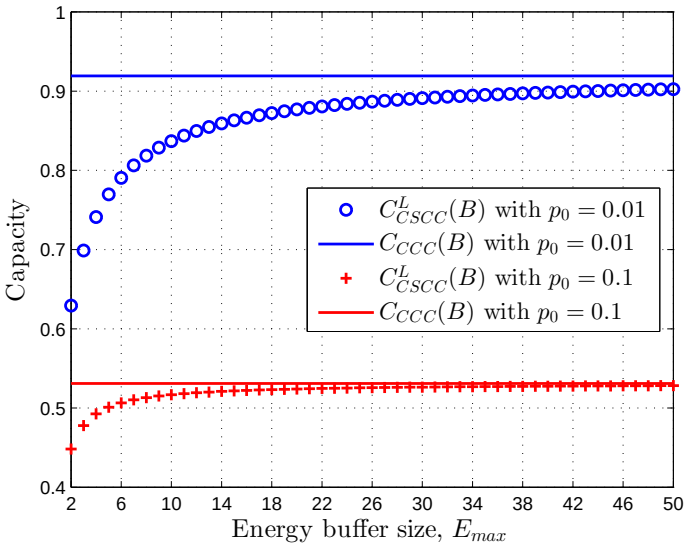

Figure 5.4: Plot of CSCC capacity versus receiver energy buffer size, E_{max}, with $B = 0.5, b(0) = 0, b(1) = 1$ for BSC with crossover probability $p_0 = \{0.01, 0.1\}$.

The non-increasing concave nature of the capacity-power function was used in [131] to show the suboptimality of a time-sharing approach to energy and information transfer.

The CSCC capacity is plotted in Fig. 5.4 for a BSC as a function of the receiver energy buffer size, E_{max}, with $B = 0.5$. The subblock length L is chosen as a function of E_{max} to satisfy (5.19). Since L increases with increasing values of E_{max}, the CSCC capacity is an increasing function of E_{max}. For $p_0 = 0.1$, the CSCC capacity is limited by the relatively high value of the crossover probability, rather than the subblock length, with capacity remaining almost constant as E_{max} is increased beyond 10. On the other hand, for $p_0 = 0.01$, the CSCC capacity is limited by the subblock length (since 'noise' is weak). From (5.19) we observe that the subblock length tends to infinity as E_{max} tends to infinity, and hence the CSCC capacity corresponding to $E_{max} \to \infty$ is equal to $C_{CCC}(B)$.

Fig. 5.5 plots the rate penalty incurred by using CSCC instead of CCC, for a BSC with crossover probability p_0, $L = 16$, and $\Pr(0) = \Pr(1) = 0.5$. As discussed

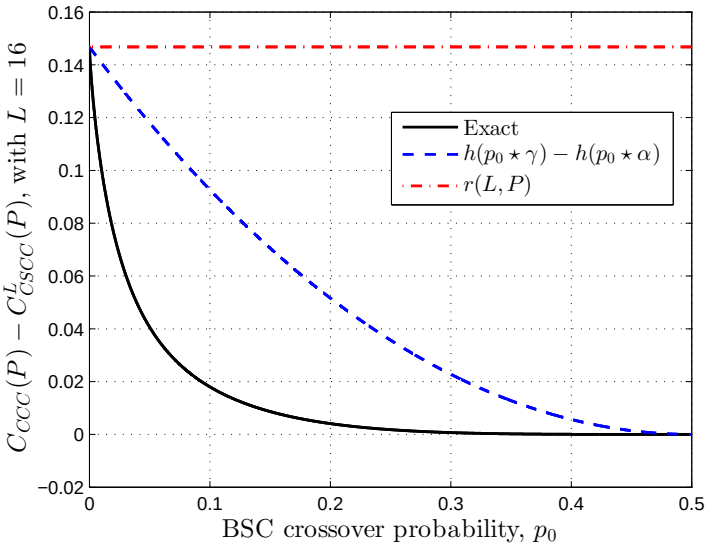

Figure 5.5: Plot of $C_{CCC}(P) - C^L_{CSCC}(P)$ as a function of BSC crossover probability p_0 for $L = 16$ and $\Pr(0) = \Pr(1) = 0.5$.

in Sec. 5.3.1, the upper bound on the rate penalty given by $r(L, P)$ is shown to be close to the exact value when $p_0 \approx 0$. Note that $r(L, P)$ is independent of the underlying channel. A tighter bound on the rate penalty given by $h(p_0 \star \gamma) - h(p_0 \star \alpha)$ is also plotted (see Theorem 5). These bounds are useful in estimating the rate penalty for large values of L when the computational complexity of $C^L_{CSCC}(P)$ becomes high. The bounds on rate penalty may also be used to bound the exact value of $C^L_{CSCC}(P)$ for large L.

Fig. 5.6 compares achievable rates using local subblock decoding (LSD) with rates using joint subblock decoding for a BSC with crossover probability $p_0 = 0.11$ when each subblock has equal number of zeros and ones (that is, $P(0) = P(1) = 0.5$). In case of CSCC with LSD, each subblock may itself be viewed as a codeword and so the achievable rate is approximated by (5.41) with $n = L$. The achievable rates with LSD are obtained using (5.41) and seen to fall significantly as the desired probability of error, ϵ, tends to zero. The red curve plots lower bound on $C^L_{CSCC}(P)$ obtained using (5.33). Note that $C^L_{CSCC}(P)$ represents the rate

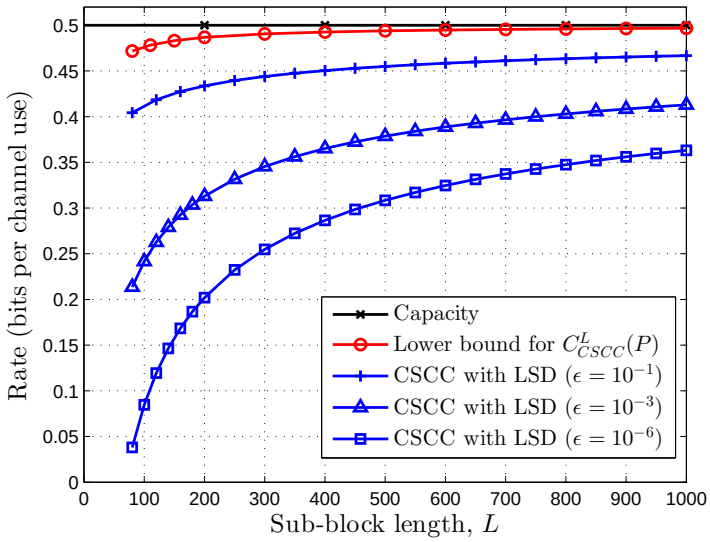

Figure 5.6: Rates for a BSC with crossover probability $p = 0.11$.

with joint subblock decoding for which the probability of error can be brought arbitrarily close to zero by increasing the number of subblocks in a codeword and then jointly decoding the subblocks.

Notice that the rate loss decreases as $\sqrt{1/L}$ with LSD whereas the rate loss with joint decoding decreases as $\log(L)/L$. Ensuring the ability to use energy in real-time imposes less of a penalty than the ability to use information in real-time.

5.6 Discussion

We proposed the use of CSCC codes for providing regular energy content in a patterned energy signal which is used for simultaneous transfer of energy and information. The subblock-composition in CSCC was chosen to maximize the rate of information transfer while ensuring that the fraction of input symbols carrying high energy within every subblock duration are sufficiently large. For characterizing the exact CSCC capacity, we employed a super-letter approach (with each subblock being viewed as a single super-letter in an induced vector-channel) and

showed that CSCC capacity computational complexity can be alleviated by exploiting certain symmetry properties.

Compared to constant composition codes (CCCs), the use of CSCCs incurs a rate loss due to the constraint restricting the subblocks to have the same composition. We showed that the CSCC error exponent is related to the CCC error exponent by the same rate loss term. We also presented numerical results for different scenarios and showed that ensuring the ability to use energy in real-time imposes less of a penalty than the ability to use information in real-time.

Chapter 6

Subblock Energy-Constrained Codes

Consider an energy-harvesting receiver that uses the same received signal both for decoding information and for harvesting energy, which is employed to power its circuitry. In the scenario where the receiver has limited battery size, a signal with bursty energy content may cause power outage at the receiver since the battery will drain during intervals with low signal energy. In Chapter 5 we studied constant subblock-composition codes (CSCCs) where all subblocks in every codeword have the same fixed composition, and this subblock-composition is chosen to maximize the rate of information transfer while meeting the energy requirement. In this chapter, we show that CSCC capacity can be improved by allowing different subblocks to have different composition while still meeting the subblock energy constraint.

We consider discrete memoryless channels (DMCs) and study *subblock energy-constrained codes* (SECCs) which by definition are codes which carry sufficient energy in every *subblock* duration. Compared to constant subblock-composition codes (CSCCs) analyzed in the previous chapter, the SECCs allow different sub-

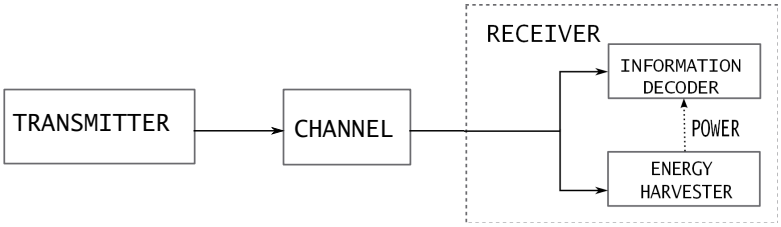

Figure 6.1: Simultaneous energy and information transfer from a transmitter to an energy-harvesting receiver

blocks to have different composition and thus achieve higher rates of information transfer while still meeting the real-time energy requirement.

Our contributions in this chapter are as follows. We provide a sufficient condition on the subblock length to avoid power outage at the receiver for all possible SECC sequences (Sec. 6.2.1). We characterize the exact SECC capacity (Sec. 6.2.2), and also provide different bounds on the SECC capacity (Sec. 6.2.3). Further, we characterize and bound the random coding error exponent for SECCs in Sec. 6.2.4, and present numerical results in Sec. 6.3.

6.1 System Model

Consider communication from a transmitter to a receiver which uses the received signal both for decoding information as well as for harvesting energy (Fig. 6.1). We model the effective channel from the output of a modulator at the transmitter to the (quantized) input to an information decoder at the receiver as a discrete memoryless channel (DMC), with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and a stochastic matrix $W = \{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ where $W(y|x)$ is the probability that the output is y when the channel input is x.

For simultaneous energy and information transfer, we consider a receiver architecture where the received signal power is split between the energy harvesting path and the information processing path with a static power-splitting ratio. We

let $b(x)$ denote the energy harvested when $x \in \mathcal{X}$ is transmitted, $b : \mathcal{X} \to [0, \infty)$. The map b is assumed to be time-invariant, and reflects the scenario where the statistical nature of the effective communication channel is due to the noise in the receiver circuitry, which does not affect the harvested energy.

To meet the real-time energy requirement at the receiver, we partition the transmitted codeword into equal-sized subblocks and require that transmitted symbols be chosen so the expected harvested energy in *each subblock* exceeds a given threshold. We assume that the subblock length, denoted L, is *fixed* while the codeword length, denoted n, can be made arbitrarily large by increasing the number of subblocks within each codeword. If a transmitted codeword is denoted $(X_1, X_2, \ldots, X_n)$, then the constraint on sufficient energy within each subblock is:

$$\frac{1}{L}\sum_{i=1}^{L} b\left(X_{(j-1)L+i}\right) \geq B,\ j = 1, 2, \ldots, k\ , \tag{6.1}$$

where j is the subblock index, B is the required energy per symbol at the receiver, and k is the number of subblocks in a codeword. We assume that $b_{\min} < B < b_{\max}$, where $b_{\min} = \min_{x\in\mathcal{X}} b(x)$, and $b_{\max} = \max_{x\in\mathcal{X}} b(x)$.

6.2 Subblock Energy-Constrained Codes

When $N_j(x)$ denotes the number of occurrences of symbol x in subblock j within a codeword, the subblock energy constraint (6.1) can equivalently be expressed as

$$\sum_{x\in\mathcal{X}} b(x)\frac{N_j(x)}{L} \geq B,\ j = 1, 2, \ldots, k. \tag{6.2}$$

A *subblock energy-constrained code* (SECC) is defined as one in which all codewords are partitioned into length-L subblocks and the composition of each subblock is chosen to satisfy (6.2). Note that if x_1^L denotes a given subblock of

length L, then the composition of x_1^L is the distribution $P_{x_1^L}$ on $\mathcal{X}$ defined by $P_{x_1^L}(x) \triangleq \frac{N(x)}{L}$, $x \in \mathcal{X}$, where $N(x)$ is the number of occurrences of symbol x in subblock x_1^L.

Let $\mathcal{P}_L$ denote the set of all compositions for input sequences of length L. Since all subblocks in SECC satisfy (6.2), the composition of each subblock belongs to the set

$$\Gamma_B^L \triangleq \{P \in \mathcal{P}_L : \sum_{x \in \mathcal{X}} b(x)P(x) \geq B\}. \tag{6.3}$$

We will let J denote the number of distinct compositions in Γ_B^L, and these compositions will be denoted as $P_j, 1 \leq j \leq J$, and so $\Gamma_B^L = \{P_1, \ldots, P_J\}$.

6.2.1 Choosing the subblock length

We impose subblock energy constraint (6.1) because a codeword satisfying only the codeword energy constraint may still cause power outage at the receiver if the energy content in the codeword is bursty, since a receiver battery with small capacity may drain during periods of low signal energy. The receiver energy update equation after i channel uses is given by $E(i+1) = \min\left(E_{max}, [E(i) + b(X_i) - B]^+\right)$, where $E(i)$ denotes the energy level at the receiver after the completion of $i-1$ channel uses, E_{max} denotes the receiver energy storage capacity, and $[z]^+ \triangleq \max(z, 0)$. We say an *outage* occurs during the ith channel use if $E(i) + b(X_i) < B$, while an *overflow* occurs if $E(i) + b(X_i) - B > E_{max}$. The following theorem gives a sufficient condition on the subblock length to avoid power outage at the receiver for all possible SECC codeword sequences, where we employ the notation: $\mathcal{X}_\triangleleft = \{x \in \mathcal{X} \,|\, b(x) < B\}$, $\mathcal{X}_\triangleright = \{x \in \mathcal{X} \,|\, b(x) \geq B\}$.

Theorem 9. *For each $P_j \in \Gamma_B^L$, $1 \leq j \leq J$, let $S_j = \sum_{x \in \mathcal{X}_\triangleleft} P_j(x)\,(B - b(x))$ and $S_{max} = \max_{j \in \{1,\ldots,J\}} S_j$. Then a sufficient set of condition to avoid outage for all*

possible SECC codeword sequences is

$$L \leq \frac{E_{max}}{2S_{max}} \ , \quad E(1) \geq LS_{max} \ , \tag{6.4}$$

where $E(1)$ is the initial level of the receive energy buffer.

Proof: Please refer to Appendix C.1. ■

The initial energy level may be ensured by transmitting a preamble, consisting of high energy symbols, before the transmission of codewords. This preamble has bounded length and hence does not affect the channel capacity.

6.2.2 SECC capacity

The set of sequences in $\mathcal{X}^L$ with composition P is denoted by $\mathcal{T}_P^L$ and is called the *type class* or *composition class* of P. When SECC constrained codes are employed on DMC $W : \mathcal{X} \to \mathcal{Y}$, we may view the L uses of the channel as a single use of the induced *vector channel* having input alphabet

$$\mathcal{A} = \bigcup_{P \in \Gamma_B^L} \mathcal{T}_P^L \ = \ \bigcup_{1 \leq j \leq J} \mathcal{T}_{P_j}^L \ , \tag{6.5}$$

and output alphabet $\mathcal{Y}^L$. Since the underlying scalar channel W is a DMC, the vector channel is also a DMC with transition probability for a pair of input and output vectors given by

$$W^L(y_1^L|x_1^L) = \prod_{i=1}^{L} W(y_i|x_i), \quad x_1^L \in \mathcal{A}, \ y_1^L \in \mathcal{Y}^L. \tag{6.6}$$

Let the codeword length, n, be of the form $n = kL$, where k is an integer denoting the number of subblocks in each codeword. We wish to quantify performance limits when the subblock length L is fixed and $k \to \infty$. For SECC, each kL-length

codeword may be viewed as an element of $\mathcal{A}^k$, and the received word belongs to the set $\left(\mathcal{Y}^L\right)^k$. A *kL-length SECC block code* for a channel $\left(W^L\right)^k : \mathcal{A}^k \to \left(\mathcal{Y}^L\right)^k$ is a pair of mappings (f, ϕ) where f maps a finite message set $\mathcal{M}$ into $\mathcal{A}^k$, and ϕ maps $\mathcal{Y}^{kL}$ into $\mathcal{M}$. The probability of erroneous transmission of message $m \in \mathcal{M}$ is $e_m \triangleq 1 - \sum_{y_1^{kL}:\phi(y_1^{kL})=m} W^{kL}(y_1^{kL}|f(m))$, the maximum probability of error of the code (f, ϕ) is $e \triangleq \max_{m\in\mathcal{M}} e_m$, while the rate of this code is $\frac{1}{kL}\log|\mathcal{M}|$. We call a kL-length SECC block code with maximum probability of error upper bounded by ϵ as a (kL, ϵ)-*SECC code*.

Definition 1. *For a fixed subblock length L, and for $0 \leq \epsilon < 1$, a non-negative number R is an* ϵ-SECC achievable rate *for the channel $W^{kL} : \mathcal{A}^k \to \mathcal{Y}^{kL}$ if for every $\delta > 0$ and every sufficiently large k there exist (kL, ϵ)-SECC codes with rate exceeding $R - \delta$. A number R is an* SECC achievable rate *if it is ϵ-SECC achievable for all $0 < \epsilon < 1$, and the supremum of SECC achievable rates is the* SECC capacity *of channel W.*

The induced vector channel W^L (6.6) is a DMC with input alphabet $\mathcal{A}$ and output alphabet $\mathcal{Y}^L$, and hence its capacity is $\max_{P_{X_1^L}:X_1^L\in\mathcal{A}} I(X_1^L; Y_1^L)$, where the maximization is over the distribution of $X_1^L \in \mathcal{A}$. Since kL uses of W correspond to k uses of W^L, the SECC capacity, denoted $C_{SECC}^L(B)$, is $1/L$ times the capacity of this vector channel, and so

$$C_{SECC}^L(B) = \max_{P_{X_1^L}:X_1^L\in\mathcal{A}} \frac{I(X_1^L; Y_1^L)}{L}, \tag{6.7}$$

where the maximization is over the distribution of subblocks over $\mathcal{A}$, where the set $\mathcal{A}$ is related to B via Γ_B^L (see (6.5)).

Remark: Note that $C_{SECC}^L(B)$ is non-increasing in B because the set Γ_B^L only becomes smaller on increasing B. Further, if $\tilde{L} = mL$, where m is a positive integer, then $\mathcal{P}_L \subset \mathcal{P}_{\tilde{L}}$ and so $\Gamma_B^L \subset \Gamma_B^{\tilde{L}}$, which implies $C_{SECC}^L(B) \leq C_{SECC}^{\tilde{L}}(B)$

because $C_{SECC}^{\tilde{L}}(B)$ is obtained by optimizing the input distribution over a richer alphabet.

Finding a capacity-achieving input distribution in (6.7) is not always straightforward, and one may have to resort to the Blahut-Arimoto algorithm [16], [17]. There even exist *symmetric* DMCs for which the corresponding induced vector channel is *not* symmetric, where a uniform distribution over $\mathcal{A}$ does not achieve capacity [143].

However, for any given channel, there always exists a SECC capacity-achieving input distribution $P_{X_1^L}^*$ satisfying $P_{X_1^L}^*(x_1^L) = P_{X_1^L}^*(\tilde{x}_1^L)$, whenever x_1^L and $\tilde{x}_1^L$ belong to the same composition class.

Theorem 10. *Let x_1^L and $\tilde{x}_1^L$ be any two vectors in $\mathcal{A}$ having the same composition. Then for a given channel W, there exists a SECC capacity-achieving* (6.7) *input distribution $P_{X_1^L}^*$ satisfying $P_{X_1^L}^*(x_1^L) = P_{X_1^L}^*(\tilde{x}_1^L)$.*

Proof: Please refer to Appendix C.2. ■

6.2.3 Bounds on SECC capacity

The elements of the random vector X_1^L, in general, are *not independent* because X_1^L belongs to the constrained set $\mathcal{A}$. The elements of X_1^L are, however, *identically* distributed.

Lemma 2. *Let $P_{X_1^L}^*$ be a SECC capacity-achieving input distribution of Theorem 10, and define*

$$c_j \triangleq \sum_{x_1^L \in \mathcal{T}_{P_j}^L} P_{X_1^L}^*(x_1^L), \quad P_j \in \Gamma_B^L,\ j \in \{1, \ldots, J\}. \tag{6.8}$$

Then each element X_i in X_1^L has identical distribution $\tilde{P}$ with

$$\tilde{P}(x) = \sum_{j=1}^{J} c_j P_j(x), \quad x \in \mathcal{X}. \tag{6.9}$$

Proof: For a given j, when all the vectors in $\mathcal{T}_{P_j}^L$ have equal probabilities, then a combinatorial argument [143] shows that for $1 \leq i \leq L$, we have $\Pr(X_i = x | X_1^L \in \mathcal{T}_{P_j}^L) = P_j(x)$. Thus (6.9) follows since $c_j = \Pr(X_1^L \in \mathcal{T}_{P_j}^L)$, and $\Pr(X_i = x) = \sum_{j=1}^{J} \Pr(X_1^L \in \mathcal{T}_P^L) \Pr(X_i = x | X_1^L \in \mathcal{T}_P^L)$. ■

The unconditional distribution $\tilde{P}$ (6.9) for each letter in a SECC codeword can be used to bound the SECC capacity.

Theorem 11. *$I(\tilde{P}, W) - \tilde{r} \leq C_{SECC}^L(B) \leq I(\tilde{P}, W)$, with*

$$I(\tilde{P}, W) \triangleq H_{\tilde{P}}(X) - H_{\tilde{P} \times W}(X|Y) = H(\tilde{P}) - H_{\tilde{P} \times W}(X|Y)$$
$$\tilde{r} \triangleq H(\tilde{P}) - \sum_{j=1}^{J} c_j \frac{\log |\mathcal{T}_{P_j}^L|}{L} - \frac{1}{L} \sum_{j=1}^{J} c_j \log \frac{1}{c_j}. \tag{6.10}$$

Proof: Please refer to Appendix C.3. ■

Next, we bound the SECC capacity using expressions which are independent of $P_{X_1^L}^*$.

The constant subblock-composition codes (CSCCs) [144] are a subclass of SECCs where all subblocks in every codeword have the same *fixed* composition. If the fixed composition is denoted P, then all the subblocks in CSCC belong to $\mathcal{T}_P^L$, and this composition meets the subblock energy constraint (6.2). The CSCC capacity, denoted $C_{CSCC}^L(B)$, can be expressed as [144]

$$C_{CSCC}^L(B) = \max_{P \in \Gamma_B^L} \frac{I\left(\mathcal{U}\left(\mathcal{T}_P^L\right), W^L\right)}{L}, \tag{6.11}$$

where $\mathcal{U}\left(\mathcal{T}_P^L\right)$ denotes a uniform distribution over $\mathcal{T}_P^L$. Now, instead of imposing

the energy constraint per subblock (6.1), if we impose the energy constraint *per codeword* $\frac{1}{L}\sum_{i=1}^{n} b(X_i) \geq B$, then the capacity, denoted $C_{CCC}(B)$, is [14]

$$C_{CCC}(B) = \max_{P_X : \mathbb{E}_{P_X}[b(X)] \geq B} I(X;Y). \tag{6.12}$$

Theorem 12. $C^L_{CSCC}(B) \leq C^L_{SECC}(B) \leq C_{CCC}(B)$.

Proof: In contrast with CSCCs where all subblocks have the same fixed composition, a general SECC has the flexibility of choosing different subblocks with different compositions. The lower bound on the SECC capacity hence follows because $C^L_{SECC}(B)$ is obtained via optimizing the subblock distribution over a richer input alphabet.

The upper bound on the SECC capacity follows since the energy constraint per subblock is stricter than the energy constraint per codeword. ■

Using $\lim_{L\to\infty} C^L_{CSCC}(B) = C_{CCC}(B)$ [144], and the above theorem, it follows that $\lim_{L\to\infty} C^L_{SECC}(B) = C_{CCC}(B)$.

6.2.4 SECC error exponent

It is well known [134], Thm. 10.2 that for every $R > \delta > 0$ there exists an n-length constant composition code of rate at least $R - \delta$ such that all codewords have composition $\tilde{P}$ and the maximum probability of error is upper bounded by $\exp[-n(E_r(R, \tilde{P}, W) - \delta)]$ for every DMC W, whenever n is sufficiently large. Here $E_r(R, \tilde{P}, W)$, characterizing the exponential rate of decay of the probability of error with the blocklength, is called the *random coding exponent function* of channel W with input distribution $\tilde{P}$, and is defined as [134]

$$E_r(R, \tilde{P}, W) \triangleq \min_{V} \left(D(V||W|\tilde{P}) + [I(\tilde{P}, V) - R]^+ \right), \tag{6.13}$$

V ranging over all channels $V : \mathcal{X} \to \mathcal{Y}$.

As discussed in Sec. 6.2.2, the L uses of channel W with SECC may be viewed as a single use of the *vector channel* W^L (6.6). Thus, each n-length SECC codeword may be viewed as a sequence of n/L *super-letters* to be transmitted on vector channel W^L. Since rate R for the *scalar* channel corresponds to rate LR for the *vector* channel, there exists a SECC code for which the maximum probability of error on W^L, with codewords comprising of n/L *super-letters*, is bounded as [134]

$$P_e \leq \exp[-\frac{n}{L}\left(E_r(LR, P^*_{X_1^L}, W^L) - L\delta\right)], \tag{6.14}$$

$$= \exp[-n\left(\frac{E_r(LR, P^*_{X_1^L}, W^L)}{L} - \delta\right)], \tag{6.15}$$

where $P^*_{X_1^L}$ is a SECC capacity-achieving distribution of Thm. 10, and $E_r(LR, P^*_{X_1^L}, W^L)/L$ is the *random coding error exponent* for SECCs with rate R over channel W . The following theorem bounds the error exponent $E_r(LR, P^*_{X_1^L}, W^L)/L$ in terms of $E_r(R, \tilde{P}, W)$ and $\tilde{r}$ (6.10).

Theorem 13. $\dfrac{E_r(LR, P^*_{X_1^L}, W^L)}{L} \geq E_r(R + \tilde{r}, \tilde{P}, W).$

Proof: Please refer to Appendix C.4. ■

6.3 Numerical Results and Discussion

In this section, we provide examples highlighting the tradeoff between delivery of sufficient energy to the receiver and achieving high information transfer rates. These results are used to draw meaningful insights into choice of subblock length and subblock composition as a function of required energy per symbol at the receiver.

Fig. 6.2 compares the capacity of CSCC and SECC for a noiseless binary channel with $b(0) = 0, b(1) = 1$ and subblock length $L = 8$. Note that the

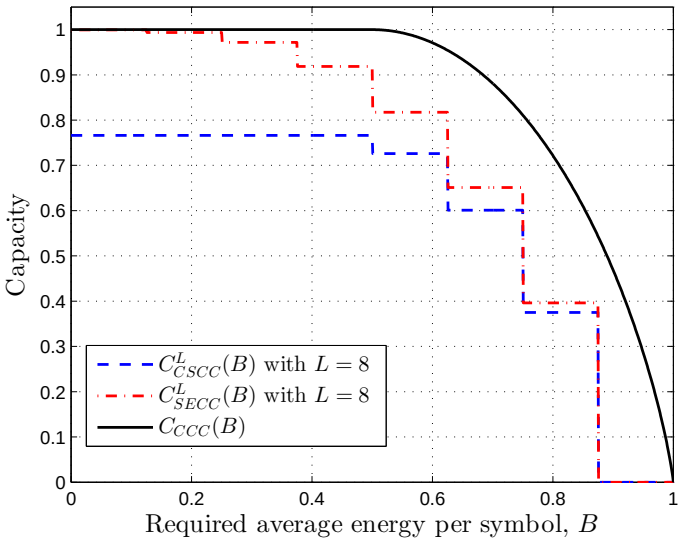

Figure 6.2: Comparison of capacity of different schemes for a noiseless binary channel with $b(0) = 0, b(1) = 1$.

capacity curve for CCC may be viewed as the CSCC capacity curve corresponding to $L = \infty$. Fig. 6.2 highlights the potential of improving the CSCC capacity by using SECCs and allowing different subblocks to have different compositions while still meeting the subblock energy constraint (6.1). Note that the SECC capacity for a noiseless channel is achieved by a uniform distribution of input vectors and is given by $(\log |\mathcal{A}|)/L$, where $\mathcal{A}$ is given by (6.5).

Fig. 6.3 compares capacity of different schemes for $L = 8$ and $B = 0.6$, as a function of BSC crossover probability p_0. It shows that for $p_0 < 0.05$, the capacity with uniform distribution over the set of length L vectors which satisfy the subblock energy constraint (6.1), is higher compared to CSCC capacity. However, $C^L_{U_{\mathcal{A}}}(B) < C^L_{CSCC}(B)$ for relative higher values of p_0. This observation emphasizes the fact that merely adding more types is not sufficient to increase capacity compared to CSCC; we need to choose an appropriate distribution over the enlarged alphabet as well. In Fig. 6.3, we used the Blahut-Arimoto algorithm [16], [17] to compute the exact SECC capacity, $C^L_{SECC}(B)$.

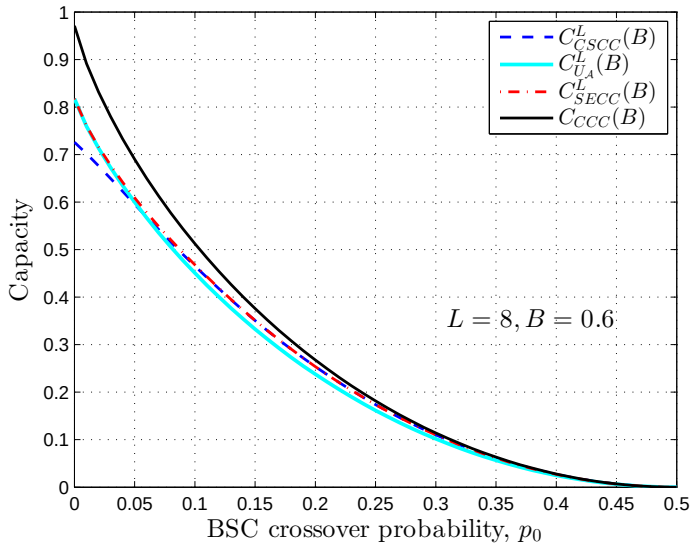

Figure 6.3: Comparison of capacity of different schemes for $L = 8, B = 0.6$, as a function of BSC crossover probability p_0 and $b(0) = 0, b(1) = 1$.

6.4 Discussion

In this chapter, we proposed the use of SECCs for real-time simultaneous energy and information transfer, and gave a sufficient condition on the subblock length to avoid receiver power outage. We characterized the SECC capacity and the SECC random coding error exponent, and also presented bounds for these values. Compared to the constant subblock composition codes (CSCCs) discussed in Chapter 5, we showed that the SECCs achieve higher information rates due to the flexibility of allowing different compositions for different subblocks.

We denoted by $b(x)$, the amount of energy harvested when symbol x is transmitted, and assumed that b is time-invariant, a reasonable assumption for short-range line-of-sight wireless power transfer. As an extension, one may consider a time-varying b corresponding to wireless communication with fading.

Chapter 7

Codes Satisfying Sliding Window Energy Constraint

We consider the problem of binary code design for simultaneous energy and information transfer where the receiver completely relies on the received signal for fulfilling its real-time power requirements. When on-off keying is employed (where "1" (resp. "0") is represented by the presence (resp. absence) of a carrier), a majority transmission of "1" indicates a greater opportunity for the receiver to use the signal to fulfill its power requirements.

In order to provide real-time energy transfer to the receiver, a *subblock-constrained* approach was adopted in Chapters 5 and 6. In this approach, each codeword was divided into subblocks and each subblock was constrained to carry sufficient energy. Another approach to providing real-time energy transfer is to ensure that each codeword carries sufficient energy within a moving time window.

In this chapter, we study binary codes in which each codeword is constrained to have at least d ones in a *sliding window* of $W = d + 1$ consecutive bits. This constraint is equivalent to having at least d ones between successive zeros, which in turn defines a Type-1 (d, ∞) run-length limited (RLL) code. Note that, in

general, a Type-1 (d, k) RLL code is one in which the number of ones between successive zeros in each codeword is at least d and at most k.

We give a probabilistic proof that the noiseless capacity of a (d, ∞) RLL code can be achieved by using a $d+1$ state Markovian chain. The state transition probabilities for this Markov chain are explicitly provided and any sequence obtained from these state transitions satisfies the given codeword constraint. We also give analytical expressions for achievable rates when these constrained codes are used on the (i) binary symmetric channel (BSC), (ii) Z-channel and (iii) binary erasure channel (BEC). Although a few of these results are well known for run-length codes used for data storage, they do not seem to appear in literature in the form presented here.

Type-0 (d, k) RLL codes (where the number of zeros between successive ones are at least d and at most k) have been used for magnetic and optical recording and researchers usually refer to Type-0 (d, k) RLL codes simply as (d, k) codes [78], [79], [145], [146]. However, unless specified otherwise in this chapter, by (d, k) codes and (d, ∞) codes we shall mean Type-1 (d, k) RLL codes and Type-1 (d, ∞) RLL codes, respectively.

Note that apart from the physical interpretation of constraints, there is no combinatorial difference between Type-0 and Type-1 RLL codes when they are used on symmetric channels. However differences arise when these codes are employed on asymmetric channels, like the Z-channel.

A (d, ∞) code can be represented by transitions between $d+1$ states as shown by the finite-state machine in Fig. 7.1 [147], [146]. We remark that a (d, k) code can be represented by a state machine with $k+1$ states, but it is not desirable to view a (d, ∞) code as a special case of (d, k) code with $k = \infty$ since the state space becomes infinite.

The use of (d, k) codes for simultaneous energy and information transfer has

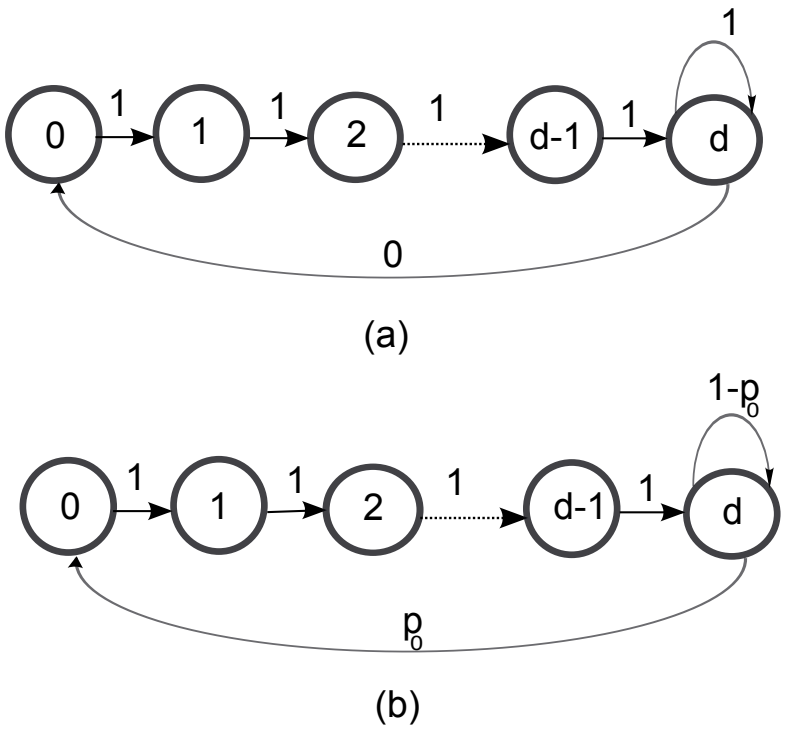

Figure 7.1: State transition diagram of a (d, ∞) code. (a) State transitions labeled by output bit, (b) State transitions labeled by transition probabilities.

been proposed in [81], [82]. In [81], the tag-to-reader channel in RFID systems is modeled as a *discretized Gaussian shift* channel and the frame error rate of finite blocklength (d, k) codes is compared through simulations. In [82], the communication is assumed over a Z-channel and the receiver is assumed to be equipped with a finite energy buffer. The performance of different (d, k) codes is compared through numerical optimization over state transition probabilities. In comparison, in this chapter we derive analytical expressions for state transition probabilities which lead to maximization of information rates using (d, ∞) codes over BSC, BEC, Z-channel.

We analyze the noiseless capacity in Section 7.1 while achievable rates for noisy channels are investigated in Section 7.2. Numerical results and conclusions are presented in Sections 7.3 and 7.4, respectively.

7.1 (d, ∞) Code Capacity

If M_N denotes the maximum number of distinct binary sequences of length N satisfying the (d, ∞) constraint, then the (d, ∞) code capacity is given by

$$C_0 = \lim_{N \to \infty} \frac{\log_2 M_N}{N}. \tag{7.1}$$

This was first studied by Shannon [1] and the code capacity is given by the logarithm of the largest root of the following characteristic equation (see [79] and [148])

$$z^{d+1} - z^d - 1 = 0. \tag{7.2}$$

It is interesting to note that solutions to the above equation, for different values of d, are related to certain constants, called Meru constants, obtained from recurrence relations studied by the ancient mathematician Pingala in his work on rhythm and meter in Sanskrit poetry [149]. In particular, the first, second, and fourth Meru constants can be shown to be equal to the largest root of (7.2) for $d = 1$, $d = 2$, and $d = 3$, respectively.

Let S_n denote the state of the Markov machine in Fig. 7.1 at time n where $S_n \in \{0, 1, \ldots, d\}$. A transition from state S_{n-1} to state S_n produces bit X_n. From Fig. 7.1(a) it follows that $X_n = 0$ when the machine transitions from state d to state 0 and $X_n = 1$ for all other state transitions. From Fig. 7.1(b) we note that the transition probability from state d to state 0 is denoted by p_0 and the transition probability from state d to itself is equal to $1 - p_0$.

This Markov machine is irreducible and aperiodic. Thus, it has a stationary probability distribution which we denote by $\{\pi_j\}_{j=0}^{d}$ where π_j is the steady state probability of the Markov chain being in state j. For this Markov chain, the

entropy rate (or information rate) is

$$R = H(S_2|S_1) = \pi_d \, H(p_0), \tag{7.3}$$

where $H(p_0) = 1 + p_0 \log_2 p_0 + (1 - p_0) \log_2(1 - p_0)$.

The value of R is uniquely given by the choice of d and the transition probability p_0. We will now prove that the sequences generated using the finite state machine given by Fig. 7.1 achieve the binary (d, ∞) code capacity through an appropriate choice of the transition probability p_0. That is, for a given d, the following relation is satisfied

$$C_0 = \max_{p_0} R. \tag{7.4}$$

The asymptotic equipartition property (AEP) holds for Markov sources and the number of typical sequences of length n is approximately given by $2^{nH(S_2|S_1)}$ where $H(S_2|S_1)$ is the entropy rate (or information rate) of the Markov source [20]. In general, if a constrained code is represented by a finite Markov model then, using the AEP, it can be proved that there exist state transition probabilities such that the entropy rate of the Markov source matches the constrained code capacity [146]. A similar result is also obtained in [145] by enumerating the distinct sequences that a Markov source can generate using its associated *connection matrix*.

Here, we give a simple proof of (7.4) when (d, ∞) constrained codes are represented by Markov model in Fig. 7.1. A useful outcome of our proof is that closed-form expressions for the optimized transition probabilities are explicitly presented. This seems not to have appeared in the literature before.

Theorem 14. *The maximum information rate of the Markov source governed by the state machine given in Fig. 7.1 is equal to the (d, ∞) code capacity given by the logarithm of the largest real root of the following characteristic equation $z^{d+1} - z^d - 1 = 0$. This capacity is achieved by choosing p_0 as the largest real value*

for which $p_0 = (1 - p_0)^{d+1}$.

Proof: The steady state probability distribution satisfies

$$[\pi_0 \, \pi_1 \cdots \pi_d] = [\pi_0 \, \pi_1 \cdots \pi_d] \, A, \tag{7.5}$$

where A denotes the transition probability matrix for the finite state machine in Fig. 7.1. The (i, j) entry of A is the transition probability from state i to state j and A is given by

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & 0 & \ddots & 1 \\ p_0 & 0 & 0 & 0 & \ddots & 1 - p_0 \end{bmatrix} \tag{7.6}$$

The diagonal of A is all zeros except the bottom corner, which is $1 - p_0$. Solving (7.5), we get

$$\pi_0 = \pi_1 = \cdots = \pi_{d-1} = p_0 \, \pi_d. \tag{7.7}$$

Since the steady state probabilities sum to one, we have

$$\pi_d = \frac{1}{1 + p_0 d} \qquad \text{and} \qquad \pi_0 = \cdots = \pi_{d-1} = \frac{p_0}{1 + p_0 d}. \tag{7.8}$$

Using (7.3) and (7.8), the information rate is given by

$$R = \frac{H(p_0)}{1 + p_0 d}. \tag{7.9}$$

To solve for p_0 which maximizes R, we equate the derivative of R with respect to

p_0 to zero

$$\frac{\partial R}{\partial p_0} = \frac{(1+p_0 d)\log_2\left(\frac{1-p_0}{p_0}\right) - dH(p_0)}{(1+p_0 d)^2} = 0. \tag{7.10}$$

The above equation yields

$$p_0 = (1-p_0)^{d+1}. \tag{7.11}$$

Substituting (7.11) in (7.9), we get

$$R = \log_2(1-p_0)^{-1}. \tag{7.12}$$

From (7.11), we see that $(1-p_0)$ satisfies the equation $z^{d+1} = 1-z$. Equivalently, $(1-p_0)^{-1}$ satisfies the equation

$$1 - z^{-1} - z^{-(d+1)} = 0 \quad \Leftrightarrow \quad \frac{z^{d+1} - z^d - 1}{z^{d+1}} = 0. \tag{7.13}$$

Thus, the maximum information rate is given by the logarithm of the largest real root of the equation $z^{d+1} - z^d - 1 = 0$. ■

7.2 Achievable rate using (d, ∞) code on memoryless channels

Consider a memoryless channel with input sequence $\mathbf{X}^N$ (satisfying the (d,∞) constraint) and output sequence $\mathbf{Y}^N = (Y_1, \ldots, Y_N)$. The channel capacity in this scenario is equal to [78], [79]

$$C = \lim_{N\to\infty} \sup_{P(\mathbf{X}^N)} \frac{I(\mathbf{X}^N; \mathbf{Y}^N)}{N} \tag{7.14}$$

$$= \lim_{N\to\infty} \sup_{P(\mathbf{S}^N)} \frac{I(\mathbf{S}^N; \mathbf{Y}^N)}{N}, \tag{7.15}$$

where, in the first equality the supremum is taken over all probabilities $P(\mathbf{X}^N)$ for the input sequence. In the second equality, the supremum is taken over all probabilities $P(\mathbf{S}^N)$ for the sequence of states. The second equality follows since given the initial state, the sequences $\mathbf{X}^N$ and $\mathbf{S}^N$ are in one-to-one correspondence, and the initial states does not affect the average mutual information [79].

Although the channel capacity using (d, ∞) codes given by (7.15) is difficult to obtain for noisy channels, a useful lower bound on the capacity for a stationary Markovian source over memoryless channels is given as [78]

$$C \geq C_{LB} = \sup_{P(S_1,S_2)} I(S_2; Y_2|S_1). \tag{7.16}$$

In general for a constrained code, an analytical expression for C_{LB} is not available and thus its computation is performed either through numerical optimization [78] or through approximation [79]. In this work we obtain analytical expressions for C_{LB} when the finite state machine in Fig. 7.1 generates the (d, ∞) code for the following channels:

1. Binary Symmetric Channel

2. Z-Channel

3. Binary Erasure Channel

7.2.1 Binary Symmetric Channel (BSC)

A BSC is a binary-input binary-output memoryless channel with an associated "crossover" probability, $Pr(1|0) = Pr(0|1)$, which we denote by q. The crossover probability represents the probability of bit error by a hard-decision information decoder at the receiver due to channel noise.

It may be tempting to interpret the transition of information bit 1 to information bit 0 due to channel noise as an energy loss, but the energy harvester at the receiver harvests energy radiated by the transmitter independent of the information decoder. Thus imposing the (d, ∞) code constraint at the transmitter helps to meet the energy requirement at the receiver even on noisy channels.

The following proposition evaluates the achievable rate given by (7.16) for a BSC using (d, ∞) constrained codes.

Proposition 8. *The lower bound for capacity on BSC with crossover probability q when the Markovian state machine in Fig. 7.1 is used to generate the (d, ∞) code is given by*

$$C_{LB} = \frac{H(p_0 + q - 2pq) - H(q)}{1 + p_0 d}, \tag{7.17}$$

where p_0 satisfies the equation

$$(1 - p_0 - q + 2pq)^{1+d-2q-dq} = q^{dq}(1-q)^{d-dq}(p_0 + q - 2p_0 q)^{1-2q-dq}. \tag{7.18}$$

Proof: The average conditional mutual information in this case is

$$I(S_2; Y_2|S_1) = H(S_2|S_1) + H(Y_2|S_1) - H(S_2, Y_2|S_1) \tag{7.19}$$

$$= \pi_d \left(H(p_0 + q - 2p_0 q) - H(q)\right) \tag{7.20}$$

$$= \frac{H(p_0 + q - 2p_0 q) - H(q)}{1 + p_0 d}, \tag{7.21}$$

where the last equality above follows from (7.8).

To solve for p_0 which maximizes $I(S_2; Y_2|S_1)$, we equate its derivative with

respect to p_0 to zero

$$\frac{\partial I}{\partial p_0} = \frac{(1+p_0 d)(1-2q)\log_2\left(\frac{1-r}{r}\right) - d\left(H(r) - H(q)\right)}{(1+p_0 d)^2} = 0, \tag{7.22}$$

where $r = p_0 + q - 2p_0 q$. Using the relation

$$H(r) = r\log_2\left(\frac{1-r}{r}\right) - \log_2(1-r), \tag{7.23}$$

equation (7.22) can be simplified to obtain (7.18). Finally, (7.17) follows from (7.16) and (7.21). ■

Remark 1. Following observations can be made on the application of Prop. 8 for the special case of $q = 0$ and $d = 0$.

- When the BSC crossover probability $q = 0$, (7.17) and (7.18) reduce to (7.9) and (7.11), respectively. In this case, C_{LB} corresponds to the (d, ∞) code capacity C_0. For this reason, C_0 is also called the noiseless capacity under the (d, ∞) RLL code constraint.

- When $d = 0$, the code becomes unconstrained and the probability of number of zeros in a codeword is represented by p_0. In this case, (7.18) reduces to $p_0 = 0.5$ and (7.17) corresponds to the unconstrained BSC capacity $1 - H(q)$.

7.2.2 Z-Channel

The Z-channel is memoryless with input alphabet $\mathcal{X} = \{0, 1\}$, output alphabet $\mathcal{Y} = \{0, 1\}$ and satisfies $Pr(0|0) = 1$. We denote the probability $Pr(0|1)$ by q.

The following proposition gives an explicit expression for the achievable rate C_{LB}, given by (7.16), for the Z-channel.

Proposition 9. *When the Markovian state machine in Fig. 7.1 is used to generate*

the (d, ∞) code for a Z-channel with $q = Pr(0|1)$, the achievable rate C_{LB} is given by

$$C_{LB} = -\log_2(1-p_0) - q\log_2\left(1 + \frac{p_0}{q(1-p_0)}\right), \tag{7.24}$$

where p_0 satisfies the equation

$$(1-p_0)^{(d+1)(1-q)} = (q(1-p_0)+p_0)^{1-(d+1)q}\ q^{(d+1)q}. \tag{7.25}$$

Proof: The average conditional mutual information in this case is

$$I(S_2; Y_2|S_1) = \frac{H\left((1-p_0)(1-q)\right) - (1-p_0)H(1-q)}{1+p_0 d}. \tag{7.26}$$

To solve for p_0 which maximizes $I(S_2; Y_2|S_1)$, we equate its derivative with respect to p_0 to zero and simplify to get (7.25). We get (7.24) by substituting the constraint (7.25) in (7.26). ■

Remark 2. Following observations can be made on the application of Prop. 9 to the special case of $q = 0$ and $d = 0$.

- When $q = 0$, (7.24) and (7.25) reduce to (7.12) and (7.11), respectively, and C_{LB} becomes equal to C_0.
- When $d = 0$, the code becomes unconstrained and the probability of number of zeros in a codeword is represented by p_0. In this case, (7.25) reduces to

$$p_0 = 1 - \frac{1}{(1-q)\left(1 + 2^{H(q)/(1-q)}\right)}, \tag{7.27}$$

which is equal to the probability for the occurrence of 0 for achieving the unconstrained capacity on a Z-channel.

7.2.3 Binary Erasure Channel

In this subsection we consider the BEC, a memoryless channel with input alphabet $\mathcal{X} = \{0, 1\}$, output alphabet $\mathcal{Y} = \{0, \epsilon, 1\}$ and transition probabilities

$$Pr(\epsilon|0) = q \;\; ; \;\; Pr(0|0) = 1 - q,$$
$$Pr(\epsilon|1) = q \;\; ; \;\; Pr(1|1) = 1 - q,$$

where q is called the erasure probability.

The following proposition evaluates the achievable rate for the BEC using (d, ∞) constrained codes.

Proposition 10. *When the Markovian state machine in Fig. 7.1 is used to generate the (d, ∞) code for a BEC with erasure probability q, the achievable rate C_{LB} is given by*

$$C_{LB} = (1 - q)C_0, \tag{7.28}$$

where C_0 is the noiseless code capacity for a (d, ∞) code.

Proof: The average conditional mutual information in this case can be shown to satisfy

$$I(S_2; Y_2|S_1) = \frac{(1 - q)H(p_0)}{1 + p_0 d}. \tag{7.29}$$

To solve for p_0 which maximizes $I(S_2; Y_2|S_1)$, we equate its derivative with respect to p_0 to zero and simplify to get (7.11). Substituting (7.11) in (7.29),

$$C_{LB} = -(1 - q)\log_2(1 - p_0) = (1 - p_0)C_0. \tag{7.30}$$

■

Remark 3. Following observations can be made from Prop. 10 and its proof.

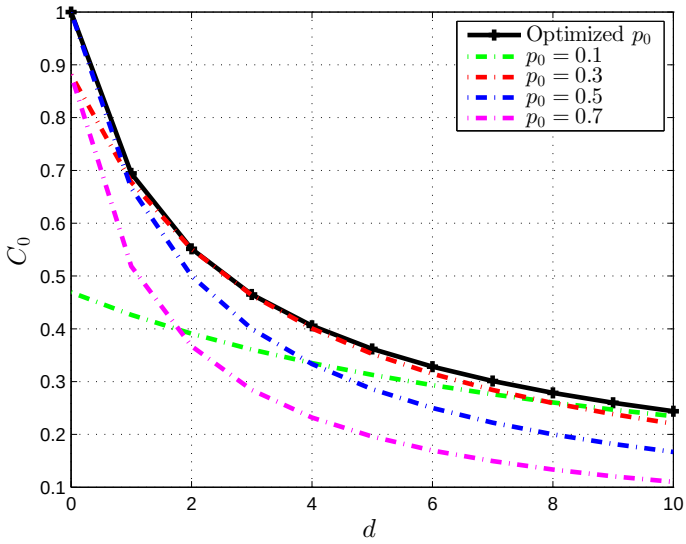

Figure 7.2: C_0 as a function of d for (d, ∞) codes.

- The optimized value of p_0 in this case which maximizes the average conditional mutual information is independent of q. For a given d, this p_0 is equal to the corresponding value for the noiseless case.

- The lower bound C_{LB} is tight for the case $q = 0$ and is equal to the noiseless code capacity C_0.

7.3 Numerical Example

The noiseless (d, ∞) code capacity C_0 is plotted in Fig. 7.2 as a function of d. The dotted curves correspond to evaluation of the entropy rate, given by (7.9), for fixed values of the state transition probability p_0. The optimized value of p_0 which achieves C_0 is obtained by solving $p_0 = (1 - p_0)^{d+1}$ and is tabulated in Table 7.1. A higher value of d implies increased transmission of ones in every codeword and hence greater opportunity for the receiver to use the received signal to fulfill its energy requirements.

Fig. 7.3 plots C_{LB}, the lower bound on the achievable rate using (d, ∞) code,

d	0	1	2	5	10
C_0	1	0.6942	0.5515	0.3620	0.2440
p_0	0.5	0.3820	0.3177	0.2219	0.1556

TABLE 7.1: TABLE OF C_0 AND OPTIMIZED p_0 AS A FUNCTION OF d

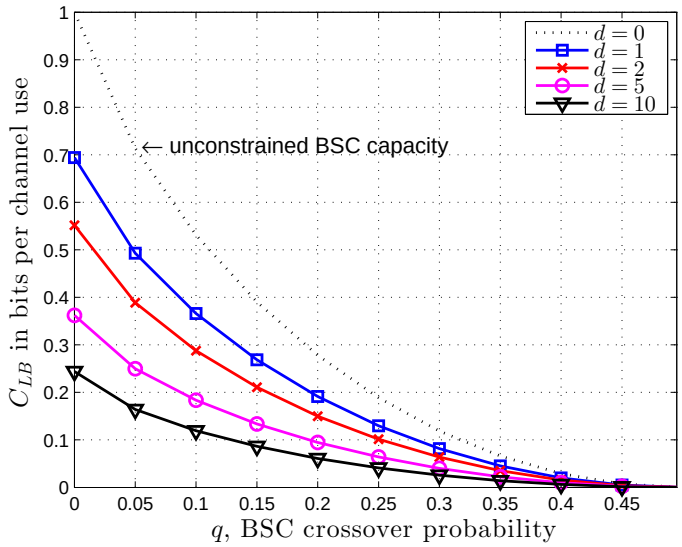

Figure 7.3: C_{LB} as a function of the BSC crossover probability

versus the BSC crossover probability q. For $q = 0$, the lower bound is tight and is equal to the noiseless code capacity C_0. The lower bound is also tight for the case $d = 0$, in which case it is equal to the unconstrained BSC capacity. The optimized value of p_0 which satisfies the constraint given by (7.18) and maximizes the average conditional mutual information $I(S_2; Y_2|S_1)$ for BSC is plotted in Fig. 7.4. We observe that the optimized value of p_0 varies both with d and with BSC crossover probability q.

The lower bound on the achievable rate using (d, ∞) code on the Z-channel is plotted in Fig. 7.5. This lower bound is tight for $d = 0$ in which case it is equal to the unconstrained Z-channel capacity. The optimized value of p_0 which satisfies (7.25) and maximizes $I(S_2; Y_2|S_1)$ is shown in Fig. 7.6. The curve corresponding to $d = 0$ in Fig. 7.6 depicts the value of probability of occurrence of zeros for

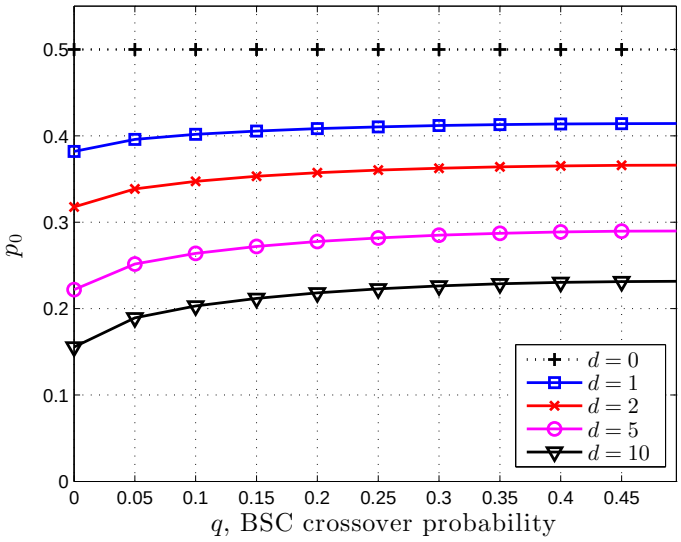

Figure 7.4: Optimized p_0 for BSC

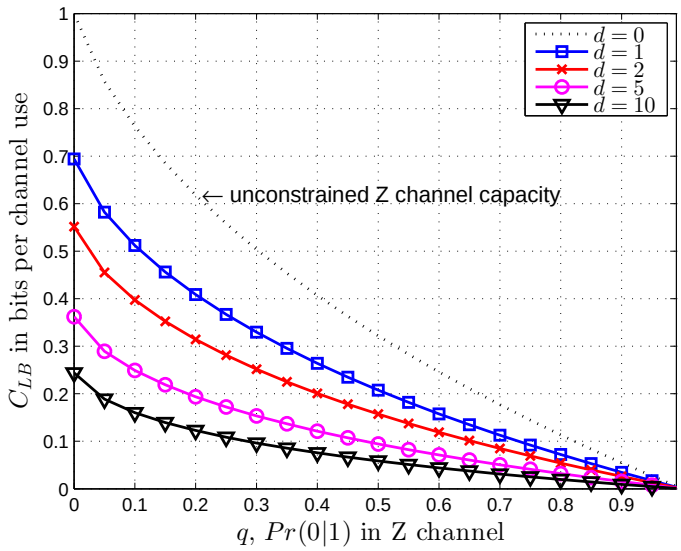

Figure 7.5: C_{LB} for Z-channel

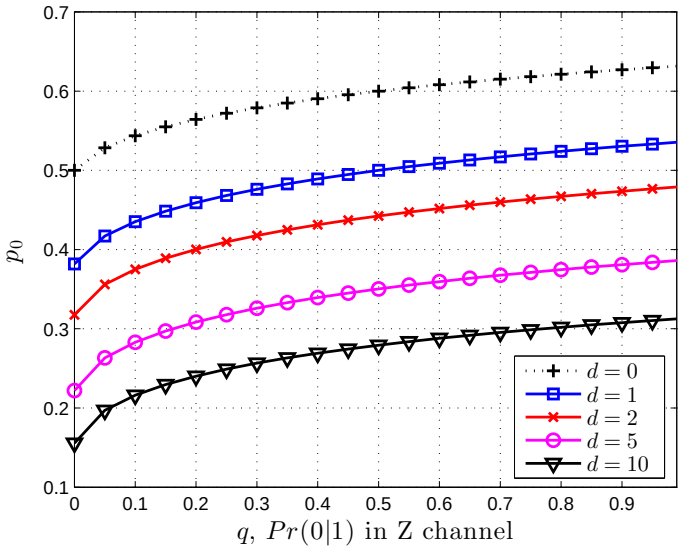

Figure 7.6: Optimized p_0 for Z-channel

achieving the unconstrained capacity for a Z-channel.

The lower bound on the achievable rate using (d, ∞) code on BEC is plotted in Fig. 7.5. The optimized value of p_0 in this case satisfies the constraint (7.11) and is independent of the probability of erasure. These optimized values of p_0 are same as in the noiseless case and are tabulated in Table 7.1 as a function of d.

7.4 Discussion

We analyzed achievable rates using Type-1 (d, ∞) run-length limited codes for noiseless and noisy channels. The impact of increasing d on the achievable information rate was presented through optimization of a single parameter. The relation which this parameter, denoting the state transition probability, satisfies in order to maximize information rate was given explicitly for different channel models.

The use of (d, ∞) codes was motivated by the codeword constraint of having at least d ones in a moving window of size $W = d + 1$ bits. The case when binary

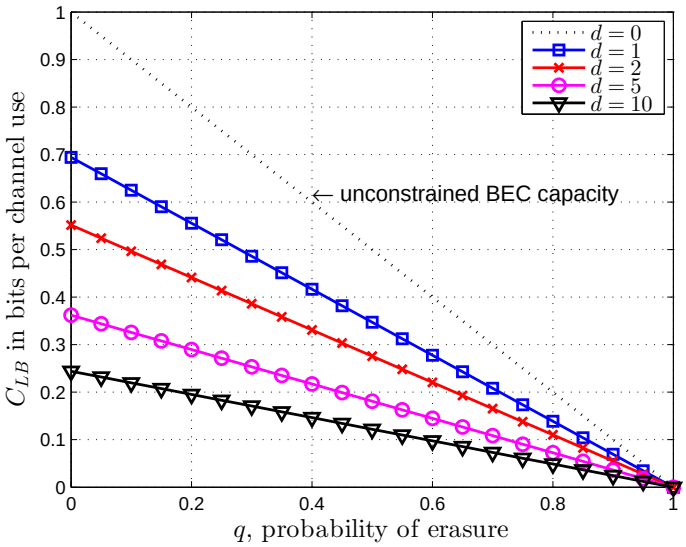

Figure 7.7: C_{LB} for BEC

codewords are constrained to have at least d ones in a window of size $W > d + 1$ opens interesting problems on quantifying achievable rates for different channels. These problems may be generalized to the study of constrained codewords over alphabets of size greater than two.

In this chapter, the sliding window constraint was imposed as a means to enable real-time energy transfer for the application of simultaneous energy and information transfer. The sliding window constraint provide an alternate codeword constraint compared to the subblock-based codeword constraints analyzed in Chapters 5 and 6.

Chapter 8

Reflections & Future Research

8.1 Reflections

In this thesis, we analyzed the performance limits for three different energy-constrained communication systems. The main contributions of this thesis and the corresponding insights are summarized below.

- In Chapter 3, we investigated the reduction in the average packet delay and the average energy usage, relative to an uncoded system, through joint use of forward error correction (FEC) codes and automatic repeat request (ARQ). For a polling based multiaccess system, we established that it is sufficient to reduce the average service time (AST) using FEC in order to achieve lower average packet delay. When the switching time from one polled node to another is negligible, we proved that the percentage reduction in average packet delay is at least as much as the percentage reduction in AST obtained using channel coding. We showed that the percentage reduction in AST due to coding translates into an equal percentage reduction in average transmit energy when the energy per coded bit is fixed. We analyzed the tradeoff between reduction in packet retransmissions and increase in transmission

time due to addition of redundancy, and quantified the achievable reduction in AST using the best possible FEC codes. It was shown that instead of trying to make the packet error probability (PEP) close to zero via FEC codes, we should aim to bring it close to 10^{-2} using as high a coding rate as possible. In particular, we showed that a coding scheme with PEP 10^{-2} and coding rate 0.9 results in lower AST compared to a code with PEP 10^{-6} and coding rate 0.89. Additionally, we showed that the average packet delay can be further reduced in certain cases by opportunistically combining and encoding several packets jointly.

- Next, in Chapter 4, we considered an energy harvesting transmitter which uses the harvested energy for transmission of data packets. We formulated a two stage *virtual queueing system* which decouples the wait stages for the energy arrival process and the service process. This decoupling of the wait stages was used to obtain closed-form expressions for the average packet delay and the probability of data packet loss due to buffer overflow for different energy arrival statistics. After obtaining these expressions for single-source energy harvesting, we extended the results to the important case of multi-source energy harvesting. We showed that the derived expressions are useful in making a judicious choose of system design parameters under given quality of service (QoS) constraints. The derived expressions were shown to be exact when the service time is negligible. Even for relative high values of the average service time, the robustness of these expressions was shown via Monte Carlo simulations.

- As a third energy-constrained communication system, we considered an energy harvesting receiver which uses the received signal to simultaneously harvest energy as well as to decode the information embedded in the sig-

nal. Here, we investigated achievable rates under three different classes of constrained codes for enabling *real-time energy transfer*. The first class of constrained codes, called the constant subblock-composition codes, were analyzed in Chapter 5 and required that each subblock within every codeword has the same composition. The second class of constrained codes, called the subblock energy-constrained codes, were proposed and examined in Chapter 6. The subblock energy-constrained codes allowed different subblocks to have different composition while ensuring sufficient energy within each subblock. In Chapter 7, we analyzed the third class of constrained codes which ensured that sufficient energy is carried within a *sliding time window*. For each of these three classes of constrained codes, we provided capacity bounds and analyzed the tradeoff between delivery of sufficient energy and achieving high information transfer rates.

8.2 Directions for Future Research

In this section, we discuss several future work directions towards extending the results presented in this thesis. A list of possible extensions, worthy of further investigation, are presented below.

1. The closed-form expression for the average packet delay at an energy harvesting transmitter was derived in Chapter 4 for a benign transmission channel where retransmissions were not required. However, in practice, the received packets may be decoded incorrectly due to factors such as noise, fading, and interference. In this scenario, an exciting future work direction is to derive the average packet delay at an energy harvesting transmitter in the presence of packet retransmissions for ensuring error-free information transfer. Further, following the approach adopted in Chapter 3, it will be interesting

to quantify the possible reduction in delay via joint use of FEC codes and ARQ at the energy harvesting transmitter.

2. The average packet delay was derived in Chapter 4 under the assumption that the data packet size is fixed and that the transmission of each data packet requires a constant amount of energy. Future work may consider the case where different data packets may have different sizes, and hence different energy requirements. This scenario corresponds to the analysis of a multi-class queueing system with a common queue, where packets of the same size belong to the same traffic class.

3. Constant subblock composition codes and subblock energy-constrained codes were introduced and analyzed in Chapter 5 and Chapter 6, respectively, for enabling real-time energy transfer. The capacity of these subblock constrained codes was derived under the assumption that the number of subblocks in a codeword could be arbitrarily large. However, large codeword lengths entail large transmission and decoding delay. Thus, in order to bound the overall delay, it will be interesting to provide capacity bounds for these subblock constrained codes when the number of subblocks in a codeword are kept fixed.

4. The *subblock* constrained codes (analyzed in Chapters 5 and 6) and the *sliding window* constrained codes (analyzed in Chapter 7) were proposed as a means for enabling real-time energy to a receiver with a small energy buffer size. An interesting future work direction for simultaneous energy and information transfer is a unified study of constrained codes which avoid energy outage at a receiver with a given small energy buffer. A related direction is the capacity comparison for the subblock constrained codes and the sliding window constrained codes for a given set of receiver energy requirements.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, no. 27, pp. 379–423, 1948.

[2] A. Ephremides and B. Hajek, "Information theory and communication networks: an unconsummated union," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2416–2434, Oct 1998.

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

[4] M. Tomamichel and V. Y. F. Tan, "A tight upper bound for the third-order asymptotics for most discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7041–7051, Nov. 2013.

[5] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1746–1755, 2004.

[6] I. E. Telatar and R. G. Gallager, "Combining queueing theory with information theory for multiaccess," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 6, pp. 963–969, Aug. 1995.

[7] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002.

[8] A. Sinha and A. Chandrakasan, "Dynamic power management in wireless sensor networks," *IEEE Des. Test. Comput.*, vol. 18, no. 2, pp. 62–74, Mar. 2001.

[9] V. Raghunathan, S. Ganeriwal, and M. Srivastava, "Emerging techniques for long lived wireless sensor networks," *IEEE Commun. Mag.*, vol. 44, no. 4, pp. 108–114, Apr. 2006.

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

[11] S. Ulukus, A. Yener, E. Erkip, O. Simeone, M. Zorzi, P. Grover, and K. Huang, "Energy harvesting wireless communications: A review of recent advances," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 360–381, Mar. 2015.

[12] J. Tan, B. T. Swapna, and N. B. Shroff, "Retransmission delays with bounded packets: Power-law body and exponential tail," *IEEE/ACM Trans. Netw.*, vol. 22, no. 1, pp. 27–38, Feb 2014.

[13] William E. Ryan and Shu Lin, *Channel Codes: Classical and Modern.* Cambridge University Press, 2009.

[14] L. R. Varshney, "Transporting information and energy simultaneously," in *Proc. 2008 IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 1612–1616.

[15] Dimitri P. Bertsekas, and Robert Gallager, *Data Networks (2nd ed.)*. Prentice Hall, 1992.

[16] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[17] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[18] E. Berlekamp, "The technology of error-correcting codes," *Proc. IEEE*, vol. 68, no. 5, pp. 564–593, May 1980.

[19] J. Proakis, *Digital Communications (4th ed.)*. McGraw-Hill, 2000.

[20] R. G. Gallager, *Principles of Digital Communication*. Cambridge University Press, 2008.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006.

[22] W. Wesley Peterson and E. J Weldon, Jr., *Error-Correcting Codes (2nd ed.)*. The M.I.T. Press, 1972.

[23] K. S. Andrews, D. Divsalar, S. Dolinar, J. Hamkins, C. R. Jones, and F. Pollara, "The development of turbo and ldpc codes for deep-space applications," *Proc. IEEE*, vol. 95, no. 11, pp. 2142–2156, Nov. 2007.

[24] A. Tanenbaum, *Computer Networks*, 4th ed. Prentice Hall Professional Technical Reference, 2002.

[25] Robert B. Cooper, *Introduction to Queueing Theory (2nd ed.)*. North Holland, 1981.

[26] V. Srivastava and M. Motani, "Coding meets queueing: Quantifying the impact of forward error correction on higher layers," in *Proc. Int. Symp. Inf. Theory Applications (ISITA)*, 2004.

[27] M. C. Vuran and I. F. Akyildiz, "Error control in wireless sensor networks: A cross layer analysis," *IEEE/ACM Trans. Netw.*, vol. 17, no. 4, pp. 1186–1199, Aug 2009.

[28] J. J. Metzner, "An improved broadcast retransmission protocol," *IEEE Trans. Commun.*, vol. 32, no. 6, pp. 679–683, Jun 1984.

[29] J. Nonnenmacher, E. W. Biersack, and D. Towsley, "Parity-based loss recovery for reliable multicast transmission," *IEEE/ACM Trans. Netw.*, vol. 6, no. 4, pp. 349–361, Aug 1998.

[30] L. Huang, S. Pawar, H. Zhang, and K. Ramchandran, "Codes can reduce queueing delay in data centers," in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 2766–2770.

[31] D. Kim, B. C. Jung, H. Lee, D. K. Sung, and H. Yoon, "Optimal modulation and coding scheme selection in cellular networks with hybrid-ARQ error control," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5195–5201, 2008.

[32] X. Wang, Q. Liu, and G. B. Giannakis, "Analyzing and optimizing adaptive modulation coding jointly with ARQ for QoS-guaranteed traffic," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 710–720, March 2007.

[33] P. Wu and N. Jindal, "Coding versus ARQ in fading channels: How reliable should the PHY be?" *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3363–3374, 2011.

[34] H.-C. Yang and S. Sasankan, "Analysis of channel-adaptive packet transmission over fading channels with transmit buffer management," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 404–413, Jan 2008.

[35] J. Ramis and G. Femenias, "Cross-layer design of adaptive multirate wireless networks using truncated HARQ," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 944–954, March 2011.

[36] R. Zhang and L. Cai, "Joint AMC and packet fragmentation for error control over fading channels," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 3070–3080, July 2010.

[37] B. Shrader and A. Ephremides, "Queueing delay analysis for multicast with random linear coding," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 421–429, Jan 2012.

[38] P. Parag, J. Chamberland, H. D. Pfister, and K. Narayanan, "Code-rate selection, queueing behavior, and the correlated erasure channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 397–407, 2013.

[39] Y. Yang, J. Tan, N. B. Shroff, and H. El Gamal, "Delay asymptotics with retransmissions and incremental redundancy codes over erasure channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1932–1944, March 2014.

[40] E. Uysal-Biyikoglu, B. Prabhakar, and A. El Gamal, "Energy-efficient packet transmission over a wireless link," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 487–499, 2002.

[41] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," vol. 4, no. 5, pp. 2349–2360, 2005.

[42] C. Joo, X. Lin, and N. B. Shroff, "Understanding the capacity region of the greedy maximal scheduling algorithm in multihop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 4, pp. 1132–1145, Aug 2009.

[43] A. Dimakis and J. Walrand, "Sufficient conditions for stability of longest-queue-first scheduling: Second order properties using fluid limits," *Adv. Appl. Probab.*, vol. 38, no. 2, pp. 505–521, 2006.

[44] M. Leconte, J. Ni, and R. Srikant, "Improved bounds on the throughput efficiency of greedy maximal scheduling in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 709–720, June 2011.

[45] J. Yang and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system," *IEEE Trans. Commun.*, vol. 60, no. 1, pp. 220–230, Jan. 2012.

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

[47] O. Ozel, J. Yang, and S. Ulukus, "Optimal broadcast scheduling for an energy harvesting rechargeable transmitter with a finite capacity battery," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2193–2203, Jun. 2012.

[48] B. Devillers and D. Gunduz, "Energy harvesting communication system with battery constraint and leakage," in *Proc. IEEE GLOBECOM Workshops*, Dec. 2011, pp. 383–388.

[49] P. Mitran, "On optimal online policies in energy harvesting systems for compound poisson energy arrivals," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 960–964.

[50] R. Vaze, R. Garg, and N. Pathak, "Dynamic power allocation for maximizing throughput in energy-harvesting communication system," *IEEE/ACM Trans. Netw.*, vol. 22, no. 5, pp. 1621–1630, Oct. 2014.

[51] R.-S. Liu, K.-W. Fan, Z. Zheng, and P. Sinha, "Perpetual and fair data collection for environmental energy harvesting sensor networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 4, pp. 947–960, Aug. 2011.

[52] L. Huang and M. J. Neely, "Utility optimal scheduling in energy-harvesting networks," *IEEE/ACM Trans. Netw.*, vol. 21, no. 4, pp. 1117–1130, Aug. 2013.

[53] D. Niyato and P. Wang, "Delay-limited communications of mobile node with wireless energy harvesting: Performance analysis and optimization," *IEEE Trans. Veh. Technol.*, vol. 63, no. 4, pp. 1870–1885, May 2014.

[54] T. Zhang, W. Chen, Z. Han, and Z. Cao, "A cross-layer perspective on energy-harvesting-aided green communications over fading channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1519–1534, Apr. 2015.

[55] R. Srivastava and C. Koksal, "Basic performance limits and tradeoffs in energy-harvesting sensor nodes with finite data and energy storage," *IEEE/ACM Trans. Netw.*, vol. 21, no. 4, pp. 1049–1062, Aug. 2013.

[56] S. Peng and C. P. Low, "Throughput optimal energy neutral management for energy harvesting wireless sensor networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC 2012)*, Apr. 2012, pp. 2347–2351.

[57] V. Joseph, V. Sharma, and U. Mukherji, "Optimal sleep-wake policies for an energy harvesting sensor node," in *Proc. IEEE Int. Conf. Commun. (ICC 2009)*, Jun. 2009, pp. 1–6.

[58] A. Seyedi and B. Sikdar, "Modeling and analysis of energy harvesting nodes in wireless sensor networks," in *Proc. 46th Annu. Allerton Conf. Commun. Control Comput.*, Sept. 2008, pp. 67–71.

[59] S. Zhang, A. Seyedi, and B. Sikdar, "An analytical approach to the design of energy harvesting wireless sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4010–4024, Aug. 2013.

[60] A. Schlichting, R. Tiwari, and E. Garcia, "Passive multi-source energy harvesting schemes," *J. Intell. Mater. Syst. Struct.*, vol. 23, no. 17, pp. 1921–1935, Nov. 2012.

[61] N. J. Guilar, R. Amirtharajah, P. J. Hurst, and S. H. Lewis, "An energy-aware multiple-input power supply with charge recovery for energy harvesting applications," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC 2009)*, Feb. 2009, pp. 298–299,299a.

[62] S. Hui, "Planar wireless charging technology for portable electronic products and Qi," *Proceedings of the IEEE*, vol. 101, no. 6, pp. 1290–1301, Jun. 2013.

[63] A. Yakovlev, S. Kim, and A. Poon, "Implantable biomedical devices: Wireless powering and communication," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 152–159, Apr. 2012.

[64] W. C. Brown, "The history of power transmission by radio waves," *IEEE Trans. Microw. Theory Tech.*, vol. 32, no. 9, pp. 1230–1242, Sept. 1984.

[65] M. Radovsky, *Alexander Popov: Inventor of the Radio.* Moscow: Foreign Language Publishing House, 1957.

[66] D. T. Emerson, "The work of Jagadis Chandra Bose: 100 years of millimeter-wave research," *IEEE Trans. Microw. Theory Tech.*, vol. 45, no. 12, pp. 2267–2273, Dec 1997.

[67] G. Marconi, "Wireless telegraphic communication," *Nobel Lectures*, Dec. 1909.

[68] D. D. Hoolihan, "The Titanic and radio frequency interference," *IEEE Electromagn. Compat. Mag.*, vol. 1, no. 2, July 2012.

[69] P. Grover and A. Sahai, "Shannon meets Tesla: Wireless information and power transfer," in *Proc. 2010 IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 2363–2367.

[70] P. Popovski, A. M. Fouladgar, and O. Simeone, "Interactive joint transfer of energy and information," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 2086–2097, May 2013.

[71] Z. Ding, C. Zhong, D. W. K. Ng, M. Peng, H. A. Suraweera, R. Schober, and H. V. Poor, "Application of smart antenna technologies in simultaneous wireless information and power transfer," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 86–93, Apr. 2015.

[72] S. Timotheou, I. Krikidis, S. Karachontzitis, and K. Berberidis, "Spatial domain simultaneous information and power transfer for MIMO channels," accepted for publication in *IEEE Trans. Wireless Commun.*, 2015.

[73] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher, "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 123–133, 2007.

[74] G. Yilmaz, O. Atasoy, and C. Dehollain, "Wireless energy and data transfer for in-vivo epileptic focus localization," *IEEE Sensors J.*, vol. 13, no. 11, pp. 4172–4179, Nov. 2013.

[75] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for digital recorders," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2260–2299, Oct. 1998.

[76] K. A. S. Immink, "Runlength-limited sequences," *Proc. IEEE*, vol. 78, no. 11, pp. 1745–1759, Nov. 1990.

[77] R. D. Cideciyan, F. Dolivo, R. Hermann, W. Hirt, and W. Schott, "A PRML system for digital magnetic recording," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 1, pp. 38–56, Jan. 1992.

[78] E. Zehavi and J. K. Wolf, "On runlength codes," *IEEE Trans. Inf. Theory*, vol. 34, no. 1, pp. 45–54, 1988.

[79] S. Shamai and Y. Kofman, "On the capacity of binary and Gaussian channels with run-length-limited inputs," *IEEE Trans. Commun.*, vol. 38, no. 5, pp. 584–594, 1990.

[80] P. Jacquet and W. Szpankowski, "Noisy constrained capacity for BSC channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5412–5423, Nov. 2010.

[81] Á. I. Barbero, E. Rosnes, G. Yang, and Ø. Ytrehus, "Constrained codes for passive RFID communication," in *Proc. 2011 Inf. Theory Appl. Workshop*, Feb. 2011.

[82] A. M. Fouladgar, O. Simeone, and E. Erkip, "Constrained codes for joint energy and information transfer," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 2121–2131, Jun. 2014.

[83] Y. M. Chee, Z. Cherif, J.-L. Danger, S. Guilley, H. M. Kiah, J.-L. Kim, P. Sole, and X. Zhang, "Multiply constant-weight codes and the reliability of loop physically unclonable functions," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7026–7034, Nov. 2014.

[84] Z. Cherif, J.-L. Danger, S. Guilley, J.-L. Kim, and P. Sole, "Multiply constant weight codes," in *Proc. 2013 IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 306–310.

[85] S. Johnson, "Upper bounds for constant weight error correcting codes," *Discrete Math.*, vol. 3, no. 1-3, pp. 109–124, 1972.

[86] E. Ordentlich and R. M. Roth, "Two-dimensional weight-constrained codes through enumeration bounds," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1292–1301, Jul. 2000.

[87] Y. M. Chee, H. M. Kiah, and P. Purkayastha, "Matrix codes and multitone frequency shift keying for power line communications," in *Proc. 2013 IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2870–2874.

[88] N. Pavlidou, A. J. H. Vinck, J. Yazdani, and B. Honary, "Power line communications: state of the art and future trends," *IEEE Commun. Mag.*, vol. 41, no. 4, pp. 34–40, Apr. 2003.

[89] W. Chu, C. J. Colbourn, and P. Dukes, "Constructions for permutation codes in powerline communications," *Des. Codes Cryptog.*, vol. 32, pp. 51–64, 2004.

[90] ——, "On constant composition codes," *Discrete Appl. Math.*, vol. 154, no. 6, pp. 912–929, Apr. 2006.

[91] Y. M. Chee, H. M. Kiah, P. Purkayastha, and C. Wang, "Importance of symbol equity in coded modulation for power line communications," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 661–665.

[92] O. Ozel and S. Ulukus, "AWGN channel under time-varying amplitude constraints with causal information at the transmitter," in *Conf. Rec. 45th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2011, pp. 373–377.

[93] ——, "Achieving AWGN capacity under stochastic energy harvesting," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6471–6483, Oct. 2012.

[94] Y. Dong and A. Ozgur, "Approximate capacity of energy harvesting communication with finite battery," in *Proc. 2014 IEEE Int. Symp. Inf. Theory*, June 2014, pp. 801–805.

[95] V. Jog and V. Anantharam, "An energy harvesting AWGN channel with a finite battery," in *Proc. 2014 IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 806–810.

[96] R. Rajesh, V. Sharma, and P. Viswanath, "Capacity of Gaussian channels with energy harvesting and processing cost," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2563–2575, May 2014.

[97] W. Mao and B. Hassibi, "On the capacity of a communication system with energy harvesting and a limited battery," in *Proc. 2013 IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 1789–1793.

[98] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 74–80, 2003.

[99] G. Dimic, N. D. Sidiropoulos, and R. Zhang, "Medium access control - physical cross-layer design," *IEEE Signal Process. Mag.*, vol. 21, no. 5, pp. 40–50, 2004.

[100] V. Srivastava and M. Motani, "Cross-layer design: A survey and the road ahead," *IEEE Commun. Mag.*, vol. 43, no. 12, pp. 112–119, 2005.

[101] J. B. B. Neto, P. F. R. Neto, and R. M. C. Andrade, "A delay-sensitive strategy for real-time monitoring in wireless sensor networks," in *Proc. IEEE Int. Conf. Telecomm. (ICT)*, 2010, pp. 814–821.

[102] N. Cai et al., "Introduction to the special issue on networking and information theory," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2285–2288, June 2006.

[103] R. G. Gallager, *Information Theory and Reliable Communication.* New York: John Wiley and Sons, Inc., 1968.

[104] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 2012.

[105] *Specification of the Bluetooth System*, Bluetooth Special Interest Group Std., Rev. 2.0 + EDR.

[106] M. van der Schaar and D. S. Turaga, "Cross-layer packetization and retransmission strategies for delay-sensitive wireless multimedia transmission," *IEEE Trans. Multimedia*, vol. 9, no. 1, pp. 185–197, 2007.

[107] S.-T. Cheng and M. Wu, "Contention-polling duality coordination function for IEEE 802.11 WLAN family," *IEEE Trans. Commun.*, vol. 57, no. 3, pp. 779–788, 2009.

[108] A. K. Jacob and L. Jacob, "A discrete time polling protocol for wireless body area network," in *Proc. 2014 IEEE Int. Adv. Computing Conf. (IACC)*, Feb 2014, pp. 294–299.

[109] A. Vardy, "Even more efficient bounded-distance decoding of the hexacode, the Golay code, and the Leech lattice," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1495–1499, 1995.

[110] A. Brouwer and T. Verhoeff, "An updated table of minimum-distance bounds for binary linear codes," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 662–677, 1993.

[111] M. Grassl. (2007) Bounds on the minimum distance of linear codes and quantum codes. Accessed on 2013-5-13. [Online]. Available: www.codetables.de

[112] G. Polytrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1284–1292, 1994.

[113] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.

[114] M. Tomamichel and V. Y. F. Tan, "A tight upper bound for the third-order asymptotics of discrete memoryless channels," 2013, arXiv:1212.3689v2 [cs.IT].

[115] A. E. Ashikhmin, A. Barg, and S. N. Litsyn, "A new upper bound on the reliability function of the Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1945–1961, Sep 2000.

[116] A. D. Wyner, "Capabilities of bounded discrepancy decoding," *Bell System Technical Journal*, vol. 44, pp. 1061–1122, 1965.

[117] J. M. Jornet, "A joint energy harvesting and consumption model for self-powered nano-devices in nanonetworks," in *Proc. IEEE Int. Conf. Commun. (ICC 2012)*, Jun. 2012, pp. 6151–6156.

[118] J. M. Jornet and I. F. Akyildiz, "Joint energy harvesting and communication analysis for perpetual wireless nanosensor networks in the terahertz band," *IEEE Trans. Nanotechnol.*, vol. 11, no. 3, pp. 570–580, May 2012.

[119] N. A. Khovanova and I. A. Khovanov, "The role of excitations statistic and nonlinearity in energy harvesting from random impulsive excitations," *Appl. Phys. Lett.*, vol. 99, no. 14, pp. 144 101 – 144 101–3, 2011.

[120] S. Zhang and A. Seyedi, "Analysis and design of energy harvesting wireless sensor networks with linear topology," in *Proc. IEEE Int. Conf. Commun. (ICC 2011)*, Jun. 2011, pp. 1–5.

[121] J. Lei, R. Yates, and L. Greenstein, "A generic model for optimizing single-hop transmission policy of replenishable sensors," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 547–551, Feb. 2009.

[122] T. Henderson. (2009, Aug.) Energy harvesting roads in Israel. Accessed on 2014-10-02. [Online]. Available: http://www.energyharvestingjournal.com/articles/energy-harvesting-roads-in-israel-00001589.asp

[123] Y. Ma and J. H. Aylor, "System lifetime optimization for heterogeneous sensor networks with a hub-spoke technology," *IEEE Trans. Mobile Comput.*, vol. 3, no. 3, pp. 286–294, Jul. 2004.

[124] M. A. Zafer and E. Modiano, "A calculus approach to minimum energy transmission policies with quality of service guarantees," vol. 1, Mar. 2005, pp. 548–559 vol. 1.

[125] Z. Liang, S. Feng, D. Zhao, and X. Shen, "Delay performance analysis for supporting real-time traffic in a cognitive radio sensor network," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 325–335, Jan. 2011.

[126] S. Xu, J. Hansen, and Z. L. Wang, "Piezoelectric-nanowire-enabled power source for driving wireless microelectronics," *Nat. Commun.*, vol. 1, no. 7, pp. 1–5, Oct. 2010.

[127] J. M. Harrison, "Assembly-like queues," *J. Appl. Prob.*, vol. 10, no. 2, pp. 354–367, Jun. 1973.

[128] E. H. Lipper and B. Sengupta, "Assembly-like queues with finite capacity: Bounds, asymptotics and approximations," *Queueing Systems*, vol. 1, no. 1, pp. 67–83, Jun. 1986.

[129] U. N. Bhat, "Finite capacity assembly like queues," *Queueing Systems*, vol. 1, no. 1, pp. 85–101, 1986.

[130] M. Manitz, "Assembly-like queues with finite capacity: Bounds, asymptotics and approximations," *Comput. Oper. Res.*, vol. 35, no. 8, pp. 2520–2536, 2008.

[131] L. R. Varshney, "On energy/information cross-layer architectures," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 1361–1365.

[132] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-energy tradeoff," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4754–4767, Nov. 2013.

[133] R. M. Fano, *Transmission of Information: A Statistical Theory of Communications*. Cambridge, MA: MIT Press, 1961.

[134] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems (2nd ed)*. Cambridge Univ. Press, 2011.

[135] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications–I," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 6, pp. 769–772, Nov. 1973.

[136] N. Chayat and S. Shamai, "Extension of an entropy property for binary input memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1077–1079, Sept. 1989.

[137] R. Ahlswede and J. Körner, "On the connection between the entropies of input and output distributions of discrete memoryless channels," in *Proc. 5th Conf. Probability Theory, Brasov 1974*. Ed. Acad. Republicii Socialiste România, 1977, pp. 13–22.

[138] S. Yekhanin, "Locally decodable codes," *Found. Trends Theor. Comput. Sci.*, vol. 7, no. 1, pp. 1–117, 2011.

[139] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. dissertation, Princeton University, Nov. 2010.

[140] P. Moulin, "The log-volume of optimal constant-composition codes for memoryless channels, within $O(1)$ bits," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 826–830.

[141] L. Weiss, "On the strong converse of the coding theorem for symmetric channels without memory," *Q. Appl. Math.*, vol. 18, no. 3, pp. 209–214, Oct. 1960.

[142] V. Strassen, "Asymptotische abschätzungen in Shannon's informationstheorie," in *Trans. 3rd Prague Conf. Inf. Theory*, Prague, 1962, pp. 689–723.

[143] A. Tandon, M. Motani, and L. R. Varshney, "Subblock-constrained codes for real-time simultaneous energy and information transfer," May 2015, arXiv:1506.00213 [cs.IT].

[144] ——, "Real-time simultaneous energy and information transfer," in *Proc. 2015 IEEE Int. Symp. Inf. Theory*, Jun. 2015, pp. 1124–1128.

[145] K. A. S. Immink, *Codes for Mass Data Storage Systems.* Shannon Foundation Publishers, The Netherlands, 1999.

[146] B. H. Marcus, R. M. Roth, and P. H. Siegel. (2001, Oct.) An introduction to coding for constrained systems. [Online]. Available: http://www.math.ubc.ca/~marcus/Handbook/index.html

[147] C. V. Freiman and A. D. Wyner, "Optimum block codes for noiseless input restricted channels," *Inf. Control*, vol. 7, no. 3, pp. 398–415, 1964.

[148] J. J. Ashley and P. H. Siegel, "A note on the Shannon capacity of run-length-limited codes," *IEEE Trans. Inf. Theory*, vol. 33, no. 4, pp. 601–605, 1987.

[149] L. R. Varshney, "Local fidelity, constrained codes, and the Meru Prastāra," *IEEE Potentials*, vol. 27, no. 2, pp. 27–32, Mar. 2008.

[150] R. B. Ash, *Information Theory.* New York: Dover Publications, 1965.

[151] M. Smieja and J. Tabor, "Entropy of the mixture of sources and entropy dimension," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2719–2728, May 2012.

[152] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, Jul. 1974.

List of Publications

Journals

(J1) A. Tandon and M. Motani, "A Cross-Layer Approach to Reducing Packet Delay in Polling Based Multiuser Systems," accepted, *IEEE Trans. Veh. Tech.*, 2016, DOI:10.1109/TVT.2016.2568578.

(J2) A. Tandon and M. Motani, "Diphase: Characterizing Packet Delay in Multi-Source Energy Harvesting Systems," accepted, *IEEE Trans. Commun.*, 2016, DOI:10.1109/TCOMM.2016.2594230.

(J3) A. Tandon, M. Motani, and L. R. Varshney, "Subblock-Constrained Codes for Real-Time Simultaneous Energy and Information," in *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4212–4227, July 2016.

Conferences

(C1) A. Tandon, M. Motani, and L. R. Varshney, "Subblock Energy-Constrained Codes for Simultaneous Energy and Information Transfer," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1969-1973.

(C2) A. Tandon, M. Motani, and L. R. Varshney, "Real-time Simultaneous Energy

and Information Transfer," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 1124-1128.

(C3) A. Tandon and M. Motani, "Has green energy arrived? Delay analysis for energy harvesting communication systems," in *Proc. IEEE Conf. Sensing, Commun. and Netw. (SECON)*, June 2014, pp. 582–590.

(C4) A. Tandon, M. Motani, and L. R. Varshney, "Constant Subblock Composition Codes for Simultaneous Energy and Information Transfer," in *Proc. IEEE SECON Workshop on Energy Harvesting Communications*, June 2014, pp. 45–50.

(C5) A. Tandon, M. Motani, and L. R. Varshney, "On Code Design for Simultaneous Energy and Information Transfer," in *Proc. 2014 Inf. Theory Appl. Workshop (ITA)*, Feb. 2014.

(C6) A. Tandon, M. Motani, and V. Srivastava, "On the Impact of Channel Coding on Average Packet Delay in a Multiuser Environment," in *Proc. IEEE Wireless Comm. Networking Conf. (WCNC)*, April 2013, pp. 499–504.

Appendix A

Proofs Related to Chapter 3

A.1 Proof of Theorem 1

Let $\overline{\tilde{T}_w}$ and $\overline{\hat{T}_w}$ denote the average waiting time for the two respective coded systems, and define

$$\begin{aligned}
\tilde{Y} &= N\lambda\overline{\tilde{T}_s^2} + (N + N\lambda\overline{\tilde{T}_s})\overline{T_v} + \sigma_v^2 N\lambda, \\
\tilde{Z} &= 2(1 - N\lambda\overline{\tilde{T}_s} - N\lambda\overline{T_v}), \\
\hat{Y} &= N\lambda\overline{\hat{T}_s^2} + (N + N\lambda\overline{\hat{T}_s})\overline{T_v} + \sigma_v^2 N\lambda, \\
\hat{Z} &= 2(1 - N\lambda\overline{\hat{T}_s} - N\lambda\overline{T_v}), \\
K &= \frac{\sigma_v^2}{2\overline{T_v}}.
\end{aligned}$$

Then $\overline{\tilde{T}_w} = K + (\tilde{Y}/\tilde{Z})$ and $\overline{\hat{T}_w} = K + (\hat{Y}/\hat{Z})$. Hence,

$$\overline{\tilde{T}_s} < \overline{\hat{T}_s},\ \overline{\tilde{T}_s^2} < \overline{\hat{T}_s^2} \implies \tilde{Y} < \hat{Y},\ \ \tilde{Z} > \hat{Z} \implies \overline{\tilde{T}_w} < \overline{\hat{T}_w}. \tag{A.1}$$

Finally, the claim in (3.7) is proved by using (A.1) and the fact that $\overline{\tilde{T}_D} = \overline{\tilde{T}_s} + \overline{\tilde{T}_w}$ and $\overline{\hat{T}_D} = \overline{\hat{T}_s} + \overline{\hat{T}_w}$. ■

A.2 Proof of Proposition 1

When $\overline{T_v} \to 0$ and $\sigma_v^2 \to 0$, then from (3.5), it follows that

$$\overline{T_w} = \frac{N\lambda\overline{T_s^2}}{2(1 - N\lambda\overline{T_s})} \tag{A.2}$$

$$= \frac{N\lambda\left(\overline{T_s}\right)^2(1+p)}{2(1 - N\lambda\overline{T_s})}. \tag{A.3}$$

Now,

$$\frac{\overline{\tilde{T}_w}}{\overline{T_w}} = \frac{\left(\overline{\tilde{T}_s}\right)^2(1+\tilde{p})}{\left(\overline{T_s}\right)^2(1+p)} \times \frac{(1 - N\lambda\overline{T_s})}{(1 - N\lambda\overline{\tilde{T}_s})}$$

$$< \frac{\left(\overline{\tilde{T}_s}\right)^2(1+\tilde{p})}{\left(\overline{T_s}\right)^2(1+p)}, \qquad (\text{since } \overline{\tilde{T}_s} < \overline{T_s}). \tag{A.4}$$

Further, from (3.11) we have $\overline{\tilde{T}_s} < \overline{T_s} \implies \tilde{p} < p$, and hence

$$\frac{\overline{\tilde{T}_w}}{\overline{T_w}} < \frac{\left(\overline{\tilde{T}_s}\right)^2}{\left(\overline{T_s}\right)^2} < \frac{\overline{\tilde{T}_s}}{\overline{T_s}}. \tag{A.5}$$

Finally, we have

$$\frac{\overline{\tilde{T}_D}}{\overline{T_D}} = \frac{\overline{\tilde{T}_s} + \overline{\tilde{T}_w}}{\overline{T_s} + \overline{T_w}} < \frac{\overline{\tilde{T}_s}}{\overline{T_s}}. \tag{A.6}$$

where the last inequality follows using the fact that if A, B, C, D are positive real numbers and $\frac{C}{D} < \frac{A}{B}$ then $\frac{A+C}{B+D} < \frac{A}{B}$. ■

A.3 Proof of Proposition 2

Let p_k and p_{mk} denote the PEP for the uncoded system when the packet lengths are k and mk, respectively. The uncoded packet of length mk may be viewed

as a packet consisting of m subblocks, each of length k. The probability that the uncoded packet of length mk is decoded correctly is equal to the probability that each of the m uncoded subblocks are decoded correctly. Thus, $(1-p_{mk}) = (1-p_k)^m$.

For a fixed k, let $\mathcal{C}^{\star}_{n,k}$ be the code which results in minimum value of $\nu(\mathcal{C}_{n,k})$ over all block codes having 2^k codewords. Thus we have $\nu^{\star}(k) = \nu\left(\mathcal{C}^{\star}_{n,k}\right)$. Then we can encode a packet of length mk as follows: (i) divide the input packet into m sub-blocks each of length k, (ii) encode each subblock using the code $\mathcal{C}^{\star}_{n,k}$, and (iii) obtain the complete codeword of length mn by concatenating the coded subblocks (each of length n) and denote the derived code by $\mathcal{C}_{mn,mk}$.

Let $\tilde{p}_k$ and $\tilde{p}_{mk}$ denote the PEP using codes $\mathcal{C}^{\star}_{n,k}$ and $\mathcal{C}_{mn,mk}$, respectively. The probability of correctly decoding a packet encoded using $\mathcal{C}_{mn,mk}$ is equal to the probability that each of the coded sub-block is decoded correctly. Thus, $(1-\tilde{p}_{mk}) = (1-\tilde{p}_k)^m$. Hence if we let $R_c = k/n \leq 1$, then for this code we have

$$\begin{aligned}
\nu(\mathcal{C}_{mn,mk}) &= \frac{mn(1-p_{mk})}{mk(1-\tilde{p}_{mk})} \\
&= \frac{n(1-p_k)^m}{k(1-\tilde{p}_k)^m} \\
&= R_c^{m-1}\left(\frac{n(1-p_k)}{k(1-\tilde{p}_k)}\right)^m \\
&= R_c^{m-1}\left(\nu\left(\mathcal{C}^{\star}_{n,k}\right)\right)^m = R_c^{m-1}\left(\nu^{\star}(k)\right)^m . \qquad \text{(A.7)}
\end{aligned}$$

The assertion in (3.19) follows because $R_c \leq 1$ and by definition of $\nu^{\star}(mk)$ we have $\nu^{\star}(mk) \leq \nu(\mathcal{C}_{mn,mk})$. ■

A.4 Proof of Proposition 5

Consider a scenario where a node has m packets waiting for transmission at time $t = 0$. Let D_m denote the mean delay for these m waiting packet, relative to $t = 0$, when each packet is encoded and transmitted independently. Let $D_{m,\text{super}}$ denote the mean delay for these m waiting packet, relative to $t = 0$, when these packets are combined into a super-packet and encoded jointly. Now the condition $D_{m,\text{super}} < D_m$ implies that the mean delay corresponding to m waiting packets is lower with opportunistic joint encoding. This condition, in turn, implies that the delay per packet, averaged over *all* data packets, will only be lower when opportunistic joint encoding is employed. We now show that the condition given by (3.41) is equivalent to $D_{m,\text{super}} < D_m$, and hence implies that opportunistic combining results in lower average packet delay.

When each packet is encoded and transmitted individually, the minimum achievable AST is equal to $T/(R_1(1-p_1))$. Since the second data packet is waiting while the first one is being transmitted, the average delay for the second packet (relative to $t = 0$) is equal to

$$\frac{T}{R_1(1-p_1)} + \frac{T}{R_1(1-p_1)} = \frac{2T}{R_1(1-p_1)}. \tag{A.8}$$

Similarly, the average delay for the i^{th} packet, for $1 \leq i \leq m$, is given by

$$\frac{(i-1)T}{R_1(1-p_1)} + \frac{T}{R_1(1-p_1)} = \frac{iT}{R_1(1-p_1)}. \tag{A.9}$$

Hence D_m can be expressed as

$$D_m = \frac{1}{m}\sum_{i=1}^{m} \frac{iT}{R_1(1-p_1)} = \frac{(m+1)T}{2R_1(1-p_1)}. \tag{A.10}$$

On the other hand, the super-packet formed by combining the m waiting packets has length equal to

$$mk - (m-1)k_O = (m-1)k_I + k, \tag{A.11}$$

as only one set of header and CRC bits are required for the super-packet. Further, since $\theta = k_I/k$, the transmission time for $(m-1)k_I + k$ bits is equal to

$$((m-1)\theta + 1)T. \tag{A.12}$$

In this case, when the super-packet is encoded jointly and transmitted, the average delay relative to $t = 0$ is same for each of the m packets and is given by

$$D_{m,\text{super}} = \frac{((m-1)\theta + 1)\,T}{R_m(1-p_m)}. \tag{A.13}$$

From (A.10) and (A.13) it follows that

$$\begin{aligned} D_{m,\text{super}} < D_m &\iff \frac{((m-1)\theta + 1)\,T}{R_m(1-p_m)} < \frac{(m+1)T}{2R_1(1-p_1)} \\ &\iff \frac{R_m(1-p_m)}{R_1(1-p_1)} > \frac{2(m-1)\theta + 2}{m+1}. \end{aligned}$$

■

Appendix B

Proofs Related to Chapter 5

B.1 Proof of Theorem 2

We will prove Theorem 2 by first proving some simple lemmas and employing Gallager's definition of a symmetric channel [103].

If π denotes any permutation on L letters with

$$\pi(x_1^L) = \pi(x_1, x_2, \ldots, x_L) \triangleq (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(L)}), \tag{B.1}$$

then for a DMC, we have

$$W^L\left(\pi(y_1^L)|\pi(x_1^L)\right) = \prod_{i=1}^{L} W\left(y_{\pi(i)}|x_{\pi(i)}\right) = \prod_{i=1}^{L} W(y_i|x_i) = W^L\left(y_1^L|x_1^L\right). \tag{B.2}$$

Let the composition of the output vector $y_1^L \in \mathcal{Y}^L$ be Q and let $\mathcal{T}_Q^L$ be the set of all output vectors of length L having composition Q. We now prove a useful lemma.

Lemma 3. *The following sets are equal*

$$\{\pi(x_1^L)|x_1^L \in \mathcal{T}_P^L\} = \mathcal{T}_P^L \tag{B.3}$$

$$\{W^L\left(\pi(y_1^L)|x_1^L\right) : x_1^L \in \mathcal{T}_P^L\} = \{W^L\left(y_1^L|x_1^L\right) : x_1^L \in \mathcal{T}_P^L\} \tag{B.4}$$

$$\{W^L\left(y_1^L|\pi(x_1^L)\right) : y_1^L \in \mathcal{T}_Q^L\} = \{W^L\left(y_1^L|x_1^L\right) : y_1^L \in \mathcal{T}_Q^L\} \tag{B.5}$$

Proof: A permutation preserves the composition of a sequence. Thus, π may be viewed as a map $\pi : \mathcal{T}_P^L \to \mathcal{T}_P^L$. This map is injective by definition of a permutation. Since the set $\mathcal{T}_P^L$ is finite, this map is also surjective and hence (B.3) follows. From (B.2) we have $W^L\left(\pi(y_1^L)|x_1^L\right) = W^L\left(y_1^L|\pi^{-1}(x_1^L)\right)$. Now (B.4) follow from (B.3). Similar to (B.3), we can show $\{\pi(y_1^L)|y_1^L \in \mathcal{T}_Q^L\} = \mathcal{T}_Q^L$ and use (B.2) to prove (B.5). ■

We recall Gallager's definition [103] of a *symmetric* DMC.

Definition 2. *A DMC is* symmetric *if the set of outputs can be partitioned into subsets in such a way that for each subset the matrix of transition probabilities (using inputs as rows and outputs of the subsets as columns) has the property that each row is a permutation of each other row and each column (if more than 1) is a permutation of each other column.*

We will show that when CSCC is employed on a DMC, the induced vector-channel is symmetric. Note that the underlying (scalar) channel can be any arbitrary DMC (not necessarily symmetric).

Lemma 4. *When CSCC with subblock length L is employed on any DMC, the induced vector-channel (obtained from L uses of the DMC) is symmetric.*

Proof: The lemma will be proved if we can partition the outputs into subsets

such that for each subset the matrix of transition probabilities has the property that each row (column) is a permutation of each other row (column).

We now show that if we partition the outputs into subsets such that each subset contains all the outputs of a given composition, then the symmetry conditions will be satisfied.

If $y_1^L \in \mathcal{T}_Q^L$ and $\tilde{y}_1^L \in \mathcal{T}_Q^L$ for a given composition Q, then since y_1^L and $\tilde{y}_1^L$ have the same composition, we have $\tilde{y}_1^L = \pi(y_1^L)$ for some permutation π. Let $\mathcal{T}_P^L$ be the input alphabet for the induced vector channel using CSCC with subblock-composition P. Then using (B.4), we note that the columns of the vector-channel transition matrix corresponding to output subset $\mathcal{T}_Q^L$ are permutations of each other. Similarly, using (B.5) we can prove that the corresponding rows are permutations of each other. ■

Theorem 15 ([103], p. 94)**.** *For a symmetric discrete memoryless channel, capacity is achieved by using the inputs with equal probability.*

Finally, Theorem 2 follows directly from Lemma 4 and Theorem 15. ■

B.2 Proof of Proposition 6

Since

$$P_{XY}(X_i = x, Y_i = y) \;=\; \Pr(X_i = x)W(y|x), \tag{B.6}$$

the claim will be proved if we show

$$\Pr(X_i = x) = N(x)/L = P(x)\ ,\ 1 \le i \le L. \tag{B.7}$$

As X_1^L is uniformly distributed over $\mathcal{T}_P^L$, the $\Pr(X_i = x)$ is equal to the ratio of the number of input vectors with x at index i to the total number of vectors in

$\mathcal{T}_P^L$. Since

$$|\mathcal{T}_P^L| = \frac{L!}{\prod_{x\in\mathcal{X}} N(x)!}, \tag{B.8}$$

and the number of sequences in $\mathcal{T}_P^L$ with x at index i is

$$\frac{(L-1)!}{(N(x)-1)!\prod_{\tilde{x}\neq x} N(\tilde{x})!}, \tag{B.9}$$

the ratio of (B.9) to (B.8) is equal to $\Pr(X_i = x) = N(x)/L$. ■

B.3 Proof of Theorem 3

When L satisfies (5.19), then $E_{max} \geq 2G$. Since $E(1) \geq G$, the energy level at the start of *every* subblock is at least G (by recursive application of Lemma 1(c)) and sufficiency follows from Lemma 1(b).

Now let $L_1 = \sum_{x\in\mathcal{X}_\triangleleft} LP(x)$, and define

$$P_1(x) = \begin{cases} \frac{P(x)}{\sum_{x\in\mathcal{X}_\triangleleft} P(x)}, & \text{if } x \in \mathcal{X}_\triangleleft \\ 0, & \text{if } x \in \mathcal{X}_\triangleright \end{cases}, \quad P_2(x) = \begin{cases} 0, & \text{if } x \in \mathcal{X}_\triangleleft \\ \frac{P(x)}{\sum_{x\in\mathcal{X}_\triangleright} P(x)}, & \text{if } x \in \mathcal{X}_\triangleright \end{cases}$$

$$S_1 = \{x_1^L | x_1^{L_1} \in \mathcal{T}_{P_1}^{L_1}, x_{L_1+1}^L \in \mathcal{T}_{P_2}^{L-L_1}\}\ ,\ S_2 = \{x_1^L | x_1^{L-L_1} \in \mathcal{T}_{P_2}^{L-L_1}, x_{L-L_1+1}^L \in \mathcal{T}_{P_1}^{L_1}\}.$$

Clearly $S_1 \subset \mathcal{T}_P^L$, $S_2 \subset \mathcal{T}_P^L$, where S_1 (resp. S_2) denotes the set of subblocks of length L with first (resp. last) L_1 input symbols belonging to $\mathcal{X}_\triangleleft$. Note that $E(1) \geq G$ is necessary to avoid outage because if $E(1) < G$, then outage results when the first subblock in a codeword belongs to S_1. To prove that (5.19) is necessary, we will show that when

$$L > \frac{E_{max}}{\sum_{x\in\mathcal{X}_\triangleleft} 2P(x)\,(B - b(x))}, \tag{B.10}$$

then CSCC codewords exist which will result in energy outage at the receiver. Here we have

$$G = \sum_{x \in \mathcal{X}_\triangleleft} LP(x)\,(B - b(x)) > \frac{E_{max}}{2}\ . \tag{B.11}$$

Let the first subblock in a given codeword belong to S_2. Since the last L_1 symbols (within the first subblock) belong to $\mathcal{X}_\triangleleft$, we have $E(L+1) = |E(L-L_1+1) - G|^+$. If there is no outage during the reception of the first subblock,

$$E(L+1) = E(L-L_1+1) - G \leq E_{max} - G < E_{max}/2, \tag{B.12}$$

where the last inequality follows from (B.11). Now let the second subblock belong to S_1. There is no energy outage during the reception of first L_1 symbols within the second subblock if and only if $E(L+1) \geq G$. However, from (B.12) and (B.11) it follows that $E(L+1) < E_{max}/2 < G$, and hence outage cannot be avoided in the second subblock. ■

B.4 Proof of Theorem 5

The strict inequality $0 < C_{CCC}(P) - C^L_{CSCC}(P)$ follows for BSC with crossover probability $0 < p_0 < 0.5$ because

$$\begin{aligned}
C^L_{CSCC}(P) &= \frac{1}{L}\left[H(Y_1^L) - H(Y_1^L|X_1^L)\right] && \text{(B.13)}\\
&= \frac{1}{L}\left[\sum_{i=1}^{L} H(Y_i|Y_1^{i-1}) - \sum_{i=1}^{L} H(Y_i|X_i)\right] && \text{(B.14)}\\
&\overset{(a)}{<} \frac{1}{L}\left[\sum_{i=1}^{L} H(Y_i) - \sum_{i=1}^{L} H(Y_i|X_i)\right] && \text{(B.15)}\\
&= C_{CCC}(P), && \text{(B.16)}
\end{aligned}$$

where $Y_1^{i-1} = Y_1 \dots Y_{i-1}$, the strict inequality (a) follows since Y_i is related to Y_1^{i-1} via X_1^{i-1} and X_i. The last equality above follows from Prop. 6 and (5.20).

For subblock-composition P with $0 < \gamma = \min(P(0), P(1)) \leq 0.5$, the output entropy on a BSC is $H(Y) = h(p_0 \star \gamma)$ and hence

$$C_{CCC}(P) = h(p_0 \star \gamma) - h(p_0). \tag{B.17}$$

For CSCC, from (5.21) and definition of α, it follows that

$$\frac{1}{L} H(X_1^L) = H(P) - r(L, P) \tag{B.18}$$

$$= h(\gamma) - r(L, P) = h(\alpha). \tag{B.19}$$

Now using (B.19) and applying Mrs. Gerber's Lemma [135],

$$\frac{1}{L} H(Y_1^L) \geq h(p_0 \star \alpha), \tag{B.20}$$

and hence

$$C_{CSCC}^L(P) = \frac{1}{L} \left[H(Y_1^L) - \sum_{i=1}^{L} H(Y_i | X_i) \right] \tag{B.21}$$

$$\geq h(p_0 \star \alpha) - h(p_0). \tag{B.22}$$

Using (B.17) and (B.22) we have

$$C_{CCC}(P) - C_{CSCC}^L(P) \leq h(p_0 \star \gamma) - h(p_0 \star \alpha) \tag{B.23}$$

We only have to show that $h(p_0 \star \gamma) - h(p_0 \star \alpha) < r(L, P)$ for completing the proof. Towards this we first observe that when $0 < x \leq 0.5$ and $0 < p_0 < 0.5$,

then $p_0 \star x \geq x$. Next we note that the derivative of $h(x)$ satisfies

$$h'(x) = \log \frac{1-x}{x}, \tag{B.24}$$

and hence $h'(x)$ is a monotonically decreasing function of x for $0 < x \leq 0.5$.

Since $h(\alpha) = h(\gamma) - r(L, P)$, we have

$$\begin{aligned} h(p_0 \star \gamma) - h(p_0 \star \alpha) < r(L,P) \iff \\ h(p_0 \star \gamma) - h(\gamma) < h(p_0 \star \alpha) - h(\alpha). \end{aligned} \tag{B.25}$$

If we define $f(x) = h(p_0 \star x) - h(x)$ for $0 \leq x \leq 0.5$, then we have

$$f'(x) = (1 - 2p_0)h'(p_0 \star x) - h'(x). \tag{B.26}$$

Hence $f'(x) < 0$ for $0 < x \leq 0.5$ since $h'(x)$ is monotonically decreasing in x and $p_0 \star x \geq x$. This in turn implies that $f(x)$ is a strictly monotonically decreasing function of x. It follows that $f(\gamma) < f(\alpha)$ (since $\alpha < \gamma$) and (B.25) is satisfied. ■

B.5 Proof of Theorem 6

For a BEC with erasure probability ϵ, and $\gamma = P(0)$,

$$C_{CCC}(P) = (1 - \epsilon)h(\gamma). \tag{B.27}$$

If α is chosen such that $h(\alpha) = h(\gamma) - r(L, P)$, then from (5.21) it follows that $H(X_1^L)/L = h(\alpha)$. Now applying an extension of MGL for binary input symmetric

channels [136], we get $H(Y_1^L)/L \geq (1-\epsilon)h(\alpha) + h(\epsilon)$. Thus,

$$C_{CSCC}^{L}(P) = \frac{1}{L}\left[H(Y_1^L) - \sum_{i=1}^{L} H(Y_i|X_i)\right] \geq (1-\epsilon)h(\alpha), \tag{B.28}$$

and (5.35) follows from (B.27), (B.28), and definition of α. ■

B.6 Proof of Theorem 7

For a Z-channel with $\gamma = \Pr(X=1),\ p_0 = \Pr(1 \to 0)$,

$$C_{CCC}(P) = h\left(\gamma(1-p_0)\right) - \gamma h(p_0). \tag{B.29}$$

If $0 \leq \alpha \leq 0.5$ is chosen such that $h(\alpha) = h(\gamma) - r(L,P)$, then from (5.21) it follows that $H(X_1^L)/L = h(\alpha)$. Now applying the extension of MGL for memoryless asymmetric binary-input, binary-output channels [137], we get $H(Y_1^L)/L \geq h\left(\alpha(1-p_0)\right)$. Thus,

$$C_{CSCC}^{L}(P) = h\left(\alpha(1-p_0)\right) - \gamma h(p_0), \tag{B.30}$$

and (5.36) follows from (B.29) and (B.30). ■

B.7 Proof of Theorem 8

As discussed in Sec. 5.2.2, L uses of the channel for CSCC using subblock-composition P, subblock length L, and codeword length n on DMC W, may be viewed as a single use of a *vector channel* having input alphabet $\mathcal{T}_P^L$, output alphabet $\mathcal{Y}^L$, length-L product channel W^L, and codeword length equal to n/L

super-letters. By Theorem 2, the input distribution is

$$P^L(x_1^L) = \begin{cases} \frac{1}{|\mathcal{T}_P^L|}, & \text{if } x_1^L \in \mathcal{T}_P^L \\ 0, & \text{otherwise} \end{cases} \tag{B.31}$$

Since rate R for the scalar channel corresponds to rate LR for the vector channel, the random coding bound on the maximum probability of error for the induced *vector channel* is [134]

$$P_e \leq \exp[-\frac{n}{L}\left(E_r(LR, P^L, W^L) - L\delta\right)], \tag{B.32}$$

where the exponent $E_r(LR, P^L, W^L)$ is given by

$$E_r(LR, P^L, W^L) = \min_{V^L} D(V^L||W^L|P^L) + [I(P^L, V^L) - LR]^+, \tag{B.33}$$

V^L ranging over all channels $V^L : \mathcal{T}_P^L \to \mathcal{Y}^L$. The first term on the right side of (B.33) is

$$\begin{aligned} D(V^L||W^L|P^L) &= \sum_{x_1^L, y_1^L} P^L(x_1^L) V^L(y_1^L|x_1^L) \log \frac{V^L(y_1^L|x_1^L)}{W^L(y_1^L|x_1^L)} && \text{(B.34)} \\ &= -H_{P^L \times V^L}(Y_1^L|X_1^L) + \mathbb{E}_{P^L \times V^L}\left[\log \frac{1}{W^L(Y_1^L|X_1^L)}\right] && \text{(B.35)} \\ &\overset{(a)}{\geq} \sum_{i=1}^{L} \left(-H_{P \times V_i}(Y|X) + \mathbb{E}_{P \times V_i}\left[\log \frac{1}{W(Y_i|X_i)}\right]\right) && \text{(B.36)} \\ &= \sum_{i=1}^{L} D(V_i||W|P), && \text{(B.37)} \end{aligned}$$

where V_i is the marginal distribution of V^L corresponding to the ith symbol, and (a) follows from (B.7), the memoryless property of W, and the fact conditioning

only reduces the entropy. The term $I(P^L, V^L)$ in (B.33) is

$$I(P^L, V^L) = H_{P^L}(X_1^L) - H_{P^L \times V^L}(X_1^L | Y_1^L) \tag{B.38}$$

$$\overset{(b)}{\geq} LH_P(X) - Lr(L,P) - \sum_{i=1}^{L} H_{P \times V_i}(X_i | Y_i) \tag{B.39}$$

$$= \sum_{i=1}^{L} \left(H_P(X_i) - H_{P \times V_i}(X_i | Y_i) - r(L,P) \right) \tag{B.40}$$

$$= \sum_{i=1}^{L} \left(I(P, V_i) - r(L,P) \right), \tag{B.41}$$

where (b) follows using (5.21), (B.7), and the fact conditioning only reduces entropy. Now combining (B.33), (B.37), and (B.41), we get

$$E_r(LR, P^L, W^L) \geq \min_{V^L} \sum_{i=1}^{L} \left(D(V_i || W | P) + [I(P, V_i) - r(L,P) - R]^+ \right) \tag{B.42}$$

$$\geq \sum_{i=1}^{L} \min_{V_i} \left(D(V_i || W | P) + [I(P, V_i) - r(L,P) - R]^+ \right) \tag{B.43}$$

$$= L\, E_r(R + r(L,P), P, W), \tag{B.44}$$

where the last equality follows from the definition of the random coding exponent (5.38). The theorem is proved by applying (B.44) in (B.32). ■

Appendix C

Proofs Related to Chapter 6

C.1 Proof of Theorem 9

Let X_i denote the transmitted symbol in the ith channel use, $I = \{1, 2, \ldots, L\}$, and $I_< = \{i \in I | X_i \in \mathcal{X}_\triangleleft\}$. The level in the energy buffer decreases during the ith channel use if and only if $i \in I_<$, and the corresponding decrease in energy level is $B - b(X_i)$. Let the composition of the first subblock be $P_j \in \Gamma_B^L$. Then the sum of energy decrements over the reception of the first subblock is $\sum_{i \in I_<} (B - b(X_i)) = LS_j$, and there will be no outage during the reception of the first subblock because $E(1) \geq LS_{max}$.

Next we show that $E(L+1) \geq LS_{max}$, and the theorem will be proved by recursive application of the above argument. We note that if there is no energy overflow during the reception of the first subblock, then $E(L+1) \geq E(1) \geq LS_{max}$ as $P_j \in \Gamma_B^L$. In case there is overflow in the ith channel use for any $i \in I$, then $E(i+1) = E_{max} \geq 2LS_{max}$ and hence $E(L+1) \geq E_{max} - LS_j \geq LS_{max}$. ■

C.2 Proof of Theorem 10

When x_1^L and $\tilde{x}_1^L$ have the same composition, then the rows $W^L(\cdot|x_1^L)$ and $W^L(\cdot|\tilde{x}_1^L)$, of the induced vector channel W^L, are permutations of each other. The proof is then complete by observing that if the Blahut-Arimoto algorithm [16] (for the finding the capacity-achieving input distribution for the DMC W^L (6.6) having input alphabet $\mathcal{A}$ and output alphabet $\mathcal{Y}^L$), is initialized with a uniform distribution over $\mathcal{A}$, then the probabilities corresponding to x_1^L and $\tilde{x}_1^L$ remain equal after every iteration of the algorithm. (A brief overview of the Blahut-Arimoto algorithm is presented in Sec. 2.2.2). ■

C.3 Proof of Theorem 11

When $P^*_{X_1^L}$ is a capacity-achieving distribution of Thm. 10, we have, from (6.8) and the grouping axiom [150], p. 8,

$$H_{P^*_{X_1^L}}(X_1^L) = \sum_{j=1}^{J} c_j \log |\mathcal{T}_{P_j}^L| + \sum_{j=1}^{J} -c_j \log c_j. \tag{C.1}$$

$$\stackrel{(i)}{=} LH(\tilde{P}) - L\tilde{r}, \tag{C.2}$$

where (i) follows from (6.10). Now,

$$\begin{aligned} C_{SECC}^L(B) &\stackrel{(a)}{=} \left(H_{P^*_{X_1^L}}(X_1^L) - H_{P^*_{X_1^L}\times W^L}(X_1^L|Y_1^L) \right) / L \\ &\stackrel{(b)}{\geq} \left(LH(\tilde{P}) - L\tilde{r} - \sum_{i=1}^{L} H_{\tilde{P}\times W}(X_i|Y_i) \right) / L \\ &= I(\tilde{P}, W) - \tilde{r}, \end{aligned}$$

where (a) follows from (6.7), and (b) follows from (C.2) and the fact conditioning only reduces entropy.

Towards proving the SECC capacity upper bound, note that since constant composition codes achieve capacity on a DMC [134], the SECC capacity is achieved by codewords having empirical distribution $P^*_{X_1^L}$ with respect to alphabet $\mathcal{A}$. These codewords, when viewed as a sequence of symbols from $\mathcal{X}$ have empirical distribution $\tilde{P}$. Thus, $C^L_{SECC}(B) \leq I(\tilde{P}, W)$ since achievable rates using codewords having constant composition $\tilde{P}$ is upper bounded by $I(\tilde{P}, W)$. ■

C.4 Proof of Theorem 13

$E_r(LR, P^*_{X_1^L}, W^L)$ can be expressed as [134]

$$\min_{V^L} \left(D(V^L||W^L|P^*_{X_1^L}) + [I(P^*_{X_1^L}, V^L) - LR]^+ \right), \tag{C.3}$$

V^L ranging over all channels $V^L : \mathcal{A} \to \mathcal{Y}^L$. The term $D(V^L||W^L|P^*_{X_1^L})$ in (C.3), by definition, is

$$\begin{aligned}
&= \sum_{x_1^L \in \mathcal{A}} P^*_{X_1^L}(x_1^L) \sum_{y_1^L \in \mathcal{Y}^L} V^L(y_1^L|x_1^L) \log \frac{V^L(y_1^L|x_1^L)}{W^L(y_1^L|x_1^L)} \\
&= -H_{P^*_{X_1^L} \times V^L}(Y_1^L|X_1^L) + \mathbb{E}_{P^*_{X_1^L} \times V^L} \left[\log \frac{1}{W^L(Y_1^L|X_1^L)} \right] \\
&\overset{(a)}{\geq} \sum_{i=1}^{L} \left(-H_{\tilde{P} \times V_i}(Y|X) + \mathbb{E}_{\tilde{P} \times V_i} \left[\log \frac{1}{W(Y_i|X_i)} \right] \right) \\
&= \sum_{i=1}^{L} D(V_i||W|\tilde{P}),
\end{aligned} \tag{C.4}$$

where V_i is the marginal distribution of V^L corresponding to the ith symbol, and (a) follows from (6.6), (6.9), and the fact conditioning only reduces the entropy.

The term $I(P^*_{X_1^L}, V^L)$ in (C.3) satisfies

$$\begin{aligned} I(P^*_{X_1^L}, V^L) &= H_{P^*_{X_1^L}}(X_1^L) - H_{P^*_{X_1^L}\times V^L}(X_1^L|Y_1^L) \\ &\overset{(b)}{\geq} LH_{\tilde{P}}(X) - L\tilde{r} - \sum_{i=1}^{L} H_{\tilde{P}\times V_i}(X_i|Y_i) \\ &= \sum_{i=1}^{L} I(\tilde{P}, V_i) - \ L\tilde{r}, \end{aligned} \tag{C.5}$$

where (b) follows using (C.2) and the fact conditioning only reduces entropy. Let $\widehat{V}^L$ denote that V^L which achieves the minimum in (C.3). Then $\widehat{V}^L$ has the following form [143], [152]

$$\widehat{V}^L(y_1^L|x_1^L) = \frac{W^L(y_1^L|x_1^L)^{1-s} Q_{Y_1^L}(y_1^L)^s}{\sum_{\tilde{y}_1^L \in \mathcal{Y}^L} W^L(\tilde{y}_1^L|x_1^L)^{1-s} Q_{Y_1^L}(\tilde{y}_1^L)^s}, \tag{C.6}$$

where $Q_{Y_1^L}$ satisfies the set of simultaneous equations

$$Q_{Y_1^L}(y_1^L) = \sum_{x_1^L \in \mathcal{A}} \frac{P^*_{X_1^L}(x_1^L) W^L(y_1^L|x_1^L)^{1-s} Q_{Y_1^L}(y_1^L)^s}{\sum_{\tilde{y}_1^L \in \mathcal{Y}^L} W^L(\tilde{y}_1^L|x_1^L)^{1-s} Q_{Y_1^L}(\tilde{y}_1^L)^s}, \tag{C.7}$$

and $s \in [0,1]$ is chosen as a function of rate R. Now, if π is an arbitrary permutation on L letters, then $W^L(\pi(y_1^L)|\pi(x_1^L)) = W^L(y_1^L|x_1^L)$ (see (6.6)) and $P^*_{X_1^L}(\pi(x_1^L)) = P^*_{X_1^L}(x_1^L)$ (see Thm. 10). Thus, from (C.7) it follows that $Q_{Y_1^L}(\pi(y_1^L)) = Q_{Y_1^L}(y_1^L)$ and hence $\widehat{V}^L(\pi(y_1^L)|\pi(x_1^L)) = \widehat{V}^L(y_1^L|x_1^L)$. In particular, $\widehat{V}^L(\pi(y_1^L)|\pi(x_1^L)) = \widehat{V}^L(y_1^L|x_1^L)$ when π corresponds to a transposition which interchanges the symbols at the first and the ith index, and hence $\widehat{V}_i$, the marginal distribution of $\widehat{V}^L$ corresponding to the ith symbol, has distribution identical to $\widehat{V}_1$, where $1 < i \leq L$, .

Because $\widehat{V}^L$ achieves the minimum in (C.3), we have

$$\begin{aligned} E_r(LR, P^*_{X_1^L}, W^L) &= D(\widehat{V}^L||W^L|P^*_{X_1^L}) + [I(P^*_{X_1^L}, \widehat{V}^L) - LR]^+ \\ &\overset{(c)}{\geq} L\, D(\widehat{V}_1||W|\tilde{P}) + [L\, I(\tilde{P}, \widehat{V}_1) - L\tilde{r} - LR]^+ \\ &= L\left(D(\widehat{V}_1||W|\tilde{P}) + [I(\tilde{P}, \widehat{V}_1) - (\tilde{r} + R)]^+\right) \\ &\geq L \min_V \left(D(V||W|\tilde{P}) + [I(\tilde{P}, V) - (\tilde{r} + R)]^+\right) \\ &= L\, E_r(R + \tilde{r}, \tilde{P}, W), \end{aligned}$$

where (c) follows from (C.4), (C.5), and the fact $\widehat{V}_i, 1 \leq i \leq L$ are identically distributed. ■