UNIVERSITY OF CALIFORNIA SAN DIEGO

Energy Optimization for Hybrid ARQ

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering
(Communication Theory and Systems)

by

Bentao Zhang

Committee in charge:

Professor Pamela Cosman, Chair
Professor William S. Hodgkiss
Professor Tara Javidi
Professor Laurence B. Milstein
Professor Steven Swanson

2020

The dissertation of Bentao Zhang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I am greatly indebted to my advisors, Professor Pamela Cosman and Professor Laurence B. Milstein, for their guidance and support throughout my PhD study. I thank them for for their effort and patience so that I could overcome the difficulties in my research.

I would like to thank Professor Tara Javidi, Professor William S. Hodgkiss, and Professor Steven Swanson for serving on my committee and their feedback during my PhD study.

Chapter 2, in part, is a reprint of the material as it appears in the papers: B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization For Hybrid ARQ With Turbo Coding: Rate Adaptation and Allocation," submitted to *IEEE Transactions on Vehicular Technology*, and B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization for Incremental Redundancy Hybrid-ARQ," In *2019 53rd Asilomar Conference on Signals, Systems, and Computers. IEEE, 2019*. The dissertation author was the primary investigator and author of these papers.

Chapter 3, in part, is a reprint of the material as it appears in the papers: B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization for Wireless Video Transmission Employing Hybrid ARQ," *IEEE Transactions on Vehicular Technology (2019)*, and B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization for Hybrid-ARQ and AMC" In *2017 51st Asilomar Conference on Signals, Systems, and Computers. IEEE, 2017*. The dissertation author was the primary investigator and author of these papers.

VITA

2014	Bachelor of Electronics Engineering, Tsinghua University, China
2017	Master of Science, University of California San Diego
2020	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization for Wireless Video Transmission Employing Hybrid ARQ," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5606–5617, June 2019.

B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization For Hybrid ARQ With Turbo Coding: Rate Adaptation and Allocation," submitted to *IEEE Transactions on Vehicular Technology*.

B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization for Hybrid-ARQ and AMC" in *2017 51st Asilomar Conference on Signals, Systems, and Computers. IEEE*, pp. 1796-1800.

B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization for Incremental Redundancy Hybrid-ARQ," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers. IEEE*.

ABSTRACT OF THE DISSERTATION

Energy Optimization for Hybrid ARQ

by

Bentao Zhang

Doctor of Philosophy in Electrical Engineering
(Communication Theory and Systems)

University of California San Diego, 2020

Professor Pamela Cosman, Chair

Hybrid automatic repeat request (HARQ) [1] plays an important role in providing reliable and efficient data transmission. In wireless communications, the wireless channel may vary fast, due to the mobility of the transmitter/receiver and the channel. Forward error correction (FEC) and automatic repeat request (ARQ) are two basic techniques to control errors. FEC employs error correction coding, by adding parity bits to the information bits, to combat channel errors. ARQ allows the receiver to request a retransmission of the packet when an error is detected in the received packet. HARQ gives protection to the wireless transmission by combining FEC and ARQ. In typical HARQ systems, redundancy is added to the information bits, and a retransmission is

performed until either the packet is successfully decoded, or a maximum number of transmissions is reached.

The motivation to optimize the energy consumption of HARQ is the high energy consumption of wireless communications on mobile devices. Wireless devices usually have a limited battery life, and wireless communications consume the majority of the battery energy of mobile devices. One example is that 3G and Wifi units consume more than 50% of the energy for some smart phones [2]. Another example is that battery depletion has been identified as one of the primary factors that limit the lifetime of wireless sensor networks [3].

Previous works on HARQ mainly use information-theoretic approach, which assumes that the number of bits in each transmission round is sufficiently large. This assumption does not necessarily hold for actual codes with finite length. Therefore, in this dissertation, we consider HARQ with actual codes. We use turbo-coded HARQ, since turbo codes are well-known capacity-approaching codes [4] and widely used in standards such as 3GPP Long-Term Evolution (LTE) [5]. We study the energy optimization for HARQ in two scenarios: the energy optimization for incremental redundancy (IR) HARQ, and the energy optimization for HARQ in wireless video transmission. For IR HARQ, each retransmission contains additional parity bits beyond those of the previous transmissions. For the first scenario, we consider different cases of channel state information (CSI) at the transmitter: the transmitter has no knowledge of any CSI, or knows the CSI in previous transmission rounds through a perfect feedback channel, or knows both current and previous CSI. The transmitter decides the forward error correction code rate based on the CSI it has. We minimize the energy consumption of turbo-coded HARQ, subject to a packet loss rate constraint. Numerical results show that the energy consumption of HARQ decreases when more CSI information is available at the transmitter. We also compare IR combining with both Chase combining and the system without combining, and IR combining yields the least energy consumption. For the second scenario, we formulate the problem as maximizing the video quality, subject to a constraint on the wireless transmission energy consumption. We consider multiple

parameters in multiple layers in a wireless video transmission system: transmit power, alphabet size, FEC code rate, maximum number of transmissions and unequal video data importance. An analytical framework is proposed to include these parameters, which allows us to divide this problem into two sub-problems: data transmission and unequal error protection (UEP) for video content. The problem is tackled by solving the two sub-problems, which are done by exhaustive search and convex optimization, respectively. Simulations of different videos show that the proposed scheme outperforms methods using conventional data transmission and/or unequal error protection. For example, in the low SNR region, there is a total gain of 4.8 to 5.6dB on the peak signal-to-noise ratio of the received video compared to video transmission using conventional HARQ without any video UEP.

Chapter 1

Introduction

1.1 Hybrid Automatic Repeat Request

Hybrid automatic repeat request (HARQ) is a hybrid scheme combining forward error correction (FEC) and automatic repeat request (ARQ). FEC uses error correction coding to protect a transmission by adding redundant bits. ARQ retransmits a packet when the previous attempt failed. A cyclic redundancy check is usually used to detect the error in a packet. In an HARQ system, upon receiving a packet, the receiver decodes the packet, and it may or may not use previously received packets. If the decoding is successful, the receiver sends an acknowledgement (ACK) to the transmitter. If the decoding is unsuccessful, the receiver sends a negative acknowledgement (NACK) to the transmitter, and then the transmitter will perform a retransmission of the packet. The retransmission might be identical to the previous transmission, or might contain additional information about the packet. The process continues until either the decoding is successful at the receiver or a maximum number of transmissions is reached. Pure FEC may introduce unnecessary redundancy, and pure ARQ may require many retransmissions due to heavy losses for each single transmission. The authors in [6, 7] suggest that HARQ outperforms pure FEC and pure ARQ for wireless transmission. HARQ is suited for

error-sensitive and delay-tolerant services, and it has become a standard and fundamental tool in modern cellular wireless communications such as 3GPP LTE [8].

1.1.1 Classification of HARQ

There are three types of HARQ based on the combining scheme: no combining, Chase combining (CC) and incremental redundancy (IR) combining.

- No combining: Each transmission round is self-decodable and decoded independently. This means the receiver does not use the packets that were received in previous rounds.

- CC HARQ: All transmission rounds are identical, i.e., the same FEC code is used for each transmission and the packet in each transmission contains the same bits [9, 10]. Each transmission contains both information bits and parity bits, and is self-decodable. Since each retransmission is identical to the first transmission, CC combining is also called repetition diversity HARQ. The received packets are decoded together at the receiver through maximum ratio combining (MRC) to improve the quality of the decoding. The MRC method is used because it is superior to other combining techniques such as selective combining and equal-gain combining for independent Gaussian channels.

- IR HARQ: Each retransmission contains additional parity bits beyond those of the previous transmissions. This means in the first transmission, information bits and parity bits are transmitted. If the first transmission fails, only additional parity bits are transmitted in the second transmission. The receiver combines the bits received in the first transmission and the second transmission for the decoding. Note that only the first transmission is self-decodable since it contains information bits, whereas the retransmissions are not self-decodable since they contain only parity bits.

IR HARQ is more spectrally efficient than CC HARQ due to a larger coding gain. However, IR HARQ is more complex and requires more hardware resources than CC HARQ

from an implementation perspective [11]. For IR HARQ, the receiver needs to keep records of all received packets and requires a larger buffer. The received packets are combined during the decoding of a subsequent packet. But for CC HARQ, the received packets can be combined with MRC at each transmission round, so the receiver only needs a buffer with length of a single transmission. For no combining and CC HARQ, any error correction code can be used for FEC. Representative examples are convolutional codes, turbo codes and low-density parity-check (LDPC) codes. However, for IR HARQ, only codes that are decodable with only a fraction of the bits can be used, such as turbo and LDPC codes.

In the literature, the HARQ schemes are also divided into three categories: type-I, type-II and type-III.

- Type-I: This is HARQ without combining: the receiver discards the received packet if the decoding is unsuccessful, and it requests another retransmission until either the packet is successfully decoded or a maximum number of transmissions is reached.

- Type-II: The receiver combines all the received packets for decoding the overall packet. But the retransmissions are not self-decodable. IR HARQ is a type-II HARQ, since the retransmissions contain only addition parity bits but no information bits.

- Type-III: The receiver combines all the received packets for decoding the overall packet. The retransmissions are self-decodable. CC HARQ is a type-III HARQ, since all the transmissions contain information bits and are self-decodable.

1.1.2 Analysis Methods

The existing works on HARQ mainly use an information-theoretic approach [12–20]. The information-theoretic approach is based on [21], where the assumption is that the number of bits in each transmission round is sufficiently large. Under this assumption, for IR HARQ, the decoding is successful if the average accumulated mutual information at the receiver is larger

than the overall transmission rate. In the case of CC HARQ, the decoding is successful if the accumulated SNR is larger than an SNR threshold. However, this assumption does not necessarily hold for actual codes with finite length. Since the assumption does not hold for actual codes which have finite length, we consider HARQ with actual codes, such as turbo codes, in this dissertation. In Chapter 2, we consider turbo-coded IR HARQ. In Chapter 3, we apply turbo-coded HARQ to wireless video transmission. Note that some studies [22, 23] also evaluate the performance of HARQ using an actual code such as a convolutional code. However, no combining was considered.

1.1.3 Link Adaptation

Since the wireless channel varies with time and frequency, link adaptation can be used to improve the quality of wireless communications in a single transmission. When the channel can be estimated, the channel state information (CSI) can be estimated at the receiver and sent back to the transmitter. The transmitter can use this information to determine various parameters. A power adaptation transmission scheme was proposed in [24]. Later, other transmission parameters were investigated in the adaptation schemes. For example, symbol transmission rate [25], FEC coding rate [26] and alphabet size [27] were considered to adapt according to CSI. Link adaptation is helpful to energy efficiency for wireless communications. When the channel condition is good, the transmitter leverages it by using a low power consumption or high FEC code rate. When the channel condition is bad, the transmitter uses a high power consumption or low FEC code rate. The transmitter chooses the optimal parameters based on the CSI in order to achieve satisfactory performance with constrained energy consumption. In Chapter 3, we investigate how multiple parameters (power consumption, alphabet size, FEC code rate and maximum number of transmissions) can be adapted to CSI.

Link adaptation is helpful to HARQ not only because it allows the transmitter to adapt parameters based on the CSI in the current transmission as in the systems without HARQ, but also

because the CSI in the previous transmissions could help in some HARQ systems. In the systems without HARQ, previous CSI is useful only when there is correlation between the previous and current channel conditions. Otherwise, previous CSI does not give any information about the current transmission. But in CC and IR HARQ systems, since the received packets in previous transmission rounds are still used when decoding the current packet, the previous CSI gives information about how close we were to a successful decoding. This is true even when the channels are not correlated for the previous and current transmissions. If the previous CSI was good but the decoding failed, a small amount of resources, e.g., low power or a smaller number of additional bits, is required in the current transmission. If the previous transmission was in a deep fade, the previous transmission does not help the decoding in the current transmission. But the knowledge that the previous CSI was bad still helps the transmitter to allocate resources more correctly, e.g., high power consumption or a large number of additional bits, to provide the required reliability. Consider the extreme case where the previous CSI was so bad that only noise was received in the previous transmission. In this case, the transmitter knows that it should act as if the previous transmission did not occur because the received packet in the previous transmission will be of no use. In Chapter 2, different models of CSI are studied: no CSI available at the transmitter, only previous CSI available at the transmitter, and both current and previous CSI available at the transmitter.

1.1.4 Unequal Energy Allocation

Conventional HARQ systems allocate equal amounts of energy in different transmission rounds of the same packet [22]. Here, energy can be affected by different transmission parameters, such as power consumption, FEC code rate and alphabet size. For example, a lower FEC code rate leads to a longer transmission duration and thus larger energy consumption. Similar reasoning applies for alphabet size: smaller alphabet size leads to a longer transmission duration and thus larger energy consumption. Therefore, a transmission consumes more energy if it uses high power

consumption, low FEC code rate and small alphabet size. In this dissertation, we show that a key factor to improving the performance of HARQ systems is unequal resource allocation. Indeed, early transmissions should consume less energy and later transmissions should consume more energy. If the system is lucky in the cheap early transmissions, the later expensive transmissions are avoided and energy is saved. Otherwise, later transmissions should use more energy to meet the reliability requirement of the system. Different systems allow different parameters to vary among the transmissions. For example, some systems fix the FEC code rate and alphabet size, and only vary power consumption [13, 14]. Some papers studied unequal energy allocation in a heuristic way [28, 29]. Some studies analyzed the optimal energy allocation for different transmission rounds using the information-theoretic approach [12–20]. In this dissertation, we propose an analytical framework to study unequal energy allocation for turbo-coded HARQ. In Chapter 2, we study turbo-coded IR HARQ, where the FEC code rate is allowed to change. In Chapter 3, we study turbo-coded HARQ without combining, but multiple parameters, such as power consumption, FEC code rate, alphabet size and maximum number of transmissions, are allowed to change.

1.2 Video Transmission With HARQ

Mobile video transmission is an important topic in wireless communications because of the large data rate compared to other data formats, such as text and voice. According to Cisco's forecast [30], mobile video will increase 9-fold between 2017 and 2022, accounting for 79 percent of total mobile data traffic by 2022. The large volume of wireless video traffic also challenges the energy efficiency of mobile devices, such as mobile phones and the sensors in wireless sensor networks. Therefore, energy efficient wireless video transmission schemes are critical to prolong the lifetime of mobile devices.

1.2.1 Video Encoding Structure

One of the reasons that video transmission is different from plain data transmission is the unequal importance of the video data. This is due to the video encoding/decoding structure. In the H.264 standard, the encoder can select between intra-coding and inter-coding for block-shaped regions of each picture. Intra-coding uses spatial prediction modes to reduce spatial redundancy in a single source picture. Inter-coding uses inter-prediction of each block region from some previous decoded pictures. The frames in a video sequence can be divided into the following three types:

- I picture or I frame: Intra-coded which is independent of all other frames.
- P picture or P frame: Contains inter-coding difference information relative to previously encoded pictures.
- B picture or B frame: Contains inter-coding difference information relative to previously encoded pictures from two directions.

Fig. 1.1 is an example of an IBBP encoding structure. The I frame is used to predict the P frame, and both of them are used to predict the B frames in between. A group of pictures (GOP) is a collection of successive frames in a video and specifies the order in which the frames are encoded. Since frames or regions in the frames can be lost during transmission, the lost regions impair the video quality at the receiver. Moreover, this error may propagate and affect the image quality in other frames. For example, if there is error during the transmission of an I frame, the error propagates to P and B frames. On the other hand, if there is error during the transmission of a B frame, the error stays in this frame and does not propagate, because no other frame is predicted using B frames. Therefore, errors in I frames are more influential to the decoded video quality than errors in B frames. This motivates unequal error protection (UEP) for the frames in a GOP. Since the resources such as energy and bandwidth are limited, we give more protection (or more

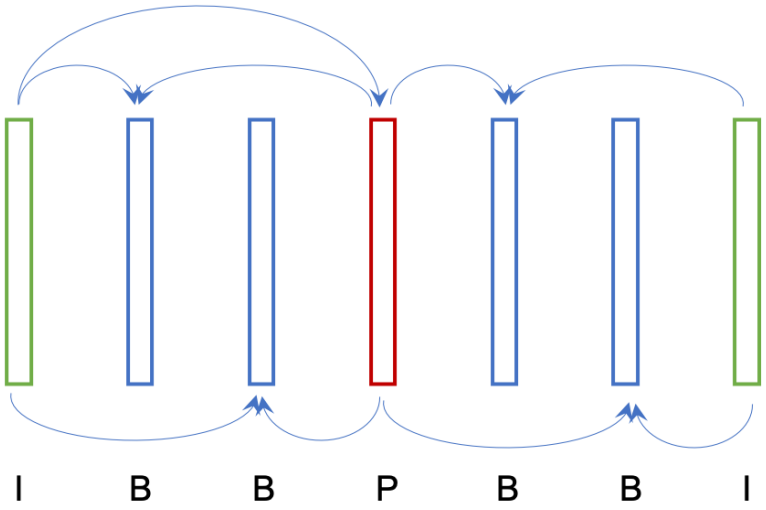

Figure 1.1: A typical illustration of IBBP encoding structure.

resources) to I frames so that errors happen less frequently in I frames, and give less protection (or less resources) to B frames, so that the overall decoded video quality is improved.

1.2.2 Combining HARQ with UEP

The existing schemes on video transmission employing HARQ can be categorized into two groups. For the first group, multiple transmissions of the same video content are treated equally, and the authors leverage the unequal importance of video contents which arises from the coding/decoding structure [46], as shown in Fig. 1.1. For the second group, the multiple transmissions of the same video content have unequal importance, but there lacks an analytical framework to provide the optimal strategy in each transmission. Most studies proposed heuristic energy allocation among different transmission rounds of the same video content, which were not optimal. Therefore, in Chapter 3 of this dissertation, we jointly consider the energy allocation among different transmission rounds of the same video content and UEP for different frames. We divide the video transmission problem into two sub-problems: the first sub-problem is the data transmission with turbo-coded HARQ, and the second sub-problem is the optimal UEP for

different frames.

1.3 Dissertation Overview

This dissertation is organized as follows.

In Chapter 2, we IR HARQ over independent block-fading channels with turbo coding. We consider different cases of CSI at the transmitter: the transmitter has no knowledge of any CSI, or knows the CSI in previous transmission rounds through a perfect feedback channel, or knows both current and previous CSI. The transmitter decides the forward error correction code rate based on the CSI it has. We minimize the energy consumption of turbo-coded HARQ, subject to a packet loss rate constraint. We also compare IR combining with both Chase combining and the system without combining.

In Chapter 3, we investigate energy-optimized wireless video transmission employing HARQ. We formulate the problem as maximizing the video quality, subject to a constraint on the wireless transmission energy consumption. We consider multiple parameters in multiple layers in a wireless video transmission system: transmit power, alphabet size, FEC code rate, maximum number of transmissions, and unequal video data importance. An analytical framework is proposed to include these parameters, which allows us to divide this problem into two sub-problems: data transmission and unequal error protection. The problem is tackled by solving the two sub-problems, which are done by exhaustive search and convex optimization, respectively.

In Chapter 4, we draw conclusions of the dissertation and list possible future work.

Chapter 2

Energy Optimization For IR HARQ With Turbo Coding

HARQ is a fundamental and standard technology for modern communications systems. IR HARQ is an important category of HARQ because of its advanced combining technique. However, the previous works on IR HARQ mainly used an information-theoretic approach, which assumes the number of bits in each transmission is infinite. In this chapter, we consider IR HARQ over independent block-fading channels with turbo coding. We consider different cases of CSI at the transmitter: the transmitter either has no knowledge of any CSI, or knows the CSI in previous transmission rounds through a perfect feedback channel, or knows both current and previous CSI. The transmitter decides the forward error correction code rate based on the CSI it has. We minimize the energy consumption of turbo-coded HARQ, subject to a packet loss rate constraint. We also compare IR combining with both Chase combining and the system without combining.

2.1 Introduction

For IR HARQ, the code rate in each transmission can be varied. There are two kinds of IR HARQ code rate selection algorithms [20]: rate allocation and rate adaptation. For rate allocation, the code rate in each transmission round is predetermined. For rate adaptation, the code rate in each transmission round is determined by the previous (and current) CSI. Because the previous transmitted packets are also used in the decoding process, the previous CSI provides information about how many additional bits are needed in the current transmission.

In [22], the authors studied the performance of HARQ with convolutional codes. However, no combining technique was used and each transmission round was identical. In [23], power and rate adaptation were presented for HARQ with MQAM, but no combining was considered. The authors in [12] considered the combination of adaptive modulation and coding and HARQ, using an information-theoretic approach. The state of the convolutional decoder was used to determine the optimal code rate for HARQ in [31]. In [13–15], the optimal power assignment across the transmission rounds was investigated for CC HARQ, under different channel models. In [16, 17], the optimal power assignment for IR HARQ was derived. In [18], the optimal rate in different transmission rounds for IR HARQ was studied. The authors in [19] generalized the power allocation and adaptation problem for CC and IR HARQ, and built a framework for close-form solutions. In [32], the authors built a framework to analytically express the throughput of HARQ systems. The performance of HARQ with imperfect feedback was studied in [33]. The influence of time correlations of wireless channels in different HARQ transmission rounds and signaling overhead were considered in [34]. The authors in [35] showed that correlated channels may yield higher throughput than independent channels for HARQ.

We compare different models of CSI availability at the transmitter: the transmitter either has no knowledge of any CSI, or knows the CSI in previous transmission rounds through a perfect feedback channel, or knows both the CSI in the current transmission round and the previous CSI.

The theoretical analyses focus on the first two models. The scheme without CSI is called rate allocation because the FEC code rates are predetermined. The schemes with CSI are called rate adaptation because the FEC code rates depend on CSI. We investigate the optimal strategy in each transmission round of the IR HARQ: which FEC code rate should the transmitter use for the transmission, based on the available CSI the transmitter has. The optimization problem is to minimize the energy consumption of HARQ, subject to a packet loss rate (PLR) constraint. A packet loss can happen either when the maximum number of retransmissions is reached or the transmitter decides to discard the packet. The PLR is defined as the probability that a packet is not successfully decoded by the receiver after retransmissions.

The rest of this chapter is organized as follows. In Section 2.2, we introduce the system model. In Section 2.3, we formulate and solve the problem. In Section 2.4, we show the numerical results. Conclusions are drawn in Section 2.5.

2.2 System Model

Suppose we have an (N_t, N_b) block turbo code, which is called the mother code. Every N_b information bits are encoded into N_t bits, and this codeword is transmitted with IR HARQ. We set $N_b = 256$ in this chapter. Consider the transmission process for every N_b information bits. Let N_i be the number of bits transmitted in the i-th transmission for the N_b information bits, where i = 1, 2, ..., K, and K is the maximum number of transmissions. We have $\sum_{i=1}^{K} N_i \leq N_t$. We will discuss how N_i is determined later in this section. In the first transmission, the transmitter punctures the mother code and transmits N_1 bits, including N_b information bits and $N_1 - N_b$ parity bits. The receiver decodes with N_1 received bits, and sends an ACK/NACK back to the transmitter through a perfect feedback channel based on the decoding result. If the transmitter receives a NACK, the next N_2 bits in the mother code are transmitted in the second transmission. The receiver decodes with the received $N_1 + N_2$ bits. This process continues until the packet

is successfully decoded or the maximum number of transmissions K is reached. We consider different cases of CSI at the transmitter which correspond to different assumptions about delay and complexity.

(1) The transmitter does not have knowledge of the CSI: The number of bits in the i-th transmission N_i is predetermined and is a function of only the transmission round i. This assumption is as in [14, 36]. This corresponds to the case where the complexity associated with sending the CSI or using it at the transmitter is not considered acceptable.

(2) The transmitter has knowledge of the CSI in the previous transmission rounds: We assume the receiver can perfectly estimate the channel state and send this information back to the transmitter along with the NACK through a perfect feedback channel when a transmission fails, as in [19, 20]. The transmitter has only the CSI corresponding to the first $i-1$ transmissions at the time of the i-th transmission. Therefore, N_1 is predetermined, and N_i depends on the CSI in the first $i-1$ transmissions and on $N_1, ..., N_{i-1}$. This corresponds to the case where the communication delay is too large for the transmitter to have CSI for the current packet.

(3) The transmitter has knowledge of both current and previous CSI: We assume the receiver can perfectly estimate the channel state and send this information back to the transmitter along with the NACK through a perfect and delay-free feedback channel when a transmission fails, as in [15, 17, 22, 37]. Therefore, N_i depends on the CSI in the first i transmissions and $N_1, ..., N_{i-1}$. This corresponds to the case where the communication delay is negligible and where the complexity associated with sending and using current CSI is considered acceptable.

In cases (2) and (3), if the CSI was/is too bad, the transmitter is allowed to abandon the opportunity to transmit the packet in order to save energy, since the probability that the packet will be transmitted successfully is not high enough. In other words, N_i is allowed to be zero. We define a packet loss to be the event that the packet is not successfully decoded after K transmission opportunities, including the ones that the transmitter chooses to abandon.

The code rate after the i-th transmission is

$$r_i = \frac{N_b}{\sum_{j=1}^{i} N_j}. \tag{2.1}$$

We have $r_j \leq r_i$ when $j > i$, since $N_{i+1}, N_{i+2}, ..., N_j$ are non-negative. For simplicity, we only allow r_i to be chosen from the following code rate set: $\{r\} = \{r^{(1)} = 1/5, r^{(2)} = 1/3, r^{(3)} = 2/5, r^{(4)} = 1/2, r^{(5)} = 2/3\}$. Once r_i is chosen, then N_i is determined by

$$N_i = \begin{cases} \frac{N_b}{r_1}, & i = 1 \\ \frac{N_b}{r_i} - N_1 - ... - N_{i-1}, & 2 \leq i \leq K \end{cases}. \tag{2.2}$$

In other words, N_i should be chosen such that

$$N_1 + ... + N_i = \frac{N_b}{r_i} \in \mathbb{Z} \tag{2.3}$$

where $r_i \in \{r\}$. A special case occurs when the transmitter abandons all the opportunities to transmit a packet. In that case, $\sum_{j=1}^{i} N_j = 0$, and we define $r_i = 0$.

We assume the transmission duration of a packet (excluding retransmissions) is much smaller than the channel coherence time so that the channel is constant during this time period. In the LTE standard for example, the time slot duration is 0.5 millisecond (ms) [38]. In [39], the authors showed that in a system with a carrier frequency of 2.5 GHz and a receiver moving with speeds of 2 km/h, 45 km/h, and 100 km/h, the coherence times are 200 ms, 10 ms, and 4 ms, respectively. Thus, the transmission duration is much smaller than channel coherence time for a wide range of receiver speed. We further assume the different transmission rounds corresponding to the same mother code experience independent fading, as in [14, 20, 38, 40]. Constant power S_0 and BPSK with symbol duration T_s are used. The noise power spectral density is N_0. Let the channel gain in the i-th transmission be γ_i, and γ_i's are i.i.d. Rayleigh distributed. The received

signal-to-noise ratio (SNR) in the i-th transmission is $\Gamma_i = \frac{S_0 T_s}{N_0}\gamma_i^2$. The pdf of Γ_i, $f_{\Gamma_i}(\Gamma_i)$, is exponential and the joint pdf is $f_{\Gamma_1,\Gamma_2,...,\Gamma_j}(\Gamma_1,\Gamma_2,...,\Gamma_j) = \prod_{i=1}^{j} f_{\Gamma_i}(\Gamma_i)$, where $1 \leq j \leq K$.

2.3 Problem Formulation

2.3.1 Optimization Problem

The optimization problem is

$$\begin{aligned} \min \quad & \overline{\mathcal{E}} \\ \text{s.t.} \quad & \overline{P_L} \leq P_{const} \\ & \frac{N_b}{N_1 + ... + N_i} \in \{r\} \\ \text{variables:} \quad & N_i,\ i = 1, 2, ..., K \end{aligned} \tag{2.4}$$

where $\overline{\mathcal{E}}$ is the average overall energy consumption of a packet, $\overline{P_L}$ is the average overall PLR, and K is the maximum number of transmissions. Here, the term "overall" refers to all transmissions of a packet. In other words, $\overline{\mathcal{E}}$ is the sum of energy consumption in all transmissions until the packet is transmitted successfully or the maximum number of transmissions K is reached, and $\overline{P_L}$ is the probability that a packet cannot be successfully transmitted after K transmissions. Equation (2.4) is the optimization problem for all the CSI availability models in Sections 2.3.2 to 2.3.4, although the method to solve the problem is different. For simplicity, we let the maximum number of transmissions K be 2 in this section. It can be extended to arbitrary K.

2.3.2 Without CSI

The transmitter only receives ACK or NACK without CSI. Since incremental redundancy is used, when the first transmission fails, the transmitter may choose to transmit additional parity bits after receiving the NACK from the receiver. The transmitter does not have any CSI, so the

optimal FEC code rates in both transmissions are predetermined, instead of evaluated at the time of transmission. This means N_2 does not depend on N_1 or Γ_1. The average overall energy consumption $\overline{\mathcal{E}}$ is

$$\overline{\mathcal{E}} = S_0 T_s \left(N_1 + \int_0^\infty N_2 P(e_1; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right), \tag{2.5}$$

where the first term in the parentheses is the *constant* number of bits in the first transmission, and the second term in the parentheses is the *average* number of bits in the second transmission since the second transmission may or may not happen. The term $P(e_1; N_1|\Gamma_1)$ is the conditional PER in the first transmission, conditioned on Γ_1. We use $P(e_1; N_1|\Gamma_1)$ and $P(e_1; r_1|\Gamma_1)$ interchangeably. The average overall PLR is

$$\overline{P_L} = \int_0^\infty \int_0^\infty P(e_1, e_2; N_1, N_2|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 d\Gamma_2, \tag{2.6}$$

where e_i is the event that the i-th transmission fails and $P(e_1, e_2; N_1, N_2|\Gamma_1, \Gamma_2)$ is the conditional probability that both transmissions fail, conditioned on Γ_1 and Γ_2. We use $P(e_1, e_2; N_1, N_2|\Gamma_1, \Gamma_2)$ and $P(e_1, e_2; r_1, r_2|\Gamma_1, \Gamma_2)$ interchangeably. In the Appendix, we show that since incremental redundancy is used, the probability that both transmissions fail can be approximated by the probability that the second transmission fails, regardless of the result in the first transmission. The assumption for this approximation is that the probability that the second transmission fails, if the first one was successful, is very small compared to the probability that both transmissions fail. The intuition for this assumption is that if the first transmission was successfully decoded, the realization of the channel, i.e., the combination of the channel gain and noise, in the second transmission has to be exceedingly bad to make the second decoding fail, which is of small probability. Then,

$$\overline{P_L} \approx \int_0^\infty \int_0^\infty P(e_2; N_1, N_2|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 d\Gamma_2, \tag{2.7}$$

where $P(e_2; N_1, N_2|\Gamma_1, \Gamma_2)$ is the conditional error probability of the second transmission, conditioned on Γ_1 and Γ_2. Because of incremental redundancy, the second decoding is performed for a codeword with length $N_1 + N_2$, and with channel state Γ_1 for the first N_1 bits and Γ_2 for the next N_2 bits. We give an analytical expression to approximate $P(e; N_1, N_2|\Gamma_1, \Gamma_2)$ in the Appendix.

The problem is as follows:

$$\begin{aligned} \text{min} \quad & S_0 T_s \left(N_1 + \int_0^\infty N_2 P(e; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right) \\ \text{s.t.} \quad & \int_0^\infty \int_0^\infty P(e_2; N_1, N_2|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 d\Gamma_2 \leq P_{const} \\ & \frac{N_b}{N_1 + N_2} \in \{r\} \\ \text{variables:} \quad & N_1, N_2 \end{aligned} \tag{2.8}$$

Next we want to find the analytical expression for $P(e_1; N_1|\Gamma_1)$ and $P(e_2; N_1, N_2|\Gamma_1, \Gamma_2)$ to solve Equation (2.8). As introduced in [22] and used in [41],

$$P(e_1; N_1|\Gamma_1) \approx \min(1, a_1 e^{-b_1 \Gamma_1}), \tag{2.9}$$

and the parameters a_1 and b_1 are obtained through curve fitting.

Since N_2 does not depend on N_1, we can use an exhaustive search to find the solution to the problem. We try all possible (N_1, N_2) to find the one that satisfies the constraint in Equation (2.8) with the minimum energy consumption.

2.3.3 With Previous CSI

As no CSI is available for the first transmission, N_1 is predetermined as in Section 2.3.2. Since the bits transmitted in the first transmission are also used in the decoding of the second transmission, N_2 should be a function of N_1 and Γ_1, i.e., N_2 can be written as $N_2(N_1, \Gamma_1)$. When the first transmission fails, the channel state Γ_1 and number of bits N_1 can provide information

about how many extra bits are needed in the second transmission. For example, if the channel was bad and N_1 is small, then it is likely that many extra bits are needed for the second transmission to be successful, although the transmitter knows nothing about the CSI in the second transmission; if the channel was good and N_1 is larger, then it is likely that a small N_2 would be sufficient in the second transmission. Let t_1 be an index that says which of the available FEC code rates is chosen for the first packet. Then $N_1 = N_b/r_1 = N_b/r^{(t_1)}$, where $2 \leq t_1 \leq |\{r\}|$ and $|\cdot|$ is the cardinality of a set. The reason that t_1 is constrained to be greater than or equal to two is that $t_1 = 1$ means the strongest FEC code rate $r^{(1)}$ is used for the first transmission, in which case there is no possibility of sending additional incremental bits in the second transmission, even if the first transmission fails. Since there would be no possibility of a second transmission, even if the first one fails, the maximum number of transmissions would be one, instead of two. Similar to Section 2.3.2, this problem can be solved exhaustively for each N_1. For a fixed t_1 or N_1, the number of bits in the second transmission $N_2(N_1, \Gamma_1)$ is determined by the SNR boundaries $\Gamma^{1,1}, \Gamma^{1,2}..., \Gamma^{1,t_1-1}$ as follows

$$N_2(N_1, \Gamma_1) = \begin{cases} 0 \quad \text{when } 0 \leq \Gamma_1 < \Gamma^{1,1} \\ N_b\left(\frac{1}{r_2} - \frac{1}{r_1}\right) = N_b\left(\frac{1}{r^{(t_2)}} - \frac{1}{r^{(t_1)}}\right) \quad \text{when } \Gamma^{1,t_2} \leq \Gamma_1 < \Gamma^{1,t_2+1} \end{cases} \tag{2.10}$$

where $1 \leq t_2 \leq t_1 - 1$, and $\Gamma^{1,t_1} = \infty$. When $0 \leq \Gamma_1 < \Gamma^{1,1}$, the transmitter discards the packet in the second transmission to save energy because the deep fade in the first transmission decreases the probability of a successful decoding in the second transmission. As Γ_1 increases,

N_2 decreases. The average energy consumption $\overline{\mathcal{E}}$ is

$$\begin{aligned}\overline{\mathcal{E}} &= S_0 T_s \left(N_1 + \int_{\Gamma^{1,1}}^{\infty} N_2(N_1, \Gamma_1) P(e_1; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right) \\ &= S_0 T_s \left(N_1 + \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} N_2(N_1, \Gamma_1) P(e_1; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right) \\ &= S_0 T_s N_b \left(\frac{1}{r^{(t_1)}} + \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \left(\frac{1}{r^{(i)}} - \frac{1}{r^{(t_1)}} \right) P(e_1; r^{(t_1)}|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right)\end{aligned} . \tag{2.11}$$

The overall PLR is

$$\begin{aligned}\overline{P_L} \approx & \int_0^{\Gamma_1^{(1)}} P(e_1; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \\ & + \int_{\Gamma_1^{(1)}}^{\infty} \int_0^{\infty} P(e_2; N_1, N_2(N_1, \Gamma_1)|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_2 d\Gamma_1 \\ = & \int_0^{\Gamma_1^{(1)}} P(e_1; r^{(t_1)}|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \\ & + \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \int_0^{\infty} P(e_2; r^{(t_1)}, r^{(i)}|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_2 d\Gamma_1\end{aligned} , \tag{2.12}$$

where the first term corresponds to the situation when the first transmission fails and then the packet is discarded by the transmitter because of the deep fade, and the second term corresponds to the case when both transmissions fail.

We use the Lagrangian multiplier method. The Lagrangian function is

$$
\begin{aligned}
L =& \overline{\mathcal{E}} + \lambda(\overline{P_L} - P_{const}) \\
=& S_0 T_s N_b \left(\frac{1}{r^{(t_1)}} + \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \left(\frac{1}{r^{(i)}} - \frac{1}{r^{(t_1)}} \right) P(e_1; r^{(t_1)} | \Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right) \\
& + \lambda \int_0^{\Gamma_1^{(1)}} P(e_1; r^{(t_1)} | \Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 + \\
& \lambda \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \int_0^{\infty} P(e_2; r^{(t_1)}, r^{(i)} | \Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_2 d\Gamma_1 \\
& - \lambda \overline{P}_{const}
\end{aligned}
\quad , \tag{2.13}
$$

which is obtained by plugging in Equations (2.11) and (2.12). We can get the optimal SNR boundaries by setting $\frac{\partial L}{\partial \lambda} = 0$ and $\frac{\partial L}{\partial \Gamma(1,j)} = 0$ for $j = 1, ..., t_1 - 1$, where

$$
\frac{\partial L}{\partial \Gamma_1^{(j)}} = h(j) + g(j), \tag{2.14}
$$

$$
\begin{aligned}
h(j) =& \frac{\partial \{ S_0 T_s N_b \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \left(\frac{1}{r^{(i)}} - \frac{1}{r^{(t_1)}} \right) P(e_1; r^{(t_1)} | \Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \}}{\partial \Gamma_1^{(j)}} \\
=& \underbrace{S_0 T_s N_b \left(\frac{1}{r^{(j-1)}} - \frac{1}{r^{(t_1)}} \right) P(e_1; r^{(t_1)} | \Gamma^{1,j}) f_{\Gamma_1}(\Gamma^{1,j})}_{w.r.t.\ i=j-1} \\
& - \underbrace{S_0 T_s N_b \left(\frac{1}{r^{(j)}} - \frac{1}{r^{(t_1)}} \right) P(e_1; r^{(t_1)} | \Gamma^{1,j}) f_{\Gamma_1}(\Gamma^{1,j})}_{w.r.t.\ i=j} \\
=& S_0 T_s N_b \left(\frac{1}{r^{(j-1)}} - \frac{1}{r^{(j)}} \right) P(e_1; r^{(t_1)} | \Gamma^{1,j}) f_{\Gamma_1}(\Gamma^{1,j})
\end{aligned}
\quad , \tag{2.15}
$$

and

$$
\begin{aligned}
g(j) =& \lambda \frac{\partial \int_0^{\Gamma_1^{(1)}} P(e_1; r^{(t_1)}|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1}{\partial \Gamma_1^{(j)}} \\
&+ \lambda \frac{\partial \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \int_0^{\infty} P(e_2; r^{(t_1)}, r^{(i)}|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_2 d\Gamma_1}{\partial \Gamma_1^{(j)}} \\
=& \begin{cases}
\lambda(P(e_1; r^{(t_1)}|\Gamma^{1,1}) f_{\Gamma_1}(\Gamma^{1,1}) - \int_0^{\infty} P(e_2; r^{(t_1)}, r^{(1)}|\Gamma^{1,1}, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma^{1,1}, \Gamma_2) d\Gamma_2), \ j = 1 \\
\lambda(\int_0^{\infty} P(e_2; r^{(t_1)}, r^{(j-1)}|\Gamma^{1,j-1}, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma^{1,j-1}, \Gamma_2) d\Gamma_2 - \\
\int_0^{\infty} P(e_2; r^{(t_1)}, r^{(j)}|\Gamma^{1,j}, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma^{1,j}, \Gamma_2) d\Gamma_2), \quad 2 \le j \le t_1 - 1
\end{cases}
\end{aligned}
\tag{2.16}
$$

where $h(j)$ is the derivative of the second line in Equation (2.13), and $g(j)$ is the derivative of the third and fourth lines in Equation (2.13). Equation (2.13) is solved for each N_1, and then we find the minimum energy for all possible values of N_1, where there are $|\{r\}| - 1$ possibilities.

2.3.4 With Current and Previous CSI

Both current and previous CSI are available at the transmitter, and they are used to determine the FEC code rate. This means $N_1(\Gamma_1)$ depends on Γ_1 and $N_2(N_1, \Gamma_1, \Gamma_2)$ depends on N_1, Γ_1 and Γ_2. Since $N_1(\Gamma_1)$ is just a function of Γ_1, $N_2(\Gamma_1, \Gamma_2)$ is fully determined by Γ_1 and Γ_2. For simplicity of notation, we discuss the FEC code rates $r_1(\Gamma_1)$ and $r_2(\Gamma_1, \Gamma_2)$. The relationship between r_i and N_i is in Equation (2.1).

The code rate in the first transmission is determined by the SNR boundaries $\Gamma^{1,1}, ..., \Gamma^{1,5}$

as follows

$$r_1(\Gamma_1) = \begin{cases} 0, \quad 0 \le \Gamma_1 < \Gamma^{1,1} \\ r^{(1)} = 1/5, \quad \Gamma^{1,1} \le \Gamma_1 < \Gamma^{1,2} \\ r^{(2)} = 1/3, \quad \Gamma^{1,2} \le \Gamma_1 < \Gamma^{1,3} \\ r^{(3)} = 2/5, \quad \Gamma^{1,3} \le \Gamma_1 < \Gamma^{1,4} \\ r^{(4)} = 1/2, \quad \Gamma^{1,4} \le \Gamma_1 < \Gamma^{1,5} \\ r^{(5)} = 2/3, \quad \Gamma^{1,5} \le \Gamma_1 < \Gamma^{1,6} = \infty \end{cases} \tag{2.17}$$

When $0 \le \Gamma_1 < \Gamma^{1,1}$, $r_1(\Gamma_1) = 0$ means the transmitter does not transmit anything since the success probability is not high enough, so the transmitter will wait and decide the code rate again in the second transmission.

If the first transmission succeeds, there is no second transmission. If the first transmission fails, the FEC code rate in the second transmission $r_2(\Gamma_1, \Gamma_2)$ is determined as shown in Fig. 2.1.

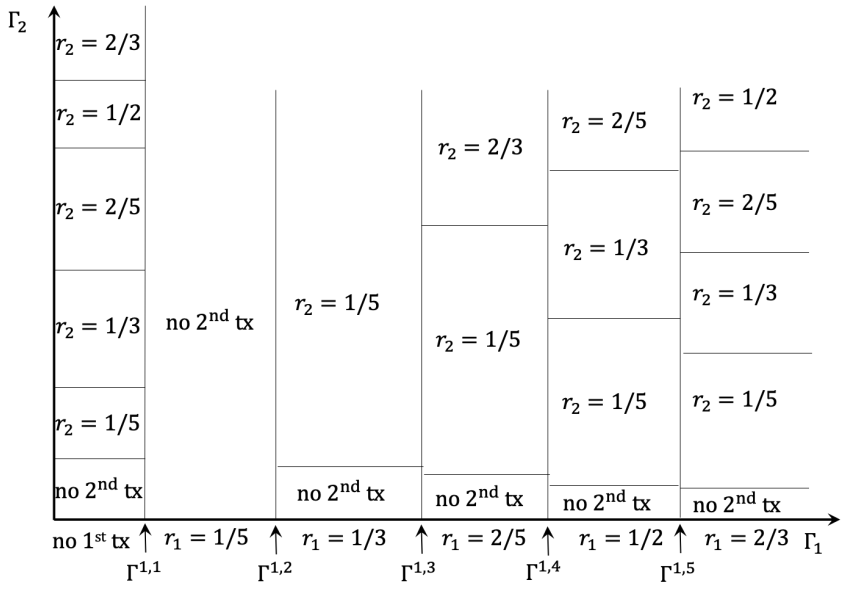

Figure 2.1: SNR boundaries for two transmissions.

The detailed explanation of Fig. 2.1 is as follows. The x and y axes correspond to the received SNR for the first and second transmissions. If the channel for the first packet is bad,

corresponding to the leftmost portion of the figure, i.e,. $0 \leq \Gamma_1 < \Gamma^{1,1}$, the transmitter chooses not to transmit the packet. Since nothing was transmitted in the first transmission, the transmitter can use any code rate in the second transmission, including discarding the packet. Therefore, when $0 \leq \Gamma_1 < \Gamma^{1,1}$,

$$r_2(\Gamma_1, \Gamma_2) = \begin{cases} 0, \quad 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad 0 \leq \Gamma_2 < \Gamma^{2,1} \\ r^{(1)} = 1/5, \quad 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad \Gamma^{2,1} \leq \Gamma_2 < \Gamma^{2,2} \\ r^{(2)} = 1/3, \quad 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad \Gamma^{2,2} \leq \Gamma_2 < \Gamma^{2,3} \\ r^{(3)} = 2/5, \quad 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad \Gamma^{2,3} \leq \Gamma_2 < \Gamma^{2,4} \\ r^{(4)} = 1/2, \quad 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad \Gamma^{2,4} \leq \Gamma_2 < \Gamma^{2,5} \\ r^{(5)} = 2/3, \quad 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad \Gamma^{2,5} \leq \Gamma_2 < \Gamma^{2,6} = \infty \end{cases} \tag{2.18}$$

When $\Gamma^{1,1} \leq \Gamma_1 < \Gamma^{1,2}$, FEC code rate 1/5 is used in the first transmission. If the first transmission fails, $r_2(\Gamma_1, \Gamma_2)$ = 0, i.e., nothing is transmitted, since the lowest code rate was used in the first transmission.

When $\Gamma^{1,2} \leq \Gamma_1 < \Gamma^{1,3}$, FEC code rate 1/3 is used in the first transmission. If the first transmission fails, the transmitter can choose to either discard the packet or use FEC code rate 1/5.

$$r_2(\Gamma_1, \Gamma_2) = \begin{cases} 0, \quad \Gamma^{1,2} \leq \Gamma_1 < \Gamma^{1,3}, \quad 0 \leq \Gamma_2 < \Gamma^{2,7} \\ r^{(1)} = 1/5, \quad \Gamma^{1,2} \leq \Gamma_1 < \Gamma^{1,3}, \quad \Gamma^{2,7} \leq \Gamma_2 < \Gamma^{2,8} = \infty \end{cases} \tag{2.19}$$

When $\Gamma^{1,3} \leq \Gamma_1 < \Gamma^{1,4}$, FEC code rate 2/5 is used in the first transmission. If the first transmission fails, the transmitter can choose to either discard the packet, or use FEC code rate

1/3, or use FEC code rate 1/5.

$$r_2(\Gamma_1,\Gamma_2)=\begin{cases}0, & \Gamma^{1,3}\le\Gamma_1<\Gamma^{1,4}, \quad 0\le\Gamma_2<\Gamma^{2,9}\\ r^{(1)}=1/5, & \Gamma^{1,3}\le\Gamma_1<\Gamma^{1,4}, \quad \Gamma^{2,9}\le\Gamma_2<\Gamma^{2,10}\\ r^{(2)}=1/3, & \Gamma^{1,3}\le\Gamma_1<\Gamma^{1,4}, \quad \Gamma^{2,10}\le\Gamma_2<\Gamma^{2,11}=\infty\end{cases} \tag{2.20}$$

Similarly, when $\Gamma^{1,4}\le\Gamma_1<\Gamma^{1,5}$,

$$r_2(\Gamma_1,\Gamma_2)=\begin{cases}0, & \Gamma^{1,4}\le\Gamma_1<\Gamma^{1,5}, \quad 0\le\Gamma_2<\Gamma^{2,12}\\ r^{(1)}=1/5, & \Gamma^{1,4}\le\Gamma_1<\Gamma^{1,5}, \quad \Gamma^{2,12}\le\Gamma_2<\Gamma^{2,13}\\ r^{(2)}=1/3, & \Gamma^{1,4}\le\Gamma_1<\Gamma^{1,5}, \quad \Gamma^{2,13}\le\Gamma_2<\Gamma^{2,14}\\ r^{(3)}=2/5, & \Gamma^{1,4}\le\Gamma_1<\Gamma^{1,5}, \quad \Gamma^{2,14}\le\Gamma_2<\Gamma^{2,15}=\infty\end{cases} \tag{2.21}$$

When $\Gamma^{1,5}\le\Gamma_1<\Gamma^{1,6}=\infty$,

$$r_2(\Gamma_1,\Gamma_2)=\begin{cases}0, & \Gamma^{1,5}\le\Gamma_1<\infty, \quad 0\le\Gamma_2<\Gamma^{2,16}\\ r^{(1)}=1/5, & \Gamma^{1,5}\le\Gamma_1<\infty, \quad \Gamma^{2,16}\le\Gamma_2<\Gamma^{2,17}\\ r^{(2)}=1/3, & \Gamma^{1,5}\le\Gamma_1<\infty, \quad \Gamma^{2,17}\le\Gamma_2<\Gamma^{2,18}\\ r^{(3)}=2/5, & \Gamma^{1,5}\le\Gamma_1<\infty, \quad \Gamma^{2,18}\le\Gamma_2<\Gamma^{2,19}\\ r^{(4)}=1/2, & \Gamma^{1,5}\le\Gamma_1<\infty, \quad \Gamma^{2,19}\le\Gamma_2<\Gamma^{2,20}=\infty\end{cases} \tag{2.22}$$

Therefore, $r_1(\Gamma_1)$ and $r_2(\Gamma_1,\Gamma_2)$ are determined by the SNR boundaries $\Gamma^{1,1},...,\Gamma^{1,5}$, $\Gamma^{2,1}$, ..., $\Gamma^{2,20}$, where some of them are ∞ for simplicity of notation in the following derivation.

The average overall energy consumption is

$$
\begin{aligned}
\overline{\mathcal{E}} =& S_0 T N_b \sum_{i=1}^{5} \int_{0}^{\Gamma^{1,1}} \int_{\Gamma^{2,i}}^{\Gamma^{2,(i+1)}} \frac{1}{r^{(i)}} f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& S_0 T N_b \int_{\Gamma^{1,1}}^{\Gamma^{1,2}} \frac{1}{r^{(1)}} f_{\Gamma_1}(\Gamma_1) d\Gamma_1 + \\
& S_0 T N_b \int_{\Gamma^{1,2}}^{\Gamma^{1,3}} \int_{0}^{\Gamma^{2,7}} \frac{1}{r^{(2)}} f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& S_0 T N_b \int_{\Gamma^{1,2}}^{\Gamma^{1,3}} \int_{\Gamma^{2,7}}^{\Gamma^{2,8}} \left(\frac{1}{r^{(2)}} + P(e_1; r^{(2)}|\Gamma_1) \left(\frac{1}{r^{(1)}} - \frac{1}{r^{(2)}} \right) \right) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& S_0 T N_b \int_{\Gamma^{1,3}}^{\Gamma^{1,4}} \int_{0}^{\Gamma^{2,9}} \frac{1}{r^{(3)}} f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& S_0 T N_b \sum_{i=1}^{2} \int_{\Gamma^{1,3}}^{\Gamma^{1,4}} \int_{\Gamma^{2,(i+8)}}^{\Gamma^{2,(i+9)}} \left(\frac{1}{r^{(3)}} + P(e_1; r^{(3)}|\Gamma_1) \left(\frac{1}{r^{(i)}} - \frac{1}{r^{(3)}} \right) \right) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& S_0 T N_b \int_{\Gamma^{1,4}}^{\Gamma^{1,5}} \int_{0}^{\Gamma^{2,12}} \frac{1}{r^{(4)}} f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& S_0 T N_b \sum_{i=1}^{3} \int_{\Gamma^{1,4}}^{\Gamma^{1,5}} \int_{\Gamma^{2,(i+11)}}^{\Gamma^{2,(i+12)}} \left(\frac{1}{r^{(4)}} + P(e_1; r^{(4)}|\Gamma_1) \left(\frac{1}{r^{(i)}} - \frac{1}{r^{(4)}} \right) \right) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 \\
& S_0 T N_b \int_{\Gamma^{1,5}}^{\infty} \int_{0}^{\Gamma^{2,16}} \frac{1}{r^{(5)}} f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& S_0 T N_b \sum_{i=1}^{4} \int_{\Gamma^{1,5}}^{\infty} \int_{\Gamma^{2,(i+15)}}^{\Gamma^{2,(i+16)}} \left(\frac{1}{r^{(5)}} + P(e_1; r^{(5)}|\Gamma_1) \left(\frac{1}{r^{(i)}} - \frac{1}{r^{(5)}} \right) \right) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1
\end{aligned}
\tag{2.23}
$$

The average overall packet loss rate is

$$
\begin{aligned}
\overline{P_L} = & \int_0^{\Gamma^{1,1}} \int_0^{\Gamma^{2,1}} f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& \sum_{i=1}^{5} \int_0^{\Gamma^{1,1}} \int_{\Gamma^{2,i}}^{\Gamma^{2,(i+1)}} P(e_1; r^{(i)}|\Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& \int_{\Gamma^{1,1}}^{\Gamma^{1,2}} P(e_1; r^{(1)}|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 + \\
& \int_{\Gamma^{1,2}}^{\Gamma^{1,3}} \int_0^{\Gamma^{2,7}} P(e_1; r^{(2)}|\Gamma_1) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& \int_{\Gamma^{1,2}}^{\Gamma^{1,3}} \int_{\Gamma^{2,7}}^{\Gamma^{2,8}} P(e_2; r^{(2)}, r^{(1)}|\Gamma_1,\Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& \int_{\Gamma^{1,3}}^{\Gamma^{1,4}} \int_0^{\Gamma^{2,9}} P(e_1; r^{(3)}|\Gamma_1) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& \sum_{i=1}^{2} \int_{\Gamma^{1,3}}^{\Gamma^{1,4}} \int_{\Gamma^{2,(i+8)}}^{\Gamma^{2,(i+9)}} P(e_2; r^{(3)}, r^{(i)}|\Gamma_1,\Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& \int_{\Gamma^{1,4}}^{\Gamma^{1,5}} \int_0^{\Gamma^{2,12}} P(e_1; r^{(4)}|\Gamma_1) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& \sum_{i=1}^{3} \int_{\Gamma^{1,4}}^{\Gamma^{1,5}} \int_{\Gamma^{2,(i+11)}}^{\Gamma^{2,(i+12)}} P(e_2; r^{(4)}, r^{(i)}|\Gamma_1,\Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& \int_{\Gamma^{1,5}}^{\Gamma^{1,6}} \int_0^{\Gamma^{2,16}} P(e_1; r^{(5)}|\Gamma_1) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1 + \\
& \sum_{i=1}^{4} \int_{\Gamma^{1,5}}^{\Gamma^{1,6}} \int_{\Gamma^{2,(i+15)}}^{\Gamma^{2,(i+16)}} P(e_2; r^{(5)}, r^{(i)}|\Gamma_1,\Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2) d\Gamma_2 d\Gamma_1
\end{aligned}
\tag{2.24}
$$

Let the Lagrangian function be $L = \overline{\mathcal{E}} + \lambda(\overline{P_L} - P_{const})$. We can solve the problem by setting $\frac{\partial L}{\partial \lambda} = 0$ and $\frac{\partial L}{\partial \Gamma^{i,j}} = 0$, for i = 1, j = 1,..., 5 and i = 2, j = 1, 2, 3, 4, 5, 8, 11, 12, 15, 16, 17, 20, 21, 22, 23. We skip the derivation of the derivatives since it is similar to Section 2.3.3.

2.4 Numerical Results

We compare the proposed IR HARQ to the schemes without combining and with Chase combining. For the comparison schemes, we show the derivation for $K = 2$ transmission rounds although it can be easily extended to arbitrary K. Since the lowest code rate in $\{r\}$ is 1/5, the maximum number of coded bits that can be transmitted in K transmission rounds for IR combining is $5N_b$. We limit the maximum number of coded bits in the comparison schemes to the same value. For the schemes without combining and with Chase combining, we add the constraint that $N_1 + N_2 + ... + N_K \leq 5N_b$, where N_i is the number of coded bits in the i-th transmission. This means $\frac{1}{r_1} + \frac{N_b}{r_2} + ... + \frac{N_b}{r_K} \leq 5N_b$, where r_i is the FEC code rate *in* the i-th transmission and $r_i \in \{r\}$. Note that the definition of r_i is different for the IR combining and the comparison schemes. For the IR combining, r_i is the FEC code rate *after* the i-th transmission, because the incremental bits cannot be independently decoded, and the FEC code rate *in* the i-th transmission is not meaningful.

2.4.1 Without combining

Each transmission round is decoded independently. This is a baseline scheme and we only discuss the case where CSI is not available at the transmitter. Since the fading is independent in the transmissions and no combining is used, the FEC code rates for all the transmission rounds should be predetermined, i.e, the FEC code rate r_i is a function of only the transmission round

number i. The optimization problem is as follows:

$$\begin{aligned}
\text{min} \quad & S_0 T_s \left(N_1 + \int_0^\infty N_2 P(e_1; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right) \\
\text{s.t.} \quad & \int_0^\infty \int_0^\infty P(e_1, e_2; N_1, N_2|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 d\Gamma_2 \leq P_{const} \\
& \frac{N_b}{N_1} \in \{r\}, \frac{N_b}{N_2} \in \{r\} \\
& N_1 + N_2 \leq 5N_b \\
\text{variables:} \quad & N_1, N_2
\end{aligned} \tag{2.25}$$

Since the two transmissions are independent, we have

$$\begin{aligned}
& P(e_1, e_2; N_1, N_2|\Gamma_1, \Gamma_2) \\
=& P(e_1; N_1, N_2|\Gamma_1, \Gamma_2) P(e_2; N_1, N_2|e_1, \Gamma_1, \Gamma_2) \\
=& P(e_1; N_1|\Gamma_1) P(e_2; N_2|\Gamma_2)
\end{aligned} \tag{2.26}$$

where $P(e_1; N_1|\Gamma_1)$ and $P(e_2; N_2|\Gamma_2)$ are the conditional packet error rates in the first and second transmissions, conditioned on Γ_1 and Γ_2, respectively. The second line in Equation (2.26) is from Bayes rule. In the third line, $P(e_1; N_1|\Gamma_1) = P(e_1; N_1, N_2|\Gamma_1, \Gamma_2)$ since the error probability in the first transmission does not depend on either the future CSI or the code rate in the second transmission, and $P(e_2; N_2|\Gamma_2) = P(e_2; N_1, N_2|e_1, \Gamma_1, \Gamma_2)$ since the error probability in the second transmission is totally determined by the CSI and code rate in the second transmission, and does not depend on anything in the first transmission, e.g., the CSI, the code rate or the decoding result (success or failure). The terms $P(e_1; N_1|\Gamma_1)$ and $P(e_2; N_2|\Gamma_2)$ are approximated by Equation (2.9). We exhaustively search all the possible (N_1, N_2), and find the one which satisfies the constraint in Equation (2.25) and yields the least energy consumption.

2.4.2 Chase combining

Each transmission repeats the FEC code rate from the first transmission, and the packets are combined for decoding at the receiver. So the FEC code rate in all transmissions is determined by the first transmission. We discuss two kinds of CSI availability models: the transmitter has no CSI or has the current CSI. If the transmitter has only previous CSI, then the system cannot utilize this information because the first transmission does not have any CSI and the later transmissions are forced to use the same FEC code rate as the first transmission.

Without CSI

The optimization problem is as follows:

$$\begin{aligned} \min \quad & S_0 T_s \left(N_1 + \int_0^\infty N_1 P(e_1; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right) \\ \text{s.t.} \quad & \int_0^\infty \int_0^\infty P(e_1, e_2; N_1, N_1|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 d\Gamma_2 \leq P_{const} \\ & \frac{N_b}{N_1} \in \{r\} \\ & N_1 + N_1 \leq 5N_b \\ \text{variable:} \quad & N_1 \end{aligned} \tag{2.27}$$

where $P(e_1, e_2; N_1, N_1|\Gamma_1, \Gamma_2)$ is the conditional probability that the decoding fails for both transmissions, conditioned on Γ_1 and Γ_2. Using the same reasoning as for the IR combining, the probability that the first transmission succeeds and the second transmission fails is much smaller than the probability that the second transmission fails, so we have $P(e_1, e_2; N_1, N_1|\Gamma_1, \Gamma_2) \approx P(e_2; N_1, N_1|\Gamma_1, \Gamma_2)$. With maximum ratio combining, and assuming the receiver can perfectly estimate the channel state, this is equivalent to decoding a single transmission with received SNR $\Gamma_1 + \Gamma_2$. So $P(e_2; N_1, N_1|\Gamma_1, \Gamma_2) = P(e_2; N_1|\Gamma_1 + \Gamma_2)$, where $P(e_2; N_1|\Gamma_1 + \Gamma_2)$ is the error probability of decoding a packet with N_1 bits and received SNR $\Gamma_1 + \Gamma_2$, and can be approximated

by Equation (2.9). An exhaustive search is used to find the N_1 that satisfies the constraint in Equation (2.27) and yields the least energy consumption.

With current CSI

With the current CSI available at the transmitter, the FEC code rate in the first transmission depends on Γ_1. The optimization problem is as follows:

$$\begin{aligned}
\text{min} \quad & \overline{\mathcal{E}} = S_0 T_s \int_0^\infty \left(N_1(\Gamma_1) + N_1(\Gamma_1) P(e_1; (\Gamma_1)|\Gamma_1)\right) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \\
\text{s.t.} \quad & \overline{P_L} = \int_0^\infty \int_0^\infty P(e_1, e_2; N_1(\Gamma_1), N_1(\Gamma_1)|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 d\Gamma_2 \leq P_{const} \\
& \frac{N_b}{N_1(\Gamma_1)} \in \{r\} \\
& N_1(\Gamma_1) + N_1(\Gamma_1) \leq 5N_b \\
\text{variable:} \quad & N_1(\Gamma_1)
\end{aligned} \tag{2.28}$$

The FEC code rate in the first transmission $r_1(\Gamma_1)$ can be determined by a set of SNR boundaries $\Gamma^1, ..., \Gamma^5$ as follows:

$$r_1(\Gamma_1) = \begin{cases} 0, & 0 \leq \Gamma_1 < \Gamma^1 \\ r^{(1)} = 1/5, & \Gamma^1 \leq \Gamma_1 < \Gamma^2 \\ r^{(2)} = 1/3, & \Gamma^2 \leq \Gamma_1 < \Gamma^3 \\ r^{(3)} = 2/5, & \Gamma^3 \leq \Gamma_1 < \Gamma^4 \\ r^{(4)} = 1/2, & \Gamma^4 \leq \Gamma_1 < \Gamma^5 \\ r^{(5)} = 2/3, & \Gamma^5 \leq \Gamma_1 < \infty \end{cases} \tag{2.29}$$

From Equation (2.29), we can easily get $N_1(\Gamma_1)$. The Lagrangian function is $L = \overline{\mathcal{E}} + \lambda(\overline{P_L} - P_{const})$. The optimal solution can be obtained by setting $\frac{\partial L}{\partial \lambda} = 0$ and $\frac{\partial L}{\partial \Gamma^i} = 0$ for i = 1, 2, ..., 5. We skip the derivation of the derivatives since it is similar to Section 2.3.3.

2.4.3 Performance Comparison

Fig. 2.2 shows the overall average energy consumption vs. average channel SNR for different cases when the maximum number of transmissions K is two. The average channel SNR is defined as $\mathbb{E}[\Gamma_i] = \frac{S_0 T_s}{N_0}\mathbb{E}[\gamma_i^2]$, where $\mathbb{E}[\cdot]$ is the expectation operation. The PLR constraint is 0.01. For all the scenarios, the energy consumption decreases with increasing average channel SNR because it is possible to use a high FEC code rate at high channel SNR to achieve the PLR constraint, and thus save energy. For IR combining, the scheme with both current and previous CSI yields the least energy consumption, and the scheme without any CSI consumes the most energy. For an average channel SNR of 7dB, the scheme with both current and previous CSI consumes 5% less energy than the scheme using only previous CSI, and 19% less energy than the scheme without any CSI. For a given CSI availability, e.g., without any CSI, IR combining outperforms Chase combining. For a channel SNR of 8dB, IR combining consumes 21% less energy than Chase combining assuming no CSI is available. Note that Chase combining sometimes performs worse than no combining, even if Chase combining utilizes current CSI. The reason is that the scheme without combining is allowed to use different FEC code rates in different transmissions. As shown in [42], the multiple transmission opportunities can be leveraged by using a high FEC code rate in the first transmission, thus saving energy, while later transmissions use low FEC code rates to increase the success probability. If the first transmission is successful, the later, more energy-consuming transmissions, are avoided. However, the Chase combining scheme is forced to use the same FEC code rate in each transmission, and thus loses the advantage of unequal energy allocation among multiple transmissions.

Note that some of the curves are jagged because the FEC code rate set $\{r\}$ is discrete. As the channel SNR changes, the system may jump from one FEC code rate option to another, thus yielding different energy consumption. The curves would be less jagged if $\{r\}$ contains more rate options. The left end of each curve corresponds to the minimum channel SNR such that the PLR constraint can be achieved. For the schemes with IR combining, more CSI information at the

transmitter allows the system to achieve the PLR constraint at a lower channel SNR. For a given CSI availability, e.g., no CSI at the transmitter, IR combining can achieve the PLR constraint at a lower channel SNR than can both Chase combining and the system without combining.

Fig. 2.3 shows $\mathcal{E}_i$ for the IR combining with both current and previous CSI when $K = 2$, where $\mathcal{E}_i$ is the average energy consumption in the i-th transmission given that this transmission happens. So $\mathcal{E}_i$ can be written as follows:

$$\mathcal{E}_i = \begin{cases} \int_0^\infty S_0 T_s N_1(\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1, & i = 1 \\ \frac{S_0 T_s}{\int_0^\infty P(e_1; r_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1} \int_0^\infty \int_0^\infty P(e_1; r_1|\Gamma_1) N_2(\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 d\Gamma_2, & i = 2 \end{cases}. \tag{2.30}$$

The first transmission uses a high FEC code rate to save energy, whereas the second transmission uses a low FEC code rate to meet the PLR constraint of the system.

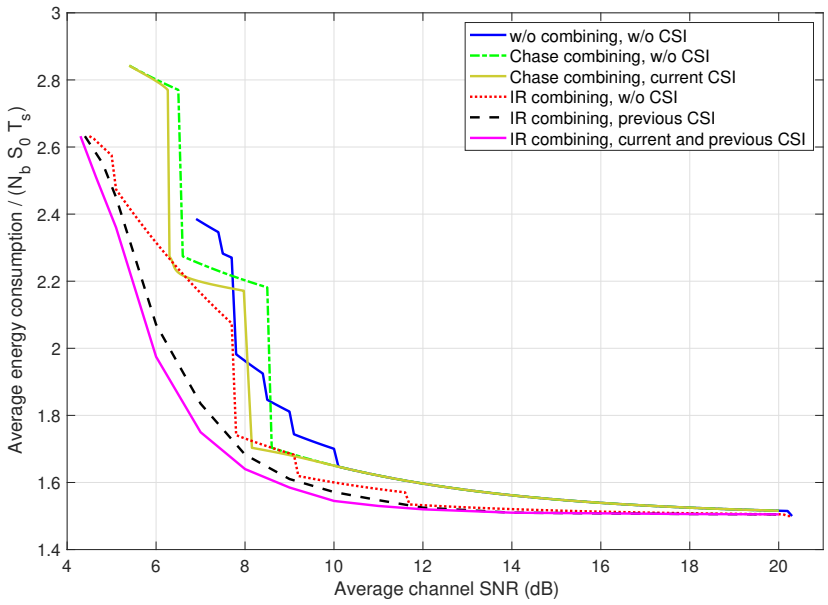

Figure 2.2: Energy consumption vs. average channel SNR. The maximum number of transmissions is two.

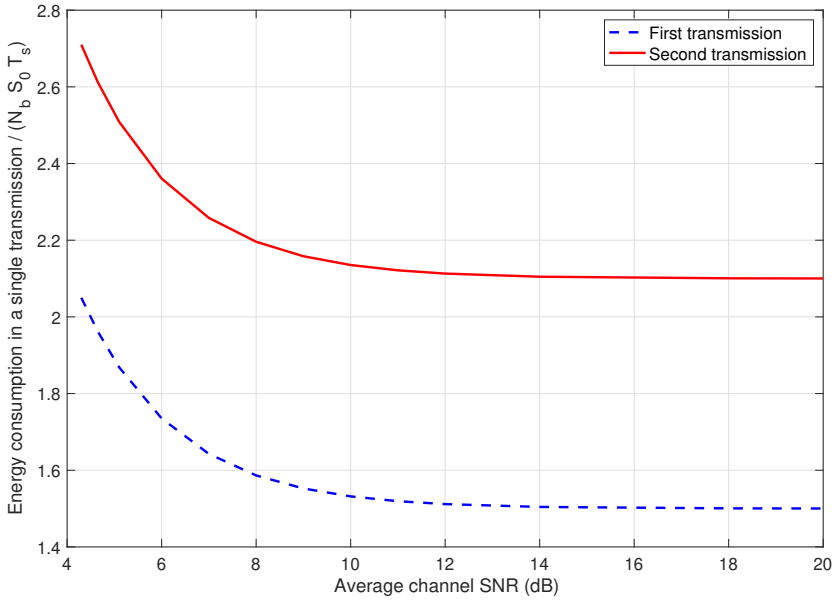

Figure 2.3: Energy consumption in each transmission. The maximum number of transmissions is two.

Fig. 2.4 shows the overall average energy consumption vs. average channel SNR for different cases when the maximum number of transmissions K is three. The PLR constraint is 0.01. The trends are similar to $K = 2$. For Chase combining, only one curve is shown, because all the transmissions have to use the FEC code rate 2/3 to ensure that the total number of transmitted bits does not exceed $5N_b$. At an average channel SNR of 4dB, IR combining with both current and previous CSI consumes 7% less energy consumption than IR combining with only previous CSI, and 10% less energy than IR combining without CSI. For all the scenarios, the energy consumption is decreased compared to the case where $K = 2$ in Fig. 2.2. For IR combining with both current and previous CSI, the energy consumption is decreased by 26% compared to $K = 2$ at an average channel SNR of 5dB, and the minimum average channel SNR such that the PLR can be achieved is 2.4dB smaller than that for $K = 2$. When more transmission opportunities are available, i.e., K is larger, less energy is consumed in the early transmission rounds. If the channel happens to be good and the packet is successfully transmitted, no additional energy is consumed;

If the packet fails, later transmissions use more energy to provide sufficient reliability. As seen in Figs. 2.2 and 2.4, the energy consumption is significantly reduced by having more transmission opportunities, as the system has multiple chances to get the packet through inexpensively before paying the higher energy cost on the final attempt to ensure the overall reliability.

Fig. 2.5 shows the $\mathcal{E}_i$ for IR combining with both current and previous CSI when $K = 3$. The trend is similar to $K = 2$: early transmissions use high FEC code rates and low energy consumption, whereas subsequent transmissions use low FEC code rates to provide the required PLR for the system.

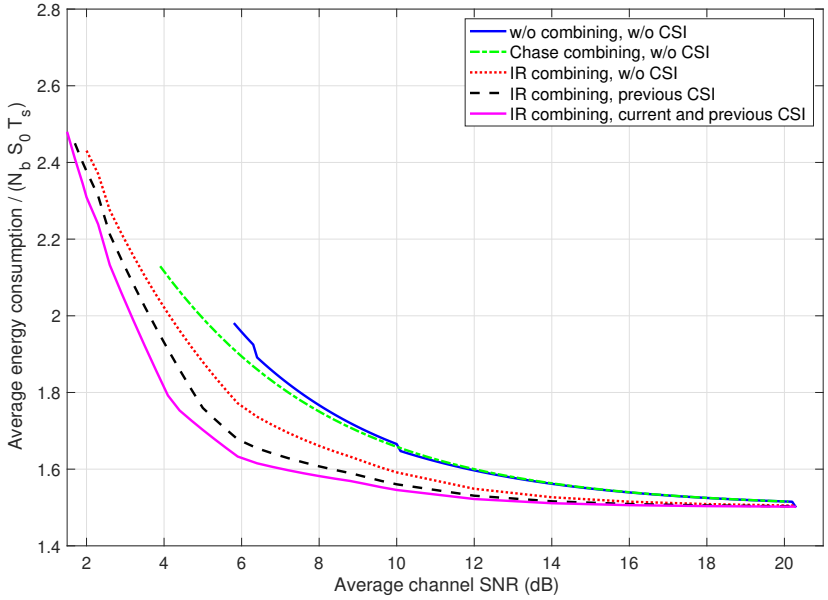

Figure 2.4: Energy consumption vs. average channel SNR. The maximum number of transmissions is three.

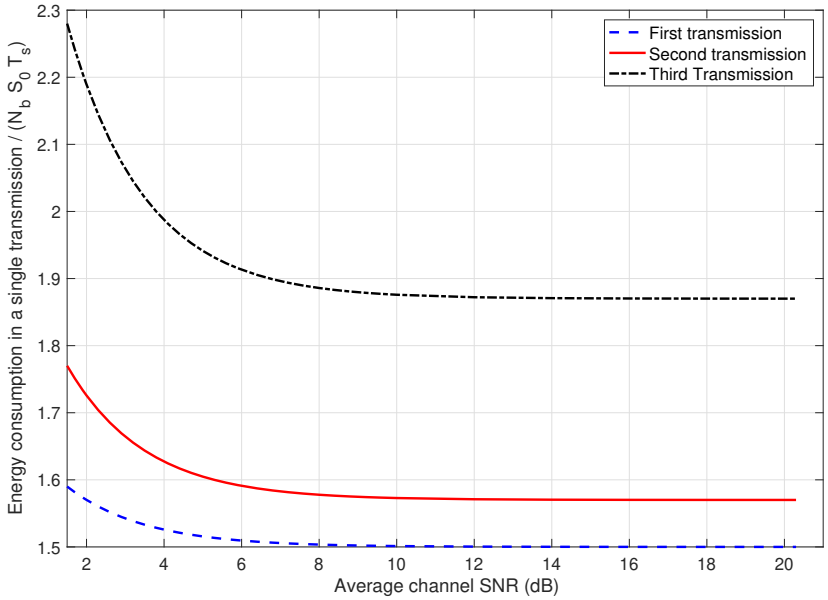

Figure 2.5: Energy consumption in each transmission. The maximum number of transmissions is three.

2.5 Conclusion

In this chapter, we consider rate allocation and rate adaptation for IR HARQ over independent block-fading channels with turbo coding. We minimize the energy consumption of HARQ, subject to a packet loss rate constraint. We investigate the influence of different CSI availabilities at the transmitter and compare different combining techniques. The key factor to reduce energy consumption for an IR HARQ system is unequal energy allocation among the multiple transmissions, which is similar to the finding for the HARQ system without combining in [42]. This explains why the energy consumption significantly decreases when the maximum number of transmissions is larger: having more transmission opportunities allows the system to consume less energy in early transmissions, and thus saves energy. It also explains why more CSI information at the transmitter helps to reduce energy consumption, but the difference between different CSI availabilities is not significant: even if early transmissions do not have the CSI for

that transmission, or even no CSI at all, they can consume low energy consumption (by using a high FEC code rate) to save energy, and later transmissions can adjust the energy consumption based on the channel states of the transmitted bits. This also explains why Chase combining sometimes performs worse than the system without combining: the former one is forced to use the same FEC code rate in all transmissions, whereas the latter one is allowed to use different FEC code rates in the transmissions. In addition, numerical results show that IR combining consumes less energy than both the Chase combining and the HARQ system without combining.

Chapter 2, in part, is a reprint of the material as it appears in the papers: B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization For Hybrid ARQ With Turbo Coding: Rate Adaptation and Allocation," submitted to *IEEE Transactions on Vehicular Technology*, and B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization for Incremental Redundancy Hybrid-ARQ," In *2019 53rd Asilomar Conference on Signals, Systems, and Computers. IEEE, 2019*. The dissertation author was the primary investigator and author of these papers.

Chapter 3

Energy Optimization for Wireless Video Transmission Employing Hybrid ARQ

We investigate energy-optimized wireless video transmission employing HARQ. We formulate the problem as maximizing the video quality, subject to a constraint on the wireless transmission energy consumption. We consider multiple parameters in multiple layers in a wireless video transmission system: transmit power, alphabet size, FEC code rate, maximum number of transmissions and unequal video data importance. An analytical framework is proposed to include these parameters, which allows us to divide this problem into two sub-problems: data transmission and unequal error protection. The problem is tackled by solving the two sub-problems, which are done by exhaustive search and convex optimization, respectively.

3.1 Introduction

In this chapter, we consider a cross-layer optimization with a maximum number of retransmissions. We consider both unequal importance for the transmissions of the same video content, and unequal importance for different video contents, and we include the analysis of both

HARQ and adaptive modulation and coding (AMC). We consider multiple parameters in multiple layers in a wireless video transmission system: transmit power, alphabet size, FEC code rate, maximum number of transmissions and unequal frame importance. We decouple the problem into two sub-problems, data transmission and video UEP, which are solved by exhaustive search and convex optimization, respectively. Finally we conduct simulations to compare the performance of the proposed scheme to existing data HARQ and/or video UEP schemes. The results show that the proposed scheme significantly outperforms comparison HARQ and UEP methods.

The contributions of this chapter can be summarized as follows:

- We establish a framework to optimize video quality, subject to an energy consumption constraint, which considers both the unequal importance of multiple transmissions and the unequal importance of different video contents.

- We decouple a non-convex problem into two sub-problems, where the first one is non-convex and solved by exhaustive search, and the second one is convex and solved using a Lagrangian multiplier.

- We conduct simulations on different videos, and both simulation and theoretical results show that the proposed scheme significantly outperforms the baseline scheme which either takes into account only one of the unequal importances considered in this paper or does not consider any unequal importance at all.

- We analyze the performance of the algorithm using a specific set of parameters for all videos, instead of using the optimal parameters for each video. Results show that the specific set of parameters can provide close-to-optimal video quality, and thus the proposed algorithm can be run offline to generate a lookup table. During the video transmission, the parameters are obtained from the table, and the power and extra latency due to computation are avoided. The complexity of the offline algorithm is analyzed.

The rest of the chapter is organized as follows. The related work is introduced in Section 3.2. In Section 3.3, we formulate the problem and propose an algorithm. In Sections 3.4 and 3.5, we introduce the two procedures in the algorithm. We show simulation results in Section 3.6, and conclusions are drawn in Section 3.7.

3.2 Related Work

3.2.1 Energy Optimization For Video Transmission

Energy optimization for video transmission has been studied by many researchers. In [43], the authors developed a framework to estimate the energy consumption of the video based on the channel state and video characteristics. However, the unequal importance of multiple transmissions was not fully investigated. In [44], the authors optimized the energy efficiency at the application layer. They adapt video encoder parameters based on the channel state and drop less important packets to avoid network congestion. In [45], a quasi-quadrature modulation is used to minimize the power consumption of video transmission. In [46], the authors proposed an algorithm to assign different forward error correction (FEC) code rates for the frames in a group of pictures (GOP) and drop packets with low priority. In [47], the authors studied coding unit level prioritization and FEC code rate selection in a wireless relay network. The authors in [48] studied the trade-off between energy saving and video quality and proposed an algorithm to assign different AMC modes to different video encoding layers. In [49], the authors maximized video quality, subject to an energy constraint in a D2D network, by optimizing the bit rate and FEC code rate on the paths between nodes in the D2D network. In [50], the optimal power allocation and the link adaptation algorithms are derived analytically, subject to a quality-of-service constraint. However, the fact that the video contents in a video sequence can require different quality-of-service, which reflects the unequal importance of video contents, is not taken into account. In [51], the source and channel coding are jointly optimized to achieve

the best video quality, under an energy constraint, which is suitable for video streaming scenarios. In [52], the authors optimized energy efficiency for MIMO-OFDM multimedia communication with quality-of-service constraints. The trade-off of the energy between video encoding and transmission was discussed in [53].

3.2.2 Video Transmission With HARQ

We summarize the work on video transmission employing HARQ in this section. In [54], the authors proposed an APP/MAC/PHY cross-layer framework to optimize perceptual video quality. The maximum number of retransmissions varies with the layer in which the packet is encoded. If a transmission fails, the next retransmission is assigned a lower order modulation and coding scheme to provide a satisfactory probability of success. A limited-retransmission priority encoding transmission (LR-PET) scheme investigated PET with multiple transmissions [55]. The optimal protection depends on both the importance of the stream content and its behavior in future transmissions. The authors in [56] showed that under a scenario where the video contents have the same decoding deadline, the optimal strategy concludes that the most important packets should be retransmitted as often as needed and the less important ones may get discarded. A prioritized retransmission scheme was proposed based on the error propagation effect of the lost packet [57]. The authors in [58] assessed the impact of the lost macroblocks on the reconstructed frame and the ones with the highest impact are prioritized to be retransmitted. In [59], a video flow is subdivided into independent and incrementally encoded packets and the HARQ scheme privileges the retransmission of independent packets. The maximum number of transmissions is changed based on the importance of the packet for LTE networks in [60, 61]. The authors in [62] considered interlayer FEC and HARQ, and found the best FEC code rate distribution among the video layers to minimize the video distortion. In [63], retransmissions are used for the base layer to reduce network congestion. The authors in [64] studied unequal error protection (UEP), retransmission and GOP-level interleaving. They proposed segment-wise and

byte-wise retransmission based on different types of receiver feedback. When the round trip time fluctuates significantly, the authors in [65] proposed a retransmission scheme to adaptively control the retransmission window size. The transmit power optimization problem was solved in [66] for multimedia applications, based on the observation that some performance metrics such as throughput, delay, and peak signal-to-noise ratio, may exhibit a staircase behavior for particular systems with HARQ. The authors in [67] proposed a scheduling scheme based on channel quality in LTE networks. In [68], the video source is able to adjust the video layers based on the channel estimator at the receiver. In [69], an adaptive unequal video protection method was designed for peer-to-peer video streaming over mobile wireless mesh networks, which assigns different priorities to the frames. A priority-based key frame protection method, which utilizes the idea of unequal error protection, was studied on a 5G network in [70]. A joint source-channel resource allocation problem was investigated in [71] to achieve adaptive error protection for video transmission.

The existing schemes on video retransmission employing HARQ can be categorized into two groups. For the first group, multiple transmissions of the same video content are treated equally, and the authors leverage the unequal importance of video contents which arises from the coding/decoding structure [72]. For example, for an IPPP encoding structure, the frames at the beginning of a GOP have higher importance than later frames because of the decoding dependency of later frames on previous frames. Thus, different video contents should have unequal protection during transmission to maximize the decoded video quality. Examples of leveraging unequal importance of video contents include assigning different maximum numbers of transmissions, assigning different priorities for retransmission, and using different FEC codes. We analyze this unequal importance quantitatively in Section 3.3.3, using the model in [73] and [74].

For the second group, the multiple transmissions of the same video content have unequal importance, but there lacks an analytical framework to provide the optimal strategy in each transmission. This unequal importance of multiple transmissions was studied using a single

alphabet size and variable power [13, 14]. The authors derived the optimal transmit power for each transmission round and this unequal power allocation outperforms the scheme in which the power is kept constant in each transmission. The optimal rate in each transmission for incremental redundancy was analyzed in [20, 40]. However, most of the existing data transmission HARQ schemes are based on information-theoretical analysis, which assumes there is no error when the SNR is larger than a threshold and assumes there is always an error when the SNR is smaller than the threshold. In our previous work [75], we used multiple alphabet sizes and variable power and investigated both the optimal transmit power and the optimal alphabet selection algorithm across the transmissions of the same data content, based on the actual packet error rate (PER) performance of a turbo code.

3.3 Problem Setup and Formulation

3.3.1 System Setup and Assumptions

A source encoder encodes a GOP, which contains T frames, with IPPP structure. Each frame is encoded into one or more slices. A slice is encoded by the FEC encoder and becomes a *codeword*, and is then sent through the wireless channel. A slice contains only information bits, a codeword is a slice plus parity bits, and a *packet* is any realization of a codeword. That is, a codeword can be transmitted multiple times, and the realization of each transmission is a packet. Slice copy is used at the decoder for error concealment. The term "packet error rate/probability" (PER) denotes the probability of error of any single transmission or retransmission of a codeword, and "overall packet error rate/probability" denotes the probability of error after a maximum number of transmissions of a codeword. For simplicity, the slices within frame i use the identical FEC code rate r_i, maximum number of transmissions N_i, power allocation algorithm and alphabet size mapping algorithm, where $i = 1, 2, ..., T$. The available FEC code rates are 1/2, 1/3 and 1/5, and the retransmissions of a codeword use the same FEC code rate as the first transmission.

The maximum number of transmissions $N_i \in \{1, 2, ...N_{max}\}$. For power allocation, the transmit power varies with retransmission number of a codeword/slice, but does not vary with the slice number in a frame. Thus, in a given frame, the power depends on only the retransmission number of a codeword/slice, but not which slice it is. The maximum transmit power is S_{max}. For alphabet size mapping, adaptive M-PSK is used, and available modulations are BPSK, QPSK, 8PSK and 16PSK. Thus, retransmissions are allowed to use different alphabet sizes from the first transmission. The alphabet mapping algorithm will be explained in Section 3.3.2. All packets contain L bits (including FEC, but excluding CRC bits and tail bits) and we ignore the influence of CRC bits and tail bits since their number is small relative to L. The symbol duration of the system, T_s, is kept constant, and a rectangular pulse shape is used. The number of symbols for an M-PSK packet is $L/log_2(M)$, and the number of information bits in a codeword in frame i is Lr_i. The receiver is a matched filter and coherent detection is used. The system block diagram is shown in Fig. 3.1.

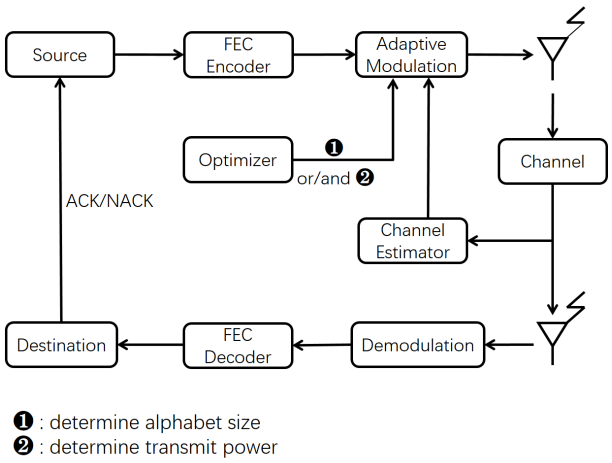

Figure 3.1: System Block Diagram.

The instantaneous channel gain for the j-th transmission of a codeword in frame i, $\gamma_{j,i}$, is assumed to be Rayleigh distributed, and the pdf of $\gamma_{j,i}$ does not change with time. The fades of different transmissions of a codeword are assumed to be independent, and we assume perfect CSI

is available at the transmitter.

The symbols and functions used in this chapter are listed in Table 3.1.

3.3.2 Alphabet Size Mapping

Since adaptive modulation is used, we introduce the alphabet size mapping in this subsection before we formulate the problem. The instantaneous received signal-to-noise ratio (SNR) for the j-th transmission of a codeword in frame i is $\Gamma_{j,i} = \gamma_{j,i}^2 S_{j,i} T_s / N_0$, where $S_{j,i}$ is the transmit power for the j-th transmission of a codeword in frame i and N_0 is the spectral density of the additive Gaussian noise. Since the packet error probability for turbo codes cannot be expressed analytically, we fit the packet error probability with an exponential function as in [22]. For a given code rate, we have

$$\psi_M(\Gamma_{j,i}) = \begin{cases} 1, & 0 < \Gamma_{j,i} < \Gamma_M^{min} \\ a_{log_2 M} e^{-b_{log_2 M} \Gamma_{j,i}}, & \Gamma_{j,i} \geq \Gamma_M^{min}, \end{cases} \tag{3.1}$$

where $\psi_M(\Gamma_{j,i})$ is the conditional single-transmission packet error probability, conditioned on $\Gamma_{j,i}$, for M-PSK, and $a_n > 0, b_n > 0, n = 1, 2, 3, 4$. The parameters a_n, b_n and Γ_M^{min} are obtained through simulation and curve fitting and they depend on the code rate. The simulated packet error rate for a rate 1/3 turbo code is shown in Fig. 3.2.

We map alphabet sizes as in [22]. We set a PER "upper bound" $U_{j,i}$ for the j-th transmission of the codewords in the i-th frame. $U_{j,i}$ and $U_{k,i}$ can be the same or different for $j \neq k$. Let the instantaneous received SNR boundaries for the j-th transmission of the slices in the i-th frame be $(\Gamma_{j,i}^{(1)}, \Gamma_{j,i}^{(2)}, \Gamma_{j,i}^{(3)}, \Gamma_{j,i}^{(4)}, \Gamma_{j,i}^{(5)})$, and choose M-PSK when $\Gamma_{j,i}^{(log_2 M)} \leq \Gamma_{j,i} < \Gamma_{j,i}^{(log_2 M+1)}$. The SNR boundaries for the j-th transmission of the codewords in the i-th frame are determined by

Table 3.1: Table of variables.

$\{\cdot\}$	the set of a variable
T_s	symbol duration
N_0	spectral density of Gaussian noise
L	number of coded bits in a packet
T	number of frames in a GOP
n_i	number of slices/codewords in frame i
N_i	maximum number of transmissions for a codeword in frame i
r_i	code rate for a codeword in frame i
$\gamma_{j,i}$	instantaneous channel gain for the j-th transmission of a codeword in frame i
$\Gamma_{j,i}$	instantaneous received SNR for the j-th transmission of a codeword in frame i
$f_{\Gamma_{j,i}}(\Gamma_{j,i})$	pdf of $\Gamma_{j,i}$ for the j-th transmission of a codeword in frame i
$S_{j,i}$	power for the j-th transmission of a codeword in frame i
$U_{j,i}$	upper bound on conditional PER for the j-th transmission of a codeword in frame i
$\Gamma^{(j)}$	SNR boundary for alphabet size
$\psi_M(\Gamma)$	conditional PER for M-PSK using turbo code
P_{ei}	average overall PER for a codeword in frame i
$MSE(\{N_i\},\{r_i\},\{S_{j,i}\},\{U_{j,i}\})$	function of the system parameters to calculate the MSE over GOP
$MSE(\{P_{ei}\})$	function of the average frame PER to calculate the MSE over GOP
$\mathcal{E}^{ave,N_i,r_i}(S_{1,i},...,S_{N_i,i},U_{1,i},...,U_{N_i,i})$	function to calculate the average overall energy for a codeword in frame i
$P^{ave,N_i,r_i}(S_{1,i},...,S_{N_i,i},U_{1,i},...,U_{N_i,i})$	function to calculate the average overall PER for a codeword in frame i
$\mathcal{E}_{min}^{N_i,r_i}(P_{ei})$	minimum average energy to achieve an average overall PER of P_{ei}, with given N_i and r_i

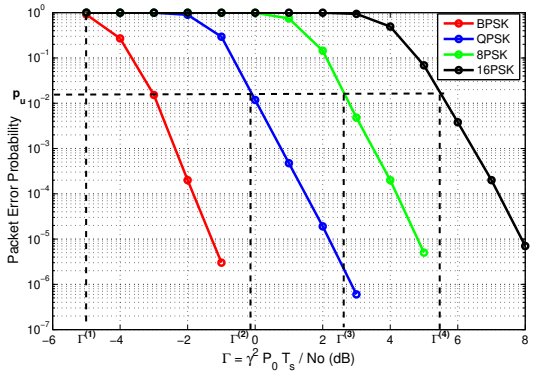

Figure 3.2: Packet error rate vs. instantaneous received SNR.

setting the left hand side of Equation (3.1) equal to $U_{j,i}$, where $\Gamma_{j,i}^{(1)} = \Gamma_2^{min}$, $\Gamma_{j,i}^{(5)} = \infty$:

$$\Gamma_{j,i}^{(n)} = \frac{1}{b_n} ln(\frac{a_n}{U_{j,i}}), \quad n = 2, 3, 4. \tag{3.2}$$

The terms $\Gamma_{j,i}^{(2)}, \Gamma_{j,i}^{(3)}$ and $\Gamma_{j,i}^{(4)}$ are determined by $U_{j,i}$. Thus, we aim to choose the largest alphabet size so that the curve-fitted conditional PER, conditioned on $\Gamma_{j,i}$, is lower than $U_{j,i}$ when $\Gamma_{j,i} > \Gamma_{j,i}^{(2)}$, and we choose BPSK when $\Gamma_{j,i}^{(1)} < \Gamma_{j,i} \leq \Gamma_{j,i}^{(2)}$, i.e., the curve fit to the conditional PER is lower than 1. The system does not transmit when $\Gamma_{j,i} \leq \Gamma_{j,i}^{(1)}$, i.e., the fitted conditional PER is 1. Although $U_{j,i}$ is not a strict upper bound on the conditional PER over the whole range where BPSK is used, we use the term "upper bound" since the conditional PER is lower than $U_{j,i}$ for all of the channel states which are above some threshold.

3.3.3 Problem Formulation

We want to minimize the mean square error (MSE) in the GOP of the given video sequence, subject to overall energy and maximum power constraints. The system parameters we can directly control are the maximum number of transmissions, FEC code rate, power, and PER "upper bound". According to [73] and [74], the MSE directly depends on the average PERs for the frames (although the average PER further depends on the system parameters above). Here

the "average PER" is the instantaneous PER averaged over the Rayleigh fading channel. If the average PER for frame i in a GOP is denoted by P_{ei}, where $i = 1, 2, ..., T$, then the MSE incurred in frame i is

$$\sigma_0^2 \sum_{\tau=0}^{i} P_{e\tau} \frac{1}{1 + \alpha(i - \tau)}, \tag{3.3}$$

where σ_0^2 are α are parameters which depend on the video. The above model assumes that 1) the pixel errors incurred in different frames are independent and 2) the pixel errors incurred in a frame, but propagated from different frames, are uncorrelated. The first assumption is valid in our model, since a packet can only contain contents from one frame, and different packets experience independent fading, thus the pixel errors in different frames are independent. The second assumption is realistic when P_{ei} is small. The accuracy of the model is evaluated in Section 3.5. Therefore, the MSE over the GOP is

$$\begin{aligned} MSE(\{P_{ei}\}) &= \sigma_0^2 \sum_{i=1}^{T} \sum_{\tau=1}^{i} P_{e\tau} \frac{1}{1 + \alpha(i - \tau)} \\ &= \sigma_0^2 \sum_{i=1}^{T} P_{ei} \sum_{\tau=0}^{T-i} \frac{1}{1 + \alpha\tau} \end{aligned}, \tag{3.4}$$

where $\{P_{ei}\} = (P_{e1}, P_{e2}, ..., P_{eT})$ is the set of PERs for T frames. Note that the MSE over the GOP in Equation (3.4) is just the sum of the MSE of every frame in the GOP, so Equation (3.4) does not require further assumptions beyond those made for Equation (3.3).

The optimization problem is formulated as follows:

$$\begin{aligned} min \quad & MSE(\{P_{ei}\}) \\ s.t. \quad & \sum_{i=1}^{T} n_i \mathcal{E}(P_{ei}) = min(\mathcal{E}_c, \tilde{\mathcal{E}}_c) \\ & 0 \leq S_{j,i} \leq S_{max} \quad \text{for } i = 1, 2, ..., T \text{ and} \\ & \qquad\qquad\qquad\qquad\qquad\qquad j = 1, 2, ..., N_i \\ variables: \quad & \{P_{ei}\} \end{aligned} \tag{3.5}$$

where n_i is the number of information bits in frame i, $\mathcal{E}_c$ is the energy constraint for the entire GOP, $\tilde{\mathcal{E}}_c = (\sum_{i=1}^{T} n_i) S_{max} T_s N_{max} / r_{min}$ represents the maximum energy that a GOP could consume. Here, r_{min} is the lowest FEC code rate (r_{min} = 1/5 in this paper). The maximum energy occurs when the FEC is the strongest possible and the maximum number of transmissions is used. The term $\mathcal{E}(P_{ei})$ is the energy consumption per information bit in frame i to achieve an average PER of P_{ei}, i.e., the energy consumption of a codeword divided by the number of information bits in the codeword. The units of $\mathcal{E}_c$, $\tilde{\mathcal{E}}_c$ and $\mathcal{E}(P_{ei})$ are Joules. We define $\mathcal{E}(P_{ei})$ in this way because the number of information bits in frame i is determined by the video encoder, which does not depend on the FEC rate r_i. Thus, $\mathcal{E}(P_{ei})$ is a fair comparison for different FEC rates. The set operator is $\{\cdot\}$. The maximum energy that an information bit in frame i can possibly use is $S_{max} T_s N_{max} / r_{min}$, i.e., the codeword uses BPSK, FEC rate r_{min}, maximum power S_{max} and gets transmitted N_{max} times. The energy constraint $\mathcal{E}_c$ is used to minimize MSE when $\mathcal{E}_c \leq \tilde{\mathcal{E}}_c$, and the energy consumption cannot exceed $\tilde{\mathcal{E}}_c$ when $\mathcal{E}_c > \tilde{\mathcal{E}}_c$. In Equation (3.5), the function $\mathcal{E}(P_{ei})$ is not uniquely defined in the above formulation since there can be multiple sets of parameters $(N_i, r_i, S_{j,i}, U_{j,i})$ which generate the same P_{ei}, but do not result in the same energy consumption. Thus, for a given P_{ei}, there can be multiple values of energy $\mathcal{E}(P_{ei})$.

As a consequence, for any P_{ei} being considered, we choose the energy-minimizing $\mathcal{E}_{min}(P_{ei})$ to achieve P_{ei}, where $\mathcal{E}_{min}(P_{ei})$ is defined as the least energy consumption per informa-

tion bit to achieve an average PER of P_{ei}. Notice that $\mathcal{E}_{min}(P_{ei})$ is a function of P_{ei}, but not the system parameters $(N_i, r_i, S_{j,i}, U_{j,i})$. Intuitively, if a particular packet is to be transmitted with a certain PER, it does not make sense to achieve that PER using a parameter set $(N_i, r_i, S_{j,i}, U_{j,i})$ that uses a higher energy if the same PER can be achieved using a parameter set that uses a lower energy. Hence in the following, as the algorithm searches through the possible values of PER, for each value of PER, only the energy-minimizing value of achieving that PER, $\mathcal{E}_{min}(P_{ei})$, will be considered.

Note that there are two constraints in Equation (3.5): energy constraint $\mathcal{E}_c$ and power constraint S_{max}. The energy $\mathcal{E}_c$ constraint is determined by the video application, and is determined by such factors as the desired video quality level and the battery status. The power constraint S_{max} is determined by the hardware, e.g., the power amplifier. These two constraints are determined independently.

3.3.4 Algorithm

Equation (3.5) is a non-convex optimization problem due to the non-convexity of the term $\mathcal{E}(P_{ei})$, so we solve it by using a combination of exhaustive search and Lagrangian multiplier method. First, we quantize P_{ei}. For each quantized value of P_{ei}, we then quantize the variables $S_{1,i}, ..., S_{N_i,i}, U_{1,i}, ..., U_{N_i,i}$ and exhaustively search all possible tuples of $N_i, r_i, S_{1,i}, ..., S_{N_i,i}$, $U_{1,i}, ..., U_{N_i,i}$ to find all the tuples which yield PER P_{ei}. Among these, we find the one which has the minimum energy consumption $\mathcal{E}_{min}(P_{ei})$. After examining all quantized values of P_{ei} and obtaining the corresponding minimum energy, we get the function $\mathcal{E}_{min}(P_{ei})$. Then we use the Lagrangian multiplier method to solve Equation (3.5).

Before introducing the details of the algorithm, we define $\mathcal{E}^{ave}(N_i, r_i, S_{1,i}, ..., S_{N_i,i}, U_{1,i}, ..., U_{N_i,i})$ to be the average overall energy consumption (including retransmissions) of an information bit in frame i, i.e., the overall energy consumption of a codeword divided by the number of information bits in the codeword, and we define $P^{ave}(N_i, r_i, S_{1,i}, ..., S_{N_i,i}, U_{1,i}, ..., U_{N_i,i})$ to be

the average overall PER for a codeword in frame i. Notice that the codewords in one frame have the same $N_i, r_i, S_{j,i}, U_{j,i}$, but the modulation alphabet is allowed to vary from one codeword to the next based on the channel state.

3.4 Proposed HARQ

3.4.1 Problem Formulation

The procedure which obtains $\mathcal{E}_{min}(P_{ei})$ solves the HARQ problem by optimizing the parameters in each transmission, and does not depend on frame number i, so we simplify the notation by removing the dependence on i. Then the problem can be formulated as follows:

$$\begin{aligned} min \quad & \mathcal{E}^{ave}(N, r, \{S\}, \{U\}) \\ s.t. \quad & P^{ave}(N, r, \{S\}, \{U\}) = P_e \\ & 0 \leq S_j \leq S_{max} \quad j = 1, 2, ..., N \\ variables: \quad & N, r, \{S\}, \{U\} \end{aligned}, \tag{3.6}$$

where N is the maximum number of transmissions of the codeword, r is the FEC code rate of the codeword, $\{S\} = (S_1, ..., S_N)$ with S_j being the transmit power for the j-th transmission of the codeword and $\{U\} = (U_1, ..., U_N)$ with U_j being the PER "upper bound" for the j-th transmission of the codeword. $\mathcal{E}^{ave}(N, r, \{S\}, \{U\})$ is the average overall energy consumption (including retransmissions) of an information bit in a codeword, $P^{ave}(N, r, \{S\}, \{U\})$ is the average overall PER for a codeword, and P_e is the average overall PER constraint. We define $\{\Gamma\} = (\Gamma_1, \Gamma_2, ..., \Gamma_M)$ as the set of channel states for M transmissions of the codeword, where $M \leq N$, and $\{X\} = (X_1, X_2, ..., X_M)$ as the set of outcomes for M transmissions of the codeword. X_j = 1 means the j-th transmission of the codeword is successful, and X_j = 0 means the j-th transmission is not successful. Let $P_e^{cndt}(N, r, \{S\}, \{U\} \mid \{\Gamma\})$ be the conditional

codeword error probability, conditioned on $\{\Gamma\}$, i.e., the probability that the codeword cannot be successfully decoded after N transmissions with the given channel states $\{\Gamma\}$ in the transmissions. Let $\mathcal{E}^{cndt}(N, r, \{S\}, \{U\} \mid \{\Gamma\}, \{X\})$ be the conditional energy consumption of an information bit, conditioned on $\{\Gamma\}$ and $\{X\}$, i.e., the sum of the energy consumptions of all transmissions of the codeword with the given channel states $\{\Gamma\}$, divided by the number of information bits in the codeword. Then we have the following equations:

$$\begin{aligned} & P_e^{cndt}(N, r, \{S\}, \{U\} \mid \{\Gamma\}) \\ =& \mathcal{P}(X_1 = 0, X_2 = 0, ..., X_N = 0 \mid \{\Gamma\}) \end{aligned} \tag{3.7}$$

$$\begin{aligned} & \mathcal{E}^{cndt}(N, r, \{S\}, \{U\} \mid \{\Gamma\}, \{X\}) \\ =& \begin{cases} \sum_{j=1}^{k} e_j(\Gamma_j) & \text{if } k \text{ is the smallest number such that} \\ & \qquad X_j = 1, k \leq N \\ \sum_{j=1}^{N} e_j(\Gamma_j) & \text{if } X_j = 0 \text{ for all } j \in [1, N-1] \end{cases}, \end{aligned} \tag{3.8}$$

where $e_j(\Gamma_j)$ is the energy consumption during the j-th transmission of an information bit, given that the channel state is Γ_j.

3.4.2 Special Case

For simplicity of presentation, consider the special case where $N = 2$. Other cases can be shown to follow by similar analysis. The average overall PER is

$$
\begin{aligned}
& P^{ave}(2, r, \{S\}, \{U\}) \\
=& E_{\{\Gamma\}}[P_e^{cndt}(2, r, \{S\}, \{U\} \mid \{\Gamma\})] \\
=& \int_0^\infty \int_0^\infty P_e^{cndt}(2, r, \{S\}, \{U\} \mid \{\Gamma\}) f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 d\Gamma_2 \\
=& \int_0^\infty \int_0^\infty P_{e1}^{cndt}(r, S_1, U_1 \mid \Gamma_1) P_{e2}^{cndt}(r, S_2, U_2 \mid \Gamma_2) \times \\
& f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 d\Gamma_2 \\
=& \int_0^\infty P_{e1}^{cndt}(r, S_1, U_1 \mid \Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \times \\
& \int_0^\infty P_{e2}^{cndt}(r, S_2, U_2 \mid \Gamma_2) f_{\Gamma_2}(\Gamma_2) d\Gamma_2
\end{aligned}
\tag{3.9}
$$

where $f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2)$ is the joint pdf of the channel states for the two transmissions, $f_{\Gamma_j}(\Gamma_j)$ is the pdf of the channel state for the j-th transmission, and $P_{ej}^{cndt}(r, S_j, U_j \mid \Gamma_j)$ is the conditional PER in the j-th transmission, conditioned on channel state Γ_j. The SNR boundaries in the j-th transmission ($\Gamma_j^{(1)}, \Gamma_j^{(2)}, \Gamma_j^{(3)}, \Gamma_j^{(4)}, \Gamma_j^{(5)}$) are obtained from the method in Section 3.3.2. With the SNR boundaries, we have

$$
\begin{aligned}
P_{ej}^{cndt}(r, S_j, U_j \mid \Gamma_j) = \psi_{2^i}(\Gamma_j) = a_i e^{-b_i \Gamma_j} \\
\text{when } \Gamma_j^{(i)} < \Gamma_j < \Gamma_j^{(i+1)}, i = 1, 2, 3, 4
\end{aligned}
\tag{3.10}
$$

where the curve-fitted parameters a_i and b_i depend on r. Let A_j be the average SNR in the j-th transmission, where $A_j = E[\gamma^2 S_j T_s / N_0]$. Then we have

$$
\begin{aligned}
& P^{ave}(N, r, \{S\}, \{U\}) \\
=& \frac{1}{A_1 A_2} \left[\int_0^{\Gamma_1^{(1)}} e^{-\frac{\Gamma_1}{A_1}} d\Gamma_1 + \sum_{i=1}^{4} a_i \int_{\Gamma_1^{(i)}}^{\Gamma_1^{(i+1)}} e^{-(b_i + \frac{1}{A_1})\Gamma_1} d\Gamma_1 \right] \times \\
& \left[\int_0^{\Gamma_2^{(1)}} e^{-\frac{\Gamma_2}{A_2}} d\Gamma_2 + \sum_{i=1}^{4} a_i \int_{\Gamma_2^{(i)}}^{\Gamma_2^{(i+1)}} e^{-(b_i + \frac{1}{A_2})\Gamma_2} d\Gamma_2 \right]
\end{aligned}
\tag{3.11}
$$

The average overall energy per information bit is in Equation (3.12), where $\mathcal{E}_j^{cndt}(r, S_j, U_j \mid \Gamma_j)$ is

the conditional energy consumption during the j-th transmission of the information bit, conditioned on channel state Γ_j.

$$
\begin{aligned}
&\mathcal{E}^{ave}(2, r, \{S\}, \{U\}) \\
=&E_{\{\Gamma\},\{X\}}[\mathcal{E}^{cndt}(2, r, \{S\}, \{U\} \,|\, \{\Gamma\}, \{X\})] \\
=&E_{\{\Gamma\}}[\mathcal{P}(X_1 = 1)\mathcal{E}^{cndt}(2, r, \{S\}, \{U\} \,|\, \{\Gamma\}, X_1 = 1)+ \\
&\mathcal{P}(X_1 = 0)\mathcal{E}^{cndt}(2, r, \{S\}, \{U\} \,|\, \{\Gamma\}, X_1 = 0)] \\
=&E_{\{\Gamma\}}\big[\mathcal{P}(X_1 = 1)\mathcal{E}_1^{cndt}(r, S_1, U_1 \,|\, \Gamma_1)+ \\
&\mathcal{P}(X_1 = 0)\left(\mathcal{E}_1^{cndt}(r, S_1, U_1 \,|\, \Gamma_1) + \mathcal{E}_2^{cndt}(r, S_2, U_2 \,|\, \Gamma_2)\right)\big] \\
=&\int_0^\infty \int_0^\infty \big[\mathcal{E}_1^{cndt}(r, P_1, U_1 \,|\, \Gamma_1)+ \\
&P_{e1}^{cndt}(r, P_1, U_1 \,|\, \Gamma_1)\mathcal{E}_2^{cndt}(r, S_2, U_2 \,|\, \Gamma_2)\big] f_{\Gamma_1,\Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 d\Gamma_2 \\
=&\int_0^\infty \mathcal{E}_1^{cndt}(r, S_1, U_1 \,|\, \Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 + \\
&\int_0^\infty P_{e1}^{cndt}(r, S_1, U_1 \,|\, \Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \int_0^\infty \mathcal{E}_2^{cndt}(r, S_2, U_2 \,|\, \Gamma_2) f_{\Gamma_2}(\Gamma_2) d\Gamma_2 \\
=&\frac{1}{A_1 L r} \sum_{i=1}^{4} \int_{\Gamma_1^{(i)}}^{\Gamma_1^{(i+1)}} \frac{S_1 T_s}{i} e^{-\frac{\Gamma_1}{A_1}} d\Gamma_1 + \\
&\frac{1}{A_1 A_2 L r} \left[\int_0^{\Gamma_1^{(1)}} e^{-\frac{\Gamma_1}{A_1}} d\Gamma_1 + \sum_{i=1}^{4} a_i \int_{\Gamma_1^{(i)}}^{\Gamma_1^{(i+1)}} e^{-(b_i + \frac{1}{A_1})\Gamma_1} d\Gamma_1\right] \left[\sum_{i=1}^{4} \int_{\Gamma_2^{(i)}}^{\Gamma_2^{(i+1)}} \frac{S_2 T_s}{i} e^{-\frac{\Gamma_2}{A_2}} d\Gamma_2\right]
\end{aligned}
\tag{3.12}
$$

With the above expressions, we would like to solve Equation (3.6). However, the problem is non-convex, even when $N = 2$. We use an exhaustive search to find the optimal solution. First, we quantize P_e. For each quantized value of P_e, we then quantize the variables $S_1, ..., S_N, U_1, ..., U_N$ and exhaustively search all possible tuples of $N, r, S_1, ..., S_N, U_1, ..., U_N$ to find the one which gives average PER P_e and has the minimum energy consumption $\mathcal{E}_{min}(P_e)$. We quantize three continuous parameters: P_e, S and U. In the numerical exhaustive search, the step size for $log_{10}(P_e)$ is 0.02, the step size for $10log_{10}(S)$ is 0.02 and the step size for $log_{10}(U)$

is 0.02. The range for P_e is from the minimum achievable PER to 1. The minimum achievable PER depends on the channel statistics, e.g., 4×10^{-3} when $E[\gamma^2 S_0 T_s/N_0] = 1$. The range for S is from 0 to $S_{max} = 2S_0$. The range for U is from 10^{-6} to 1.

Figs. 3.3 and 3.4 show the minimum energy consumption per bit vs. PER for different (N, r) combinations when $E[\gamma^2 S_0 T_s/N_0]$ = 0dB and 10dB, respectively. The energy consumption per bit is normalized by $S_0 T_s$, i.e., the energy consumption per bit (in units of Joules) for BPSK without FEC in a single transmission. Thus, the y-axis is dimensionless. In this paper, we let the available protection combinations be $\{N\} = \{3, 2, 1\}, \{r\} = \{1/5, 1/3, 1/2\}$, so there are 9 possible protections. In both figures, each line color corresponds to a value of N: black for N = 3, green for N = 2, and blue for N = 1, and each line type corresponds to a value of r: dashed for r = 1/5, solid for r = 1/3, and dotted for r = 1/2. Then $\mathcal{E}_{min}(P_e)$ is the minimum among these curves. From both figures, we find that, given N, energy decreases with the decreasing of r, which is due to the coding gain; given r, energy decreases with the increasing of N, which is due to the diversity in the transmissions. The influence of N is larger than the influence of r, especially when the channel SNR is large. No matter what the channel SNR is, $(N = 3, r = 1/5)$ results in the minimum energy due to the best coding gain and diversity in transmissions.

Figs. 3.5 and 3.6 show the power and PER "upper bound" distributions for $(N = 3, r = 1/5)$ when $E[\gamma^2]$ = 10dB. In Fig. 3.5, the y-axis is the ratio of the transmit power to the constant power S_0 in dB. In Fig. 3.6, the y-axis is the parameter PER "upper bound", which determines the alphabet size based on the instantaneous CSI. The first transmission has the least power and the largest PER "upper bound" among the three transmissions, and the opposite holds for the third transmission. The conclusion is if there are three transmission opportunities, the first transmission should be sent at a low cost, because if it happens to succeed, then the second and third transmissions are not necessary and energy consumption is low; however, if the first transmission fails, we need to expend additional power and transmission duration on the second transmission to guarantee the PER; if the second transmission also fails, the third transmission

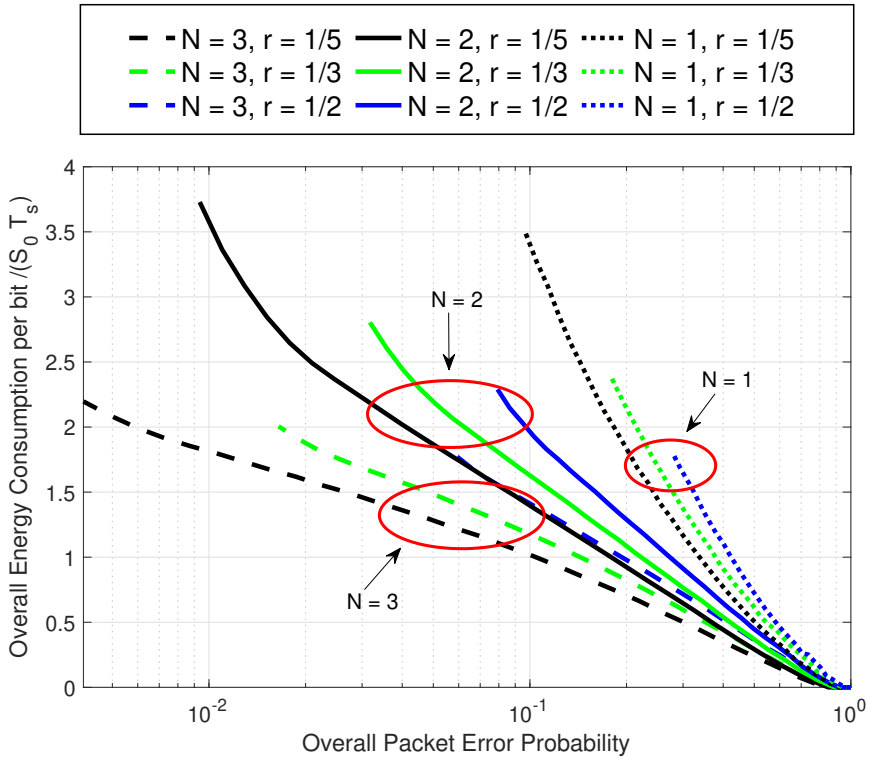

Figure 3.3: Energy consumption for different FEC rates and maximum numbers of transmissions. Average channel SNR is 0dB.

should use the most resources since it is the last transmission opportunity and its transmission failure would result in the failure of the codeword.

3.4.3 Complexity Analysis

Since the algorithm uses an exhaustive search, the complexity of the algorithm is $O(\sum_{n=1}^{N_{max}} |\{r\}| \cdot (M_S M_U)^n) = O(|\{r\}| \cdot \frac{(M_S M_U)^{N_{max}+1}-1}{M_S M_U - 1})$, where $|\{r\}|$ is the number of rate options, and M_S and M_U are the number of discrete values of S and U in the exhaustive search, respectively. This number is exponentially increasing with N_{max}. Note, however, this optimization does not depend on the video and can be done offline. The results for the optimization, i.e., $N, r, \{S\}, \{U\}$, can be stored, and we only need a lookup table during the video transmission.

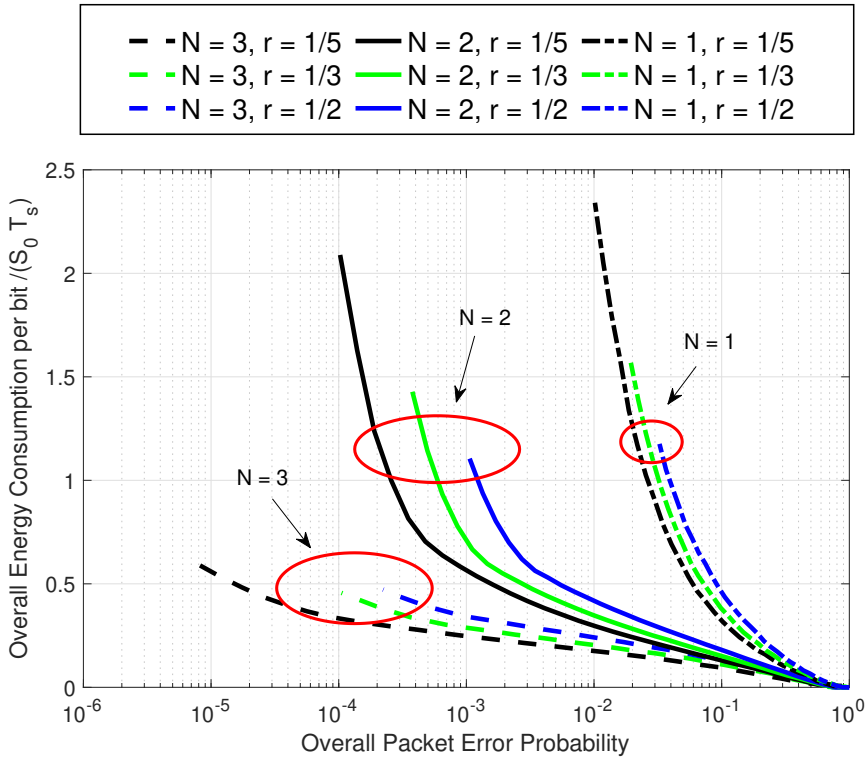

Figure 3.4: Energy consumption for different FEC rates and maximum numbers of transmissions. Average channel SNR is 10dB.

3.5 Proposed UEP

3.5.1 Problem Formulation

After obtaining $\mathcal{E}_{min}(P_{ei})$, we solve the video content UEP problem, since the PERs for each frame are optimized, and thus different frames have unequal error protection. This procedure

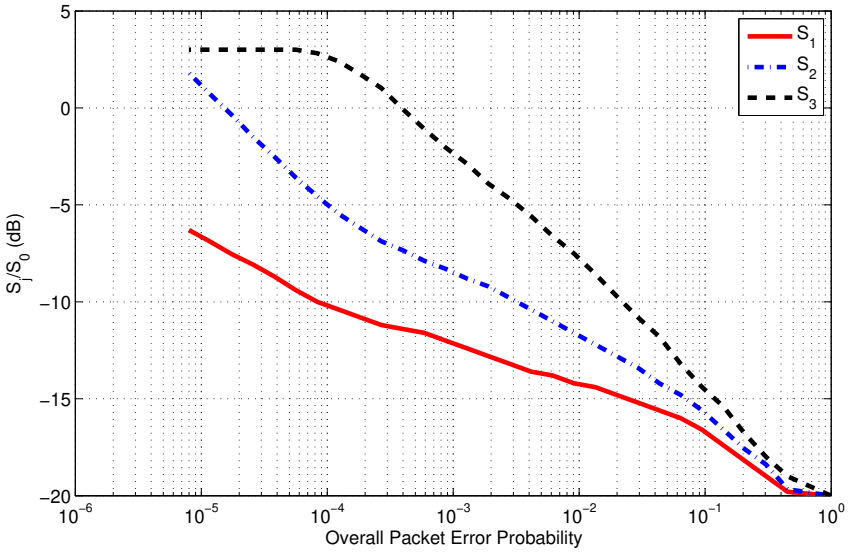

Figure 3.5: Power distribution for $N = 3$ and $r = 1/5$. Average channel SNR is 10dB.

can be formulated as the following problem:

$$\begin{aligned} min \quad & MSE(\{P_{ei}\}) \\ s.t. \quad & \sum_{i=1}^{T} n_i \mathcal{E}_{min}(P_{ei}) = min(\mathcal{E}_c, \tilde{\mathcal{E}}_c) \\ & 0 \leq S_{j,i} \leq S_{max} \quad \text{for } i = 1, 2, ..., T \text{ and } \\ & \qquad\qquad\qquad\qquad\qquad\qquad j = 1, 2, ..., N_i \\ variables: \quad & \{P_{ei}\} \end{aligned} \tag{3.13}$$

Let the average overall PERs for the frames be $P_{e1}, P_{e2}, ..., P_{eT}$. The task is to find the optimal average PER for each frame, where frame i uses energy $\mathcal{E}_{min}(P_{ei})$ to achieve this PER. It is seen in Equation (3.3) that the error in the previous frames can propagate to later frames, so the frames at the front should have lower PER than frames at the end, i.e., the PER sequence should be in ascending order, i.e., $P_{ej} \leq P_{ek}$ where $j \leq k$.

Section 3.4 showed that all frames should use the strongest protection, i.e, $N = 3$ and $r = 1/5$, to minimize the energy consumption. For simplicity of notation, let $g(P_{ei}) = \mathcal{E}_{min}(P_{ei})$.

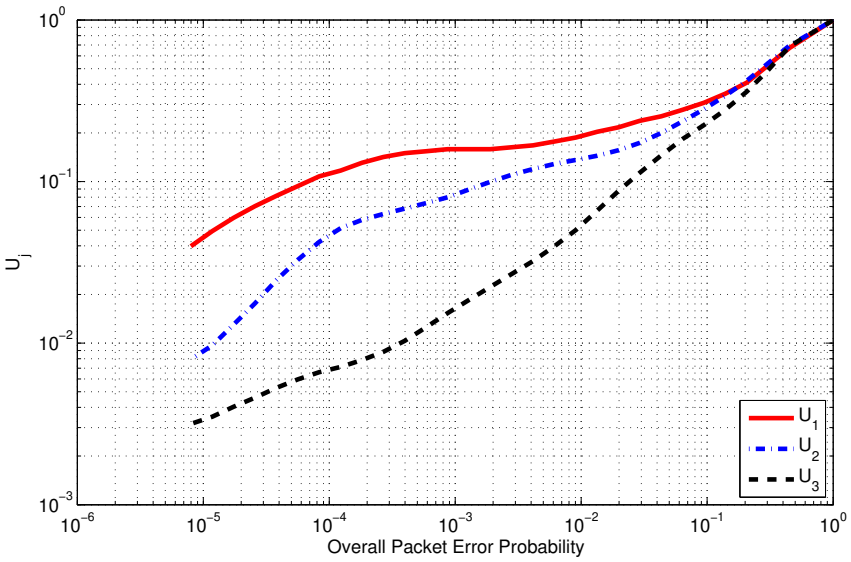

Figure 3.6: PER "upper bound" distribution for $N = 3$ and $r = 1/5$. Average channel SNR is 10dB.

As stated in Section 3.3.3, we know the solution to Equation (3.5) when $\mathcal{E}_c \geq \tilde{\mathcal{E}}_c$: the codeword uses BPSK, FEC rate r_{min}, maximum power S_{max}, and transmits N_{max} times. So we consider the case $\mathcal{E}_c < \tilde{\mathcal{E}}_c$. Then using Equation (3.4), Equation (3.13) can be written as

$$\begin{aligned} min \quad & \sigma_0^2 \sum_{i=1}^{T} P_{ei} \sum_{\tau=0}^{T-i} \frac{1}{1+\alpha\tau} \\ s.t. \quad & \sum_{i=1}^{T} n_i g(P_{ei}) = \mathcal{E}_c \end{aligned} . \tag{3.14}$$

where the constraint on $S_{j,i}$ has been taken into account in the procedure that yields $\mathcal{E}_{min}(P_{ei})$.

We obtained the numerical results of $g(P_{ei})$ through exhaustive search. We approximate the second order derivative by $\frac{g(P_{ei}+h)-2g(P_{ei})+g(P_{ei}-h)}{h^2}$, where h is a sufficiently small value. We find that the approximate second derivative is positive, so $g(P_{ei})$ is convex. Since the objective function in Equation (3.14) is a linear function of P_{ei}, it is a convex function; since $g(P_{ei})$ is convex, the constraint is convex, so this problem is convex and we can use the Lagrange multiplier

method to get the globally optimal solution. The Lagrange function is

$$L = \sigma_0^2 \sum_{i=1}^{T} P_{ei} \sum_{\tau=0}^{T-i} \frac{1}{1+\alpha\tau} + \lambda(\sum_{i=1}^{T} n_i g(P_{ei}) - \mathcal{E}_c) \tag{3.15}$$

Letting $\frac{\partial L}{\partial P_{ei}} = 0$ for $i = 1, 2, ..., T$, we have

$$\sigma_0^2 P_{ei} \sum_{\tau=0}^{T-i} \frac{1}{1+\alpha\tau} + \lambda n_i g'(P_{ei}) = 0, \tag{3.16}$$

where $g'(P_{ei}) \triangleq \frac{g(P_{ei}+h)-g(P_{ei})}{h}$. Letting $\frac{\partial L}{\partial \lambda} = 0$, we have

$$n_i g(P_{ei}) - \mathcal{E}_c = 0. \tag{3.17}$$

Equations (3.16) and (3.17) together give the solution to Equation (3.14).

3.5.2 Complexity Analysis

The complexity of the proposed UEP algorithm is composed of two parts: parameter estimation of σ_0^2 and α, and solving Equations (3.16) and (3.17). The complexity for the first part depends on the algorithm of curve fitting and the complexity for the second part depends on the algorithm of solving the convex problem. This work does not focus on the complexity analysis of these two parts. However, as shown in Section 3.6.3, the UEP algorithm can use fixed parameters without much loss of accuracy and performance and thus can be solved offline. Therefore, the results can be stored, and a lookup table is sufficient during video transmission.

Table 3.2: Simulation parameters.

T	30
number of information bits in a packet	256
T_s	10^{-6}s
S_0	5mW
N_0	3.98×10^{-21}W/Hz
N	3
r	1/5
video size	1024x576
number of frames in video	1903 to 2229
$\mathcal{E}_c$	20mJ or 10mJ

3.6 Numerical Results

3.6.1 Comparison Algorithms

The comparison HARQ scheme [13, 14] uses a single alphabet size and variable power, whereas the proposed HARQ uses variable alphabet sizes and variable power. The comparison UEP is similar to the method in [72], but with a modification that improves its performance. In [72], the authors assume constant power and a single alphabet size without retransmission, and they use three rates of FEC for the frames in a GOP. The video quality depends on the choice of these three values. The number of FEC code rates is limited due to the implementation complexity. We improve upon their method by allowing variable power and variable alphabet size and using the strongest FEC and the maximum number of transmissions for all frames, and assign three levels of energy consumption for the frames. As discussed in Section 3.4, using variable power with a single strong FEC performs better than using constant power and three different FEC values as done in [72]. Since the optimal energy levels are hard to obtain, we determine three non-overlapping ranges by trial and error, and randomly choose the three energy values from these non-overlapping ranges. The best ranges we found are ($1.2E_0$, $1.5E_0$], ($0.8E_0$, $1.2E_0$] and ($0.5E_0$, $0.8E_0$], where E_0 is a parameter dependent on the energy constraint. Since the video

quality depends on the selection of these three values, we repeat this experiment 20 times and show the average Y-PSNR, where Y-PSNR is the peak signal-to-noise ratio (PSNR) for the Y channel in the YUV video. Here, PSNR = $10log_{10}(V_{max}^2/MSE)$, V_{max} is the maximum pixel value, and MSE is the mean square error on the Y channel.

3.6.2 Simulation Environment

We use JM 15.0 software to achieve H.264 video coding. The previous frame is used for error concealment. We use video from a mobile video database [76, 77]. The video scenes are a football game with high motion, a person skating with medium motion, and an interview scene with slow motion. The videos are of size 1024×576 with 1903 frames, 2022 frames and 2229 frames. The video frame rate is 30 fps, so their lengths are approximately 60 to 70 seconds. We repeat the experiment 10 times for statistical convergence.

3.6.3 Discussion of Results

We show the Y-PSNR vs. channel SNR for different cases in Figs. 3.7 to 3.10, where the solid lines correspond to simulation results and dashed lines correspond to theoretical results. In Fig. 3.7, the video is *skate* and the energy constraint for a GOP is $\mathcal{E}_c$ = 20 mJ. The maximum energy that a GOP can possibly use with BPSK, FEC rate r_{min}, maximum power S_{max} and N_{max} transmissions is 75 mJ. For all cases, the Y-PSNR increases with channel SNR. When using the proposed UEP, the gap between the proposed HARQ and comparison HARQ is about 4 dB in the low SNR region, which shows a significant advantage for the proposed HARQ. When using the proposed HARQ, the gap between the proposed UEP and EEP is about 1.6dB in the low SNR region, and the gap between the proposed UEP and comparison UEP is about 0.7dB in the low SNR region. The 0.7dB gap means that it is not desirable to divide the whole GOP into three groups as in the comparison UEP. But some conventional UEP schemes, e.g., those which use

different FEC for different parts of the video data, only allow a few protection levels due to the complexity. However, the proposed UEP is able to assign different protection for each frame, thus achieving higher video quality. In Fig. 3.8, the energy constraint $\mathcal{E}_c$ = 10 mJ. The curves show a similar trend to Fig. 3.7. In both figures, the theoretical analysis and simulation are close except for the low SNR region. This is because the PER is large in the low SNR region, so the model is less accurate because the errors that occur in a frame, but which propagate from different frames, are not uncorrelated.

Figs. 3.9 and 3.10 show the Y-PSNR for *football* and *interview* with energy constraint $\mathcal{E}_c$ = 20 mJ. Comparing Figs. 3.9 and 3.10 to 3.7, we find that for a given scheme, Y-PSNR depends on the video motion. With higher motion, Y-PSNR is lower because frame copy is less effective for high motion video. The difference between UEP and EEP, which is defined as the UEP gain, depends on video motion. With higher motion, the gain is smaller. However, the difference between the proposed HARQ and comparison HARQ is almost the same for all videos.

Next we evaluate the influence of the parameters σ_0^2 and α, which depend on the video, on the UEP gain. Let the PER sequence with respect to EEP be $P_{e0}, P_{e0}, ..., P_{e0}$, and let the PER sequence with respect to UEP be $P_{e1}, P_{e2}, ..., P_{eT}$. Then the UEP gain (dB) is

$$10log_{10}\left(\frac{\sigma_0^2\sum_{i=1}^{T}P_{e0}\sum_{\tau=0}^{T-i}\frac{1}{1+\alpha\tau}}{\sigma_0^2\sum_{i=1}^{T}P_{ei}\sum_{\tau=0}^{T-i}\frac{1}{1+\alpha\tau}}\right)=10log_{10}\left(\frac{P_{e0}\sum_{i=1}^{T}\sum_{\tau=0}^{T-i}\frac{1}{1+\alpha\tau}}{\sum_{i=1}^{T}P_{ei}\sum_{\tau=0}^{T-i}\frac{1}{1+\alpha\tau}}\right) \tag{3.18}$$

From Equations (3.16) and (3.17), we find that the unequal PER sequence $P_{e1}, P_{e2}, ..., P_{eT}$ does not depend on σ_0^2, since σ_0^2 is absorbed by the Lagrangian multiplier λ. Thus, the UEP gain in Equation (3.18) only depends on α. The α values for *football*, *skate* and *interview* are 0.18, 0.0065, 0.005, respectively. In Fig. 3.11, the blue solid curve shows the UEP gain vs. video parameter α. We curve fit the parameter α based on the video and use this correct value for the problem solving. The gain decreases with the increase of α, and high motion video has large α, so the gain is smaller for high motion video. This is also demonstrated from Figs. 3.7 to 3.10.

The red dashed curve shows the UEP gain vs. video parameter α, but fixed $\alpha = 0.1$ is used for problem solving. There is a small loss due to unmatched parameter α for low motion video, and the loss is negligible for medium and high motion video.

Comparing Figs. 3.7, 3.9 and 3.10, we find that the performance gain for the proposed HARQ over the comparison HARQ is similar for videos with different degrees of motion. This can be demonstrated by Equation (3.4). Supposing EEP is used, we let the PER sequence with respect to the comparison HARQ be $P_{e0}, P_{e0}, ..., P_{e0}$, and let the PER sequence with respect to the proposed HARQ be $P'_{e0}, P'_{e0}, ..., P'_{e0}$. Then the HARQ gain (dB) when using EEP is

$$10log_{10}\left(\frac{\sigma_0^2\sum_{i=1}^{T}P_{e0}\sum_{\tau=0}^{T-i}\frac{1}{1+\alpha\tau}}{\sigma_0^2\sum_{i=1}^{T}P'_{e0}\sum_{\tau=0}^{T-i}\frac{1}{1+\alpha\tau}}\right) = 10log_{10}\left(\frac{P_{e0}}{P'_{e0}}\right) \tag{3.19}$$

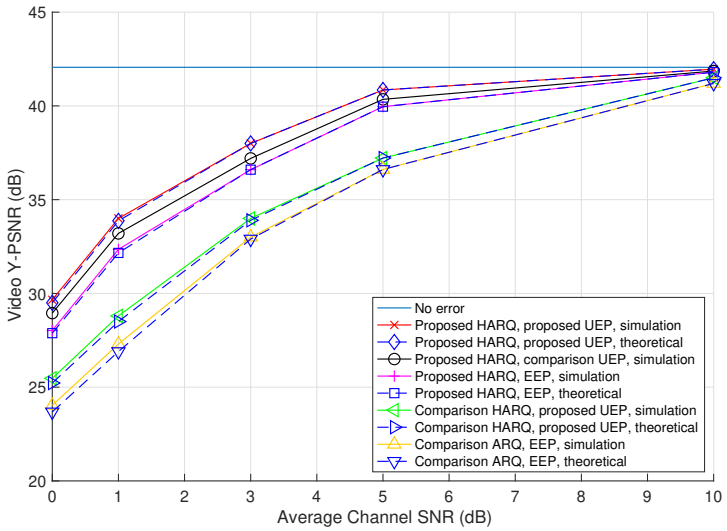

Figure 3.7: Y-PSNR for *skate*. High energy constraint.

This gain does not depend on σ_0^2 and α. When using UEP, this gain is also largely independent of σ_0^2 and α according to numerical results. Thus, the gain of the proposed HARQ system is almost constant for different videos, and fixed parameters provide close-to-optimal performance. This means the proposed algorithm has the important advantage that if optimization

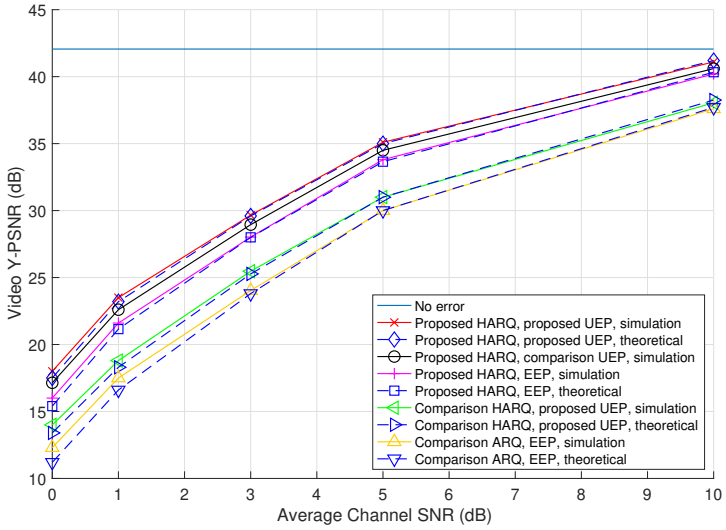

Figure 3.8: Y-PSNR for *skate*. Low energy constraint.

results are obtained offline, then minimal computation is caused by the proposed algorithm during transmission. This reduces the computational power and delay, and allows the algorithm to be real-time.

3.7 Conclusion

In this chapter, we investigate energy-optimized wireless video transmission employing HARQ and AMC. We consider both the unequal importance for multiple transmissions of the same video content, and the unequal importance of different video contents. We divide the problem into two sub-problems which solve the unequal importance for multiple transmissions and solve the UEP for different video content, respectively. The unequal importance for multiple transmissions shows that the transmissions should be sent with increasing order of cost to minimize the overall energy consumption. The video UEP shows that the frames in a GOP should have increasing order of packet error rate to minimize the MSE in the GOP. We compare the proposed scheme to the HARQ that uses single alphabet size and variable power, whereas the proposed scheme uses

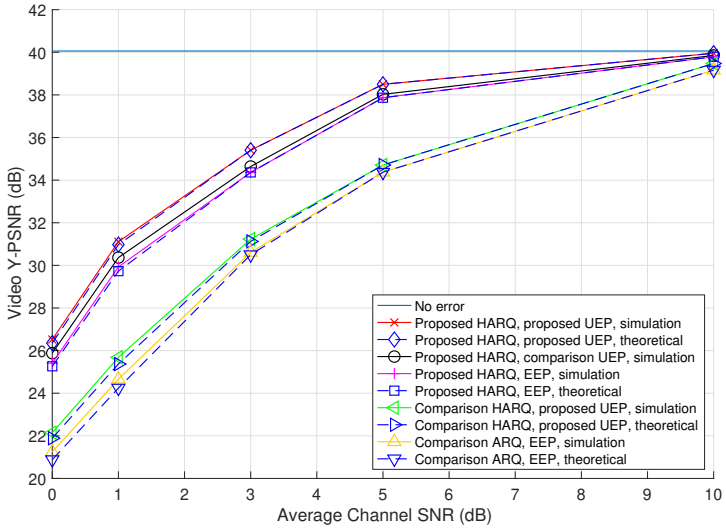

Figure 3.9: Y-PSNR for *football*. High energy constraint.

variable alphabet size and variable power. Compared to the video UEP scheme which is only allowed to have three (or few) levels of protections, the proposed scheme is able to assign optimal protection for each frame in the GOP. Simulations show that in the low SNR region, the proposed scheme outperforms the comparison HARQ by about 4dB, and outperforms the comparison video UEP scheme by 0.8 to 1.6dB depending on the video motion. There is a total gain of 4.8 to 5.6dB compared to video transmission using conventional HARQ without any video UEP.

Chapter 3, in part, is a reprint of the material as it appears in the papers: B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization for Wireless Video Transmission Employing Hybrid ARQ," *IEEE Transactions on Vehicular Technology (2019)*, and B. Zhang, P. Cosman and L. B. Milstein, "Energy Optimization for Hybrid-ARQ and AMC" In *2017 51st Asilomar Conference on Signals, Systems, and Computers. IEEE, 2017*. The dissertation author was the primary investigator and author of these papers.

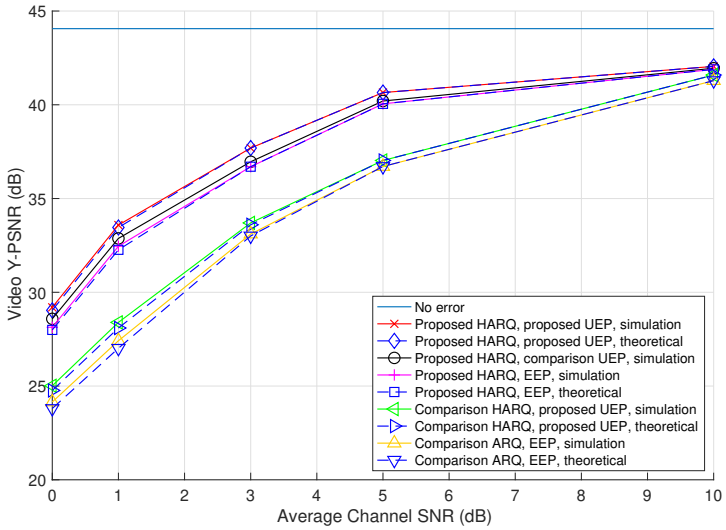

Figure 3.10: Y-PSNR for *interview*. High energy constraint.

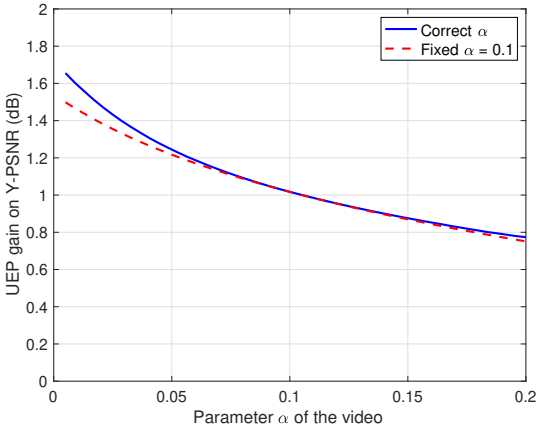

Figure 3.11: UEP gain on Y-PSNR vs. α.

Chapter 4

Conclusion and Future work

In this dissertation, we study energy optimizatioThe influence of different CSI availabilities at the transmitter are investigated and different combining techniques are compared. n for turbo-coded HARQ. First, we consider rate allocation and rate adaptation for IR HARQ. We minimize the energy consumption of HARQ, subject to a packet loss rate constraint. Numerical results show that the energy consumption of HARQ decreases when more CSI information is available at the transmitter. When the maximum number of transmissions is three, at an average channel SNR of 4dB, IR combining with both current and previous CSI consumes 10% less energy than IR combining without CSI. We also compare IR combining with both Chase combining and the system without combining, and IR combining yields the least energy consumption.

Second, we investigate energy-optimized wireless video transmission employing HARQ and AMC. We consider both the unequal importance for multiple transmissions of the same video content, and the unequal importance of different video contents. We divide the problem into two sub-problems which solve the unequal importance for multiple transmissions and solve the UEP for different video content, respectively. The video UEP shows that the frames in a GOP should have increasing order of packet error rate to minimize the MSE in the GOP. We compare the proposed scheme to the HARQ that uses single alphabet size and variable power,

whereas the proposed scheme uses variable alphabet size and variable power. Compared to the video UEP scheme which is only allowed to have three (or fewer) levels of protections, the proposed scheme is able to assign optimal protection for each frame in the GOP. Simulations show that in the low SNR region, the proposed scheme outperforms the comparison HARQ by about 4dB, and outperforms the comparison video UEP scheme by 0.8 to 1.6dB depending on the video motion. There is a total gain of 4.8 to 5.6dB compared to video transmission using conventional HARQ without any video UEP. For both IR HARQ systems and the HARQ systems without combining, a key factor to reduce energy consumption in both systems is unequal energy allocation among the multiple transmissions. Early transmissions should consume less energy and later transmissions should consume more energy. If the system is lucky in the cheap early transmissions, the later expensive transmissions are avoided and energy is saved. Therefore, the energy consumption significantly decreases when the maximum number of transmissions is larger: having more transmission opportunities allows the system to consume less energy in early transmissions, and thus saves energy.

The extension of this work might consider the following:

1) Imperfect CSI. This may occur due to many reasons, such as imperfect channel estimation at the receiver, discretized CSI feedback, and the difficulty of obtaining accurate current CSI at the transmitter. The last one happens when the round trip time is larger than or comparable to the channel coherence time.

2) IR HARQ with AMC. Multiple parameters can be variable, such as power and alphabet size.

3) Applying HARQ to multiple parallel channels, such as multiple-input multiple-output.

Appendix A

Error Probability of IR HARQ

A.1 Approximation

The error probability of IR HARQ is the probability that both transmissions fail, and we approximate it with the probability that the second transmission fails (regardless of the result in the first transmission). Let e_1 and e_2 be the events that the decoding fails after the first and second transmission, respectively. The error probability of the HARQ system is $P_{HARQ} = P(e_1 \cap e_2)$, where $\cap$ denotes the intersection of sets. We have $e_2 = (e_1 \cup \overline{e}_1) \cap e_2 = (e_1 \cap e_2) \cup (\overline{e}_1 \cap e_2)$, where $\cup$ denotes the union of sets, and $\overline{e_i}$ is the complement of e_i, i.e., the decoding is successful after the i-th transmission. Then $P(e_2) = P(e_1 \cap e_2) + P(\overline{e}_1 \cap e_2)$. Thus, $P_{HARQ} = P(e_1 \cap e_2) = P(e_2) - P(\overline{e}_1 \cap e_2)$. The term $P(e_2)$ is the probability that the second transmissions fails, irrespective of the result in the first transmission, and $P(\overline{e}_1 \cap e_2)$ is the probability that the first transmission is successfully decoded, whereas the second transmission fails. Note that although the event $\overline{e}_1 \cap e_2$ does not actually happen in the HARQ system, since there is no need for the second transmission if the first one succeeds, the above equations still hold. Now we need to show that the probability of this event is small, i.e., $P(\overline{e}_1 \cap e_2) << P(e_2)$, so that $P_{HARQ} \approx P(e_2)$. We would like to justify this approximation using simulation results and an information-theoretic

approach.

Fig. A.1 shows the simulated $P(e_2)$ and $P(\overline{e}_1 \cap e_2)$ where the FEC code rate in the second transmission is r_2 = 1/ 3. We consider the three cases of r_1 = 2 / 5, r_1 = 1 / 2, and r_1 = 2 / 3. For all the cases, $P(\overline{e}_1 \cap e_2)$ is two to three orders of magnitude smaller than $P(e_2)$. We also examined $r_2 = 1/2, 2/5$ and $1/5$, and the results were similar.

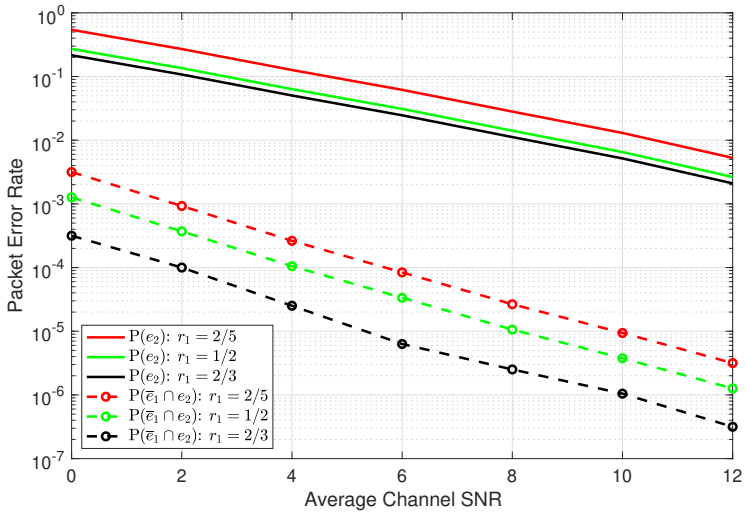

Figure A.1: Simulated Error Probability.

In [14, 18–20, 38, 78], the authors use an information-theoretic approach to study HARQ. The analysis in these papers is based on [21], where the assumption is that the number of bits in each transmission round is sufficiently large. In this approach, the transmission succeeds if the accumulated mutual information is larger than a threshold, and fails otherwise. Therefore, if the first transmission succeeds, the accumulated mutual information is already larger than the threshold, and the second transmission must also succeed because the accumulated mutual information does not decrease. So the error probability for the HARQ system is the probability of error in the final transmission round.

A.2 Analytical Expression

There is no analytical expression for $P(e_2)$ because it is the error probability of a turbo code in which the first N_1 bits and the next N_2 bits experience independent channels. We use the following expression to approximate

$$P(e_2; r_1, r_2|\Gamma_1, \Gamma_2) \approx \min\left(1, a_2 e^{-b_{2,1}\Gamma_1} e^{-b_{2,2}\Gamma_2} e^{-c_2\left(\frac{1}{\Gamma_1}+\frac{1}{\Gamma_2}\right)^{-1}}\right), \tag{A.1}$$

where $a_2, b_{2,1}, b_{2,2}$ and c_2 are positive and obtained through curve fitting. The first two exponential terms correspond to the first two transmissions by themselves and are similar to Equation (2.9). The third exponential term corresponds to the correlation between two transmissions. The reason for this term is that the channel state in each transmission affects the decoding of other transmissions because of incremental redundancy. This term becomes smaller as either Γ_1 or Γ_2 increases and is 1 when either Γ_1 or Γ_2 is 0.

For a maximum number of K transmissions, we use the following expression

$$P(e_K; r_1, r_2, ..., r_K|\Gamma_1, \Gamma_2, ..., \Gamma_K) \approx \min\left(1, a_K\left(\prod_{i=1}^{K} e^{-b_{K,i}\Gamma_i}\right) e^{-c_K\left(\sum_{i=1}^{K}\frac{1}{\Gamma_i}\right)^{-1}}\right). \tag{A.2}$$

The accuracy of the curve fitting is summarized in the following table, where RMSE is defined as the root mean square error of $\log_{10}(P(e_K; r_1, r_2, ..., r_K|\Gamma_1, \Gamma_2, ..., \Gamma_K))$.

Table A.1: RMSE of curve fitting.

	RMSE	adjusted R squared
$K = 2$	0.110	0.9814
$K = 3$	0.115	0.9810

Figs. A.2 to A.6 show an example of the curve fitting for $K = 2$, where $r_1 = 2/5$ and $r_2 = 1/3$. The dots represent the simulated PER, and the curve represents Equation (A.1) with different values of Γ_2.

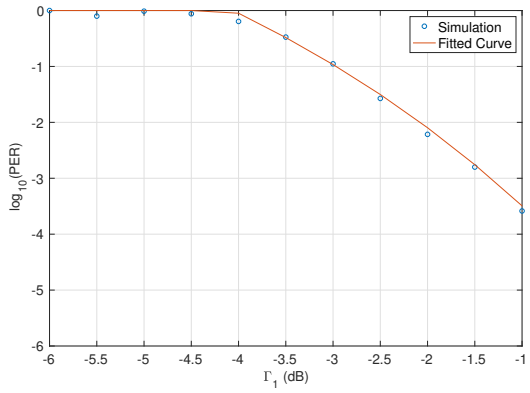

Figure A.2: Error probability curve fitting for $K = 2$, where $r_1 = 2/5$ and $r_2 = 1/3$. $\Gamma_2 = -6dB$.

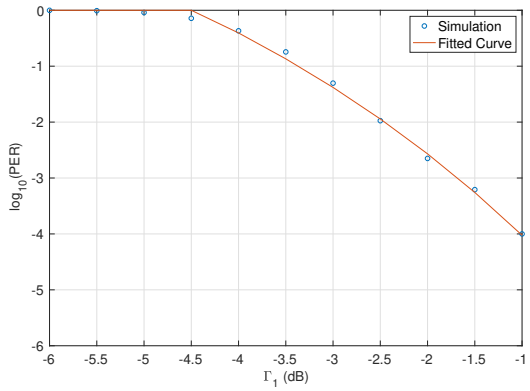

Figure A.3: Error probability curve fitting for $K = 2$, where $r_1 = 2/5$ and $r_2 = 1/3$. $\Gamma_2 = -4dB$.

Since the fit is good, we use Equations (A.1) and (A.2) to approximate the error probability of an IR HARQ system, which allows us to numerically evaluate the energy consumption and packet loss rate.

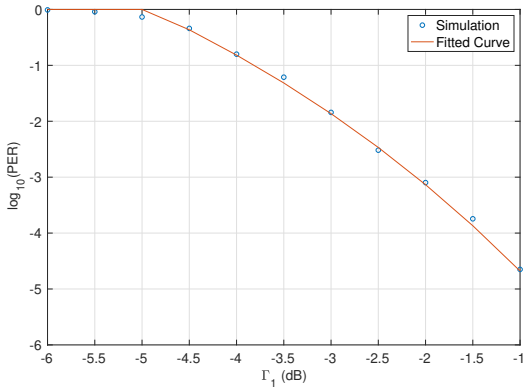

Figure A.4: Error probability curve fitting for $K = 2$, where $r_1 = 2/5$ and $r_2 = 1/3$. $\Gamma_2 = -2dB$.

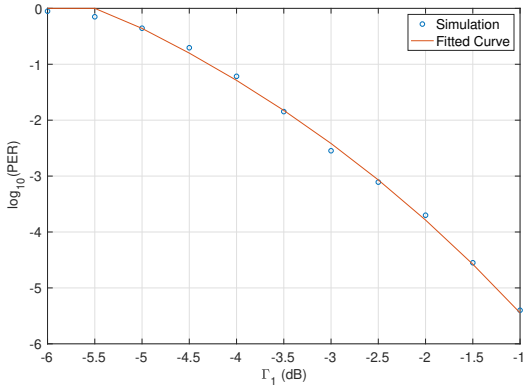

Figure A.5: $\Gamma_2 = 0dB$.

Figure A.6: Error probability curve fitting for $K = 2$, where $r_1 = 2/5$ and $r_2 = 1/3$. $\Gamma_2 = 0dB$.

Bibliography

[1] D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Prentice Hall, 1983.

[2] M. Tawalbeh, A. Eardley *et al.*, "Studying the energy consumption in mobile devices," *Procedia Computer Science*, vol. 94, pp. 183–189, 2016.

[3] R. Verdone, D. Dardari, G. Mazzini, and A. Conti, *Wireless sensor and actuator networks: technologies, analysis and design*. Academic Press, 2010.

[4] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes." in *ICC'93-IEEE International Conference on Communications*, vol. 2. IEEE, 1993, pp. 1064–1070.

[5] 3GPP TS 36.212 v8.0.0 (2007-09), "Multiplexing and channel coding (FDD) (Release 8)."

[6] X. Zhang and Q. Du, "Adaptive low-complexity erasure-correcting code-based protocols for QoS-driven mobile multicast services over wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 5, pp. 1633–1647, 2006.

[7] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-distortion optimized hybrid error control for real-time packetized video transmission," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 40–53, 2006.

[8] A. M. Cipriano, P. Gagneur, G. Vivier, and S. Sezginer, "Overview of ARQ and HARQ in beyond 3G systems," in *2010 IEEE 21st International Symposium on Personal, Indoor and Mobile Radio Communications Workshops*. IEEE, 2010, pp. 424–429.

[9] G. Benelli, "An ARQ scheme with memory and soft error detectors," *IEEE Transactions on Communications*, vol. 33, no. 3, pp. 285–288, 1985.

[10] D. Chase, "Code combining—A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Transactions on Communications*, vol. 33, no. 5, pp. 385–393, 1985.

[11] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with chase combining and incremental redundancy for HSDPA," in *IEEE 54th Vehicular Technology Conference. VTC Fall 2001. Proceedings (Cat. No. 01CH37211)*, vol. 3. IEEE, 2001, pp. 1829–1833.

[12] T. Villa, R. Merz, R. Knopp, and U. Takyar, "Adaptive modulation and coding with hybrid-ARQ for latency-constrained networks," in *Wireless Conference (European Wireless), 2012 18th European*. VDE, 2012, pp. 1–8.

[13] W. Su, S. Lee, D. A. Pados, and J. D. Matyjas, "Optimal power assignment for minimizing the average total transmission power in hybrid-ARQ Rayleigh fading links," *IEEE Transactions on Communications*, vol. 59, no. 7, pp. 1867–1877, 2011.

[14] T. V. Chaitanya and E. G. Larsson, "Optimal power allocation for hybrid ARQ with chase combining in i.i.d. Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 61, no. 5, pp. 1835–1846, 2013.

[15] C. Ji, J. Cao, and G. Zhang, "Optimization of power allocation for chase combining hybrid ARQ," *IEICE Transactions on Communications*, vol. 102, no. 3, pp. 613–622, 2019.

[16] C. Shen, T. Liu, and M. P. Fitz, "On the average rate performance of hybrid-ARQ in quasi-static fading channels," *IEEE Transactions on Communications*, vol. 57, no. 11, 2009.

[17] C. Ji, D. Wang, N. Liu, and X. You, "On power allocation for incremental redundancy Hybrid ARQ," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1506–1518, 2015.

[18] L. Szczecinski, C. Correa, and L. Ahumada, "Variable-rate transmission for incremental redundancy hybrid ARQ," in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*. IEEE, 2010, pp. 1–5.

[19] M. Jabi, L. Szczecinski, M. Benjillali, and F. Labeau, "Outage minimization via power adaptation and allocation in truncated hybrid ARQ," *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 711–723, 2015.

[20] L. Szczecinski, S. R. Khosravirad, P. Duhamel, and M. Rahman, "Rate allocation and adaptation for incremental redundancy truncated HARQ," *IEEE Transactions on Communications*, vol. 61, no. 6, pp. 2580–2590, 2013.

[21] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Transactions on Information Theory*, vol. 47, no. 5, pp. 1971–1988, 2001.

[22] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1746–1755, 2004.

[23] A. K. Karmokar, D. V. Djonin, and V. K. Bhargava, "Delay constrained rate and power adaptation over correlated fading channels," in *2004 IEEE Global Telecommunications Conference, GLOBECOM'04*, vol. 6. IEEE, 2004, pp. 3448–3453.

[24] J. Hayes, "Adaptive feedback communications," *IEEE Transactions on Communication Technology*, vol. 16, no. 1, pp. 29–34, 1968.

[25] J. Cavers, "Variable-rate transmission for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 20, no. 1, pp. 15–22, 1972.

[26] B. Vucetic, "An adaptive coding scheme for time-varying channels," *IEEE Transactions on Communications*, vol. 39, no. 5, pp. 653–663, 1991.

[27] W. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Transactions on Communications*, vol. 43, no. 7, pp. 2223–2230, 1995.

[28] W. Hong, Y. ShuYa, H. Ning, and Z. Liang, "A cross-layer design combining method of AMC with HARQ based on LDPC codes," in *2008 International Conference on Communications, Circuits and Systems*. IEEE, 2008, pp. 355–359.

[29] A. A. Khalek, C. Caramanis, and R. W. Heath, "A cross-layer design for perceptual optimization of H. 264/SVC with unequal error protection," *IEEE Journal on selected areas in Communications*, vol. 30, no. 7, pp. 1157–1171, 2012.

[30] C. V. N. Index, "Global mobile data traffic forecast update, 2017–2022 white paper," *Cisco: San Jose, CA, USA*, 2017.

[31] E. Visotsky, Y. Sun, V. Tripathi, M. L. Honig, and R. Peterson, "Reliability-based incremental redundancy with convolutional codes," *IEEE Transactions on communications*, vol. 53, no. 6, pp. 987–997, 2005.

[32] P. Larsson, L. K. Rasmussen, and M. Skoglund, "Throughput analysis of hybrid-ARQ—A matrix exponential distribution approach," *IEEE Transactions on Communications*, vol. 64, no. 1, pp. 416–428, 2016.

[33] B. Makki and T. Eriksson, "On hybrid ARQ and quantized CSI feedback schemes in quasi-static fading channels," *IEEE Transactions on Communications*, vol. 60, no. 4, pp. 986–997, 2012.

[34] H. Jin, C. Cho, N.-O. Song, and D. K. Sung, "Optimal rate selection for persistent scheduling with HARQ in time-correlated Nakagami-m fading channels," *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 637–647, 2011.

[35] H. Zhuang and V. Sethuraman, "Hybrid ARQ: Does fading diversity help?" *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 210–213, 2017.

[36] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of the incremental redundancy HARQ," *IEEE Wireless Communications Letters*, vol. 3, no. 5, pp. 529–532, 2014.

[37] J. S. Harsini, F. Lahouti, M. Levorato, and M. Zorzi, "Analysis of non-cooperative and cooperative type II hybrid ARQ protocols with AMC over correlated fading channels," *IEEE Transactions on Wireless Communications*, vol. 10, no. 3, pp. 877–889, 2011.

[38] A. Chelli and M.-S. Alouini, "Performance of hybrid-ARQ with incremental redundancy over relay channels," in *2012 Globecom Workshops (GC Wkshps)*. IEEE, 2012, pp. 116–121.

[39] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: understanding broadband wireless networking*. Pearson Education, 2007.

[40] M. Jabi, M. Benjillali, L. Szczecinski, and F. Labeau, "Energy efficiency of adaptive HARQ." *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 818–831, 2016.

[41] N.-S. Vo, T. Q. Duong, H.-J. Zepernick, and M. Fiedler, "A cross-layer optimized scheme and its application in mobile multimedia networks with QoS provision," *IEEE Systems Journal*, vol. 10, no. 2, pp. 817–830, 2016.

[42] B. Zhang, P. C. Cosman, and L. Milstein, "Energy optimization for wireless video transmission employing hybrid ARQ," *IEEE Transactions on Vehicular Technology*, 2019.

[43] A. Aqil, A. O. Atya, S. V. Krishnamurthy, and G. Papageorgiou, "Streaming lower quality video over LTE: How much energy can you save?" in *2015 IEEE 23rd International Conference on Network Protocols (ICNP)*. IEEE, 2015, pp. 156–167.

[44] M. E. El Dien, A. A. Youssif, and A. Z. Ghalwash, "Energy efficient and QoS aware framework for video transmission over wireless sensor networks," *Wireless Sensor Network*, vol. 8, no. 03, p. 25, 2016.

[45] T. Maksymyuk, L. Han, X. Ge, H.-H. Chen, and M. Jo, "Quasi-quadrature modulation method for power-efficient video transmission over LTE networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 5, pp. 2083–2092, 2014.

[46] J. Wu, B. Cheng, M. Wang, and J. Chen, "Energy-efficient bandwidth aggregation for delay-constrained video over heterogeneous wireless networks." *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 1, pp. 30–49, 2017.

[47] Q. Wang, T. M. Steinman, and W. Wang, "Quality driven modulation rate optimization for energy efficient wireless video relays," *Computer Communications*, vol. 115, pp. 2–9, 2018.

[48] C. Singhal, S. De, R. Trestian, and G.-M. Muntean, "Joint optimization of user-experience and energy-efficiency in wireless multimedia broadcast," *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, pp. 1522–1535, 2014.

[49] T. Q. Duong, N.-S. Vo, T.-H. Nguyen, M. Guizani, and L. Shu, "Energy-aware rate and description allocation optimized video streaming for mobile D2D communications," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 6791–6796.

[50] L. Liu, Y. Yi, J.-F. Chamberland, and J. Zhang, "Energy-efficient power allocation for delay-sensitive multimedia traffic over wireless systems." *IEEE Transactions on Vehicular Technology*, vol. 63, no. 5, pp. 2038–2047, 2014.

[51] Q. Wang, W. Wang, S. Jin, and H. Zhu, "Cross-layer source-channel control for future wireless multimedia services: energy, latency, and quality investigation," *IET Communications*, vol. 11, no. 17, pp. 2575–2584.

[52] X. Ge, X. Huang, Y. Wang, M. Chen, Q. Li, T. Han, and C.-X. Wang, "Energy-efficiency optimization for MIMO-OFDM mobile multimedia communication systems with QoS constraints," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 5, pp. 2127–2138, 2014.

[53] Z. Ye, R. Hegazy, W. Zhou, P. Cosman, and L. Milstein, "Joint energy optimization of video encoding and transmission," in *2018 Picture Coding Symposium (PCS)*. IEEE, 2018, pp. 116–120.

[54] A. A. Khalek, C. Caramanis, and R. Heath, "A cross-layer design for perceptual optimization of H.264/SVC with unequal error protection," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 7, pp. 1157–1171, 2012.

[55] R. Xiong, D. S. Taubman, and V. Sivaraman, "PET protection optimization for streaming scalable videos with multiple transmissions," *IEEE Transactions on Image Processing*, vol. 22, no. 11, pp. 4364–4379, 2013.

[56] M. van der Schaar and D. S. Turaga, "Cross-layer packetization and retransmission strategies for delay-sensitive wireless multimedia transmission," *IEEE Transactions on Multimedia*, vol. 9, no. 1, pp. 185–197, 2007.

[57] H.-C. Wei, Y.-C. Tsai, and C.-W. Lin, "Prioritized retransmission for error protection of video streaming over WLANs," in *2004 International Symposium on Circuits and Systems, ISCAS'04*, vol. 2. IEEE, 2004, pp. II–65.

[58] T. P. Fowdur, D. Indoonundon, and S. K. Soyjaudah, "An enhanced framework for H.264 video transmission with joint prioritisation of retransmission and concealment order," in *2014 9th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)*. IEEE, 2014, pp. 634–639.

[59] V. Vadori, A. V. Guglielmi, and L. Badia, "Markov analysis of video transmission based on differential encoded HARQ," in *2016 IEEE 17th International Symposium on World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2016, pp. 1–9.

[60] B. Tirouvengadam, R. Radhakrishnan, and A. Nayak, "CAAHR: Content aware adaptive HARQ retransmission scheme for 4G/LTE network," in *2012 Fourth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2012, pp. 456–461.

[61] A. Rapaport, W. Liu, L. Ma, G. S. Sternberg, A. Ziera, and A. Balasubramanian, "Adaptive HARQ and scheduling for video over LTE," in *2013 Asilomar Conference on Signals, Systems and Computers*. IEEE, 2013, pp. 1584–1588.

[62] C. Zhu, Y. Huo, B. Zhang, R. Zhang, M. El-Hajjar, and L. Hanzo, "Adaptive-truncated-HARQ-aided layered video streaming relying on interlayer FEC coding," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1506–1521, 2016.

[63] D. T. Nguyen and J. Ostermann, "Congestion control for scalable video streaming using the scalability extension of H.264/AVC," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 2, pp. 246–253, 2007.

[64] T. Gan, L. Gan, and K.-K. Ma, "Reducing video-quality fluctuations for streaming scalable video using unequal error protection, retransmission, and interleaving," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 819–832, 2006.

[65] L. Han, S.-S. Kang, H. Kim, and H. P. In, "Adaptive retransmission scheme for video streaming over content-centric wireless networks," *IEEE Communications Letters*, vol. 17, no. 6, pp. 1292–1295, 2013.

[66] H. Mukhtar, A. Al-Dweik, M. Al-Mualla, and A. Shami, "Low complexity power optimization algorithm for multimedia transmission over wireless networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 1, pp. 113–124, 2015.

[67] R. Radhakrishnan and A. Nayak, "Cross layer design for efficient video streaming over LTE using scalable video coding," in *Communications (ICC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 6509–6513.

[68] J. Yang, Q. Zheng, H. Xi, and L. Hanzo, "Receiver-driven adaptive enhancement layer switching algorithm for scalable video transmission over link-adaptive networks," *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 47–50, 2013.

[69] H. R. Ghaeini, B. Akbari, B. Barekatain, and A. Trivino-Cabrera, "Adaptive video protection in large scale peer-to-peer video streaming over mobile wireless mesh networks," *International Journal of Communication Systems*, vol. 29, no. 18, pp. 2580–2603, 2016.

[70] J.-H. Lee, G.-S. Hong, Y.-W. Lee, C.-K. Kim, N. Park, and B.-G. Kim, "Design of efficient key video frame protection scheme for multimedia internet of things (IoT) in converged 5G network," *Mobile Networks and Applications*, pp. 1–13, 2018.

[71] W. Wang and A. Kwasinski, "Adaptive learning for scalable video transmission with HARQ over dynamic wireless channels," in *Communications (ICC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3094–3099.

[72] F. Marx and J. Farah, "A novel approach to achieve unequal error protection for video transmission over 3G wireless networks," *Signal Processing: Image Communication*, vol. 19, no. 4, pp. 313–323, 2004.

[73] B. Girod and N. Fiarber, "Compressed video over networks," *Telecommunications Laboratory, University of Erlangen, Nuremberg*, 1999.

[74] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1012–1032, 2000.

[75] B. Zhang, P. Cosman, and L. Milstein, "Energy optimization for hybrid-ARQ and AMC," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2017, pp. 1796–1800.

[76] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A subjective and objective study of stalling events in mobile streaming videos," *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[77] J. P. D. Ghadiyaram and A. C. Bovik, "LIVE Mobile Stall Video Database-II." [Online]. Available: http://live.ece.utexas.edu/research/LIVEStallStudy/index.html

[78] N. Gopalakrishnan and S. Gelfand, "Rate selection algorithms for IR hybrid ARQ," in *2008 IEEE Sarnoff Symposium*. IEEE, 2008, pp. 1–6.