Cross-layer optimization of cooperative and coordinative schemes for next generation cellular networks

Alaa Khreis

université PARIS-SACLAY

NNT : 2018SACLT011

Thèse de doctorat

TELECOM ParisTech

Cross-layer Optimization of Cooperative and Coordinative Schemes for Next Generation Cellular Networks

PhD Thesis of University of Paris-Saclay prepared at Télécom ParisTech

Doctoral School 580 on "Sciences et Technologies de l'Information et de la Communication" (STIC)

Speciality : « Réseaux, Information, Communications »

Thesis defended at Télécom ParisTech, on 6 November 2018, by

Alaa KHREIS

Committee:

Inbar FIJALKOW, Professor, ENSEA	**President**
Jean-François HELARD, Professor, INSA Rennes	**Reviewer**
Charly POULLIAT, Professor, ENSEEIHT	**Reviewer**
Claire GOURSAUD, Associate Professor, INSA Lyon	**Examiner**
Raphaël LE BIDAN, Associate Professor, IMT Atlantique	**Examiner**
Philippe CIBLAT, Professor, Telecom ParisTech	**Supervisor**
Pierre DUHAMEL, Senior Researcher, CNRS/L2S	**Supervisor**
Francesca BASSI, Associate Professor, ESME/SUDRIA	**Invited**

Contents

List of Acronyms

5G	5th Generation wireless systems
ACK	ACKnowledgment
AF	Amplify-and-Forward
ARQ	Automatic Repeat reQuest
AWGN	Additive White Gaussian Noise
BPSK	Binary Phase Shift Keying
CC	Chase Combining
CRC	Cyclic Redundancy Check
CSI	Channel State Information
DCF	Decode-and-Forward
DMF	Demodulate-and-Forward
FEC	Forward Error Correction
GBN	Go-Back-N
HARQ	Hybrid Automatic Repeat reQuest
HD	Half Duplex
ID	IDentifier
IR	Incremental Redundancy
ISI	Inter-Symbol Interference
LLR	Log Likelihood Ratio
LTE	Long Term Evolution
MAC	Multiple Access Channel
MCS	Modulation and Coding Scheme
MER	Message Error Rate
MISO	Multiple Input Single Output
MMSE	Minimum Mean Square Error
NACK	Negative ACKnowledgment
NDF	Non-orthogonal Decode-and-Forward
NDI	New Data Indicator
NOMA	Non Orthogonal Multiple Access
QAM	Quadrature Amplitude Modulation
RCPC	Rate-Compatible Punctured Convolutional

RTT	Round-Trip Time
RV	Redundancy Version
SAF	Slotted Amplify-and-Forward
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SR	Selective Repeat
SAW	Stop-And-Wait
TCP	Transport Control Protocol
TDMA	Time Division Multiple Access
QPSK	Quadrature Phase-shift Keying

List of Symbols

Notation	Definition
C	Maximum number of allocated time-slots to transmit one message
R	FEC coding rate ($R < 1$)
N	Number of time-instants (channel uses) in one time-slot
μ	HARQ message error rate
η	HARQ throughput
δ	HARQ delay
t	Time-slot index
n	Index of the time-instant within a time-slot
k	Message index
ℓ	Index of the HARQ round ($\ell \in \{1, \ldots, C\}$)
$\mathbf{m}_k$	Message of index k, composed of information bits and CRC bits
$\mathbf{p}_k(\ell)$	Packet corresponding to message $\mathbf{m}_k$ and HARQ round ℓ, composed of modulation symbols
$p_k(\ell, n)$	n-th symbol of packet $\mathbf{p}_k(\ell)$
E_s	Energy per symbol
E_b	Energy per information bit

General Introduction

The work presented in this Ph.D. thesis has been done in the "Communications et Electronique" (COMELEC) department of Télécom ParisTech (Paris, France) and was supported by the Labex Digicosme PhD scholarship from Université Paris-Saclay under the grant called "Cross-layer Optimization of Cooperative and Coordinative Schemes for Next Generation Cellular Networks" (Coccinelle). The thesis started in October 2015.

Problem statement

Hybrid Automatic Repeat reQuest (HARQ) has become an important research field in the wireless digital communications area during the last years. From a cross-layer point of view, data link layer approaches such as HARQ improve the robustness of the communication against channel-induced packet losses. Different HARQ strategies can be envisaged to communicate efficiently. In particular, HARQ processes share the system resources, including the time-slots. A simple way to decouple different HARQ retransmission packets is to mimic Time Division Multiple Access (TDMA). Nevertheless, TDMA is not the optimal multiple access technique in terms of data rate as it is outperformed by Non Orthogonal Multiple Access (NOMA). More precisely, NOMA enables the superposition of different packets belonging to different users. **The main idea in this thesis is to improve the time-slotted HARQ in a single-user context by mimicking NOMA, which means by allowing the superposition of packets.** We propose to apply this idea in two configurations:

- The first configuration is a single-user relay-assisted HARQ communication. In this setting, **we propose HARQ protocols using the help of a relay to improve the transmission rate and reliability**. One of these protocols allows the simultaneous transmission of different packets from the source and the relay using superposition coding. As in NOMA, the objective of using non-orthogonal transmission is to increase the rate of the communication system.

- The second configuration is a single-user HARQ communication where the feedback is delayed by several time-slots. In this setting, **we propose a new multi-layer HARQ protocol in which additional redundant packets are sent preemptively**

before receiving the acknowledgement, and in superposition to other HARQ processes. Once again, the objective is to reduce the delivery delay and boost the rate.

Outline and contributions

This Section depicts the thesis outline and gives some insights on the main contributions. The thesis is organized into two Chapters: the first one is dedicated to relay assisted HARQ, whereas the second one focuses on HARQ with delayed feedback.

In Chapter 1, we investigate point-to-point relay assisted HARQ wireless communication systems. Firstly, we describe HARQ mechanisms and briefly review the state of the art. We propose afterwards a system model to study relay assisted HARQ. We present next various HARQ protocols for the wireless communication system model, and we propose to extend a protocol that relies on non-orthogonal transmission combined with HARQ. In particular, we allow the source to transmit a new message during the same time-slot in which the relay is retransmitting a previous message. Packets corresponding to both messages are superposed. Afterwards, we define the performance metrics and design a Markov Chain model to study this protocol. Based on the system model, we derive using an information theoretic approach the analytical expressions of the performance metrics. In addition to analyzing its performance with capacity achieving codes, we compare this protocol to other protocols via simulations where we use practical codes with an interference canceler at the decoder. Although suboptimal, we show the benefits of using Successive Interference Cancellation (SIC) with this protocol.

In Chapter 2, we propose a protocol for HARQ schemes with feedback delay. We firstly present related existing works to improve the conventional HARQ mechanism. These works are either based on multi-layer HARQ or time-sharing with rate adaption policies. However, none of the previous existing works on multi-layer HARQ considered the delayed feedback. To address this problem, we define the system model consisting of a point-to-point wireless communication system using an HARQ mechanism with delayed ACKnowledgment (ACK)/Negative ACKnowledgment (NACK) feedback. Then, we propose a multi-packet HARQ protocol for this system. This protocol, also called superposition coding or multi-layer HARQ, superposes an additional layer of redundant packets to the layer of parallel Stop-and-Wait HARQ processes, with different power fractions, in order to improve the user's delay distribution and increase the throughput. Moreover, we discuss the design choices of the protocol at the transmitter side. Also, we present an in-depth analysis of the receiver, including an information theoretic characterization. Next, we define the performance metrics to evaluate the protocol and compare it to conventional HARQ using numerical results. The main advantage of this protocol

is providing better delay distribution, higher throughput and lower message error rate compared to the conventional parallel Stop-and-Wait HARQ, at the expense of increased decoding complexity.

Publications

International Conferences

C1. A. Khreis, P. Ciblat, F. Bassi, and P. Duhamel: "Throughput-efficient Relay assisted Hybrid ARQ", *in proc. of 15th International Symposium on Wireless Communication Systems (ISWCS)*, Lisbon (Portugal), August 2018.

C2. A. Khreis, P. Ciblat, F. Bassi, and P. Duhamel: "Multi-Packet HARQ with Delayed Feedback", *in proc. of 29th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Bologna (Italy), September 2018.

Chapter 1

Relay assisted Hybrid ARQ

1.1 Introduction

Reliable data transmission within wireless communication systems can be obtained via various means including HARQ mechanisms. In next generation wireless communication systems such as 5th Generation wireless systems (5G), additional nodes called relays may also help the transmission. While HARQ allows the retransmission of unsuccessfully decoded packets, the relays may contribute as well by retransmitting these packets. Therefore, an efficient use of the relays to enhance HARQ mechanisms is of great interest.

More precisely, we consider a wireless communication system where a message is sent from one source to one destination. When the decoding fails at the destination side, a relay along the route between the source and the destination retransmits a packet related to this message. The relay mimics an Automatic Repeat reQuest (ARQ) retransmission mechanism. Our main objective is to find efficient HARQ protocols taking into account the presence of one relay. This objective is addressed by analyzing existing relay assisted HARQ protocols and by proposing a new protocol in order to increase the system throughput. The main idea is to allow non-orthogonal transmission between source and relay combined with HARQ. The proposed HARQ protocol is modeled using Markov chains to theoretically investigating it by assuming capacity-achieving codes. We also simulate and compare these HARQ protocols using practical codes.

The Chapter is organized as follows. In Section 1.2, we describe HARQ mechanisms and relay assisted communication while briefly reviewing the state of the art. The system model and the performance evaluation framework for HARQ protocols are depicted in Section 1.3. Section 1.4 is devoted to relay assisted HARQ protocols. We describe these protocols along with the investigated protocol. Also, we derive in this Section a Markov chain model of the investigated protocol and calculate its performance metrics. Numerical results and comparison of the protocols, with capacity-achieving and with practical codes, are provided in Section 1.6. Finally, concluding remarks are drawn in Section 1.7.

1.2 From ARQ to Relay assisted Hybrid ARQ

Retransmission mechanisms in wireless communications existed since the invention of ARQ by Van Duuren [Van Duuren, 1943] in the 1940s. The reliability of these mechanisms is improved thanks to Forward Error Correction (FEC) techniques which combined with retransmission mechanisms gave birth to HARQ. In parallel, relay networks were developed regardless of the presence of an HARQ mechanism. Few recent works, including our contribution, combine HARQ with relay-assisted communication to improve the performance metrics of the system. In this Section, we give the necessary fundamentals of the study of relay assisted HARQ to understand the work that will be presented throughout this Chapter.

1.2.1 Automatic Repeat reQuest (ARQ)

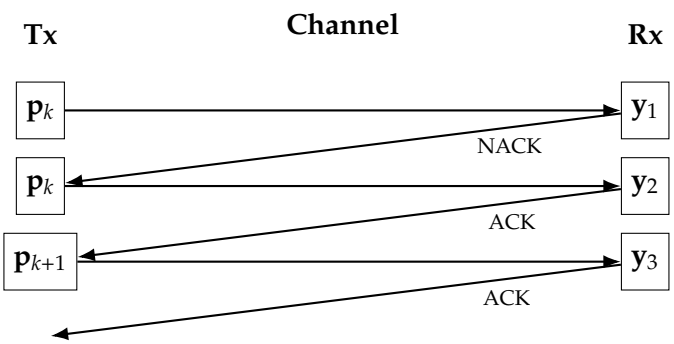


Figure 1.1 – Stop-And-Wait ARQ scheme.

ARQ is an error-control method for data transmission based on a feedback mechanism that informs the transmitter whether a transmitted packet is correctly received or not. An ACK or a NACK is sent back to the transmitter accordingly. In ARQ systems, the transmitter encodes the data with an error detection code, then transmits, after modulation, the resulting packet $\mathbf{p}_k$, where $k \in \mathbb{N}^+$, via the noisy channel, as depicted in Fig. 1.1. The receiver attempts to decode the received signal $\mathbf{y}_t$ at time-slot t, $t \in \mathbb{N}^+$, (also called ARQ round t) and checks the decoding output using the error detection code, then feeds back to the transmitter an ACK or a NACK accordingly. If a NACK occurs, the receiver discards the received signal and the transmitter retransmits the same packet $\mathbf{p}_k$. Otherwise, the receiver sends an ACK feedback and the next packet $\mathbf{p}_{k+1}$ is transmitted.

1.2.2 Hybrid ARQ (HARQ)

The error detection code that is used in ARQ, usually Cyclic Redundancy Check (CRC), can detect but not correct errors. In other words, the transmitted packets using ARQ are

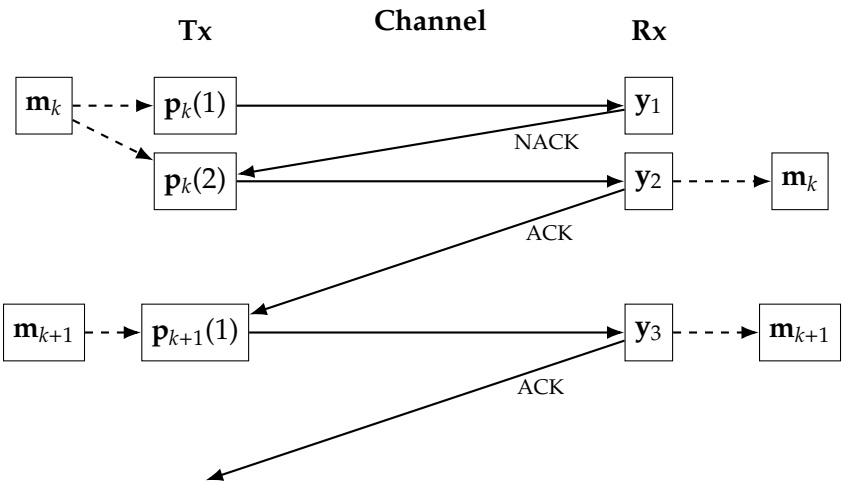


Figure 1.2 – Hybrid ARQ scheme.

constituted by the information bits and are not protected against channel-induced errors during the transmission. To overcome this issue, FEC techniques were developed along with the packet repetition mechanism, which gave birth to HARQ [Lin et al., 1984]. Many HARQ mechanisms exist and can be classified based on their error correction capabilities or based on the presence of a memory at the receiver side, for packet recombination purposes. In the following Sections, we present the state of the art HARQ mechanisms.

1.2.2.1 Type-I HARQ

Type-I HARQ describes an ARQ mechanism for which each message $\mathbf{m}_k$ is channel encoded by a FEC code and modulated into a packet $\mathbf{p}_k$ of rate R. In the example provided in Fig. 1.2, $\mathbf{p}_k = \mathbf{p}_k(1) = \mathbf{p}_k(2)$ for type-I HARQ. This packet is sent repetitively through a propagation channel using the HARQ mechanism. At the receiver side, each received version of the packet is decoded separately. In this way, the receiver does not combine the newly received signal with previously received signals. In case of failed decoding, the received signal is discarded and the receiver responds with a NACK. Otherwise, if the message is successfully decoded, the receiver responds with an ACK, and the HARQ process restarts. In truncated HARQ, the message is repeated at most C times. If the message is still detected in error and the transmission limit is reached, *i.e.* , all C packets were transmitted, the message $\mathbf{m}_k$ is dropped. Hence, Type-I HARQ schemes provide a constant error correction capability along the retransmissions and require no memory at the receiver. The correction capability of the FEC in HARQ enables the recovery of the information bits in noisy conditions, which decreases the retransmission probability.

1.2.2.2 Type-II HARQ

Type-II HARQ stores the received signals within each ARQ round at the receiver side and combines those signals for decoding, which allows to increase the correction capability of the code and the coding gain. Thus, the code rate is adapted to the current channel conditions. In order to decode the message $\mathbf{m}_k$, depicted in Fig. 1.2, the received signal corresponding to packet $\mathbf{p}_k(1)$ is stored at the receiver, then it is combined with the received signal corresponding to packet $\mathbf{p}_k(2)$ for decoding. Type-II HARQ can be implemented using Chase Combining (CC) [Chase, 1985] (also known as code combining) or using Incremental Redundancy (IR) [Mandelbaum, 1974].

Chase Combining (CC-HARQ) Type-II HARQ with CC, or CC-HARQ for short, is a scheme where the same encoded packet is retransmitted if requested. At the transmitter, the same operations are done as for Type-I HARQ, *i.e.* encoding and modulation of the message $\mathbf{m}_k$ into a packet of information rate R. However, if a NACK occurs, the received version of the transmitted packet is kept at the receiver's buffer. Then, the transmitter sends the same encoded packet, which is combined at the receiver with the previous packets in memory, using the so-called CC scheme. In Fig. 1.2,

$$\mathbf{p}_k(1) = \mathbf{p}_k(2) \qquad \text{for CC-HARQ.} \tag{1.1}$$

In most implementations, code combining is the maximal ratio combining of the received signals, corresponding to message $\mathbf{m}_k$, before entering the soft channel decoder. The coding gain brought by CC is due to the increasing correction capability at each retransmission: at the ℓ-th transmission, CC yields a virtual coding rate $R_\ell = R/\ell$, where R is the information rate.

Incremental Redundancy (IR-HARQ) Type-II HARQ with IR, or IR-HARQ for short, is a scheme where the redundancy is sent progressively upon error detection. IR-HARQ is possible thanks to the discovery of rate-compatible codes [Hagenauer, 1988; Kim et al., 2006] which enable substantial gains in received information bits per accepted packet. A message $\mathbf{m}_k$ is channel encoded at the transmitter via a mother code of rate R_0 and then punctured into C modulated packets $\mathbf{p}_k(\ell)$, $\ell \in \{1...C\}$. The transmitter transmits sequentially the packets upon error detection at the receiver side (NACK feedback). At the receiver side, the first packet of the sequence $\mathbf{p}_k(1)$, of information rate R, is simply decoded, and the decoder checks if the message can be recovered without error or not. If an ACK occurs, the HARQ process is restarted with the next message. Otherwise, the next packet in the sequence, $\mathbf{p}_k(2)$, is transmitted over the channel and so on. In Fig. 1.2,

$$\mathbf{p}_k(1) \neq \mathbf{p}_k(2) \qquad \text{for IR-HARQ.} \tag{1.2}$$

The received signal is combined to the previously received signals of the sequence, the aggregation of packets is decoded and ACK/NACK is sent back depending on the decoding

outcome. The increasing correction capability given by IR is the result of the decreasing code rate obtained after incremental combination of the packets at the receiver.

1.2.3 Relay assisted communication

Relay-assisted wireless communication schemes have been proposed in various contexts [Cover and Gamal, 1979; Laneman et al., 2004; Kramer et al., 2005; Yang and Belfiore, 2007; Guan and Chen, 2013; Nazer and Gastpar, 2008] regardless of the presence of an HARQ mechanism. We consider a point-to-point Half Duplex (HD) channel with one relay, *i.e.* , the system consists of three nodes: one source, one relay, and one destination as shown in Fig. 1.3.

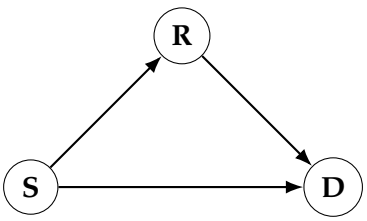


Figure 1.3 – Relay assisted point-to-point communication scheme.

The relay-assisted schemes are classified into orthogonal and non-orthogonal schemes depending on whether the source and relay share the same resources simultaneously or not.

1.2.3.1 Orthogonal relaying schemes

In orthogonal relaying schemes the source remains silent when the relay is active. Orthogonal relaying strategies include Amplify-and-Forward (AF), Demodulate-and-Forward (DMF), and Decode-and-Forward (DCF). The AF strategy allows the relay to amplify the received signal from the source and to forward it to the destination. AF requires low computing power as no decoding or quantizing operation is performed at the relay. When using the DMF strategy, the relay demodulates the received signal from the source and retransmits the signal to the destination. The DMF strategy mitigates the received noise and interference by performing a simple signal processing without decoding the overheard transmission. A relay following the DCF strategy overhears transmissions from the source, decodes them and in case of correct decoding, forwards them to the destination. In case the relay fails to decode the overheard signal, the relay remains silent. Using DCF, the relay can regenerate the message perfectly if the overheard transmission is decoded correctly.

1.2.3.2 Non-Orthogonal relaying schemes

In non-orthogonal relaying schemes the source might transmit packets when the relay is active as in the Slotted Amplify-and-Forward (SAF) [Yang and Belfiore, 2007] or the Non-orthogonal Decode-and-Forward (NDF) [Nabar et al., 2004; Prasad and Varanasi, 2004; Guan and Chen, 2013]. The received signal at the destination using SAF strategy is a linear combination of the amplified signal by the relay and transmitted signal by the source. In NDF, the relay does not completely take over the message transmission, even if the relay has correctly decoded the message, since the source continues to broadcast the same message through other packets.

1.2.4 Relay assisted Hybrid ARQ

In relay assisted HARQ protocols, the relay participates in the HARQ mechanism by transmitting the redundant packets. Only few works have considered the use of relays to support the HARQ mechanism. These works, and our work as well, consider HD relays with a DCF strategy. Hereby, we present a brief review of the key papers on relay assisted HARQ.

- In [Chelli and Alouini, 2013], a distributed Alamouti relaying scheme is investigated. The authors propose to use the Alamouti code during the retransmission phase if the relay successfully decodes the overheard message while the destination fails to decode this message. The objective is to increase the transmit diversity using the Alamouti code. We provide in-depth analysis of this protocol in Section 1.4.3.

- Network coding approach for relay assisted HARQ has been developed in [Hong and Chung, 2010] where the relay sends a XOR combination of bits of multiple overheard and decoded packets at the relay, that are not correctly received at the destination.

- In [Larsson and Vojcic, 2005; Chaitanya and Larsson, 2011], a superimposed modulation (linear combination of symbols) is proposed in a multi-user context where a node transmits simultaneously its own data packet and the packet of another transmitter for which it acts as a relay.

- In [Zhang et al., 2010; Ma et al., 2013], a protocol in single-user context combining HARQ and DCF relay is introduced. The idea is to let the source send a new message while the relay is retransmitting a previous one. In both papers, this idea is only considered for a CC-HARQ with one retransmission credit. In addition, only a Minimum Mean Square Error (MMSE) receiver is employed for handling the interference between both simultaneously received messages at the destination. More precisely, in [Zhang et al., 2010], the authors considered a dynamic protocol where the source sends a new message rather than a space-time coded version

of the previous message as soon as the received Signal-to-Interference-plus-Noise Ratio (SINR) is higher than a predefined threshold. This implies that i) the source only seldom uses the ability to send a new message, ii) the source has the channel state information of the future channels! In [Ma et al., 2013], the authors considered only Binary Phase Shift Keying (BPSK)-modulated signals. In both papers, the gain in throughput is marginal. We suspect that the throughput gain is small due to the choice of the sub-optimal MMSE receiver.

- The idea of receiving more than one packet simultaneously (incoming from the source and relay) is close to the multi-packet reception, for the random access protocol, in [Samano-Robles et al., 2015]. Their paper presents signal processing tools that significantly improve the multi-packet reception such as SIC. These tools will be discussed in Section 1.6.6.

The list of works presented in this Section is non-exhaustive, but shows the major approaches of using the relay to support the HARQ mechanism.

1.3 Performance evaluation framework for HARQ protocols

In this Section, we define the performance metrics and present the model of the communication system for HARQ protocols. In addition, we explain the theoretical tools that enable the analysis of communication systems using HARQ, including relay assisted HARQ protocols.

1.3.1 HARQ performance metrics

The throughput, Message Error Rate (MER) and delay are the performance metrics to evaluate in order to provide a complete overview of the system [Lin et al., 1984].

1.3.1.1 Throughput

The throughput, denoted by η, is the average number of correctly received information bits at the destination per channel use. In practice, each channel use corresponds to one time-instant which is the allocated time to transmit one symbol over the channel. Therefore, the throughput is calculated as the average ratio of correctly received information bits at the destination per time-instant to the total number of information bits that could be transmitted per time-instant.

1.3.1.2 Message Error Rate (MER)

The MER, denoted by μ, is the average ratio of the number of dropped messages over the number of sent messages. In (H)ARQ retransmission mechanisms, a message is dropped in case of a NACK response on the last transmission.

1.3.1.3 Delay

The delay, denoted by δ, is the average number of elapsed time-slots since the first transmission of each message until its successful decoding at the receiver, considering only the successfully decoded messages. Notice that the delay is not proportional to the inverse of the throughput.

1.3.2 System model

As already said, we consider a three nodes system with one source, one relay, and one destination as described in Section 1.2.3. Each link (source-relay, source-destination, relay-destination) is modeled as an independent Rayleigh flat fading channel. The coherence time of each link is equal to the time-slot duration containing N time-instants. Let $h_{sr}(t)$, $h_{sd}(t)$, and $h_{rd}(t)$ be the fading components for the source-relay, source-destination, and relay-destination links at the t-th time-slot respectively. The associated gains are $g_{sr}(t) = |h_{sr}(t)|^2$, $g_{sd}(t) = |h_{sd}(t)|^2$, and $g_{rd}(t) = |h_{rd}(t)|^2$ with variance $\sigma_{sr}^2 = \mathbb{E}[g_{sr}(t)]$, $\sigma_{sd}^2 = \mathbb{E}[g_{sd}(t)]$, $\sigma_{rd}^2 = \mathbb{E}[g_{rd}(t)]$. At time-slot t, the received signal at the destination is denoted by $\mathbf{y}_t$ and the Additive White Gaussian Noise (AWGN) vector is denoted by $\mathbf{w}_t$. The relay is assumed to work in a HD, DCF mode. After each time-slot, the relay and destination send an instantaneous error-free ACK heard by all the nodes if they succeeded to decode their message or a NACK otherwise. To clarify furthermore, the relay sends its feedback to the source while the destination sends its feedback to both the relay and the source. Moreover, perfect Channel State Information (CSI) at each receiver (relay or destination) is assumed available. Each message $\mathbf{m}_k$ contains NR information bits. We remind that, in CC-HARQ, each message $\mathbf{m}_k$ is channel encoded and modulated into a packet $\mathbf{p}_k$ of information rate R. This packet is sent through a propagation channel during C time-slots at most, using the truncated HARQ mechanism. In IR-HARQ, each message $\mathbf{m}_k$ is encoded via a mother code of rate R_0 and then punctured into C codeword chunks of index ℓ, $\ell \in \{1, 2, \ldots, C\}$. The ℓ-th codeword chunk is modulated into a packet of length N, denoted by $\mathbf{p}_k(\ell)$.

1.3.3 Theoretical analysis tools for HARQ

HARQ protocols can be modeled, as in [Badia et al., 2008; Chen et al., 2013], using Markov chains whose transition probabilities enable the evaluation of the performance metrics provided in Section 1.3.1. Other tools, such as renewal reward processes in [Caire and Tuninetti, 2001], have been used to compute the throughput of HARQ protocols. Hereby, we describe the process of modeling HARQ protocols via Markov chains and deriving their performance metrics. To illustrate the concept, the throughput of an HARQ protocol is derived using a Markov chain model and compared to the results in [Caire and Tuninetti, 2001]. In Section 1.4, we will use Markov chain models to derive the performance metrics of relay assisted HARQ protocols.

1.3.3.1 Markov chain models for HARQ protocols

In Markov chains, the probability of being in a state only depends on the state reached in the previous time. To model an HARQ protocol as a Markov chain we i) describe the protocol and define the states of the corresponding Markov chain, ii) characterize the transition probabilities, iii) calculate the steady state probabilities and deduce the performance metrics of the protocol. Analytical expressions of the transition probabilities are derived using an information theoretic approach. Following this approach, successful decoding of a message corresponds to comparing the accumulated mutual information of a specific HARQ process to the coding rate R.

Markov chain model of IR-HARQ with 2 transmission credits Hereby, we provide an example of deriving the performance metrics of an HARQ protocol based on a Markov chain model. Then, we compare the result to the one obtained in [Caire and Tuninetti, 2001] based on the renewal-reward theorem. For clarity purposes, we choose the example of point-to-point transmission using IR-HARQ with 2 transmission credits ($C = 2$). The source transmits each message to the destination through 2 packets at most, using IR. Moreover if the destination decodes the message successfully, the source transmits the next message. In case of a NACK of the first packet, the source transmits the second packet. In case of NACK of the second packet, the message is dropped and the source restarts the process with next message. These events can be described by three states as follows.
$\mathbf{S}_a$: following the ACK of the previous message, the source node transmits the first packet of a new message.
$\mathbf{S}_b$: the source node transmits the first packet of a new message following a drop, *i.e.* a NACK of the previous message and the HARQ transmission credit is exhausted.
$\mathbf{S}_c$: the source node retransmits a message through a second packet after a NACK.
Consequently, the states of the Markov chain and transitions between these states are represented in Fig. 1.4.

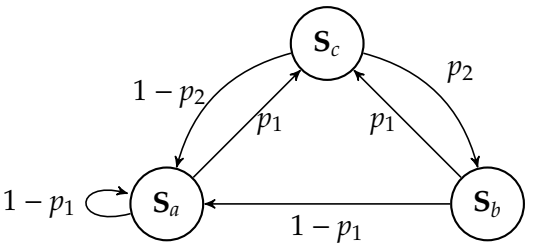


Figure 1.4 – Markov chain model of HARQ with 2 transmission credits.

For IR-HARQ, the transition probabilities p_1 and p_2 can be written as in Eq. (1.3) and Eq. (1.5), respectively. We denote by $\Pr(\Omega)$ be the probability of the event Ω. In this way,

the probability of not decoding the first transmission is:

$$\begin{aligned} p_1 &= \Pr\left(\log_2(1+g_{sd}(1)) < R\right) \\ &= \Pr\left(g_{sd}(1) < 2^R - 1\right) \\ &= \frac{1}{\sigma_{sd}^2}\int_0^{2^R-1} e^{-g/\sigma_{sd}^2} dg \\ &= 1 - e^{-(2^R-1)/\sigma_{sd}^2}. \end{aligned} \tag{1.3}$$

The probability of not decoding the second transmission knowing that the first transmission was not decoded is:

$$\begin{aligned} p_2 &= \Pr\left(\log_2(1+g_{sd}(1)) + \log_2(1+g_{sd}(2)) < R \,\middle|\, \log_2(1+g_{sd}(1)) < R\right) \\ &= \Pr\left(g_{sd}(2) < \frac{2^R}{1+g_1} - 1 \,\middle|\, g_{sd}(1) < 2^R - 1\right). \end{aligned}$$

According to Bayes' rule,

$$\Pr\left(\Omega_1 \,|\, \Omega_2\right) = \frac{\Pr\left(\Omega_1, \Omega_2\right)}{\Pr\left(\Omega_2\right)}. \tag{1.4}$$

Therefore, p_2 can be written as:

$$\begin{aligned} p_2 &= \frac{1}{p_1}\Pr\left(g_{sd}(2) < \frac{2^R}{1+g_{sd}(1)} - 1, g_{sd}(1) < 2^R - 1\right) \\ &= \frac{1}{p_1}\left(\frac{1}{\sigma_{sd}^4}\int_{g_1=0}^{2^R-1}\int_{g_2=0}^{\frac{2^R}{1+g_1}-1} e^{-g_1/\sigma_{sd}^2} e^{-g_2/\sigma_{sd}^2} dg_2 dg_1\right) \\ &= 1 - \frac{1}{p_1}\frac{e^{1/\sigma_{sd}^2}}{\sigma_{sd}^2}\int_{g_1=0}^{2^R-1} e^{-\left(\frac{2^R}{1+g_1}+g_1\right)/\sigma_{sd}^2} dg_1. \end{aligned} \tag{1.5}$$

Moreover, this expression can be evaluated using the generalized Fox's H function as in [Chelli and Alouini, 2013].

These transitions can also be represented by the transition matrix given in Eq. (1.6) as:

$$\mathbf{T}_{(\text{HARQ},C=2)} = \begin{bmatrix} 1-p_1 & 1-p_1 & 1-p_2 \\ 0 & 0 & p_2 \\ p_1 & p_1 & 0 \end{bmatrix}. \tag{1.6}$$

The corresponding steady state probabilities vector $\mathbf{\Pi}$ is obtained by solving Eq. (1.7a) and Eq. (1.7b) given by:

$$\mathbf{T}_{(\text{HARQ},C=2)}\mathbf{\Pi} = \mathbf{\Pi}, \tag{1.7a}$$

$$\sum_{i=a,b,c} \pi_i = 1. \tag{1.7b}$$

Hence, we obtain $\pi_a = \frac{1-p_1p_2}{1+p_1}$, $\pi_b = \frac{p_1p_2}{1+p_1}$ and $\pi_c = \frac{p_1}{1+p_1}$. According to the definitions of the performance metrics above, we find the expressions of the throughput and MER as follows:

$$\eta_{(\text{HARQ},C=2)} = \pi_a = \frac{1-p_1p_2}{1+p_1}, \tag{1.8a}$$

$$\mu_{(\text{HARQ},C=2)} = \frac{\pi_b}{\pi_b+\pi_a} = p_1p_2. \tag{1.8b}$$

The expression of the throughput can also be obtained by applying [Caire and Tuninetti, 2001, Eq. (17)] for the IR scheme as follows.

$$\eta_C = RG\frac{\left[1-\sum_{i=0}^{C-1}\binom{C}{i}(1-p_t)^{C-i}p_t^i p(i) - \sum_{i=M}^{C}\binom{C}{i}(1-p_t)^{C-i}p_t^i p(C)\right]}{\sum_{\ell=0}^{C-1}p(\ell)\left[1-\sum_{i=0}^{\ell}\binom{C+1}{i}(1-p_t)^{C+1-i}p_t^i - \binom{C}{\ell}(1-p_t)^{C-\ell}p_t^{\ell+1}\right]},$$

where p_t is the probability that the source transmits a signal, $p(\ell)$ is the probability of having failed decoding with ℓ transmitted packets, G is the number of users transmitting over a time-slot. In our case, we have $C = 2$, $p_t = 1$, $G = 1$, and $p(1) = p_1$ and $p(2) = p_1p_2$. Consequently, we obtain

$$\eta_2 = \frac{1-p(2)}{1+p(1)} = \frac{1-p_1p_2}{1+p_1} = \eta_{(\text{HARQ},C=2)}, \tag{1.9}$$

which is equal to the calculated throughput using the Markov chain model.

1.4 Relay assisted HARQ protocols

Relay assisted HARQ protocols are described in this Section. In the presence of a relay, the source performs the first HARQ transmission of a message to the destination while the relay overhears this transmission. As follows, the relay assisted HARQ protocols differ by the retransmission strategy. We remind that the relay operates in HD mode which allows either the reception or the transmission of a packet during each time-slot. Also the relay uses a DCF strategy which means that the relay is active only upon successful decoding of the overheard message. Moreover, we consider truncated HARQ where C time-slots at most are allocated to transmit each message.

1.4.1 Orthogonal relay retransmission

If the relay succeeds to decode the message, the relay retransmits the remaining packets to the destination, while the source remains silent, as depicted in Fig. 1.5.

Using this protocol, the relay overhears the transmissions of the source and tries to decode the overheard messages. In case of successful decoding of message $\mathbf{m}_k$ by the relay node and failed decoding by the destination node, the relay sends an ACK to the

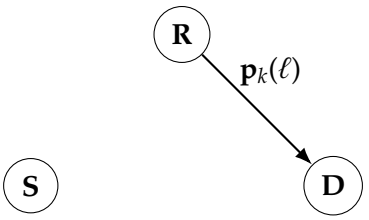


Figure 1.5 – Orthogonal relay retransmission of message $\mathbf{m}_k$ at time-slot t.

source while the destination sends a NACK feedback to both the source and relay nodes. Thus, the source decides to remain silent and allows the relay to retransmit the remaining redundant packets to the destination. In other words, if the relay decodes packet $\mathbf{p}_k(\ell-1)$, with $1 < \ell < C$, and the destination fails to decode it, the relay transmits packet $\mathbf{p}_k(\ell)$ to the destination while the source remains silent. This protocol is beneficial if the relay-to-destination channel is statistically better than the source-to-destination channel. More precisely, based on the channel model, this protocol is beneficial if the distance between the relay and destination is smaller than the distance between the source and destination.

1.4.2 Non-orthogonal relay retransmission

If the relay succeeds to decode the message, the relay retransmits the remaining packets to the destination along with the source, as depicted in Fig. 1.6.

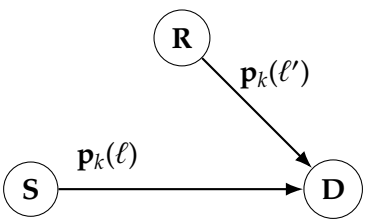


Figure 1.6 – Non-orthogonal retransmission of message $\mathbf{m}_k$ at time-slot t.

Using this protocol, the source does not remain silent when the relay retransmits the redundant packet to the destination. Moreover, the source retransmits simultaneously a redundant packet of message $\mathbf{m}_k$ as the relay. Hence the source is always active and the relay is active upon successful decoding of the overheard message. In other words, if the relay decodes packet $\mathbf{p}_k(\ell-1)$, with $1 < \ell < C$, and the destination fails to decode it, the relay and the source transmit packets corresponding to message $\mathbf{m}_k$, simultaneously. Moreover, the relay might transmit $\mathbf{p}_k(\ell') = \mathbf{p}_k(\ell)$ or $\mathbf{p}_k(\ell') \neq \mathbf{p}_k(\ell)$, depending on the considered non-orthogonal relay retransmission protocol. Notice that the simultaneous transmission of $\mathbf{p}_k(\ell)$ and $\mathbf{p}_k(\ell')$ corresponds to one HARQ round since one time-slot only is allocated for this transmission. The main advantage of non-orthogonal relay retransmission is to provide larger Signal-to-Noise Ratio (SNR) during the retransmissions in

comparison to orthogonal relay retransmission. However, this protocol uses more transmission power since both the source and relay transmit simultaneously the redundant packets during the HARQ retransmission rounds.

1.4.3 Alamouti retransmission

If the relay succeeds to decode the message, the relay retransmits the *Alamouti conjugate* of the same packet that is transmitted simultaneously by the source, as depicted in Fig. 1.7.

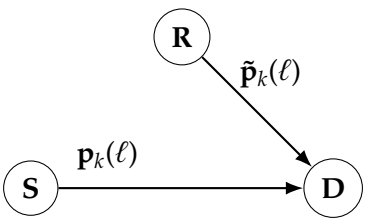


Figure 1.7 – Alamouti retransmission of message $\mathbf{m}_k$ at time-slot t.

This strategy is suggested in [Chelli and Alouini, 2013] and relies on the Alamouti space-time coding [Alamouti, 1998]. As explained in Section 1.3.2, each packet $\mathbf{p}_k(\ell)$ is composed of N symbols where N is even, *i.e.* ,

$$\mathbf{p}_k(\ell) = \left[p_{k,1}(\ell), p_{k,2}(\ell), \cdots, p_{k,N}(\ell)\right]. \tag{1.10}$$

The *Alamouti conjugate* of packet $\mathbf{p}_k(\ell)$, denoted by $\tilde{\mathbf{p}}_k(\ell)$, is given by:

$$\tilde{\mathbf{p}}_k(\ell) = \left[-\overline{p}_{k,2}(\ell), \overline{p}_{k,1}(\ell), -\overline{p}_{k,4}(\ell), \cdots, \overline{p}_{k,N-1}(\ell)\right]. \tag{1.11}$$

If the destination fails to decode the message of index k after $\ell - 1$ transmissions, where $\ell > 1$, while the relay succeeds to decode this message, then the relay sends $\tilde{\mathbf{p}}_k(\ell)$ while the source sends packet $\mathbf{p}_k(\ell)$. This simultaneous transmission by the source and relay corresponds to one HARQ round since it occupies one time-slot only. Hence the received signal by the receiver at time-slot t is $\mathbf{y}_t$, where

$$\mathbf{y}_t = [y_t[1], y_t[2], \cdots, y_t[N]]^T. \tag{1.12}$$

The received sample at each time-instant, using Alamouti retransmission, is given by Eq. (1.13) as:

$$y_t[1] = \begin{bmatrix} h_{sd}(t) & h_{rd}(t) \end{bmatrix} \begin{bmatrix} p_{k,1}(\ell) \\ -\overline{p}_{k,2}(\ell) \end{bmatrix} + w_t[1], \tag{1.13a}$$

$$y_t[2] = \begin{bmatrix} h_{sd}(t) & h_{rd}(t) \end{bmatrix} \begin{bmatrix} p_{k,2}(\ell) \\ \overline{p}_{k,1}(\ell) \end{bmatrix} + w_t[2], \tag{1.13b}$$

$$\cdots$$

$$y_t[N] = \begin{bmatrix} h_{sd}(t) & h_{rd}(t) \end{bmatrix} \begin{bmatrix} p_{k,N}(\ell) \\ \overline{p}_{k,N-1}(\ell) \end{bmatrix} + w_t[N]. \tag{1.13c}$$

During time-slot t, the source-to-destination and relay-to-destination channels remain constant due to Rayleigh flat fading (also called slow fading).

1.4.3.1 Decoder of Alamouti retransmission

Following the reception of samples $y_t[1]$ and $y_t[2]$, the destination node performs the operations described in this Section to cancel the effect of the channel fading. The received samples, during the first two time-instants of time-slot t, can be written as:

$$\begin{bmatrix} y_t[1] \\ \overline{y}_t[2] \end{bmatrix} = \begin{bmatrix} h_{sd}(t) & -h_{rd}(t) \\ \overline{h}_{rd}(t) & \overline{h}_{sd}(t) \end{bmatrix} \begin{bmatrix} p_{k,1}(\ell) \\ \overline{p}_{k,2}(\ell) \end{bmatrix} + \begin{bmatrix} w_t[1] \\ \overline{w}_t[2] \end{bmatrix}. \tag{1.14}$$

We define H as:

$$\mathrm{H} = \begin{bmatrix} h_{sd}(t) & -h_{rd}(t) \\ \overline{h}_{rd}(t) & \overline{h}_{sd}(t) \end{bmatrix}. \tag{1.15}$$

The destination node computes the pseudo-inverse of H, denoted by H^{+}, as:

$$\mathrm{H}^{+} = (\mathrm{H}^{H}\mathrm{H})^{-1}\mathrm{H}^{H}, \tag{1.16}$$

such that $\mathrm{H}^{+}\mathrm{H} = \mathrm{I}_d$. Hence the channel fading is canceled by multiplying both sides of Eq. (1.14) by H^{+} as shown in Eq. (1.17):

$$\begin{bmatrix} \hat{p}_{k,1}(\ell) \\ \hat{\overline{p}}_{k,2}(\ell) \end{bmatrix} = \begin{bmatrix} p_{k,1}(\ell) \\ \overline{p}_{k,2}(\ell) \end{bmatrix} + \mathrm{H}^{+} \begin{bmatrix} w_t[1] \\ \overline{w}_t[2] \end{bmatrix}, \tag{1.17}$$

where $\hat{p}_{k,1}(\ell)$ and $\hat{\overline{p}}_{k,2}(\ell)$ are the received versions of $p_{k,1}(\ell)$ and $\overline{p}_{k,2}(\ell)$ respectively. This operation is also performed on each pair of received samples $(y_t[3], y_t[4])$, $(y_t[5], y_t[6])$,...,$(y_t[N-1], y_t[N])$. Then the estimated version of the packet $\mathbf{p}_k(\ell)$ is passed to the channel decoder. The main benefit of using Alamouti retransmission is to reduce the probability of failed decoding, by providing a diversity of 2 in each time-slot when the relay is active. Alamouti retransmission can be seen as a virtual Multiple Input Single Output (MISO) system where the source node and relay node cooperate to transmit a packet to the destination node. As a result, this protocol achieves a diversity gain and is expected to help reducing the MER of the system.

1.4.4 Proposed protocol

Once the relay has decoded the message $\mathbf{m}_k$, it is able to help the source by transmitting the appropriate packets related to the message $\mathbf{m}_k$ to the destination. Hence, the source trusts the relay in managing the message $\mathbf{m}_k$, and it decides to send a new message $\mathbf{m}_{k+1}$ in parallel, as depicted in Fig. 1.8.

We hope that this protocol will provide a better throughput, but the protocol corresponds to the following trade off: *i)* the data rate should increase since two messages are being sent simultaneously, but *ii)* more retransmissions are likely to occur due to the

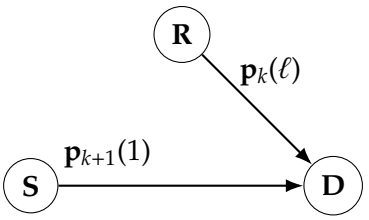


Figure 1.8 – Superposition of the transmitted packets at time-slot t using the investigated protocol.

interference between these two simultaneous incoming packets from the relay and the source, which degrades the detection at the destination.

More precisely, assuming that we are starting a protocol session, the source transmits and the relay overhears a sequence of packets $\mathbf{p}_1(1)$, $\mathbf{p}_1(2)$,$\cdots$,$\mathbf{p}_1(\ell-1)$, where $\ell-1$ is the HARQ round from which the relay successfully decodes the message $\mathbf{m}_1$ while the destination fails to decode this message. At round ℓ, if $\ell \leq C$ the relay transmits packet $\mathbf{p}_1(\ell)$ to the destination while the source transmits the first packet of the next message, *i.e.*, $\mathbf{p}_2(1)$. The destination tries to decode both messages $\mathbf{m}_1$ and $\mathbf{m}_2$ simultaneously. A pair of ACK/NACK is sent by the destination to the relay and the source according to the success of decoding of each message. Let (N)ACK$_k$ be the (non)acknowledgment message associated with the message $\mathbf{m}_k$, $k \in \{1,2\}$.

- If ACK$_1$/ACK$_2$: the source sends $\mathbf{p}_3(1)$, the relay overhears,
- if ACK$_1$/NACK$_2$: the source sends $\mathbf{p}_2(2)$, the relay overhears,
- if NACK$_1$/ACK$_2$: the source sends $\mathbf{p}_3(1)$ and the relay transmits $\mathbf{p}_1(\ell+1)$ (if $\ell+1 \leq C$, else the relay overhears),
- if NACK$_1$/NACK$_2$: the source sends $\mathbf{p}_2(2)$ and the relay transmits $\mathbf{p}_1(\ell+1)$ (if $\ell+1 \leq C$, else the relay overhears),

and so on.

1.4.4.1 Decoder for the proposed protocol

Assuming that the destination is receiving during the t-th time-slot, the decoder relies on the set of observations $\mathcal{O}_t$ given by the B last time-slots, *i.e.*, from $t-B+1$ to t. The contributions of the decoded messages in previous time-slots are removed from the observations. In other words, the previously decoded messages are known to the receiver and are not considered as interference. Let the set $\mathcal{M}_t$ designate all the messages that are not decoded until time-slot t and having at least one transmitted packet within the observations set $\mathcal{O}_t$. Within $\mathcal{M}_t$, we select the subset of messages $\mathcal{D}_t$ whose first transmission is less old than D time-slots. For instance, if a message has been sent at least

once before the $(t - D + 1)$-th time-slot, the message is considered to be delayed and we stop trying to decode it. Consequently, based on the observations $\mathcal{O}_t$, we have a Multiple Access Channel (MAC) where the messages to decode are $\mathcal{D}_t$ and the other messages $\mathcal{M}_t/\mathcal{D}_t$ are not to decode but are seen as a structured interference and not as a Gaussian noise. Thus, the probability of successfully decoding messages in $\mathcal{D}_t$ is the probability that the rates belong to a MAC capacity region since we assume capacity-achieving codes.

A realization of the investigated protocol with $C = 2$, $B = 2$ and $D = 2$ is provided in Fig. 1.9.

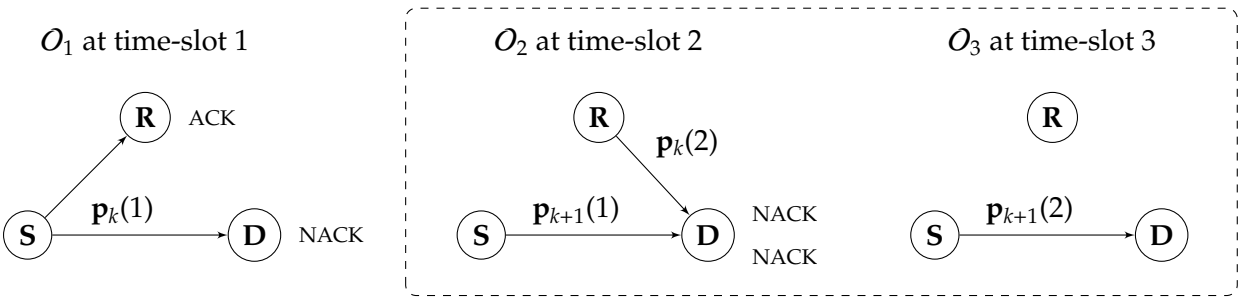


Figure 1.9 – Observation window at $t = 3$.

At $t = 3$, the decoder relies on the set of observations $\{\mathcal{O}_2, \mathcal{O}_3\}$ given by the last $B = 2$ time-slots. The received signals by the destination during this observation window are:

$$\mathbf{y}_2 = h_{sd}(2)\mathbf{p}_{k+1}(1) + h_{rd}(2)\mathbf{p}_k(2) + \mathbf{w}_2, \tag{1.18}$$

$$\mathbf{y}_3 = h_{sd}(3)\mathbf{p}_{k+1}(2) + \mathbf{w}_3. \tag{1.19}$$

In this case, there are no previously decoded messages. Furthermore, the set of undecoded messages in the observation window is $\mathcal{M}_t = \{\mathbf{m}_k, \mathbf{m}_{k+1}\}$. Notice that, at $t = 3$, $\mathbf{m}_k$ is older than $D = 2$ time-slots since it was transmitted first at $t = 1$ through $\mathbf{p}_k(1)$. Moreover, the destination tries to decode $\mathcal{D}_t = \{\mathbf{m}_{k+1}\}$ at $t = 3$ using the set of observations $\{\mathcal{O}_2, \mathcal{O}_3\}$, while $\mathbf{p}_k(2)$ is seen as structured interference since $\mathcal{M}_t/\mathcal{D}_t = \{\mathbf{m}_k\}$. Notice that $\mathbf{p}_k(2)$ might be decoded at $t = 3$, according to [Bandemer et al., 2012]. However, the corresponding message $\mathbf{m}_k$ is in time-out and will be seen by the upper layer as a dropped message anyway. Consequently, we distinguish four different events at $t = 3$, these are:

- θ_1: the destination decodes $\mathbf{m}_{k+1}$ and the structured interference $\mathbf{p}_k(2)$, corresponding to the message $\mathbf{m}_k$.
- θ_2: the destination decodes $\mathbf{m}_{k+1}$, but fails to decode the structured interference $\mathbf{p}_k(2)$.
- θ_3: the destination fails to decode $\mathbf{m}_{k+1}$, but decodes the structured interference $\mathbf{p}_k(2)$.

- θ_4: the destination fails to decode $\mathbf{m}_{k+1}$ and fails to decode the structured interference $\mathbf{p}_k(2)$.

The decoding outcome depends on the operating rate region of the system, as shown in Fig. 1.10 [Bandemer et al., 2012].

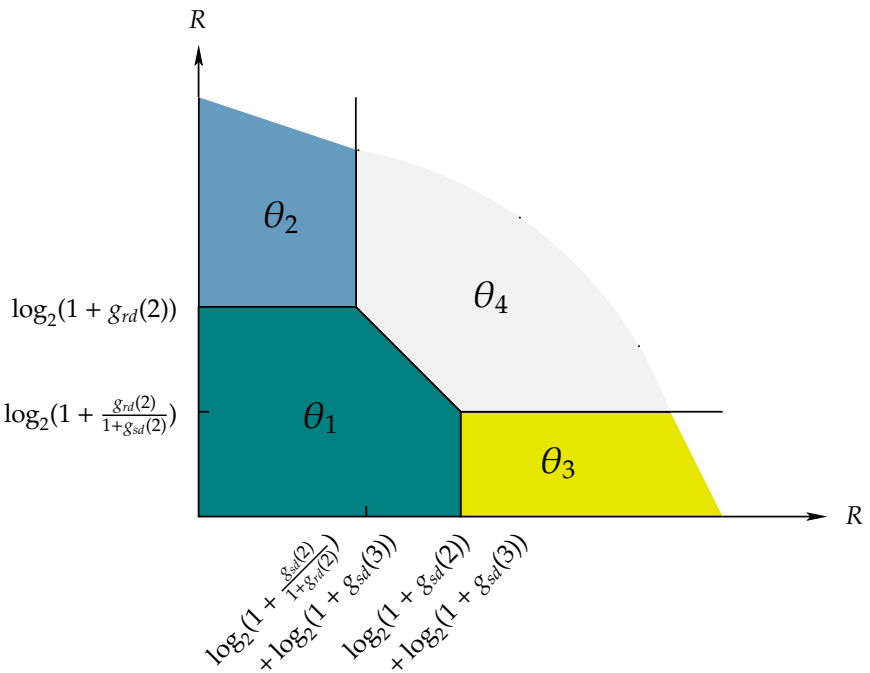


Figure 1.10 – The rate regions at $t = 3$.

In case of θ_1 or θ_2, an ACK feedback corresponding to message $\mathbf{m}_{k+1}$ is sent by the destination at $t = 3$. Otherwise, a NACK feedback corresponding to $\mathbf{m}_{k+1}$ is sent by the destination. Notice that $\mathbf{m}_k$ is dropped at $t = 3$ due to time-out.

1.5 Markov chain model of the investigated protocol

We model the investigated protocol as a Markov chain to derive the analytical expressions of its performance metrics, namely the throughput and MER.

Hereafter we will i) describe the corresponding Markov chain, ii) characterize the transition probabilities (assuming decoders described in Section 1.4.4.1), iii) derive the throughput and MER based on these transition probabilities. Unless otherwise stated, we assume $C = 2$, $B = 2$ and $D = 2$. These parameters are chosen to enable a trackable theoretical analysis, while providing a realistic overview of the system performance.

1.5.1 States of the Markov Chain

We provide the 8 states of the Markov chain needed for describing the protocol in Fig. 1.11.

(a) States $\mathbf{S}_1$ (left) and $\mathbf{S}_2$ (right).

(b) States $\mathbf{S}_3$ (left) and $\mathbf{S}_4$ (right).

(c) States $\mathbf{S}_5$ (left) and $\mathbf{S}_6$ (right).

(d) States $\mathbf{S}_7$ (left) and $\mathbf{S}_8$ (right).

Figure 1.11 – States of the Markov chain model of the investigated protocol.

Moreover, the states with odd indexes, correspond to a transmission by the source only.

- $\mathbf{S}_1$ is the transmission of packet $\mathbf{p}_k(1)$ or $\mathbf{p}_k(2)$ by the source and an ACK feedback by the destination.
- $\mathbf{S}_3$ is the transmission of packet $\mathbf{p}_k(2)$ (the last packet of message $\mathbf{m}_k$) by the source and NACK feedback by the destination.
- $\mathbf{S}_5$ is the transmission of packet $\mathbf{p}_k(1)$ by the source and NACK feedback by the relay and the destination.
- $\mathbf{S}_7$ is the transmission of packet $\mathbf{p}_k(1)$ by the source, an ACK feedback by the relay

and NACK feedback by the destination.

The states with even indexes, correspond to a transmission by the source and the relay, simultaneously. More precisely, the source transmits packet $\mathbf{p}_{k+1}(1)$ while the relay transmits packet $\mathbf{p}_k(2)$. These states differ by the feedback of the destination as follows.

- $\mathbf{S}_2$ corresponds to an ACK feedback of both packets by the destination.
- $\mathbf{S}_4$ corresponds to an ACK feedback of packet $\mathbf{p}_{k+1}(1)$ and NACK feedback of packet $\mathbf{p}_k(2)$ by the destination.
- $\mathbf{S}_6$ corresponds to a NACK feedback of packet $\mathbf{p}_{k+1}(1)$ and ACK feedback of packet $\mathbf{p}_k(2)$ by the destination.
- $\mathbf{S}_8$ corresponds to a NACK feedback of both packets by the destination.

In this way, each state characterizes one time-slot t. In addition, each state provides information about what happens in the previous time-slot $t-1$ and in the following time-slot $t+1$. For instance, we make the following observations.

- In $\mathbf{S}_1$ and $\mathbf{S}_3$, $\mathbf{p}_{k+1}(1)$ is necessarily sent at time-slot $t+1$.
- In $\mathbf{S}_2$ and $\mathbf{S}_4$, $\mathbf{p}_{k+2}(1)$ is necessarily sent at time-slot $t+1$.
- In $\mathbf{S}_5$, $\mathbf{p}_k(2)$ is necessarily sent at time-slot $t+1$.
- In $\mathbf{S}_7$, the relay and the source send $\mathbf{p}_k(2)$ and $\mathbf{p}_{k+1}(1)$ respectively at time-slot $t+1$.
- In $\mathbf{S}_6$ and $\mathbf{S}_8$, the source sends $\mathbf{p}_{k+1}(2)$ at time-slot $t+1$.

Moreover, in $\mathbf{S}_2$, $\mathbf{S}_4$, $\mathbf{S}_6$ and $\mathbf{S}_8$, the system was in $\mathbf{S}_7$ at time-slot $t-1$.

The possible transitions between these states are depicted in the state diagram in Fig. 1.12.

Extension to any value C can be done at the expense of the number of states. For instance, we succeeded to exhibit the states for $C=3$ and $C=4$ where their numbers are 19 and 49 respectively.

1.5.2 Transition matrix of the Markov Chain

Let $t_{i,j}$ denote the transition probability from $\mathbf{S}_j$ to $\mathbf{S}_i$. Consequently,

$$t_{i,j} = \Pr\left(\mathbf{S}_i \,\middle|\, \mathbf{S}_j\right), \tag{1.20}$$

and $\sum_{i=1}^{8} t_{i,j} = 1$. The transition matrix of the Markov chain describing the investigated protocol is denoted by $\mathbf{T} = (t_{i,j})_{(i,j)\in\{1,\cdots,8\}^2}$. It describes the transitions between the states

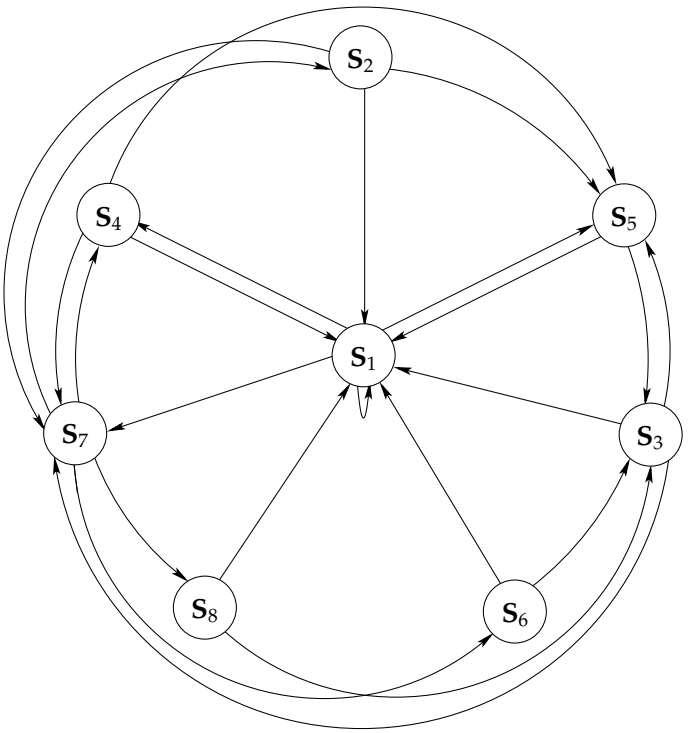


Figure 1.12 – States diagram of the Markov chain model of the investigated protocol.

of the Markov chain as in the states diagram in Fig. 1.12. This transition matrix is given in Eq. (1.21) as:

$$\mathbf{T} = \begin{array}{c} \begin{matrix} \mathbf{S}_1 & \mathbf{S}_2 & \mathbf{S}_3 & \mathbf{S}_4 & \mathbf{S}_5 & \mathbf{S}_6 & \mathbf{S}_7 & \mathbf{S}_8 \end{matrix} \\ \begin{pmatrix} t_{1,1} & t_{1,2} & t_{1,3} & t_{1,4} & t_{1,5} & t_{1,6} & 0 & t_{1,8} \\ 0 & 0 & 0 & 0 & 0 & 0 & t_{2,7} & 0 \\ 0 & 0 & 0 & 0 & t_{3,5} & t_{3,6} & 0 & t_{3,8} \\ 0 & 0 & 0 & 0 & 0 & 0 & t_{4,7} & 0 \\ t_{5,1} & t_{5,2} & t_{5,3} & t_{5,4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & t_{6,7} & 0 \\ t_{7,1} & t_{7,2} & t_{7,3} & t_{7,4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & t_{8,7} & 0 \end{pmatrix} \end{array}. \tag{1.21}$$

According to Bayes' rule ,

$$t_{i,j} = \frac{\Pr\left(\mathbf{S}_i, \mathbf{S}_j\right)}{\Pr\left(\mathbf{S}_j\right)}. \tag{1.22}$$

Except for $t_{1,8}$, the closed-form expressions of all other transition probabilities are provided in Eqs. (1.23), (1.24), (1.25), (1.27), (1.29), (1.30), (1.31) and (1.32). The analytical expression of $t_{1,8}$ is partially provided. This is due to improper integrals obtained in the derivation of the analytical expression of $t_{1,8}$. These improper integrals can only be

evaluated numerically via Monte Carlo simulations. The evaluation of $t_{1,8}$ is discussed furthermore in Section 1.6. The following Sections are dedicated to the derivation of the transition probabilities.

1.5.2.1 Transitions from $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$ and $\mathbf{S}_4$

From the states $\mathbf{S}_j$ where $j \in \{1, \dots 4\}$, the probability to go to states $\mathbf{S}_1$, $\mathbf{S}_5$ and $\mathbf{S}_7$ does not depend on the incoming state.

Derivations of $t_{1,j}$ for $j = 1, \cdots, 4$: $\mathbf{S}_1$ is the state where the transmitted packet by the source is successfully decoded at the destination. Therefore,

$$\begin{aligned} t_{1,j} &= \Pr\left(R < \log_2(1 + g_{sd}(1))\right) \\ &= \Pr\left(g_{sd}(1) > 2^R - 1\right) \\ &= \frac{1}{\sigma_{sd}^2} \int_{2^R-1}^{\infty} e^{-g/\sigma_{sd}^2} dg. \end{aligned}$$

Hence,

$$t_{1,j} = e^{-(2^R-1)/\sigma_{sd}^2}, j = 1, \cdots, 4. \tag{1.23}$$

Derivations of $t_{5,j}$ for $j = 1, \cdots, 4$: $\mathbf{S}_5$ is the state where the relay and the destination respond with NACK after the transmission of the first packet associated with a message. Therefore,

$$\begin{aligned} t_{5,j} &= \Pr\left(R > \log_2(1 + g_{sd}(1)), R > \log_2(1 + g_{sr}(1))\right) \\ &= \Pr\left(R > \log_2(1 + g_{sd}(1))\right) \Pr\left(R > \log_2(1 + g_{sr}(1))\right) \\ &= (1 - \Pr\left(R < \log_2(1 + g_{sd}(1))\right))(1 - \Pr\left(R < \log_2(1 + g_{sr}(1))\right)). \end{aligned}$$

Hence,

$$t_{5,j} = \left(1 - e^{-(2^R-1)/\sigma_{sd}^2}\right)\left(1 - e^{-(2^R-1)/\sigma_{sr}^2}\right), j = 1, \cdots, 4. \tag{1.24}$$

Derivations of $t_{7,j}$ for $j = 1, \cdots, 4$: $\mathbf{S}_7$ is the state where the relay succeeded but the destination failed to decode the first packet associated with a message. Therefore,

$$\begin{aligned} t_{7,j} &= \Pr\left(R > \log_2(1 + g_{sd}(1)), R < \log_2(1 + g_{sr}(1))\right) \\ &= \Pr\left(R > \log_2(1 + g_{sd}(1))\right) \Pr\left(R < \log_2(1 + g_{sr}(1))\right) \\ &= (1 - \Pr\left(R < \log_2(1 + g_{sd}(1))\right))\Pr\left(R < \log_2(1 + g_{sr}(1))\right). \end{aligned}$$

Hence,

$$t_{7,j} = \left(1 - e^{-(2^R-1)/\sigma_{sd}^2}\right) e^{-(2^R-1)/\sigma_{sr}^2} = 1 - t_{1,j} - t_{5,j}, j = 1, \cdots, 4. \tag{1.25}$$

1.5.2.2 Transitions from $\mathbf{S}_5$

We remind that, in state $\mathbf{S}_5$, the first packet associated with a message has not been decoded neither at the relay nor at the destination. This implies, in the second time-slot, a retransmission of a second packet associated with the same message from the source to the destination and the relay. The following observations are available to the destination at state $\mathbf{S}_5$:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} h_{sd}(1)\mathbf{1}_N & 0 \\ 0 & h_{sd}(2)\mathbf{1}_N \end{bmatrix} \begin{bmatrix} \mathbf{p}_k(1) \\ \mathbf{p}_k(2) \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}.$$

Thus, only transitions to $\mathbf{S}_1$ or $\mathbf{S}_3$ are possible. In the following, we derive $t_{1,5}$. We also have $t_{3,5} = 1 - t_{1,5}$.

Derivations of $t_{1,5}$: In this term, the probability of going to state $\mathbf{S}_1$ depends on state $\mathbf{S}_5$. By using the Bayes' rule, we obtain:

$$t_{1,5} = \frac{\Pr(\mathbf{S}_1, \mathbf{S}_5)}{\Pr(\mathbf{S}_5)},$$

where

$$\begin{aligned} \Pr(\mathbf{S}_5) &= \Pr\left(R > \log_2(1+g_{sd}(1)), R > \log_2(1+g_{sr}(1))\right) \\ &= \left(1 - e^{-(2^R-1)/\sigma_{sd}^2}\right)\left(1 - e^{-(2^R-1)/\sigma_{sr}^2}\right) \end{aligned}$$

and

$$\Pr(\mathbf{S}_1, \mathbf{S}_5) = \Pr\begin{pmatrix} R < \log_2(1+g_{sd}(1)) + \log_2(1+g_{sd}(2), \\ R > \log_2(1+g_{sd}(1)), R > \log_2(1+g_{sr}(1)) \end{pmatrix}.$$

By rewriting the inequalities in $\Pr(\mathbf{S}_1, \mathbf{S}_5)$, we obtain:

$$\begin{aligned} \Pr(\mathbf{S}_1, \mathbf{S}_5) &= \Pr\left((1+g_{sd}(1))(1+g_{sd}(2)) > 2^R, g_{sd}(1) < 2^R - 1, g_{sr}(1) < 2^R - 1\right) \\ &= \Pr\left((1+g_{sd}(1))(1+g_{sd}(2)) > 2^R, g_{sd}(1) < 2^R - 1\right) \Pr\left(g_{sr}(1) < 2^R - 1\right), \end{aligned}$$

since $g_{sr}(1)$ is independent of the other gains. Moreover,

$$\Pr\left(g_{sr}(1) < 2^R - 1\right) = \left(1 - e^{-(2^R-1)/\sigma_{sr}^2}\right).$$

To simplify the notations, we set $q_1 = g_{sd}(1)$ and $q_2 = g_{sd}(2)$. Then, the expression $\Pr\left((1+g_{sd}(1))(1+g_{sd}(2)) > 2^R, g_{sd}(1) < 2^R - 1\right)$ can be rewritten as:

$$\begin{aligned} \Pr\left((1+q_1)(1+q_2) > 2^R, q_1 < 2^R - 1\right) &= \Pr\left(q_2 > \frac{2^R}{1+q_1} - 1, q_2 > 0, q_1 < 2^R - 1\right) \\ &= \Pr\left(q_2 > \left(\frac{2^R}{1+q_1} - 1\right)^+, q_1 < 2^R - 1\right). \end{aligned}$$

Hence,

$$\Pr\Big((1+q_1)(1+q_2) > 2^R, q_1 < 2^R - 1\Big) = \int_{0<q_1<2^R-1} \Pr\left(q_2 > \left(\frac{2^R}{1+q_1} - 1\right)^+ \middle| q_1\right) p(q_1) dq_1.$$

One can notice that $\left(\frac{2^R}{1+q_1} - 1\right)^+ = \frac{2^R}{1+q_1} - 1$ for $q_1 < 2^R - 1$. Thus,

$$\begin{aligned}
\Pr\Big((1+q_1)(1+q_2) > 2^R, q_1 < 2^R - 1\Big) &= \int_{0<q_1<2^R-1} \int_{\frac{2^R}{1+q_1}-1<q_2} p(q_2)p(q_1)dq_2dq_1 \\
&= \int_{0<q_1<2^R-1} \int_{\frac{2^R}{1+q_1}-1<q_2} \frac{1}{\sigma_{sd}^2} e^{-q_2/\sigma_{sd}^2} dq_2 p(q_1)dq_1 \\
&= \int_{0<q_1<2^R-1} \left[-e^{-q_2/\sigma_{sd}^2}\right]_{\frac{2^R}{1+q_1}-1}^{\infty} p(q_1)dq_1 \\
&= \int_{0<q_1<2^R-1} e^{-(\frac{2^R}{1+q_1}-1)/\sigma_{sd}^2} \frac{1}{\sigma_{sd}^2} e^{-q_1/\sigma_{sd}^2} dq_1 \\
&= \frac{1}{\sigma_{sd}^2} e^{1/\sigma_{sd}^2} \int_0^{2^R-1} e^{-\frac{2^R}{\sigma_{sd}^2(1+q)} - \frac{q}{\sigma_{sd}^2}} dq.
\end{aligned}$$

By setting

$$U(\alpha, \beta, \gamma) = \frac{1}{\gamma} e^{1/\gamma} \int_0^{\alpha} e^{-\frac{\beta}{1+q} - \frac{q}{\gamma}} dq, \tag{1.26}$$

which can be evaluated as a generalized Fox's H function [Yilmaz and Alouini, 2010], we obtain:

$$\Pr\Big((1+q_1)(1+q_2) > 2^R, q_1 < 2^R - 1\Big) = U(2^R - 1, 2^R/\sigma_{sd}^2, \sigma_{sd}^2).$$

Hence,

$$t_{1,5} = \frac{U(2^R - 1, 2^R/\sigma_{sd}^2, \sigma_{sd}^2)}{1 - e^{-(2^R-1)/\sigma_{sd}^2}}. \tag{1.27}$$

1.5.2.3 Transitions from $\mathbf{S}_6$

State $\mathbf{S}_6$ corresponds to the transmission of the second packet associated with message $\mathbf{m}_k$ by the relay (which has decoded $\mathbf{m}_k$ in the previous time-slot $t-1$ while the destination failed to decode it) and the transmission of the first packet of $\mathbf{m}_{k+1}$ at time-slot t by the source. At time-slot $t+1$, only the second packet of the message $\mathbf{m}_{k+1}$, *i.e.* $\mathbf{p}_{k+1}(2)$, is transmitted by the source. The successful decoding of $\mathbf{m}_{k+1}$ corresponds to a transition from $\mathbf{S}_6$ to $\mathbf{S}_1$, and occurs with probability $t_{1,6}$. Otherwise, a NACK feedback corresponds to a transition from $\mathbf{S}_6$ to $\mathbf{S}_3$, and occurs with probability $t_{3,6} = 1 - t_{1,6}$. Using Bayes' rule,

$$t_{1,6} = \frac{\Pr(\mathbf{S}_1, \mathbf{S}_6)}{\Pr(\mathbf{S}_6)}.$$

At state $\mathbf{S}_6$, the following observations are available to the destination:

$$\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \end{bmatrix} = \begin{bmatrix} h_{sd}(0)\mathbf{1}_N & 0 & 0 \\ 0 & h_{rd}(1)\mathbf{1}_N & h_{sd}(1)\mathbf{1}_N \end{bmatrix} \begin{bmatrix} \mathbf{p}_k(1) \\ \mathbf{p}_k(2) \\ \mathbf{p}_{k+1}(1) \end{bmatrix} + \begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_1 \end{bmatrix}.$$

Moreover, $\mathbf{S}_6$ is described by the following inequalities:

$$\begin{cases} R & > & \log_2(1+g_{sd}(0)); \\ R & < & \log_2(1+g_{sr}(0)); \\ R & < & \log_2(1+g_{sd}(0)) + \log_2(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}); \\ R & > & \log_2(1+g_{sd}(1)). \end{cases}$$

Hence,

$$\Pr(\mathbf{S}_6) = \Pr\left(\begin{array}{c} R > \log_2(1+g_{sd}(0)), R < \log_2(1+g_{sr}(0)), \\ R < \log_2(1+g_{sd}(0)) + \log_2\left(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}\right), R > \log_2(1+g_{sd}(1)) \end{array}\right).$$

The term $g_{sr}(0)$ is independent of the other gains. Thus,

$$\begin{aligned} \Pr(\mathbf{S}_6) &= \Pr\left(R < \log_2(1+g_{sr}(0))\right) \Pr\left(\begin{array}{c} R > \log_2(1+g_{sd}(0)), \\ R < \log_2(1+g_{sd}(0)) + \log_2\left(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}\right), \\ R > \log_2(1+g_{sd}(1)) \end{array}\right) \\ &= \mathfrak{p}_1 \Pr\left(R < \log_2(1+g_{sr}(0))\right) \end{aligned}$$

with

$$\Pr\left(R < \log_2(1+g_{sr}(0))\right) = \Pr\left(g_{sr}(0) > 2^R - 1\right) = e^{-(2^R-1)/\sigma_{sr}^2},$$

and

$$\mathfrak{p}_1 = \Pr\left(g_{sd}(0) < 2^R - 1, (1+g_{sd}(0))\left(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}\right) > 2^R, g_{sd}(1) < 2^R - 1\right).$$

We set $q_1 = g_{sd}(0)$, $q_2 = g_{sd}(1)$ and $q_3 = g_{rd}(1)$. Then,

$$\begin{aligned} \mathfrak{p}_1 &= \Pr\left(q_1 < 2^R - 1, (1+q_1)\left(1+\frac{q_3}{1+q_2}\right) > 2^R, q_2 < 2^R - 1\right) \\ &= \Pr\left(q_1 < 2^R - 1, q_2 < 2^R - 1, q_3 > (1+q_2)(\frac{2^R}{1+q_1} - 1)^+\right). \end{aligned}$$

We define

$$V(\alpha, \beta_1, \beta_2, \beta_3, \gamma) = \frac{1}{\gamma} e^{1/\gamma} \int_0^\alpha \frac{e^{-\frac{\beta_1}{1+q}}}{\frac{\beta_2}{1+q} + \beta_3} e^{-q/\gamma} dq. \tag{1.28}$$

Also $\frac{2^R}{1+q_1} - 1 > 0$ for $q_1 < 2^R - 1$. Consequently,

$$\begin{aligned}
\mathfrak{p}_1 &= \Pr\left(q_1 < 2^R - 1, q_2 < 2^R - 1, q_3 > (1+q_2)(\frac{2^R}{1+q_1} - 1)\right) \\
&= \frac{1}{\sigma_{sd}^4 \sigma_{rd}^2} \int_{q_1=0}^{2^R-1} \int_{q_2=0}^{2^R-1} \int_{q_3>(1+q_2)(\frac{2^R}{1+q_1}-1)}^{\infty} e^{-q_1/\sigma_{sd}^2} e^{-q_2/\sigma_{sd}^2} e^{-q_3/\sigma_{rd}^2} dq_1 dq_2 dq_3 \\
&= \frac{1}{\sigma_{sd}^4} \int_{q_1=0}^{2^R-1} \int_{q_2=0}^{2^R-1} e^{-q_1/\sigma_{sd}^2} e^{-q_2/\sigma_{sd}^2} e^{-(1+q_2)(\frac{2^R}{1+q_1}-1)/\sigma_{rd}^2} dq_1 dq_2 \\
&= \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4} \int_{q_1=0}^{2^R-1} e^{-q_1/\sigma_{sd}^2} e^{-\frac{2^R}{\sigma_{rd}^2(1+q_1)}} \int_{q_2=0}^{2^R-1} e^{-q_2/\sigma_{sd}^2} e^{-q_2(\frac{2^R}{1+q_1}-1)/\sigma_{rd}^2} dq_1 dq_2 \\
&= \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4} \int_0^{2^R-1} e^{-\frac{2^R}{\sigma_{rd}^2(1+q)}} \frac{1}{\frac{2^R}{\sigma_{rd}^2(1+q)} + 1/\sigma_{sd}^2 - 1/\sigma_{rd}^2} e^{-q/\sigma_{sd}^2} dq \\
&- \frac{e^{1/\sigma_{sd}^2} e^{2^R(1/\sigma_{rd}^2 - 1/\sigma_{sd}^2)}}{\sigma_{sd}^4} \int_0^{2^R-1} \frac{1}{\frac{2^R}{\sigma_{rd}^2(1+q)} + 1/\sigma_{sd}^2 - 1/\sigma_{rd}^2} e^{-\frac{2^{2R}}{\sigma_{rd}^2(1+q)}} e^{-q/\sigma_{sd}^2} dq \\
&= \frac{e^{(1/\sigma_{rd}^2 - 1/\sigma_{sd}^2)}}{\sigma_{sd}^2} V(2^R - 1, \frac{2^R}{\sigma_{rd}^2}, \frac{2^R}{\sigma_{rd}^2}, 1/\sigma_{sd}^2 - 1/\sigma_{rd}^2, \sigma_{sd}^2) \\
&- \frac{e^{2^R(1/\sigma_{rd}^2 - 1/\sigma_{sd}^2)}}{\sigma_{sd}^2} V(2^R - 1, \frac{2^{2R}}{\sigma_{rd}^2}, \frac{2^R}{\sigma_{rd}^2}, 1/\sigma_{sd}^2 - 1/\sigma_{rd}^2, \sigma_{sd}^2).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pr(\mathbf{S}_6) &= e^{-(2^R-1)/\sigma_{sr}^2} \Big(\frac{e^{(1/\sigma_{rd}^2 - 1/\sigma_{sd}^2)}}{\sigma_{sd}^2} V(2^R - 1, \frac{2^R}{\sigma_{rd}^2}, \frac{2^R}{\sigma_{rd}^2}, 1/\sigma_{sd}^2 - 1/\sigma_{rd}^2, \sigma_{sd}^2) \\
&- \frac{e^{2^R(1/\sigma_{rd}^2 - 1/\sigma_{sd}^2)}}{\sigma_{sd}^2} V(2^R - 1, \frac{2^{2R}}{\sigma_{rd}^2}, \frac{2^R}{\sigma_{rd}^2}, 1/\sigma_{sd}^2 - 1/\sigma_{rd}^2, \sigma_{sd}^2) \Big).
\end{aligned}$$

Derivations of $\Pr(\mathbf{S}_1, \mathbf{S}_6)$: A transition from state $\mathbf{S}_6$ to state $\mathbf{S}_1$ occurs if $\mathbf{m}_{k+1}$ is successfully decoded at the destination. The message $\mathbf{m}_k$ has been previously decoded and can be omitted in the equations of the observations. Thus, the available observations at the destination are given by:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} h_{sd}(1)\mathbf{1}_N & 0 \\ 0 & h_{sd}(2)\mathbf{1}_N \end{bmatrix} \begin{bmatrix} \mathbf{p}_{k+1}(1) \\ \mathbf{p}_{k+1}(2) \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}.$$

Moreover, the transition from $\mathbf{S}_6$ to $\mathbf{S}_1$ occurs if:

$$\begin{cases}
R > \log_2(1 + g_{sd}(0)); \\
R < \log_2(1 + g_{sr}(0)); \\
R < \log_2(1 + g_{sd}(0)) + \log_2\left(1 + \frac{g_{rd}(1)}{1+g_{sd}(1)}\right); \\
R > \log_2(1 + g_{sd}(1)); \\
R < \log_2(1 + g_{sd}(1)) + \log_2(1 + g_{sd}(2)).
\end{cases}$$

The first four inequalities correspond to being in state $\mathbf{S}_6$ while the fifth inequality describes the transition to $\mathbf{S}_1$. Therefore,

$$\Pr(\mathbf{S}_1, \mathbf{S}_6) = \mathfrak{p}_2 \Pr\left(R < \log_2(1 + g_{sr}(0))\right),$$

with

$$\mathfrak{p}_2 = \Pr\begin{pmatrix} R > \log_2(1 + g_{sd}(0)), R < \log_2(1 + g_{sd}(0)) + \log_2\left(1 + \frac{g_{rd}(1)}{1+g_{sd}(1)}\right), \\ R > \log_2(1 + g_{sd}(1)), R < \log_2(1 + g_{sd}(1)) + \log_2(1 + g_{sd}(2)) \end{pmatrix}.$$

We set $q_1 = g_{sd}(0)$, $q_2 = g_{sd}(1)$, $q_3 = g_{rd}(1)$ and $q_4 = g_{sd}(2)$. Then,

$$\begin{aligned}
\mathfrak{p}_2 &= \Pr\begin{pmatrix} R > \log_2(1 + q_1), R < \log_2(1 + q_1) + \log_2\left(1 + \frac{q_3}{1+q_2}\right), R > \log_2(1 + q_2), \\ R < \log_2(1 + q_2) + \log_2(1 + q_4) \end{pmatrix} \\
&= \Pr\left(q_1 < 2^R - 1, q_2 < 2^R - 1, q_3 > (1 + q_2)(\frac{2^R}{1 + q_1} - 1)^+, q_4 > (\frac{2^R}{1 + q_2} - 1)^+\right) \\
&= \Pr\left(q_1 < 2^R - 1, q_2 < 2^R - 1, q_3 > (1 + q_2)(\frac{2^R}{1 + q_1} - 1), q_4 > \frac{2^R}{1 + q_2} - 1\right) \\
&= \frac{1}{\sigma_{sd}^6 \sigma_{rd}^2} \int_{q_1=0}^{2^R-1} \int_{q_2=0}^{2^R-1} \int_{q_3=(1+q_2)(\frac{2^R}{1+q_1}-1)}^{\infty} \int_{q_4=\frac{2^R}{1+q_2}-1}^{\infty} e^{-q_1/\sigma_{sd}^2} e^{-q_2/\sigma_{sd}^2} e^{-q_3/\sigma_{rd}^2} e^{-q_4/\sigma_{sd}^2} dq_1 dq_2 dq_3 dq_4 \\
&= \frac{e^{1/\sigma_{sd}^2}}{\sigma_{sd}^4 \sigma_{rd}^2} \int_{q_1=0}^{2^R-1} \int_{q_2=0}^{2^R-1} \int_{q_3=(1+q_2)(\frac{2^R}{1+q_1}-1)}^{\infty} e^{-q_1/\sigma_{sd}^2} e^{-q_2/\sigma_{sd}^2} e^{-q_3/\sigma_{rd}^2} e^{-\frac{2^R}{\sigma_{sd}^2(1+q_2)}} dq_1 dq_2 dq_3 \\
&= \frac{e^{1/\sigma_{sd}^2}}{\sigma_{sd}^4} \int_{q_1=0}^{2^R-1} \int_{q_2=0}^{2^R-1} e^{-q_1/\sigma_{sd}^2} e^{-q_2/\sigma_{sd}^2} e^{-\frac{2^R}{\sigma_{sd}^2(1+q_2)}} e^{-(1+q_2)(\frac{2^R}{1+q_1}-1)/\sigma_{rd}^2} dq_1 dq_2 \\
&= \frac{e^{(1/\sigma_{sd}^2+1/\sigma_{rd}^2)}}{\sigma_{sd}^4} \int_{q_2=0}^{2^R-1} e^{-\frac{2^R}{\sigma_{sd}^2(1+q_2)}} e^{-q_2(1/\sigma_{sd}^2-1/\sigma_{rd}^2)} \int_{q_1=0}^{2^R-1} e^{-q_1/\sigma_{sd}^2} e^{-(1+q_2)\frac{2^R}{\sigma_{rd}^2(1+q_1)}} dq_1 dq_2 \\
&= \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^2} \int_0^{2^R-1} U(2^R - 1, \frac{2^R(1 + q)}{\sigma_{rd}^2}, \sigma_{sd}^2) e^{-\frac{2^R}{\sigma_{sd}^2(1+q)}} e^{-q(1/\sigma_{sd}^2-1/\sigma_{rd}^2)} dq.
\end{aligned}$$

Hence,

$$t_{1,6} = \frac{e^{\frac{1}{\sigma_{rd}^2}} \int_0^{2^R-1} U(2^R - 1, \frac{2^R(1+q)}{\sigma_{rd}^2}, \sigma_{sd}^2) e^{-\frac{2^R}{\sigma_{sd}^2(1+q)}} e^{-q(\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2})} dq}{e^{(\frac{1}{\sigma_{rd}^2}-\frac{1}{\sigma_{sd}^2})} V(2^R - 1, \frac{2^R}{\sigma_{rd}^2}, \frac{2^R}{\sigma_{rd}^2}, \frac{1}{\sigma_{sd}^2} - \frac{1}{\sigma_{rd}^2}, \sigma_{sd}^2) - e^{2^R(\frac{1}{\sigma_{rd}^2}-\frac{1}{\sigma_{sd}^2})} V(2^R - 1, \frac{2^{2R}}{\sigma_{rd}^2}, \frac{2^R}{\sigma_{rd}^2}, \frac{1}{\sigma_{sd}^2} - \frac{1}{\sigma_{rd}^2}, \sigma_{sd}^2)}. \quad (1.29)$$

1.5.2.4 Transitions from $\mathbf{S}_7$

In $\mathbf{S}_7$, the relay succeeded to decode the first packet of the message while the destination failed to decode this packet. More precisely, the following observation is available to the destination at $\mathbf{S}_7$:

$$\mathbf{y}_1 = h_{sd}(1)\mathbf{1}_N \mathbf{p}_k(1) + \mathbf{w}_1.$$

Hence, $\mathbf{S}_7$ is described by the following inequalities:

$$\begin{cases} R & > & \log_2(1+g_{sd}(1)); \\ R & < & \log_2(1+g_{sr}(1)). \end{cases}$$

Therefore,

$$\begin{aligned} \Pr(\mathbf{S}_7) &= \Pr\left(R > \log_2(1+g_{sd}(1)), R < \log_2(1+g_{sr}(1))\right) \\ &= \Pr\left(R > \log_2(1+g_{sd}(1)\right) \Pr\left(R < \log_2(1+g_{sr}(1))\right) \\ &= \left(1-e^{-(2^R-1)/\sigma_{sd}^2}\right) e^{-(2^R-1)/\sigma_{sr}^2}. \end{aligned}$$

Consequently, during the second time-slot, the relay retransmits the second packet of the message $\mathbf{m}_k$, *i.e.* $\mathbf{p}_k(2)$, while the source transmits the first packet of a new message $\mathbf{m}_{k+1}$, *i.e.* $\mathbf{p}_{k+1}(1)$. Thus, a transition to one of the states $\mathbf{S}_2$, $\mathbf{S}_4$, $\mathbf{S}_6$, or $\mathbf{S}_8$ is possible. In the following, we derive $t_{2,7}$, $t_{4,7}$ and $t_{6,7}$. We also have $t_{8,7} = 1 - t_{2,7} - t_{4,7} - t_{6,7}$. By using Bayes' rule,

$$t_{j,7} = \frac{\Pr\left(\mathbf{S}_j, \mathbf{S}_7\right)}{\Pr(\mathbf{S}_7)}, j = 2,4,7,8.$$

Derivations of $t_{2,7}$: The transition from state $\mathbf{S}_7$ to state $\mathbf{S}_2$ occurs if both incoming messages from the source and relay are successfully decoded at the destination. Following the transmission of $\mathbf{p}_{k+1}(1)$ by the source and $\mathbf{p}_k(2)$ by the relay, the available observations at the destination are:

$$\begin{aligned} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} &= \begin{bmatrix} h_{sd}(1)\mathbf{1}_N & 0 & 0 \\ 0 & h_{rd}(2)\mathbf{1}_N & h_{sd}(2)\mathbf{1}_N \end{bmatrix} \begin{bmatrix} \mathbf{p}_k(1) \\ \mathbf{p}_k(2) \\ \mathbf{p}_{k+1}(1) \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} \\ &= \begin{bmatrix} h_{sd}(1)\mathbf{1}_N & 0 \\ 0 & h_{rd}(2)\mathbf{1}_N \end{bmatrix} \begin{bmatrix} \mathbf{p}_k(1) \\ \mathbf{p}_k(2) \end{bmatrix} + \begin{bmatrix} 0 \\ h_{sd}(2)\mathbf{1}_N \end{bmatrix} \mathbf{p}_{k+1}(1) + \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}. \end{aligned}$$

This is a MISO MAC where a transition from $\mathbf{S}_7$ to $\mathbf{S}_2$ occurs if:

$$\begin{cases} R & > & \log_2(1+g_{sd}(1)); \\ R & < & \log_2(1+g_{sr}(1)); \\ R & < & \log_2(1+g_{sd}(1)) + \log_2(1+g_{rd}(2)); \\ R & < & \log_2(1+g_{sd}(2)); \\ 2R & < & \log_2(1+g_{sd}(1)) + \log_2(1+g_{rd}(2)+g_{sd}(2)). \end{cases}$$

The first two inequalities correspond to being in state $\mathbf{S}_7$ while the last three inequalities describe the transition to $\mathbf{S}_2$. Consequently,

$$\Pr(\mathbf{S}_2, \mathbf{S}_7) = \Pr\begin{pmatrix} R < \log_2(1+g_{sd}(2)), R < \log_2(1+g_{sd}(1)) + \log_2(1+g_{rd}(2)), \\ 2R < \log_2(1+g_{sd}(1)) + \log_2(1+g_{rd}(2)+g_{sd}(2)), \\ R > \log_2(1+g_{sd}(1)), R < \log_2(1+g_{sr}(1)) \end{pmatrix}.$$

Since $g_{sr}(1)$ is independent of other gains,

$$\Pr(\mathbf{S}_2, \mathbf{S}_7) \quad = \quad \mathfrak{p}_3 \times \Pr\left(R < \log_2(1 + g_{sr}(1))\right),$$

where $\Pr\left(R < \log_2(1 + g_{sr}(1))\right) = e^{-(2^R-1)/\sigma_{sr}^2}$, and

$$\begin{aligned}\mathfrak{p}_3 &= \Pr\begin{pmatrix} R < \log_2(1 + g_{sd}(2)), R < \log_2(1 + g_{sd}(1)) + \log_2(1 + g_{rd}(2)), \\ 2R < \log_2(1 + g_{sd}(1)) + \log_2(1 + g_{rd}(2) + g_{sd}(2), R > \log_2(1 + g_{sd}(1)) \end{pmatrix} \\ &= \Pr\begin{pmatrix} 2^R - 1 < g_{sd}(2), 2^R < (1 + g_{sd}(1))(1 + g_{rd}(2)), \\ 2^{2R} < (1 + g_{sd}(1))(1 + g_{rd}(2) + g_{sd}(2)), 2^R - 1 > g_{sd}(1) \end{pmatrix}.\end{aligned}$$

We simplify the notations using $q_1 = g_{sd}(1)$, $q_2 = g_{sd}(2)$ and $q_3 = g_{rd}(2)$. Therefore,

$$\begin{aligned}\mathfrak{p}_3 &= \Pr\left(2^R - 1 < q_2, 2^R < (1 + q_1)(1 + q_3), 2^{2R} < (1 + q_1)(1 + q_3 + q_2), 2^R - 1 > q_1\right) \\ &= \Pr\left(q_1 < 2^R - 1, q_2 > 2^R - 1, q_3 > \max(2^R/(1 + q_1) - 1, 2^{2R}/(1 + q_1) - 1 - q_2)\right) \\ &= \int_{q_1=0}^{2^R-1} \int_{q_2=2^R-1}^{\infty} \int_{q_3=\max(2^R/(1+q_1)-1, 2^{2R}/(1+q_1)-1-q_2)}^{\infty} p(q_3)p(q_2)p(q_1)dq_3dq_2dq_1.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathfrak{p}_3 &= \frac{1}{\sigma_{sd}^4\sigma_{rd}^2} \int_{q_1=0}^{2^R-1} \int_{q_2=2^R-1}^{\infty} \int_{q_3=(\max(2^R/(1+q_1), 2^{2R}/(1+q_1)-q_2)-1))^+}^{\infty} e^{-q_3/\sigma_{rd}^2 - (q_1+q_2)/\sigma_{sd}^2} dq_3dq_2dq_1 \\ &= \frac{1}{\sigma_{sd}^4} \int_{q_1=0}^{2^R-1} \int_{q_2=2^R-1}^{\infty} e^{-(\max(2^R/(1+q_1), 2^{2R}/(1+q_1)-q_2)-1))^+/\sigma_{rd}^2} e^{-(q_1+q_2)/\sigma_{sd}^2} dq_2dq_1.\end{aligned}$$

In order to simplify $\mathfrak{p}_3$, we analyze the following term:

$$\left(\max\left(\frac{2^R}{1 + q_1} - 1, \frac{2^{2R}}{1 + q_1} - q_2 - 1\right)\right)^+$$

when $q_1 < 2^R - 1$ and $q_2 > 2^R - 1$.

Since $q_1 < 2^R - 1$, the term $\frac{2^R}{1+q_1} - 1$ is positive. Thus,

$$\left(\max\left(\frac{2^R}{1 + q_1} - 1, \frac{2^{2R}}{1 + q_1} - q_2 - 1\right)\right)^+ = \max\left(\frac{2^R}{1 + q_1} - 1, \frac{2^{2R}}{1 + q_1} - q_2 - 1\right).$$

The first term is the maximum if:

$$\begin{aligned}\frac{2^R}{1 + q_1} - 1 > \frac{2^{2R}}{1 + q_1} - q_2 - 1 \quad &\Leftrightarrow \quad \frac{2^R}{1 + q_1} > \frac{2^{2R}}{1 + q_1} - q_2 \\ &\Leftrightarrow \quad q_2 > \frac{2^R}{1 + q_1}\left(2^R - 1\right).\end{aligned}$$

Since $q_1 < 2^R - 1$, we have $2^R/(1 + q_1) > 1$. Therefore,

- if $q_2 \in [\frac{2^R}{1+q_1}\left(2^R - 1\right), \infty]$, then $\max\left(\frac{2^R}{1+q_1} - 1, \frac{2^{2R}}{1+q_1} - q_2 - 1\right) = \frac{2^R}{1+q_1} - 1$;

- if $q_2 \in [2^R-1, \frac{2^R}{1+q_1}\left(2^R-1\right)]$, then $\max\left(\frac{2^R}{1+q_1}-1, \frac{2^{2R}}{1+q_1}-q_2-1\right) = \frac{2^{2R}}{1+q_1}-q_2-1$.

Consequently,

$$\begin{aligned}
\mathfrak{p}_3 &= \frac{1}{\sigma_{sd}^4\sigma_{rd}^2}\int_{q_1=0}^{2^R-1}\int_{q_2=\frac{2^R}{1+q_1}(2^R-1)}^{\infty}\int_{q_3=2^R/(1+q_1)-1}^{\infty} e^{-q_3/\sigma_{rd}^2-(q_1+q_2)/\sigma_{sd}^2}dq_3dq_2dq_1 \\
&+ \frac{1}{\sigma_{sd}^4\sigma_{rd}^2}\int_{q_1=0}^{2^R-1}\int_{q_2=2^R-1}^{\frac{2^R}{1+q_1}(2^R-1)}\int_{q_3=2^R/(1+q_1)-q_2-1}^{\infty} e^{-q_3/\sigma_{rd}^2-(q_1+q_2)/\sigma_{sd}^2}dq_3dq_2dq_1 \\
&= \frac{1}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1}\int_{q_2=\frac{2^R}{1+q_1}(2^R-1)}^{\infty} e^{-(2^R/(1+q_1)-1)/\sigma_{rd}^2}e^{-(q_1+q_2)/\sigma_{sd}^2}dq_2dq_1 \\
&+ \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1}\int_{q_2=2^R-1}^{\frac{2^R}{1+q_1}(2^R-1)} e^{-(2^{2R}/(\sigma_{rd}^2(1+q_1))-q_1/\sigma_{sd}^2}e^{-q_2(1/\sigma_{sd}^2-1/\sigma_{rd}^2)}dq_2dq_1.
\end{aligned}$$

Then, we integrate the terms respectively to q_2, and we set $q = q_1$ for clarity purposes as follows:

$$\begin{aligned}
\mathfrak{p}_3 &= \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^2}\int_0^{2^R-1} e^{-2^R/(1+q)/\sigma_{rd}^2}e^{-\frac{2^R}{1+q}(2^R-1)/\sigma_{sd}^2}e^{-q/\sigma_{sd}^2}dq \\
&+ \frac{e^{1/\sigma_{rd}^2}e^{-(2^R-1)(1/\sigma_{sd}^2-1/\sigma_{rd}^2)}\sigma_{rd}^2}{\sigma_{sd}^2(\sigma_{rd}^2-\sigma_{sd}^2)}\int_0^{2^R-1} e^{-\frac{2^{2R}}{\sigma_{rd}^2(1+q)}}e^{-q/\sigma_{sd}^2}dq \\
&- \frac{e^{1/\sigma_{rd}^2}\sigma_{rd}^2}{\sigma_{sd}^2(\sigma_{rd}^2-\sigma_{sd}^2)}\int_0^{2^R-1} e^{-2^R/(1+q)/\sigma_{rd}^2}e^{-\frac{2^R}{1+q}(2^R-1)/\sigma_{sd}^2}e^{-q/\sigma_{sd}^2}dq \\
&= \frac{e^{1/\sigma_{rd}^2}e^{-(2^R-1)(1/\sigma_{sd}^2-1/\sigma_{rd}^2)}\sigma_{rd}^2}{\sigma_{sd}^2(\sigma_{rd}^2-\sigma_{sd}^2)}\int_0^{2^R-1} e^{-\frac{2^{2R}}{\sigma_{rd}^2(1+q)}}e^{-q/\sigma_{sd}^2}dq \\
&+ \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^2-\sigma_{rd}^2}\int_0^{2^R-1} e^{-2^R/(1+q)/\sigma_{rd}^2}e^{-\frac{2^R}{1+q}(2^R-1)/\sigma_{sd}^2}e^{-q/\sigma_{sd}^2}dq \\
&= \frac{\sigma_{rd}^2e^{2^R(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\sigma_{rd}^2-\sigma_{sd}^2}U\left(2^R-1, \frac{2^{2R}}{\sigma_{rd}^2}, \sigma_{sd}^2\right) \\
&+ \frac{\sigma_{sd}^2e^{(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\sigma_{sd}^2-\sigma_{rd}^2}U\left(2^R-1, \frac{2^R}{\sigma_{rd}^2}+\frac{2^R(2^R-1)}{\sigma_{sd}^2}, \sigma_{sd}^2\right).
\end{aligned}$$

Hence,

$$t_{2,7} = \frac{\sigma_{sd}^2e^{(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}U\left(2^R-1, \frac{2^R}{\sigma_{rd}^2}+\frac{2^R(2^R-1)}{\sigma_{sd}^2}, \sigma_{sd}^2\right) - \sigma_{rd}^2e^{2^R(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}U\left(2^R-1, \frac{2^{2R}}{\sigma_{rd}^2}, \sigma_{sd}^2\right)}{\left(\sigma_{sd}^2-\sigma_{rd}^2\right)\left(1-e^{-(2^R-1)/\sigma_{sd}^2}\right)}. \tag{1.30}$$

Derivations of $t_{4,7}$: The transition from state $\mathbf{S}_7$ to $\mathbf{S}_4$ occurs if message $\mathbf{m}_k$, coming from the relay via packet $\mathbf{p}_k(2)$, has not been decoded while message $\mathbf{m}_{k+1}$, coming from the

source via packet $\mathbf{p}_{k+1}(1)$, has been successfully decoded. We remind that following the transmission of $\mathbf{p}_{k+1}(1)$ by the source and $\mathbf{p}_k(2)$ by the relay, the available observations at the destination are:

$$\begin{aligned}\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} &= \begin{bmatrix}h_{sd}(1)\mathbf{1}_N & 0 & 0\\ 0 & h_{rd}(2)\mathbf{1}_N & h_{sd}(2)\mathbf{1}_N\end{bmatrix}\begin{bmatrix}\mathbf{p}_k(1)\\ \mathbf{p}_k(2)\\ \mathbf{p}_{k+1}(1)\end{bmatrix}+\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{bmatrix}\\ &= \begin{bmatrix}h_{sd}(1)\mathbf{1}_N & 0\\ 0 & h_{rd}(2)\mathbf{1}_N\end{bmatrix}\begin{bmatrix}\mathbf{p}_k(1)\\ \mathbf{p}_k(2)\end{bmatrix}+\begin{bmatrix}0\\ h_{sd}(2)\mathbf{1}_N\end{bmatrix}\mathbf{p}_{k+1}(1)+\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{bmatrix}.\end{aligned}$$

In this MISO MAC, the accumulated mutual information on the message $\mathbf{m}_k$ is smaller than the decoding rate while it is greater in the case of the message $\mathbf{m}_{k+1}$. Therefore, the transition from $\mathbf{S}_7$ to $\mathbf{S}_4$ occurs if:

$$\begin{cases} R > \log_2(1+g_{sd}(1)); \\ R < \log_2(1+g_{sr}(1)); \\ R > \log_2(1+g_{sd}(1)) + \log_2(1+g_{rd}(2)); \\ R < \log_2(1+\frac{g_{sd}(2)}{1+g_{rd}(2)}). \end{cases}$$

The first two inequalities correspond to being in state $\mathbf{S}_7$ while the last two inequalities describe the transition to $\mathbf{S}_4$. Thus,

$$\Pr(\mathbf{S}_4,\mathbf{S}_7) = \Pr\begin{pmatrix} R>\log_2(1+g_{sd}(1))+\log_2(1+g_{rd}(2)), R<\log_2(1+\frac{g_{sd}(2)}{1+g_{rd}(2)}), \\ R>\log_2(1+g_{sd}(1)), R<\log_2(1+g_{sr}(1)) \end{pmatrix}.$$

Also, $g_{sr}(1)$ is independent of the other gains. Consequently,

$$\Pr(\mathbf{S}_4,\mathbf{S}_7) = \mathfrak{p}_4 \Pr\left(R<\log_2(1+g_{sr}(1))\right),$$

with

$$\begin{aligned}\mathfrak{p}_4 &= \Pr\begin{pmatrix} R>\log_2(1+g_{sd}(1))+\log_2(1+g_{rd}(2)), R<\log_2(1+\frac{g_{sd}(2)}{1+g_{rd}(2)}), \\ R>\log_2(1+g_{sd}(1)) \end{pmatrix}\\ &= \Pr\left((1+g_{sd}(1))(1+g_{rd}(2))<2^R, \frac{g_{sd}(2)}{1+g_{rd}(2)}>2^R-1, g_{sd}(1)<2^R-1\right).\end{aligned}$$

Once again, we set $q_1=g_{sd}(1)$, $q_2=g_{sd}(2)$ and $q_3=g_{rd}(2)$. Therefore,

$$\begin{aligned}\mathfrak{p}_4 &= \Pr\left(q_3<\left(\frac{2^R}{1+q_1}-1\right)^+, q_2>(2^R-1)(1+q_3), q_1<2^R-1\right)\\ &= \int_{q_1=0}^{2^R-1}\int_{q_3=0}^{\left(\frac{2^R}{1+q_1}-1\right)^+}\int_{q_2=(2^R-1)(1+q_3)}^{\infty} p(q_1)p(q_2)p(q_3)dq_1dq_2dq_3.\end{aligned}$$

We have $\frac{2^R}{1+q_1} - 1 > 0$ for $q_1 < 2^R - 1$. Thus,

$$\begin{aligned}
\mathfrak{p}_4 &= \frac{1}{\sigma_{sd}^4\sigma_{rd}^2}\int_{q_1=0}^{2^R-1}\int_{q_3=0}^{\frac{2^R}{1+q_1}-1}\int_{q_2=(2^R-1)(1+q_3)}^{\infty} e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}e^{-q_3/\sigma_{rd}^2}dq_1dq_2dq_3 \\
&= \frac{e^{-(2^R-1)/\sigma_{sd}^2}}{\sigma_{sd}^2\sigma_{rd}^2}\int_{q_1=0}^{2^R-1} e^{-q_1/\sigma_{sd}^2}\int_{q_3=0}^{\frac{2^R}{1+q_1}-1} e^{-((2^R-1)/\sigma_{sd}^2+1/\sigma_{rd}^2)q_3}dq_1dq_3 \\
&= \frac{e^{-(2^R-1)/\sigma_{sd}^2}}{\sigma_{sd}^2\sigma_{rd}^2((2^R-1)/\sigma_{sd}^2+1/\sigma_{rd}^2)}\int_0^{2^R-1} e^{-q/\sigma_{sd}^2}dq \\
&- \frac{e^{-(2^R-1)/\sigma_{sd}^2}}{\sigma_{sd}^2\sigma_{rd}^2((2^R-1)/\sigma_{sd}^2+1/\sigma_{rd}^2)}\int_0^{2^R-1} e^{-q/\sigma_{sd}^2}e^{-((2^R-1)/\sigma_{sd}^2+1/\sigma_{rd}^2)(\frac{2^R}{1+q}-1)}dq \\
&= \frac{e^{-(2^R-1)/\sigma_{sd}^2}}{1+(2^R-1)\sigma_{rd}^2/\sigma_{sd}^2}\left(1-e^{-(2^R-1)/\sigma_{sd}^2}\right) \\
&- \frac{e^{(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{1+(2^R-1)\sigma_{rd}^2/\sigma_{sd}^2}U\left(2^R-1,\frac{2^{2R}}{\sigma_{sd}^2}+2^R\left(1/\sigma_{rd}^2-1/\sigma_{sd}^2\right),\sigma_{sd}^2\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
t_{4,7} &= \frac{e^{-(2^R-1)/\sigma_{sd}^2}}{1+(2^R-1)\sigma_{rd}^2/\sigma_{sd}^2} \\
&- \frac{e^{(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{(1+(2^R-1)\sigma_{rd}^2/\sigma_{sd}^2)(1-e^{-(2^R-1)/\sigma_{sd}^2})}U\left(2^R-1,\frac{2^{2R}}{\sigma_{sd}^2}+2^R\left(1/\sigma_{rd}^2-1/\sigma_{sd}^2\right),\sigma_{sd}^2\right).
\end{aligned} \tag{1.31}$$

Derivations of $t_{6,7}$: The transition to state $\mathbf{S}_6$ occurs if message $\mathbf{m}_k$, coming from the relay via packet $\mathbf{p}_k(2)$, has been successfully decoded while message $\mathbf{m}_{k+1}$, coming from the source via packet $\mathbf{p}_{k+1}(1)$, has not been decoded. Once again, we remind that following the transmission of $\mathbf{p}_{k+1}(1)$ by the source and $\mathbf{p}_k(2)$ by the relay, the available observations at the destination are:

$$\begin{aligned}
\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} &= \begin{bmatrix} h_{sd}(1)\mathbf{1}_N & 0 & 0 \\ 0 & h_{rd}(2)\mathbf{1}_N & h_{sd}(2)\mathbf{1}_N\end{bmatrix}\begin{bmatrix}\mathbf{p}_k(1)\\ \mathbf{p}_k(2)\\ \mathbf{p}_{k+1}(1)\end{bmatrix}+\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{bmatrix} \\
&= \begin{bmatrix} h_{sd}(1)\mathbf{1}_N & 0 \\ 0 & h_{rd}(2)\mathbf{1}_N\end{bmatrix}\begin{bmatrix}\mathbf{p}_k(1)\\ \mathbf{p}_k(2)\end{bmatrix}+\begin{bmatrix}0\\ h_{sd}(2)\mathbf{1}_N\end{bmatrix}\mathbf{p}_{k+1}(1)+\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{bmatrix}.
\end{aligned}$$

We notice that, unlike the state $\mathbf{S}_4$, the message $\mathbf{m}_k$ has been transmitted only once. In this MISO MAC, the message $\mathbf{m}_{k+1}$ has a rate smaller than the SIC corner point while the message $\mathbf{m}_k$ has a rate larger than the point where it is the only successfully decoded packet. Therefore the transition from $\mathbf{S}_7$ to $\mathbf{S}_6$ corresponds to the following inequalities:

$$\begin{cases}
R > \log_2(1+g_{sd}(1)); \\
R < \log_2(1+g_{sr}(1)); \\
R < \log_2(1+g_{sd}(1)) + \log_2\left(1+\frac{g_{rd}(2)}{1+g_{sd}(2)}\right); \\
R > \log_2(1+g_{sd}(2)).
\end{cases}$$

The first two inequalities correspond to being in state $\mathbf{S}_7$ while the last two inequalities describe the transition to $\mathbf{S}_6$. Thus,

$$\Pr(\mathbf{S}_6, \mathbf{S}_7) = \Pr\left(\begin{matrix} R < \log_2(1+g_{sd}(1)) + \log_2\left(1+\frac{g_{rd}(2)}{1+g_{sd}(2)}\right), R > \log_2(1+g_{sd}(2)), \\ R > \log_2(1+g_{sd}(1)), R < \log_2(1+g_{sr}(1)) \end{matrix}\right).$$

Once again, $g_{sr}(1)$ is independent of the other gains. Consequently,

$$\Pr(\mathbf{S}_6, \mathbf{S}_7) = \mathfrak{p}_5 \Pr\left(R < \log_2(1+g_{sr}(1))\right),$$

with

$$\begin{aligned} \mathfrak{p}_5 &= \Pr\left(\begin{matrix} R < \log_2(1+g_{sd}(1)) + \log_2\left(1+\frac{g_{rd}(2)}{1+g_{sd}(2)}\right), \\ R > \log_2(1+g_{sd}(2)), R > \log_2(1+g_{sd}(1)) \end{matrix}\right) \\ &= \Pr\left(2^R < (1+g_{sd}(1))\left(1+\frac{g_{rd}(2)}{1+g_{sd}(2)}\right), g_{sd}(2) < 2^R-1, g_{sd}(1) < 2^R-1\right). \end{aligned}$$

We set $q_1 = g_{sd}(1)$, $q_2 = g_{sd}(2)$ et $q_3 = g_{rd}(2)$. Therefore,

$$\begin{aligned} \mathfrak{p}_5 &= \Pr\left(q_3 > (1+q_2)\left(\frac{2^R}{1+q_1}-1\right), q_2 < 2^R-1, q_1 < 2^R-1\right) \\ &= \Pr\left(q_3 > (1+q_2)\left(\frac{2^R}{1+q_1}-1\right)^+, q_2 < 2^R-1, q_1 < 2^R-1\right) \\ &= \int_{q_1=0}^{2^R-1}\int_{q_2=0}^{2^R-1}\int_{q_3=(1+q_2)\left(\frac{2^R}{1+q_1}-1\right)^+}^{\infty} p(q_1)p(q_2)p(q_3)dq_1dq_2dq_3 \\ &= \frac{1}{\sigma_{sd}^4\sigma_{rd}^2}\int_{q_1=0}^{2^R-1}\int_{q_2=0}^{2^R-1}\int_{q_3=(1+q_2)\left(\frac{2^R}{1+q_1}-1\right)^+}^{\infty} e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}e^{-q_3/\sigma_{rd}^2}dq_1dq_2dq_3. \end{aligned}$$

And $\frac{2^R}{1+q_1}-1 > 0$, for $q_1 < 2^R-1$. Thus,

$$\begin{aligned} \mathfrak{p}_5 &= \frac{1}{\sigma_{sd}^4\sigma_{rd}^2}\int_{q_1=0}^{2^R-1}\int_{q_2=0}^{2^R-1}\int_{q_3=(1+q_2)(\frac{2^R}{1+q_1}-1)}^{\infty} e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}e^{-q_3/\sigma_{rd}^2}dq_1dq_2dq_3 \\ &= \frac{1}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1}\int_{q_2=0}^{2^R-1} e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}e^{-(1+q_2)(\frac{2^R}{1+q_1}-1)/\sigma_{rd}^2}dq_1dq_2 \\ &= \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1} e^{-q_1/\sigma_{sd}^2}e^{-\frac{2^R}{\sigma_{rd}^2(1+q_1)}}\int_{q_2=0}^{2^R-1} e^{-q_2\left(\frac{2^R}{\sigma_{rd}^2(1+q_1)}+(1/\sigma_{sd}^2-1/\sigma_{rd}^2)\right)}dq_1dq_2 \\ &= \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\int_0^{2^R-1}\frac{e^{-\frac{2^R}{\sigma_{rd}^2(1+q)}}}{\frac{2^R}{\sigma_{rd}^2(1+q)}+(1/\sigma_{sd}^2-1/\sigma_{rd}^2)}e^{-q/\sigma_{sd}^2}dq \\ &\quad - \frac{e^{1/\sigma_{sd}^2}e^{2^R(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\sigma_{sd}^4}\int_0^{2^R-1}\frac{e^{-\frac{2^{2R}}{\sigma_{rd}^2(1+q)}}}{\frac{2^R}{\sigma_{rd}^2(1+q)}+(1/\sigma_{sd}^2-1/\sigma_{rd}^2)}e^{-q/\sigma_{sd}^2}dq. \end{aligned}$$

Using Eq. (1.28),

$$\begin{aligned}\mathfrak{p}_5 &= \frac{e^{(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\sigma_{sd}^2}V\left(2^R-1,\frac{2^R}{\sigma_{rd}^2},\frac{2^R}{\sigma_{rd}^2},1/\sigma_{sd}^2-1/\sigma_{rd}^2,\sigma_{sd}^2\right)\\ &- \frac{e^{2^R(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\sigma_{sd}^2}V\left(2^R-1,\frac{2^{2R}}{\sigma_{rd}^2},\frac{2^R}{\sigma_{rd}^2},1/\sigma_{sd}^2-1/\sigma_{rd}^2,\sigma_{sd}^2\right).\end{aligned}$$

Hence,

$$t_{6,7}=\frac{e^{\left(\frac{1}{\sigma_{rd}^2}-\frac{1}{\sigma_{sd}^2}\right)}V\left(2^R-1,\frac{2^R}{\sigma_{rd}^2},\frac{2^R}{\sigma_{rd}^2},\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2},\sigma_{sd}^2\right)-e^{\left(\frac{2^R}{\sigma_{rd}^2}-\frac{2^R}{\sigma_{sd}^2}\right)}V\left(2^R-1,\frac{2^{2R}}{\sigma_{rd}^2},\frac{2^R}{\sigma_{rd}^2},\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2},\sigma_{sd}^2\right)}{\sigma_{sd}^2\left(1-e^{-(2^R-1)/\sigma_{sd}^2}\right)}. \tag{1.32}$$

1.5.2.5 Transitions from $\mathbf{S}_8$

State $\mathbf{S}_8$ corresponds to the transmission of the second packet of the message $\mathbf{m}_k$ by the relay (which has decoded $\mathbf{m}_k$ at time-slot t but the decoding has failed at the destination) and the transmission of the first packet of the message $\mathbf{m}_{k+1}$ at time-slot $t+1$. Moreover, none of these two messages has been decoded at the destination. In this way, the following observations are available to the destination at $\mathbf{S}_8$:

$$\begin{aligned}\begin{bmatrix}\mathbf{y}_0\\ \mathbf{y}_1\end{bmatrix} &= \begin{bmatrix}h_{sd}(0)\mathbf{1}_N & 0 & 0\\ 0 & h_{rd}(1)\mathbf{1}_N & h_{sd}(1)\mathbf{1}_N\end{bmatrix}\begin{bmatrix}\mathbf{p}_k(1)\\ \mathbf{p}_k(2)\\ \mathbf{p}_{k+1}(1)\end{bmatrix}+\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_1\end{bmatrix}\\ &= \begin{bmatrix}h_{sd}(0)\mathbf{1}_N & 0\\ 0 & h_{rd}(1)\mathbf{1}_N\end{bmatrix}\begin{bmatrix}\mathbf{p}_k(1)\\ \mathbf{p}_k(2)\end{bmatrix}+\begin{bmatrix}0\\ h_{sd}(1)\mathbf{1}_N\end{bmatrix}\mathbf{p}_{k+1}(1)+\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_1\end{bmatrix}.\end{aligned}$$

Hence, $\mathbf{S}_8$ is described by the following inequalities:

$$\begin{cases}R &> \log_2(1+g_{sd}(0));\\ R &< \log_2(1+g_{sr}(0));\\ R &> \log_2(1+g_{sd}(0))+\log_2(1+\frac{g_{rd}(1)}{1+g_{sd}(1)});\\ 2R &> \log_2(1+g_{sd}(0))+\log_2(1+g_{rd}(1)+g_{sd}(1));\\ R &> \log_2(1+\frac{g_{sd}(1)}{1+g_{rd}(1)}).\end{cases}$$

Therefore,

$$\Pr(\mathbf{S}_8) = \Pr\begin{pmatrix}R>\log_2(1+g_{sd}(0)), R<\log_2(1+g_{sr}(0)),\\ R>\log_2(1+g_{sd}(0))+\log_2(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}),\\ 2R>\log_2(1+g_{sd}(0))+\log_2(1+g_{rd}(1)+g_{sd}(1)), R>\log_2(1+\frac{g_{sd}(1)}{1+g_{rd}(1)})\end{pmatrix}.$$

At time-slot $t+2$, only the second packet of the message $\mathbf{m}_{k+1}$, *i.e.* $\mathbf{p}_{k+1}(2)$, is transmitted by the source. If $\mathbf{m}_{k+1}$ is successfully decoded, a transition from $\mathbf{S}_8$ to $\mathbf{S}_1$ occurs. Otherwise, a transition from $\mathbf{S}_8$ to $\mathbf{S}_3$ occurs. We derive $t_{1,8}$ and we have $t_{3,8}=1-t_{1,8}$. Using Bayes' rule,

$$t_{1,8}=\frac{\Pr(\mathbf{S}_1,\mathbf{S}_8)}{\Pr(\mathbf{S}_8)}.$$

Derivation of $\mathbf{Pr}(\mathbf{S}_8)$: We remind that,

$$\Pr(\mathbf{S}_8) = \Pr\begin{pmatrix} R > \log_2(1+g_{sd}(0)), R < \log_2(1+g_{sr}(0)), \\ R > \log_2(1+g_{sd}(0)) + \log_2(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}), \\ 2R > \log_2(1+g_{sd}(0)) + \log_2(1+g_{rd}(1)+g_{sd}(1)), R > \log_2(1+\frac{g_{sd}(1)}{1+g_{rd}(1)}) \end{pmatrix}.$$

The term $g_{sr}(0)$ is independent of the other gains. Thus,

$$\Pr(\mathbf{S}_8) = \Pr\left(R < \log_2(1+g_{sr}(0))\right) \mathfrak{p}_6,$$

with

$$\Pr\left(R < \log_2(1+g_{sr}(0))\right) = \Pr\left(g_{sr}(0) > 2^R - 1\right) = e^{-(2^R-1)/\sigma_{sr}^2},$$

and

$$\begin{aligned} \mathfrak{p}_6 &= \Pr\begin{pmatrix} R > \log_2(1+g_{sd}(0)), \\ R > \log_2(1+g_{sd}(0)) + \log_2(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}), \\ 2R > \log_2(1+g_{sd}(0)) + \log_2(1+g_{rd}(1)+g_{sd}(1)), \\ R > \log_2(1+\frac{g_{sd}(1)}{1+g_{rd}(1)}) \end{pmatrix} \\ &= \Pr\begin{pmatrix} g_{sd}(0) < 2^R - 1, (1+g_{sd}(0))(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}) < 2^R, \\ (1+g_{sd}(0))(1+g_{rd}(1)+g_{sd}(1)) < 2^{2R}, 1+\frac{g_{sd}(1)}{1+g_{rd}(1)} < 2^R \end{pmatrix}. \end{aligned}$$

We set $q_1 = g_{sd}(0)$, $q_2 = g_{sd}(1)$ and $q_3 = g_{rd}(1)$. Then,

$$\begin{aligned} \mathfrak{p}_6 &= \Pr\left(q_1 < 2^R - 1, (1+q_1)(1+\frac{q_3}{1+q_2}) < 2^R, (1+q_1)(1+q_2+q_3) < 2^{2R}, 1+\frac{q_2}{1+q_3} < 2^R\right) \\ &= \Pr\left(q_1 < 2^R - 1, q_3 < (1+q_2)(\frac{2^R}{1+q_1} - 1), q_3 < \frac{2^{2R}}{1+q_1} - q_2 - 1, q_3 > \frac{q_2}{2^R-1} - 1\right) \\ &= \Pr\left(q_1 < 2^R - 1, q_3 < \left(\min\left((1+q_2)(\frac{2^R}{1+q_1} - 1), \frac{2^{2R}}{1+q_1} - q_2 - 1\right)\right)^+, q_3 > \left(\frac{q_2}{2^R-1} - 1\right)^+\right). \end{aligned}$$

We analyze the following term:

$$\left(\min\left((1+q_2)(\frac{2^R}{1+q_1} - 1), \frac{2^{2R}}{1+q_1} - q_2 - 1\right)\right)^+$$

with the constraint on q_1 ($< 2^R - 1$). Moreover,

$$\begin{aligned} (1+q_2)(\frac{2^R}{1+q_1} - 1) < \frac{2^{2R}}{1+q_1} - q_2 - 1 &\Leftrightarrow (1+q_2)\frac{2^R}{1+q_1} < \frac{2^{2R}}{1+q_1} \\ &\Leftrightarrow q_2 < (2^R - 1). \end{aligned}$$

Consequently,

$$\begin{aligned} \mathfrak{p}_6 &= \Pr\left(q_1 < 2^R - 1, q_2 < 2^R - 1, q_3 < (1+q_2)(\frac{2^R}{1+q_1} - 1), q_3 > \left(\frac{q_2}{2^R-1} - 1\right)^+\right) \\ &+ \Pr\left(q_1 < 2^R - 1, q_2 > 2^R - 1, q_3 < \left(\frac{2^{2R}}{1+q_1} - q_2 - 1\right)^+, q_3 > \left(\frac{q_2}{2^R-1} - 1\right)^+\right). \end{aligned}$$

Also $\left(\frac{q_2}{2^R-1}-1\right)^+ = 0$ for $q_2 < 2^R - 1$ which implies:

$$\begin{aligned}\mathfrak{p}_6 &= \Pr\left(q_1 < 2^R-1, q_2 < 2^R-1, q_3 < (1+q_2)(\frac{2^R}{1+q_1}-1)\right)\\ &+ \Pr\left(q_1 < 2^R-1, q_2 > 2^R-1, q_3 < \left(\frac{2^{2R}}{1+q_1}-q_2-1\right)^+, q_3 > \frac{q_2}{2^R-1}-1\right).\end{aligned}$$

The first term is easy to derive. The second term is non-null only if:

$$0 < \frac{q_2}{2^R-1} - 1 < \frac{2^{2R}}{1+q_1} - q_2 - 1.$$

The first inequality is satisfied due to the constraint on q_2. The second inequality leads to:

$$q_2 < (2^R-1)\frac{2^R}{1+q_1},$$

which is greater than $2^R - 1$ due to the constraint on q_1. Thus,

$$\mathfrak{p}_6 = \mathfrak{p}_{6,1} + \mathfrak{p}_{6,2},$$

with

$$\begin{aligned}\mathfrak{p}_{6,1} &= \Pr\left(q_1 < 2^R-1, q_2 < 2^R-1, q_3 < (1+q_2)(\frac{2^R}{1+q_1}-1)\right),\\ \mathfrak{p}_{6,2} &= \Pr\left(q_1 < 2^R-1, q_2 > 2^R-1, q_2 < (2^R-1)\frac{2^R}{1+q_1}, q_3 < \frac{2^{2R}}{1+q_1}-q_2-1, q_3 > \frac{q_2}{2^R-1}-1\right).\end{aligned}$$

We derive $\mathfrak{p}_{6,1}$ as follows:

$$\begin{aligned}\mathfrak{p}_{6,1} &= \frac{1}{\sigma_{sd}^4\sigma_{rd}^2}\int_{q_1=0}^{2^R-1}\int_{q_2=0}^{2^R-1}\int_{q_3=0}^{(1+q_2)(2^R/(1+q_1)-1)} e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}e^{-q_3/\sigma_{rd}^2}dq_1dq_2dq_3\\ &= \frac{1}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1}\int_{q_2=0}^{2^R-1} e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}\left(1-e^{-(1+q_2)(2^R/(1+q_1)-1)/\sigma_{rd}^2}\right)dq_1dq_2\\ &= \frac{1}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1}\int_{q_2=0}^{2^R-1} e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}dq_1dq_2\\ &- \frac{1}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1}\int_{q_2=0}^{2^R-1} e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}e^{-(1+q_2)(2^R/(1+q_1)-1)/\sigma_{rd}^2}dq_1dq_2\\ &= \left(1-e^{-(2^R-1)/\sigma_{sd}^2}\right)^2\\ &- \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1}\int_{q_2=0}^{2^R-1} e^{-2^R/(\sigma_{rd}^2(1+q_1))}e^{-q_1/\sigma_{sd}^2}e^{-q_2(1/\sigma_{sd}^2-1/\sigma_{rd}^2)}e^{-q_2\frac{2^R}{\sigma_{rd}^2(1+q_1)}}dq_1dq_2,\end{aligned}$$

which implies that:

$$\begin{aligned}
\mathfrak{p}_{6,1} &= \left(1-e^{-(2^R-1)/\sigma_{sd}^2}\right)^2 \\
&- \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\int_0^{2^R-1}\frac{1}{(1/\sigma_{sd}^2-1/\sigma_{rd}^2)+\frac{2^R}{\sigma_{rd}^2(1+q)}}e^{-2^R/(\sigma_{rd}^2(1+q))}e^{-q/\sigma_{sd}^2}dq \\
&+ \frac{e^{1/\sigma_{sd}^2}e^{2^R(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\sigma_{sd}^4}\int_0^{2^R-1}\frac{1}{(1/\sigma_{sd}^2-1/\sigma_{rd}^2)+\frac{2^R}{\sigma_{rd}^2(1+q)}}e^{-\frac{2^{2R}}{\sigma_{rd}^2(1+q)}}e^{-\frac{q}{\sigma_{sd}^2}}dq.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathfrak{p}_{6,1} &= \left(1-e^{-(2^R-1)/\sigma_{sd}^2}\right)^2 \\
&- \frac{e^{1/\sigma_{rd}^2-1/\sigma_{sd}^2}}{\sigma_{sd}^2}V(2^R-1,\frac{2^R}{\sigma_{rd}^2},\frac{2^R}{\sigma_{rd}^2},\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2},\sigma_{sd}^2) \\
&+ \frac{e^{2^R(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\sigma_{sd}^2}V(2^R-1,\frac{2^{2R}}{\sigma_{rd}^2},\frac{2^R}{\sigma_{rd}^2},\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2},\sigma_{sd}^2).
\end{aligned}$$

Then, we derive $\mathfrak{p}_{6,2}$ as follows:

$$\begin{aligned}
\mathfrak{p}_{6,2} &= \frac{1}{\sigma_{sd}^4\sigma_{rd}^2}\int_{q_1=0}^{2^R-1}\int_{q_2=2^R-1}^{(2^R-1)2^R/(1+q_1)}\int_{q_3=\frac{q_2}{2^R-1}-1}^{\frac{2^{2R}}{1+q_1}-q_2-1}e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}e^{-q_3/\sigma_{rd}^2}dq_1dq_2dq_3 \\
&= \frac{1}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1}\int_{q_2=2^R-1}^{(2^R-1)2^R/(1+q_1)}e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}e^{-(\frac{q_2}{2^R-1}-1)/\sigma_{rd}^2}dq_1dq_2 \\
&- \frac{1}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1}\int_{q_2=2^R-1}^{(2^R-1)2^R/(1+q_1)}e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}e^{-(\frac{2^{2R}}{1+q_1}-q_2-1)/\sigma_{rd}^2}dq_1dq_2 \\
&= \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1}\int_{q_2=2^R-1}^{(2^R-1)2^R/(1+q_1)}e^{-q_1/\sigma_{sd}^2}e^{-q_2/\sigma_{sd}^2}e^{-\frac{q_2}{\sigma_{rd}^2(2^R-1)}}dq_1dq_2 \\
&- \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\int_{q_1=0}^{2^R-1}\int_{q_2=2^R-1}^{(2^R-1)2^R/(1+q_1)}e^{-\frac{2^{2R}}{\sigma_{rd}^2(1+q_1)}}e^{-q_1/\sigma_{sd}^2}e^{-q_2(1/\sigma_{sd}^2-1/\sigma_{rd}^2)}dq_1dq_2 \\
&= \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\frac{1}{\frac{1}{\sigma_{sd}^2}+\frac{1}{\sigma_{rd}^2(2^R-1)}}\int_0^{2^R-1}e^{-q/\sigma_{sd}^2}e^{-(2^R-1)/\sigma_{sd}^2}e^{-\frac{2^R-1}{\sigma_{rd}^2(2^R-1)}}dq \\
&- \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\frac{1}{\frac{1}{\sigma_{sd}^2}+\frac{1}{\sigma_{rd}^2(2^R-1)}}\int_0^{2^R-1}e^{-q/\sigma_{sd}^2}e^{-(2^R-1)2^R/(1+q)/\sigma_{sd}^2}e^{-\frac{(2^R-1)2^R/(1+q)}{\sigma_{rd}^2(2^R-1)}}dq \\
&- \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\frac{1}{\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2}}\int_0^{2^R-1}e^{-\frac{2^{2R}}{\sigma_{rd}^2(1+q)}}e^{-q/\sigma_{sd}^2}e^{-(2^R-1)(1/\sigma_{sd}^2-1/\sigma_{rd}^2)}dq \\
&+ \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\frac{1}{\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2}}\int_0^{2^R-1}e^{-\frac{2^{2R}}{\sigma_{rd}^2(1+q)}}e^{-q/\sigma_{sd}^2}e^{-(2^R-1)2^R/(1+q)(1/\sigma_{sd}^2-1/\sigma_{rd}^2)}dq.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathfrak{p}_{6,2} &= \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^2}\frac{e^{-(2^R-1)/\sigma_{sd}^2}e^{-\frac{2^R-1}{\sigma_{rd}^2(2^R-1)}}}{\frac{1}{\sigma_{sd}^2}+\frac{1}{\sigma_{rd}^2(2^R-1)}}\left(1-e^{-(2^R-1)/\sigma_{sd}^2}\right)\\
&- \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\frac{1}{\frac{1}{\sigma_{sd}^2}+\frac{1}{\sigma_{rd}^2(2^R-1)}}\int_0^{2^R-1} e^{-\frac{(2^R-1)2^R}{\sigma_{sd}^2(1+q)}}e^{-\frac{2^R}{\sigma_{rd}^2(1+q)}}e^{-q/\sigma_{sd}^2}dq\\
&- \frac{e^{1/\sigma_{sd}^2}}{\sigma_{sd}^4}\frac{e^{2^R(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2}}\int_0^{2^R-1} e^{-\frac{2^{2R}}{\sigma_{rd}^2(1+q)}}e^{-q/\sigma_{sd}^2}dq\\
&+ \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^4}\frac{1}{\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2}}\int_0^{2^R-1} e^{-\frac{2^{2R}}{\sigma_{rd}^2(1+q)}}e^{-\frac{(2^R-1)2^R(1/\sigma_{sd}^2-1/\sigma_{rd}^2)}{(1+q)}}e^{-q/\sigma_{sd}^2}dq.
\end{aligned}$$

We can write the last equation with respect to U as:

$$\begin{aligned}
\mathfrak{p}_{6,2} &= \frac{e^{1/\sigma_{rd}^2}}{\sigma_{sd}^2}\frac{e^{-(2^R-1)/\sigma_{sd}^2}e^{-\frac{2^R-1}{\sigma_{rd}^2(2^R-1)}}}{\frac{1}{\sigma_{sd}^2}+\frac{1}{\sigma_{rd}^2(2^R-1)}}\left(1-e^{-(2^R-1)/\sigma_{sd}^2}\right)\\
&- \frac{e^{1/\sigma_{rd}^2-1/\sigma_{sd}^2}}{\sigma_{sd}^2}\frac{1}{\frac{1}{\sigma_{sd}^2}+\frac{1}{\sigma_{rd}^2(2^R-1)}}U(2^R-1,\frac{2^{2R}}{\sigma_{sd}^2}+\frac{2^R}{\sigma_{rd}^2}-\frac{2^R}{\sigma_{sd}^2},\sigma_{sd}^2)\\
&- \frac{e^{2^R(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\sigma_{sd}^2}\frac{1}{\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2}}U(2^R-1,\frac{2^{2R}}{\sigma_{rd}^2},\sigma_{sd}^2)\\
&+ \frac{e^{1/\sigma_{rd}^2-1/\sigma_{sd}^2}}{\sigma_{sd}^2}\frac{1}{\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2}}U(2^R-1,\frac{2^{2R}}{\sigma_{sd}^2}+\frac{2^R}{\sigma_{rd}^2}-\frac{2^R}{\sigma_{sd}^2},\sigma_{sd}^2),
\end{aligned}$$

which leads to:

$$\begin{aligned}
\mathfrak{p}_{6,2} &= \frac{1}{\sigma_{sd}^2}\frac{e^{-(2^R-1)/\sigma_{sd}^2}}{\frac{1}{\sigma_{sd}^2}+\frac{1}{\sigma_{rd}^2(2^R-1)}}\left(1-e^{-(2^R-1)/\sigma_{sd}^2}\right)\\
&+ e^{1/\sigma_{rd}^2-1/\sigma_{sd}^2}\frac{2^R\sigma_{sd}^2\sigma_{rd}^2}{(\sigma_{rd}^2-\sigma_{sd}^2)(\sigma_{rd}^2(2^R-1)+\sigma_{sd}^2)}U(2^R-1,\frac{2^{2R}}{\sigma_{sd}^2}+\frac{2^R}{\sigma_{rd}^2}-\frac{2^R}{\sigma_{sd}^2},\sigma_{sd}^2)\\
&- \frac{e^{2^R(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\sigma_{sd}^2}\frac{1}{\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2}}U(2^R-1,\frac{2^{2R}}{\sigma_{rd}^2},\sigma_{sd}^2).
\end{aligned}$$

By gathering the derivations above, we obtain:

$$\Pr(\mathbf{S}_8) = e^{-(2^R-1)/\sigma_{sr}^2}\mathfrak{p}_6$$

with

$$\begin{aligned}
\mathfrak{p}_6 &= \left(1-\frac{\sigma_{sd}^2 e^{-(2^R-1)/\sigma_{sd}^2}}{\sigma_{rd}^2(2^R-1)+\sigma_{sd}^2}\right)\left(1-e^{-(2^R-1)/\sigma_{sd}^2}\right)\\
&- \frac{e^{1/\sigma_{rd}^2-1/\sigma_{sd}^2}}{\sigma_{sd}^2}V(2^R-1,\frac{2^R}{\sigma_{rd}^2},\frac{2^R}{\sigma_{rd}^2},\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2},\sigma_{sd}^2)\\
&+ \frac{e^{2^R(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\sigma_{sd}^2}V(2^R-1,\frac{2^{2R}}{\sigma_{rd}^2},\frac{2^R}{\sigma_{rd}^2},\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2},\sigma_{sd}^2)\\
&+ e^{1/\sigma_{rd}^2-1/\sigma_{sd}^2}\frac{2^R\sigma_{sd}^2\sigma_{rd}^2}{(\sigma_{rd}^2-\sigma_{sd}^2)(\sigma_{rd}^2(2^R-1)+\sigma_{sd}^2)}U(2^R-1,\frac{2^{2R}}{\sigma_{sd}^2}+\frac{2^R}{\sigma_{rd}^2}-\frac{2^R}{\sigma_{sd}^2},\sigma_{sd}^2)\\
&- \frac{e^{2^R(1/\sigma_{rd}^2-1/\sigma_{sd}^2)}}{\sigma_{sd}^2}\frac{1}{\frac{1}{\sigma_{sd}^2}-\frac{1}{\sigma_{rd}^2}}U(2^R-1,\frac{2^{2R}}{\sigma_{rd}^2},\sigma_{sd}^2).
\end{aligned}$$

Derivations of $\Pr(\mathbf{S}_1,\mathbf{S}_8)$: The transition from state $\mathbf{S}_8$ to state $\mathbf{S}_1$ occurs if $\mathbf{m}_{k+1}$ is successfully decoded at the destination. The available observations for decoding this message are the following:

$$\begin{aligned}
\begin{bmatrix}\mathbf{y}_0\\ \mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} &= \begin{bmatrix} h_{sd}(0)\mathbf{1}_N & 0 & 0 & 0\\ 0 & h_{rd}(1)\mathbf{1}_N & h_{sd}(1)\mathbf{1}_N & 0\\ 0 & 0 & 0 & h_{sd}(2)\mathbf{1}_N\end{bmatrix}\begin{bmatrix}\mathbf{p}_k(1)\\ \mathbf{p}_k(2)\\ \mathbf{p}_{k+1}(1)\\ \mathbf{p}_{k+1}(2)\end{bmatrix}+\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_1\\ \mathbf{w}_2\end{bmatrix}\\
&= \begin{bmatrix} h_{sd}(0)\mathbf{1}_N & 0\\ 0 & h_{rd}(1)\mathbf{1}_N\\ 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{p}_k(1)\\ \mathbf{p}_k(2)\end{bmatrix}+\begin{bmatrix} 0 & 0\\ h_{sd}(1)\mathbf{1}_N & 0\\ 0 & h_{sd}(2)\mathbf{1}_N\end{bmatrix}\begin{bmatrix}\mathbf{p}_{k+1}(1)\\ \mathbf{p}_{k+1}(2)\end{bmatrix}+\begin{bmatrix}\mathbf{w}_0\\ \mathbf{w}_1\\ \mathbf{w}_2\end{bmatrix}.
\end{aligned}$$

As $B=2$, we keep only the last two time-slots for decoding, therefore the transition to $\mathbf{S}_1$ is determined by the following (reduced) observations:

$$\begin{aligned}
\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} &= \begin{bmatrix} 0 & h_{rd}(1)\mathbf{1}_N & h_{sd}(1)\mathbf{1}_N & 0\\ 0 & 0 & 0 & h_{sd}(2)\mathbf{1}_N\end{bmatrix}\begin{bmatrix}\mathbf{p}_k(1)\\ \mathbf{p}_k(2)\\ \mathbf{p}_{k+1}(1)\\ \mathbf{p}_{k+1}(2)\end{bmatrix}+\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{bmatrix}\\
&= \begin{bmatrix} h_{rd}(1)\mathbf{1}_N\\ 0\end{bmatrix}\mathbf{p}_k(2)+\begin{bmatrix} h_{sd}(1)\mathbf{1}_N & 0\\ 0 & h_{sd}(2)\mathbf{1}_N\end{bmatrix}\begin{bmatrix}\mathbf{p}_{k+1}(1)\\ \mathbf{p}_{k+1}(2)\end{bmatrix}+\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{bmatrix}.
\end{aligned}$$

Moreover, the constraint $D=2$ is important in the previous equation as the destination may decode the message $\mathbf{m}_k$, but it will be delayed with 3 time-slots. Hence, $\mathbf{m}_k$ will be seen by the upper layer as a dropped message. Consequently, the transition from $\mathbf{S}_8$ to $\mathbf{S}_1$ corresponds to the occurrence of the event: i) Ω_1': the destination decodes both messages of $\mathbf{m}_k$ and $\mathbf{m}_{k+1}$, or ii) Ω_1'': the destination decodes only the message of index $\mathbf{m}_{k+1}$, while the message $\mathbf{m}_k$ is still in error. Therefore,

$$\Pr(\mathbf{S}_1,\mathbf{S}_8)=\Pr\left(\Omega_1',\mathbf{S}_8\right)+\Pr\left(\Omega_1'',\mathbf{S}_8\right).$$

Derivations of $\Pr\left(\Omega_1', \mathbf{S}_8\right)$

The event Ω_1' can be written as:

$$\Omega_1' = \begin{cases} R & < & \log_2(1+g_{rd}(1)); \\ R & < & \log_2(1+g_{sd}(1)) + \log_2(1+g_{sd}(2)); \\ 2R & < & \log_2(1+g_{rd}(1)+g_{sd}(1)) + \log_2(1+g_{sd}(2)). \end{cases}$$

In order to calculate $\Pr\left(\Omega_1', \mathbf{S}_8\right)$, we consider the inequalities that define the state $\mathbf{S}_8$ and the event Ω_1' as follows:

$$\Pr\left(\Omega_1', \mathbf{S}_8\right) = \Pr\begin{pmatrix} R < \log_2(1+g_{rd}(1)), R < \log_2(1+g_{sd}(1)) + \log_2(1+g_{sd}(2)), \\ 2R < \log_2(1+g_{rd}(1)+g_{sd}(1)) + \log_2(1+g_{sd}(2)), R > \log_2(1+g_{sd}(0)), \\ R < \log_2(1+g_{sr}(0)), R > \log_2(1+g_{sd}(0)) + \log_2(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}), \\ 2R > \log_2(1+g_{sd}(0)) + \log_2(1+g_{rd}(1)+g_{sd}(1)), R > \log_2(1+\frac{g_{sd}(1)}{1+g_{rd}(1)}) \end{pmatrix}.$$

Once again, $g_{sr}(0)$ is independent of other channel gains. Therefore,

$$\Pr\left(\Omega_1', \mathbf{S}_8\right) = \Pr\left(R < \log_2(1+g_{sr}(0))\right) \times \mathfrak{p}_7,$$

where

$$\Pr\left(R < \log_2(1+g_{sr}(0))\right) = e^{-(2^R-1)/\sigma_{sr}^2},$$

and

$$\mathfrak{p}_7 = \Pr\begin{pmatrix} R < \log_2(1+g_{rd}(1)), R < \log_2(1+g_{sd}(1)) + \log_2(1+g_{sd}(2)), \\ 2R < \log_2(1+g_{rd}(1)+g_{sd}(1)) + \log_2(1+g_{sd}(2)), R > \log_2(1+g_{sd}(0)), \\ R > \log_2(1+g_{sd}(0)) + \log_2(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}), \\ 2R > \log_2(1+g_{sd}(0)) + \log_2(1+g_{rd}(1)+g_{sd}(1)), R > \log_2(1+\frac{g_{sd}(1)}{1+g_{rd}(1)}) \end{pmatrix}.$$

We set $q_1 = g_{sd}(0)$, $q_2 = g_{sd}(1)$, $q_3 = g_{rd}(1)$ and $q_4 = g_{sd}(2)$. Then,

$$\mathfrak{p}_7 = \Pr\begin{pmatrix} 2^R - 1 < q_3, 2^R < (1+q_2)(1+q_4), 2^{2R} < (1+q_3+q_2)(1+q_4), \\ q_1 < 2^R - 1, 2^R > (1+q_1)(1+\frac{q_3}{1+q_2}), \\ 2^{2R} > (1+q_1)(1+q_2+q_3), 2^R - 1 > \frac{q_2}{1+q_3} \end{pmatrix}.$$

The expression of $\mathfrak{p}_7$ is evaluated via Monte Carlo simulations. Further simplification of this expression is possible but is useless since it leads to improper multiple integrals that can only be evaluated via Monte Carlo simulations.

Derivations of $\Pr\left(\Omega_1'', \mathbf{S}_8\right)$

The event Ω_1'' can be written as:

$$\Omega_1'' = \begin{cases} R & < & \log_2(1+\frac{g_{sd}(1)}{1+g_{rd}(1)}) + \log_2(1+g_{sd}(2)); \\ R & > & \log_2(1+g_{rd}(1)). \end{cases}$$

In order to calculate $\Pr\left(\Omega_1'', \mathbf{S}_8\right)$, we consider the inequalities that define the state $\mathbf{S}_8$ and the event Ω_1'' as follows:

$$\Pr\left(\Omega_1'', \mathbf{S}_8\right) = \Pr\begin{pmatrix} R < \log_2(1+\frac{g_{sd}(1)}{1+g_{rd}(1)}) + \log_2(1+g_{sd}(2)), R > \log_2(1+g_{rd}(1)), \\ R > \log_2(1+g_{sd}(0)), R < \log_2(1+g_{sr}(0)), \\ R > \log_2(1+g_{sd}(0)) + \log_2(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}), \\ 2R > \log_2(1+g_{sd}(0)) + \log_2(1+g_{rd}(1)+g_{sd}(1)), R > \log_2(1+\frac{g_{sd}(1)}{1+g_{rd}(1)}) \end{pmatrix}.$$

Once again, $g_{sr}(0)$ is independent of other gains. Therefore,

$$\Pr\left(\Omega_1'', \mathbf{S}_8\right) = \Pr\left(R < \log_2(1+g_{sr}(0)\right) \times \mathfrak{p}_8,$$

where

$$\Pr\left(R < \log_2(1+g_{sr}(0)\right) = e^{-(2^R-1)/\sigma_{sr}^2},$$

and

$$\mathfrak{p}_8 = \Pr\begin{pmatrix} R < \log_2(1+\frac{g_{sd}(1)}{1+g_{rd}(1)}) + \log_2(1+g_{sd}(2)), R > \log_2(1+g_{rd}(1)), \\ R > \log_2(1+g_{sd}(0)), R > \log_2(1+g_{sd}(0)) + \log_2(1+\frac{g_{rd}(1)}{1+g_{sd}(1)}), \\ 2R > \log_2(1+g_{sd}(0)) + \log_2(1+g_{rd}(1)+g_{sd}(1)), R > \log_2(1+\frac{g_{sd}(1)}{1+g_{rd}(1)}) \end{pmatrix}.$$

We set $q_1 = g_{sd}(0)$, $q_2 = g_{sd}(1)$, $q_3 = g_{rd}(1)$, and $q_4 = g_{sd}(2)$. If $q_4 > 2^R - 1$ then $2^R < (1+\frac{q_2}{1+q_3})(1+q_4)$, else $2^R < (1+\frac{q_2}{1+q_3})(1+q_4)$ for $q_3 < \frac{q_2}{\frac{2^R}{1+q_4}-1} - 1$. Therefore,

$$\mathfrak{p}_8 = \Pr\begin{pmatrix} 2^R < (1+\frac{q_2}{1+q_3})(1+q_4), 2^R - 1 > q_3, 2^R - 1 > q_1, \\ 2^R > (1+q_1)(1+\frac{q_3}{1+q_2}), 2^{2R} > (1+q_1)(1+q_2+q_3), 2^R - 1 > \frac{q_2}{1+q_3} \end{pmatrix}.$$

The expression of $\mathfrak{p}_8$ is also evaluated via Monte Carlo simulations. Further simplification of this expression is also possible but useless since it leads to improper multiple integrals that can only be evaluated via Monte Carlo simulations. Hence,

$$t_{1,8} = \frac{\Pr(S_1, S_8)}{\Pr(S_8)} = \frac{e^{-(2^R-1)/\sigma_{sr}^2}(\mathfrak{p}_7 + \mathfrak{p}_8)}{e^{-(2^R-1)/\sigma_{sr}^2}\mathfrak{p}_6} = \frac{\mathfrak{p}_7 + \mathfrak{p}_8}{\mathfrak{p}_6}, \tag{1.33}$$

with

$$\begin{aligned}
\mathfrak{p}_6 &= \Pr\begin{pmatrix} q_1 < 2^R - 1, (1+q_1)(1+\frac{q_3}{1+q_2}) < 2^R, \\ (1+q_1)(1+q_2+q_3) < 2^{2R}, 1+\frac{q_2}{1+q_3} < 2^R \end{pmatrix} \qquad (1.34)\\
&= \left(1 - \frac{\sigma_{sd}^2 e^{-(2^R-1)/\sigma_{sd}^2}}{\sigma_{rd}^2(2^R-1)+\sigma_{sd}^2}\right)\left(1 - e^{-(2^R-1)/\sigma_{sd}^2}\right)\\
&- \frac{e^{1/\sigma_{rd}^2 - 1/\sigma_{sd}^2}}{\sigma_{sd}^2} V(2^R-1, \frac{2^R}{\sigma_{rd}^2}, \frac{2^R}{\sigma_{rd}^2}, \frac{1}{\sigma_{sd}^2} - \frac{1}{\sigma_{rd}^2}, \sigma_{sd}^2)\\
&+ \frac{e^{2^R(1/\sigma_{rd}^2 - 1/\sigma_{sd}^2)}}{\sigma_{sd}^2} V(2^R-1, \frac{2^{2R}}{\sigma_{rd}^2}, \frac{2^R}{\sigma_{rd}^2}, \frac{1}{\sigma_{sd}^2} - \frac{1}{\sigma_{rd}^2}, \sigma_{sd}^2)\\
&+ e^{1/\sigma_{rd}^2 - 1/\sigma_{sd}^2} \frac{2^R \sigma_{sd}^2 \sigma_{rd}^2}{(\sigma_{rd}^2 - \sigma_{sd}^2)(\sigma_{rd}^2(2^R-1)+\sigma_{sd}^2)} U(2^R-1, \frac{2^{2R}}{\sigma_{sd}^2} + \frac{2^R}{\sigma_{rd}^2} - \frac{2^R}{\sigma_{sd}^2}, \sigma_{sd}^2)\\
&- \frac{e^{2^R(1/\sigma_{rd}^2 - 1/\sigma_{sd}^2)}}{\sigma_{sd}^2} \frac{1}{\frac{1}{\sigma_{sd}^2} - \frac{1}{\sigma_{rd}^2}} U(2^R-1, \frac{2^{2R}}{\sigma_{rd}^2}, \sigma_{sd}^2), \qquad (1.35)
\end{aligned}$$

$$\mathfrak{p}_7 = \Pr\begin{pmatrix} 2^R - 1 < q_3, 2^R < (1+q_2)(1+q_4), 2^{2R} < (1+q_3+q_2)(1+q_4), \\ q_1 < 2^R - 1, 2^R > (1+q_1)(1+\frac{q_3}{1+q_2}), \\ 2^{2R} > (1+q_1)(1+q_2+q_3), 2^R - 1 > \frac{q_2}{1+q_3} \end{pmatrix}, \qquad (1.36)$$

and

$$\mathfrak{p}_8 = \Pr\begin{pmatrix} 2^R < (1+\frac{q_2}{1+q_3})(1+q_4), 2^R - 1 > q_3, \\ 2^R - 1 > q_1, 2^R > (1+q_1)(1+\frac{q_3}{1+q_2}), \\ 2^{2R} > (1+q_1)(1+q_2+q_3), 2^R - 1 > \frac{q_2}{1+q_3} \end{pmatrix}. \qquad (1.37)$$

This concludes the derivations of the transition probabilities. The following step is to deduce the steady state vector that enables the computation of the performance metrics of the investigated protocol.

1.5.3 Analytical expressions of the performance metrics using the Markov chain model

Based on the transition matrix $\mathbf{T}$, we find the steady state vector $\boldsymbol{\pi}$ by solving $\mathbf{T}\boldsymbol{\pi} = \boldsymbol{\pi}$ and $\sum_{k=1}^{8} \pi_k = 1$, where $\boldsymbol{\pi} = [\pi_1, \cdots, \pi_8]$ denotes the steady state vector of the Markov chain

that describes the investigated protocol. The equation $\mathbf{T}\boldsymbol{\pi} = \boldsymbol{\pi}$ can be rewritten as:

$$\begin{bmatrix} t_{1,1} & t_{1,2} & t_{1,3} & t_{1,4} & t_{1,5} & t_{1,6} & 0 & t_{1,8} \\ 0 & 0 & 0 & 0 & 0 & 0 & t_{2,7} & 0 \\ 0 & 0 & 0 & 0 & t_{3,5} & t_{3,6} & 0 & t_{3,8} \\ 0 & 0 & 0 & 0 & 0 & 0 & t_{4,7} & 0 \\ t_{5,1} & t_{5,2} & t_{5,3} & t_{5,4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & t_{6,7} & 0 \\ t_{7,1} & t_{7,2} & t_{7,3} & t_{7,4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & t_{8,7} & 0 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \\ \pi_6 \\ \pi_7 \\ \pi_8 \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \\ \pi_6 \\ \pi_7 \\ \pi_8 \end{bmatrix}. \tag{1.38}$$

By solving this equation (numerically), we obtain the steady state probabilities. Then we can find the throughput and MER. The throughput, provided in Eq. (1.39), is obtained by summing the steady state probabilities of the states that lead to an ACK feedback. We note that π_2 is multiplied by 2 since $\mathbf{S}_2$ leads to acknowledge two messages simultaneously. The MER, provided in Eq. (1.40), is the ratio of the sum of the steady state probabilities of the states that lead to drop a message over the sum of the steady state probabilities of the states that lead to generate a new message. Hence,

$$\eta = R\left(\pi_1 + 2\pi_2 + \pi_4 + \pi_6\right), \tag{1.39}$$

$$\mu = \frac{\pi_3 + \pi_4 + \pi_8}{\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_6 + \pi_8}. \tag{1.40}$$

In this Section, we derived a Markov chain model that characterizes the investigated protocol and enables the derivation of its performance metrics. These performance metrics were derived using an information theoretic approach. A numerical validation follows in Section 1.6.

1.6 Numerical results and comparison of the protocols

We verify the analytical expressions of the performance metrics and illustrate numerically the interest of the investigated protocol. For this purpose, we describe in Section 1.6.1 the path loss model for numerical evaluation of the protocols. In Section 1.6.2, we validate numerically the correctness of the Markov chain model that describes the investigated protocol and its derived metrics. Then we compare, in Section 1.6.3, this protocol to other protocols presented throughout this Chapter. These numerical results provide insights on the usefulness of the investigated protocol in comparison to other relay assisted HARQ protocols. In addition, we study the performance of the investigated protocol with different relay positions in Section 1.6.4. Furthermore, numerical optimization of power allocation when using the investigated protocol is provided in Section 1.6.5. Afterwards in Section 1.6.6, we show simulation results of the protocols using practical codes.

1.6.1 Path loss model for numerical evaluation of the protocols

The path loss is dependent on the position of the relay relatively to the source and destination. According to [Proakis and Salehi, 2007], the path loss in a mobile radio channel is inversely proportional to d^p, where d is the distance between the transmitter and the receiver, and $2 \leq p \leq 4$. While the case of $p = 2$ corresponds to waves propagating through free space, propagation in mobile radio channel is generally neither free space nor line of sight, with $p = 4$ being the worst-case model. Therefore, we analyze the protocols with $p = 4$. The distances from the source to destination, source to relay and relay to destination are denoted by d_{sd}, d_{sr} and d_{rd}, respectively. In this way, the variance of the source to destination channel is given by:

$$\sigma_{sd}^2 = \left(\frac{c}{d_{sd}^2}\right)^2, \tag{1.41}$$

where c is a constant. Likewise, the variances of the source to relay and relay to destination channels are given by $\sigma_{sr}^2 = \left(\frac{c}{d_{sr}^2}\right)^2$ and $\sigma_{rd}^2 = \left(\frac{c}{d_{rd}^2}\right)^2$, respectively. The distances between the source, relay and destination are depicted in one dimension of space, as in Fig. 1.13.

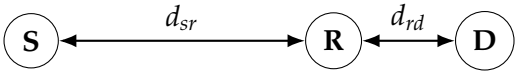


Figure 1.13 – Relay position relatively to the source and destination.

Let u be a unit of distance. In order to mimic the distances in mobile wireless communication systems, the constant c is fixed as $c = 400u^2$ where $1u$ is equivalent to $1m$ (meter). For analysis purposes, we are interested in studying the protocols for the three positions of the relay, given as follows.

- **Relay in the middle**: $d_{sd} = 15u$, $d_{sr} = d_{rd} = 7.5u$;
- **Relay closer to the source**: $d_{sd} = 15u$, $d_{sr} = 5u$ and $d_{rd} = 10u$;
- **Relay closer to the destination**: $d_{sd} = 15u$, $d_{sr} = 10u$ and $d_{rd} = 5u$.

1.6.2 Validation of the Markov chain model of the investigated protocol

We provide numerical results to validate the Markov chain model, depicted in Section 1.5, of the investigated protocol. For this purpose, the throughput and MER of the protocol are numerically evaluated via three different methods, as follows. In all cases, we consider the investigated protocol with IR-HARQ using capacity-achieving codes where $C = 2$ and $R = 0.8$. The relay is located halfway between the source and the destination as explained in the previous Section 1.6.1 (relay in the middle).

Protocol simulation designates the simulation of the investigated protocol as described in Section 1.4.4. Since we use capacity-achieving codes, successful (or failed) decoding is determined by comparing the accumulated mutual information at the receiver to the coding rate. We use counters (of the number of ACKs, time-slots, generated messages and dropped messages) to evaluate the performance metrics (throughput and MER).

Monte Carlo evaluation of T designates the evaluation of each transition probability, given by Eq. (1.20), based on the randomly generated channels. Then, the steady state probabilities are evaluated as in Section 1.5.3, and the performance metrics are obtained using Eq. (1.39) and Eq. (1.40).

Analytical form of T designates the evaluation of the closed-form expression of each transition probability except for $t_{1,8}$, whose expression cannot be written in closed-form. These expressions were derived in Section 1.5.2.

The corresponding plots of the throughput and MER are provided in Fig. 1.14 and Fig. 1.15, respectively. For validation purposes, these performance metrics are evaluated in function of the energy per symbol to noise ratio (in dB), which is denoted by E_s/N_0. Actually the source as well as the relay transmit their symbols with energy E_s.

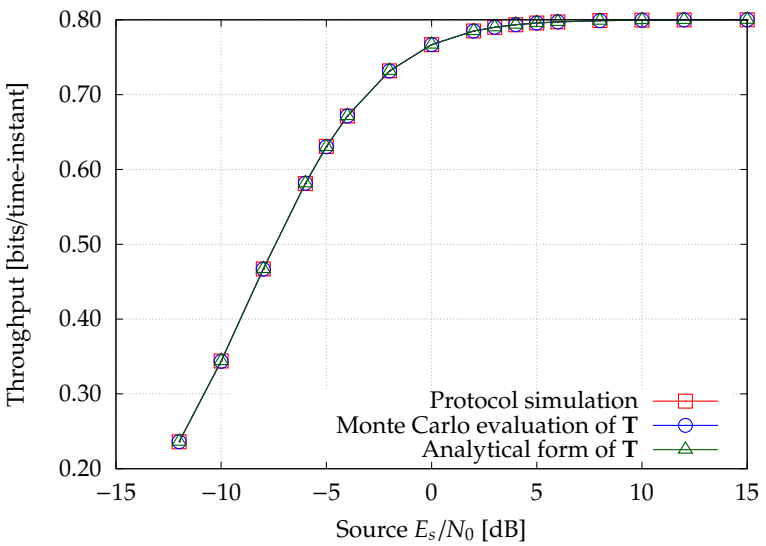


Figure 1.14 – Throughput validation of the investigated protocol.

The plots of the performance metrics using the three methods described above match almost perfectly. This verifies the correctness of the Markov chain model and the calculation of the transition matrix **T**, provided in Section 1.5.2. Furthermore, we compare the transition matrix that is obtained by Monte Carlo evaluation with the one obtained by evaluation of the analytical form, in the case of $E_s/N_0 = -2$dB.

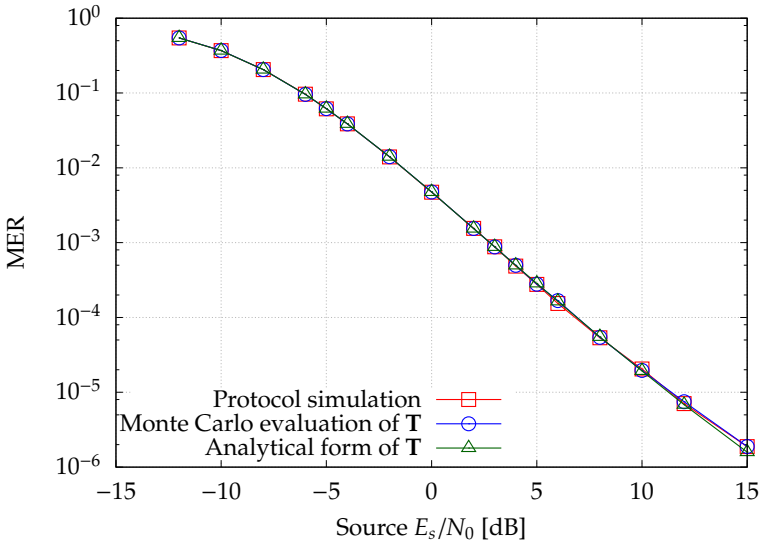


Figure 1.15 – MER validation of the investigated protocol.

The obtained transition matrix by Monte Carlo evaluation of **T** is:

$$\mathbf{T}_{\text{Monte-Carlo}} = \begin{bmatrix} 0.689 & 0.689 & 0.689 & 0.689 & 0.854 & 0.854 & 0 & 0.864 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.682 & 0 \\ 0 & 0 & 0 & 0 & 0.146 & 0.146 & 0 & 0.136 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.007 & 0 \\ 0.007 & 0.007 & 0.007 & 0.007 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.306 & 0 \\ 0.303 & 0.303 & 0.303 & 0.303 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.005 & 0 \end{bmatrix}, \tag{1.42}$$

while the obtained transition matrix by evaluation of the analytical form of **T** is:

$$\mathbf{T}_{\text{Analytical-form}} = \begin{bmatrix} 0.690 & 0.690 & 0.690 & 0.690 & 0.854 & 0.854 & 0 & 0.871 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.682 & 0 \\ 0 & 0 & 0 & 0 & 0.146 & 0.146 & 0 & 0.129 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.006 & 0 \\ 0.007 & 0.007 & 0.007 & 0.007 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.306 & 0 \\ 0.303 & 0.303 & 0.303 & 0.303 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.005 & 0 \end{bmatrix}. \tag{1.43}$$

Concerning the evaluation of $t_{1,8}$, we use the closed-form expression of $\mathfrak{p}_6$ and Monte Carlo simulation of $\mathfrak{p}_7$ and $\mathfrak{p}_8$. This induces a precision error in the evaluation of the analytical form of $t_{1,8}$. However, this precision error is acceptable and has only a marginal impact on the obtained steady state vector and the performance metrics. The obtained

steady state vector by Monte Carlo simulation of $\mathbf{T}$, *i.e.* , using $\mathbf{T}_{\text{Monte-Carlo}}$, is:

$$\boldsymbol{\pi}_{\text{Monte-Carlo}} = \begin{bmatrix} 0.5525 \\ 0.1472 \\ 0.0105 \\ 0.0014 \\ 0.0051 \\ 0.0662 \\ 0.2159 \\ 0.0011 \end{bmatrix}, \tag{1.44}$$

while the obtained steady state vector by evaluation of the analytical form of $\mathbf{T}$, *i.e.* , using $\mathbf{T}_{\text{Analytical-form}}$, is:

$$\boldsymbol{\pi}_{\text{Analytical-form}} = \begin{bmatrix} 0.5526 \\ 0.1473 \\ 0.0105 \\ 0.0014 \\ 0.0051 \\ 0.0661 \\ 0.2159 \\ 0.0011 \end{bmatrix}. \tag{1.45}$$

In this example $\eta_{\text{Monte-Carlo}} = 0.731656$ while $\eta_{\text{Analytical-form}} = 0.731742$, according to Eq. (1.39) . Using the numerical results above, we verified the correctness of the Markov chain model of the investigated protocol.

1.6.3 Comparison of the protocols

To ensure a fair comparison, we evaluate the performance metrics of the protocols in function of the average energy consumed for sending one information bit, denoted by E_b. In other words, E_b designates the energy consumed to transmit one information bit including the energy consumption due to HARQ retransmissions and taking into account the energy consumption of the relay. We remind that E_s/N_0 designates the allocated energy per symbol to noise ratio. Unless otherwise stated (such as in Section 1.6.5 which is dedicated to power allocation), we assume equal energy per symbol for the transmissions of the source and the relay. In order to obtain E_b/N_0, we denote by κ the average time (in time-slots) required to transmit a message, *i.e.* , κ is the average number of time-slots dedicated to transmit one message regardless whether this message is successfully decoded or not at the destination. In addition, we denote by ρ the ratio of time-slots during which the relay is active, *i.e.* , the relay is transmitting a packet to the destination.

In this way, E_b/N_0 (in dB) is given by:

$$E_b/N_0 = \begin{cases} \kappa \frac{E_s/N_0}{R} & \text{using no relay retransmission,} \\ \kappa \frac{E_s/N_0}{R} & \text{using orthogonal relay retransmission,} \\ (1+\rho)\kappa \frac{E_s/N_0}{R} & \text{using other protocols.} \end{cases} \tag{1.46}$$

Hereby, we provide a comparison of the performance metrics of i) IR-HARQ (without a relay), ii) Orthogonal relay retransmission [1], iii) Non-orthogonal relay retransmission, iv) Alamouti retransmission, and v) the investigated protocol.

In Fig. 1.16, Fig. 1.17 and Fig. 1.18, we compare the throughput, MER and delay of these protocols versus E_b/N_0 for capacity-achieving codes with $C = 2$ and $R = 0.8$. As for the path loss model, we place the relay in the middle.

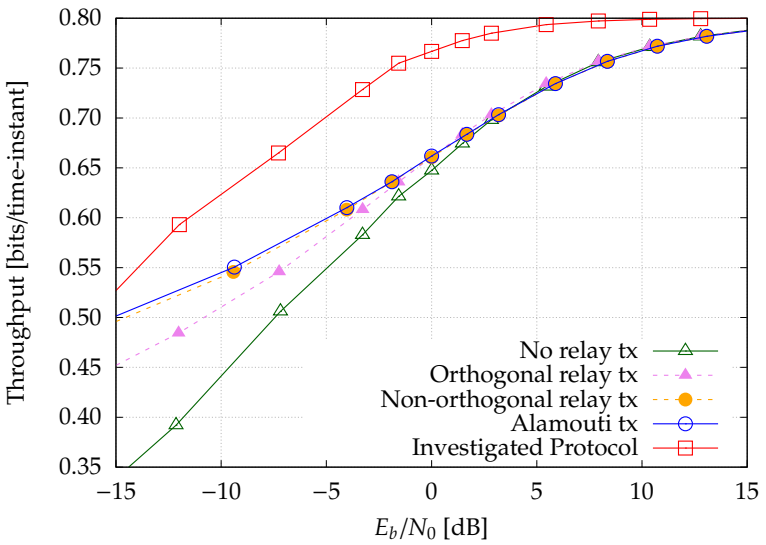


Figure 1.16 – Throughput comparison of the protocols using capacity-achieving codes, with $C = 2$ and a relay in the middle.

The investigated protocol achieves a significant gain in throughput, of more than 5dB, particularly for medium an high SNR, in comparison to other protocols. On the other hand, Alamouti retransmission is less prone to message errors since it provides a higher diversity (of 3 when $C = 2$). This gain in diversity is due to the use of Alamouti code in the second transmission, which provides a diversity of 2 since the same message is sent through two independent channels, in addition to the first transmission. However, the diversity gain in Alamouti retransmission, which provides a significant gain in MER, has no significant impact on the throughput. Hence, the investigated protocol is of high interest as it provides a good trade-off, especially for throughput demanding applications, by

[1] The word "retransmission" is abbreviated as "tx" in the legend of the plots due to space limitations.

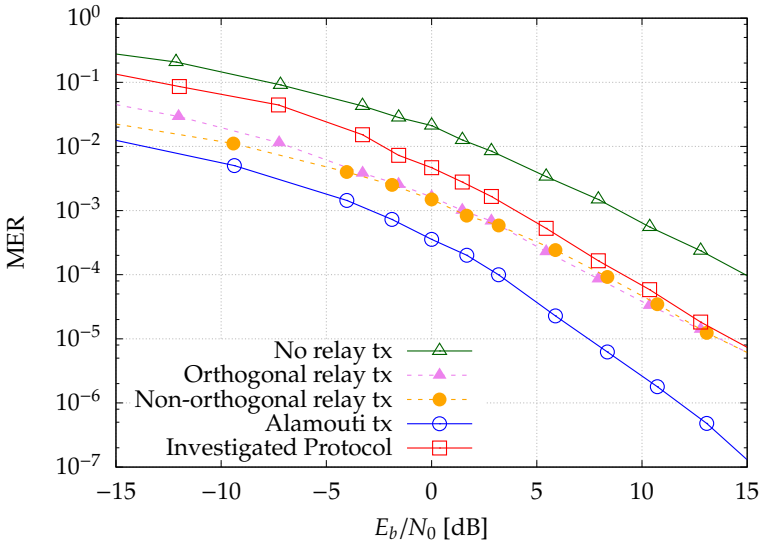


Figure 1.17 – MER comparison of the protocols using capacity-achieving codes, with $C = 2$ and a relay in the middle.

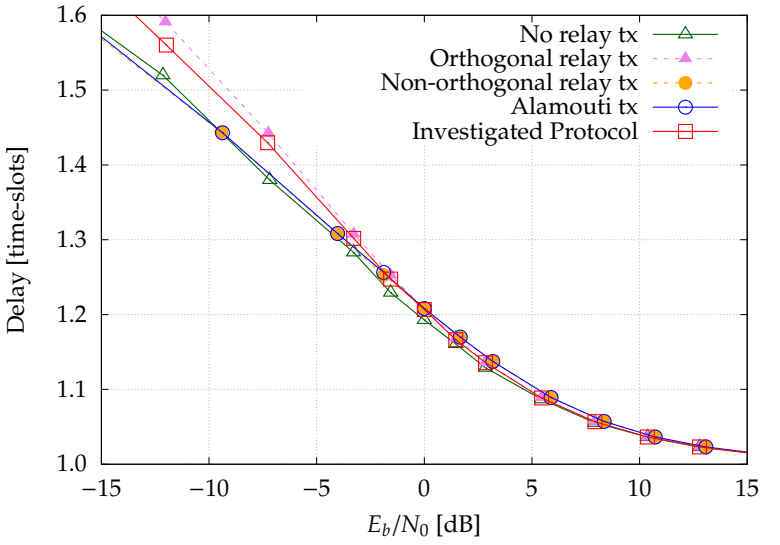
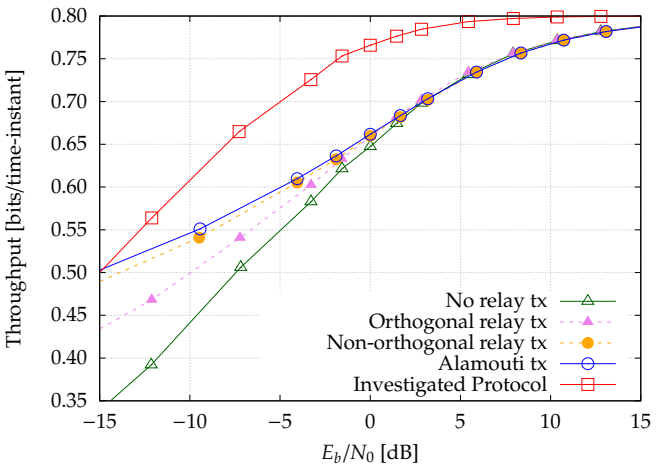


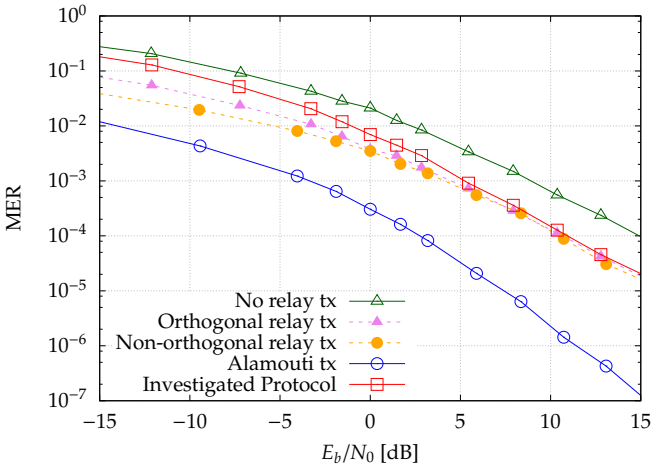
Figure 1.18 – Delay comparison of the protocols using capacity-achieving codes, with $C = 2$ and a relay in the middle.

achieving a significantly higher throughput than any other protocol while *sufficiently* protecting the transmitted messages against channel fluctuations. In addition, no significant difference between the compared protocols is observed in terms of delay.

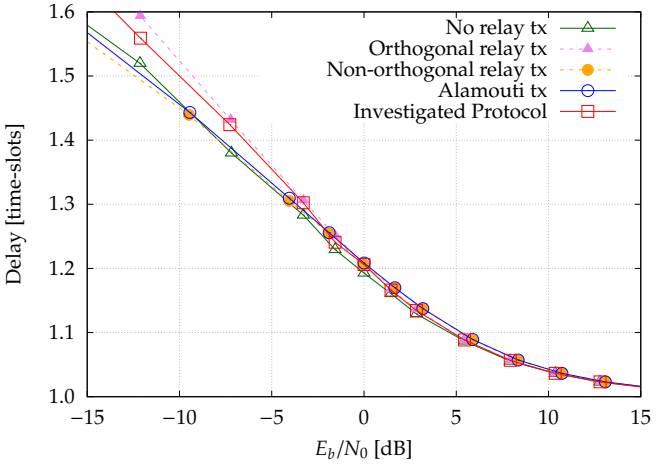
The protocols are compared for other relay positions, as in Fig. 1.19 where we place the relay closer to the source, and in Fig. 1.20 where we place the relay closer to the destination. We notice, in Fig. 1.20b, that the Alamouti retransmission protocol is profitless when the relay is close to the destination, since the channel between the relay and destination becomes less prone to errors due to proximity, according to the path loss model. Based on these numerical results, we deduce that the investigated protocol offers significantly higher throughput than existing ones for any SNR and any relay position. This gain is explained by the fact that the source transmits a new message while the relay retransmits the previous one. The investigated protocol has higher MER than the Alamouti retransmission protocol due to the interference at the destination between messages coming from the source and the relay. Nevertheless, this loss in MER does not damage the throughput when using an ideal decoder.



(a) Throughput using capacity-achieving codes.

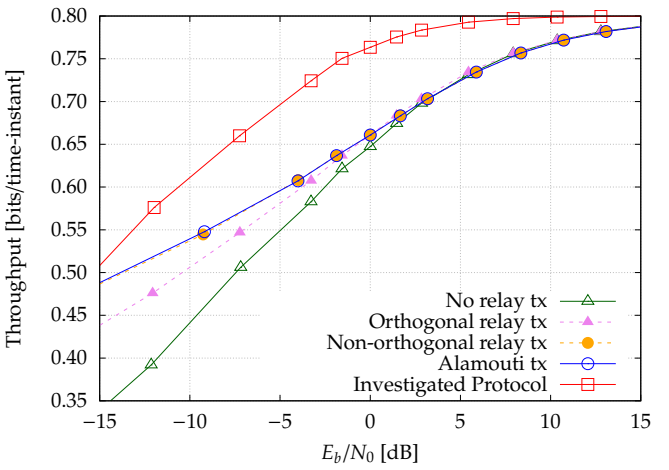


(b) MER using capacity-achieving codes.

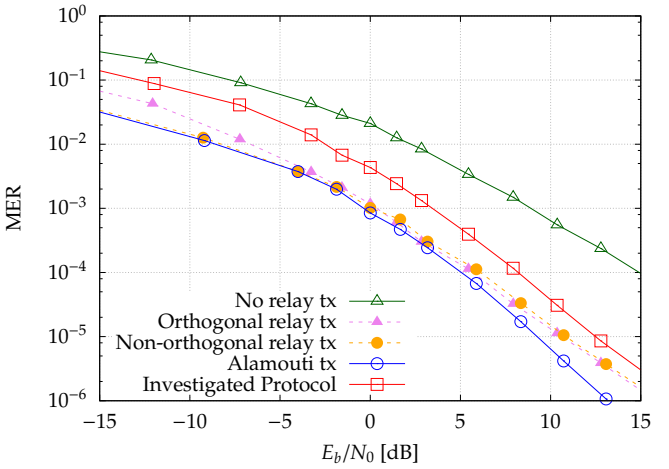


(c) Delay using capacity-achieving codes.

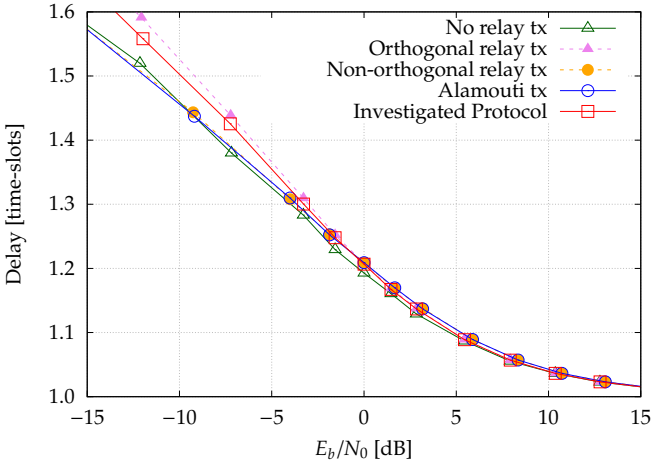
Figure 1.19 – Performance metrics comparison of the protocols, with $C = 2$ and a relay closer to the source.



(a) Throughput using capacity-achieving codes.



(b) MER using capacity-achieving codes.

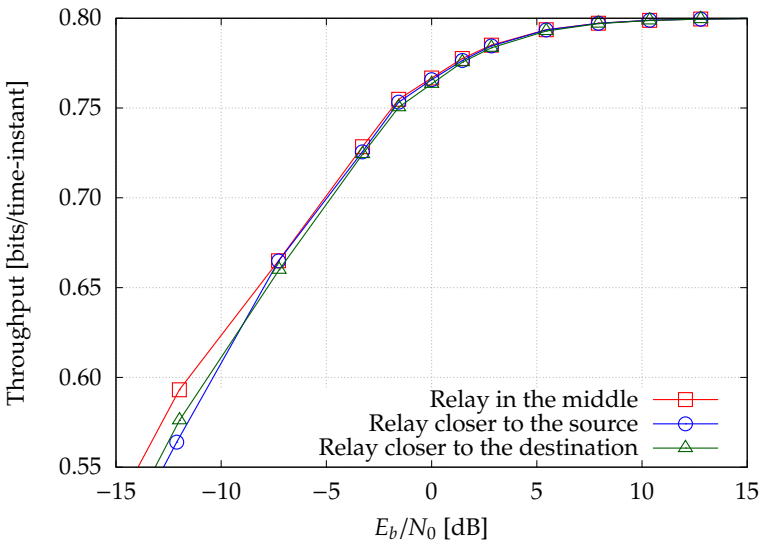


(c) Delay using capacity-achieving codes.

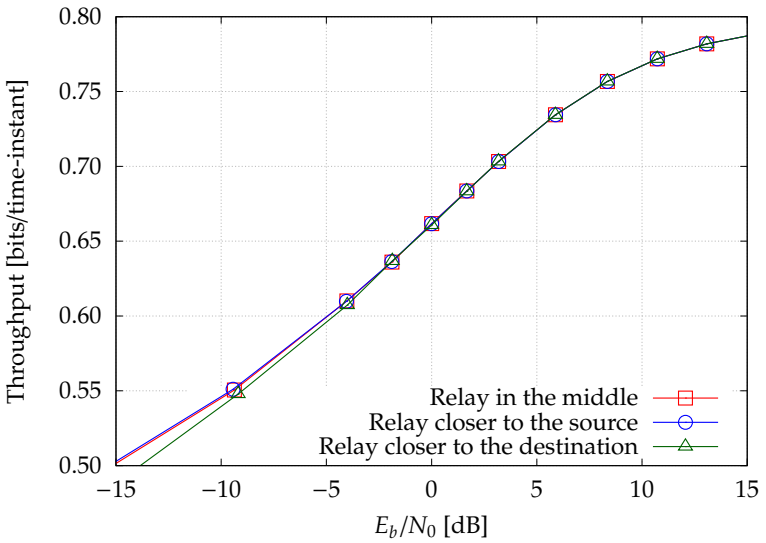
Figure 1.20 – Performance metrics comparison of the protocols, with $C = 2$ and a relay closer to the destination.

1.6.4 Impact of the relay position on the performance of the investigated protocol

We compare the performance of the investigated protocol with the Alamouti retransmission protocol for various relay positions, with $C = 2$. The throughput is compared in Fig. 1.21, the delay is compared in Fig. 1.22, and the MER is compared in Fig. 1.23.



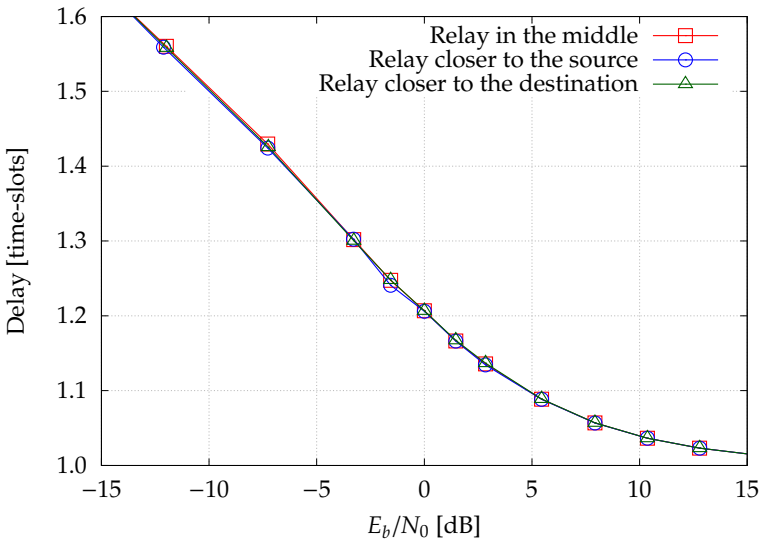
(a) Throughput using the investigated protocol.



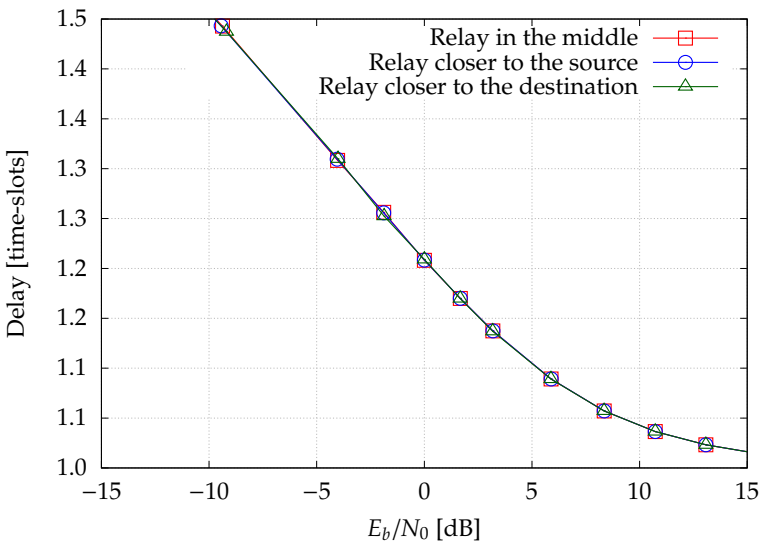
(b) Throughput using the Alamouti retransmission.

Figure 1.21 – Throughput comparison of the protocols for various relay positions.

Fig. 1.21 shows that the relay position has no significant impact neither on the throughput of the Alamouti retransmission nor on the throughput of the investigated protocol.



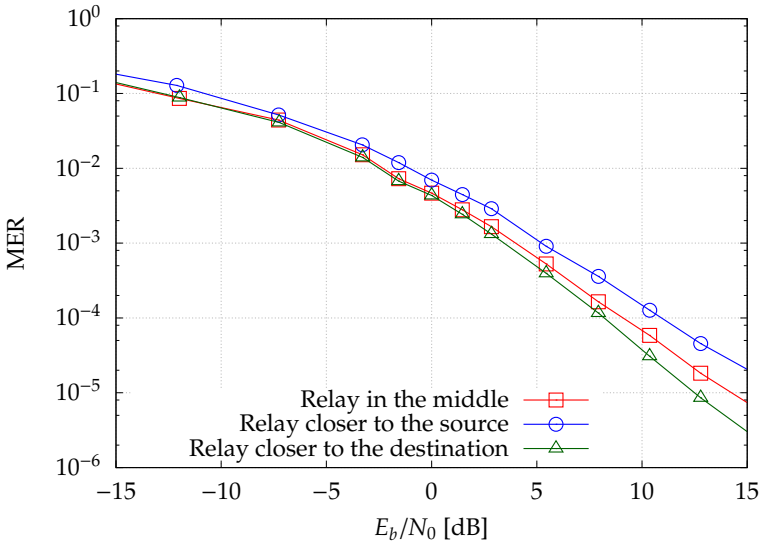
(a) Delay using the investigated protocol.



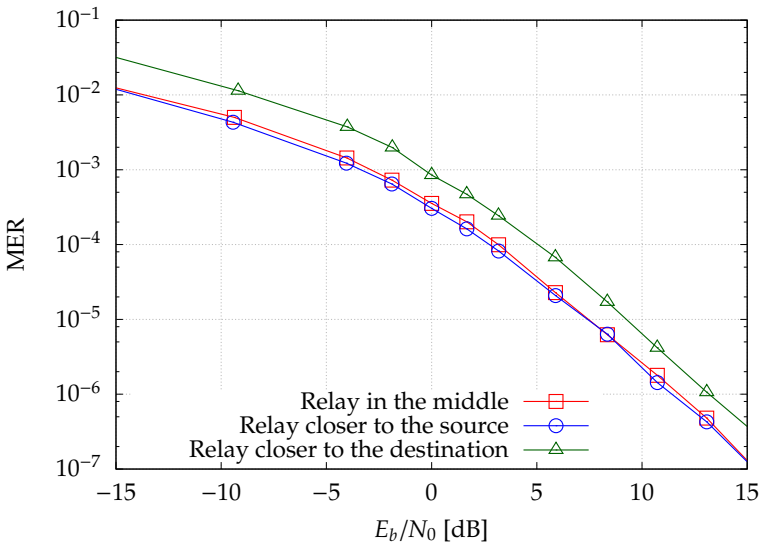
(b) Delay using the Alamouti retransmission.

Figure 1.22 – Delay comparison of the protocols for various relay positions.

No significant impact of the relay position on the delay can be observed in Fig. 1.22 when using either the Alamouti retransmission protocol or the investigated protocol.



(a) MER using the investigated protocol.



(b) MER using the Alamouti retransmission.

Figure 1.23 – MER comparison of the protocols for various relay positions.

However, a slightly lower MER is obtained when the relay is closer to the source in the case of Alamouti retransmission, while the MER is lower when the relay is closer to the destination in the case of the investigated protocol, as can be seen in Fig. 1.23. In contrast to the Alamouti retransmission protocol, the investigated protocol superposes the transmitted packets, thus creates interference. In other words, the transmitted signal

using superposition of packets is more vulnerable to errors than the one obtained using the Alamouti code. Therefore, it is slightly more beneficial to have the relay closer to the source when using the Alamouti retransmission protocol, while it is more beneficial to have the relay closer to the destination when using the investigated protocol.

1.6.5 Power allocation for the investigated protocols

We propose to optimize the power allocation for the source and relay nodes when using the investigated protocol. Differently from previous Sections, we set E_1 to be the energy per symbol transmitted by the source while E_2 is the energy per symbol transmitted by the relay, in this Section. Notice that the case of $E_2 = E_1 = E_s$ corresponds to equal power allocation, as in the previous Sections. In order to optimize power allocation, we run several instances of the simulation for each quantized pair of values (E_1,E_2). Each simulation provides the value of ρ, which designates the ratio of time-slots during which the relay is active, along with the performance metrics. Following each instance of the simulation, the average energy per symbol consumed by the transmitters at each channel use (time-instant), denoted by E_t, is computed as $E_t = E_1 + \rho E_2$. Finally, the optimal power allocation is obtained by finding the pair of values (E_1,E_2) that achieves the highest throughput for each value of E_t/N_0. Hereby, we compare the obtained performance metrics of the investigated protocol using this optimal power allocation to the case of equal power allocation. The throughput, MER and delay are shown in Figs. 1.24, 1.25 and 1.26, respectively.

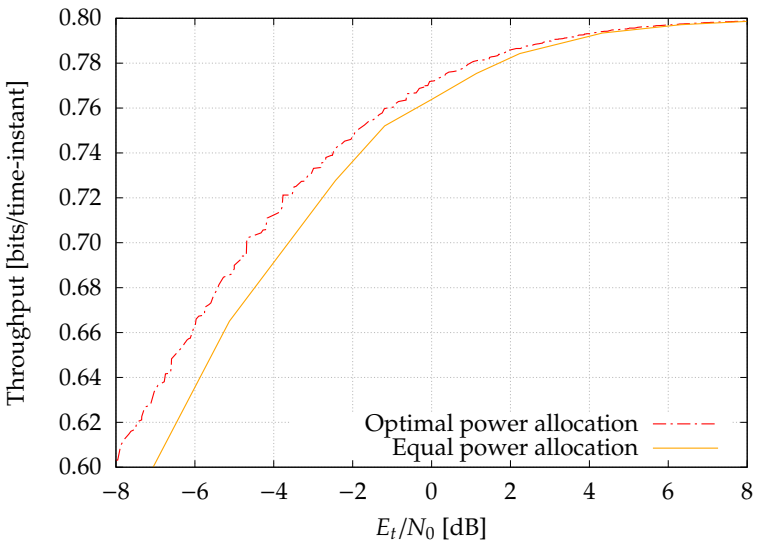


Figure 1.24 – Throughput of the investigated protocol using power allocation, with $C = 2$ and a relay in the middle.

According to Fig. 1.24, numerical power allocation improves the throughput of the investigated protocol.

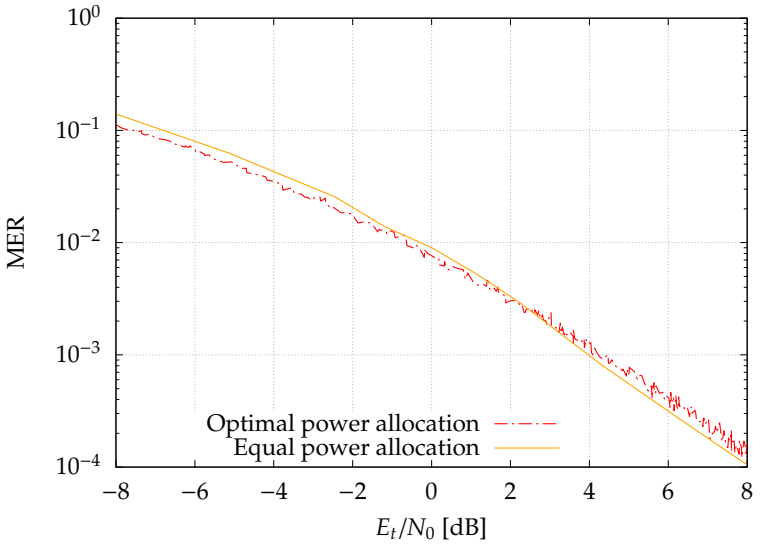


Figure 1.25 – MER of the investigated protocol using power allocation, with $C = 2$ and a relay in the middle.

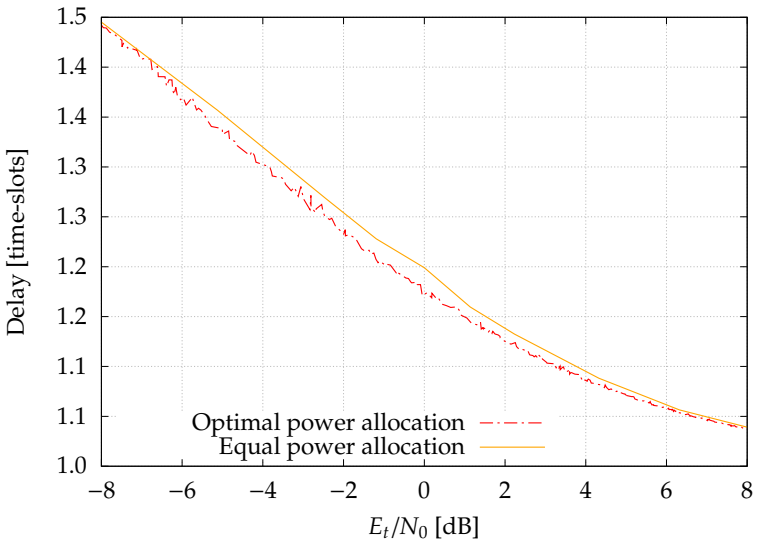
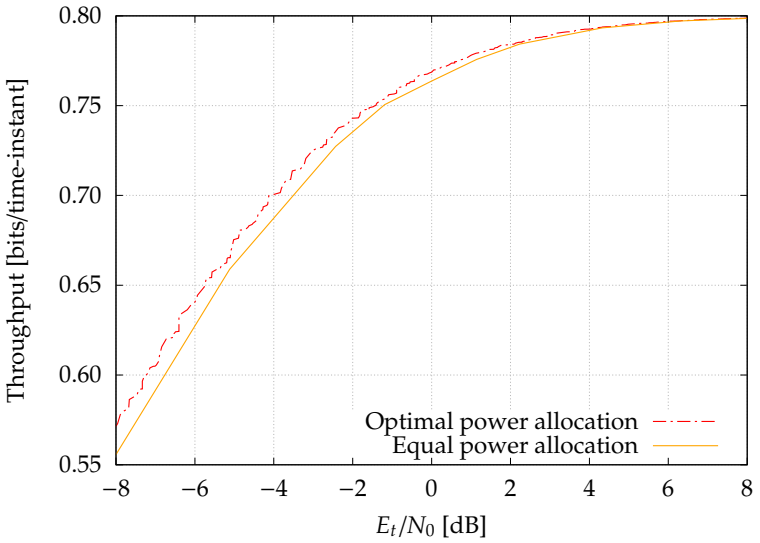


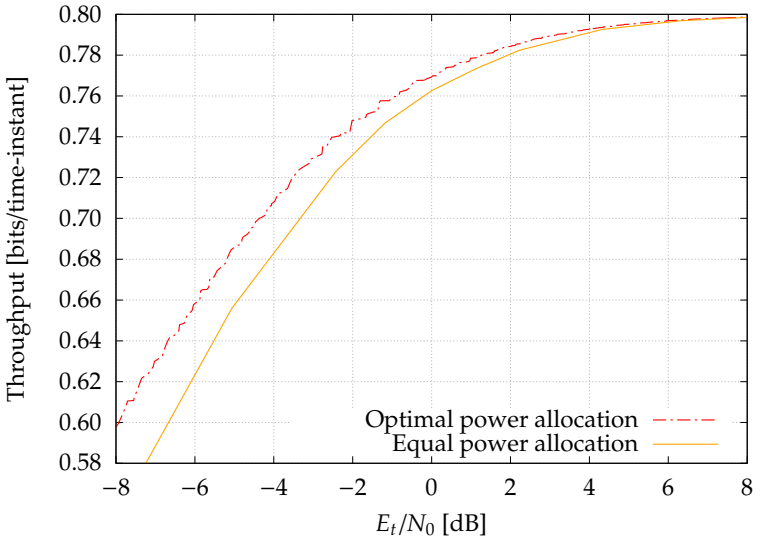
Figure 1.26 – Delay of the investigated protocol using power allocation, with $C = 2$ and a relay in the middle.

Notice in Fig. 1.25 that achieving the highest throughput does not imply necessarily obtaining the lowest MER. Instead, the MER is almost the same in both cases of power allocation. In this optimization, the gain in throughput is due to smaller delay in comparison to equal power allocation according to Fig. 1.26 at (almost) the same MER. Alternatively, one could choose the values of α that achieve the lowest MER or the smallest delay depending on the application requirements.

Moreover, the impact of power allocation on the throughput of the investigated protocol for other relay positions is depicted in Fig. 1.27. The plots indicate that power allocation improves the throughput of the investigated protocol. More specifically, this improvement is significant when the relay is in the middle or closer to the destination. Hence, it is beneficial to numerically optimize the power allocation at the transmitters, for each SNR and depending on the relay position, when using the investigated protocol. Power allocation could be numerically optimized to improve other performance metrics such as the MER or delay, or a combination of these performance metrics, depending on the application and requirements of the system.



(a) Throughput with a relay closer to the source.



(b) Throughput with a relay closer to the destination.

Figure 1.27 – Throughput of the investigated protocol with various relay positions using power allocation, with $C = 2$.

1.6.6 Comparison of the protocols using practical codes

We discuss the implementation of the HARQ protocols using practical codes. In this work, we use Rate-Compatible Punctured Convolutional (RCPC) codes as channel codes, as explained in Section 1.6.6.1. The performance of (non-orthogonal) relay assisted HARQ protocols, particularly when using the investigated protocol, depends on the interference handling at the decoder. Therefore, we illustrate the decoder that we use in conjunction with the investigated protocol in Section 1.6.6.2. Then, we compare simulation results of the discussed HARQ protocols, using practical codes, in Section 1.6.6.3.

1.6.6.1 Implementation of IR-HARQ using RCPC codes

In this work, IR-HARQ is implemented using RCPC codes of coding rate 0.8, as FEC codes. More precisely, the message bits are encoded using RCPC codes to obtain a mother-codeword of rate 1/4. Then, we use puncturing tables of memory 4 and period 8, as defined in [Hagenauer, 1988], to create three codeword chunks of equal size and successive rates of 8/10, 8/20, and 8/30. When using IR-HARQ with $C = 2$, the coding rate of the codeword chunk in the first round is 8/10. If a second round is needed, additional redundancy bits are transmitted in the second round such that the resulting coding rate is 8/20. When using IR-HARQ with $C = 3$, the successive coding rates are 8/10, 8/20, and 8/30, as more transmissions are needed. The codeword chunks are modulated using BPSK prior to transmission and the resulting packets are sent over the channel.

1.6.6.2 Implementation of the decoder for the investigated protocol

When using the investigated protocol, the interference due to superposed packets, coming from the source and relay, should be handled properly at the destination. The destination node receives the superposed packets, then yields Log Likelihood Ratio (LLR)s for each packet prior to channel decoding. These LLRs are calculated, using the interference demodulator, and assuming interference as a noise. More precisely, the interference demodulator generates separate LLR streams for the transmitted packets by the source and relay. On a side note, the interference demodulator is generalized here to Quadrature Amplitude Modulation (QAM), however we implement BPSK modulation, or equivalently 2-QAM, in this work. Assume that the system is in one of the states (of even index) $\mathbf{S}_2$, $\mathbf{S}_4$, $\mathbf{S}_6$ or $\mathbf{S}_8$, at time-slot t. At time-instant n ($n \in \{1, \ldots, N\}$) of time-slot t, the source and relay transmit simultaneously the QAM symbols (2-QAM in our work) $p_{k+1,n}(1)$ and $p_{k,n}(2)$, respectively. The received sample at the destination, $y_t[n]$, is given by:

$$y_t[n] = h_{sd}(t)p_{k+1,n}(1) + h_{rd}(t)p_{k,n}(2) + w_t[n]. \tag{1.47}$$

Notice that $p_{k+1,n}(1)$ and $p_{k,n}(2)$ are independent symbols since the incoming packets, $\mathbf{p}_{k+1}(1)$ and $\mathbf{p}_k(2)$, from the source and relay respectively correspond to different messages

($\mathbf{m}_{k+1}$ and $\mathbf{m}_k$). The LLR of $b_{k+1,n}(1)$, which is the transmitted bit by the source via the symbol $p_{k+1,n}(1)$, is given by:

$$LLR(b_{k+1,n}(1)) = \log \frac{P(y_t[n]|b_{k+1,n}(1)=0)}{P(y_t[n]|b_{k+1,n}(1)=1)} \tag{1.48a}$$

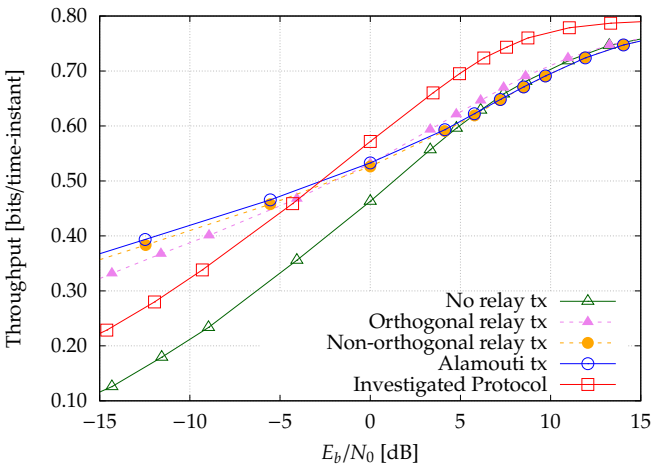
$$= \log \frac{\sum_{\substack{\alpha\in Q_0 \\ \beta\in Q}} \exp\left(-\frac{|y_t[n]-h_{sd}(t)\alpha-h_{rd}(t)\beta|^2}{N_0}\right)}{\sum_{\substack{\alpha\in Q_1 \\ \beta\in Q}} \exp\left(-\frac{|y_t[n]-h_{sd}(t)\alpha-h_{rd}(t)\beta|^2}{N_0}\right)}, \tag{1.48b}$$

where Q is the set of all symbols in the considered QAM modulation. Q_0 and Q_1 are the set of symbols corresponding to bits 0 and 1, respectively. The value of $LLR(b_{k,n}(2))$, *i.e.*, the bit sent by the relay, can be found using the same method. Once the LLRs are evaluated, they are passed to the channel decoder (Viterbi decoder in our case). Notice that the decoder could be improved by applying SIC in an iterative manner, *i.e.*, the demodulator regenerates the LLR of $p_{k+1,n}(1)$ in case the destination successfully decodes the retransmitted message by the relay $\mathbf{m}_k$, and vice versa. Albeit suboptimal, this decoder is simple to implement. Hence, we suggest to use it in conjunction with the investigated protocol to decode superposed packet.

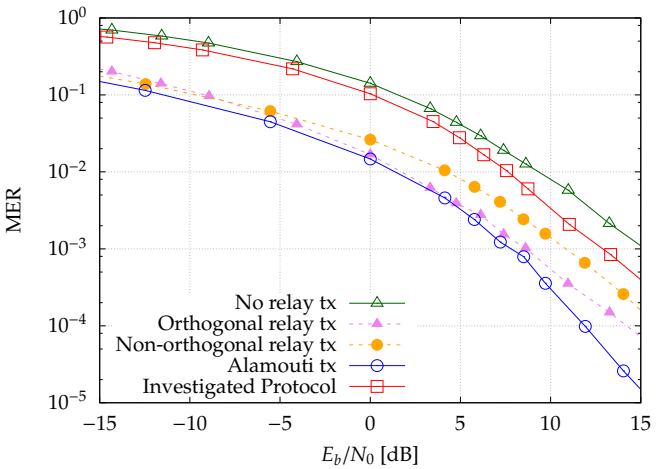
1.6.6.3 Simulation results

The performance of the investigated protocol, using practical codes and the suggested decoder (without applying SIC), is compared to other protocols. The results are shown in Fig. 1.28. We notice that the gain is much higher than those seen in [Zhang et al., 2010; Ma et al., 2013] due to the considered decoder. Hence, the decoder above is more efficient than the MMSE one for such a protocol. As a consequence, the protocol is powerful and of great interest. Once again, the investigated protocol outperforms the existing ones in terms of throughput.

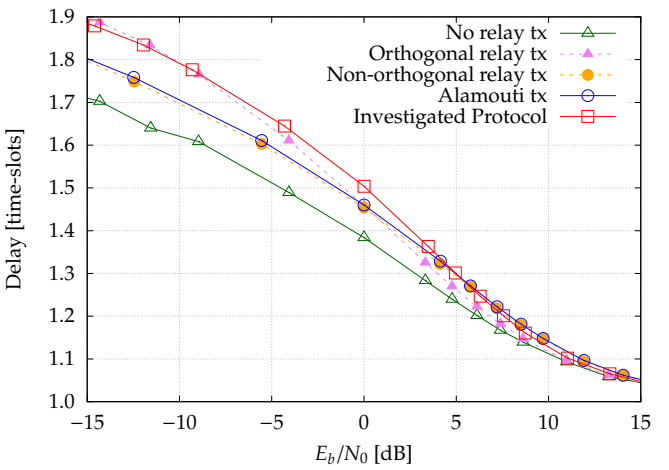
A similar conclusion is obtained when comparing the protocols with $C = 3$, as can be seen in in Fig. 1.29. The obtained simulation results show that the investigated protocol is significantly more throughput-efficient than other relay assisted HARQ protocols, when using practical codes.



(a) Throughput using practical codes.

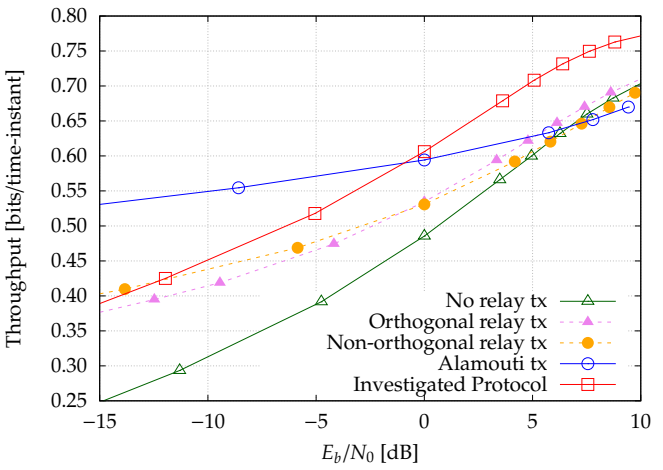


(b) MER using practical codes.

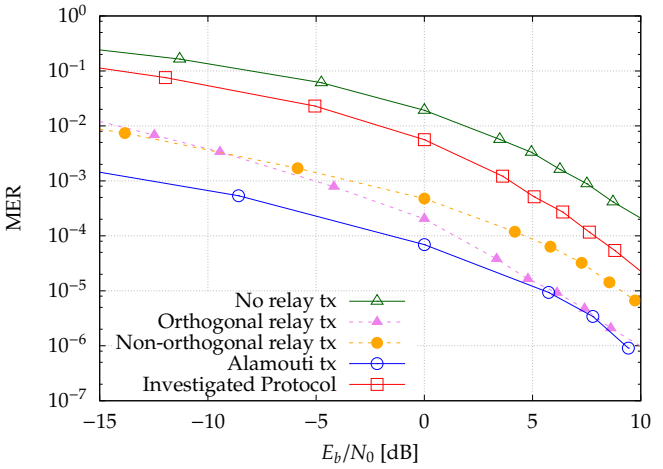


(c) Delay using practical codes.

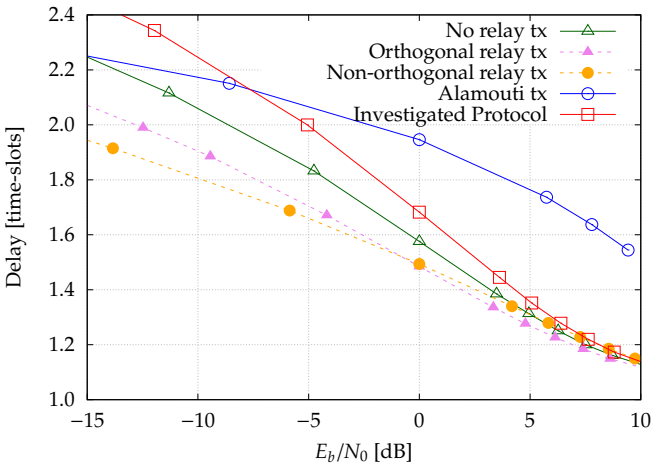
Figure 1.28 – Performance metrics comparison of the protocols using practical codes, with $C = 2$ and a relay in the middle.



(a) Throughput using practical codes.



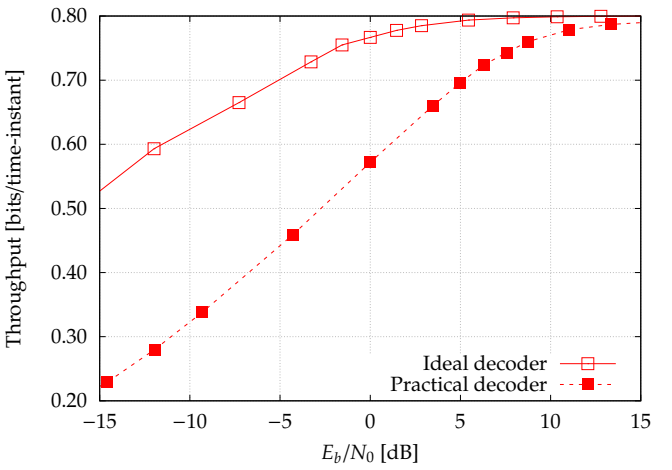
(b) MER using practical codes.



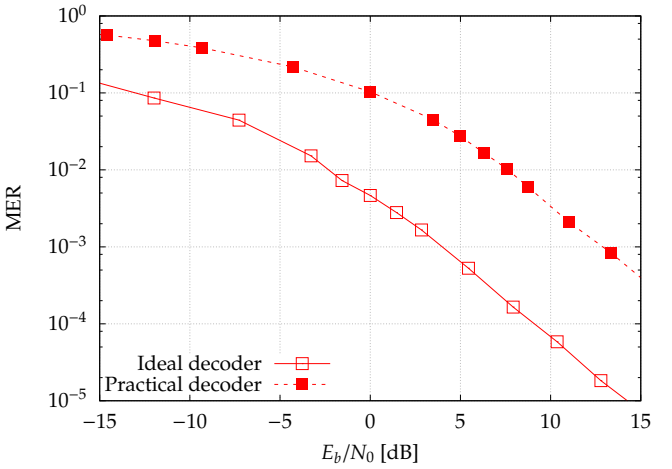
(c) Delay using practical codes.

Figure 1.29 – Performance metrics comparison of the protocols using practical codes, with $C = 3$ and a relay in the middle.

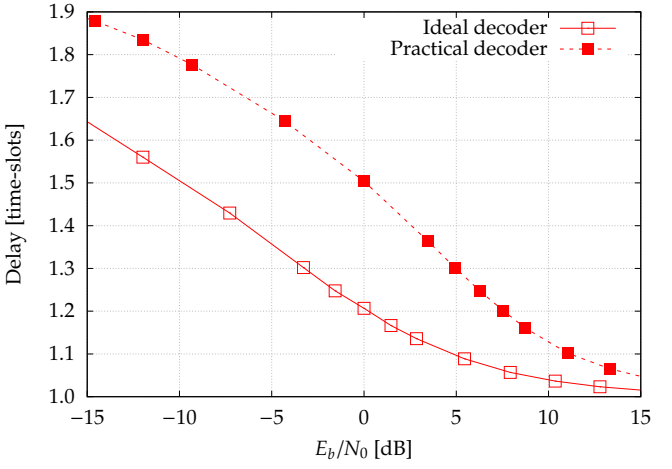
Albeit simple to implement and practical, the suggested decoder is sub-optimal, as can be seen in Fig. 1.30, where we compare the performance metrics of this practical decoder to the ideal decoder. This comparison is carried out by using practical codes with the practical decoder and capacity-achieving codes with the ideal decoder, with $C = 2$ and a relay in the middle. Hence the performance of the investigated protocol can be improved furthermore by conceiving a better practical decoder.



(a) Throughput of the investigated protocol.



(b) MER of the investigated protocol.



(c) Delay of the investigated protocol.

Figure 1.30 – Performance metrics comparison of the practical and ideal decoders.

1.7 Conclusion

In this Chapter, we investigated retransmission techniques combined with superposition of packets. We started from the basic concepts of retransmission techniques, namely ARQ. Then, we presented various HARQ mechanisms that improve the reliability of the communication. In addition, we showed that a relay node could improve the system's performance furthermore by retransmitting unsuccessfully decoded packets. Moreover, we analyzed various protocols that use a relay in conjunction with HARQ. The core of this Chapter was investigating a specific relay assisted HARQ protocol that uses non-orthogonal transmission combined with HARQ. This investigation included

- theoretical analysis, in which we derived analytical expressions of the performance metrics of the investigated protocol based on a Markov chain model;
- numerical evaluation, in which we evaluated the performance metrics when using capacity achieving codes with optimal information theoretic decoders;
- practical simulation, in which we simulated the investigated protocol with channel codes and real decoders.

Comparing this protocol to other relay assisted HARQ protocols showed that this protocol is of great interest for throughput demanding applications. More precisely, this protocol is more throughput-efficient than other relay assisted HARQ protocols. The trade-off is a small loss in MER that does not impact the throughput. As a result, the investigated protocol achieves better performance than other relay assisted HARQ protocols for most applications that require reliable retransmissions at high rates. Hence, the combination of non-orthogonal transmission with HARQ is of great interest, as it provides simultaneously a throughput-efficient and reliable communication. Also, part of this work has been published in [C1]. In the next Chapter, we show furthermore that non-orthogonal transmission with HARQ improves the performance of wireless communication systems.

Chapter 2

Multi-Packet Hybrid ARQ with Delayed Feedback

2.1 Introduction

In current wireless communication systems, the feedback required by the HARQ mechanism is received with a delay of multiple time-slots at the transmitter side. The communication system is idle in-between if Stop-And-Wait (SAW) HARQ protocol is used [Lin and Costello, 2004]. To alleviate this issue and enable continuous transmission, up to eight parallel SAW HARQ processes are used in Long Term Evolution (LTE) [Sesia et al., 2009]. However, using parallel SAW HARQ does not mitigate the delay of the received messages. In this Chapter, we propose a multi-packet HARQ protocol (also called superposition coding or multi-layer HARQ) to improve the user's delay distribution and increase the throughput, without any additional feedback such as CSI.

In IR-HARQ, the transmitter's decision to send a packet depends on the receiver's feedback, as explained in Chapter 1. Upon the transmission of a packet, the transmitter stops and waits for the reception of the receiver's feedback in order to decide the next packet to send. This mechanism, called SAW HARQ, is used along with IR-HARQ, in which the throughput and delay metrics are dependent on the feedback delay. Due to propagation time, processing time and reverse link scheduling, the feedback associated with a sent packet is available to the transmitter with a delay of multiple time-slots. For instance, the Round-Trip Time (RTT) in LTE is 8 time-slots, *i.e.* , 8ms [Sesia et al., 2009]. In other words, the feedback delay degrades the performance of wireless communication systems using HARQ.

A conventional approach to compensate for this issue is to initiate, during the unused time-slots, several SAW HARQ processes corresponding to other messages. This scheme, called parallel SAW HARQ [Lin and Costello, 2004; Sesia et al., 2009], allows the transmission of multiple messages in parallel, each of them employing an independent SAW HARQ process. Parallel SAW HARQ improves the throughput but does not mitigate the

delay of the received messages, which is related to the maximum retransmission credit and to the feedback delay. The proposed protocol in this Chapter enables the transmitter to anticipate the feedback by sending, in advance to its reception, data related to unacknowledged messages.

More precisely, the main contribution of this Chapter is to propose a protocol that superposes an additional layer of redundant packets to the parallel SAW HARQ protocol. In the superposed layer, the transmitter may perform retransmissions of a message even before having received any feedback about it. The selection of the superposed packets is based solely on the delayed ACK/NACK feedback (no additional CSI). We provide performance analysis of the proposed protocol, from an information-theoretic point-of-view, and compare its delay distribution, throughput and MER to parallel SAW HARQ.

This Chapter is organized as follows: Section 2.2 presents a brief overview of ARQ mechanisms with delayed feedback and describes recent works on multi-packet HARQ. Then, we explain in Section 2.3 the proposed HARQ protocol, and its corresponding receiver is analyzed from an information theoretic point of view. In Section 2.4, numerical evaluation of the proposed protocol is provided and compared to parallel SAW HARQ . Concluding remarks are drawn in Section 2.5.

2.2 From Stop-and-Wait HARQ to Multi-packet HARQ

HARQ uses packet retransmissions and channel coding to achieve reliable communication. If the feedback of the receiver arrives with a feedback delay T (also referred to as RTT), various protocols can be adopted by the transmitter. The most common protocols in wireless communication systems are SAW [Lin and Costello, 2004] and parallel SAW [Sesia et al., 2009], which is also known as Selective Repeat (SR) in [Lin and Costello, 2004]. These retransmission protocols can be implemented in conjunction with type-I HARQ or type-II HARQ (with CC or IR), as explained in Chap. 1. Furthermore, the retransmission protocols can be improved by combining multi-packet transmission with HARQ. Hereafter in Section 2.2.1, we provide an overview of the SAW protocol. Then, we explain parallel SAW in Section 2.2.2, and its implementation in LTE in Section 2.2.3. Whereas, Section 2.2.4 is devoted to discuss recent works on multi-packet HARQ.

2.2.1 Stop-And-Wait Hybrid ARQ

The SAW protocol consists of the transmitter being idle during the RTT, while waiting for the feedback, as depicted in Fig. 2.1. We denote by $\{\cdot\}_A$ (respectively $\{\cdot\}_N$) the set of message indexes triggering the ACK (respectively NACK) feedback. The notation $\mathcal{F}_t$ stands for the ACK/NACK feedbacks of the receiver at time-slot t. Moreover, $\mathcal{F}_t$ will be available at the transmitter side with a delay of T time-slots, where $T = 3$ in this example. Also, we assume that $T \geq 1$. Notice that for $T = 1$, the receiver's feedback is available

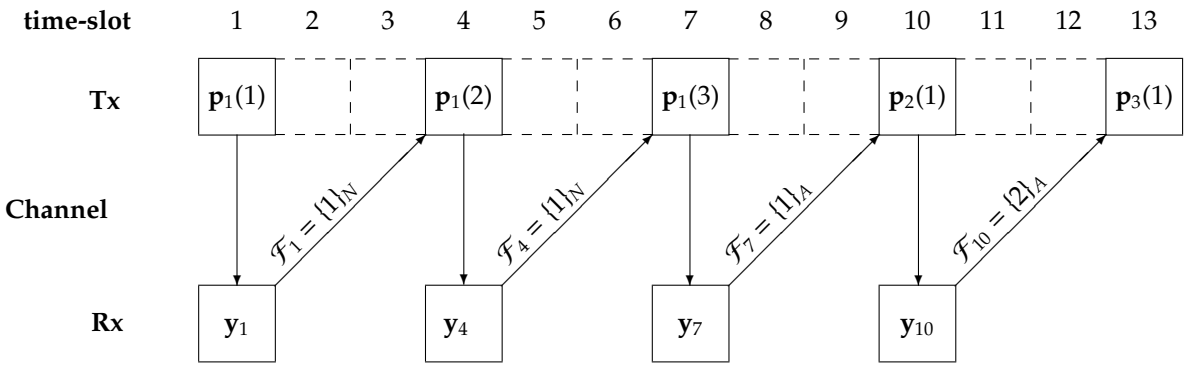


Figure 2.1 – Stop-And-Wait Hybrid ARQ.

to the transmitter at the beginning of the next time-slot, *i.e.*, with no delay. Due to this feedback delay, the transmitter is idle $100\frac{T-1}{T}\%$ of the time. In this instance of the SAW HARQ protocol, the transmitter is idle 66.67% of the time. Also, we remind that C time-slots at most can be allocated to transmit a single message in the case of truncated HARQ, where $C = 3$ in this example. In this way, the maximum number of elapsed time-slots since the first transmission of a message until its successful decoding at the receiver, called maximum delay, is $(C-1)T+1$. In this instance of the SAW HARQ protocol, $\mathbf{m}_1$ is decoded with a delay of 7 time-slots, which is the maximum delay in this example. Moreover, if a message remains undecoded after its first transmission, the receiver has to wait T time-slots to receive a redundant packet related to this message. Then, the receiver has to wait T additional time-slots for each requested retransmission of the message. Hence, SAW HARQ is the simplest retransmission protocol, but it is inefficient in terms of throughput and it induces large delays in delay sensitive applications.

2.2.2 Parallel Stop-And-Wait Hybrid ARQ

Parallel SAW HARQ (also called SR) allows the transmitter to continuously send messages over the channel, even during the RTT, as shown in Fig. 2.2. Upon a NACK feedback, only the erroneous messages are selectively requested to be retransmitted (hence this protocol is also known as SR). In case of delayed feedback, each HARQ operation uses a SAW protocol. Notice that when using the SAW protocol, the system is idle during the RTT while waiting for the feedback, as explained in the previous Section 2.2.1, which is not throughput efficient. However, the parallel SAW protocol enables the transmitter to initiate, during the unused time-slots, parallel SAW HARQ processes corresponding to other messages. In other words, this protocol allows the transmission of multiple messages in parallel, each of them employing an independent SAW HARQ process. Each HARQ process is responsible for a separate SAW operation and manages a separate buffer. Notice that the SAW and parallel SAW protocols become equivalent when the RTT is one ($T = 1$). Also, both HARQ protocols have the same MER and delay, but

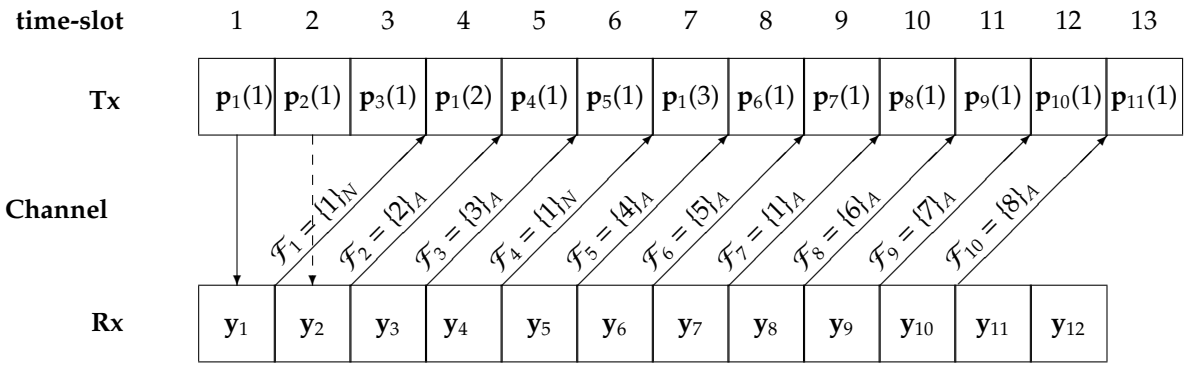


Figure 2.2 – Parallel Stop-And-Wait Hybrid ARQ.

their performance differs in terms of throughput (according to the definition of the MER, delay and throughput given in Chapter 1). Moreover, the parallel SAW protocol provides higher throughput than the SAW protocol.

In the example in Fig. 2.2, the messages $\mathbf{m}_1$, $\mathbf{m}_2$ and $\mathbf{m}_3$ are sent through packets $\mathbf{p}_1(1)$, $\mathbf{p}_2(1)$ and $\mathbf{p}_3(1)$ at time-slots 1, 2 and 3 respectively, in three parallel processes. The NACK feedback corresponding to the transmission of $\mathbf{p}_1(1)$, *i.e.* $\mathcal{F}_1 = \{1\}_N$, is available to the transmitter just before the beginning of time-slot 4, due to the feedback delay of $T = 3$ time-slots. Consequently, $\mathbf{p}_1(2)$ is transmitted at time-slot 4. Upon the occurrence of an ACK feedback, such as the reception of $\mathcal{F}_3 = \{3\}_A$ at the beginning of time-slot 6, a new SAW HARQ process corresponding to the transmission of a new message ($\mathbf{m}_4$ in this example) is initiated. Moreover, a new SAW HARQ process is initiated as well in case of time-out. In this realization of the protocol, $\mathbf{m}_1$ will be in time-out at the beginning of time-slot 10.

The parallel SAW HARQ protocol is currently used in most wireless communication systems such as LTE.

2.2.3 Hybrid ARQ in LTE

In this Section, we provide an overview of HARQ schemes in LTE based on [Sesia et al., 2009].

The HARQ entity in LTE is implemented at the Medium Access Control layer, and is responsible for the transmit and receive HARQ operations. Moreover, HARQ is implemented with IR, and is used for both the downlink and uplink channels. The typical HARQ RTT is 8ms, corresponding to $T = 8$ time-slots. Consequently, up to eight SAW HARQ processes are initiated in parallel. Each HARQ process corresponds to the transmission of a different message and requires a separate buffer allocation in the receiver for the purpose of combining the retransmissions. The HARQ process to which a packet belongs is identified by a unique HARQ IDentifier (ID). Furthermore, the following attributes, called control information, are additional signaling fields that aid the LTE

downlink HARQ operation:

- New Data Indicator (NDI): toggled whenever a new message transmission begins;
- Redundancy Version (RV): indicates the index ℓ of the HARQ round;
- Modulation and Coding Scheme (MCS): indicates the modulation and coding scheme.

The implementation of HARQ differs in the downlink and uplink. More precisely, the HARQ schemes can be either *synchronous* or *asynchronous*, with the retransmissions in each case being either *adaptive* or *non-adaptive*.

Synchronous vs. asynchronous HARQ In synchronous HARQ schemes, the retransmissions of each message $\mathbf{m}_k$ for each process, if requested by the receiver, occur at predefined time-slots relative to the initial transmission, *i.e.* , each T time-slots. In this way, their is no need to indicate the HARQ process ID to the receiver along with the transmission, as this can be inferred form the transmission timing. In contrast, asynchronous HARQ schemes have retransmissions that can occur at any time-slot relative to the initial transmission, usually upon reception of the receiver's feedback where the feedback delay varies at each round. In this case, additional signaling is required, so that the receiver can correctly associate each retransmission packet $\mathbf{p}_k(\ell)$, where $2 \leq \ell \leq C$, with its corresponding initial transmission $\mathbf{p}_k(1)$. Also, a scheduling algorithm is required at the transmitter to handle the retransmissions. Hence, synchronous HARQ schemes reduce the signaling overhead, while asynchronous HARQ schemes enables to have a flexible scheduling.

Adaptive vs. non-adaptive HARQ In adaptive HARQ schemes, the MCS and the transmission resource allocation in the frequency domain can be changed at each retransmission in response to variations in the radio channel conditions. In case of a non-adaptive HARQ scheme, the retransmissions are performed without explicit signaling of new transmission attributes (either by using the same transmission attributes as in the previous transmission, or by changing the attributes according to a predefined rule). Consequently, a gain can be obtained using adaptive HARQ schemes at the expense of increased signaling overhead.

In LTE, asynchronous adaptive HARQ is used for the downlink. Therefore, every downlink transmission is accompanied by explicit signaling of control information. The uplink HARQ is synchronous in LTE. On the other hand, the retransmissions may be either adaptive or non-adaptive in the uplink, depending on whether new signaling of transmission attributes is provided. The uplink non-adaptive HARQ requires an ordered RV sequence, *i.e.* , the packet index ℓ is incremental at each transmission ($p_k(1)$ at the first transmission, $p_k(2)$ at the second transmission, and so on). However, for the adaptive HARQ uplink, the index ℓ of the packet is explicitly signaled.

2.2.4 Multi-packet Hybrid ARQ

Superposition of packets, also called multi-packet or multi-layer transmission, enables the transmitter to send simultaneously multiple packets corresponding to different messages. When combined with HARQ, superposition of packets enables the (re)transmission of several packets, in different layers, during the same time-slot. Hereafter, we present a brief overview of recent works on multi-packet transmission in the context of HARQ.

- In [Steiner and Shamai, 2008], the authors propose a multi-layer HARQ scheme where every layer supports HARQ independently. In this way, the system performs CC-HARQ or IR-HARQ for each layer separately. The receiver's feedback (considered there with no feedback delay) indicates the highest successfully decoded layer, which implicitly designates the channel fading gain. Thus, the retransmission consists of additional parity bits only for undecoded layers. Moreover, a constant power is allocated per transmission. Using an information theoretic approach, this paper shows that combining layering and HARQ achieves high throughput with small delays.

- A multi-layer CC-HARQ transmission is adopted in [Assimi et al., 2009], where the layers are transmitted over the channel using a linear superposition of modulated symbols. Each layer is allocated a different power and transmission rate under the constraint of a fixed total power. Furthermore, the authors of this paper notice that multi-layer transmission can be viewed as a MISO system where the number of transmit antenna is equal to the number of layers. According to the instantaneous receiver's feedback, the transmitter resends the same packets for the layers in error and sends new packets for successfully decoded layers. On the other hand, the receiver uses previously received signals (stored in the buffer) in addition to the most recent signal to decode the messages. Moreover, the receiver uses SIC to decode the superposed layers. If a layer is successfully decoded, the buffer is updated by removing the contribution of the decoded layer from all the buffered signals. In summary, the main idea in this paper is that performing a joint detection of superposed layers using multiple signals containing that layer would enhance the detection performance, hence the overall system performance.

- The protocol in [Zhang and Hanzo, 2009] considers feedback on the Transport Control Protocol (TCP) layer, where each TCP frame corresponds to a pre-defined number of retransmissions with superposed packets. The main idea is to superpose a new HARQ packet with possibly multiple HARQ packets that are about to be retransmitted. In order to avoid large delays, multiple HARQ rounds are accumulated and transmitted at once in one composite packet of multiple superimposed layers at the TCP level. Their work elaborates on type-I HARQ, and the performance metrics are viewed at the TCP layer.

- In the conventional HARQ scheme, a retransmission packet is sent upon a NACK feedback which results in decreasing the transmission data rate. In order to provide a constant data rate, the authors in [Takahashi and Higuchi, 2010] propose a simple superposition modulation (for HARQ with a single possible retransmission) to combine the retransmission packet with the initial packet. More precisely, the initial and retransmission packets are firstly channel coded and interleaved separately. Then, the Quadrature Phase-shift Keying (QPSK)-modulated retransmission packet is superimposed on the QPSK-modulated initial packet. In this way, the overall transmission bandwidth is constant at the cost of the interference between the initial and retransmission packets. The receiver uses iterative interference cancellation where the retransmission packet is decoded first and the initial packet is decoded next. In this context, the proposed predetermined-rate HARQ scheme, based on superposition coding, achieves better throughput than the conventional HARQ scheme.

- A multi-packet IR-HARQ protocol is proposed in [Aoun et al., 2010] where the redundancy packets are created by jointly encoding several messages. Using this protocol, the transmitter sends successively a group of packets (one packet per time-slot), where each packet corresponds to a different message. Following the transmission of this group of packets, the receiver replies with a NACK feedback, without delay, if any of those packets is not successfully decoded. In this case, each retransmission consists of a jointly encoded redundancy packet corresponding to this group of messages. Moreover, these packets are encoded with incremental redundancy. In this way, the receiver performs joint decoding at each multi-packet retransmission. An extension of this protocol is proposed in [Aoun et al., 2011] where a perfect estimate of the CSI is available to the transmitter. Depending on the available CSI (average SNR estimation or instantaneous SNR estimation), different multi-packet HARQ protocols are proposed and analyzed using a Markov chain model.

- An information theoretic study in [Bandemer et al., 2012] considers multiple transmitters sending data on a broadcast channel to multiple receivers, where the encoding functions of each transmitter are restricted to superposition coding and time sharing. The main result is that *simultaneous nonunique decoding*, in which each receiver attempts to recover the unique codeword of its intended transmitter along with codewords from interfering senders, achieves the optimal rate region. One of this result's implications is useful in multi-packet transmission as it allows us to determine the decoding rate regions of each superposed packet at the receiver. This can be achieved by viewing multi-layer transmission as a MISO system where the number of transmit antenna is equal to the number of layers, as previously noted in [Assimi et al., 2009].

- The objective of the work in [Szczecinski et al., 2013] is to optimize rate allocation when using IR-HARQ over a Rayleigh block fading channel (whose realizations are independent during each time-slot), which enables the transmission of multiple packets in one time-slot using time sharing, accordingly. The main challenge is that rate adaption policies require necessarily some sort of CSI feedback to be available at the transmitter. In this paper, it is assumed that the transmitter has a perfect knowledge of the accumulated mutual information (of each HARQ process) at the receiver following each HARQ transmission and without delay. Based on this knowledge (which is a partial outdated CSI), the transmitter decides the transmission rate of each redundancy packet during the next transmission.

- In [Hamss et al., 2014; Jabi et al., 2015], a multi-packet IR-HARQ scheme where the transmitter allows two different packets to share the same time-slot is analyzed. The transmitter's role is to decide, at each time-slot, whether to use superposition coding or time sharing or send a single packet. This decision is based on an instantaneous multi-bit ACK/NACK feedback in addition to the accumulated mutual information at the receiver (which is sent to the transmitter through the feedback channel). In order to simplify the decoding of superposed packets, a sub-optimal SIC decoder is employed. Although, the authors discuss the possibility of joint decoding and interference removal.

- The IR-HARQ protocols in [Trillingsgaard and Popovski, 2018] are based on rate adaption policies using delayed (outdated) CSI at the transmitter. The delayed CSI is used to estimate the amount of unresolved information at the receiver. Accordingly, the transmitter decides the number of redundant information bits to send in addition to the new information bits, using joint encoding, in such a way that all information bits can be jointly decoded.

This non-exhaustive overview of the state of the art shows that multi-packet HARQ is of great interest, since it improves the throughput relatively to conventional HARQ. In summary, several strategies of multi-packet transmission exist. Those are:

- Joint encoding: different messages are jointly encoded (before modulation), then the obtained codeword is modulated into a single packet and sent by the transmitter;

- Time sharing: each message is separately encoded and modulated using a specific rate allocation policy in order to share a time-slot (in the time domain);

- Superposition coding: each message is separately encoded and modulated using a fixed rate. The resulting packets are superposed in one time-slot using a specific power allocation;

- Any combination of the previous strategies: some protocols propose to combine these strategies according to a specific transmit algorithm.

Notice that time sharing requires a rate adaption policy which usually requires some sort of CSI at the transmitter, provided via a feedback channel. Also, the efficiency of these policies depends on the quality of the CSI estimation. However, superposition coding does not necessarily require CSI at the transmitter. Another main advantage of superposition coding is to exploit MAC communication. Moreover, multi-packet transmission using superposition coding can be viewed as a MISO transmission. To the best of our knowledge, no previous works exist on multi-packet HARQ with a delayed feedback. Also, the previous works focus mainly on optimizing the throughput. Whereas, the delay is also a critical metric in several applications.

2.3 Proposed Protocol

We propose a protocol that enables the transmitter to anticipate the HARQ feedback by sending, in advance to its reception, packets related to unacknowledged messages using superposition coding. In contrast to previous works, where HARQ is considered without a delayed feedback of multiple time-slots, the proposed protocol is designed to counteract this feedback delay as well as to improve the throughput. The implemented HARQ protocol in LTE, *i.e.* parallel SAW HARQ, enables continuous transmission but does not mitigate the delay of the received messages as in the proposed protocol. Hereafter in Section 2.3.1, we present the system model. Then, we detail the proposed protocol in Section 2.3.2. Afterwards, a description of its corresponding decoder is provided in Section 2.3.3, and Section 2.3.4 is devoted to analyze the receiver's performance from an information theoretic point of view.

2.3.1 System model

We consider slotted point-to-point transmission where each time-slot corresponds to N channel uses. During each time-slot t, the transmitter sends N symbols stacked in vector $\mathbf{x}_t$. This vector may be composed of a packet or a superposition of packets, as it will be explained in Section 2.3.2. As in Chapter 1, the received signal at time-slot t is:

$$\mathbf{y}_t = h(t)\mathbf{x}_t + \mathbf{w}_t, \tag{2.1}$$

where $\mathbf{w}_t$ is an additive white Gaussian noise vector, with zero-mean and variance per component equal to N_0. We consider a Rayleigh flat fading channel with coherence time equal to the time-slot duration. The channel is fixed during a time-slot, but has independent realizations at each time-slot. We also assume perfect CSI at the receiver. The channel gain is denoted by $g(t)$ where $g(t) = \frac{|h(t)|^2}{N_0}$. Moreover, we use IR-HARQ with the same notations for the messages and packets as in Chapter 1. The feedback is error-free and only composed of ACK or NACK of the considered messages. We assume a feedback delay of T time-slots, which means that the feedback related to a transmission

performed in time-slot t is received by the transmitter just before the beginning of time-slot $t+T$. The case $T=1$ corresponds then to a no-delay feedback. We assume moreover that this delay is due to the return channel and not to the decoding time at the receiver, which means that the receiver knows at the end of time-slot t if the messages related to the packets transmitted at time-slot t are successfully decoded or not. A message is said in *timeout* if it is not ACKed by CT time-slots after its first transmission, corresponding to the timeout in parallel SAW HARQ.

2.3.2 Transmitter strategy

In the proposed protocol, at each time-slot the transmitter selects a packet $\mathbf{p}_k(\ell)$, based on the ACK/NACK feedbacks, as in parallel SAW HARQ. The transmitter may superpose to $\mathbf{p}_k(\ell)$ a second packet $\mathbf{p}_{k'}(\ell')$, with $k' \neq k$, even if there is not any feedback on previous transmissions of message $\mathbf{m}_{k'}$ yet. The idea is to send a redundant packet without waiting for the feedback to arrive at the transmitter side, which enables the receiver to possibly decode $\mathbf{m}_{k'}$ without waiting for the next SAW HARQ round. Accordingly, the transmission occurs in two layers where:

- Layer 1 acts as the parallel SAW HARQ protocol;
- Layer 2 corresponds to the transmission of additional redundant packets.

In order to keep the same energy at each time-slot, the superposed packet, belonging to the second layer, uses $100(1-\alpha)\%$ of the predefined energy per time-slot, while the packet sent by the first layer uses $100\alpha\%$ of the energy, with $\alpha \in [0,1]$. The influence of α will be investigated in Section 2.4. The transmit vector $\mathbf{x}_t$ is given by:

$$\mathbf{x}_t = \begin{cases} \mathbf{p}_k(\ell), & \text{if no superposition,} \\ \sqrt{\alpha}\mathbf{p}_k(\ell) + \sqrt{1-\alpha}\mathbf{p}_{k'}(\ell'), & \text{if superposition.} \end{cases} \tag{2.2}$$

We note that the case of $\alpha = 1$ corresponds to the parallel SAW HARQ. Notice also that superposing many HARQ packets in the same time-slot is limited by interference between the superposed packets. Although more layers could be superposed in theory, we superpose two layers at most to avoid increasing the decoder complexity in practical systems, since the decoder has to manage the interference between the superposed layers.

At the beginning of time-slot t the transmitter knows the ACK/NACK related to the messages sent up to time-slot $t-T$ (because of the feedback delay). According to this knowledge, the transmitter selects the packets to include in $\mathbf{x}_t$. As anticipated, the first layer acts as parallel SAW HARQ. Therefore, if packet $\mathbf{p}_k(\ell)$ was sent in the first layer at time-slot $t-T$, the reception of a NACK relative to message $\mathbf{m}_k$ just before time-slot t triggers the transmission of another redundancy packet $\mathbf{p}_k(\ell+1)$, as long as $\ell < C$. Otherwise, the reception of an ACK of $\mathbf{m}_k$ triggers the transmission of a packet $\mathbf{p}_{k''}(1)$ associated with a new message $\mathbf{m}_{k''}$ (never transmitted before). The selection of the

superposed packet in the second layer is done according to the following principles: i) superposing packets related to the most recent messages of the first layer to reduce the delay, ii) superposing unsent redundant packets to reduce the message error by using transmit diversity. Based on these principles, we describe the selection strategy by the following rules (ordered by priority), which determine the choice of the superposed packet in the second layer:

1. A packet $\mathbf{p}_{k'}(\ell')$ cannot be superposed if message $\mathbf{m}_{k'}$ is in timeout or previously ACKed.

2. As long as there are unacknowledged messages with unsent packets, the superposed packet is the unsent packet of the lowest index ℓ' of the most recent message $\mathbf{m}_{k'}$, with $k' \neq k$ (different messages in the two layers).

3. If the transmitter already sent all the packets of all the unacknowledged messages that are not in timeout, the superposed packet is the packet with the lowest index ℓ' that was not previously sent in the second layer. (Notice that this packet has been already sent once, in the first layer).

4. No packet is superposed to a packet of the first layer that has $\ell = C$.

The first rule prevents larger delays than those provided by conventional parallel SAW HARQ. The second rule reduces the delay furthermore by sending redundant packets related to unacknowledged messages in advance to the receiver's feedback, and it provides a diversity gain. Likewise, the third rule provides more diversity gains by superposing packets related to different messages at each time-slot, in addition to sending different RVs corresponding to each message in the second layer. Moreover, the fourth rule reduces the probability to drop messages by forbidding interference during the last retransmission. Notably, this last rule is necessary in order to simplify the decoding at the receiver side by limiting the number of messages (to be decoded) in the buffer. In other words, one can check that, at each time-slot, at most T messages are not previously ACKed nor in timeout, which means also that the feedback at each time-slot contains at most T feedback bits (ACKs/NACKs).

In Sections 2.3.2.1 and 2.3.2.2, we provide two examples of our protocol, with $C = 3$ and $T = 3$. We remind that we consider instantaneous decoding at the end of the time-slot t, although $\mathcal{F}_t$ will be available at the transmitter side after T time-slots.

2.3.2.1 A realization of the proposed protocol with NACK feedbacks only

To illustrate the transmitter strategy, we show a realization of the proposed protocol with NACK feedbacks only in Fig. 2.3. For clarity reasons, we show the case with a feedback delay of $T = 3$ time-slots and $C = 3$ HARQ transmission credits. Hence, layer 1 acts as a

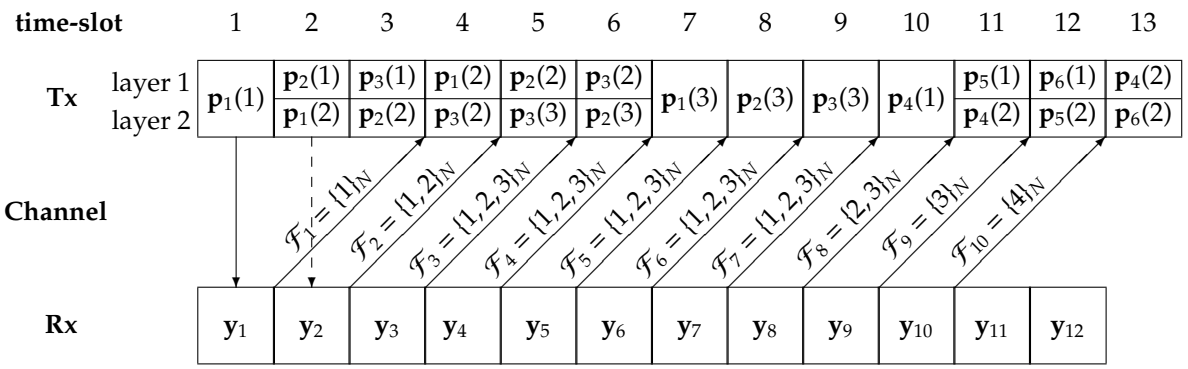


Figure 2.3 – A realization of the proposed protocol with NACK feedbacks only.

SAW HARQ mechanism with $T = 3$ parallel processes, while layer 2 corresponds to the redundant packets that are sent using superposition coding.

In layer 1, message $\mathbf{m}_1$ is sent at time-slot 1 through packet $\mathbf{p}_1(1)$. The receiver fails to decode $\mathbf{p}_1(1)$, and the corresponding NACK feedback, *i.e.* $\mathcal{F}_1 = \{1\}_N$, is available to the transmitter just before time-slot 4. Therefore, the transmitter sends $\mathbf{p}_1(2)$ at time-slot 4. The receiver fails again to decode packet $\mathbf{p}_1(2)$ of message $\mathbf{m}_1$, and the feedback $\mathcal{F}_4 = \{1, 2, 3\}_N$ is available to the transmitter just before time-slot 7. Hence, the transmitter sends $\mathbf{p}_1(3)$ at time-slot 7. Two other parallel SAW HARQ processes are initiated in this layer. The second process corresponds to sending $\mathbf{m}_2$ and occupies the time-slots 2, 5 and 8, while the third process corresponds to sending $\mathbf{m}_3$ and occupies the time-slots 3, 6 and 9. The messages $\mathbf{m}_1$, $\mathbf{m}_2$ and $\mathbf{m}_3$ are dropped, at the beginning of time-slot 10, due to time-out. Then, three new SAW HARQ processes, corresponding to messages $\mathbf{m}_4$, $\mathbf{m}_5$ and $\mathbf{m}_6$, start in layer 1 at time-slots 10, 11 and 12, respectively. In this way, the transmissions continue in layer 1 using the SAW HARQ mechanism.

Using superposition coding as in Eq. (2.2), redundant packets are sent in layer 2, based on the four previously presented rules in this Section 2.3.2.

- At time-slot 1, there is no packet to send in layer 2 according to the second rule. Therefore, the transmitter sends $\mathbf{x}_1 = \mathbf{p}_1(1)$.
- At time-slot 2, the message $\mathbf{m}_1$ remains unacknowledged, since the feedback $\mathcal{F}_1$ is not yet available to the transmitter. In addition, $\mathbf{m}_1$ is the most recent message with index $k' \neq 2$. Moreover, two packets of message $\mathbf{m}_1$ are unsent previously. Those are $\mathbf{p}_1(2)$ and $\mathbf{p}_1(3)$. Among these two packets, $\mathbf{p}_1(2)$ is the packet with the lowest index $\ell = 2$. Due to constant power allocation per time-slot, a power fraction α is assigned to the packet of layer 1, whereas the packet of layer 2 is superposed with power fraction $1-\alpha$. Therefore, the transmitted signal is $\mathbf{x}_2 = \sqrt{\alpha}\mathbf{p}_2(1) + \sqrt{1-\alpha}\mathbf{p}_1(2)$.
- At time-slot 3, $\mathbf{m}_2$ is the most recent unacknowledged message with unsent packets, different from the sent message in layer 1. Also, $\mathbf{p}_2(2)$ is the unsent packet with the lowest index $\ell = 2$ corresponding to message $\mathbf{m}_2$. Hence, $\mathbf{p}_2(2)$ is superposed to $\mathbf{p}_3(1)$

in $\mathbf{x}_3$, where $\mathbf{x}_3 = \sqrt{\alpha}\mathbf{p}_3(1) + \sqrt{1-\alpha}\mathbf{p}_2(2)$.

- At time-slot 4, the most recent unacknowledged message with unsent packets different from the sent message in layer 1 is $\mathbf{m}_3$. The unsent packet of the lowest index of this message is $\mathbf{p}_3(2)$. Therefore, the transmitted signal is $\mathbf{x}_4 = \sqrt{\alpha}\mathbf{p}_1(2) + \sqrt{1-\alpha}\mathbf{p}_3(2)$.
- At time-slot 5, the most recent unacknowledged message with unsent packets, different from the sent message in layer 1 is also $\mathbf{m}_3$. The packet $\mathbf{p}_3(1)$ was sent at time-slot 3 in the first layer. Also, $\mathbf{p}_3(2)$ was sent at time-slot 4 in the second layer. Hence, the unsent packet of the lowest index of this message is $\mathbf{p}_3(3)$. As a result, the transmitted signal is $\mathbf{x}_5 = \sqrt{\alpha}\mathbf{p}_2(2) + \sqrt{1-\alpha}\mathbf{p}_3(3)$.
- At time-slot 6, all the packets of message $\mathbf{m}_3$ are sent by the transmitter in the previous time-slots. In addition, $\mathbf{p}_3(2)$ corresponding to $\mathbf{m}_3$ is being transmitted in layer 1. Hence, the most recent unacknowledged message with unsent packets different from the sent message in layer 1 is $\mathbf{m}_2$. The unsent packet of the lowest index of this message is $\mathbf{p}_2(3)$. In this way, the transmitted signal is $\mathbf{x}_6 = \sqrt{\alpha}\mathbf{p}_3(2) + \sqrt{1-\alpha}\mathbf{p}_2(3)$.
- At time-slots 7, 8 and 9, no packets are superposed to the packets of the first layer according to the fourth rule, since $\ell = 3(= C)$. The transmitted signals at time-slots 7, 8 and 9 are $\mathbf{x}_7 = \mathbf{p}_1(3)$, $\mathbf{x}_8 = \mathbf{p}_2(3)$ and $\mathbf{x}_9 = \mathbf{p}_3(3)$, respectively.

Afterwards, the three SAW (truncated) HARQ processes of layer 1 are terminated progressively, and no packets corresponding messages $\mathbf{m}_1$, $\mathbf{m}_2$ nor $\mathbf{m}_3$ will be sent in layer 2 during future time-slots, according to the first rule. Consequently, the transmission of new messages is initiated starting from time-slot 10.

2.3.2.2 Another realization of the proposed protocol

This instance of the proposed protocol, illustrated in Fig. 2.4, clarifies furthermore the transmitter strategy by showing the transmitter's reaction to the occurrence of ACK feedbacks. Moreover, the receiver responds with ACKs to the transmission of messages $\mathbf{m}_2$ and $\mathbf{m}_3$ at time-slot $t = 3$. The corresponding feedback ($\mathcal{F}_3 = \{2,3\}_A, \{1\}_N$) is available

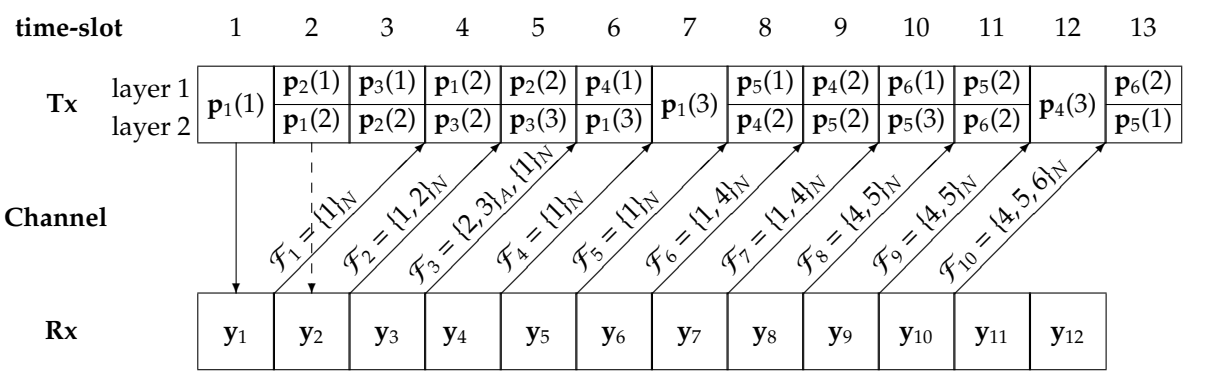


Figure 2.4 – Another realization of the proposed protocol.

to the transmitter at the beginning of time-slot 6 due to the feedback delay of $T = 3$. Hence, the transmitter decisions from $t = 1$ to $t = 5$ are the same as in the realization of the proposed protocol in Section 2.3.2.1. Notice that, due to the feedback delay of $T = 3$, $\mathbf{m}_2$ is retransmitted in layer 1 at time-slot 5 through packet $\mathbf{p}_2(2)$, although this message was successfully decoded by the receiver at the end of time-slot 3. The same remark applies to message $\mathbf{m}_3$ which is retransmitted in layer 2 via the packets $\mathbf{p}_3(2)$ and $\mathbf{p}_3(3)$ at $t = 4$ and $t = 5$, respectively. These *useless* retransmissions are inevitable since the transmitter takes decisions in advance to the reception of the receiver's feedback. However, the *early* retransmission of $\mathbf{p}_2(2)$ in layer 2 enabled the decoding of $\mathbf{m}_2$ at time-slot 3. In other words, $\mathbf{m}_2$ is delivered to the receiver with a small delay of 2 time-slots only (which is not possible using parallel SAW HARQ protocol). Before the beginning of time-slot 6, the SAW HARQ processes, in layer 1, corresponding the acknowledged messages $\mathbf{m}_2$ and $\mathbf{m}_3$ are terminated . Consequently, a new SAW HARQ process is initiated at time-slot 6 by sending a new message $\mathbf{m}_4$ through packet $\mathbf{p}_4(1)$. At time-slot 7, the SAW HARQ process corresponding to $\mathbf{m}_1$ is carried on by transmitting $\mathbf{p}_1(3)$. A new SAW HARQ process is initiated as well at time-slot 8 with message $\mathbf{m}_5$. Then, a retransmission of message $\mathbf{m}_4$ through packet $\mathbf{p}_4(2)$ occurs at time-slot 9. At the end of time-slot 9, $\mathbf{m}_1$ is in time-out. Therefore, the transmission of a new message $\mathbf{m}_6$, via $\mathbf{p}_6(1)$, starts at $t = 10$, and so on. Notice that the number of parallel SAW HARQ processes in layer 1 is always equal to the feedback delay $T = 3$.

In layer 2, we apply the four rules of the transmitter strategy. The observations during the first 5 time-slots (from $t = 1$ to $t = 5$) are similar to those in the previous example, as described in Section 2.3.2.1.

- At time-slot 6, the only unsent packet corresponding to an unacknowledged message is $\mathbf{p}_1(3)$. Therefore, this packet is chosen according to the first two rules of the transmitter strategy, in advance to its transmission in layer 1 during the next time-slot.
- At time-slot 7, the transmitted packet in layer 1, *i.e.* $\mathbf{p}_1(3)$, corresponds to the last HARQ round related to message $\mathbf{m}_1$. Hence, no packet is superposed in this transmission according to the fourth rule, and the $\mathbf{m}_1$ is in timeout at the end of this time-slot.
- At time-slot 8, the unacknowledged messages that are not in timeout are $\mathbf{m}_4$ and $\mathbf{m}_5$. By applying the second rule, the superposed packet is $\mathbf{p}_4(2)$ since it corresponds to the most recent unacknowledged message with unsent packets (which is $\mathbf{m}_4$) different from the sent message in layer 1 (which is $\mathbf{m}_5$ through packet $\mathbf{p}_5(1)$).
- At time-slot 9, $\mathbf{p}_5(2)$ is superposed based on a similar reasoning as in time-slot 8.

During the next time-slots, the superposition of packets in the second layer continues according to these rules, as depicted in Fig. 2.4 .

In this example, the proposed protocol allowed the receiver to decode $\mathbf{m}_2$ with a smaller delay of 2 time-slots. This is achieved by transmitting $\mathbf{p}_2(2)$ using superposition coding, in advance to the reception of the receiver's feedback $\mathcal{F}_2$. Furthermore, it is

sufficient to apply the four rules of the transmitter strategy to obtain the transmitter's decision regarding the superposition of a packet in layer 2, at any time-slot for any sequence of feedbacks.

2.3.3 Receiver analysis

Due to multi-packet transmission, the received signals share common information. The receiver attempts to decode multiple messages at each time-slot, using the current and previous observations.

In Section 2.3.3.1, we discuss the receiver's observation window and buffer size. Then, we explain in Section 2.3.3.2 the multi-bit ACK/NACK feedback vector $\mathcal{F}_t$, and we specify the set of messages that the receiver attempts to decode at time-slot t, denoted by $\mathcal{M}_t$.

2.3.3.1 Buffer size

A received signal at time-slot t could share common information with another received signal at any previous time-slot.

For instance, in the protocol's realization in Section 2.3.2.2, decoding $\mathbf{m}_4$ at time-slot 12 could benefit from the observations in all the previous time-slots. More precisely, at time-slot 6, $\mathbf{p}_4(1)$ is superposed to $\mathbf{p}_1(3)$. Therefore, decoding $\mathbf{m}_1$ would help in decoding $\mathbf{m}_4$ by removing the interference at time-slot 6, hence increasing the accumulated mutual information at the receiver corresponding to $\mathbf{m}_4$, which helps in decoding $\mathbf{m}_4$. Vice versa, decoding $\mathbf{m}_4$ at time-slot 12 would help in decoding $\mathbf{m}_1$, since it removes the interference due to $\mathbf{p}_4(1)$ at time-slot 6, hence it increases the accumulated mutual information corresponding to $\mathbf{m}_1$ at the receiver, which helps in decoding $\mathbf{m}_1$.

Although an optimal decoder would require an unlimited buffer size, the buffer size should be fixed for the following reasons:

- Decoding a message in timeout is not useful in most applications. If needed, a retransmission of this message could be handled by upper layer protocols.
- If a message is in timeout, it is more likely that the accumulated mutual information associated with this message at the receiver at time-slot t, which is provided by the transmissions before $t - CT$ time-slots, is low. In other words, the benefit of considering more than CT observations is low.
- In addition, the decoder becomes very complex if we consider an unlimited buffer size (both in practice and using information theoretic analysis).
- Moreover, the buffer size is limited in practice.

Therefore, the receiver's buffer consists of the last CT received signals. For decoding purposes, the receiver would consider the undecoded messages in this observation window as interference. Whereas, the decoded messages are removed from the observations,

which enhances the decoding performance. Keeping in the buffer the most recent CT observations, which correspond to the most recent CT time-slots, is a trade-off between the decoder's performance and the buffer size.

In the example in Section 2.3.2.2, only the observations from $t = 4$ to $t = 12$ (included) are kept in the buffer at $t = 12$, which corresponds to the most recent CT observations ($C = 3$ and $T = 3$). Moreover, decoding $\mathbf{m}_1$ at time-slot 12 is not useful, or could be assigned to upper layer protocols, since $\mathbf{m}_1$ is timeout at $t = 12$. Moreover, decoding $\mathbf{m}_1$ failed before time-slot 12, hence it is more likely that the accumulated mutual information at the receiver related to $\mathbf{m}_1$ that is provided by the transmissions before time-slot 4, is low.

In summary, the buffer size is fixed to CT observations which induces suboptimal decoding. However, the performance degradation due to the limited buffer size is marginal.

2.3.3.2 Feedback vector

At the end of time-slot t, the receiver considers the observations of the most recent CT time-slots. Since there are T parallel HARQ processes in this observation window, there are at most T undecoded messages (that are not in timeout). Therefore, the receiver attempts to decode these messages. Considering the observation window and superposition coding, the system is equivalent to a MISO channel with T (virtual) users, where each user is associated with a message. The output of the receiver is the feedback vector $\mathcal{F}_t$, which will be available at the transmitter at the beginning of time-slot $t + T$. The feedback vector $\mathcal{F}_t$ contains the ACK/NACK bits corresponding to the messages that i) are object of decoding at time-slot t, and ii) will not be in timeout at time-slot $t + T$. Hereafter, we show the realization of the proposed protocol that was explained in Section 2.3.2.2 from the receiver's perspective. The set of messages to decode $\mathcal{M}_t$ at time-slot t and the receiver's buffer at $t = 12$ are shown in Fig. 2.5. In this instance, $\mathbf{m}_1$ and $\mathbf{m}_4$ are object of decoding at time-slot 7, *i.e.*, $\mathcal{M}_7 = \{\mathbf{m}_1, \mathbf{m}_4\}$. Also, $\mathcal{F}_7$ contains the ACK/NACK bits corresponding to these messages. However, $\mathbf{m}_1$ will be in timeout by time-slot 10. Hence, $\mathcal{F}_8$ does not contain feedback information corresponding to $\mathbf{m}_1$. Notice that $\mathcal{F}_8$ will be available to the transmitter just before the start of time-slot 11. Moreover, attempting to decode all the messages, including the ones that will be in timeout, is beneficial because it removes the interference that is introduced by the superposition. This can be seen at time-slot 8 where the receiver attempts to decode $\mathbf{m}_1$, $\mathbf{m}_4$ and $\mathbf{m}_5$,*i.e.*, $\mathcal{M}_8 = \{\mathbf{m}_1, \mathbf{m}_4, \mathbf{m}_5\}$. Since $\mathbf{m}_1$ will be in timeout by time-slot 10, $\mathcal{F}_8$ contains only information about $\mathbf{m}_4$ and $\mathbf{m}_5$. However, attempting to decode $\mathbf{m}_1$ (which is in timeout) is beneficial since it allows to remove the interference on message $\mathbf{m}_4$ at time-slot 6.

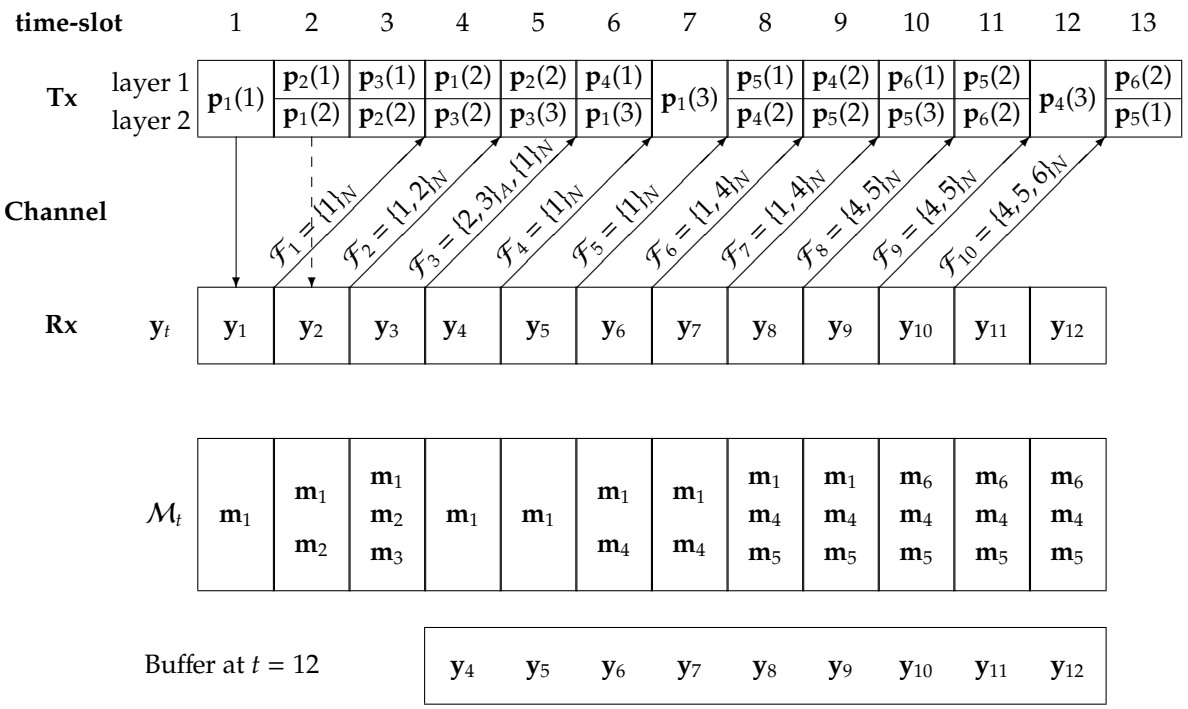


Figure 2.5 – Receiver's buffer during an instance of the proposed protocol.

2.3.4 Information theoretic characterization of the receiver

Hereby, we describe the receiver from an information theoretic point of view. We remind that $\mathcal{M}_t$ is the set of messages that the receiver is attempting to decode at time-slot t. Let $\mathcal{D}_t$ be the subset of successfully decoded messages by the receiver at time-slot t such that $\mathcal{D}_t \subseteq \mathcal{M}_t$. If the receiver successfully decodes the messages in $\mathcal{D}_t$ and none of the messages in $\mathcal{M}_t \setminus \mathcal{D}_t$, we say that the decoder operates in the rate region $\mathcal{R}_{\mathcal{D}_t}$. The set $\mathcal{D}_t$, along with the rules of the transmit protocol, allows to obtain $\mathcal{F}_t$. In order to characterize the decoding outcome, we i) evaluate the rate region $\mathcal{R}_{\mathcal{D}_t}$ for every possible $\mathcal{D}_t \subseteq \mathcal{M}_t$; and ii) determine, on the basis of the available observations, the operating rate region $\mathcal{R}_{\mathcal{D}_t}$ of the receiver. By definition, $\mathcal{R}_{\mathcal{D}_t}$ is given by the union of rate regions where the messages in $\mathcal{D}_t$ are successfully decoded (alone or jointly with other messages in $\mathcal{M}_t \setminus \mathcal{D}_t$), excluding the regions where the messages in $\mathcal{D}_t$ are jointly decoded with at least another message in $\mathcal{M}_t \setminus \mathcal{D}_t$. By construction of the system, the receiver can see the messages as users of a MAC. For a set of users $\mathcal{S}$, $\mathscr{R}_{MAC(\mathcal{S})_t}$ is the MAC rate region of users $\mathcal{S}$ at time-slot t considering the messages from users outside $\mathcal{S}$ as noise [Bandemer et al., 2012].

We consider first the case $\mathcal{D}_t \neq \emptyset$. The region where the messages in $\mathcal{D}_t$, and possibly other messages in $\mathcal{M}_t$, are successfully decoded is the union of the MAC rate regions of any set of users that includes $\mathcal{D}_t$, *i.e.*, $\bigcup_{\mathcal{D}_t \subseteq \mathcal{S}} \mathscr{R}_{MAC(\mathcal{S})_t}$ [Bandemer et al., 2012]. The region where the messages in $\mathcal{D}_t$ are successfully decoded, jointly with at least another message in $\mathcal{M}_t$, is the union of the MAC regions of any set that includes $\mathcal{D}_t$ and at least another

user from $\mathcal{M}_t \setminus \mathcal{D}_t$, *i.e.*, $\bigcup_{\mathcal{D}_t \subset \mathcal{S}, \mathcal{S} \neq \mathcal{D}_t} \mathscr{R}_{MAC(\mathcal{S})_t}$ [Bandemer et al., 2012]. We deduce the rate region in Eq. (2.3):

$$\begin{aligned} \mathcal{R}_{\mathcal{D}_t} &= \left(\bigcup_{\mathcal{D}_t \subseteq \mathcal{S}} \mathscr{R}_{MAC(\mathcal{S})_t} \right) \bigcap \left(\overline{\bigcup_{\substack{\mathcal{D}_t \subset \mathcal{S}', \\ \mathcal{S}' \neq \mathcal{D}_t}} \mathscr{R}_{MAC(\mathcal{S}')_t}} \right) \\ &= \mathscr{R}_{MAC(\mathcal{D}_t)_t} \bigcap \left(\bigcap_{\mathcal{D}_t \subset \mathcal{S}, \mathcal{S} \neq \mathcal{D}_t} \overline{\mathscr{R}_{MAC(\mathcal{S})_t}} \right). \end{aligned} \tag{2.3}$$

Since the regions $\mathcal{R}_{\mathcal{D}_t}$, for all possible $\mathcal{D}_t \subseteq \mathcal{M}_t$ form a partition by construction, the region $\mathcal{R}_{\mathcal{D}_t=\emptyset}$ is the complementary of the union of all rate regions for $\mathcal{D}_t \neq \emptyset$, *i.e.*,

$$\mathcal{R}_{\emptyset_t} = \overline{\bigcup_{\mathcal{D}_t \subseteq \mathcal{M}_t, \mathcal{D}_t \neq \emptyset} \mathcal{R}_{\mathcal{D}_t}} = \bigcap_{\mathcal{D}_t \subseteq \mathcal{M}_t, \mathcal{D}_t \neq \emptyset} \overline{\mathcal{R}_{\mathcal{D}_t}}. \tag{2.4}$$

Then, to determine whether the receiver operates in $\mathcal{R}_{\mathcal{D}_t}$, for any $\mathcal{D}_t \subseteq \mathcal{M}_t$, it is enough to verify whether the receiver operates within or outside the set of MAC regions involved in Eq. (2.3) and Eq. (2.4). The receiver operates in the MAC rate region $\mathscr{R}_{MAC(\mathcal{S})_t}$, for a set of messages $\mathcal{S}$, if the following set of inequalities is satisfied [Bandemer et al., 2012]:

$$\sum_{j \in \mathcal{T}} R_j \leq I(X_{\mathcal{T}}; Y_t | X_{\mathcal{S} \setminus \mathcal{T}}), \text{ for all } \mathcal{T} \subseteq \mathcal{S}, \tag{2.5}$$

where Y_t is the set of observations (received signals) during the most recent CT time-slots, *i.e.*, $Y_t = \{\mathbf{y}_{t-CT}, \ldots, \mathbf{y}_t\}$. $X_{\mathcal{T}}$ represents the sent packets relative to the messages in $\mathcal{T}$, and $X_{\mathcal{S} \setminus \mathcal{T}}$ is interpreted likewise. The packets relative to messages that are not in $\mathcal{S}$ are treated as interference. We also have $R_j = R$. The mutual information $I(X_{\mathcal{T}}; Y_t | X_{\mathcal{S} \setminus \mathcal{T}})$ can be calculated by reading the observations in the window of size CT, and cumulating the mutual information corresponding to the messages in $\mathcal{T}$. In this process, we need to consider that: i) some packets are superposed, and sent with different power fractions, ii) the same packet may be transmitted more than once, iii) messages which have been already decoded in the past may allow to eliminate interfering packets in the observations.

Hereby, we apply this characterization to the example in Fig. 2.5.

At time-slot 1, the receiver attempts to decode $\mathcal{M}_1 = \{\mathbf{m}_1\}$ based on the observation at $t = 1$ in Eq. (2.6) which is given by:

$$\mathbf{y}_1 = h(1)\mathbf{p}_1(1) + \mathbf{w}_1. \tag{2.6}$$

Hence, the corresponding rate region is obtained as $\mathscr{R}_{\{\mathbf{m}_1\}_{t=1}} = \mathscr{R}_{MAC(\{\mathbf{m}_1\})_{t=1}}$ by applying Eq. (2.3). In other words, the receiver would have succeeded to decode $\mathbf{m}_1$ if:

$$R \leq \log(1 + g(1)). \tag{2.7}$$

However, this inequality is not satisfied. Therefore, the receiver responds with a NACK, $\mathcal{F}_1 = \{1\}_N$.

At time-slot 2, the receiver tries to decode the set of messages $\mathcal{M}_2 = \{\mathbf{m}_1, \mathbf{m}_2\}$ based on the observations at $t = 1$ and $t = 2$. These observations are given in Eq. 2.8 as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} h(1)\mathbf{1}_N & 0 \\ 0 & \sqrt{1-\alpha}h(2)\mathbf{1}_N \end{bmatrix} \begin{bmatrix} \mathbf{p}_1(1) \\ \mathbf{p}_1(2) \end{bmatrix} + \begin{bmatrix} 0 \\ \sqrt{\alpha}h(2)\mathbf{1}_N \end{bmatrix} \mathbf{p}_2(1) + \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}. \tag{2.8}$$

Notice that, $\mathbf{p}_2(1)$ is transmitted with power fraction $\sqrt{\alpha}$ while $\mathbf{p}_1(2)$ is superposed with power fraction $\sqrt{1-\alpha}$. Hence, the MAC rate regions can be written as:

- $\mathscr{R}_{MAC(\{\mathbf{m}_1\})_{t=2}} = \Big\{R \leq \log(1 + g(1)) + \log(1 + \frac{(1-\alpha)g(2)}{1+\alpha g(2)})$,
- $\mathscr{R}_{MAC(\{\mathbf{m}_2\})_{t=2}} = \Big\{R \leq \log(1 + \frac{\alpha g(2)}{1+(1-\alpha)g(2)})$,
- $\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\})_{t=2}} = \begin{cases} R \leq \log(1 + g(1)) + \log(1 + (1-\alpha)g(2)); \\ R \leq \log(1 + \alpha g(2)); \\ 2R \leq \log(1 + g(1)) + \log(1 + g(2)). \end{cases}$

In the following, we characterize all the possible decoding outcomes, according to Eq. (2.3) and Eq. (2.4).

- The receiver succeeds to decode $\mathbf{m}_1$ and fails to decode $\mathbf{m}_2$, *i.e.* $\mathcal{D}_2 = \{\mathbf{m}_1\}$, if the decoding rate belongs to the rate region:

$$\mathscr{R}_{\{\mathbf{m}_1\}_{t=2}} = \mathscr{R}_{MAC(\{\mathbf{m}_1\})_{t=2}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\})_{t=2}}}. \tag{2.9}$$

- The receiver succeeds to decode $\mathbf{m}_2$ and fails to decode $\mathbf{m}_1$, *i.e.* $\mathcal{D}_2 = \{\mathbf{m}_2\}$, if the decoding rate belongs to the rate region:

$$\mathscr{R}_{\{\mathbf{m}_2\}_{t=2}} = \mathscr{R}_{MAC(\{\mathbf{m}_2\})_{t=2}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\})_{t=2}}}. \tag{2.10}$$

- The receiver succeeds to decode both messages $\mathbf{m}_1$ and $\mathbf{m}_2$, *i.e.* $\mathcal{D}_2 = \{\mathbf{m}_1, \mathbf{m}_2\}$, if the decoding rate belongs to the rate region:

$$\mathscr{R}_{\{\mathbf{m}_1,\mathbf{m}_2\}_{t=2}} = \mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\})_{t=2}}. \tag{2.11}$$

- The receiver fails to decode both messages, *i.e.* $\mathcal{D}_2 = \emptyset$, if the decoding rate belongs to the rate region:

$$\mathcal{R}_{\emptyset_{t=2}} = \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1\})_{t=2}}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_2\})_{t=2}}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\})_{t=2}}}. \tag{2.12}$$

Moreover, these rate regions are depicted in Fig. 2.6. In this realization of the protocol (which is depicted in Section 2.3.2.2 and in Fig. 2.5), the decoding rate belongs to the rate region $\mathcal{R}_{\emptyset_{t=2}}$. Hence, none of the messages were successfully decoded at $t = 2$.

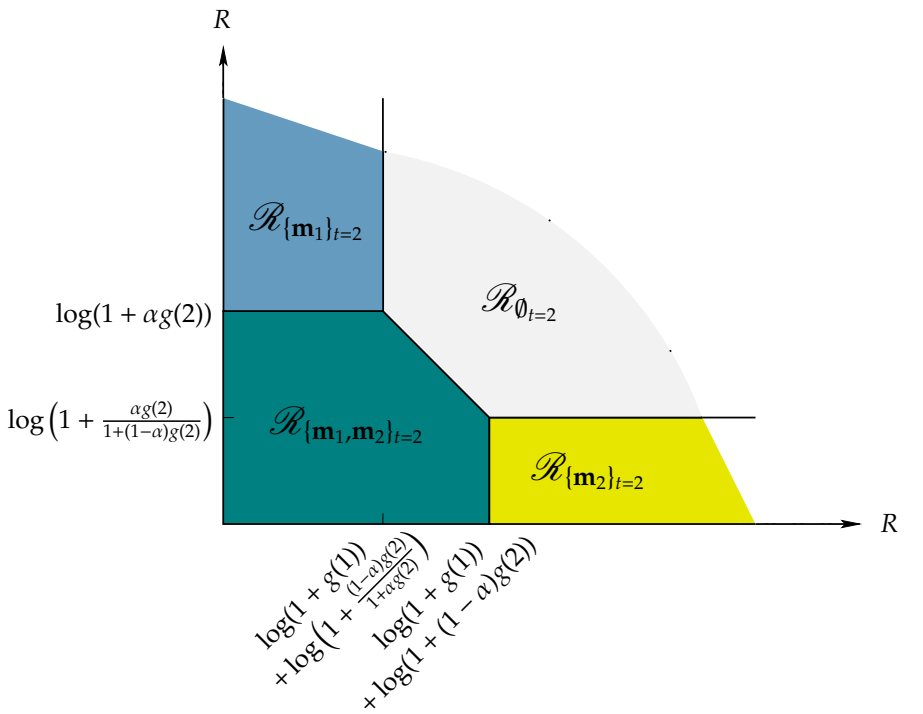


Figure 2.6 – The rate regions corresponding to the protocol realization at $t = 2$.

At time-slot 3, the receiver tries to decode the set of messages $\mathcal{M}_3 = \{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$ based on the observations at $t = 1$, $t = 2$ and $t = 3$. These observations are given in Eq. 2.13 as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix} = \begin{bmatrix} h(1)\mathbf{1}_N & 0 \\ 0 & \sqrt{1-\alpha}h(2)\mathbf{1}_N \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{p}_1(1) \\ \mathbf{p}_1(2) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \sqrt{\alpha}h(2)\mathbf{1}_N & 0 \\ 0 & \sqrt{1-\alpha}h(3)\mathbf{1}_N \end{bmatrix} \begin{bmatrix} \mathbf{p}_2(1) \\ \mathbf{p}_2(2) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \sqrt{\alpha}h(3)\mathbf{1}_N \end{bmatrix} \mathbf{p}_3(1) + \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \mathbf{w}_3 \end{bmatrix}. \tag{2.13}$$

Also, $\mathbf{p}_3(1)$ is transmitted with power fraction $\sqrt{\alpha}$ while $\mathbf{p}_2(2)$ is superposed with power fraction $\sqrt{1-\alpha}$. Hence, the MAC rate regions can be written as:

- $\mathscr{R}_{MAC(\{\mathbf{m}_1\})_{t=3}} = \Big\{ R \leq \log(1+g(1)) + \log(1+\frac{(1-\alpha)g(2)}{1+\alpha g(2)}),$
- $\mathscr{R}_{MAC(\{\mathbf{m}_2\})_{t=3}} = \Big\{ R \leq \log(1+\frac{\alpha g(2)}{1+(1-\alpha)g(2)}) + \log(1+\frac{(1-\alpha)g(3)}{1+\alpha g(3)}),$
- $\mathscr{R}_{MAC(\{\mathbf{m}_3\})_{t=3}} = \Big\{ R \leq \log(1+\frac{\alpha g(3)}{1+(1-\alpha)g(3)}),$
- $\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\})_{t=3}} = \begin{cases} R \leq \log(1+g(1)) + \log(1+(1-\alpha)g(2)); \\ R \leq \log(1+\alpha g(2)) + \log(1+\frac{(1-\alpha)g(3)}{1+\alpha g(3)}); \\ 2R \leq \log(1+g(1)) + \log(1+g(2)) + \log(1+\frac{(1-\alpha)g(3)}{1+\alpha g(3)}), \end{cases}$
- $\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_3\})_{t=3}} = \begin{cases} R \leq \log(1+g(1)) + \log(1+\frac{(1-\alpha)g(2)}{1+\alpha g(2)}); \\ R \leq \log(1+\frac{\alpha g(3)}{1+(1-\alpha)g(3)}); \\ 2R \leq \log(1+g(1)) + \log(1+\frac{(1-\alpha)g(2)}{1+\alpha g(2)}) + \log(1+\frac{\alpha g(3)}{1+(1-\alpha)g(3)}), \end{cases}$

$$\circ\ \mathscr{R}_{MAC(\{\mathbf{m}_2,\mathbf{m}_3\})_{t=3}} = \begin{cases} R \leq \log(1+\frac{\alpha g(2)}{1+(1-\alpha)g(2)}) + \log(1+(1-\alpha)g(3)); \\ R \leq \log(1+\alpha g(3)); \\ 2R \leq \log(1+\frac{\alpha g(2)}{1+(1-\alpha)g(2)}) + \log(1+g(3)), \end{cases}$$

$$\circ\ \mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2,\mathbf{m}_3\})_{t=3}} = \begin{cases} R \leq \log(1+g(1)) + \log(1+(1-\alpha)g(2)); \\ R \leq \log(1+\alpha g(2)) + \log(1+(1-\alpha)g(3)); \\ R \leq \log(1+\alpha g(3)); \\ 2R \leq \log(1+g(1)) + \log(1+g(2)) + \log(1+(1-\alpha)g(3)); \\ 2R \leq \log(1+g(1)) + \log(1+(1-\alpha)g(2)) + \log(1+g(3)); \\ 2R \leq \log(1+\frac{\alpha g(2)}{1+(1-\alpha)g(2)}) + \log(1+g(3)); \\ 3R \leq \log(1+g(1)) + \log(1+g(2)) + \log(1+g(3)). \end{cases}$$

Using these MAC regions, we characterize all the possible decoding outcomes at $t = 3$, according to Eq. (2.3) and Eq. (2.4).

- The receiver succeeds to decode $\mathbf{m}_1$, and fails to decode $\mathbf{m}_2$ and $\mathbf{m}_3$, *i.e.* $\mathcal{D}_3 = \{\mathbf{m}_1\}$, if the decoding rate belongs to the rate region:

$$\mathscr{R}_{\{\mathbf{m}_1\}_{t=3}} = \mathscr{R}_{MAC(\{\mathbf{m}_1\})_{t=3}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\})_{t=3}}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_3\})_{t=3}}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\,\mathbf{m}_3\})_{t=3}}}. \tag{2.14}$$

- The receiver succeeds to decode $\mathbf{m}_2$, and fails to decode $\mathbf{m}_1$ and $\mathbf{m}_3$, *i.e.* $\mathcal{D}_3 = \{\mathbf{m}_2\}$, if the decoding rate belongs to the rate region:

$$\mathscr{R}_{\{\mathbf{m}_2\}_{t=3}} = \mathscr{R}_{MAC(\{\mathbf{m}_2\})_{t=3}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\})_{t=3}}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_2,\mathbf{m}_3\})_{t=3}}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\,\mathbf{m}_3\})_{t=3}}}. \tag{2.15}$$

- The receiver succeeds to decode $\mathbf{m}_3$, and fails to decode $\mathbf{m}_1$ and $\mathbf{m}_2$, *i.e.* $\mathcal{D}_3 = \{\mathbf{m}_3\}$, if the decoding rate belongs to the rate region:

$$\mathscr{R}_{\{\mathbf{m}_3\}_{t=3}} = \mathscr{R}_{MAC(\{\mathbf{m}_3\})_{t=3}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_3\})_{t=3}}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_2,\mathbf{m}_3\})_{t=3}}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\,\mathbf{m}_3\})_{t=3}}}. \tag{2.16}$$

- The receiver succeeds to decode $\mathbf{m}_1$ and $\mathbf{m}_2$, and fails to decode $\mathbf{m}_3$, *i.e.* $\mathcal{D}_3 = \{\mathbf{m}_1, \mathbf{m}_2\}$, if the decoding rate belongs to the rate region:

$$\mathscr{R}_{\{\mathbf{m}_1,\mathbf{m}_2\}_{t=3}} = \mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\})_{t=3}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\,\mathbf{m}_3\})_{t=3}}}. \tag{2.17}$$

- The receiver succeeds to decode $\mathbf{m}_1$ and $\mathbf{m}_3$, and fails to decode $\mathbf{m}_2$, *i.e.* $\mathcal{D}_3 = \{\mathbf{m}_1, \mathbf{m}_3\}$, if the decoding rate belongs to the rate region:

$$\mathscr{R}_{\{\mathbf{m}_1,\mathbf{m}_3\}_{t=3}} = \mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_3\})_{t=3}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\,\mathbf{m}_3\})_{t=3}}}. \tag{2.18}$$

- The receiver succeeds to decode $\mathbf{m}_2$ and $\mathbf{m}_3$, and fails to decode $\mathbf{m}_1$, *i.e.* $\mathcal{D}_3 = \{\mathbf{m}_2, \mathbf{m}_3\}$, if the decoding rate belongs to the rate region:

$$\mathscr{R}_{\{\mathbf{m}_2,\mathbf{m}_3\}_{t=3}} = \mathscr{R}_{MAC(\{\mathbf{m}_2,\mathbf{m}_3\})_{t=3}} \bigcap \overline{\mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2\,\mathbf{m}_3\})_{t=3}}}. \tag{2.19}$$

- The receiver succeeds to decode $\mathbf{m}_1$, $\mathbf{m}_2$ and $\mathbf{m}_3$, *i.e.* $\mathcal{D}_3 = \{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$, if the decoding rate belongs to the rate region:

$$\mathscr{R}_{\{\mathbf{m}_1,\mathbf{m}_2,\mathbf{m}_3\}_{t=3}} = \mathscr{R}_{MAC(\{\mathbf{m}_1,\mathbf{m}_2,\mathbf{m}_3\})_{t=3}}. \tag{2.20}$$

Since the rate belongs to the rate region $\mathscr{R}_{\{\mathbf{m}_2,\mathbf{m}_3\}_{t=3}}$ in this example, the receiver responds with $\mathcal{F}_3 = \{2,3\}_A, \{1\}_N$.

In this way, the receiver checks the operating rate region of the system at each time-slot to decide the decoding outcome. Moreover, this characterization of the decoder provides the receiver's performance for capacity-achieving codes.

2.4 Numerical results

We evaluate the performance of the proposed protocol and compare it to conventional parallel SAW HARQ. Hereby, we present numerical results, via computer simulations, of the proposed protocol in comparison to conventional parallel SAW HARQ, both with $C = 4$, $T = 3$ and $R = 0.8$, for capacity-achieving codes. For this purpose, IR-HARQ is implemented as described in Section 2.3.1. The considered path loss model is dependent on the distance between the source and the destination $d = 15u$, where u is a unit of distance. Thus, the variance of the source to destination channel gain is given by:

$$\sigma^2 = \left(\frac{c}{d^2}\right)^2, \tag{2.21}$$

where c is a constant, fixed as $c = 400u^2$. For a fixed SNR, each symbol of the transmit vector $\mathbf{x}_t$ is transmitted with energy E_s. In case of superposition, we remind that the energy is shared between superposed symbols, as in Eq. (2.2). Accordingly, we compare the following protocols:

- IR-HARQ: the conventional parallel SAW HARQ with IR.
- Proposed protocol, $\alpha = 1$: the proposed protocol with $\alpha = 1$ is equivalent to IR-HARQ since all the transmit energy is assigned to layer 1 which acts as the parallel SAW HARQ.
- Proposed protocol, $\alpha = 0.98$: in this configuration of the proposed protocol, only 2% of the transmit energy is assigned to the second layer which includes additional redundancy packets.

- Proposed protocol, $\alpha = 0.8$: in this configuration, 80% of the transmit energy is assigned to the first layer while 20% is assigned to the second layer.
- Proposed protocol, $\alpha = 0.6$: the redundant packets in the second layer are transmitted with 40% of the transmit energy.

In Fig. 2.7, we plot the throughput which is the average number of correctly received information bits per channel use.

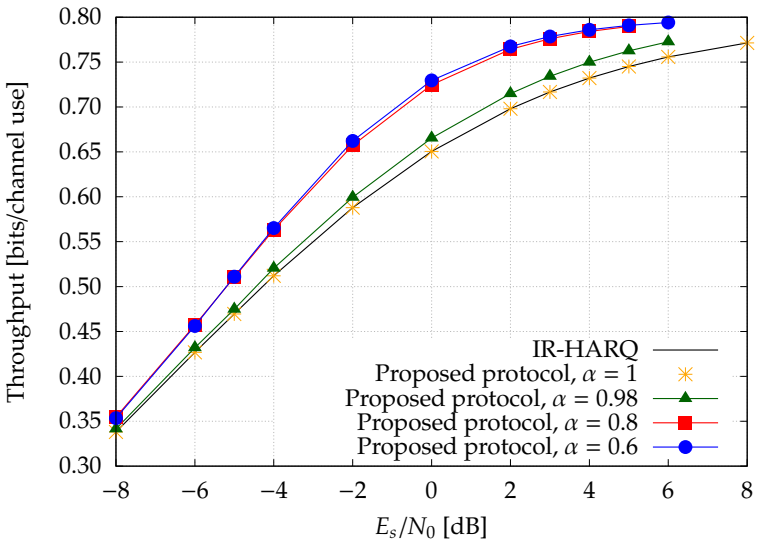


Figure 2.7 – Throughput of the proposed protocol.

The proposed protocol offers a significant throughput gain in comparison to conventional parallel SAW HARQ for any SNR at $\alpha = 0.6$ and $\alpha = 0.8$. More precisely, this gain is between 1dB and 2.5dB at moderate SNR (from −2dB to 8dB). The numerical results show 10% gain in throughput at 0dB when comparing the proposed protocol with $\alpha = 0.6$ to standard IR-HARQ. Also, the throughput converges faster to the coding rate $R = 0.8$ when using the proposed protocol.

In Fig. 2.8, we plot the MER which is defined as the average ratio of the number of dropped messages over the number of sent messages. The numerical results show that the proposed protocol achieves lower MER than the conventional IR-HARQ protocol. Moreover, the diversity gain provided by IR-HARQ with $C = 4$ is 4 considering that each message experiences at most C channel uses, as can be seen in Fig. 2.8. However, the proposed protocol achieves a larger diversity gain which is shown as a steeper slope. This gain is due to multi-layer transmission since the redundant packets of each message can be transmitted in layer 2. More precisely, those messages are transmitted using more than C channel uses.

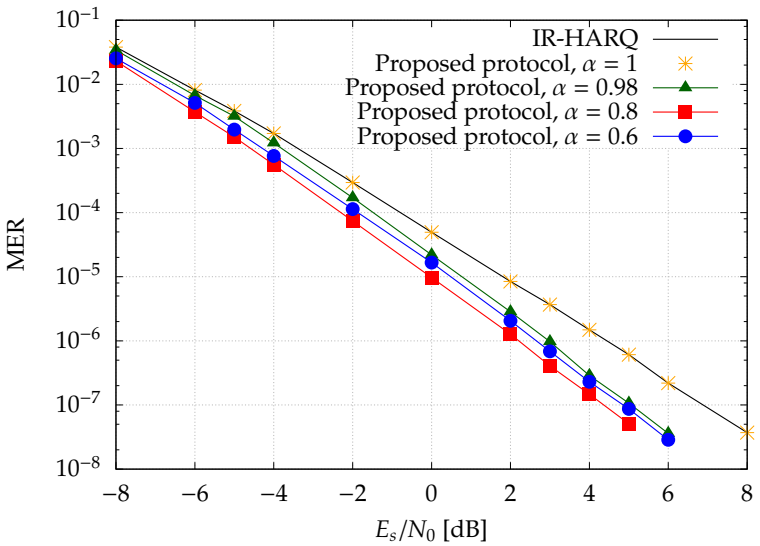


Figure 2.8 – MER of the proposed protocol.

For additional clarification on the diversity gain, check closely the example which was given in Section 2.3.2.1 and illustrated in Fig. 2.3. We remind that Fig. 2.3 shows an instance of the proposed protocol with $C = 3$. In this example, $\mathbf{m}_1$ was sent at time-slots 1, 2, 4 and 7 through the packets $\mathbf{p}_1(1)$ in layer 1, $\mathbf{p}_1(2)$ in layer 2, $\mathbf{p}_1(2)$ in layer 1 and $\mathbf{p}_1(3)$ in layer 1, respectively. Consequently, $\mathbf{m}_1$ is transmitted using 4 channel uses. However, a message sent using a conventional IR-HARQ protocol with $C = 3$ is transmitted using at most 3 channel uses.

The delay, which is the average number of elapsed time-slots until the receiver successfully decodes a message, is presented in Fig. 2.9. Notice that this delay is actually an average delay. Moreover, the maximum delay when using IR-HARQ and the proposed protocol is $(C-1)T+1$ time-slots for both protocols, which is 10 time-slots in the case where $C = 4$ and $T = 3$. This maximum delay is also equal to the delay of SAW HARQ, as explained previously in Section 2.2.1.

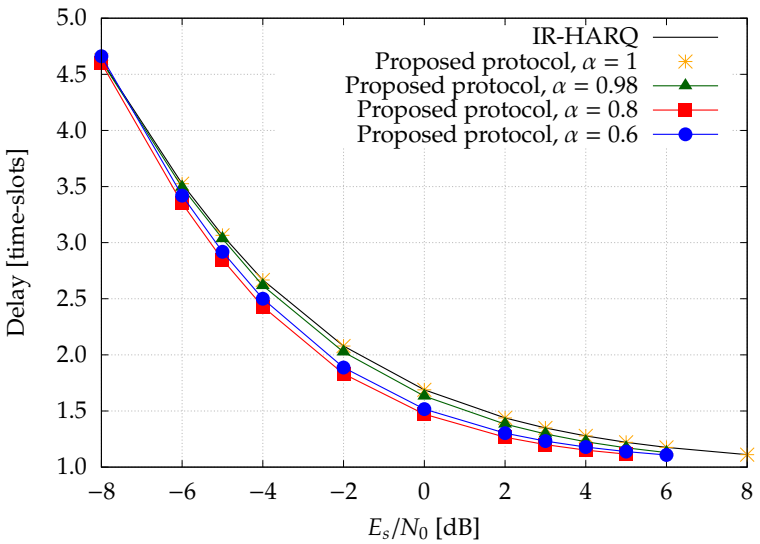


Figure 2.9 – Delay of the proposed protocol.

Moreover, the performance of the proposed protocol depends on the choice of the power fraction α. Therefore, we plot in Fig. 2.10 the throughput and MER, at $E_s/N_0 = 0$dB, versus α. We also plot the delay at $E_s/N_0 = 0$dB versus α in Fig. 2.11.

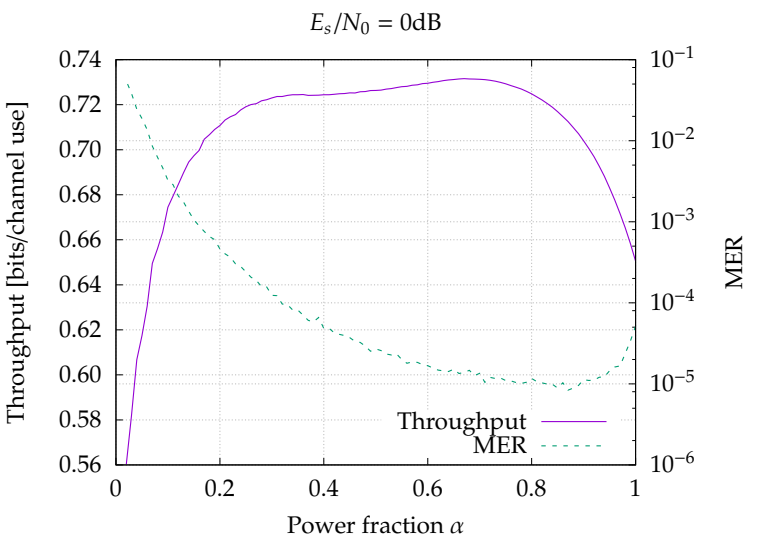


Figure 2.10 – The throughput and MER at 0dB using different power fractions.

The power fraction α can be numerically optimized and fixed for each desired SNR depending on the application requirements. Specific applications might privilege a choice of α that provides the best throughput, MER, delay or a trade-off between these metrics. At 0dB, the highest throughput is achieved at $\alpha = 0.67$ while the lowest MER is achieved

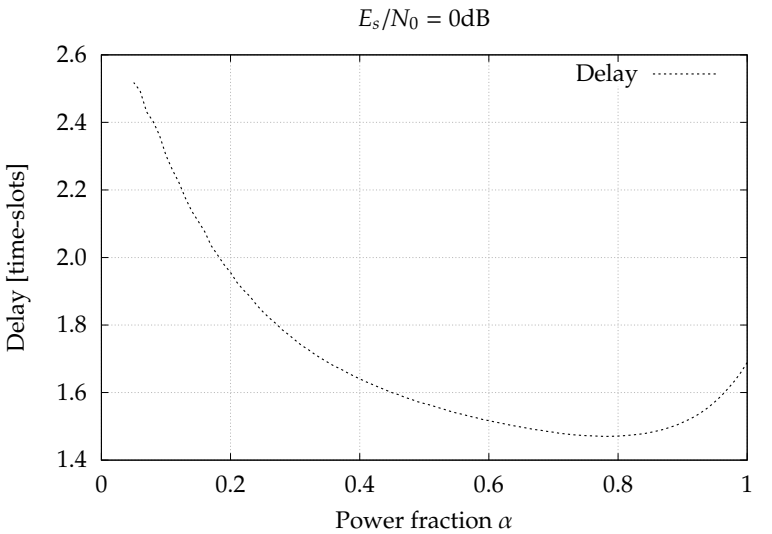
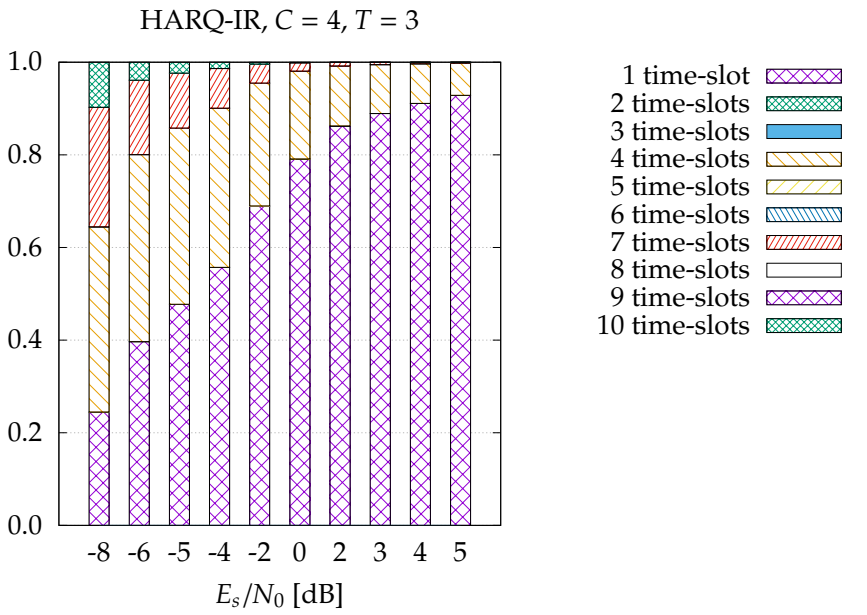


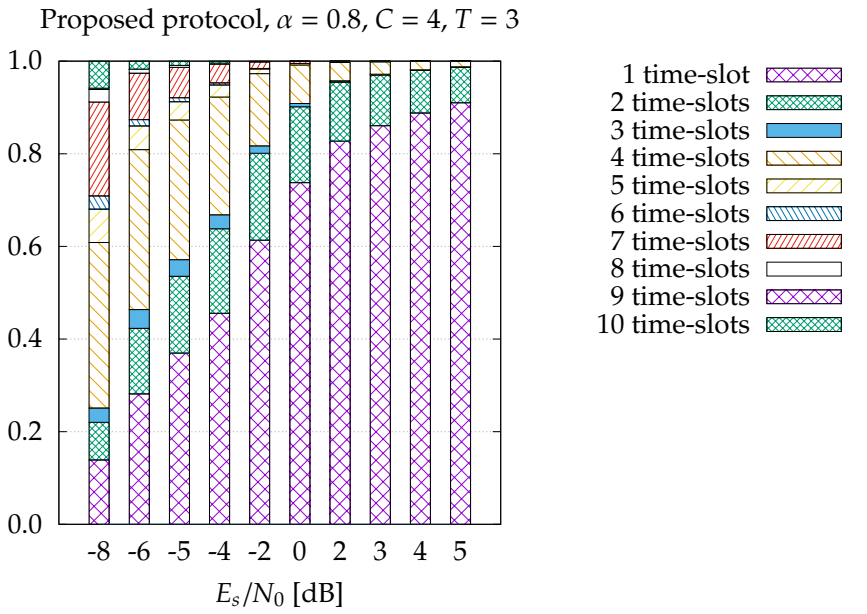
Figure 2.11 – The delay of the proposed protocol at 0dB using different power fractions.

at $\alpha = 0.87$, and the lowest delay is achieved at $\alpha = 0.78$. Moreover, $\alpha = 0.7$ is a trade-off between the performance metrics at 0dB that provides a high throughput with small MER and small delay.

In addition to small delay, the proposed protocol offers a better delay distribution. The delay distribution represents the proportion of successfully delivered messages for each value of the delay. Fig. 2.12 compares the delay distribution using IR-HARQ and using the proposed protocol, with $\alpha = 0.8$. Using IR-HARQ, a message can be decoded only with a delay of 1 (by decoding the first packet), 4 (by decoding the first retransmission), 7 (by decoding the second retransmission) or 10 time-slots (by decoding the last retransmission) for $C = 4$. Thus, the retransmissions occur every $T = 3$ time-slots. In contrast, due to superposition in the proposed protocol, the receiver can decode a message with a finer granularity of delays, and delays of 1, 2,. . . ,10 time-slots are achievable, as it can be seen in Fig. 2.12b. In addition, the transmissions with large delays, such as 4 or more as shown in Fig. 2.13, are significantly less frequent when using the proposed protocol.

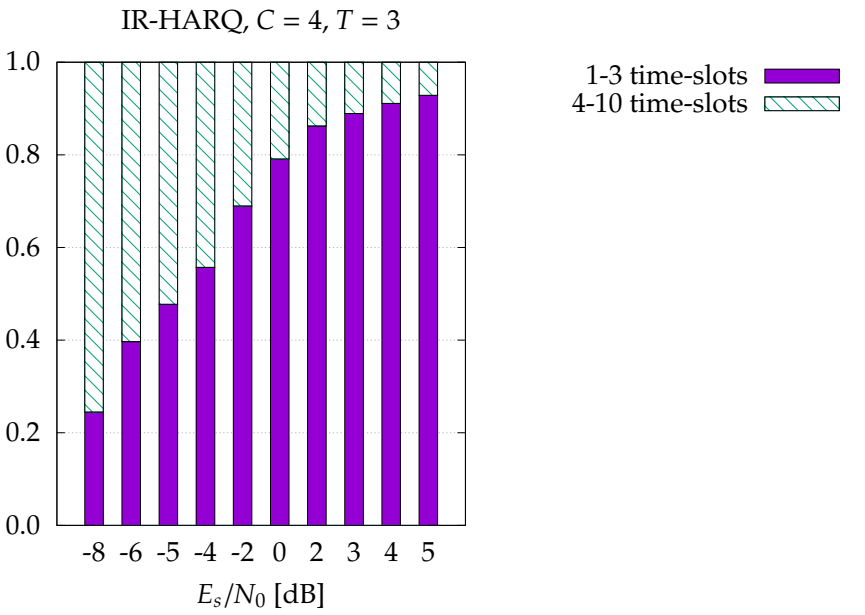


(a) Conventional parallel SAW HARQ.

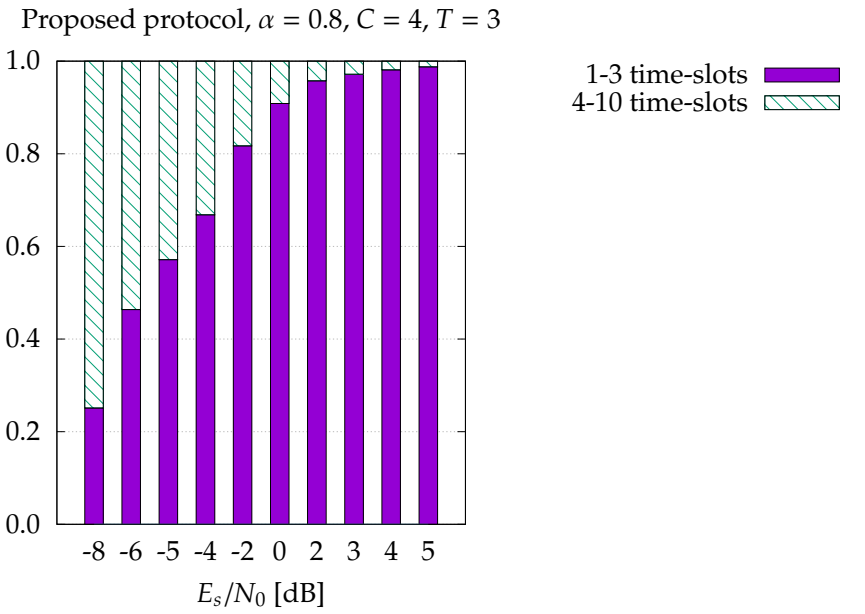


(b) Proposed protocol.

Figure 2.12 – Comparison of the delay distributions.



(a) Conventional parallel SAW HARQ.



(b) Proposed protocol.

Figure 2.13 – Comparison of the proportion of large delays.

2.5 Conclusion

This Chapter presented an enhanced version of IR-HARQ using multi-packet transmission. This proposed protocol superposes an additional layer of redundant packets to the parallel SAW HARQ protocol, according to a parameter α. In this superposed layer, the transmitter may perform retransmissions of a message even before having received any feedback about it. This preemptive selection of the superposed packets does not require any additional feedback. The proposed protocol uses multi-layer transmission, where non-orthogonal layers are superposed by the transmitter and decoded by the receiver. In other words, the receiver attempts to decode the superposed packets in the proposed protocol. Moreover, an information-theoretic characterization of the receiver is provided. The corresponding numerical results show that the proposed protocol offers a smaller average delay, a better delay distribution, a higher throughput, and a lower MER than the conventional parallel SAW HARQ. Part of this work has been published in [C2]. Possible improvements of this proposed protocol are suggested in the conclusion Chapter.

Conclusions and Perspectives

In this thesis, we improved the HARQ mechanisms - by using superposition coding - in terms of throughput and/or latency which are the bottleneck of next generation wireless communication systems. Consequently, the proposed contributions in this thesis can be employed in:

- systems enabling the use of relaying. In this case, the proposed protocol in Chapter 1 is very useful.
- any system with delayed HARQ feedback, which eventually corresponds to any practical system.

In Chapter 1, we started by reviewing the state of the art HARQ protocols. The basic types of HARQ protocols were presented, and different relay assisted HARQ protocols were discussed. This Chapter included the definition of the performance metrics and the theoretical framework to study HARQ, based on Markov chain models and information theoretic tools. Then, we proposed a new protocol and derived its performance metrics using this framework. Afterwards, the numerical results showed that the proposed protocol outperforms the other HARQ with capacity-achieving codes and with practical codes followed by an interference canceler. According to these numerical results, the investigated protocol achieved the best overall performance of the system.

In Chapter 2, we proposed an enhanced HARQ protocol adapted to delayed feedback. First, we provided an overview of the HARQ protocols in current communication systems such as LTE, where parallel HARQ processes are initiated to deal with the feedback delay of several time-slots. Then, we discussed non-orthogonal transmission and reviewed the state of the art where HARQ is combined with non-orthogonal transmission. To the best of our knowledge, our proposed protocol is the first protocol that combines HARQ, non-orthogonal transmission and preemptive scheduling of the packets. Moreover, our proposed protocol achieved simultaneously a larger throughput, lower latency and higher reliability than parallel HARQ, at the cost of increased decoding complexity. This claim has been validated numerically using capacity-achieving codes.

Both proposed protocols relied on non-orthogonal transmission, also known as superposition coding. Moreover, in Chapter 1, the superposition is due to the interference between the incoming packets from the source and the relay, whereas, in Chapter 2, the packets are superposed at the source and transmitted through the same channel. Our work shows that combining non-orthogonal transmission with HARQ improves significantly the performance of wireless communication systems.

Perspectives

We suggest for future works to extend the protocol of Chapter 1 to the case where multiple sources transmit messages to multiple destinations in the presence of relays. This extension is not straightforward since multiple choices are possible and the interference due to superposition coding has to be managed efficiently. Moreover, an end-to-end approach (with multi-hop) can be considered. In this case, the question is: how handle simultaneously HARQ and superposition coding (which is in this context strongly related to compute and forward) in a whole network.

On the other hand, the practical implementations of this protocol, as discussed in Chapter 1, could benefit from the search for adequate codes to non-orthogonal transmission. These channel codes could enable joint decoding of the messages at the receiver, which results in performance gains in comparison to SIC. Then, the performance of these codes should be compared to theoretical capacity-achieving codes, and the decoding complexity of these codes should be analyzed. Considering the significant performance gain that is provided using this HARQ protocol, the proposed extension could have a significant impact on next generation relay assisted wireless communication systems.

The main issue that we encountered when studying the protocol in Chapter 1 is the scalability of the theoretical analysis. Indeed, the derived expressions of the performance metrics are very complicated and do not provide additional insights on how to optimize the performance of the system. On the other hand, the proposed Markov chain model cannot be extended to large number of HARQ retransmissions. Therefore, we do not advocate to work this way. This explains why we did not spend time deriving the performance metrics related to the proposed protocol in Chapter 2.

We consider that the proposed HARQ protocol in Chapter 2 is of great interest for future wireless communication systems. Indeed, this proposed enhanced HARQ protocol does not require any additional signaling nor modification in the infrastructure of current wireless communication systems. Therefore, a lot of additional research on this protocol is required. First, we should implement the protocol with practical channel codes and decoders. One could consider using SIC for decoding. However, adequate channel codes and decoders that perform joint decoding (and possibly joint encoding) of the messages could approach the performance of capacity-achieving codes. Second, the choice of the

redundant packets in layer 2 in our work was rational and led to significant improvement, but this choice is not necessarily the best one for all application requirements. We could increase the number of layers at the expense of the decoding complexity. We could also find benchmarks to compare the performance of the proposed protocol and future ones to the best achievable performance, which depends on the true value of the feedback delay. From a conceptual perspective, one can remark that by using the presence of the feedback delay (which always occurs in practical systems), we manage to improve the performance and exploit an additional degree of freedom. This means that working with the simplified model without taking into account the feedback delay is restrictive and degrades the performance. A future research axis is to understand better the reason behind this gain by characterizing this additional degree of freedom.

Bibliography

S. M. Alamouti, "A simple transmit diversity technique for wireless communications," **IEEE Journal on Selected Areas in Communications**, vol. 16, no. 8, pp. 1451–1458, Oct. 1998. Cited page 17

M. E. Aoun, R. L. Bidan, X. Lagrange, and R. Pyndiah, "Multiple-packet versus single-packet incremental redundancy strategies for type-ii hybrid arq," in **2010 6th International Symposium on Turbo Codes Iterative Information Processing**, Sep. 2010, pp. 226–230. Cited page 77

M. E. Aoun, X. Lagrange, R. L. Bidan, and R. Pyndiah, "Analysis and optimization of hybrid single packet and multiple-packets incremental redundancy in the presence of channel state information," in **2011 The 14th International Symposium on Wireless Personal Multimedia Communications (WPMC)**, Oct. 2011, pp. 1–5. Cited page 77

A. N. Assimi, C. Poulliat, and I. Fijalkow, "Packet combining for multi-layer hybrid-arq over frequency-selective fading channels," in **2009 17th European Signal Processing Conference**, Aug. 2009, pp. 671–675. Cited pages 76 and 77

L. Badia, M. Levorato, and M. Zorzi, "Markov analysis of selective repeat type ii hybrid arq using block codes," **IEEE Transactions on Communications**, vol. 56, no. 9, pp. 1434–1441, Sep. 2008. Cited page 12

B. Bandemer, A. E. Gamal, and Y. H. Kim, "Simultaneous nonunique decoding is rate-optimal," in **2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)**, Oct. 2012, pp. 9–16. Cited pages 20, 21, 77, 87, and 88

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, Jul. 2001. Cited pages 12, 13, and 15

T. V. K. Chaitanya and E. G. Larsson, "Superposition modulation-based symmetric relaying with hybrid arq: Analysis and optimization," **IEEE Transactions on Vehicular Technology**, vol. 60, no. 8, pp. 3667–3683, Oct. 2011. Cited page 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, May 1985. Cited page 8

A. Chelli and M. S. Alouini, "On the performance of hybrid-arq with incremental redundancy and with code combining over relay channels," **IEEE Transactions on Wireless Communications**, vol. 12, no. 8, pp. 3860–3871, Aug. 2013. Cited pages 10, 14, and 17

H. Chen, Y. Cai, W. Yang, and Y. Hu, "A novel markov-chain-based method for throughput analysis in truncated cooperative harq systems," in **2013 International Conference on Wireless Communications and Signal Processing**, Oct. 2013, pp. 1–5. Cited page 12

T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," **IEEE Transactions on Information Theory**, vol. 25, no. 5, pp. 572–584, Sep. 1979. Cited page 9

T. Guan and L. Chen, "Opportunistic nonorthogonal cooperative communications through decode-and-forward relaying," in **2013 IEEE/CIC International Conference on Communications in China (ICCC)**, Aug. 2013, pp. 478–483. Cited pages 9 and 10

J. Hagenauer, "Rate-compatible punctured convolutional codes (rcpc codes) and their applications," **IEEE Transactions on Communications**, vol. 36, no. 4, pp. 389–400, Apr. 1988. Cited pages 8 and 63

A. E. Hamss, L. Szczecinski, and P. Piantanida, "Increasing the throughput of harq via multi-packet transmission," in **2014 IEEE Global Communications Conference**, Dec. 2014, pp. 1485–1491. Cited page 78

S. K. Hong and J. M. Chung, "Network-coding-based hybrid arq scheme for mobile relay networks," **Electronics Letters**, vol. 46, no. 7, pp. 539–541, Apr. 2010. Cited page 10

M. Jabi, A. E. Hamss, L. Szczecinski, and P. Piantanida, "Multipacket hybrid arq: Closing gap to the ergodic capacity," **IEEE Transactions on Communications**, vol. 63, no. 12, pp. 5191–5205, Dec. 2015. Cited page 78

J. Kim, W. Hur, A. Ramamoorthy, and S. W. McLaughlin, "Design of rate compatible irregular LDPC codes for incremental redundancy Hybrid ARQ systems," in **Information Theory (ISIT), 2006 International Symposium on**. Seattle, USA: IEEE, Jul. 2006. Cited page 8

G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," **IEEE Transactions on Information Theory**, vol. 51, no. 9, pp. 3037–3063, Sep. 2005. Cited page 9

J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," **IEEE Transactions on Information Theory**, vol. 50, no. 12, pp. 3062–3080, Dec. 2004. Cited page 9

E. G. Larsson and B. R. Vojcic, "Cooperative transmit diversity based on superposition modulation," **IEEE Communications Letters**, vol. 9, no. 9, pp. 778–780, Sep. 2005. Cited page 10

S. Lin and D. J. Costello, **Error Control Coding, Second Edition**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2004. Cited pages 71 and 72

S. Lin, D. J. Costello, and M. J. Miller, "Automatic-repeat-request error-control schemes," **IEEE Communications Magazine**, vol. 22, no. 12, pp. 5–17, Dec. 1984. Cited pages 7 and 11

Y. Ma, L. Li, J. Jin, and Y. Liu, "A novel network coded relay-assisted hybrid-arq scheme," in **2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)**, Jan. 2013, pp. 455–459. Cited pages 10, 11, and 64

D. Mandelbaum, "An adaptive-feedback coding scheme using incremental redundancy," **IEEE Transactions on Information Theory**, vol. 20, no. 3, pp. 388–389, May 1974. Cited page 8

R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: performance limits and space-time signal design," **IEEE Journal on Selected Areas in Communications**, vol. 22, no. 6, pp. 1099–1109, Aug. 2004. Cited page 10

B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference with structured codes," in **2008 IEEE International Symposium on Information Theory**, Jul. 2008, pp. 772–776. Cited page 9

N. Prasad and M. K. Varanasi, "Diversity and multiplexing tradeoff bounds for cooperative diversity protocols," in **International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.**, Jun. 2004, p. 268. Cited page 10

J. Proakis and M. Salehi, **Digital Communications**, 5th ed. McGraw-Hill, 2007. Cited page 47

R. Samano-Robles, D. McLernon, and M. Ghogho, "A random access protocol incorporating multi-packet reception, retransmission diversity and successive interference cancellation," in **Multiple Access Communications**. Springer International Publishing, Sep. 2015, pp. 70–86. Cited page 11

S. Sesia, I. Toufik, and M. Baker, **LTE, The UMTS Long Term Evolution: From Theory to Practice**. Wiley Publishing, 2009. Cited pages 71, 72, and 74

A. Steiner and S. Shamai, "Multi-layer broadcasting hybrid-arq strategies for block fading channels," **IEEE Transactions on Wireless Communications**, vol. 7, no. 7, pp. 2640–2650, Jul. 2008. Cited page 76

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, Jun. 2013. Cited page 78

F. Takahashi and K. Higuchi, "Harq for predetermined-rate multicast channel," in **2010 IEEE 71st Vehicular Technology Conference**, May 2010, pp. 1–5. Cited page 77

K. F. Trillingsgaard and P. Popovski, "Generalized harq protocols with delayed channel state information and average latency constraints," **IEEE Transactions on Information Theory**, vol. 64, no. 2, pp. 1262–1280, Feb. 2018. Cited page 78

H. C. A. Van Duuren, "Printing telegraph systems," U.S. Patent 2313980, Mar. 1943. Cited page 6

S. Yang and J. C. Belfiore, "Towards the optimal amplify-and-forward cooperative diversity scheme," **IEEE Transactions on Information Theory**, vol. 53, no. 9, pp. 3114–3126, Sep. 2007. Cited pages 9 and 10

F. Yilmaz and M. S. Alouini, "Outage capacity of multicarrier systems," in **2010 17th International Conference on Telecommunications**, Apr. 2010, pp. 260–265. Cited page 27

R. Zhang and L. Hanzo, "Superposition-coding-aided multiplexed hybrid arq scheme for improved end-to-end transmission efficiency," **IEEE Transactions on Vehicular Technology**, vol. 58, no. 8, pp. 4681–4686, Oct. 2009. Cited page 76

Z. Zhang, H. Yang, and D. Liyun, "A new relay assisted hybrid arq scheme," in **2010 International Conference on Wireless Communications Signal Processing (WCSP)**, Oct. 2010, pp. 1–4. Cited pages 10 and 64

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Optimisation inter-couches de schémas de coordination et de coopération pour les futurs réseaux cellulaires

Mots clés: 5G, allocation de ressources, communication coopérative, HARQ, schéma de retransmission

Résumé: Les demandes de haut débit, faible latence et grande fiabilité augmentent dans les nouvelles générations de systèmes de radiocommunications. Par conséquent, on propose de combiner la transmission non orthogonale avec les retransmissions HARQ afin de combattre les fluctuations de canal de transmission à haut débit. Dans la première partie de la thèse, on propose des protocoles de retransmissions HARQ avec l'aide d'un relais afin d'améliorer le débit et la fiabilité du système. Une nouvelle version du protocole HARQ qui prend en compte le délai de retour est proposée dans la seconde partie de la thèse.

Title: Cross-layer optimization of cooperative and coordinative schemes for next generation cellular networks

Keywords: 5G, resource allocation, cooperative communication, HARQ, retransmission scheme

Abstract: HARQ has become an important research field in the wireless digital communications area during the last years. In this thesis, we improve the HARQ mechanisms in terms of throughput and/or latency which are the bottleneck of next generation wireless communication systems. More precisely, we improve the time-slotted HARQ systems by mimicking NOMA, which means using superposed packets in a single-user context. In the first part of the thesis, we propose HARQ protocols using the help of a relay to improve the transmission rate and reliability. An enhanced HARQ protocol adapted to delayed feedback is proposed in the second part. In this new multi-layer HARQ protocol, additional redundant packets are sent preemptively before receiving the acknowledgement, and in superposition to other HARQ processes.

Université Paris-Saclay
Télécom ParisTech
46 Rue Barrault, 75013 Paris, France