UC San Diego
UC San Diego Previously Published Works

Title

Energy Optimization For Hybrid ARQ With Turbo Coding: Rate Adaptation and Allocation

Permalink

https://escholarship.org/uc/item/6t01p16w

Journal

IEEE Transactions on Vehicular Technology, 69(10)

ISSN

0018-9545

Authors

Zhang, Bentao
Milstein, Laurence B
Cosman, Pamela

Publication Date

2020

DOI

10.1109/tvt.2020.3009681

Peer reviewed

Energy Optimization For Hybrid ARQ With Turbo Coding: Rate Adaptation and Allocation

Bentao Zhang , Laurence B. Milstein, *Fellow, IEEE*, and Pamela Cosman , *Fellow, IEEE*

Abstract—We consider incremental redundancy (IR) hybrid automatic repeat request (HARQ) over independent block-fading channels with turbo coding. 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 Chase combining and the system without combining, and IR combining yields the least energy consumption.

Index Terms—Hybrid ARQ, turbo code, rate adaptation, rate allocation.

I. INTRODUCTION

HYBRID automatic repeat request (HARQ) [1] plays an important role in providing reliable and efficient data transmission. HARQ is a combination of forward error correction (FEC) and automatic repeat request (ARQ). Pure FEC may introduce unnecessary redundancy, whereas pure ARQ may require many retransmissions due to heavy losses for each single transmission. The authors in [2]–[4] suggest that HARQ outperforms pure FEC and pure ARQ for wireless video transmission. In typical HARQ systems, a retransmission is performed until either the codeword is successfully decoded, or a maximum number of transmissions is reached. There are three kinds of HARQ combining techniques: no combining, Chase combining (CC) and incremental redundancy (IR). For no combining HARQ, each transmission round is decoded independently. For CC HARQ [5], [6], all transmission rounds are identical, and the received packets are decoded together through maximum ratio combining. For IR HARQ [7], each retransmission contains additional parity bits beyond those of the previous transmissions. There are two kinds of IR HARQ code rate selection algorithms [8]: 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) channel state information (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 [9], [10], the authors studied the performance of HARQ with convolutional codes. However, no combining technique was used and each transmission round was identical. In [11], 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 [13]. In [14]–[16], the optimal power assignment across the transmission rounds was investigated for CC HARQ, under different channel models. In [17], [18], the optimal power assignment for IR HARQ was derived. In [19], the optimal rate in different transmission rounds for IR HARQ was studied. The authors in [20] generalized the power allocation and adaptation problem for CC and IR HARQ, and built a framework for close-form solutions. In [21], the authors built a framework to analytically express the throughput of HARQ systems. The performance of HARQ with imperfect feedback was studied in [22]. The influence of time correlations of wireless channels in different HARQ transmission rounds and signaling overhead were considered in [23]. The authors in [24] showed that correlated channels may yield higher throughput than independent channels for HARQ. HARQ code design for polar codes was studied in [25], and HARQ code design for low-density parity-check codes was studied in [26]. In [27], the authors surveyed HARQ code design for turbo codes. In [28], the author designed and analyzed multilevel polar coded modulation for block fading channels.

In the existing papers on HARQ using Turbo codes for fading channels, either the HARQ strategy is simple, e.g., no combining was considered in [10], [11], or they use an information-theoretic approach, e.g., [12], [14]–[20]. The information-theoretic approach is based on [29], where the assumption is 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 paper, instead of the information-theoretic approach, we consider turbo-coded IR HARQ. In addition, we compare different models of CSI availability 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 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 contributions of this paper can be summarized as follows:

- We consider IR HARQ over independent block-fading channels with turbo coding. The energy consumption of turbo-coded HARQ is minimized subject to a packet loss rate constraint.
- 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.
- An analytical expression is proposed to approximate the error probability of turbo-coded HARQ in multiple transmission rounds.
- Numerical results show that more CSI information at the transmitter helps reduce the energy consumption by assigning the FEC code rate in each transmission more accurately. We compare the energy consumption of IR HARQ to that of Chase combining and HARQ systems without combining, and IR HARQ yields the least energy consumption.

The rest of this paper is organized as follows. In Section II, we introduce the system model. In Section III, we formulate and solve the problem. In Section IV, we show the numerical results. Conclusions are drawn in Section V.

II. SYSTEM MODEL

A. Proposed 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 paper. Let N_i be the number of bits transmitted in the i-th transmission, where $i = 1, 2, ..., K$, and K is the maximum number of transmissions. 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 acknowledgement (ACK) or negative acknowledgement (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 [15], [30]. This corresponds to the case where the complexity associated with sending the CSI or using it at the transmitter is not considered acceptable. For example, some legacy user equipment of 3GPP Long-Term Evolution (LTE) [31] does not generate or receive CSI [32]; the physical uplink control channel format 1 includes only ACK/NACK information without CSI, but other formats may contain both CSI and ACK/NACK.

(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 [8], [20]. The transmitter has only the CSI in 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 $N_1, \ldots, N_{i-1}$. 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 [10], [16], [18], [33]. Therefore, N_i depends on the CSI in the first i transmissions and $N_1, \ldots, N_{i-1}$. 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.

One example of the systems which use current and/or previous CSI is multiple-input and multiple-output (MIMO). In MIMO, not only current CSI is useful, but outdated CSI also helps to enhance the system [34], [35]. Another example is 3GPP LTE, which is able to provide the CSI at the transmitter. Therefore, a transmitter can use current CSI, and even previous CSI if beneficial. 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

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

We have $r_j \leq r_i$ when $j > i$, since $N_{i+1}, N_{i+2}, \ldots, N_j$ are nonnegative. 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

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

In other words, N_i should be chosen such that

$$N_1 + \cdots + N_i = \frac{N_b}{r_i} \in \mathbb{Z} \tag{3}$$

where $r_i \in \{r\}$. A special case occurs when the transmitter abandons the opportunity 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 Long-Term Evolution (LTE) standard for example, the time slot duration is 0.5 millisecond (ms) [36]. In [37], 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 [8], [15], [36], [38]. Constant power S_0 and binary phase shift keying (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 $f_{\Gamma_1, \Gamma_2, \ldots, \Gamma_j}(\Gamma_1, \Gamma_2, \ldots, \Gamma_j) = \prod_{i=1}^{j} f_{\Gamma_i}(\Gamma_i)$, where $1 \le j \le K$.

B. Information-Theoretic Approach

In [8], [12], [14]–[20], the authors use an information-theoretic approach in the HARQ systems. We briefly describe this approach for IR HARQ in this subsection. The system model in this approach is similar to the proposed model in Section II-A, and we keep using the symbols and definitions in Section II-A. As shown in [29], [39], after k transmissions, the decoding is successful if the average accumulated mutual information at the receiver is larger than the overall transmission rate. Therefore, the condition for successful decoding after k transmissions becomes [8]

$$\frac{1}{\sum_{j=1}^{k} N_j} \sum_{j=1}^{k} c_j N_j \ge \frac{N_b}{\sum_{j=1}^{k} N_j}, \tag{4}$$

where $c_j = \log_2(1 + \Gamma_j)$ is the mutual information between the received signal and the transmitted signal. The left hand side term in Equation (4) is the average accumulated mutual information at the receiver, and the right hand side term in Equation (4) is the overall transmission rate. As shown in [29], when N_b approaches infinity, the probability of error goes to zero when Equation (4) satisfies. Therefore, the assumption in this approach is that N_b is sufficiently large. For the existing work using this information-theoretic approach, only the first two cases of CSI in Section II-A were considered: either the transmitter does not have knowledge of the CSI or the transmitter has knowledge of the CSI in the previous transmission rounds.

III. PROBLEM FORMULATION

A. Optimization Problem

The optimization problem is

$$\min \quad \overline{\mathcal{E}}$$

$$\text{s.t.} \quad \overline{P_L} \le P_{const}$$

$$\frac{N_b}{N_1 + \cdots + N_i} \in \{r\}$$

$$\text{variables:} \quad N_i, \ i = 1, 2, \ldots, K \tag{5}$$

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 (5) is the optimization problem for all the CSI availability models in Sections III-B to III-D, 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.

Note that in some literature [8], [9], [17], [25] the goal is to maximize the throughput. As shown in [8], the throughput of an HARQ system can be written as

$$\eta = \frac{\overline{N_b}}{\overline{N_s}}, \tag{6}$$

where $\overline{N_b}$ is the expected number of correctly received information bits, and $\overline{N_s}$ is the expected number of transmitted bits (after FEC). There are a certain number of information bits to be transmitted for the system, and only the correctly received packets are counted in the throughput. This means if a packet is not transmitted to save energy, then it is lost by the definition of $\overline{N_b}$. We have

$$\overline{N_b} = N_b(1 - \overline{P_L}), \tag{7}$$

where $\overline{P_L}$ is the probability of decoding failure after K transmissions, and the average overall energy consumption is

$$\overline{\mathcal{E}} = S_0 T_s \overline{N_s}. \tag{8}$$

Therefore, we have

$$\eta = \frac{S_0 T_s N_b (1 - \overline{P_L})}{\overline{\mathcal{E}}} \approx \frac{S_0 T_s N_b}{\overline{\mathcal{E}}}, \tag{9}$$

where the approximation is based on the assumption that $\overline{P_L} \ll 1$. This means minimizing the energy consumption is equivalent to maximizing the throughput for the proposed HARQ system.

B. 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

$$\overline{\mathcal{E}} = S_0 T_s \left(N_1 + \int_0^\infty N_2 P(e_1; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right), \quad (10)$$

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

$$\overline{P_L} = \int_0^\infty \int_0^\infty P(e_1, e_2; N_1, N_2|\Gamma_1, \Gamma_2) f_{\Gamma_1, \Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 \, d\Gamma_2, \quad (11)$$

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,

$$\overline{P_L} \approx \int_0^\infty \int_0^\infty P(e_2; N_1, N_2|\Gamma_1, \Gamma_2) f_{\Gamma_1, \Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 \, d\Gamma_2, \quad (12)$$

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.

$$\min \quad S_0 T_s \left(N_1 + \int_0^\infty N_2 P(e; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right)$$

$$\text{s.t.} \quad \int_0^\infty \int_0^\infty P(e_2; N_1, N_2|\Gamma_1, \Gamma_2) f_{\Gamma_1, \Gamma_2}$$

$$\times (\Gamma_1, \Gamma_2) d\Gamma_1 \, d\Gamma_2 \leq P_{const}$$

$$\frac{N_b}{N_1 + N_2} \in \{r\}$$

$$\text{variables:} \quad N_1, N_2 \quad (13)$$

The optimization problem is in Equation (13). 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 (13). As introduced in [10] and used in [40],

$$P(e_1; N_1|\Gamma_1) \approx \min(1, a_1 e^{-b_1 \Gamma_1}), \quad (14)$$

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 (13) with the minimum energy consumption.

C. With Previous CSI

As no CSI is available for the first transmission, N_1 is predetermined as in Section III-B. 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 III-B, 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 $\Gamma^{1,1}, \Gamma^{1,2} \ldots, \Gamma^{1,t_1-1}$ as follows

$$N_2(N_1, \Gamma_1) = \begin{cases} 0, & \text{when } 0 \leq \Gamma_1 < \Gamma^{1,1} \\ N_b \left(\frac{1}{r_2} - \frac{1}{r_1} \right) = N_b \left(\frac{1}{r^{(t_2)}} - \frac{1}{r^{(t_1)}} \right), \\ \quad \text{when } \Gamma^{1,t_2} \leq \Gamma_1 < \Gamma^{1,t_2+1} \end{cases} \quad (15)$$

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. We will later use a Lagrangian multiplier to obtain the optimal SNR boundaries, so that the energy consumption is minimized, subject to a packet loss rate constraint. The average energy consumption $\overline{\mathcal{E}}$ is in Equation (16). The overall PLR is in Equation (17), 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.

$$\overline{\mathcal{E}} = S_0 T_s \left(N_1 + \int_{\Gamma^{1,1}}^\infty N_2(N_1, \Gamma_1) P(e_1; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right)$$

$$= S_0 T_s \left(N_1 + \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} N_2(N_1, \Gamma_1) \right.$$

$$\times P(e_1; N_1|\Gamma_1)f_{\Gamma_1}(\Gamma_1)d\Gamma_1 \Bigg)$$

$$= S_0 T_s N_b \left(\frac{1}{r^{(t_1)}} + \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \left(\frac{1}{r^{(i)}} - \frac{1}{r^{(t_1)}} \right) \right.$$

$$\left. \times P(e_1; r^{(t_1)}|\Gamma_1)f_{\Gamma_1}(\Gamma_1)d\Gamma_1 \right). \tag{16}$$

$$\overline{P_L} \approx \int_0^{\Gamma_1^{(1)}} P(e_1; N_1|\Gamma_1)f_{\Gamma_1}(\Gamma_1)d\Gamma_1$$

$$+ \int_{\Gamma_1^{(1)}}^{\infty} \int_0^{\infty} P(e_2; N_1, N_2(N_1, \Gamma_1)|\Gamma_1, \Gamma_2)$$

$$\times f_{\Gamma_1, \Gamma_2}(\Gamma_1, \Gamma_2)d\Gamma_2\, d\Gamma_1$$

$$= \int_0^{\Gamma_1^{(1)}} P(e_1; r^{(t_1)}|\Gamma_1)f_{\Gamma_1}(\Gamma_1)d\Gamma_1$$

$$+ \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \int_0^{\infty} P(e_2; r^{(t_1)}, r^{(i)}|\Gamma_1, \Gamma_2)$$

$$\times f_{\Gamma_1, \Gamma_2}(\Gamma_1, \Gamma_2)d\Gamma_2\, d\Gamma_1 \tag{17}$$

We use the Lagrangian multiplier method. The Lagrangian function is in Equation (19), shown at the bottom of this page, which is obtained by plugging in Equations (16) and (17). 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 $j = 1, \ldots, t_1 - 1$, where

$$\frac{\partial L}{\partial \Gamma_1^{(j)}} = h(j) + g(j). \tag{18}$$

The function $h(j)$, in Equation (20), shown at the bottom of this page, is the derivative of the second line in Equation (19), and $g(j)$, in Equation (21), shown at the bottom of this page, is the derivative of the third line in Equation (19). Equation (19) 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.

D. 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 (1).

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

$$r_1(\Gamma_1) = \begin{cases} 0, & 0 \le \Gamma_1 < \Gamma^{1,1} \\ r^{(1)} = 1/5, & \Gamma^{1,1} \le \Gamma_1 < \Gamma^{1,2} \\ r^{(2)} = 1/3, & \Gamma^{1,2} \le \Gamma_1 < \Gamma^{1,3} \\ r^{(3)} = 2/5, & \Gamma^{1,3} \le \Gamma_1 < \Gamma^{1,4} \\ r^{(4)} = 1/2, & \Gamma^{1,4} \le \Gamma_1 < \Gamma^{1,5} \\ r^{(5)} = 2/3, & \Gamma^{1,5} \le \Gamma_1 < \Gamma^{1,6} = \infty \end{cases} \tag{22}$$

$$L = \overline{\mathcal{E}} + \lambda(\overline{P_L} - P_{const})$$

$$= S_0 T_s N_b \left(\frac{1}{r^{(t_1)}} + \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \left(\frac{1}{r^{(i)}} - \frac{1}{r^{(t_1)}} \right) P(e_1; r^{(t_1)}|\Gamma_1)f_{\Gamma_1}(\Gamma_1)d\Gamma_1 \right)$$

$$+ \lambda \int_0^{\Gamma_1^{(1)}} P(e_1; r^{(t_1)}|\Gamma_1)f_{\Gamma_1}(\Gamma_1)d\Gamma_1 + \lambda \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \int_0^{\infty} P(e_2; r^{(t_1)}, r^{(i)}|\Gamma_1, \Gamma_2)f_{\Gamma_1, \Gamma_2}(\Gamma_1, \Gamma_2)d\Gamma_2\, d\Gamma_1$$

$$- \lambda \overline{P}_{const} \tag{19}$$

$$h(j) = \frac{\partial \{ S_0 T_s N_b \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \left(\frac{1}{r^{(i)}} - \frac{1}{r^{(t_1)}} \right) P(e_1; r^{(t_1)}|\Gamma_1)f_{\Gamma_1}(\Gamma_1)d\Gamma_1 \}}{\partial \Gamma_1^{(j)}}$$

$$= \underbrace{S_0 T_s N_b \left(\frac{1}{r^{(j-1)}} - \frac{1}{r^{(t_1)}} \right) P(e_1; r^{(t_1)}|\Gamma^{1,j})f_{\Gamma_1}(\Gamma^{1,j})}_{w.r.t.\ i=j-1} - \underbrace{S_0 T_s N_b \left(\frac{1}{r^{(j)}} - \frac{1}{r^{(t_1)}} \right) P(e_1; r^{(t_1)}|\Gamma^{1,j})f_{\Gamma_1}(\Gamma^{1,j})}_{w.r.t.\ i=j}$$

$$= S_0 T_s N_b \left(\frac{1}{r^{(j-1)}} - \frac{1}{r^{(j)}} \right) P(e_1; r^{(t_1)}|\Gamma^{1,j})f_{\Gamma_1}(\Gamma^{1,j}), \tag{20}$$

$$g(j) = \lambda \frac{\partial \{ \int_0^{\Gamma_1^{(1)}} P(e_1; r^{(t_1)}|\Gamma_1)f_{\Gamma_1}(\Gamma_1)d\Gamma_1 + \sum_{i=1}^{t_1-1} \int_{\Gamma^{1,i}}^{\Gamma^{1,i+1}} \int_0^{\infty} P(e_2; r^{(t_1)}, r^{(i)}|\Gamma_1, \Gamma_2)f_{\Gamma_1, \Gamma_2}(\Gamma_1, \Gamma_2)d\Gamma_2\, d\Gamma_1 \}}{\partial \Gamma_1^{(j)}}$$

$$= \begin{cases} \lambda(P(e_1; r^{(t_1)}|\Gamma^{1,1})f_{\Gamma_1}(\Gamma^{1,1}) - \int_0^{\infty} P(e_2; r^{(t_1)}, r^{(1)}|\Gamma^{1,1}, \Gamma_2)f_{\Gamma_1, \Gamma_2}(\Gamma^{1,1}, \Gamma_2)d\Gamma_2), & j = 1 \\ \lambda(\int_0^{\infty} P(e_2; r^{(t_1)}, r^{(j-1)}|\Gamma^{1,j-1}, \Gamma_2)f_{\Gamma_1, \Gamma_2}(\Gamma^{1,j-1}, \Gamma_2)d\Gamma_2 - \\ \int_0^{\infty} P(e_2; r^{(t_1)}, r^{(j)}|\Gamma^{1,j}, \Gamma_2)f_{\Gamma_1, \Gamma_2}(\Gamma^{1,j}, \Gamma_2)d\Gamma_2), & 2 \le j \le t_1 - 1 \end{cases} \tag{21}$$
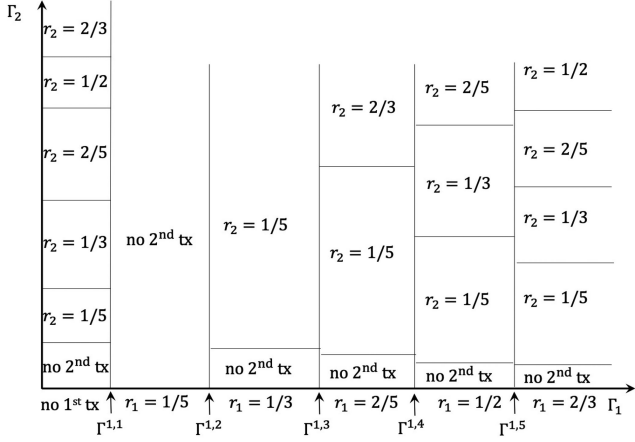
Fig. 1. SNR boundaries for two transmissions.

When $0 \leq \Gamma_1 < \Gamma^{1,1}$, $r_1(\Gamma_1) = 0$ 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. 1.

The detailed explanation of Fig. 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 $0 \leq \Gamma_1 < \Gamma^{1,1}$,

$$r_2(\Gamma_1, \Gamma_2) =$$

$$
\begin{cases}
0, & 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad 0 \leq \Gamma_2 < \Gamma^{2,1} \\
r^{(1)} = 1/5, & 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad \Gamma^{2,1} \leq \Gamma_2 < \Gamma^{2,2} \\
r^{(2)} = 1/3, & 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad \Gamma^{2,2} \leq \Gamma_2 < \Gamma^{2,3} \\
r^{(3)} = 2/5, & 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad \Gamma^{2,3} \leq \Gamma_2 < \Gamma^{2,4} \\
r^{(4)} = 1/2, & 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad \Gamma^{2,4} \leq \Gamma_2 < \Gamma^{2,5} \\
r^{(5)} = 2/3, & 0 \leq \Gamma_1 < \Gamma^{1,1}, \quad \Gamma^{2,5} \leq \Gamma_2 < \Gamma^{2,6} = \infty
\end{cases}
$$
(23)

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.

$$r_2(\Gamma_1, \Gamma_2) =$$

$$
\begin{cases}
0, & \Gamma^{1,2} \leq \Gamma_1 < \Gamma^{1,3}, \quad 0 \leq \Gamma_2 < \Gamma^{2,7} \\
r^{(1)} = 1/5, & \Gamma^{1,2} \leq \Gamma_1 < \Gamma^{1,3}, \quad \Gamma^{2,7} \leq \Gamma_2 < \Gamma^{2,8} = \infty
\end{cases}
$$
(24)

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.

$$r_2(\Gamma_1, \Gamma_2) =$$

$$
\begin{cases}
0, & \Gamma^{1,3} \leq \Gamma_1 < \Gamma^{1,4}, \quad 0 \leq \Gamma_2 < \Gamma^{2,9} \\
r^{(1)} = 1/5, & \Gamma^{1,3} \leq \Gamma_1 < \Gamma^{1,4}, \quad \Gamma^{2,9} \leq \Gamma_2 < \Gamma^{2,10} \\
r^{(2)} = 1/3, & \Gamma^{1,3} \leq \Gamma_1 < \Gamma^{1,4}, \quad \Gamma^{2,10} \leq \Gamma_2 < \Gamma^{2,11} = \infty
\end{cases}
$$
(25)

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

$$r_2(\Gamma_1, \Gamma_2) =$$

$$
\begin{cases}
0, & \Gamma^{1,4} \leq \Gamma_1 < \Gamma^{1,5}, \quad 0 \leq \Gamma_2 < \Gamma^{2,12} \\
r^{(1)} = 1/5, & \Gamma^{1,4} \leq \Gamma_1 < \Gamma^{1,5}, \quad \Gamma^{2,12} \leq \Gamma_2 < \Gamma^{2,13} \\
r^{(2)} = 1/3, & \Gamma^{1,4} \leq \Gamma_1 < \Gamma^{1,5}, \quad \Gamma^{2,13} \leq \Gamma_2 < \Gamma^{2,14} \\
r^{(3)} = 2/5, & \Gamma^{1,4} \leq \Gamma_1 < \Gamma^{1,5}, \quad \Gamma^{2,14} \leq \Gamma_2 < \Gamma^{2,15} = \infty
\end{cases}
$$
(26)

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

$$r_2(\Gamma_1, \Gamma_2) =$$

$$
\begin{cases}
0, & \Gamma^{1,5} \leq \Gamma_1 < \infty, \quad 0 \leq \Gamma_2 < \Gamma^{2,16} \\
r^{(1)} = 1/5, & \Gamma^{1,5} \leq \Gamma_1 < \infty, \quad \Gamma^{2,16} \leq \Gamma_2 < \Gamma^{2,17} \\
r^{(2)} = 1/3, & \Gamma^{1,5} \leq \Gamma_1 < \infty, \quad \Gamma^{2,17} \leq \Gamma_2 < \Gamma^{2,18} \\
r^{(3)} = 2/5, & \Gamma^{1,5} \leq \Gamma_1 < \infty, \quad \Gamma^{2,18} \leq \Gamma_2 < \Gamma^{2,19} \\
r^{(4)} = 1/2, & \Gamma^{1,5} \leq \Gamma_1 < \infty, \quad \Gamma^{2,19} \leq \Gamma_2 < \Gamma^{2,20} = \infty
\end{cases}
$$
(27)

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

The average overall energy consumption is in Equation (28), shown at the bottom of the next page, and the average overall packet loss rate is in Equation (29) shown at the bottom of the next page. 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 $i = 1$, $j = 1, \ldots, 4$ and $i = 2$, $j = 1, 2, 3, 4, 5, 7, 9, 10, 12, 13, 14, 16, 17, 18, 19$. We skip the derivation of the derivatives since it is similar to Section III-C.

IV. 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 $N_1 + N_2 + \cdots + N_K \leq 5N_b$, where N_i is the number of coded bits in the i-th transmission. This means $\frac{N_b}{r_1} + \frac{N_b}{r_2} + \cdots + \frac{N_b}{r_K} \leq 5N_b$, 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.

A. 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 in Equation (30), shown at the bottom of the next page.

Since the two transmissions are independent, we have

$$
\begin{aligned}
&P(e_1, e_2; N_1, N_2|\Gamma_1, \Gamma_2) \\
&= P(e_1; N_1, N_2|\Gamma_1, \Gamma_2)P(e_2; N_1, N_2|e_1, \Gamma_1, \Gamma_2) \\
&= P(e_1; N_1|\Gamma_1)P(e_2; N_2|\Gamma_2)
\end{aligned} \tag{31}
$$

$$
\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{28}
$$

$$
\begin{aligned}
\overline{P_L} = {} & \int_0^{\Gamma^{1,1}} \int_0^{\Gamma^{2,1}} f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2)d\Gamma_2 d\Gamma_1 + \sum_{i=1}^{5} \int_0^{\Gamma^{1,1}} \int_{\Gamma^{2,i}}^{\Gamma^{2,(i+1)}} P(e_1; r^{(i)}|\Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2)d\Gamma_2 d\Gamma_1 \\
& + \int_{\Gamma^{1,1}}^{\Gamma^{1,2}} P(e_1; r^{(1)}|\Gamma_1) f_{\Gamma_1}(\Gamma_1)d\Gamma_1 + \int_{\Gamma^{1,2}}^{\Gamma^{1,3}} \int_0^{\Gamma^{2,7}} P(e_1; r^{(2)}|\Gamma_1) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2)d\Gamma_2 d\Gamma_1 \\
& + \int_{\Gamma^{1,2}}^{\Gamma^{1,3}} \int_{\Gamma^{2,7}}^{\Gamma^{2,8}} P(e_2; r^{(2)}, r^{(1)}|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2)d\Gamma_2 d\Gamma_1 + \int_{\Gamma^{1,3}}^{\Gamma^{1,4}} \int_0^{\Gamma^{2,9}} P(e_1; r^{(3)}|\Gamma_1) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2)d\Gamma_2 d\Gamma_1 \\
& + \sum_{i=1}^{2} \int_{\Gamma^{1,3}}^{\Gamma^{1,4}} \int_{\Gamma^{2,(i+8)}}^{\Gamma^{2,(i+9)}} P(e_2; r^{(3)}, r^{(i)}|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2)d\Gamma_2 d\Gamma_1 + \\
& \int_{\Gamma^{1,4}}^{\Gamma^{1,5}} \int_0^{\Gamma^{2,12}} P(e_1; r^{(4)}|\Gamma_1) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2)d\Gamma_2 d\Gamma_1 + \sum_{i=1}^{3} \int_{\Gamma^{1,4}}^{\Gamma^{1,5}} \int_{\Gamma^{2,(i+11)}}^{\Gamma^{2,(i+12)}} P(e_2; r^{(4)}, r^{(i)}|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2)d\Gamma_2 d\Gamma_1 + \\
& \int_{\Gamma^{1,5}}^{\Gamma^{1,6}} \int_0^{\Gamma^{2,16}} P(e_1; r^{(5)}|\Gamma_1) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2)d\Gamma_2 d\Gamma_1 + \sum_{i=1}^{4} \int_{\Gamma^{1,5}}^{\Gamma^{1,6}} \int_{\Gamma^{2,(i+15)}}^{\Gamma^{2,(i+16)}} P(e_2; r^{(5)}, r^{(i)}|\Gamma_1, \Gamma_2) f_{\Gamma_1,\Gamma_2}(\Gamma_1,\Gamma_2)d\Gamma_2 d\Gamma_1
\end{aligned} \tag{29}
$$

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 transmission, conditioned on Γ_1 and Γ_2, respectively. The second line in Equation (31) 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 depends only on 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 (14). Note that Γ_1 and Γ_2 are not available at the transmitter, although they are conditioned on in the expression $P(e_1, e_2; N_1, N_2|\Gamma_1, \Gamma_2)$. We exhaustively search all the possible (N_1, N_2), and find the one which satisfies the constraint in Equation (30) and yields the least energy consumption.

B. 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.

1) Without CSI: The optimization problem is in Equation (32), shown at the bottom of this page, 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 (14). An exhaustive search is used

$$
\min \quad S_0 T_s \left(N_1 + \int_0^\infty N_2 P(e_1; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right)
$$

$$
\text{s.t.} \quad \int_0^\infty \int_0^\infty P(e_1, e_2; N_1, N_2|\Gamma_1, \Gamma_2) f_{\Gamma_1, \Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 \, d\Gamma_2 \leq P_{const}
$$

$$
\frac{N_b}{N_1} \in \{r\}, \frac{N_b}{N_2} \in \{r\}
$$

$$
N_1 + N_2 \leq 5N_b
$$

$$
\text{variables:} \quad N_1, N_2 \tag{30}
$$

$$
\min \quad S_0 T_s \left(N_1 + \int_0^\infty N_1 P(e_1; N_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1 \right)
$$

$$
\text{s.t.} \quad \int_0^\infty \int_0^\infty P(e_1, e_2; N_1, N_1|\Gamma_1, \Gamma_2) f_{\Gamma_1, \Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 \, d\Gamma_2 \leq P_{const}
$$

$$
\frac{N_b}{N_1} \in \{r\}
$$

$$
N_1 + N_1 \leq 5N_b
$$

$$
\text{variable:} \quad N_1 \tag{32}
$$

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

$$
\text{s.t.} \quad \overline{P_L} = \int_0^\infty \int_0^\infty P(e_1, e_2; N_1(\Gamma_1), N_1(\Gamma_1)|\Gamma_1, \Gamma_2) f_{\Gamma_1, \Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 \, d\Gamma_2 \leq P_{const}
$$

$$
\frac{N_b}{N_1(\Gamma_1)} \in \{r\}
$$

$$
N_1(\Gamma_1) + N_1(\Gamma_1) \leq 5N_b
$$

$$
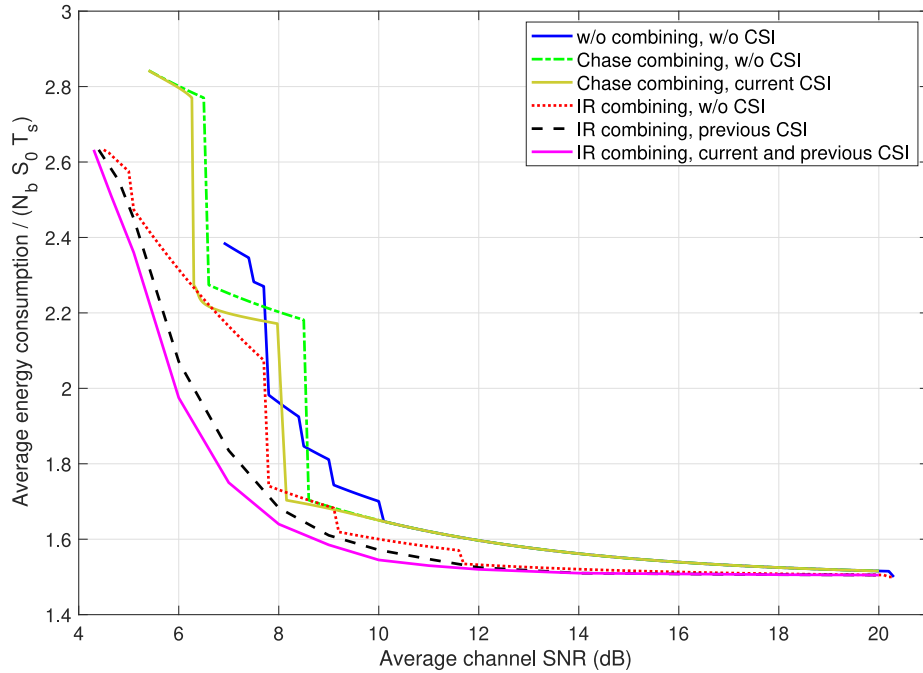\text{variable:} \quad N_1(\Gamma_1) \tag{33}
$$

Fig. 2. Energy consumption vs. average channel SNR. The maximum number of transmissions is two.

to find the N_1 that satisfies the constraint in Equation (32) and yields the least energy consumption.

2) 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 in Equation (33), shown at the bottom of the previous page.

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

$$r_1(\Gamma_1) = \begin{cases} 0, & 0 \leq \Gamma_1 < \Gamma^1 \\ r^{(1)} = 1/5, & \Gamma^1 \leq \Gamma_1 < \Gamma^2 \\ r^{(2)} = 1/3, & \Gamma^2 \leq \Gamma_1 < \Gamma^3 \\ r^{(3)} = 2/5, & \Gamma^3 \leq \Gamma_1 < \Gamma^4 \\ r^{(4)} = 1/2, & \Gamma^4 \leq \Gamma_1 < \Gamma^5 \\ r^{(5)} = 2/3, & \Gamma^5 \leq \Gamma_1 < \infty \end{cases} \quad (34)$$

From Equation (34), 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 III-C.

C. Performance Comparison

Fig. 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. As shown in [41], the impairment is almost imperceptible for audio transmission with PLR 10^{-2}. 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 7 dB, 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 8 dB, 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 utilized 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.

$$\mathcal{E}_i = \begin{cases} \int_0^\infty S_0 T_s N_1(\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1, & i = 1 \\ \frac{S_0 T_s}{\int_0^\infty P(e_1; r_1|\Gamma_1) f_{\Gamma_1}(\Gamma_1) d\Gamma_1} \int_0^\infty \int_0^\infty P(e_1; r_1|\Gamma_1) N_2(\Gamma_1, \Gamma_2) f_{\Gamma_1, \Gamma_2}(\Gamma_1, \Gamma_2) d\Gamma_1 d\Gamma_2, & i = 2 \end{cases} \quad (35)$$
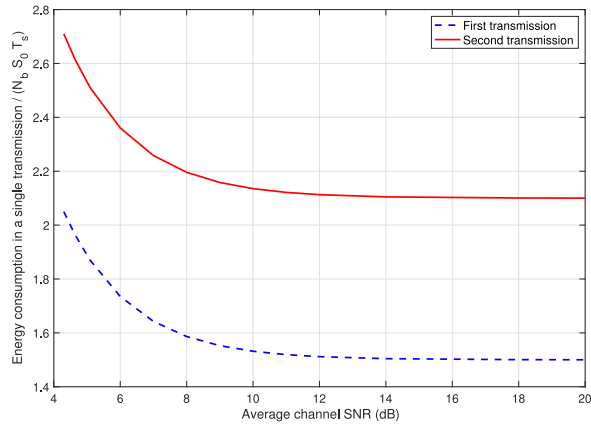
Fig. 3. Energy consumption in each transmission. The maximum number of transmissions is two.

Fig. 4. The SNR boundaries for IR combining with both current and previous CSI at an avearge channel SNR of 6 dB. The maximum number of transmissions is two.

Fig. 5. Comparison of the proposed IR combining with previous CSI to the information-theoretic approach in [8]. The maximum number of transmissions is two.

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. 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 in Equation (35) shown at the bottom of the previous page. 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.

At an average channel SNR of 6 dB, for IR combining with previous CSI, the first transmission uses FEC code rate 2/3, and the SNR boundaries are $\{\Gamma^{1,1}, \Gamma^{1,2} \ldots, \Gamma^{1,5}\} = \{-4.4 \text{ dB}, -2.9 \text{ dB}, -0.5 \text{ dB}, 1.1 \text{ dB}\}$. At an average channel SNR of 6 dB, for IR combining with both current and previous CSI, the SNR boundaries are shown in Fig. 4.

Fig. 5 compares the throughput of the proposed IR combining with previous CSI to the information-theoretic approach using rate adaptation in [8]. The throughput is the metric since energy consumption is not used in [8]. Previous CSI is also assumed in [8]. The maximum number of transmissions is two. The throughput in [8] is an upper bound of the result in this paper because 1) the condition for successful decoding is shown in Equation (4), which assumes capacity-achieving codes, whereas actual turbo codes are used in this paper, and 2) constant power and BPSK are used in this paper, but arbitrary power and continuous constellations are implicit for the information-theoretic approach, and 3) limited number of FEC code rates are used in
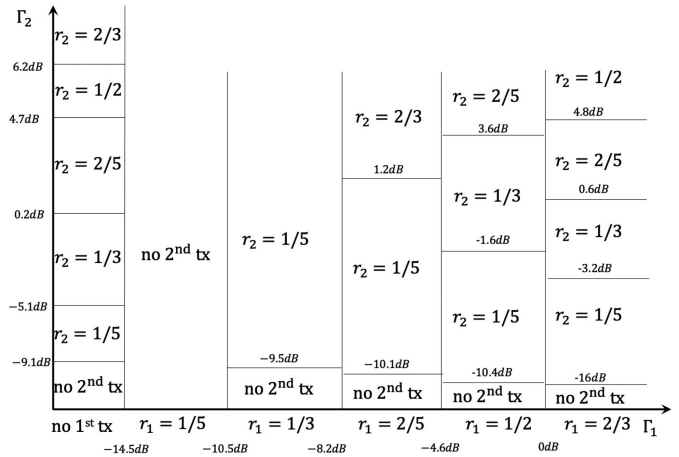
this paper, but the FEC code rates can be arbitrary in [8]. The large gap in the high SNR region is because the throughput of the proposed system is limited by the higher FEC code rate 2/3, BPSK and constant power, however, the throughput in [8] is not limited and increases with channel SNR.

Fig. 6 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 4 dB, 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. For IR combining with both current and previous CSI, the energy consumption is decreased by 26% compared to $K = 2$ at an

Fig. 6.   Energy consumption vs. average channel SNR. The maximum number of transmissions is three.

Fig. 7.   Energy consumption in each transmission. The maximum number of transmissions is three.
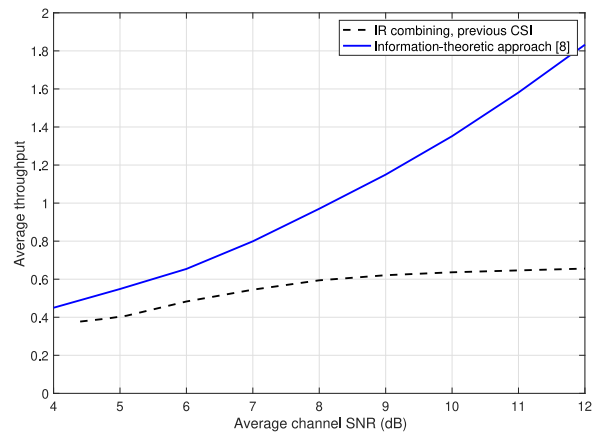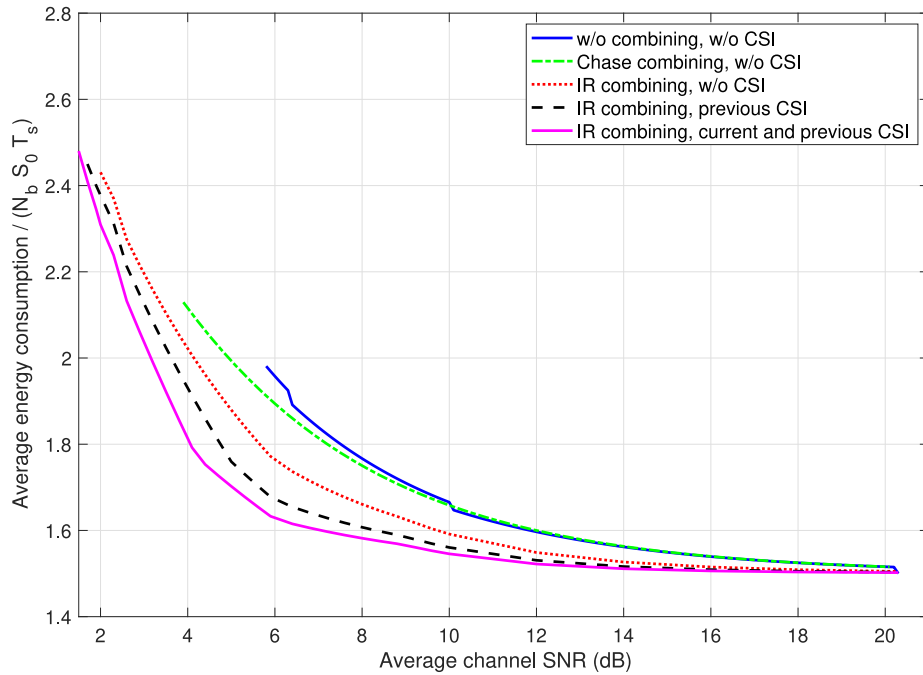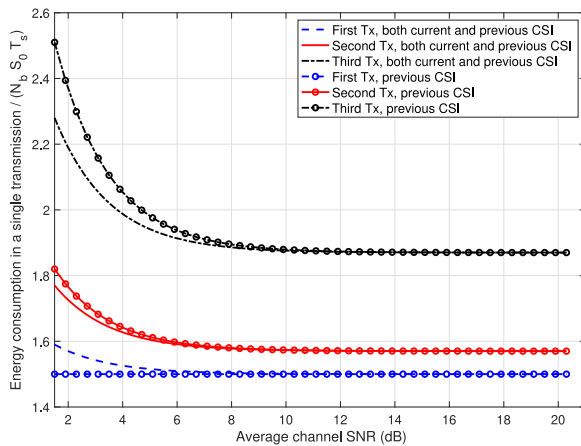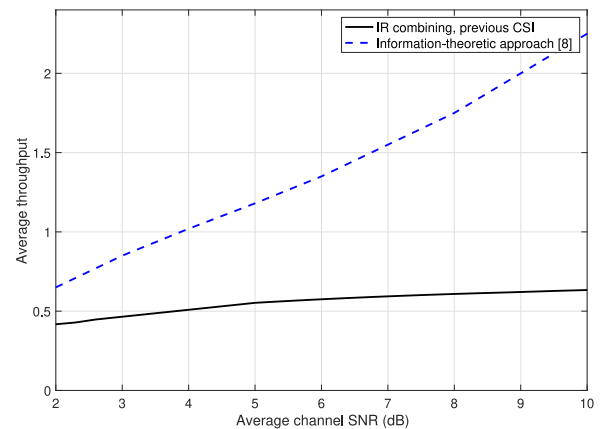

Fig. 8.   Comparison of the proposed IR combining with previous CSI to the information-theoretic approach in [8]. The maximum number of transmissions is three.

average channel SNR of 5 dB, and the minimum average channel SNR such that the PLR can be achieved is 2.4 dB 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 and 6, 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. 7 shows the $\mathcal{E}_i$ for IR combining with previous CSI and 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 a required PLR for the system. For the case with only previous CSI, the first transmission always uses FEC code rate 2/3 since no CSI is available, and the later two transmissions adjust the FEC code rate based on the previous CSI. For the case with both current and previous CSI, the FEC code rate in the first transmission is adapted to the CSI in the first transmission, which means a lower FEC code rate is used when the channel is bad in the first transmission, instead of a fixed FEC code rate. This allows the energy to be allocated more accurately in each transmission, and yields a lower overall energy consumption compared to the case where only previous CSI is available.
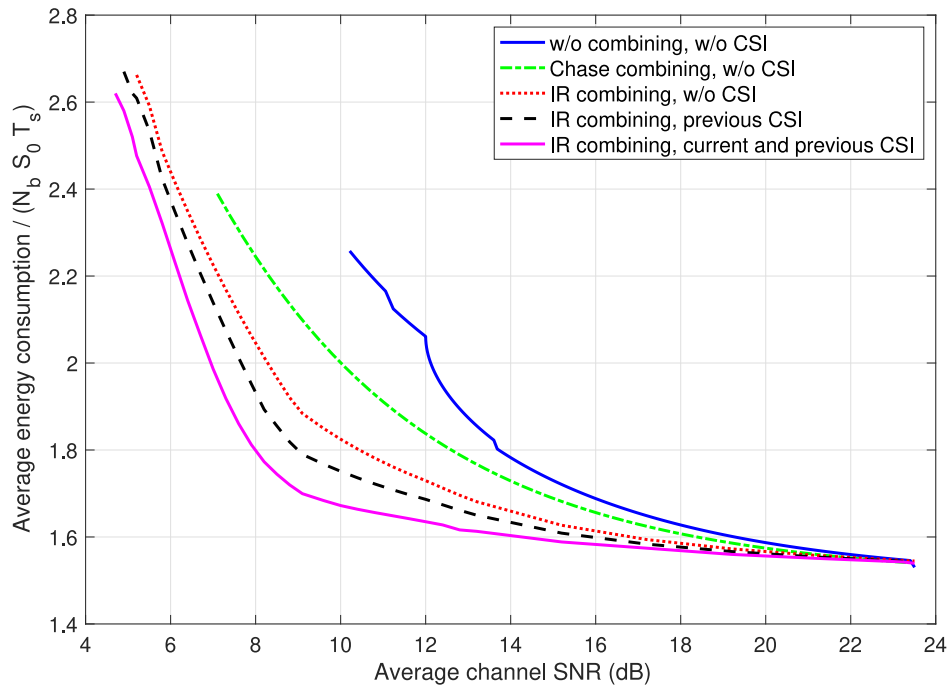
Fig. 9.   Energy consumption vs. average channel SNR. The maximum number of transmissions is three.
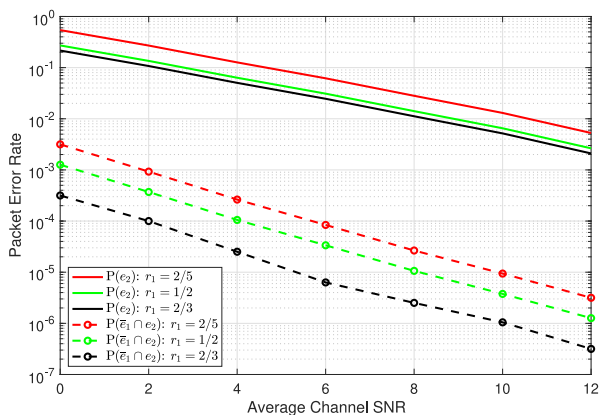

Fig. 10.   Simulated Error Probability.

Fig. 8 compares the throughput of the proposed IR combining with previous CSI to the information-theoretic approach using rate adaptation in [8]. The maximum number of transmissions is three. The trends are similar to the case where the maximum number of transmissions is two.

Video transmission usually requires a PLR smaller than 10^{-2} [43], [44]. Fig. 9 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 10^{-4}. The trends are similar to the case where the PLR constraint is 10^{-2}, except that the system operates in a higher SNR region.

V. CONCLUSION

In this paper, 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.

APPENDIX
ERROR PROBABILITY OF IR HARQ

A. 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
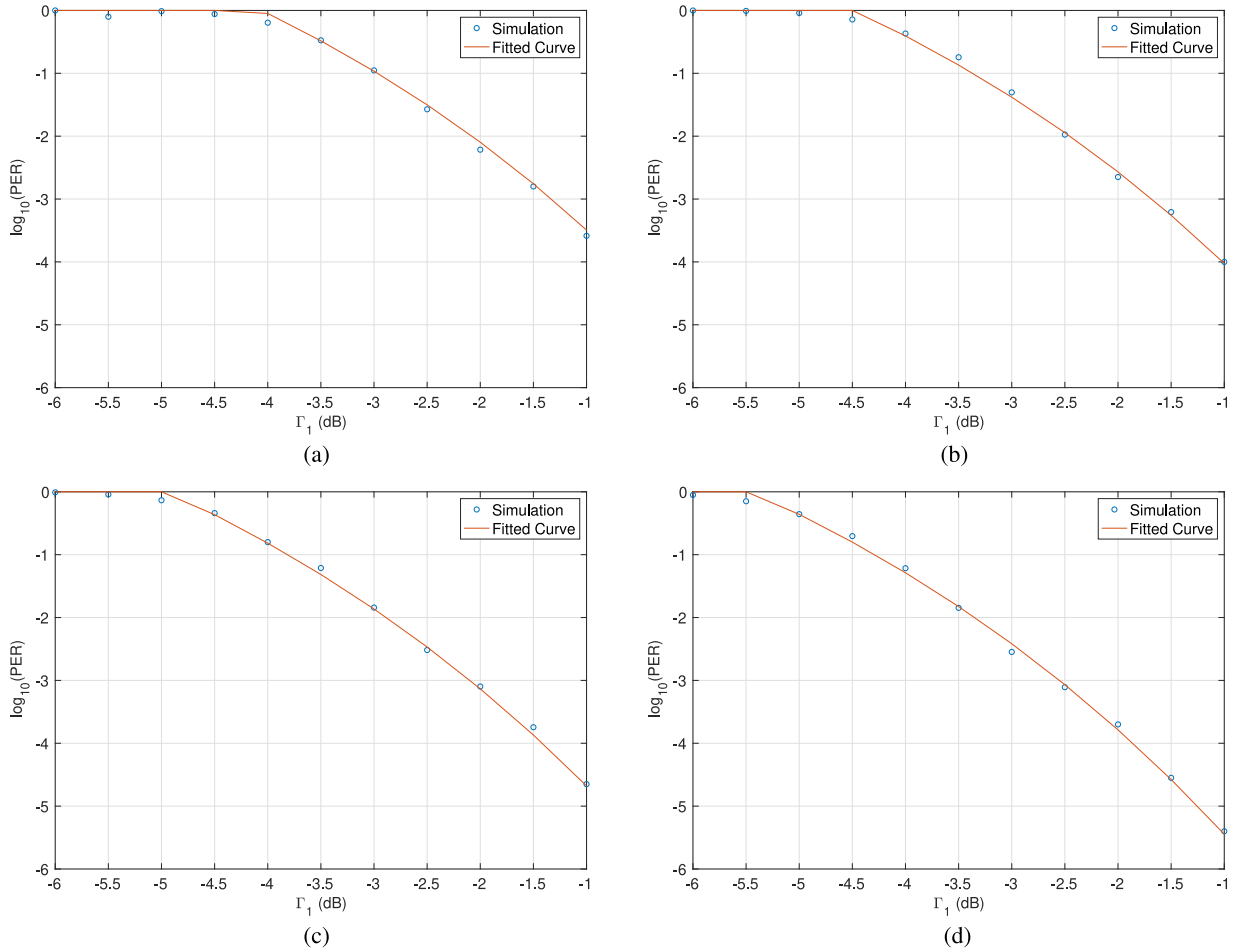
Fig. 11. Curve Fitting For Error Probability. (a) $\Gamma_2 = -6\,dB$. (b) $\Gamma_2 = -4\,dB$. (c) $\Gamma_2 = -2\,dB$. (d) $\Gamma_2 = 0\,dB$.

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. 10 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.

In [8], [15], [19], [20], [36], [45], the authors use an information-theoretic approach to study HARQ. The analysis in these papers is based on [29], 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.

B. 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

$$P(e_2; r_1, r_2 | \Gamma_1, \Gamma_2)$$

$$\approx \min\left(1, a_2 e^{-b_{2,1}\Gamma_1} e^{-b_{2,2}\Gamma_2} e^{-c_2\left(\frac{1}{\Gamma_1} + \frac{1}{\Gamma_2}\right)^{-1}}\right), \quad (36)$$

TABLE I
RMSE OF CURVE FITTING

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

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 (14). 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

$$P(e_K; r_1, r_2, \ldots, r_K | \Gamma_1, \Gamma_2, \ldots, \Gamma_K) \approx$$
$$\min\left(1, a_K \left(\prod_{i=1}^{K} e^{-b_{K,i}\Gamma_i}\right) e^{-c_K \left(\sum_{i=1}^{K} \frac{1}{\Gamma_i}\right)^{-1}}\right). \quad (37)$$

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

Fig. 11 shows 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 (36) with different values of Γ_2.

Since the fit is good, we use Equations (36) and (37) to approximate the error probability of an IR HARQ system, which allows us to numerically evaluate the energy consumption and packet loss rate.

REFERENCES

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

[2] S. Lin, D. J. Costello, and M. J. Miller, "Automatic-repeat-request error-control schemes," *IEEE Commun. Mag.*, vol. 22, no. 12, pp. 5–17, Dec. 1984.

[3] X. Zhang and Q. Du, "Adaptive low-complexity erasure-correcting code-based protocols for QoS-driven mobile multicast services over wireless networks," *IEEE Trans. Veh. Technol.*, vol. 55, no. 5, pp. 1633–1647, Sep. 2006.

[4] 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 Trans. Image Process.*, vol. 15, no. 1, pp. 40–53, Jan. 2006.

[5] G. Benelli, "An ARQ scheme with memory and soft error detectors," *IEEE Trans. Commun.*, vol. 33, no. 3, pp. 285–288, Mar. 1985.

[6] D. Chase, "Code combining–A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Trans. Commun.*, vol. 33, no. 5, pp. 385–393, May 1985.

[7] D. Mandelbaum, "An adaptive-feedback coding scheme using incremental redundancy (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 388–389, May 1974.

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

[9] R. H. Deng and M. L. Lin, "A type I hybrid ARQ system with adaptive code rates," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 733–737, Feb.–Apr. 1995.

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

[11] A. K. Karmokar, D. V. Djonin, and V. K. Bhargava, "Delay constrained rate and power adaptation over correlated fading channels," in *Proc. IEEE Global Telecommun. Conf.*, 2004, vol. 6, pp. 3448–3453.

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

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

[14] 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 Trans. Commun.*, vol. 59, no. 7, pp. 1867–1877, Jul. 2011.

[15] T. V. Chaitanya and E. G. Larsson, "Optimal power allocation for hybrid ARQ with chase combining in I.I.D. Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1835–1846, May 2013.

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

[17] C. Shen, T. Liu, and M. P. Fitz, "On the average rate performance of hybrid-ARQ in quasi-static fading channels," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3339–3352, Nov. 2009.

[18] C. Ji, D. Wang, N. Liu, and X. You, "On power allocation for incremental redundancy hybrid ARQ," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1506–1518, Mar. 2015.

[19] L. Szczecinski, C. Correa, and L. Ahumada, "Variable-rate transmission for incremental redundancy hybrid ARQ," in *Proc. IEEE Global Telecommun. Conf.*, 2010, pp. 1–5.

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

[21] P. Larsson, L. K. Rasmussen, and M. Skoglund, "Throughput analysis of hybrid-ARQ–A matrix exponential distribution approach," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 416–428, Jan. 2016.

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

[23] 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 Trans. Wireless Commun.*, vol. 10, no. 2, pp. 637–647, Feb. 2011.

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

[25] H. Khoshnevis, I. Marsland, and H. Yanikomeroglu, "Throughput-based design for polar-coded modulation," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1770–1782, Mar. 2018.

[26] T. Van Nguyen, A. Nosratinia, and D. Divsalar, "The design of rate-compatible protograph LDPC codes," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2841–2850, Oct. 2012.

[27] H. Chen, R. G. Maunder, and L. Hanzo, "A survey and tutorial on low-complexity turbo coding techniques and a holistic hybrid ARQ design example," *IEEE Commun. Surv. Tut.*, vol. 15, no. 4, pp. 1546–1566, Oct./Dec. 2013.

[28] H. Khoshnevis, "Multilevel polar coded-modulation for wireless communications," Ph.D. dissertation, Carleton University, Ottawa, ON, Canada, 2018.

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

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

[31] E. U. T. R. Access, "Physical layer procedures (Release 8)," 3GPP TS 36.213, V9. 0.1 (Dec. 2009), 2008.

[32] Y.-H. Nam, Y. Akimoto, Y. Kim, M.-I. Lee, K. Bhattad, and A. Ekpenyong, "Evolution of reference signals for LTE-advanced systems," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 132–138, Feb. 2012.

[33] 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 Trans. Wireless Commun.*, vol. 10, no. 3, pp. 877–889, Mar. 2011.

[34] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, Jul. 2012.

[35] T. Gou and S. A. Jafar, "Optimal use of current and outdated channel state information: Degrees of freedom of the MISO BC with mixed CSIT," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1084–1087, Jul. 2012.

[36] A. Chelli and M.-S. Alouini, "Performance of hybrid-ARQ with incremental redundancy over relay channels," in *Proc. IEEE Globecom Workshops*, Anaheim, CA, USA, 2012, pp. 116–121.

[37] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Pearson Education, 2007.

[38] M. Jabi, M. Benjillali, L. Szczecinski, and F. Labeau, "Energy efficiency of adaptive HARQ," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 818–831, Feb. 2016.

[39] K. D. Nguyen, L. K. Rasmussen, A. G. I. Fàbregas, and N. Letzepis, "MIMO ARQ with multibit feedback: Outage analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 765–779, Feb. 2012.

[40] 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 Syst. J.*, vol. 10, no. 2, pp. 817–830, Jun. 2016.

[41] A. Laghari, R. Laghari, A. Wagan, and A. Umrani, "Effect of packet loss and reorder on quality of audio streaming," *EAI Endorsed Trans. Scalable Inf. Syst.*, vol. 7, no. 24, 2020.

[42] B. Zhang, P. C. Cosman, and L. Milstein, "Energy optimization for wireless video transmission employing hybrid ARQ," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5606–5617, Jun. 2019.

[43] A. Biernacki and K. Tutschku, "Performance of HTTP video streaming under different network conditions," *Multimedia Tools Appl.*, vol. 72, no. 2, pp. 1143–1166, 2014.

[44] Á. Huszák and S. Imre, "Analysing GOP structure and packet loss effects on error propagation in MPEG-4 video streams," in *Proc. IEEE 4th Int. Symp. Commun., Control Signal Process.*, 2010, pp. 1–5.

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

Laurence B. Milstein (Fellow, IEEE) received the B.E.E degree from the City College of New York, New York, NY, in 1964, and the M.S. and Ph.D. degrees in electrical engineering from the Polytechnic Institute of Brooklyn, Brooklyn, NY, in 1966 and 1968, respectively. From 1968 to 1974, he was with the Space and Communications Group of Hughes Aircraft Company, and from 1974 to 1976, he was a member of the Department of Electrical and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY. Since 1976, he has been with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, where he is the Ericsson Professor of Wireless Communications and former Department Chairman, working in the area of digital communication theory with special emphasis on spread-spectrum communication systems. He has also been a consultant to both government and industry in the areas of radar and communications. Dr. Milstein was an Associate Editor for Communication Theory for the IEEE TRANSACTIONS ON COMMUNICATIONS, an Associate Editor for Book Reviews for the IEEE TRANSACTIONS ON INFORMATION THEORY, an Associate Technical Editor for the *IEEE Communications Magazine*, and the Editor-in-Chief of the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS. He was the Vice President for Technical Affairs in 1990 and 1991 of the IEEE Communications Society, and is a former Chair of the IEEE Fellow Selection Committee. He is a recipient of the 1998 Military Communications Conference Long Term Technical Achievement Award, an Academic Senate 1999 UCSD Distinguished Teaching Award, an IEEE Third Millennium Medal in 2000, the 2000 IEEE Communication Society Armstrong Technical Achievement Award, and the 2002 MILCOM Fred Ellersick Award.

Pamela Cosman (Fellow, IEEE) received the B.S. degree with Honors in electrical engineering from the California Institute of Technology in 1987, and the Ph.D. in electrical engineering from Stanford University in 1993. Following an NSF Postdoctoral Fellowship at Stanford and at the University of Minnesota (1993–1995), she joined the Faculty of the Department of Electrical and Computer Engineering at the University of California, San Diego, where she is currently a Professor.

Her research interests are in the areas of image and video compression and processing, and wireless communications. She has written over 250 technical papers in these fields, as well as one children's book, *The Secret Code Menace*, that introduces error correction coding through a fictional story. Dr. Cosman's awards include the ECE Departmental Graduate Teaching Award, Career Award from the National Science Foundation, 2017 Athena Pinnacle Award (Individual in Education), and 2018 National Diversity Award from the Electrical and Computer Engineering Department Heads Association (ECEDHA). Her administrative positions include serving as Director of the Center for Wireless Communications (2006–2008), ECE Department Vice Chair (2011–2014), and Associate Dean for Students (2013–2016).

Dr. Cosman has been a member of the Technical Program Committee or the Organizing Committee for numerous conferences, including most recently serving as Technical Program Co-Chair of ICME 2018. She was an Associate Editor of the IEEE COMMUNICATIONS LETTERS (1998–2001), and an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS (2001–2005). She was the Editor-in-Chief (2006–2009) as well as a Senior Editor (2003–2005, 2010–2013) of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. She is a member of Tau Beta Pi and Sigma Xi.

Bentao Zhang received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2014, and the M.S. degree in electrical engineering from the University of California at San Diego, La Jolla, CA, USA, in 2018, and the Ph.D. degree in electrical engineering from the University of California at San Diego, La Jolla, CA, USA, in 2020. His research interests include wireless communications and wireless video transmission.