Two-way Data Exchange in Satellite Communication

A Dissertation

Presented to

the Faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment

of the requirements for the Degree

Doctor of Philosophy (Computer Engineering)

by

Chenguang Xu

June 2013

Approval Sheet

This dissertation is submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Computer Engineering)

_____

Chenguang Xu

This dissertation has been read and approved by the Examining Committee:

_____

Stephen G. Wilson, Advisor

_____

Toby Berger, Committee Chair

_____

Malathi Veeraraghavan

_____

Alfred C. Weaver

_____

Alfredo Garcia

Accepted for the School of Engineering and Applied Science:

_____

James H. Aylor, Dean, School of Engineering and Applied Science

June 2013

Acknowledgments

I am indebted to my advisor, Professor Stephen.G Wilson, for his support, encouragement, and invaluable advice during the course of my Ph.D. study. His patience and kindness helped me through a difficult time.

I would like to thank Professors Alfred C. Weaver, Alfredo Garcia, Kamin Whitehouse, Maïté Brandt-Pearce, Malathi Veeraraghavan, and Toby Berger for serving on my proposal committee or dissertation defense committee.

Thanks to my collaborators Tingjun and John, whose suggestions and ideas benefited me a lot.

This thesis is dedicated to my parents, who gave me all they have in the past 31 years. I owe them. This thesis is also dedicated to Jessie, who brings me the happiness that I have never experienced before I met her.

Abstract

We focus on data exchange between two terminals, with primary motivations from satellite communication. The abstract communication model of the problem is a special case of the two-way relay channel (TWRC). The main work of this thesis consists of two parts, corresponding to two different channel settings: linear channel and nonlinear channel.

The first part considers a *linear* satellite channel, i.e. we assume a linear transponder. We first summarize previously proposed protocols: AF (non-orthogonal amplify and forward), MAC-NC (multiple access channel with nested coding), MAC-BSI (multiple access followed by broadcast channel with side information), MAC-XOR (multiple access followed by relay Xor), and NLC (nested lattice coding), and compare their performance in terms of achievable rate regions, maximum sum rates in both symmetric and asymmetric scenarios, and required system complexities. Within these protocols, AF is flexible on signal synchronization, has minimum code knowledge requirements on each terminal, and is compatible with legacy satellites. In addition, it can achieve good performance for all scenarios considered, so we fix on the AF protocol for later study in this thesis. Then, we consider the AF protocol from a different view that given information rate requirements and resource constraints (transponder bandwidth and power) , how to minimize the satellite transponder cost for the AF protocol over a two-way relay satellite channel? A solution to this problem is presented that determines the most efficient bandwidth and power resources requests to reduce the transponder cost, which also leads directly to proper choice for modulation and coding. Finally, we work on the extension to a three-terminal hub-remote data exchange. Two application scenarios including three information flows and four information flows are studied. By analyzing and comparing between different decoding schemes for each case, we try to fully exploit the advantage of non-orthogonal signalling.

The second part considers a *nonlinear* satellite channel (NGC). We first consider the uncoded transmission with the AF protocol over NGC. We find that large input backoff at the amplifier and incorporation of memory in the receiver are essential to the performance of symbol detection for the AF protocol. Then we study the achievable information rates for NGC. We consider both single-user and two-user (AF) cases. These studies can tell us what is the maximum allowed information rate or minimum required SNR for a reliable transmission over NGC, and help us predict the best possible performance of channel coding. Finally, we study the channel coding performance over NGC in the context of DVB-S2 modulation and coding. Both single-user and two-user (AF) scenarios are studied. For the single-user case, adjusting input backoff and ring ratio of 16APSK can significantly improve the decoding performance, while for the two-user case, a much larger backoff is preferred (-6 dB). In addition, the proposed memory-2 decoder works well, with only a tiny performance gap between it and the ideal memory-2 decoder.

Contents

List of Figures

List of Tables

List of Abbreviations

AF   (Non-orthogonal) Amplify and Forward

ARR   Achievable Rate Region

CF   Compress-forward

DC/RF   Direct Current / Radio Frequency

DVB-S2   Digital Video Broadcasting Satellite Second Generation

FMDA   Frequency Division Multiple Access

FSM   Finite State Machine

HDTV   High-definition Television

HPA   High Power Amplifier

IBO   Input Backoff

ISI   Intersymbol Interference

LDPC   Low-density Parity-check

LLR   Log-likelihood Ratio

LMS   Least Mean Squares

MAC-BSI   Multiple Access followed by Broadcast Channel with Side Information

MAC-NC   Multiple Access Channel with Nested Coding

MAC-XOR   Multiple Access followed by Relay XOR

MINLP   Mixed Integer Nonlinear Programming

MLSD   Maximum Likelihood Sequence Detection

NGC   Nonlinear Gaussian Channel

NLC   Nested Lattice Coding

NRZ   Non Return to Zero

RRC   Root-raised Cosine

SDTV   Standard-definition Television

TDMA   Time Division Multiple Access

TWRC   Two-way Relay Channel

TWTA   Traveling Wave Tube Amplifier

VSAT   Very Small Aperture Terminal

Chapter 1

Introduction

1.1 Overview

Satellites provide an essential tool for linking distributed commercial and military locations. Though limited by its long latency, low capacity, and high cost, satellite-based communication has its advantage in applications that require quick set-up or support remote regions or wide area coverage. Such application scenarios include broadcasting of standard definition and high-definition TV (SDTV and HDTV), interactive services, including Internet access, content distribution, Internet trunking, and so on [1].

We consider data exchange between two terminals via a relay terminal, with primary interest coming from satellite communication. As shown in Figure 1.1, two earth terminals T_1 and T_2 want to exchange data via satellite R, and h_{1r}, h_{2r}, h_{r1}, h_{r2} are effective complex channel gains for each link. Generally, multiple terminals can be involved in such a network, but our focus is mainly on the two-terminal problem, which is easier and can be extended to multiple terminals on a time-shared basis. Normally the bidirectional traffic flow is frequency-duplexed at the satellite, that is, uplinks to the satellite operate in one microwave band, e.g. 14 GHz, and downlinks to earth terminals operate in a disjoint band, e.g. 12 GHz. The full bands are normally segmented into transponder channels, whose bandwidth is typically on the order of 40 MHz.

T_1 and T_2 are assumed to be within the same footprint of the satellite, but not able to receive each other's signal directly, so the abstract communication model is actually a special case of a two-way relay channel, whose origins date back to Shannon's two-way channel in 1961 [2]. Traditional
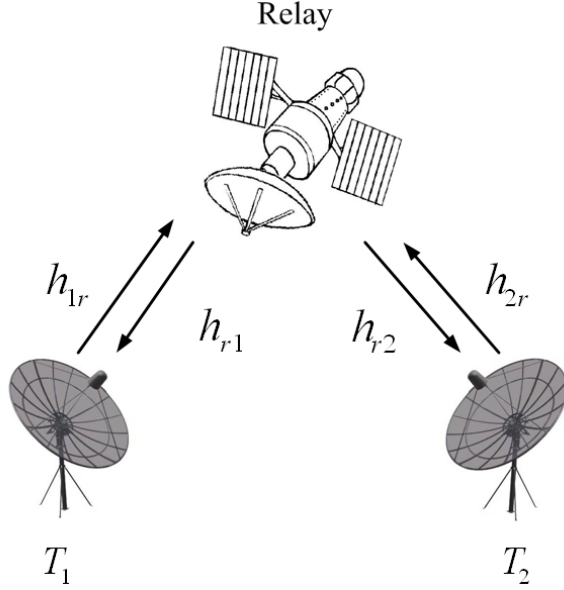
Figure 1.1: Two-terminal satellite network

solutions adopt orthogonal protocols, in which communications for T_1 and T_2 are operated in disjoint frequency bands (FDMA) or different time slots (TDMA). However, research in information theory has shown that based on the side information, non-orthogonal signalling protocols can achieve up to 100% spectrum efficiency improvement as compared with these orthogonal solutions, which constitutes the main motivation for this thesis.

1.2   Related work

1.2.1   Two-way relay protocols over linear Gaussian channel

This general problem is variously called the two-way relay channel (TWRC) [3], the separated TWRC [4], emphasizing no direct path between T_1 and T_2, or the exchange channel, [5], all of which assume linear channels with additive Gaussian noise. This rather simple model has received significant attention in the recent years from the communication theory community, mostly directed at terrestrial wireless networking where relay nodes can mitigate poor propagation or extend network size. Previously proposed efficient protocols include: decoding-based protocols,the non-orthogonal amplify-forward (AF) protocol, and the compress-forward (CF) protocol, all of which adopt non-orthogonal signalling.

In decoding-based protocols, the relay first decodes both T_1 and T_2's signals, and then there are a few options at the relay.

- Broadcast the two decoded signal via a broadcast channel [6, 7] using some superposition coding strategy [8].

- XOR the two decoded messages bit-by-bit, and broadcast this XORed *common* message to both T_1 and T_2. This protocol is often called network coding, which has received significant interest in the past few years [9–15].

- Use nested coding [16–18] at the relay, so that the information rate is no longer bottlenecked by the poorer of the two downlinks.

In the AF protocol, the relay simply amplifies and broadcasts the received signal, which is a sum of two signals from T_1 and T_2. T_1 and T_2 can eliminate the signal component they have sent respectively [19–23], and decode the desired signal component. AF is flexible on signal synchronization, has minimum code knowledge requirements on each terminal, and is compatible with legacy satellites. [19, 21, 23] consider a more general scenario when multiple relays are available, while [20, 22] focus on practical issues of interference cancellation.

If the CF protocol, instead of decoding T_1 and T_2's signals separately, the relay generates a *common* message directly from the received signal [5, 24–29]. For all channel parameters, it has been shown in [27, 28] that the achievable rate region is within $\frac{1}{2}$ bit from the capacity region for each user. However, the CF protocol has some unrealistic assumptions regarding synchronization, so we will not give further attention to the CF protocol in this thesis.

A dedicated comparison between these protocols will be presented in Chapter 3.

1.2.2 Nonlinear Gaussian channel (NGC)

Satellite channels are well-known to exhibit nonlinear behavior due to a saturating high power amplifier (HPA). The near-saturation region is the preferred region of operation since more RF power is available at the output and the DC/RF efficiency is highest. Of course for non-constant envelope carriers, or multisignal cases, nonlinear impairments may degrade performance severely. The Saleh model [30] is widely adopted to describe the instantaneous response of a nonlinear amplifier. However, in a practical system, memory effects induced by bandpass filtering and transmitter/receiver

pulse shaping may be significant, and the general Volterra series is used to model nonlinear systems with memory [31, 32].

Single-user transmission over a nonlinear channel has been widely studied for uncoded transmission. Solutions proposed to reduce the nonlinear impact mainly include linear/nonlinear equalization (see [33]) and predistortion [34–38]. But to our knowledge relatively little work exists on the information-theoretic potential of such channels, i.e. what performance is approachable if advanced channel coding and decoding is allowed? [39] indicates that under average and peak power constraints, the capacity-achieving output distribution is discrete and possesses a finite number of mass points. Other related work in this area can be found in [40–42].

For the two-way relay nonlinear channel, [43–47] try to improve the decoding performance. But these methods either assume the channel is memoryless or require the nonlinear function to be invertible, both of which are unrealistic. As far as we know there is no related work found on the two-way relay nonlinear channel from an information-theoretic perspective.

1.3 Main work and organization of this thesis

We will present the general system model in Chapter 2. Chapters 3-9 contain the main work of this thesis, which is divided into two parts: communication over linear channels and communication over nonlinear channels.

1.3.1 Two-way relay linear channel: Chapter 3-5

Chapter 3: Codification of existing protocols

Protocols applicable to bidirectional data transfer via satellite between two earth terminals are analyzed with respect to their achievable rate regions and maximum information rate sum. Protocols considered include amplify-forward (AF), requiring no special satellite processing, and several variations of decode-forward, which necessitate satellite demodulation, decoding, and re-encoding. All these protocols seek to increase the spectral efficiency by as much as 100% over conventional two-way communication. AF and a protocol called MAC-nested coding (MAC-NC) emerge as preferred candidates.

In addition to comparing the basic protocol efficiency on transfer rate limits given certain power and bandwidth resources, the protocols are characterized in terms of complexity and operational feasibility. AF is flexible on signal synchronization, has minimum code knowledge requirements on each terminal, and is compatible with legacy satellites. In addition, AF achieves good performance in all scenarios considered, so it is our main focus of this thesis. Practical issues like synchronization and channel estimation for AF protocol will also be presented in Chapter 3.

Chapter 4: Minimization of satellite transponder operating cost

The AF protocol turns out be a practical and efficient protocol. Previous studies of the AF protocol's achievable rate region (ARR) assume that the full resources (bandwidth and power) of the satellite relay are utilized. Here we take a different view, where two terminals wish to communicate with some target rates that may be well less than what the full transponder resources can support. Then, the system operator would wish to lease a fraction of a transponder, and share the remaining resources with other non-interfering services. The recurring cost for transponder services reflects the consumed resources, and it is of practical interest to minimize the recurring cost. Assuming a generic pricing structure, we model it as a joint bandwidth and power optimization problem. A solution to this problem is given in this paper that determines the most efficient bandwidth and power resources requests to reduce the transponder cost. This Chapter also considers a more general problem when multiple traffic flows are to be supported, perhaps across multiple transponders. This leads to a mixed integer non-linear programming (MINLP) problem. (The work in Chapter 4 has been presented in *ICC 2012*.)

Chapter 5: Extension to scenarios when multiple terminals are involved

We study the performance of AF protocols in a three-terminal hub-remote satellite exchange channel, which is an extension of the two-way relay channel. We want to answer two questions: 1) Can non-orthogonal signaling and partial interference cancellation bring as much improvement as what have seen in the two-way relay channel? 2) Which decoding method achieves the largest achievable rate region? Theoretical analysis and numerical results show that for the three-flow case, repetition of sending common forward message should be avoided, and at least one of the decoding schemes of three-user simultaneous signalling performs the best among all scenarios considered. For the four-

flow case, three-user simultaneous signalling only shows advantage when the hub terminal has much higher SNR than remote terminals or when remote terminal rate requirements are emphasized. In addition, the gains from three-simultaneous signalling are generally smaller than one might expect based on multi-terminal information theory, due to the constraint on the relay's power and power sharing. (The work in Chapter 5 has been presented in *ICC 2013.*)

1.3.2   Nonlinear satellite channel: Chapter 6-9

Chapter 6: Uncoded transmission of AF over nonlinear channel

Satellite communication is well known to suffer from nonlinear distortion, so one important question is how will AF perform over such a nonlinear channel. As a first step, we will focus on the uncoded transmission of AF protocol over a nonlinear satellite channel in Chapter 6, and we want to evaluate the "real" system performance in terms of symbol error ratio versus SNR. The actual channel has memory due to a combined effect of nonlinearity and pulse shaping. A memoryless detection method will be presented which is based on a finite-state channel model. However, there is severe error floor for such memoryless detector even at a large input backoff (IBO=-8 dB). Then we propose a memory-2 symbol detection scheme, which is based on the Viterbi algorithm. Simulation results show that the performance of symbol detection can be greatly improved, and for IBO=-8 dB , there is no visible error floor within the range of interest for an uncoded transmission. As we will show in subsequent chapters, coding allows operation closer to saturation.

Chapter 7: Achievable information rate over a nonlinear channel

With an eye toward coded transmission of high-level modulation on satellite links, we study the information-theoretic potential of nonlinear satellite channels, for both single-user per transponder operation and simultaneous two-way, on-frequency relaying to improve spectrum efficiency as much as 100%. Coded 16-APSK transmission is studied from the perspective of achievable information rate and channel simulation, in the context of DVB-S2 over a nonlinear satellite channel. We develop an analytic nonlinear model with finite memory for these two scenarios, and compute achievable rate regions, i.e. information rates that can be arbitrarily reliable, as a function of link SNR's and input backoff (IBO) of the nonlinear amplifier. This provides guidance on required

SNR and desired amplifier operating point for advanced coded modulation schemes. (The work in Chapter 7 has been presented in *LATINCOM 2012*.)

Chapter 8: Improving the performance of single-user transmission over nonlinear satellite channel

Studies on achievable rates in Chapter 7 provide insight into nonlinear satellite channel from information-theoretic perspective. A real satellite system adopts channel coding to obtain reliable communication between a transmitter and receiver. In Chapter 8, we will evaluate single-user nonlinear channel in the context of DVB-S2, which adopts LDPC codes as the channel coding method. We will show that modification to the constellation ring ratio for 16APSK can provide significant savings in SNR, at least for the adopted nonlinear Saleh model. In addition, a two-pass decoding that employs hard-decision feedback from the LDPC decoder is able to gain 0.4 dB in performance without significant complexity increase. Simulation results of Chapter 8 are well predicted by our achievable rate analysis in Chapter 7. (The work in Chapter 8 will be presented in *ICSSC 2013*.)

Chapter 9: Coded transmission of AF protocol over nonlinear satellite channel

We want to evaluate the performance of coded two-way communication (AF protocol) over a nonlinear channel. Two decoding schemes: memoryless decoding and memory-2 decoding are presented, and both decoding schemes favor a large IBO (-6 dB), which is quite different from the single-user case. The memory-2 scheme is based on a decoder feedback mechanism, and it avoids the complicated calculation of the BCJR algorithm [48]. Simulation results show that it can bring significant performance improvement as compared with a memoryless decoder, and there is only small performance gap between our proposed decoder and an ideal (genie-aided) memory-2 decoder. We also consider the scenario when two signals are asynchronous, say there is half symbol delay between them. Such asynchronization induces strong linear ISI that the memoryless decoding scheme achieves a severe error floor. By adopting two linear ISI compensators, we present a feedback-based memory-2 decoder which can achieve almost the same performance as in the synchronous case. (Part of the work in Chapter 9 will be presented in *ICSSC 2013*.)

Chapter 2

General Problem Model

2.1 Two-way relay model

The general system diagram is shown in Figure 2.1. Two terminals T_1 and T_2 want to exchange information via a relay R. E and D denote encoders and decoders respectively. Note the availability of decoder side-information, which is the reason that a non-orthogonal signalling protocol can achieves up to 100% spectrum improvement as compared with a orthogonal signalling protocol.
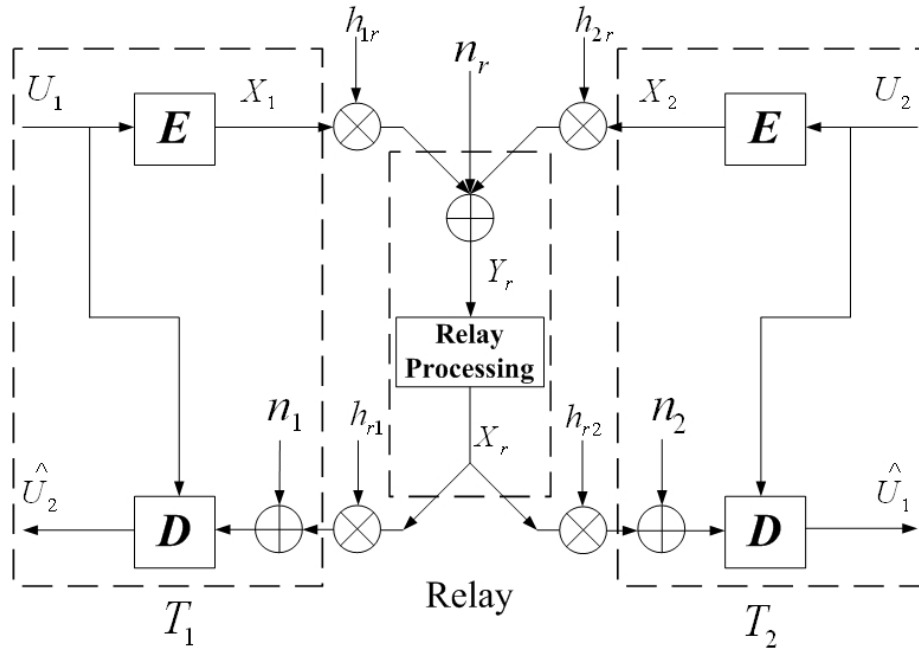
Figure 2.1: Two-way relay channel model.

We assume that each terminal is operated in full-duplexing model by using different frequency bands for sending (uplink) and receiving (downlink). We further assume the links are non-fluctuating, at least over a data frame. Important physical parameters of the system are the transmit power constraints at each terminal, denoted P_1, P_2 and P_r, the receive noise power spectral densities, η_1, η_2 and η_r, and the complex path gains between terminals, which incorporate factors of antenna gains, and free-space loss. These are denoted by h_{1r}, h_{r1} for links between T_1 and R, and similarly h_{2r}, h_{r2} for node T_2's links. For convenience, Table 2.1 lists parameters which will be used throughout this thesis.

Notation	Definition
P_1, P_2, P_r	maximum power of T_1, T_2 and the relay
X_1, X_2, X_r	data symbols sent from T_1 and T_2 and the relay
R_{12}, R_{21}	information rate from T_1 to T_2, and from T_2 to T_1
h_{1r}, h_{2r}	effective channel gain from T_1 to relay, and T_2 to relay, complex-valued
h_{r1}, h_{r2}	effective channel gain from relay to T_1, and from relay to T_2, complex-valued
n_1, n_2, n_r	white Gaussian noise at T_1, T_2, and the relay
$\eta_1, \eta_2, \eta_r, \eta$	noise spectral density at T_1, T_2, and the relay, with $\eta_1 = \eta_2 = \eta_r = \eta$
B_{up}, B_{down}	bandwidth for the uplink or downlink, with $B_{up} = B_{down} = B$

Table 2.1: General parameter definitions

As will be shown later, it is convenient to define the set of link SNR's

$$\rho_{1r} = \frac{|h_{1r}|^2 P_1}{\eta_r B}, \quad \rho_{2r} = \frac{|h_{2r}|^2 P_2}{\eta_r B}$$
$$\rho_{r1} = \frac{|h_{r1}|^2 P_r}{\eta_1 B}, \quad \rho_{r2} = \frac{|h_{r2}|^2 P_r}{\eta_2 B}$$

which are the the maximum effective signal-to-noise ratios for each link in a bandwidth of B at the relevant receiving terminal. Notice that ρ_{r1} (ρ_{r2}) defines the link SNR when all the relay power is allocated to link $R \to T_1$ ($R \to T_2$), and as will be shown later in this thesis, the relay's power is actually shared between two signals in most cases.

2.2 Memoryless nonlinear amplifier model

Satellite transponders are well-known to suffer nonlinear AM/AM and AM/PM distortions due to a traveling wave tube amplifier (TWTA). We will adopt the Saleh [30] model to describe the nonlinear characteristic, which is in fact the most widely used instantaneous model in the literature due to its accuracy and simplicity. (2.1) and (2.2) define the base-band model of TWTA as two magnitude-dependent transfer functions given by

$$A(r) = \frac{\alpha_a r}{1 + \beta_a r^2} \tag{2.1}$$

$$\Phi(r) = \frac{\alpha_\phi r^2}{1 + \beta_\phi r^2} \tag{2.2}$$

where r is the magnitude of the input signal, $A(r)$ is the AM/AM transfer function, $\Phi(r)$ is the AM/PM transfer function, and $\alpha_a, \beta_a, \alpha_\phi, \beta_\phi$ are amplifier parameters. If the instantaneous input is $r(t)e^{j\theta(t)}$, then the output of the amplifier is

$$y(t) = A(r(t))e^{j\theta(t)}e^{j\Phi(r(t))}$$

We fix $\alpha_a = 2.16$, $\beta_a = 1.15$, $\alpha_\phi = 4.00$, $\beta_\phi = 9.10$ [30] throughout this dissertation. Figure 2.2 and 2.3 present the AM/AM and AM/PM characteristics of the Saleh model for different input backoff (IBO), which is the difference in decibels between input power and saturating input power of the amplifier.

For a nonlinear channel, the SNR parameter of interest is the ratio of the saturation power at the amplifier over the power of receiver noise. We express this ratio as $E_{sat}/N_0 = P_{sat}T_{symb}/N_0$. Without loss of generality, we fix the saturation power to be 1 at the receiver, and scale the noise accordingly.
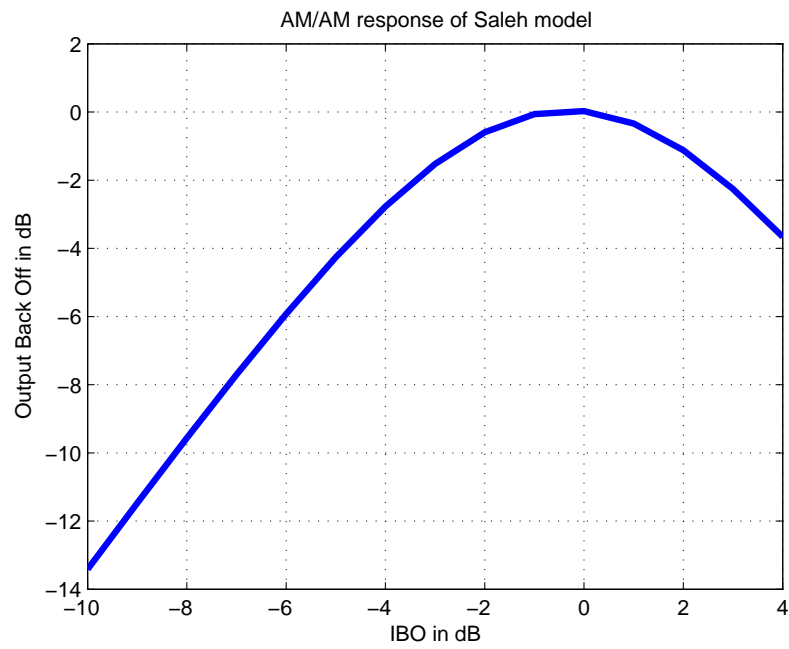
Figure 2.2: AM/AM response of Saleh model
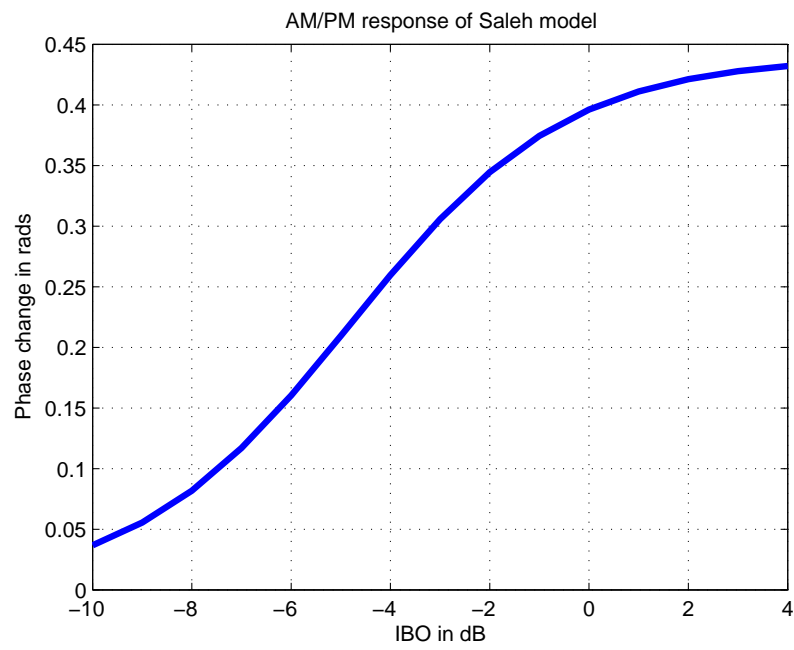
Figure 2.3: AM/PM response of Saleh model

Chapter 3

Codification of Different Two-way Relay Protocols

As have been stated in Chapter 1, the abstract communication model is a special case of a two-way relay channel, whose origins date back to Shannon's two-way channel in 1961 [2]. A list of corresponding work in this area can be found in [3, 5, 14, 15, 19–22, 24, 26, 27, 49]. The relay may be as simple as a reamplifying node, or a more sophisticated demodulate/decode/reencode terminal. Protocols studied include AF (non-orthogonal amplify-and-forward), MAC-XOR (multiple access followed by relay XOR and retransmission), MAC-BSI (multiple access followed by downlink broadcast channel with side information), MAC-NC (multiple access followed by relay nested coding), and NLC (Nested Lattice Coding). All of this literature attempts to design more efficient coding, power allocation, and signal processing schemes to maximize the achievable rate region[1], based on some strong, but maybe impractical assumptions about synchronization of user transmissions and known signal strengths. The capacity region for this channel is still unknown [3], though the cut-set bound has been shown to be approachable with certain relaying protocols.

We will speak of *achievable rate region* and *capacity region* in this thesis, which are defined below [7].

---

[1]The achievable rate region (ARR) is the set of bidirectional rate pairs for which reliable communication is possible, in the information-theoretic sense.

Definitions

- A rate pair (R_{12}, R_{21}) is *achievable* for some relaying protocols if there exists a sequence of $((2^{nR_{12}}, 2^{nR_{21}}), n)$ distributed channel codes (at T_1 and T_2) and corresponding decoders with probability of error $P_e(n) \to 0$ as $n \to \infty$, for both messages.

- The set of all achievable rate pairs for a given relaying protocol is defined as the *achievable rate region* (ARR).

- The closure of achievable rate regions over all possible relaying protocols and input probability distributions is defined as the *capacity region*.

This Chapter will summarize and evaluate these protocols for the two-way satellite transmission problem. We will evaluate the performance via theoretical analysis from the information theory perspective, specifically in describing achievable rate regions. Protocols considered include traditional orthogonal amplify-forward as a baseline for comparison, then AF, MAC-XOR, MAC-BCSI, MAC-NC, and NLC. We also consider the cut-set bound as a reference to show how efficient these protocols are. As will be shown later in this Chapter, MAC-NC is always at least as good as MAC-XOR and MAC-BSI, but nested lattice coding has impractical synchronization and channel gain requirements, so our later focus is on the comparison between AF and MAC-NC.

3.1   Relay protocols

In the sections below we summarize what is known about this problem; so far as we know the capacity region for the channel remains unknown, though as we'll see protocols have been shown to closely approach the outer-bound (cut set) on the capacity region in some scenarios.

3.1.1   Outer-bound to capacity region

Starting with the cut-set bound on information rates in general multi-user networks [7], we can obtain a simple outer-bound on the set of all achievable two-way rate pairs $\{R_{12}, R_{21}\}$. The cut-set bound is obtained by assuming there is no interference and no resource competition (bandwidth and relay's power) between the two information flows $T_1 \to R \to T_2$ and $T_2 \to R \to T_1$, that the system can effectively be divided into two separate channels as shown in Figure 3.1.
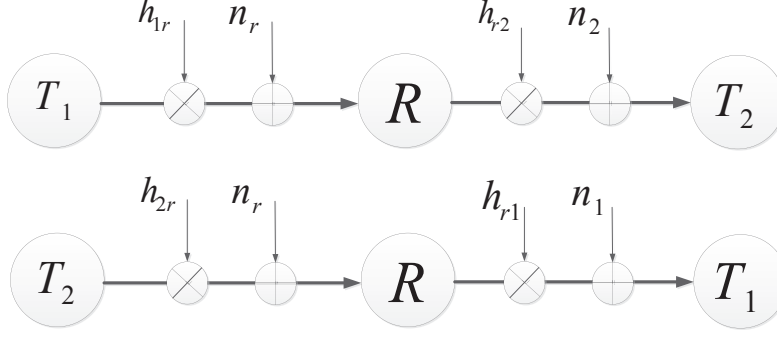
Figure 3.1: Effective channel model for cut-set analysis

Take information flow $T_1 \rightarrow R \rightarrow T_2$ as an example. Assuming an additive Gaussian channel [2], we have

$$R_{12} \leq \frac{B}{2B} \log_2(1 + \frac{|h_{1r}|^2 P_1}{\eta_r B})\text{bps/Hz}$$
$$=\leq \frac{1}{2} \log_2(1 + \rho_{1r}) \quad \text{bps/Hz}$$

which is the constraint on link $T_1 \rightarrow R$, and

$$R_{12} \leq \frac{B}{2B} \log_2(1 + \frac{|h_{r2}|^2 P_r}{\eta_2 B})\text{bps/Hz}$$
$$=\leq \frac{1}{2} \log_2(1 + \rho_{r2}) \quad \text{bps/Hz}$$

which is the constraint on link $R \rightarrow T_2$. The cut set bound for R_{12} is

$$R_{12} \leq \mathcal{MIN}\{\frac{1}{2} \log_2(1 + \rho_{1r}),\ \frac{1}{2} \log_2(1 + \rho_{r2})\} \quad \text{bps/Hz}$$

Similarly, the cut set bound for R_{21} is

$$R_{21} \leq \mathcal{MIN}\{\frac{1}{2} \log_2(1 + \rho_{2r}),\ \frac{1}{2} \log_2(1 + \rho_{r1})\} \quad \text{bps/Hz}$$

The factor 1/2 derives not from the use of 1-dimensional (real) signals, but from the duplexing penalty of using two units of bandwidth, one for uplinks and one for downlinks. Time-division-

---

[2]We assume noise at the receiver is white and Gaussian throughout this thesis

duplexing would experience the same penalty.[3] This bound on the capacity region is a rectangle in the $R_{12} - R_{21}$ plane, whose intercepts depend on the four link SNR's, and has the simple interpretation that each rate is individually bounded by the smaller of the uplink and downlink *single-user* capacities on links the flow traverses. In other words, each directional flow is limited by the smaller of the capacities of the tandem Gaussian channels the flow encounters.

3.1.2 Orthogonal amplify and forward

Orthogonal amplify and forward relaying represents the traditional approach to satellite data exchange, which is based on FDMA (Frequency Division Multiple Access). Uplink and downlink bands are similarly divided into disjoint segments, having bandwidths αB and $(1 - \alpha)B$ respectively devoted to flows $T_1 \to R \to T_2$ and $T_2 \to R \to T_1$. The parameter $0 \leq \alpha \leq 1$ represents a tuning parameter that can be optimized to best achieve a certain rate pair. Note the two signal streams will adopt different signalling rates commensurate with the allocated bandwidth.

The relay receives the sum of the two orthogonal signals, plus wideband noise over the full band. It amplifies the received two orthogonal signals and broadcasts them using bandwidths αB and $(1 - \alpha)B$ respectively. T_1 and T_2 demodulate the signal of only the desired band, and do not experience any multi-user interference, due to the assumed orthogonality.

The relay's downlink signal is, in discrete-time representation;

$$X_r = W(h_{1r}\sqrt{P_1}X_1 + h_{2r}\sqrt{P_2}X_2 + n_r)$$

where W is the signal amplifying factor, adjusted to satisfy the relay's power constraint:

$$W^2(|h_{1r}|^2 P_1 + |h_{2r}|^2 P_2 + B\eta_r) = P_r$$

Then T_1 receives

$$Y_1 = h_{r1}W(h_{1r}\sqrt{P_1}X_1 + h_{2r}\sqrt{P_2}X_2 + n_r) + n_1$$

---

[3]We prefer to describe rates in units of bits/second/Hz, rather than bits/symbol; the former is more general when terminals do not necessarily transmit at the same symbol rate.

Applying the standard formula for capacity of a bandlimited Gaussian channel and normalizing by total bandwidth, we have

$$R_{21} \leq \frac{\alpha B}{2B} \log_2(1 + \frac{|h_{r1}|^2 W^2 |h_{2r}|^2 P_2}{W^2 \alpha B \eta_r + \alpha B \eta_1})$$

$$\leq \frac{\alpha}{2} \log_2(1 + \frac{\rho_{2r}\rho_{r1}}{\alpha(1 + \rho_{1r} + \rho_{2r} + \rho_{r1})}) \quad \text{bps/Hz}$$

Similarly,

$$R_{12} \leq \frac{(1 - \alpha)}{2} \log_2(1 + \frac{\rho_{1r}\rho_{r2}}{(1 - \alpha)(1 + \rho_{1r} + \rho_{2r} + \rho_{r2})}) \quad \text{bps/Hz}$$

A typical ARR for orthogonal amplify and forward is found in Figure 3.2 for symmetric link conditions, when $\rho_{1r} = \rho_{2r} = 25$ dB and $\rho_{r1} = \rho_{r2} = 30$ dB, i.e. stronger downlinks. Clearly the ARR is far inside the cut-set bound due to power sharing and uplink noise forwarding, which in fact is the primary impetus for our study. In fairness, we note that for applications dictating highly-asymmetric rate targets, e.g. $R_{21} = 5R_{12}$, the orthogonal FDMA protocol can operate close to the outer-bound, basically by allocating nearly all bandwidth and relay power to R_{21}.
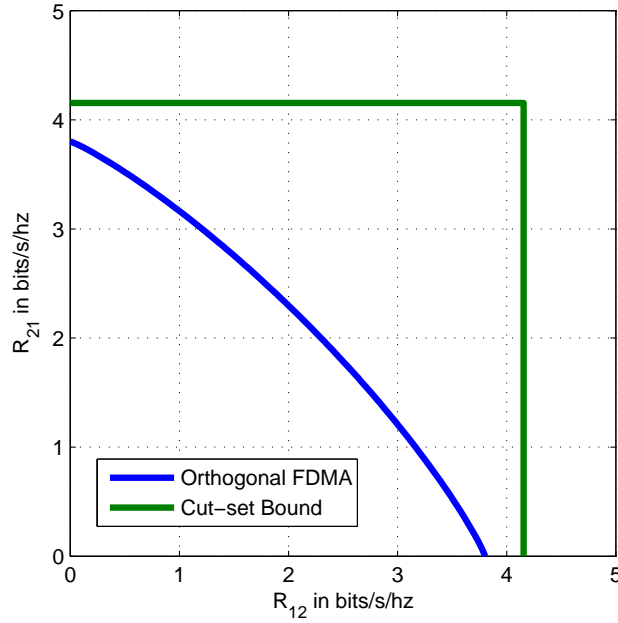
Figure 3.2: Cut-set bound vs orthogonal amplify and forward
, with $\rho_{1r} = \rho_{2r} = 25$ dB and $\rho_{r1} = \rho_{r2} = 30$ dB

3.1.3 Non-orthogonal amplify and forward (AF)

We now turn to non-orthogonal amplify-and-forward (AF), representing the first of the efficient exchange protocols. Both transmitters encode their messages into waveforms that occupy the *full* bandwidth B Hz, and thus represent interference to each other. The relay bandpass filters the sum of these, plus its own Gaussian noise, and transmits a reamplified, but power-constrained, version of this sum. Then T_1 (T_2) receives a sum of two signals originating from the two terminals over a noisy channel, and it decodes U_2 (U_1) by interference cancellation (or echo cancellation) based on the side-information of U_1 (U_2). If both terminals and the relay use the maximum power allowed, the discrete-time formation of the received signal at relay R is

$$Y_r = h_{1r}\sqrt{P_1}X_1 + h_{2r}\sqrt{P_2}X_2 + n_r$$

where n_r represents full-band (B) Gaussian noise in discrete-time. After amplification, the relay transmits

$$X_r = W(h_{1r}\sqrt{P_1}X_1 + h_{2r}\sqrt{P_2}X_2 + n_r)$$

where W is the amplifying factor at the relay, with

$$W = [\frac{P_r}{|h_{1r}|^2 P_1 + |h_{2r}|^2 P_2 + \eta_r B}]^{1/2}$$

This ensures the average relay output power is P_r. T_1 receives

$$Y_1 = h_{r1}W(h_{1r}\sqrt{P_1}X_1 + h_{2r}\sqrt{P_2}X_2 + n_r) + n_1$$

Assuming that T_1 knows W, h_{1r}, and h_{r1}, by using the side information on X_1, T_1 may cancel the interference caused by signal X_1 at Y_1, and obtain a noisy version of X_2: [4]

---

[4]Perfect interference cancellation is assumed for analysis from information theory perspective.

17

$$Y_1' = Y_1 - h_{r1}Wh_{1r}\sqrt{P_1}X_1$$

$$= Wh_{2r}h_{r1}\sqrt{P_2}X_2 + Wh_{r1}n_r + n_1$$

In the end, this is equivalent to sending T_2's coded signal to T_1 via a single-user Gaussian channel, after a power-sharing effect in the satellite. If we further assume T_1 learns h_{2r} then T_1 can decode X_2 by coherent detection. Then the information rate R_{21} from T_2 to T_1 is constrained by

$$R_{21} \leq \frac{B}{2B}\log_2\left(1 + \frac{W^2|h_{2r}|^2|h_{r1}|^2P_2}{W^2|h_{r1}|^2\eta_r B + \eta_1 B}\right) \quad \text{bps/Hz}$$

In terms of the four link SNRs, this reduces to

$$R_{21} \leq \frac{1}{2}\log_2\left(1 + \frac{\rho_{2r}\rho_{r1}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r1}}\right) \quad \text{bps/Hz} \tag{3.1}$$

Similarly for R_{12}, we have

$$R_{12} \leq \frac{1}{2}\log_2\left(1 + \frac{\rho_{1r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r2}}\right) \quad \text{bps/Hz} \tag{3.2}$$

The achievable rate region for AF is thus a *rectangle*, for any fixed vector $\bar{\rho} = (\rho_{1r}, \rho_{2r}, \rho_{r1}, \rho_{r2})$, though there is parameter coupling as seen in (3.1) and (3.2), differing with the cut-set bound. Notice that increasing one uplink strength has the effect of decreasing the downlink SNR for the flow in the opposite direction, because of relay power sharing of the total power P_r. Due to this power sharing effect in AF at the relay, there is possible benefit in uplink power backoff, depending on a desired operating point in the 2-D rate region. We can, however, regard the resource vector $\bar{\rho}$ as providing an *upper-bound* on uplink SNR, and choose to operate with lower power if desirable. The achievable rate region for AF then becomes the union of the set of all rectangular regions satisfying the above inequalities, where SNR's are at or below the constraint, as shown in Figure 3.3 and 3.4. The rate region boundary then can improve slightly over that obtained from using full power, in scenarios where rate asymmetry is desired.
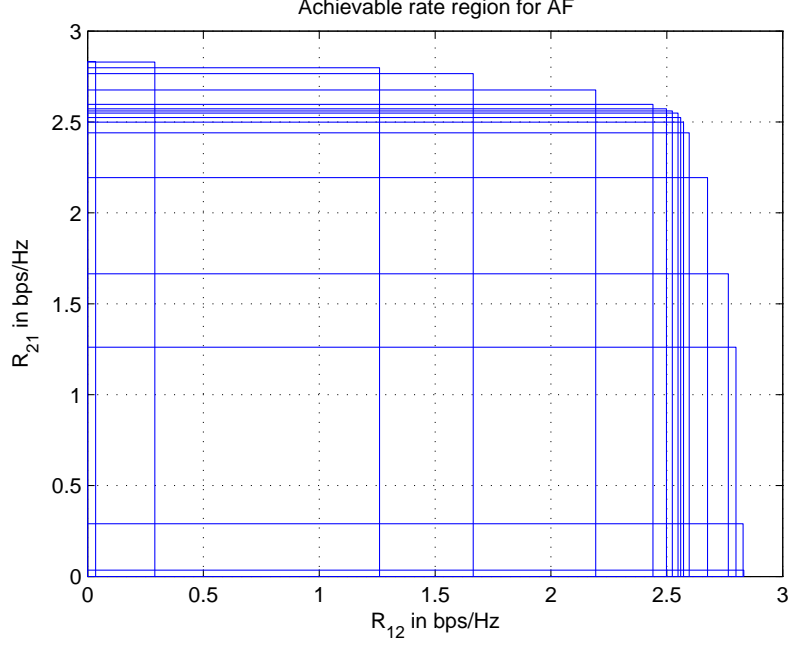
Figure 3.3: The ARR of AF is the union of rectangular regions
with $\rho_{1r} = \rho_{2r} = \rho_{r1} = \rho_{r2} = 20$ dB
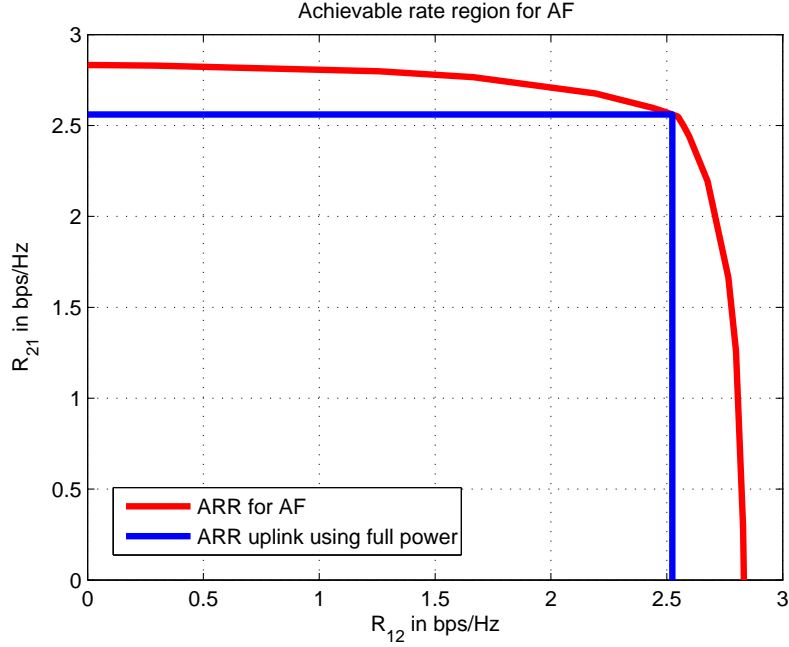
Figure 3.4: ARR vs ARR of using full uplink power
with $\rho_{1r} = \rho_{2r} = \rho_{r1} = \rho_{r2} = 20$ dB

The next series of protocols are two-phase schemes requiring relay demodulation, decoding, and reencoding. The uplink phase in each is a classic Gaussian multiple access channel (MAC) [7].

3.1.4 MAC-broadcast using side-information (MAC-BSI)

In MAC-BSI, the relay first decodes two signals, constituting a typical Gaussian Multiple Access Channel (MAC). Then the relay generates two codebooks, one with power αP_r for U_1, and the other with power $(1-\alpha)P_r$ for U_2, and broadcasts the superposition of the two codewords. T_1 and T_2 decode the superposed codewords based on the side information. (We assume both T_1 and T_2 know the codebooks used in the relay) We have the achievable rate region for MAC-BSI as below:

$$R_{12} \leq \frac{1}{2}\log(1+\rho_{1r})$$
$$R_{21} \leq \frac{1}{2}\log(1+\rho_{2r})$$
$$R_{12} + R_{21} \leq \frac{1}{2}\log(1+\rho_{1r}+\rho_{2r})$$

which are the MAC constraints for the uplinks, and

$$R_{12} \leq \frac{1}{2}\log(1+\alpha\rho_{r2})$$
$$R_{21} \leq \frac{1}{2}\log(1+(1-\alpha)\rho_{r1})$$

which are the rate constraints for the downlinks. Notice that the downlink constraints here are different from the traditional Gaussian broadcast channel [7], due to the availability of side information. The main deficiency of MAC-BSI is the power sharing for two signals at the relay in the superposition coding.

3.1.5 MAC-XOR

In MAC-XOR, after the relay successfully decodes U_1 and U_2 in the MAC phase, it XORs the two decoded data streams bit-by-bit, reminiscent of some network coding strategies. (If the two data streams have different rate, padding 0's is needed [14].) Then the relay encodes the XORed message, and broadcasts the common message. T_1 (T_2) decodes the common message and XORs the decoded common message with U_1 (U_2) to get U_2 (U_1). The information rate constraints for

20

MAC-XOR include the MAC constraints as in MAC-XOR, in addition to

$$\mathcal{MAX}\{R_{12}, R_{21}\} \leq \frac{1}{2}\mathcal{MIN}\{\log(1 + \rho_{r1}),\ \log(1 + \rho_{r2})\}$$

which ensures that *both* T_1 and T_2 can decode the common message. In MAC-XOR, the relay broadcasts only the common message, so there is no explicit power sharing problem as in MAC-BSI. However, the maximum of the two rates is constrained by the poorer of the two downlinks. Under certain circumstances, even though the quality of one downlink is very good, if the quality of the other downlink is poor, the information rate of broadcast phase is still strictly constrained by this poor downlink, which can severely degrade the performance. Practically, such asymmetric channels are common, so we need to find a more efficient protocol.

3.1.6   MAC-Nested code (MAC-NC)

MAC-XOR suffers with asymmetric dowlinks, while MAC-BSI suffers from power sharing at the relay. A more efficient protocol is MAC-NC, which actually achieves the channel capacity of the downlinks. In MAC-NC, the relay first decodes U_1 and U_2, then using two codebooks for U_1 and U_2, encodes these two messages respectively, XORs the encoded messages, and broadcasts this single codeword [16, 18]. Thus MAC-NC can be viewed as MAC-XOR with XORing done *after* channel coding at the relay. An example of MAC-NC is shown in Figure 3.5 for the binary coding case.
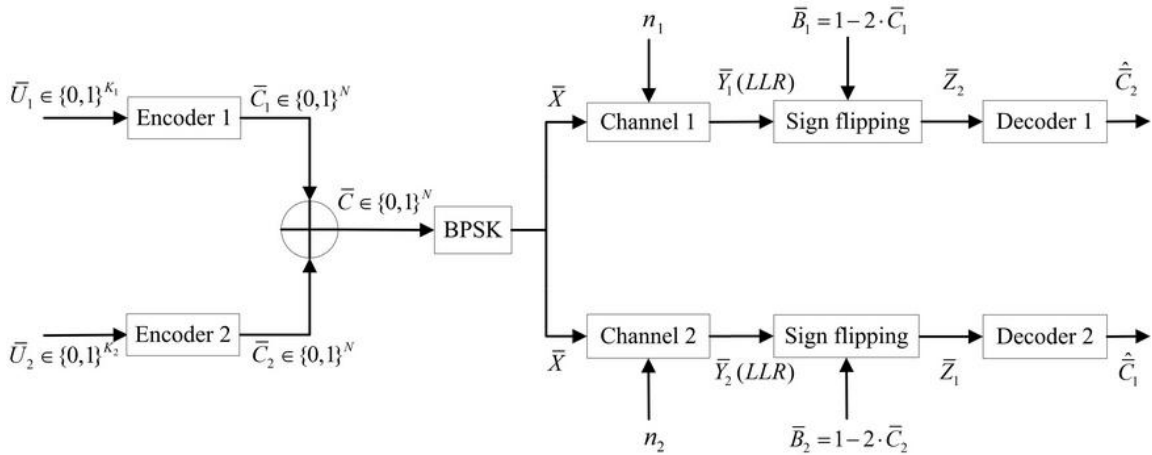
Figure 3.5: Nested coding and decoding for MAC-NC

21

The relay encodes U_1 and U_2 into C_1 and C_2, XORs C_1 and C_2, and broadcasts the modulated XORed message (not necessary to be BPSK modulation). Note that the codes $\mathcal{C}_1$ and $\mathcal{C}_2$ are nested inside $\mathcal{C}$, the set of all codewords $C_1 + C_2$. T_1 calculates log-likelihood ratios $LLR_{\bar{Y}_1}$ for the unknown sequence $C = \{C^1, C^2, ...C^N\}$. Based on the side information, T_1 can calculate $C_1 = \{C_1^1, C_1^2, ...C_1^N\}$, and flip the signs of the LLR sequence as

$$
LLR_{\bar{Z}_2}(C_2^i = 0) = \begin{cases} LLR_{\bar{Y}_1}(C^i = 0), & C_1^i = 0 \\ LLR_{\bar{Y}_1}(C^i = 1), & C_1^i = 1 \end{cases}
$$

$$
LLR_{\bar{Z}_2}(C_2^i = 1) = \begin{cases} LLR_{\bar{Y}_1}(C^i = 1), & C_1^i = 0 \\ LLR_{\bar{Y}_1}(C^i = 0), & C_1^i = 1 \end{cases}
$$

Then the LLR information is passed to a soft decoder to decode $\hat{\hat{C}}_2$, so the decoding at T_1 can be viewed as decoding on a single-user Gaussian channel without interference, and we have the following information rate constraints for downlinks

$$
R_{12} \le \frac{1}{2} \log(1 + \rho_{r2})
$$
$$
R_{21} \le \frac{1}{2} \log(1 + \rho_{r1})
$$

which is the cut-set bound imposed by downlinks.

Again, the achievable rate region of MAC-NC is the intersection of the MAC rate regions and downlink rate regions. Since MAC-NC, MAC-BSI, and MAC-XOR have the same MAC rate regions for uplinks, and MAC-NC can achieve the capacity bound for the downlink phase, MAC-NC is always as good as MAC-BSI and MAC-XOR. That is, for any given ρ, the ARR for MAC-NC is at least as large as for MAC-XOR or MAC-BCSI. Furthermore, the downlink processing for MAC-NC is not more difficult than for either of MAC-XOR or MAC-BCSI, and thus among the MAC-based protocols, MAC-NC seems the clear winner, so we focus upon it in comparison with AF in this chapter.

For completeness, we describe a protocol in the 'compress-and-forward' family, which under strict assumptions about signal synchronization, can loosen the constraint imposed by the MAC uplink phase.

3.1.7 Nested lattice coding (NLC)

The NLC protocol [28] adopts lattice codes in the uplink, and the relay does not decode both of the two messages; instead the relay just produces a common message useful for both T_1 and T_2, and broadcasts it. So there is no MAC penalty as well as power sharing penalty in MAC-NC. For a Gaussian two-way relay channel, nested lattice coding can achieve the following region [28]

$$R_{12} \leq \mathcal{MIN}\{[\frac{1}{2}\log(\frac{\rho_{1r}}{(\rho_{1r}+\rho_{2r})}) + \rho_{1r}]^+, \ \frac{1}{2}\log(1+\rho_{r2})\}$$

$$R_{21} \leq \mathcal{MIN}\{[\frac{1}{2}\log(\frac{\rho_{2r}}{(\rho_{1r}+\rho_{2r})}) + \rho_{2r}]^+, \ \frac{1}{2}\log(1+\rho_{r1})\}$$

where $[x]^+ = \mathcal{MAX}\{x, \ 0\}$. It has been shown in [28] that nested lattice coding is within $\frac{1}{2}$ bit from the cut-set bound in each rate dimension. However, nested lattice coding requires perfect symbol alignment and phase rotation for X_1 and X_2 *at the satellite*, which is generally impractical, so we will not consider it further.

3.1.8 Complexity comparison between different protocols

In this section, we summarize the implementation complexities for each of the protocols discussed before in Tables 3.1 and 3.2. [5]

Protocol	Relay knows T_1 and T_2's codebook	T_1 and T_2 know relay's codebook	T_1 and T_2 know each other's codebook
FDMA	No	No	Yes
AF	No	No	Yes
MAC-XOR	Yes	Yes	No
MAC-BSI	Yes	Yes	No
MAC-NC	Yes	Yes	No
NLC	Yes	Yes	Yes

Table 3.1: Code knowledge requirements for different protocols

---

[5]It has been shown in [50] that asynchronous MAC has the same capacity as synchronous MAC.

23

Protocol	Symbol and phase alignment between X_1 and X_2	Relay demodulation/decoding	waveform interference cancellation at T_1 and T_2
FDMA	No	No	No
AF	No	No	Yes
MAC-XOR	No	Yes	No
MAC-BSI	No	Yes	Yes
MAC-NC	No	Yes	No
NLC	Yes	Yes	No

Table 3.2: Signal synchronization and processing requirements for different protocols

Overall AF requires the least complexity, though interference, or echo cancellation is crucial. Among the MAC-based schemes, MAC-NC is probably the easiest to implement, and as argued above, its ARR is at least as large as the others. Based on the above implementation complexity and ARR comparison, we will focus on numerical study of MAC-NC and AF in the next section.

3.2 Performance comparison between AF and MAC-NC

3.2.1 Achievable rate region (ARR)

One important performance metric is the achievable rate region for each protocol. For AF, as has been described in the previous section, R_{12} may benefit from decreasing the power of P_{2r}, so that the $R \rightarrow T_2$ link gets more power from the relay, and vice versa. So the actual achievable rate region for AF is the union of ARRs for all possible power pairs (P_{1r}^*, P_{2r}^*), with $P_{1r}^* \leq P_{1r}$ and $P_{2r}^* \leq P_{2r}$, where P_{1r}^*, P_{2r}^* are the actual sending power of T_1 and T_2 respectively. For MAC-NC, it is always optimal to let all terminals use the maximum power allowed.

Figures 3.6 and 3.7 show the ARR comparison between AF and MAC-NC, with cut-set bound and orthogonal AF as references. Note scale changes between figures. Figure 3.6 is for the case of (relatively) strong uplinks. As can be seen in this scenario, the ARR of MAC-NC can achieve the cut-set bound. This is because, when uplinks are strong, the MAC phase of MAC-NC no longer constrains the ARR, while for the downlinks, as shown in previous analysis, MAC-NC can always achieve the rate bounds for the downlinks. AF suffers from the power sharing at the relay, so there is gap (of less than 1/2 bit per rate dimension) between the ARR of AF and the cut-set bound.

To contrast, Figure 3.7 assumes strong downlinks and weaker, unequal uplinks. In this situation, AF beats MAC-NC for most of the region, and it can approach (not achieve) the cut-set bound.
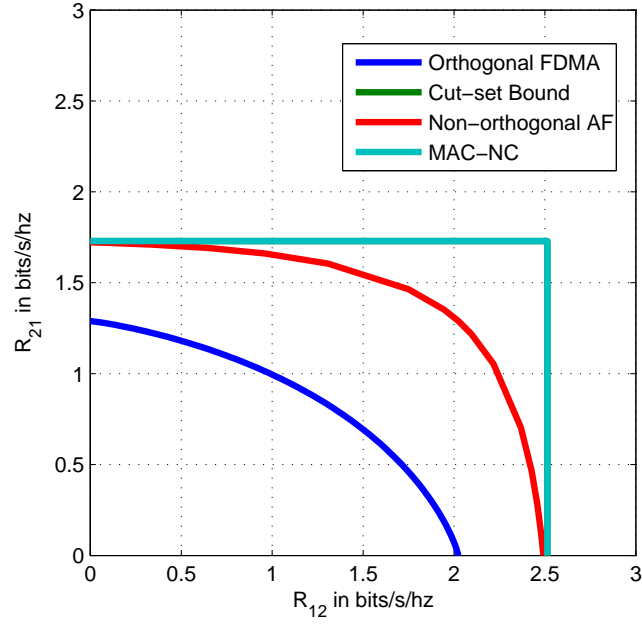
Figure 3.6: ARR comparison with strong uplink
$\rho_{1r} = 30$ dB, $\rho_{2r} = 30$ dB, $\rho_{r1} = 10$ dB, $\rho_{r2} = 15$ dB
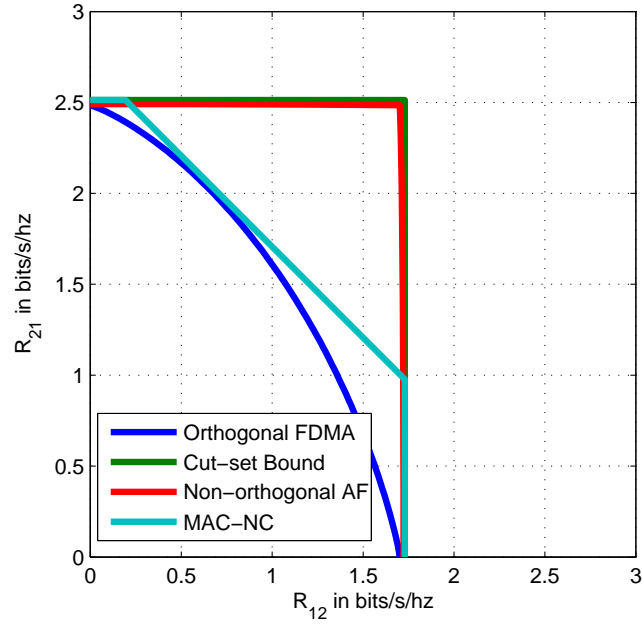
Figure 3.7: ARR comparison with strong downlink
$\rho_{1r} = 10$ dB, $\rho_{2r} = 15$ dB, $\rho_{r1} = 30$ dB, $\rho_{r2} = 30$ dB

The reason for superiority of AF here is that when downlinks are strong, power sharing at the relay is no longer a bottleneck for AF. For regions close to the axis, in which the retransmitted

uplink noise dominates the performance, MAC-NC can (very slightly) beat AF. The gap between the cut-set bound and MAC-NC is due to the fact that when uplinks are weak, the MAC phase determines the ARR of MAC-NC.

3.2.2 Optimal rate sum

We have previously compared the achievable rate regions between AF and MAC-NC. To further quantify the potential of protocols, we study another important performance metric, the maximum of the sum rate $R_{12} + R_{21}$, which shows how efficiently the allocated bandwidth is used when total throughput is of interest.

For AF, on uplinks, due to the power sharing at relay, R_{12} (R_{21}) can benefit from reducing the sending power of T_2 (T_1), while on downlinks it is always more efficient to let the relay use maximum power allowed. Let x and y be the *actual* SNRs of link $T_1 \rightarrow R$ and $T_2 \rightarrow R$ respectively, the maximum rate sum problem of AF can be stated as:

$$\mathcal{MAX} \quad R_{12} + R_{21}$$

$$s.t.$$
$$R_{21} \leq \frac{1}{2} \log_2(1 + \frac{y\rho_{r1}}{1 + x + y + \rho_{r1}}) \text{ bps/Hz}$$
$$R_{12} \leq \frac{1}{2} \log_2(1 + \frac{x\rho_{r2}}{1 + x + y + \rho_{r2}}) \text{ bps/Hz}$$
$$x \leq \rho_{1r}, \quad y \leq \rho_{2r}$$

which is a nonlinear programming problem. However, the complexity of the problem can be reduced based on the following lemma.

Lemma 1. *Let $x = \rho_{1r}^*$, $y = \rho_{2r}^*$ be the optimal SNRs pair that achieves the maximum rate sum, then either $\rho_{1r}^* = \rho_{1r}$ or $\rho_{2r}^* = \rho_{2r}$, which means at least of one of T_1 and T_2 uses the maximum allowed sending power.*

Proof. (by contradiction): Assume the maximum rate sum is achieved when $x = x_0$, $y = y_0$, with $x_0 \le \rho_{1r}$ and $y_0 \le \rho_{2r}$. Define the maximum rate sum function, MRS, as

$$
\begin{aligned}
&MRS(x,\ y)\\
&= \frac{1}{2}(\log_2(1 + \frac{x\rho_{r2}}{1 + x + y + \rho_{r2}}) + \log_2(1 + \frac{y\rho_{r1}}{1 + x + y + \rho_{r1}}))
\end{aligned}
$$

Let $\beta = y_0/x_0$, if we fix $y = \beta x$, then

$$
\begin{aligned}
&MRS(x,\ y)|_{y=\beta x}\\
&= \frac{1}{2}(\log_2(1 + \frac{x\rho_{r2}}{1 + (1+\beta)x + \rho_{r2}}) + \log_2(1 + \frac{\beta x \rho_{r1}}{1 + (1+\beta)x + \rho_{r1}}))\\
&= \frac{1}{2}(\log_2(1 + \frac{\rho_{r2}}{(1+\beta) + \frac{1+\rho_{r2}}{x}}) + \log_2(1 + \frac{\beta\rho_{r1}}{(1+\beta) + \frac{1+\rho_{r1}}{x}}))
\end{aligned}
$$

$MRS(x,\ y)|_{y=\beta x}$ increases with x. Since $x_0 < \rho_{1r}$ and $y_0 < \rho_{2r}$, there exists (x_1, y_1) that $x_1 = \gamma x_0 \le \rho_{1r}$, $y_1 = \gamma y_0 \le \rho_{2r}$, $y_1/x_1 = \beta$, with $\gamma > 1$. Then we we have

$$
MRS(x = x_1, y = y_1) > MRS(x = x_0, y = y_0),
$$

which contradicts our previous assumption that MRS is achieved when $x = x_0$, $y = y_0$. $\square$

Based on Lemma 1, we can fix $x = \rho_{1r}$ ($y = \rho_{2r}$), and vary $0 \le y \le \rho_{2r}$ ($0 \le x \le \rho_{1r}$) to find the maximum rate sum. Figure 3.8 shows the maximum rate sum as a function of SNR for symmetric up and down-link qualities, with $\rho_{1r} = \rho_{2r} = \rho_{up}$, $\rho_{r1} = \rho_{r2} = \rho_{down}$. Symmetry is enforced merely to reduce degrees of freedom.

Next consider MAC-NC. Here it is more efficient to use the maximum power for both uplinks and downlinks, so the optimal rate sum problem can be stated as

$$\mathcal{MAX} \quad R_{12} + R_{21}$$

$$\textbf{\textit{s.t.}}$$

$$R_{12} \leq \frac{1}{2}\log(1 + \rho_{1r})$$

$$R_{21} \leq \frac{1}{2}\log(1 + \rho_{2r})$$

$$R_{12} + R_{21} \leq \frac{1}{2}\log(1 + \rho_{1r} + \rho_{2r})$$

$$R_{12} \leq \frac{1}{2}\log(1 + \rho_{r2})$$

$$R_{21} \leq \frac{1}{2}\log(1 + \rho_{r1})$$

which is a linear programming problem. Figure 3.9 presents the maximum rate sum plot for MAC-NC, due to the same reason as for AF, we choose $\rho_{1r} = \rho_{2r} = \rho_{up}$, $\rho_{r1} = \rho_{r2} = \rho_{down}$. As can be seen from Figure 3.9, the maximum rate sum saturates more slowly with ρ_{up} than with ρ_{down}, which is due to MAC constraints for uplinks.
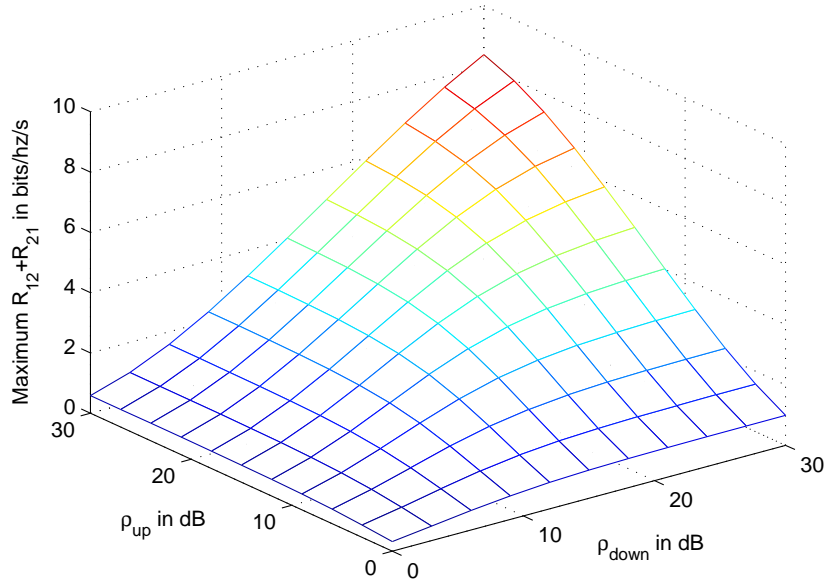
Figure 3.8: Maximum rate sum for AF

Figure 3.10 and 3.11 show the maximum rate sum comparison between AF and MAC-NC for
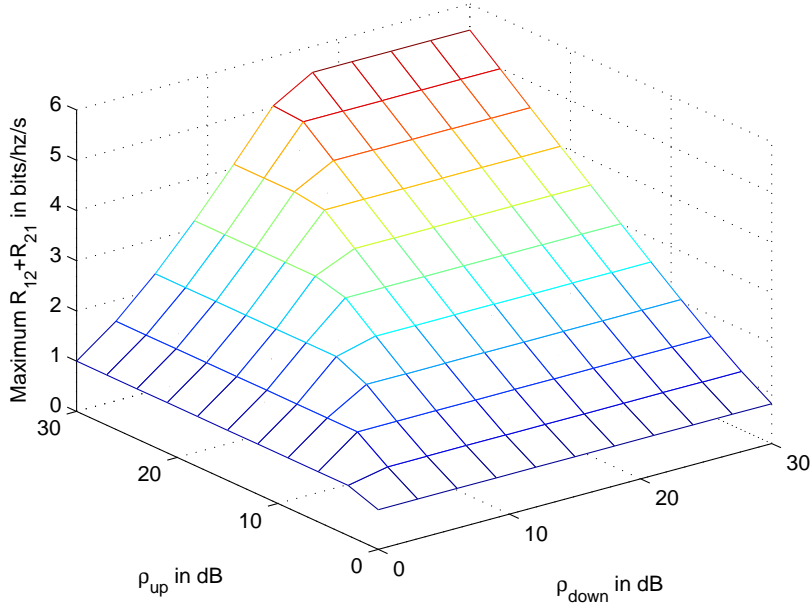
Figure 3.9: Maximum rate sum for MAC-NC

fixed ρ_{up} and ρ_{down} respectively, while the other link quality varies. Essentially these are cross-sections of Figures 3.8 and 3.9. Again, we choose $\rho_{1r} = \rho_{2r} = \rho_{up}$, $\rho_{r1} = \rho_{r2} = \rho_{down}$. In Figure 3.10, we fix $\rho_{up} = 15$ dB, and vary ρ_{down}. As can been seen, when ρ_{down} is small, MAC-NC can beat AF, and there is a 3 dB gap between AF and MAC-NC, because downlinks dominate the performance (in terms of sum rate) in this region, and AF suffers from power sharing at the relay. On the other hand, when ρ_{down} is large, uplinks dominate the performance, and AF can beat MAC-NC, for MAC-NC suffers from the MAC limitation in this region. In Figure 3.11, we fix $\rho_{down} = 15$ dB, and vary ρ_{up}. Now, when ρ_{up} is small, uplinks dominate, and AF beats MAC-NC, but when ρ_{up} is large, downlinks dominate, and MAC-NC beats AF.

3.3   Optimal rate sum with asymmetric rate requirements

In the previous subsection, we have considered the maximum rate sum problem for both AF and MAC-NC. In some scenarios, there is an extra asymmetric rate requirement, say $R_{21}/R_{12} = \alpha$, and the problem is to optimize the rate sum $R_{12} + R_{21}$, while meeting the asymmetric rate requirement $R_{21}/R_{12} = \alpha$. For AF, let x, y be the actual link SNRs for $T_1 \to R$ and $T_2 \to R$ respectively, then
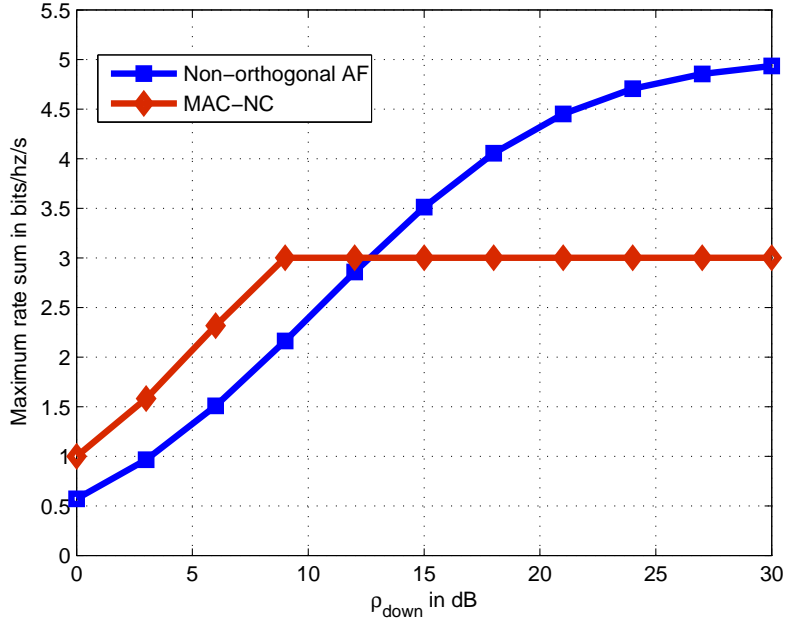
Figure 3.10: Maximum rate sum, fixing $\rho_{up} = 15$ dB
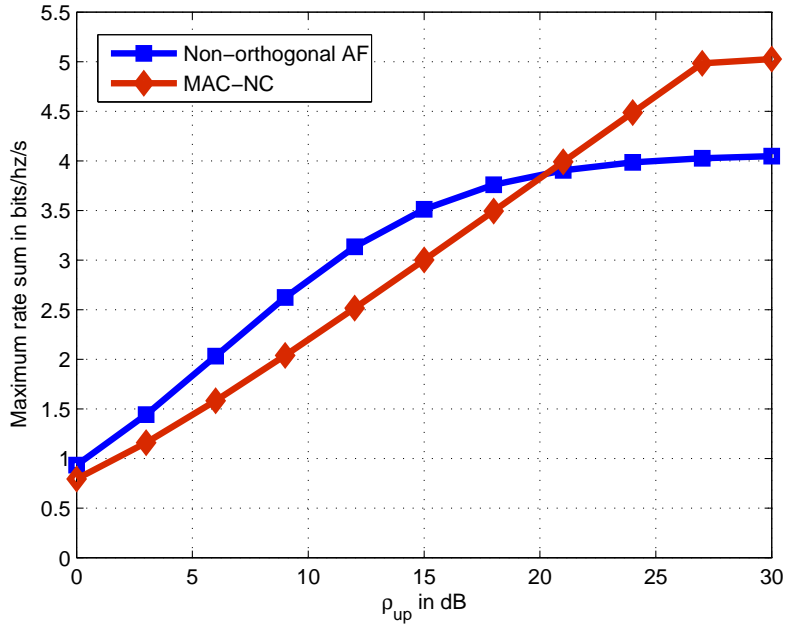
Figure 3.11: Maximum rate sum, fixing $\rho_{down} = 15$ dB

the maximum rate sum problem is

$$\mathcal{MAX} \quad R_{12} + R_{21}$$

$$\textbf{\textit{s.t.}}$$

$$R_{21} = \alpha R_{12}$$

$$R_{21} \leq \frac{1}{2} \log_2(1 + \frac{y\rho_{r1}}{1 + x + y + \rho_{r1}}) \text{ bps/Hz}$$

$$R_{12} \leq \frac{1}{2} \log_2(1 + \frac{x\rho_{r2}}{1 + x + y + \rho_{r2}}) \text{ bps/Hz}$$

$$x \leq \rho_{1r}, \quad y \leq \rho_{2r}$$

which is again a nonlinear programming problem. Similar to the previously considered maximum rate sum problem, we still have the fact that the optimal rate sum is achieved when at least one of T_1 and T_2 uses the maximum allowed sending power, based on which, we can find the maximum rate sum with asymmetric rate requirements.

For MAC-NC, the optimal rate sum with asymmetric rate requirement problem is the linear program

$$\mathcal{MAX} \quad R_{12} + R_{21}$$

$$\textbf{\textit{s.t.}}$$

$$R_{21} = \alpha R_{12}$$

$$R_{12} \leq \frac{1}{2} \log(1 + \rho_{1r})$$

$$R_{21} \leq \frac{1}{2} \log(1 + \rho_{2r})$$

$$R_{12} + R_{21} \leq \frac{1}{2} \log(1 + \rho_{1r} + \rho_{2r})$$

$$R_{12} \leq \frac{1}{2} \log(1 + \rho_{r2})$$

$$R_{21} \leq \frac{1}{2} \log(1 + \rho_{r1})$$

Figure 3.12 and 3.13 show the maximum rate sum for AF and MAC-NC with asymmetric rate requirement $R_{21} = 5R_{12}$, and we can see the impact of asymmetric rate requirements, which

decrease the maximum rate sum for any given (ρ_{up}, ρ_{down}) pair. In addition, comparing Figure 3.12 with Figure 3.8, we can see that there is no major distinction in saturation behavior between ρ_{up} and ρ_{down}. The reason is that with asymmetric rate requirement $R_{21} = 5R_{12}$, the operational rate pairs locate more closely to the R_{21} axis, and the impact of MAC constraints is no longer significant.

Figure 3.14 and 3.15 compare the performance of AF and MAC-NC where up (or down)-link quality is fixed and the other varies. In 3.14, we fix $\rho_{up} = 15$ dB, and vary ρ_{down}. We have a different result from Figure 3.10 in that the performance of AF is slightly inferior to MAC-NC for all ρ_{down} considered, which is consistent with the fact that AF is inferior to MAC-NC in the near-axis region due to the forwarding of uplink noise as shown in Figure 3.6 and 3.7. In Figure 3.15, we fix $\rho_{down} = 15$ dB, and vary ρ_{up}. The plot of Figure 3.15 is very close to that of Figure 3.14, showing that ρ_{up} and ρ_{down} are almost equally important for the maximum rate sum problem with asymmetric rate requirements.
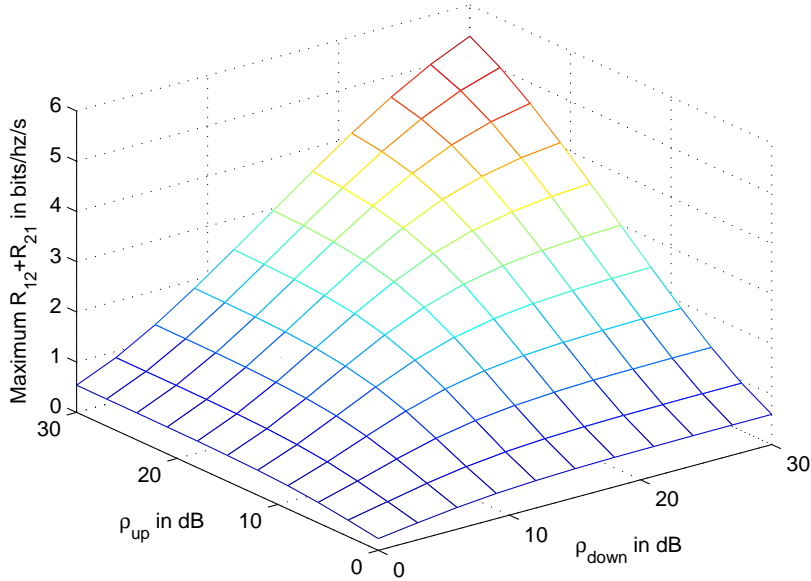
Figure 3.12: Maximum rate sum for AF, with $\alpha = \frac{R_{21}}{R_{12}} = 5$
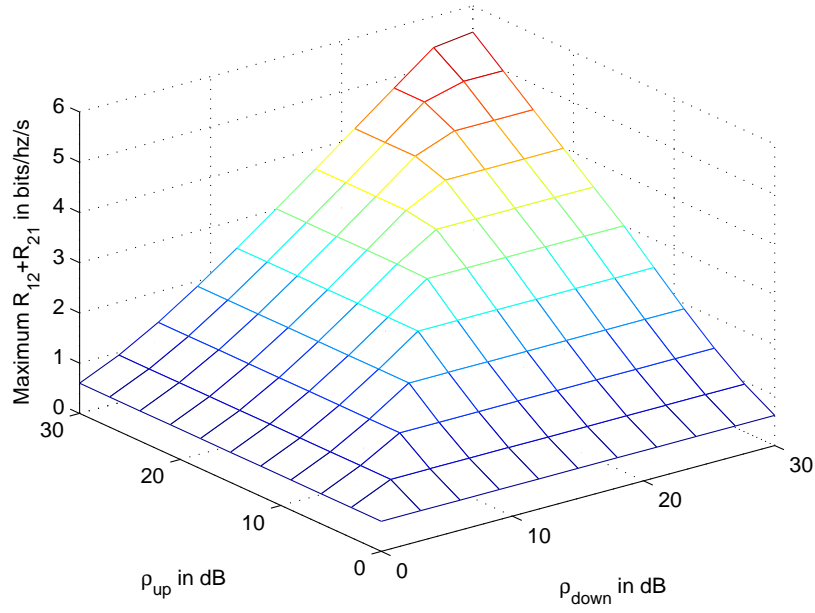
Figure 3.13: Maximum rate sum for MAC-NC, with $\alpha = 5$
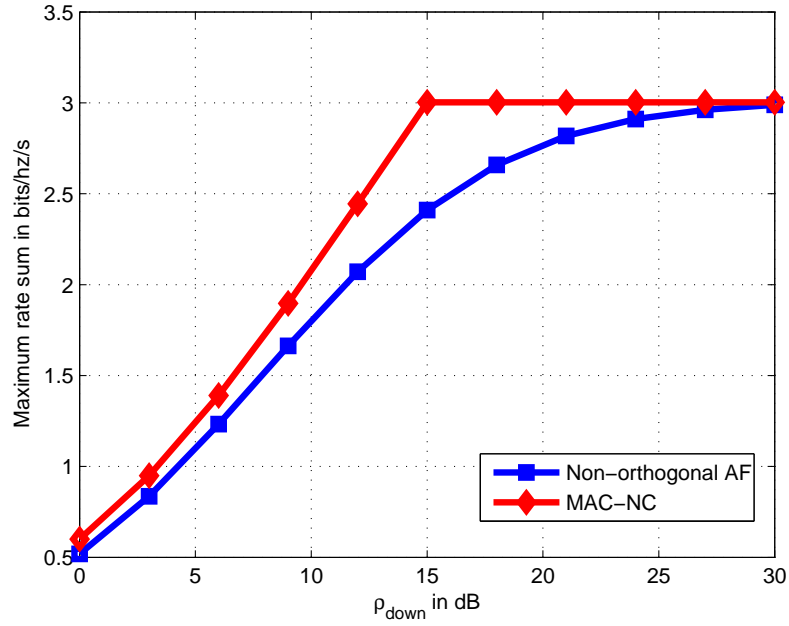
Figure 3.14: Maximum rate sum comparison for fixed ρ_{up} with $\rho_{up} = 15$ dB, and $\alpha = 5$
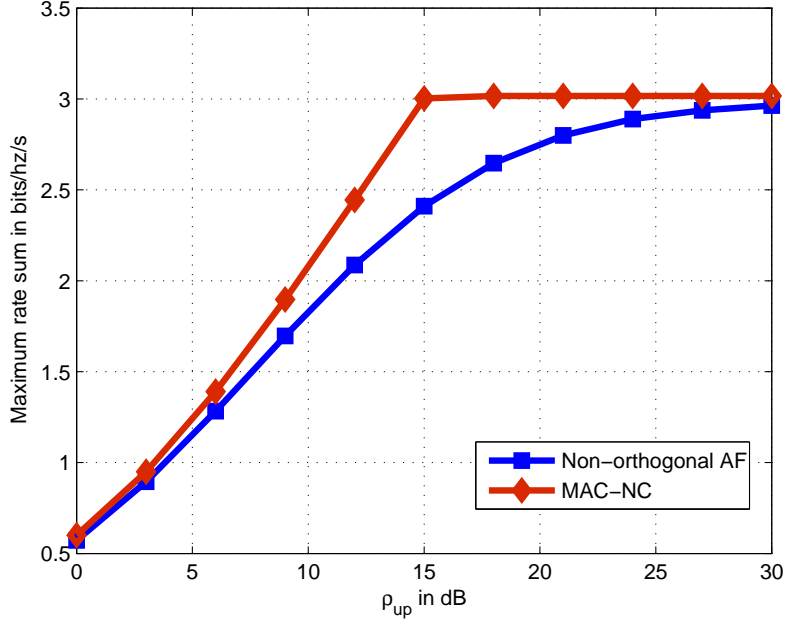
Figure 3.15: Maximum rate sum for MAC-NC, for fixed ρ_{down}
with $\rho_{down} = 15$ dB, and $\alpha = 5$

3.4 Practical issues for AF

Previous analysis has shown that AF has remarkably good performance. In addition, AF is flexible on signal synchronization, has minimum code knowledge requirements on each terminal, and is compatible with legacy non-demodulating satellites, so we will focus on AF in later part of this thesis.

In the analysis of achievable information rate of AF protocol, we have assumed perfect signal cancellation. However, practically, either time synchronization or channel estimation error can degrade the performance of signal detection. We will describe a simple signal processing procedure for the AF protocol that estimates unknown parameters such as round-trip delay and complex channel gain, then does interference cancelation, and finally data symbol detection. The signal processing procedure proposed here only shows the feasibility of AF, and it is not necessary optimal. We believe more accurate synchronization and channel estimation can be obtained if some joint estimation techniques are adopted.

Let's consider the AF protocol in (continuous) complex baseband. Assume the time delay from T_1 to relay is τ_{1r}, and the delay from T_2 to relay is τ_{2r}. Assuming for purposes here that the

frequency offset is zero, the relay receives, in complex baseband notation,

$$Y_r(t) = h_{1r}\sqrt{P_1}S_1(t - \tau_{1r})$$
$$+ h_{2r}\sqrt{P_2}S_2(t - \tau_{2r}) + n_r(t)$$

Then the relay amplifies the received signal with a scaling factor of W, and broadcasts the amplified signal to both T_1 and T_2, with W satisfying the power constraint at the relay, namely

$$|W|^2(|h_{1r}|^2P_1 + |h_{2r}|^2P_2 + N_0B) = P_r.$$

Without loss of generality, we consider the recovery of X_2 at T_1. Assuming the time delay from the relay to T_1 is τ_{r1}, then T_1 receives

$$Y_1(t) = h_{r1}(WY_r(t - \tau_{r1})) + n_1(t)$$
$$= h_{r1}Wh_{1r}\sqrt{P_1}S_1(t - (\tau_{1r} + \tau_{r1}))$$
$$+ h_{r1}Wh_{2r}\sqrt{P_2}S_2(t - (\tau_{2r} + \tau_{r1}))$$
$$+ h_{r1}Wn_r(t - \tau_{r1}) + n_1(t)$$

Let

$$a_1 = h_{r1}Wh_{1r}\sqrt{P_1}, \quad a_2 = h_{r1}Wh_{2r}\sqrt{P_2},$$

$$w_1(t) = h_{r1}Wn_r(t - \tau_{r1}) + n_1(t),$$

and

$$\tau_{s1} = \tau_{1r} + \tau_{r1}, \quad \tau_{s2} = \tau_{2r} + \tau_{r1}$$

Then

$$Y_1(t) = a_1S_1(t - \tau_{s1}) + a_2S_2(t - \tau_{s2}) + w_1(t)$$

35

where $w_1(t)$ is Gaussian noise with spectral density $(1 + |h_{r1}W|^2)N_0$.

The signal processing procedure for detecting X_2 is shown in Figure 3.16, which involves recon-stitution of the signal sent by T_1 at the proper delay, and with the appropriate complex gain, using the known side-information X_1. To do so, T_1 first estimates the delay τ_{s1} and effective channel gain a_1 for S_1. Then T_1 tries to cancel the self-interference caused by S_1. After interference cancelation, T_1 gets a noisy version of S_2, and T_1 estimates the delay τ_{s2} and effective channel gain a_2 for S_2, as conventionally required. Finally, T_1 demodulates/decodes S_2 to recover X_2.
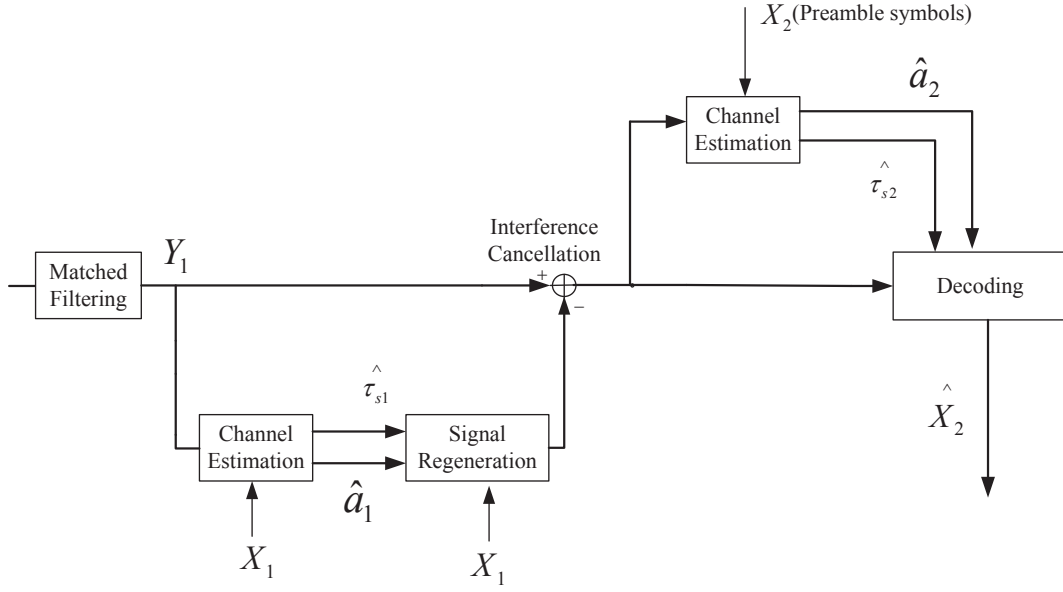
Figure 3.16: Signal processing diagram for AF

3.4.1 Delay estimation for S_1

The synchronization for S_2 follows interference cancelation, and is similar to the synchronization problem over a noisy Gaussian channel. So we will mainly discuss the synchronization for S_1.

We assume that T_1 has a rough estimate Δ for τ_{s1}, which is the round-trip delay between T_1 and the relay, based on orbit information or prior measurements. We assume the actual delay τ_{s1} satisfies $\Delta - MT \le \tau_{s1} \le \Delta + MT$, i.e. τ_{s1} is within a window of size $2MT$ seconds, or $2M$ symbols. Using an adopted delay granularity t_0, T_1 estimates τ_{s1} by finding the integer C with $\Delta - MT \le \Delta + Ct_0 \le \Delta + MT$ that maximizes the cross-correlation magnitude

$$|R_{xy}(C)| = |\sum_{n=1}^{N_w} Y_1(\Delta + Ct_0 + (n-1)T)X_{1,n}^*|$$

$$= |\sum_{n=1}^{N_w} a_1 S_1(\Delta + Ct_0 + (n-1)T - \tau_{s1})X_{1,n}^*$$

$$+ \sum_{n=1}^{N_w} (a_2 S_2(\Delta + Ct_0 + (n-1)T - \tau_{s2})X_{1,n}^*$$

$$+ \sum_{n=1}^{N_w} w_1(\Delta + Ct_0 + (n-1)T)X_{1,n}^*|$$

where N_w is the number of data symbols used in the correlation estimate at each delay. Obviously the smaller t_0 the more accurate is the possible synchronization. But small t_0 increases processing complexity, and perfect synchronization is not needed anyway. In this paper, we adopt $t_0 = \frac{T}{8}$. The second and third terms above, representing error due to co-channel interference from S_2 and noise, can be averaged out when N_w is sufficiently large, so T_1 can get an accurate estimate of delay for S_1 when is N_w is large. We will show below that $N_w = 200$ is a suitable choice.

3.4.2 Channel gain estimation for S_1

Next, we describe estimation of a_1. Assume T_1, T_2, and the relay use the same normalized root-Nyquist filter $h(t)$, with $\int h^2(t)dt = 1$. After matched filtering, if we sample at the correct time for S_1 as determined above, then the n-th sampled value $Z_{1,n}$ is

$$Z_{1,n} = a_1 p(0)X_{1,n} + a_2 \sum_k p(\tau - kT)X_{2,n+k} + w_{1,n}$$

where $w_{1,n}$ is the Gaussian additive noise with variance $(1 + |h_{r1}W|^2)N_0$, $\tau = \tau_{s2} - \tau_{s1}$, and $p = h(t) \star h(t)$, with $p(0) = 1$ and $p(-k) = p(k)$. Since T_1 has side-information of $X_{1,n}$, we can obtain an estimate of a_1 by averaging N_w measurements. First, form normalized measurements

$$V_{1,n} = \frac{Z_{1,n}}{p(0)X_{1,n}}$$

$$= a_1 + a_2 \sum_k \frac{p(-\tau - kT)X_{2,n+k}}{X_{1,n}} + \frac{w_{1,n}}{X_{1,n}},$$

37

then form the estimator

$$\hat{a}_1 = \frac{\sum_{n=1}^{N_w} V_{1,n}}{N_w}$$

$$= a_1 + \frac{\sum_{n=1}^{N_w} \left(a_2 \sum_k \frac{p(-\tau - kT)X_{2,n+k}}{X_{1,n}} + \frac{w_{1,n}}{X_{1,n}} \right)}{N_w}$$

We observe that the estimation error of channel gain again contains additive noise as well as co-channel interference terms. These can be very significant in some cases, which degrades the quality of interference cancelation, but as N_w becomes sufficiently large, the unbiased estimate of a_1 can be as accurate as required.

3.4.3 Detection of X_2

After T_1 obtains $\hat{a}_1$ and $\hat{\tau}_{s1}$, it cancels the self-interference caused by signal S_1 (see Figure 3.16.) Assuming T_1 has the correct timing for S_1, based on the estimate $\hat{a}_1$, the interference caused by signal S_1 can be greatly reduced. After the interference cancelation, T_1 uses the similar method as for S_1 with N_p training preamble symbols to get the correct timing for S_2. Assuming T_1 obtains the correct timing for S_2, then the n-th sampled value $Z_{2,n}$ is

$$Z_{2,n} = (a_1 - \hat{a}_1) \sum_k p(\tau - kT)X_{1,n+k} + a_2 X_{2,n} + w_{2,n}$$

where $w_{2,n}$ is additive independent Gaussian noise with variance $(1 + |h_{r1}W|^2)N_0$. The detection of $X_{2,n}$ from $Z_{2,n}$ is similar to the traditional signal detection. Assuming coherent detection is used, then

$$\hat{X}_{2,n} = \frac{Z_{2,n}}{a_2 p(0)}$$

$$= X_{2,n} + \frac{(\hat{a}_1 - a_1) \sum_k p(\tau - kT)X_{1,n+k} + w_{2,n}}{a_2}$$

The second term involves delay estimation error of the previous stage and noise. The above analysis is based on the assumption that T_1 gets the correct timing of both S_1 and S_2. But practically, there will be timing errors for both S_1 and S_2. The timing error of S_1 plays the key part of the overall

performance, for it can degrade the performance of subsequent interference cancelation, channel gain estimation and timing estimation for S_2. Simulation results will be presented in next section.

3.4.4   Simulation results

In this section, we present the 'real' system performance of AF, in the context of uncoded transmission. These results illuminate the real question as to whether the theoretical AF achievable rate limits can be approached with proper modulation and coding. In all cases, we let $|a_1| = |a_2|$, SNR $= E_b/N_0$ be the ratio between the received energy, per bit, of S_2 at T_1 and the noise power density of the effective noise, including both the uplink and downlink. Both T_1 and T_2 use QPSK modulation. Root-raised-cosine pulse shaping is employed, and S_1 and S_2 are simulated at 8 samples per symbol, which means $t_0 = T/8$.

Figure 3.17 and 3.18 shows the probability of symbol error for X_2. We fix N_w for synchronization and channel estimation of S_1, and vary the number of preamble training symbols N_p among $\{20, 50, 100\}$ for S_2. As can be seen, with the proposed synchronization and channel estimation scheme, $N_w = 50$ has more than 1 dB loss in SNR as compared with ideal coherent detector, while $N_w = 200$ can achieve a much better performance where SNR loss is only about 0.5 dB. In addition, for fixed N_w there is tiny performance difference between $N_p = 20$ has $N_p = 100$, so N_w determines the symbol detection performance. Fortunately, T_1 has side-information of X_1, so increasing N_w will not introduce packet overhead, though it does increase processing complexity. For the AF protocol, the proposed signal processing procedure though simple, can achieve near optimal (ideal case) performance, and we believe that performance can be further improved by adopting more complicated synchronization and channel estimation schemes [51–54].

3.5   Summary

This Chapter summarizes two-way relay protocols, including AF, MAC-BSI, MAC-XOR, MAC-NC, and NLC, in which we focus on AF and MAC-NC, with the consideration of protocol complexity and performance. We first compare the ARR between AF and MAC-NC. Then we model and solve the maximum rate sum problem, and compare the maximum achievable rate sum between AF and MAC-NC, which is a measurement of how efficiently the allocated bandwidth is used. Both

comparisons show that AF is superior when downlink SNR is strong, which is due to the power sharing at the relay, while MAC-NC is superior when uplink SNR is strong, which is due to the MAC constraints. Then, we add an extra asymmetric rate requirement to the maximum rate sum problem, and compare the performance between AF and MAC-NC. AF protocol with the least requirement at the satellite, achieves robustly good performance for all the scenarios considered, so we will focus AF protocol in this thesis. Finally, we provide a simple signal processing procedure for the AF protocol to show its feasibility.
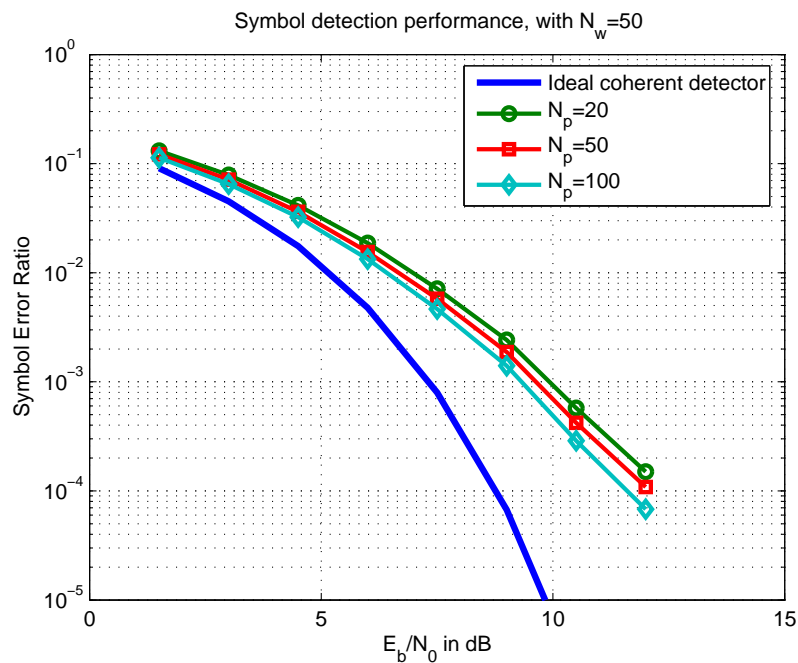
Figure 3.17: Probability of symbol error for uncoded QPSK, $N_w = 50$.
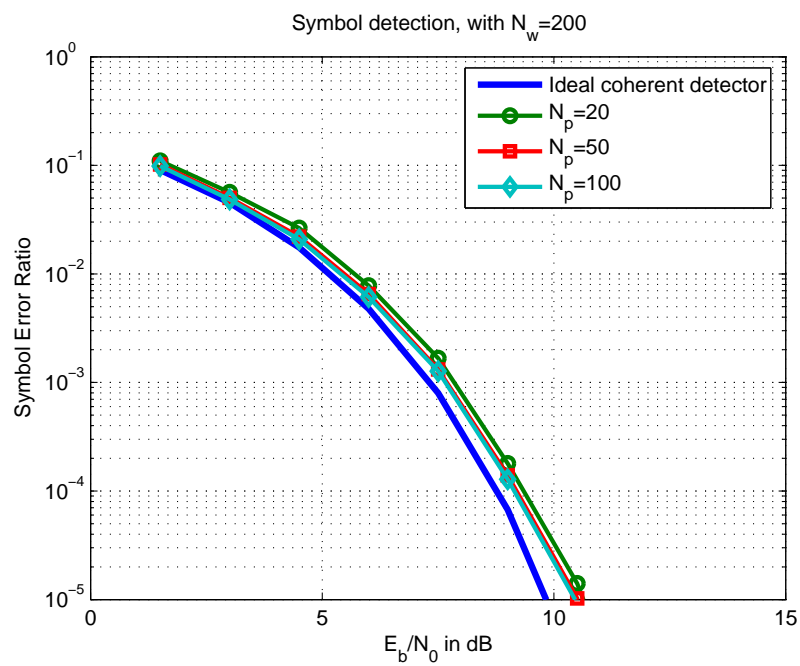
Figure 3.18: Probability of symbol error for uncoded QPSK, $N_w = 200$.

Chapter 4

Minimizing Transponder Cost for AF Protocol

In Chapter 3, we have analyzed the achievable rate region, maximum rate sum, and system complexities for different protocols, in which AF is preferred for its robust performance as well as its compatibility with legacy satellites. In the AF protocol, to achieve the *best performance* requires full utilization of the resource (bandwidth and power) at the satellite. Here we take a different view, where two terminals wish to communicate with some target rates R_{12} and R_{21} that may be well less than what the full transponder resources would support. Then, the system operator would wish to lease a fraction of a transponder, and share the remaining resources with other non-interfering services. The recurring cost for transponder services reflects the consumed resources, and it is of practical interest to minimize the recurring cost.

4.1   Problem model

The terminals T_1 and T_2 wish to exchange data, with information rates R_{12} and R_{21} respectively, using a satellite relay R. We assume both terminals are within the footprint of the satellite antenna pattern, and the four complex channel gains are h_{1r}, h_{2r}, h_{r1}, h_{r2} as before. Both uplink and downlink transmissions occur simultaneously on separate frequency bands. The available transponder bandwidth and power are B_{trans} Hz and P_{trans}. Table 4.1 lists some additional notations to Table 2.1 that will be used in this Chapter.

Notation	Definition
τ_{1r}, τ_{2r}	transmission delay from T_1 and T_2 to the relay respectively
τ_{r1}, τ_{r2}	transmission delay from the relay to T_1 (T_2)
B_s	bandwidth allocated
B_{trans}, P_{trans}	Total transponder bandwidth and power available
f_{12}, f_{21}	The information flow from T_1 to T_2, and T_2 to T_1

Table 4.1: Additional parameter definitions in Chapter 4

In satellite networks, uplinks are typically more powerful than downlinks, so we assume uplink noise may be ignored and does not consume downlink power. The satellite receives

$$Y_r(t) = h_{1r}\sqrt{P_1}S_1(t - \tau_{1r}) + h_{2r}\sqrt{P_2}S_2(t - \tau_{2r})$$

Then the relay amplifies[1] the received signal with a scaling factor of W, and broadcasts the amplified signal to both T_1 and T_2, with W satisfying the *allocated* power constraint at the relay, namely

$$|W|^2(|h_{1r}|^2 P_1 + |h_{2r}|^2 P_2) = P_r.$$

We note that P_r can be smaller than P_{trans} in fractional-transponder situations.

Without loss of generality, we consider the recovery of terminal $T_2's$ message at T_1. Assuming the time delay from relay to T_1 is τ_{r1}, then T_1 receives

$$\begin{aligned} Y_1(t) &= h_{r1}(WY_r(t - \tau_{r1})) + n_1(t) \\ &= h_{r1}Wh_{1r}\sqrt{P_1}S_1(t - (\tau_{1r} + \tau_{r1})) \\ &\quad + h_{r1}Wh_{2r}\sqrt{P_2}S_2(t - (\tau_{2r} + \tau_{r1})) + n_1(t) \end{aligned}$$

Assuming perfect synchronization, channel estimation, and interference cancelation, T_1 obtains

---

[1] We assume a linear amplifier at the satellite, typically necessitating the power amplifier be operated in backed-off condition

$$Z_1(t) = h_{r1}Wh_{2r}\sqrt{P_2}S_2(t - (\tau_{2r} + \tau_{r1})) + n_1(t)$$

Similarly, T_2 obtains

$$Z_2(t) = h_{r2}Wh_{1r}\sqrt{P_1}S_1(t - (\tau_{2r} + \tau_{r2})) + n_2(t)$$

So the rate constraint as a function of allocated power P_r and allocated bandwidth B_s for flow f_{21} is, from an information theory perspective [2]

$$R_{21} \leq B_s \ \log(1 + \frac{|h_{r1}|^2 W^2 |h_{2r}|^2 P_2}{N_1 B_s}) \quad \text{bits/sec}$$

$$= B_s \ \log(1 + \frac{\frac{|h_{2r}|^2 P_2 |h_{r1}|^2 P_r}{|h_{1r}|^2 P_1 + |h_{2r}|^2 P_2}}{N_1 B_s}) \quad \text{bits/sec}$$

Similarly, for T_2, we have

$$R_{12} \leq B_s \ \log(1 + \frac{\frac{|h_{1r}|^2 P_1 |h_{r2}|^2 P_r}{|h_{1r}|^2 P_1 + |h_{2r}|^2 P_2}}{N_2 B_s}) \quad \text{bits/sec}$$

Note that both flows are sharing the power P_r, with fraction determined by uplink power control (see below). We assume the satellite lease cost is represented by

$$C = f(max\{\frac{B_s}{B_{trans}}, \ \lambda\frac{P_r}{P_{trans}}\}),$$

where $f(.)$ is a monotone-increasing function, and λ is the ratio of power price multiplier to bandwidth price multiplier. A typical case is when $\lambda = 1$, in which the transponder cost is actually determined by the maximum of the allocated bandwidth and the power equivalent bandwidth [55]. The fundamental problem is:

Given rate requirements R_{12} and R_{21}, total transponder power P_{trans}, bandwidth B_{trans}, and effective channel gain vector $(h_{1r}, \ h_{2r}, \ h_{r1}, \ h_{r2})$, what are the optimal bandwidth B_s (or symbol

---

[2]We make the assumption that the occupied Fourier bandwidth equals the modulator symbol rate

rate R_s) and power P_r allocated so that the transponder cost is minimized?

4.2 Optimal resource allocation

This section presents a solution to minimize the transponder leasing cost. We first assume that the available transponder resources are adequate to satisfy the rate requirement (The exception is treated later). Let

$$P_{r1} = \frac{|h_{2r}|^2 P_2}{|h_{1r}|^2 P_1 + |h_{2r}|^2 P_2} P_r,$$

$$P_{r2} = \frac{|h_{1r}|^2 P_1}{|h_{1r}|^2 P_1 + |h_{2r}|^2 P_2} P_r,$$

where P_{r1} and P_{r2} are the satellite powers allocated by power sharing to flow f_{21} and f_{12} respectively. So from above

$$R_{21} \leq B_s \ \log(1 + \frac{|h_{r1}|^2 P_{r1}}{N_1 B_s}) \quad bits/second$$

$$R_{12} \leq B_s \ \log(1 + \frac{|h_{r2}|^2 P_{r2}}{N_2 B_s}) \quad bits/second$$

$$P_r = P_{r1} + P_{r2}$$

Let

$$g_1 = \frac{|h_{r1}|^2}{N_1}, \ g_2 = \frac{|h_{r2}|^2}{N_2}$$

which measure two downlinks' quality, including receiver noise level. Then we have

$$R_{21} \leq B_s \ \log(1 + \frac{g_1 P_{r1}}{B_s}) \quad bits/second$$

$$R_{12} \leq B_s \ \log(1 + \frac{g_2 P_{r2}}{B_s}) \quad bits/second$$

which give us

$$P_{r1} = B_s \frac{2^{R_{21}/B_s} - 1}{g_1}$$

$$P_{r2} = B_s \frac{2^{R_{12}/B_s} - 1}{g_2}$$

$$P_r = B_s \left(\frac{2^{R_{21}/B_s} - 1}{g_1} + \frac{2^{R_{12}/B_s} - 1}{g_2} \right) \tag{4.1}$$

In order to minimize the total cost, the usage of transponder bandwidth and power should be 'balanced' so that in the $max(\cdot, \cdot)$ expression of the cost function, the two arguments are equal, i.e.,

$$\frac{B_s}{B_{trans}} = \lambda \frac{P_r}{P_{trans}}, \tag{4.2}$$

otherwise we can reduce the transponder cost by decreasing B_s and increasing P_r, or vice-versa, while maintaining the information rate. So we require

$$\frac{P_r}{B_s} = \frac{1}{\lambda} \frac{P_{trans}}{B_{trans}} \triangleq \gamma$$

where γ reflects total power over total bandwidth available, a summary statistic for the transponder, as well as the relative pricing coefficient λ. This implies

$$\frac{2^{R_{21}/B_s} - 1}{g_1} + \frac{2^{R_{12}/B_s} - 1}{g_2} = \gamma \tag{4.3}$$

With given rate requirements $\{R_{21}, R_{12}\}$, total transponder power P_{trans} and bandwidth B_{trans}, we may solve (4.3) to get the optimal symbol rate B_s^*, which minimizes the total cost of using the transponder. Then we obtain the optimal power allocation P_r^* by

$$P_r^* = \gamma B_s^* \tag{4.4}$$

If $B_s^* > B_{trans}$ or $P_r^* > P_{trans}$ [3], then either the available bandwidth or the available power is not

---

[3]Since we have assumed that there is enough transponder resource, $B_s^* > B_{trans}$ and $P_r^* > P_{trans}$ can't happen at the same time.

adequate to met the rate requirement, while maintaining the cost-optimality of (4.2). The solution is to set $B_s^* = B_{trans}$ (or $P_r^* = P_{trans}$), and solve (4.1) to get P_r^* (or B_s^*).

After obtaining B_s^* and P_r^*, the power allocated to each flow can be calculated as below:

$$P_{r1}^* = B_s^* \frac{2^{R_{21}/B_s^*} - 1}{g_1} \tag{4.5}$$

$$P_{r2}^* = B_s^* \frac{2^{R_{12}/B_s^*} - 1}{g_2} \tag{4.6}$$

Note that with fixed P_r, the power allocations P_{r1} and P_{r2} can only be managed by uplink power management. To achieve P_{r1}^*, P_{r2}^*, the transmission power P_1 and P_2 at earth terminals should be adjusted accordingly. Since

$$P_{r1}^* = \frac{|h_{2r}|^2 P_2}{|h_{1r}|^2 P_1 + |h_{2r}|^2 P_2} P_r^*$$

$$P_{r2}^* = \frac{|h_{1r}|^2 P_1}{|h_{1r}|^2 P_1 + |h_{2r}|^2 P_2} P_r^*,$$

we have

$$\begin{aligned}
\frac{P_2}{P_1} &= \frac{|h_{1r}|^2}{|h_{2r}|^2} \frac{P_{r1}^*}{P_{r2}^*} \\
&= \frac{|h_{1r}|^2}{|h_{2r}|^2} \frac{B_s^* \frac{2^{R_{21}/B_s^*}-1}{g_1}}{B_s^* \frac{2^{R_{12}/B_s^*}-1}{g_2}} \\
&= \frac{|h_{1r}|^2}{|h_{2r}|^2} \frac{g_2}{g_1} \frac{2^{R_{21}/B_s^*} - 1}{2^{R_{12}/B_s^*} - 1} \\
&= \frac{|h_{1r}|^2 |h_{r2}|^2 N_1}{|h_{2r}|^2 |h_{r1}|^2 N_2} \frac{2^{R_{21}/B_s^*} - 1}{2^{R_{12}/B_s^*} - 1}
\end{aligned} \tag{4.7}$$

If we assume T_1 and T_2 have the same noise temperature, $N_1 = N_2$, and if the effective channels are symmetric, $|h_{1r}|^2 = |h_{r1}|^2$, $|h_{2r}|^2 = |h_{r2}|^2$, then we have[4]

---

[4]In addition to the constraint on the ratio P_1/P_2, P_1 and P_2 need to be large enough so that the uplink noise can be neglected as compared with downlink noise

$$\frac{P_2}{P_1} = \frac{2^{R_{21}/B_s^*} - 1}{2^{R_{12}/B_s^*} - 1}.$$

Figure 4.1 presents a contour plot for the achievable rate regions for R_{21} and R_{12} for different cost, expressed as a fraction of full transponder cost, when $P_{trans} = 50\ W$, $B_{trans} = 50\ MHz$, the cost function $f(.)$ is linear and $\lambda = 1$, with $g_1 = 10^6$ and $g_2 = 10^7$, which are all typical satellite system parameters. Such contour plots can help engineers to estimate the transponder cost for specific rate requirements, or find the most suitable rate pairs while keeping the transponder cost below some threshold. We observe that the 100% contour in Figure 4.1 is the achievable rate region for this scenario when a full transponder is utilized, under the assumption of strong uplinks.
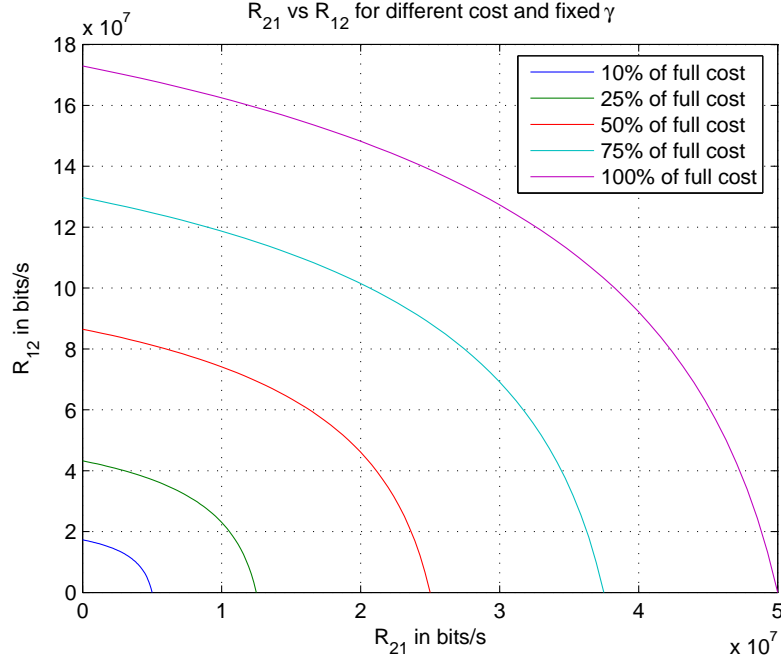
Figure 4.1: Achievable rate region for different transponder cost,
with $P_{trans} = 50\ W$, $B_{trans} = 50\ MHz$ $(\gamma = 10^{-6})$, $\lambda = 1$, $g_1 = 10^6$ and $g_2 = 10^7$

Figure 4.2 presents the procedure described above in flow chart form. Note that, we have previously assumed that the transponder is able to satisfy the information rate requirement. Practically, there can be situations in which transponder resources are not adequate. Then, T_1 and T_2 need to either reduce the transmission rate or system redesign needs to occur to increase g_1 and/or g_2. Possible rate reduction schemes include but are not limited to:

- Keep R_{12} (or R_{21}) and reduce R_{21} (or R_{12}) (Case 1 (or 2) in Figure 4.3).

- Reduce both R_{12} and R_{21} proportionally (Case 3 in Figure 4.3).

- Achieve the maximum rate sum of the two information flow (Case 4 in Figure 4.3)
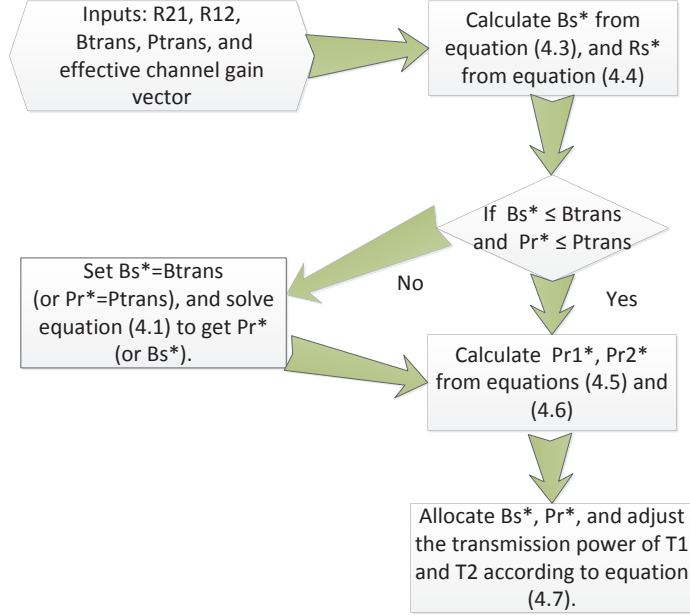
Figure 4.2: The procedure of minimizing the transponder cost

4.3 Multiple flow/multiple transponder extension

We can extend the analysis above to situations where the two terminals wish to exchange multiple traffic streams, say a video feed and a high-rate data channel. We assume these flows cannot be divided and merged, so that exactly one transponder may be consumed for example, and the remainder of the traffic be carried by another transponder. These flows may be accommodated within one transponder, or it may be that extra transponder resources must be invoked.

Suppose there are N transponders, $TR^{(1)}, TR^{(2)}, ..., TR^{(N)}$, with $B_{trans}^{(1)}, B_{trans}^{(2)}, ...B_{trans}^{(N)}$ as transponder bandwidths, $P_{trans}^{(1)}, P_{trans}^{(2)}, ..., P_{trans}^{(N)}$ as transponder powers, and $f_{(1)}, f_{(2)}, ..., f_{(N)}$ as the cost function constants, and there are L information flows $f_{12}^{(1)}, f_{12}^{(2)}, ..., f_{12}^{(L)}$ (from T_1 to T_2), and M information flows $f_{21}^{(1)}, f_{21}^{(2)}, ..., f_{21}^{(M)}$ (from T_2 to T_1) with rate requirements $R_{12}^{(1)}, R_{12}^{(2)}, ..., R_{12}^{(L)}$, $R_{21}^{(1)}, R_{21}^{(2)}, ..., R_{21}^{(M)}$ respectively, then the optimization problem can be formulated as
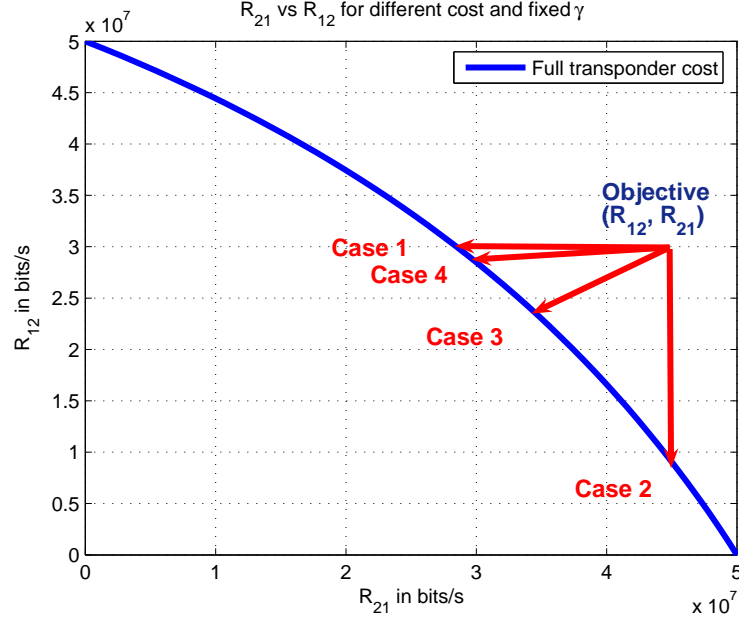
Figure 4.3: Rate adjusting for R_{21} and R_{12}, when resource is not enough, with $P_{trans} = 50\ W$, $B_{trans} = 50\ MHz$ ($\gamma = 10^{-6}$), $\lambda = 1$, $g_1 = 10^6$ and $g_2 = 10^6$.

$$\underset{B_s^{(j)},\ P_r^{(j)},\ j=1,...,N}{\text{minimize}} \quad \sum_{j=1}^{N} f_{(j)}(max\{\frac{B_s^{(j)}}{B_{trans}^{(j)}},\ \lambda\frac{P_r^{(j)}}{P_{trans}^{(j)}}\})$$

subject to:

$$P_r^{(j)} \le P_{trans}^{(j)}, \quad B_s^{(j)} \le B_{trans}^{(j)},\ \forall\ 1 \le j \le N$$

$$P_{r1}^{(j)} \ge 0,\ P_{r2}^{(j)} \ge 0, \quad P_{r1}^{(j)} + P_{r2}^{(j)} = P_r^{(j)},\ \forall\ 1 \le j \le N$$

$$x_{12}^{(i,j)} = 0\ or\ 1, \quad x_{21}^{(i,j)} = 0\ or\ 1,\ \forall\ 1 \le i \le L,\ 1 \le j \le N$$

$$\sum_{j=1}^{N} x_{12}^{(i,j)} = 1,\ \forall\ 1 \le i \le L, \quad \sum_{j=1}^{N} x_{21}^{(i,j)} = 1,\ \forall\ 1 \le i \le M$$

$$\sum_{i=1}^{L} x_{12}^{(i,j)} R_{12}^{(i)} \le B_s^{(j)} \log(1 + \frac{g_2^{(j)} P_{r2}^{(j)}}{B_s^{(j)}}),\ \forall\ 1 \le j \le N$$

$$\sum_{i=1}^{M} x_{21}^{(i,j)} R_{21}^{(i)} \le B_s^{(j)} \log(1 + \frac{g_1^{(j)} P_{r1}^{(j)}}{B_s^{(j)}}),\ \forall\ 1 \le j \le N$$

where $x_{12}^{(i,j)}$ ($x_{21}^{(i,j)}$) is an indicator function, equalling 1 when f_{12}^i (f_{21}^i) is assigned to transponder $TR^{(j)}$. If $\lambda = 1$, then the cost-optimal solution must satisfy

$$\frac{B_s^{(j)}}{B_{trans}^{(j)}} = \frac{P_r^{(j)}}{P_{trans}^{(j)}},\ \forall\ 1 \le j \le N$$

The above optimization problem can be further reduced to

$$\underset{B_s^{(j)},\ j=1,\ldots,N}{\text{minimize}} \quad \sum_{j=1}^{N} f_{(j)}\Big(\frac{B_s^{(j)}}{B_{trans}^{(j)}}\Big)$$

subject to:

$$B_s^{(j)} \leq B_{trans}^{(j)},\ \forall\ 1 \leq j \leq N$$

$$P_{r1}^{(j)} + P_{r2}^{(j)} = \gamma^{(j)} B_s^{(j)},\ \forall\ 1 \leq j \leq N$$

$$x_{12}^{(i,j)} = 0\ or\ 1,\ \forall\ 1 \leq i \leq L,\ 1 \leq j \leq N$$

$$x_{21}^{(i,j)} = 0\ or\ 1,\ \forall\ 1 \leq i \leq M,\ 1 \leq j \leq N$$

$$\sum_{j=1}^{N} x_{12}^{(i,j)} = 1,\ \forall\ 1 \leq i \leq L, \quad \sum_{j=1}^{N} x_{21}^{(i,j)} = 1,\ \forall\ 1 \leq i \leq M$$

$$\frac{2^{\sum_{i=1}^{L} x_{12}^{(i,j)} R_{12}^{(i)}/B_s^{(j)}} - 1}{g_2^{(j)}} + \frac{2^{\sum_{i=1}^{M} x_{21}^{(i,j)} R_{21}^{(i)}/B_s^{(j)}} - 1}{g_1^{(j)}} = \gamma^{(j)}, \quad \forall\ 1 \leq j \leq N$$

where

$$\gamma^{(j)} = \frac{P_{trans}^{(j)}}{B_{trans}^{(j)}},\ j = 1,\ldots,N.$$

The optimization problem modeled above is a mixed interger/non-linear programming problem (MINLP), which is generally hard to solve, but practically there are a limited of flows to accommodate and the optimal solution can be quickly achieved via numerical methods.

4.4 Summary

This Chapter focuses on minimizing the transponder cost for the amplify-forward protocol on a two-way relay satellite channel, which is modeled as a joint bandwidth and power optimization problem. A solution to this problem is given that determines the most efficient bandwidth and power resource requests to reduce the transponder cost. We also consider a more general problem when multiple traffic flows are to be supported, perhaps across multiple transponders. This leads to a mixed integer/non-linear programming problem.

The work in this Chapter is an information-theoretic formulation, but its solution leads directly to proper choices for modulation and coding, once the optimal signaling rates and then bits/symbol are identified. Allowance can, and should, be made for operation that falls short of channel capacity limits in terms of SNR, and also for Fourier bandwidth that exceeds the signaling rate by, say, 20%.

Chapter 5

Extension to Multiple-user Satellite Exchange

We next study data exchange in a three-terminal satellite network involving a hub terminal and two remote terminals, as shown in Figure 5.1. The hub terminal T_1 is typically a large earth station, and the two remote terminals T_2 and T_3 are typically very small aperture terminals (VSAT's). These terminals are connected only via the satellite R. We assume that T_1, T_2, and T_3 are located in the same footprint of the satellite, and full duplexing is achieved by use of separate transmit and receive bands at each terminal.

The three-terminal problem we consider is an extension of the two-terminal exchange channel [2, 5, 7]. There it is known that non-orthogonal signaling protocols offer up to double the spectral efficiency over orthogonal data exchange (either through TDMA or FDMA). Among these protocols, amplify-forward (AF) is preferred for a satellite communication, for AF does not require sophisticated signal processing at the relay and is compatible with legacy satellites. Commercial implementations of this technique have been produced, e.g. DoubleTalk Carrier-in-Carrier [55]. This Chapter will focus on protocols which are based on the AF protocol.

These improvements of AF in two-way relay channel derive mainly from the non-orthogonal signaling on uplinks (to the relay), and interference cancellation based on side information at receive terminals. However, as far as we know, little work has been done on a three (or more) terminal extension of the similar data exchange problem. This Chapter studies the three-terminal
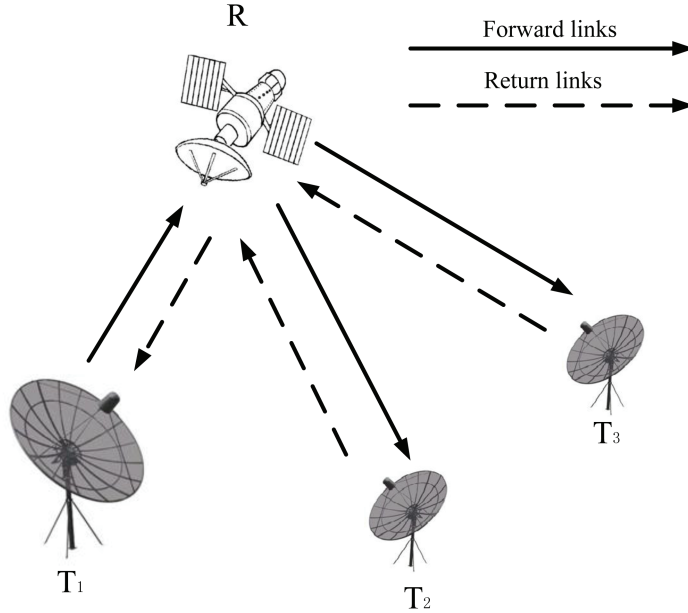
Figure 5.1: A three-terminal data exchange model

exchange channel as a first step to study the more general multiple terminal exchange channel. Fundamental questions we want to answer include: Can non-orthogonal signaling with partial self-interference cancellation bring as much efficiency improvement as what we have seen in the two-way relay channel? What decoding protocols yield the largest achievable rate regions (assuming amplify-forward at the relay)?

We study two cases of data exchange: case 1 (three flows), T_1 sends a *common* message (or *common* traffic) to both T_2 and T_3, whereas T_2 and T_3 send separate traffic to T_1; case 2 (four flows), T_1 sends *separate* messages to T_2 and T_3. We present the information rate constraints for each of the protocols and decoding schemes discussed, followed by numerical results. Performance metrics studied include achievable rate region (ARR) and maximum rate sum. Numerical results show that for ARR, three-user simultaneous signaling outperforms time-shared two-way AF for the case of three information flows, especially when remote terminal rate is a priority. However, this advantage holds only for the four information flows case when the hub terminal has much higher link quality than remote terminals. As for the maximum rate sum, time-shared two-way AF performs competitively in all the scenarios studied, though not always best.

5.1 Problem model

In the model of Figure 5.1, there are at most four information flows: f_{12}, f_{13} which are forward flows, and f_{21}, f_{31} which are return flows. (No data exchange is desired between T_2 and T_3.) The relay terminal is assumed to be a linear amplifier with average power constraint P_r. We assume additive Gaussian noise at all receiver terminals, and full-duplexing is achieved at the relay by dividing the total spectrum of bandwidth $2B$ into two disjoint bands B to both uplinks ($T_1 \rightarrow R$, $T_2 \rightarrow R$, $T_3 \rightarrow R$) and downlinks ($R \rightarrow T_1$, $R \rightarrow T_2$, $R \rightarrow T_3$). For convenience, Table 5.1 lists notation additional to Table 2.1 that will be used in this Chapter.

Notation	Definition
P_3	maximum power at T_3
P_{12}, P_{13}	power of the two signals generated at T_1, with $P_1 = P_{12}+P_{13}$ (four flows case)
X_3	data sequences sent from T_3
X_{12}, X_{13}	data sequences sent from T_1 to T_2 and T_3 respectively
$f_{12}, f_{13}, f_{21}, f_{31}$	information flows between terminals
R_{13}, R_{31}	information rate between terminals
f_1, R_1	common information flow generated at T_1 with its information rate
h_{3r}, h_{r3}	effective channel gains between terminals and the relay
n_3	Gaussian noise process at T_3
η_3	noise spectral density at T_3, with $\eta_1 = \eta_2 = \eta_3 = \eta_r = \eta$

Table 5.1: Additional parameter definitions in Chapter 5

As in previous Chapters, it is convenient to define the set of link SNR constraints as

$$\rho_{3r} = \frac{|h_{3r}|^2 P_3}{\eta_r B}, \quad \rho_{r3} = \frac{|h_{r3}|^2 P_r}{\eta_3 B}$$

$$\rho_{12r} = \frac{|h_{1r}|^2 P_{12}}{\eta_r B}, \quad \rho_{13r} = \frac{|h_{1r}|^2 P_{13}}{\eta_r B}, \quad \text{with} \quad \rho_{1r} = \rho_{12r} + \rho_{13r}$$

as all achievable rates can be expressed in these terms.

5.2 Hub terminal broadcasts common data

One special case of three-user data exchange involves T_1 broadcasting a common message to both T_2 and T_3, while T_2 and T_3 each send their own message to T_1. There are thus three information flows for this case, denoted f_1, f_{21}, and f_{31}. We can identify three signalling protocols for the three-flow case. The first two are time-sharing extensions of the two-user protocol studied early.

5.2.1 Time-shared protocol 1

In this protocol, data exchange is time-shared between two communication sessions $T_1 \leftrightarrow T_2$ and $T_1 \leftrightarrow T_3$. For each session, the amplify-forward protocol is used, and one of the remote terminals is silent, as shown in Figure 5.2.
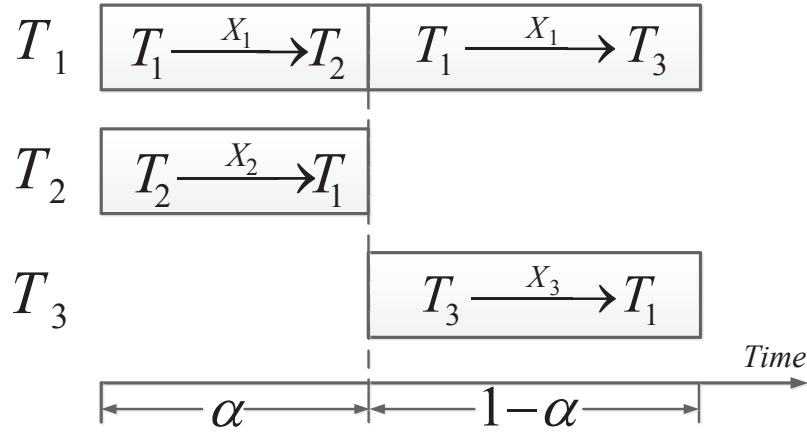
Figure 5.2: Time-shared protocol 1 for three-flow case

We assume perfect interference cancellation of self-interference in each of the two-user exchanges, exploiting side-information that makes the two-terminal exchange protocol efficient. For session $T_1 \leftrightarrow T_2$, we have, as seen earlier

$$R_{12} \leq \frac{1}{2} \log(1 + \frac{\rho_{1r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r2}}) \quad \text{bps/Hz}$$

$$R_{21} \leq \frac{1}{2} \log(1 + \frac{\rho_{2r}\rho_{r1}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r1}})$$

(Information rate is normalized by $2B$ Hz)

55

There are similar information rate constraints imposed on R_{13} and R_{31} by T_3. These rates would then be multiplied by α and $1 - \alpha$ respectively to obtain achievable rate combinations.

5.2.2 Time-shared protocol 2

In this protocol, T_2 and T_3 again time-share for *return* links, but T_1 sends the *common* forward message only once with information rate R_1. Now T_2 (T_3) experiences two decoding scenarios. While T_2 (T_3) is transmitting, it can use interference cancellation to process T_1's signal, but while T_3 (T_2) is transmitting, T_2 (T_3) sees T_3's (T_2's) signal competing with the desired message from T_1, as shown in Figure 5.3.
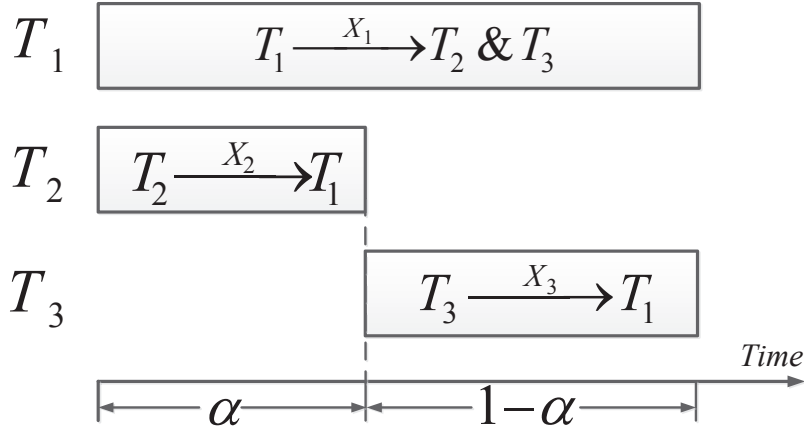
Figure 5.3: Time-shared protocol 2 for three-flow case

Take the the interval when T_1 and T_2 are transmitting as an example. We first have the information rate constraints of traditional amplify-forward scheme listed below

$$R_1 \leq \frac{1}{2} \log(1 + \frac{\rho_{1r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r2}}) \quad \text{bps/Hz}$$

$$R_{21} \leq \frac{1}{2} \log(1 + \frac{\rho_{2r}\rho_{r1}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r1}})$$

These are the rate constraints associated with a two-terminal AF exchange protocol, identical to those above. However, T_3 also needs to decode the forward link common information during this

session, and this places additional rate constraints on R_1 and R_{21}, depending on signal processing at terminal T_3, as now developed. T_3 receives in complex baseband notation[1]

$$Y_3 = h_{r3}W(h_{1r}\sqrt{P_1}X_1 + h_{2r}\sqrt{P_2}X_2 + n_r) + n_3$$
$$= h_{r3}Wh_{1r}\sqrt{P_1}X_1 + h_{r3}Wh_{2r}\sqrt{P_2}X_2 + h_{r3}Wn_r + n_3$$

where

$$W = \sqrt{\frac{P_r}{|h_{1r}|^2 P_1 + |h_{2r}|^2 P_2 + \eta_r B}}$$

is the amplifying factor at the relay, chosen to satisfy the power constraint P_r. We observe that T_3 cannot perform an interference cancellation, in contrast with T_2. T_3 has two possible decoding options, so there are actually two information rate constraint sets at T_3.

- Option 1: T_3 views T_2's signal as noise;

- Option 2: T_3 jointly decodes X_1 and X_2 as in a traditional 2-user MAC.

For the former option, we have

$$\begin{bmatrix} & view\ X_2\ as\ noise & \end{bmatrix}$$
$$R_1 \le \frac{1}{2}\log(1 + \frac{\rho_{1r}\rho_{r3}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r3} + \rho_{2r}\rho_{r3}}) \quad \text{bps/Hz}$$

For the latter option, we have

$$\begin{bmatrix} & joint\ decoding & \end{bmatrix}$$
$$R_1 \le \frac{1}{2}\log(1 + \frac{\rho_{1r}\rho_{r3}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r3}})$$
$$R_{21} \le \frac{1}{2}\log(1 + \frac{\rho_{2r}\rho_{r3}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r3}})$$
$$R_1 + R_{21} \le \frac{1}{2}\log(1 + \frac{(\rho_{1r} + \rho_{2r})\rho_{r3}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r3}})$$

---

[1]We suppress propagation delay in this and subsequent expressions for received signals, with no loss of generality.

The achievable rate region is the set of rate pairs (R_1, R_{21}) for which constraints at T_1, T_2, and T_3 are all satisfied (at least one of the constraints sets at T_3 needs to be satisfied). Similar constraints apply for the other time-sharing session, and by adjusting the return link time-sharing fractions, we may achieve certain rate triples (R_1, R_{21}, R_{31}).

5.2.3 Three-simultaneous signaling

In this protocol, T_1, T_2, T_3 all send *simultaneously*, with no time-sharing between T_2 and T_3, as shown in Figure 5.4.
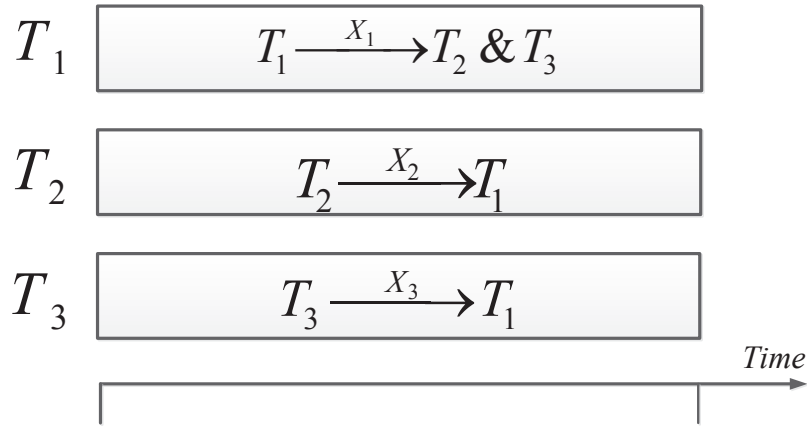
Figure 5.4: Three-simultaneous signalling protocol for three-flow case

The relay receives the discrete-time complex baseband signal

$$Y_r = h_{1r}\sqrt{P_1}X_1 + h_{2r}\sqrt{P_2}X_2 + h_{3r}\sqrt{P_3}X_3 + n_r$$

and broadcasts

$$X_r = W(h_{1r}\sqrt{P_1}X_1 + h_{2r}\sqrt{P_2}X_2 + h_{3r}\sqrt{P_3}X_3 + n_r)$$

where W is a signal amplifying factor, satisfying

$$W = \sqrt{\frac{P_r}{|h_{1r}|^2 P_1 + |h_{2r}|^2 P_2 + |h_{3r}|^2 P_3 + \eta_r B}}$$

T_1 processing:

The hub terminal T_1 receives

$$Y_1 = h_{r1}X_r + n_1$$
$$= h_{r1}W(h_{1r}\sqrt{P_1}X_1 + h_{2r}\sqrt{P_2}X_2 + h_{3r}\sqrt{P_3}X_3 + n_r) + n_1$$

After assumed perfect interference cancellation of T_1's own signal, T_1 receives

$$Y_1' = h_{r1}W(h_{2r}\sqrt{P_2}X_2 + h_{3r}\sqrt{P_3}X_3 + n_r) + n_1$$

Applying the standard constraints for capacity of a bandlimited two-user Gaussian multiple access channel and normalizing by total bandwidth $2B$, we have

$$R_{21} = \frac{1}{2}\log(1 + \frac{\rho_{2r}\rho_{r1}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r1}}) \quad \text{bps/Hz}$$
$$R_{31} \leq \frac{1}{2}\log(1 + \frac{\rho_{3r}\rho_{r1}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r1}})$$
$$R_{21} + R_{31} \leq \frac{1}{2}\log(1 + \frac{(\rho_{2r} + \rho_{3r})\rho_{r1}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r1}})$$

We observe these constraints are coupled with the constraint on R_1 through the SNRs' ρ_{1r} and ρ_{r1}.

T_2 processing:

The remote terminal T_2 receives, after interference cancellation of its own uplink signal,

$$Y_2' = h_{r2}W(h_{1r}\sqrt{P_1}X_1 + h_{3r}\sqrt{P_3}X_3 + n_r) + n_1$$

Similar to the time sharing protocol 2, T_2 has two options: view interference as noise or decode a 2-user MAC, so this defines two constraint sets:

$$[\quad view\ X_3\ as\ noise\quad]$$
$$R_1 \leq \frac{1}{2}\log(1 + \frac{\rho_{1r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r2} + \rho_{3r}\rho_{r2}})$$

or

$$[\quad joint \; decoding \quad]$$

$$R_1 \leq \frac{1}{2} \log(1 + \frac{\rho_{1r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r2}})$$

$$R_{31} \leq \frac{1}{2} \log(1 + \frac{\rho_{3r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r2}})$$

$$R_1 + R_{31} \leq \frac{1}{2} \log(1 + \frac{(\rho_{1r} + \rho_{3r})\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r2}})$$

In the second case, we simply discard the message sent by T_3 and T_1. There are similar information rate constraints imposed on R_1 and R_{21} by T_3. The set of all these constraints defines an achievable rate region in 3-D space, for each of the decoding options.

5.3 Hub sends different messages

A more general case is when T_1 sends different traffic to T_2 and T_3, which means there are four information flows, $f_{12}, f_{13}, f_{21}, f_{31}$. Effectively there are two separate two-user exchange channels sharing a common relay, but the third terminal is now 'competing' for resources.

5.3.1 Time-sharing protocol

As with the three-flow problem, we may extend the two-user exchange protocol by adding time sharing. We define two time-sharing communication sessions $T_1 \leftrightarrow T_2$ and $T_1 \leftrightarrow T_3$, occupying time fractions α and $1 - \alpha$ respectively, as shown in Figure 5.5.

T_2 (T_3) no longer experiences interference from the other remote in its own session. Also relay power sharing is between only two terminals. For session $T_1 \leftrightarrow T_2$, we have the information rate constraints for amplify-forward on a two-way relay channel as

$$R_{12} \leq \frac{1}{2} \log(1 + \frac{\rho_{1r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r2}}) \qquad \text{bps/Hz}$$

$$R_{21} \leq \frac{1}{2} \log(1 + \frac{\rho_{2r}\rho_{r1}}{1 + \rho_{1r} + \rho_{2r} + \rho_{r1}})$$
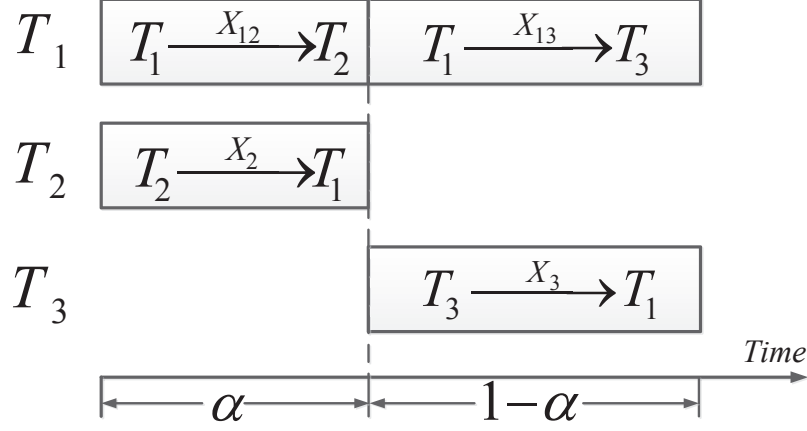
Figure 5.5: Time-shared protocol for four-flow case

We have similar results for session $T_1 \leftrightarrow T_3$. The achievable rate region (now in 4-D) is determined by these rate constraints as well as the time-sharing factor α.

5.3.2 Three-simultaneous signaling

Here, as before, all three terminals send simultaneously, but now T_1 superposes two codewords X_{12} and X_{13} [7], with power P_{12} and P_{13} respectively ($P_1 = P_{12} + P_{13}$), and the remote terminals encode and modulate their own traffic for the hub. The relay amplifies the received signal, which is a sum of three signals (four codewords), and broadcasts it to T_1, T_2 and T_3, as shown in Figure 5.6.
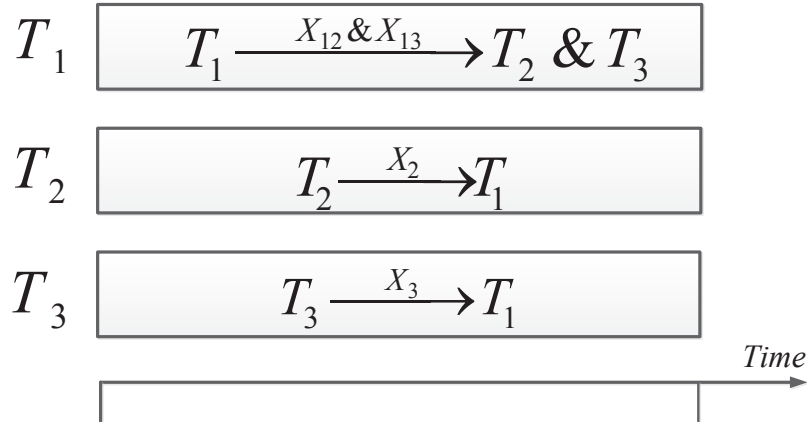
Figure 5.6: Three simultaneous signalling protocol for four-flow case

61

T$_1$ processing:

T_1, after cancelling its self-interfering signal, obtains

$$Y_1' = h_{r1}W(h_{2r}\sqrt{P_2}X_2 + h_{3r}\sqrt{P_3}X_3 + n_r) + n_1$$
$$= h_{r1}Wh_{2r}\sqrt{P_2}X_2 + h_{r1}Wh_{3r}\sqrt{P_3}X_3 + h_{r1}Wn_r + n_1$$

where

$$W = \sqrt{\frac{P_r}{|h_{1r}|^2 P_{12} + |h_{1r}|^2 P_{13} + |h_{2r}|^2 P_2 + |h_{3r}|^2 P_3 + \eta B}}$$

T_1 decodes a traditional 2-user MAC, so the return-link flows must satisfy

$$R_{21} \leq \frac{B}{2B}\log(1 + \frac{|h_{r1}|^2 W^2 |h_{2r}|^2 P_2}{|h_{r1}|^2 W^2 \eta_r B + \eta_1 B})$$
$$\leq \frac{1}{2}\log(1 + \frac{\rho_{2r}\rho_{r1}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r1}})$$
$$R_{31} \leq \frac{1}{2}\log(1 + \frac{\rho_{3r}\rho_{r1}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r1}})$$
$$R_{21} + R_{31} \leq \frac{1}{2}\log(1 + \frac{(\rho_{2r} + \rho_{3r})\rho_{r1}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r1}})$$

T$_2$ processing:

T_2, after cancelling its self-interfering signal, obtains

$$Y_2' = h_{r2}W(h_{1r}\sqrt{P_{12}}X_{12} + h_{1r}\sqrt{P_{13}}X_{13} + h_{3r}\sqrt{P_3}X_3 + n_r) + n_2$$
$$= h_{r2}Wh_{1r}\sqrt{P_{12}}X_{12} + h_{r2}Wh_{1r}\sqrt{P_{13}}X_{13}$$
$$+ h_{r2}Wh_{3r}\sqrt{P_3}X_3 + h_{r2}Wn_r + n_2$$

T_2 is interested in X_{12}, and can view X_{13} and X_3 as noise, in which case we have:

$$[\quad view\ both\ X_{13}\ and\ X_3\ as\ noise\quad]$$
$$R_{12} \leq \frac{1}{2}\log(1 + \frac{\rho_{12r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{13r}\rho_{r2} + \rho_{3r}\rho_{r2} + \rho_{r2}})$$

However, viewing other signals as noise is sometimes pessimistic, especially when these signals carry low-rate messages. Another option for T_2 is to view X_3 as noise, and jointly decode both X_{12} and X_{13} on a 2-user MAC channel. We have

$$[\quad view\ X_3\ as\ noise \quad]$$

$$R_{12} \leq \frac{1}{2} \log(1 + \frac{\rho_{12r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{3r}\rho_{r2} + \rho_{r2}})$$

$$R_{13} \leq \frac{1}{2} \log(1 + \frac{\rho_{13r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{3r}\rho_{r2} + \rho_{r2}})$$

$$R_{12} + R_{13} \leq \frac{1}{2} \log(1 + \frac{\rho_{1r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{3r}\rho_{r2} + \rho_{r2}})$$

T_2 can also view X_{13} as noise, and jointly decode both X_{12} and X_3 via a 2-user MAC channel. We have

$$[\quad view\ X_{13}\ as\ noise \quad]$$

$$R_{12} \leq \frac{1}{2} \log(1 + \frac{\rho_{12r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{13r}\rho_{r2} + \rho_{r2}})$$

$$R_{31} \leq \frac{1}{2} \log(1 + \frac{\rho_{3r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{13r}\rho_{r2} + \rho_{r2}})$$

$$R_{12} + R_{31} \leq \frac{1}{2} \log(1 + \frac{(\rho_{12r} + \rho_{3r})\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{13r}\rho_{r2} + \rho_{r2}})$$

Finally, T_2 can jointly decode X_{12}, X_{13}, X_3 via a 3-user MAC, and we have

$$[\quad decode\ three\text{-}user\ MAC \quad]$$

$$R_{12} \leq \frac{1}{2}\log(1 + \frac{\rho_{12r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r2}})$$

$$R_{13} \leq \frac{1}{2}\log(1 + \frac{\rho_{13r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r2}})$$

$$R_{31} \leq \frac{1}{2}\log(1 + \frac{\rho_{3r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r2}})$$

$$R_{12} + R_{13} \leq \frac{1}{2}\log(1 + \frac{\rho_{1r}\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r2}})$$

$$R_{12} + R_{31} \leq \frac{1}{2}\log(1 + \frac{(\rho_{12r} + \rho_{3r})\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r2}})$$

$$R_{13} + R_{31} \leq \frac{1}{2}\log(1 + \frac{(\rho_{13r} + \rho_{3r})\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r2}})$$

$$R_{12} + R_{13} + R_{31} \leq \frac{1}{2}\log(1 + \frac{(\rho_{1r} + \rho_{3r})\rho_{r2}}{1 + \rho_{1r} + \rho_{2r} + \rho_{3r} + \rho_{r2}})$$

Again, a similar set of constraints would be imposed at T_3.

5.4 Numerical results

The general achievable rate region (ARR) is either three or four-dimensional, and to reduce the complexity of visualization, we impose symmetry by assuming that $\rho_{2r} = \rho_{3r} = \rho_{r2} = \rho_{r3}$, and $\rho_{1r} = \rho_{r1}$, i.e. maximum uplink and downlink SNR's are equal for the remote terminals, and a different equal value for T_1. We also assume both return link rates are equal, i.e. $R_{31} = R_{21}$, and in the four-flow case, both forward link rates are equal, $R_{12} = R_{13}$ ($\alpha = 1/2$ for time-shared protocols).

We first consider the achievable rate region for each protocol in the three-flow problem. While not discussed explicitly in this Chapter, we note from Chapter 3 that backing off of uplink power from one terminal can help competing signals from other terminals to obtain more power at the relay, so the achievable rate region is actually the union of achievable rate regions over all possible uplink power triples allowed by the maximum power constraints at each terminal. (For downlinks, it is always best for the relay to use maximum power.)

Figures 5.7, 5.8, and 5.9 present the achievable rate region of each option discussed for the three

information flows case, in which $R_{rem} = R_{21} + R_{31}$, and $R_{hub} = R_1$. As can be seen from Figures 5.7, 5.8, and 5.9, a three-simultaneous signaling protocol can outperform the time-sharing protocol 2. As for the two decoding options at remote terminals, option 2 (decode 2-user MAC) is favored with high SNR conditions between $T_2(T_3)$ and the relay, while option 1 (view the other signal as noise) is better when $T_2(T_3)$ has weaker links to the satellite relay. One interesting phenomenon seen from Figure 5.8 and 5.9 is that time sharing protocol 1, though apparently inefficient in repeating forward link data, has a larger ARR for some rate pairs (R_{rem}, R_{hub}) than does both protocol 2's. This is because the penalty of uncancelled interference can exceed the penalty of repeating forward links data in that region. On the other hand, when R_{hub} is given priority, time-sharing protocol 1 is inefficient due to its repetition in sending X_1.

Figure 5.7: Achievable rate region for three information flows, scenario 1
with $\rho_{1r} = 30$ dB, and $\rho_{2r} = 10$ dB

Figures 5.10, 5.11, and 5.12 present the achievable rate region for four information flows case, in which $R_{hub} = R_{12} + R_{13}$, and $R_{rem} = R_{21} + R_{31}$. First, we observe that time-sharing of two-terminal exchange via AF is preferred for all but small regions of the ARR. The region of rate pairs for which three-simultaneous operation outperforms time-sharing diminishes as $\rho_{2r} \to \rho_{1r}$. In addition, within the three-simultaneous signalling protocol, there are four possible decoding options

Figure 5.8: Achievable rate region for three information flows, scenario 2
with $\rho_{1r} = 30$ dB, and $\rho_{2r} = 20$ dB

Figure 5.9: Achievable rate region for three information flows, scenario 3
with $\rho_{1r} = 30$ dB, and $\rho_{2r} = 30$ dB

for remote nodes. For the case considered (symmetric rate), at terminal T_2, option 3 (decode 3-user

MAC) dominates option 2 (view X_{13} as noise and decode 2-user MAC), and option 1 (view X_{31}

as noise and decode 2-user MAC) dominates option 4 (view both X_{13} and X_{31} as noise). However, there is no universal protocol winner between option 1 and option 3 for all rate pairs.
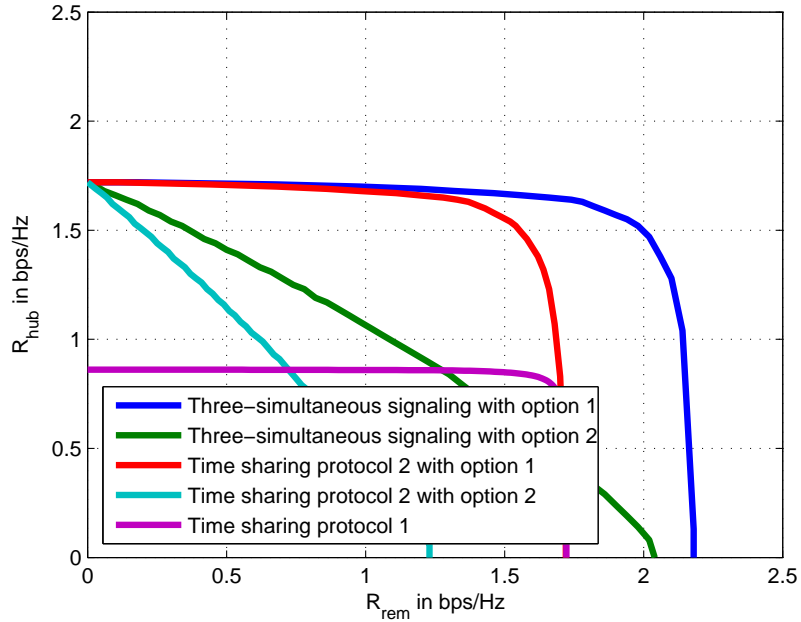

Figure 5.10: Achievable rate region for four information flows, scenario 1 with $\rho_{1r} = 30$ dB, and $\rho_{2r} = 10$ dB

Figures 5.13, 5.14, and 5.15 present the maximum *rate sum* for the four information flows case, where $R_{sum} = R_{12} + R_{13} + R_{21} + R_{31}$. We fix $\rho_{1r} = \rho_{r1}$ and $\rho_{2r} = \rho_{r2} = \rho_{3r} = \rho_{r3}$, and vary ρ_{2r} to get the corresponding maximum allowable rate sum for each protocol and decoding scheme. As can be seen from Figures 5.13, 5.14, and 5.15, the time sharing protocol performs very well in all the three cases studied. Considering the simplicity of implementation, time sharing with a two-user exchange using amplify-forward will be a good choice if achieving large rate sum is the main objective. For the simultaneous signalling protocol, depending on the difference between ρ_{1r} and ρ_{2r}, the performance is quite different between the four decoding schemes. Option 3, which decodes a 3-user MAC, outperforms other protocols when $\rho_{1r} = \rho_{2r}$, while for the other 2 scenarios, option 1, which views signal generated by the other remote terminal as noise, outperforms the other three protocols.
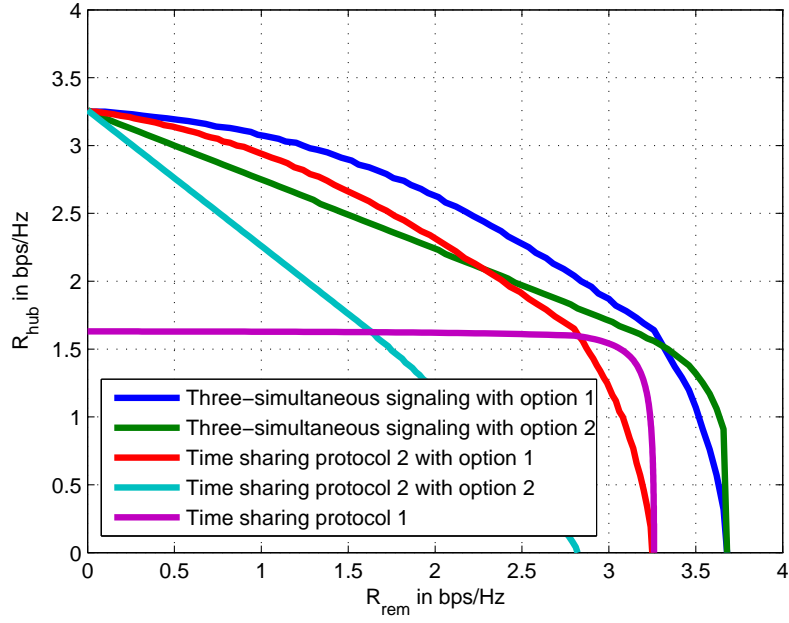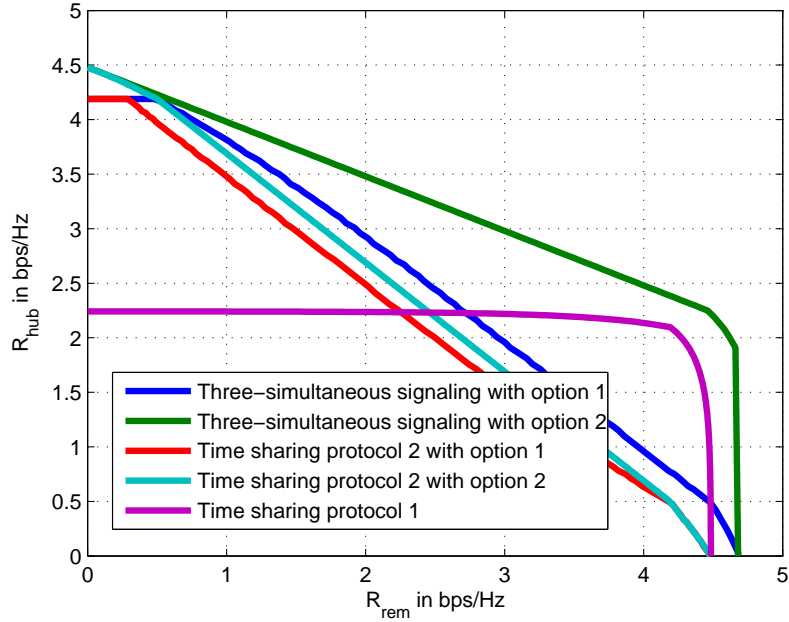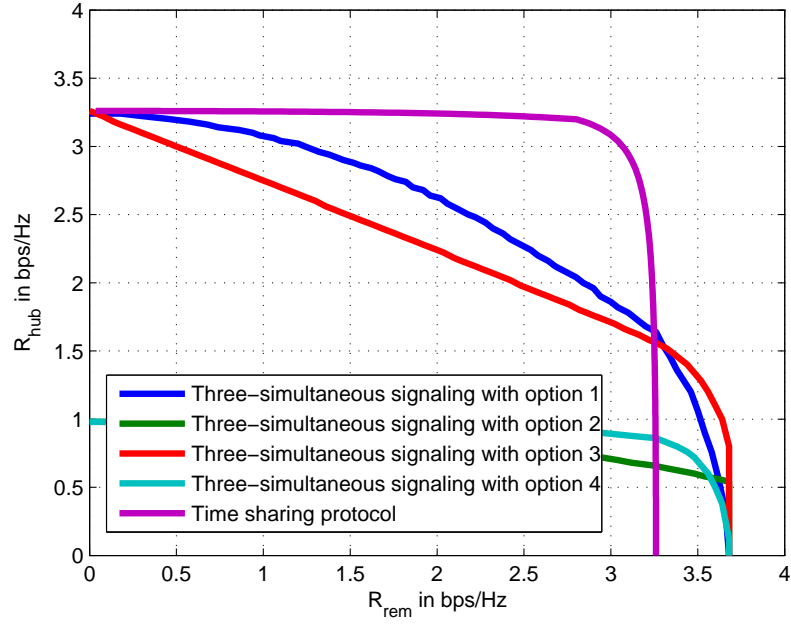
Figure 5.11: Achievable rate region for four information flows, scenario 2 with $\rho_{1r} = 30$ dB, and $\rho_{2r} = 20$ dB
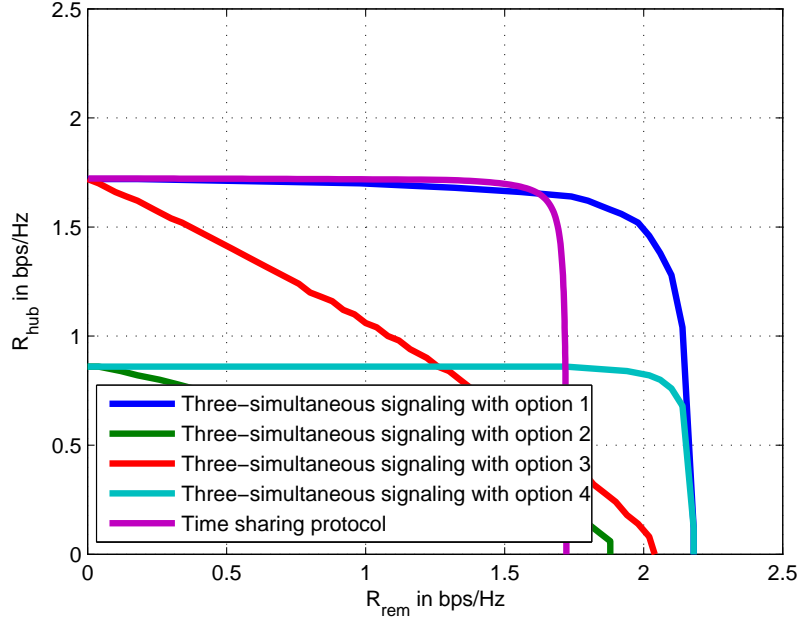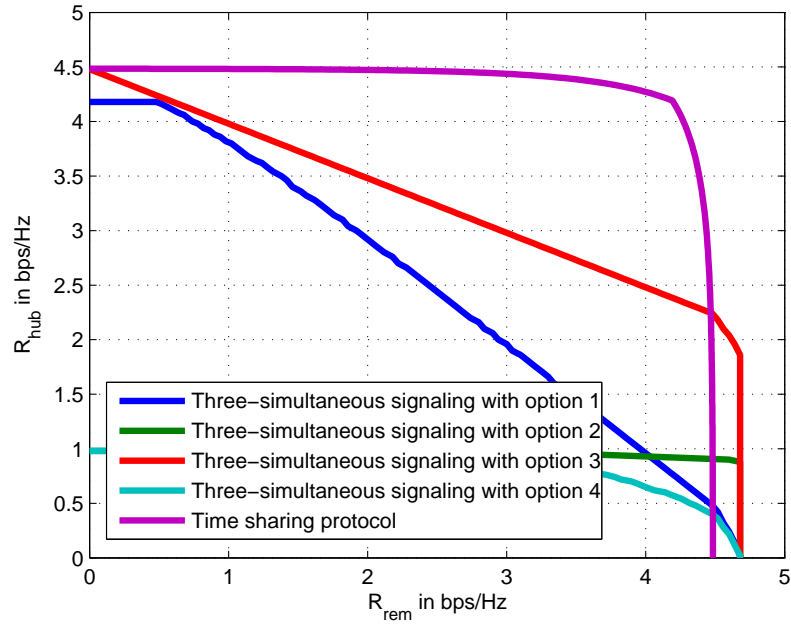

Figure 5.12: Achievable rate region for four information flows, scenario 3 with $\rho_{1r} = 30$ dB, and $\rho_{2r} = 30$ dB
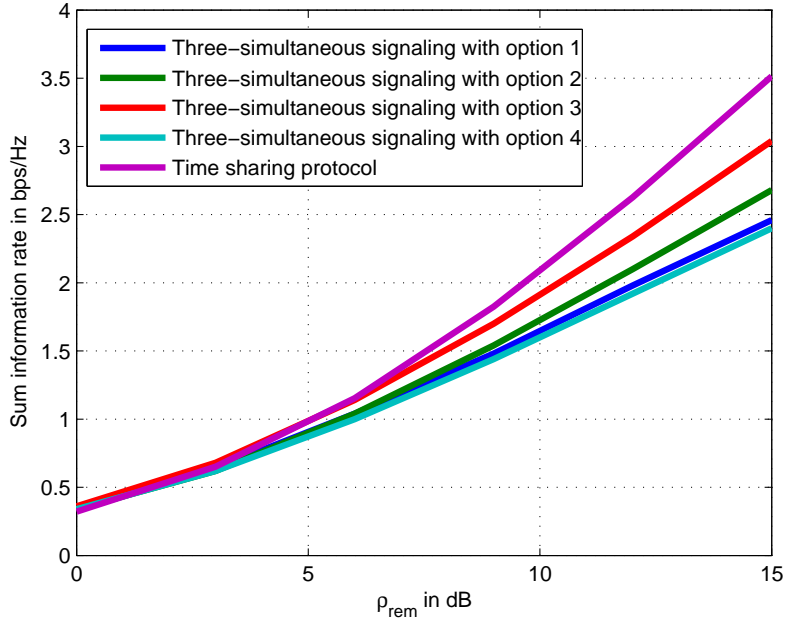
Figure 5.13: Maximum rate sum, four information flows, with $\rho_{1r} = \rho_{2r}$
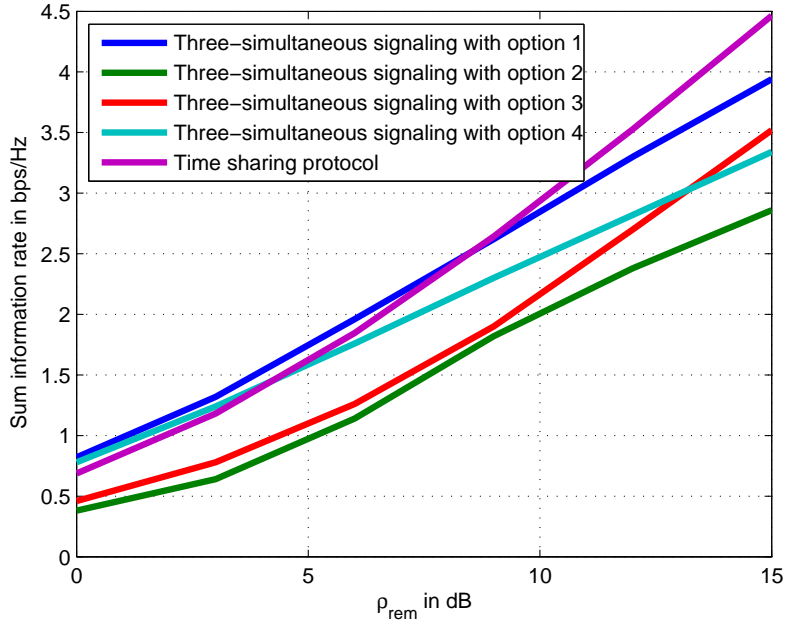
Figure 5.14: Maximum rate sum, four information flows, with $\rho_{1r} = \rho_{2r} + 10$ dB

5.5   Summary

We have considered a simple extension of two-terminal data exchange via a satellite relay to a three-terminal case. Both broadcasting of common information to remotes and separate traffic
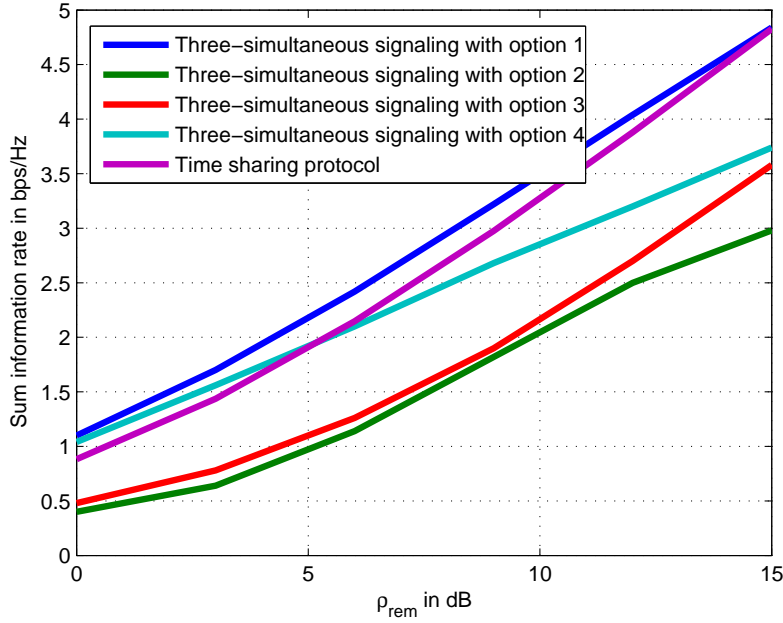
Figure 5.15: Maximum rate sum, four information flows, with $\rho_{1r} = \rho_{2r} + 20$ dB

from hub-to-remotes are analyzed from an information-theoretic perspective. The relay merely amplifies and forwards uplink signals in all cases. Our conclusions are that for the three-flow case, three-simultaneous signaling has a larger throughput than time-shared protocols in some scenarios, while for the four-flow case, three-simultaneous signaling is preferred only if there is significant link quality difference between hub and remote terminals, or when remote terminal rate requirements are emphasized.

The benefits of three-simultaneous signaling are not surprising to those familiar with multi-terminal information theory. However the gains are generally smaller than one might expect. This is primarily due to the constraint on relay power, and power competition effects are important here. Also, we note that the gains for non-orthogonal signaling are less pronounced in the three-terminal problem than with two terminal exchange, since total interference cancellation is not possible here.

In summary, time-sharing of two-user sessions makes good practical sense, as it provides competitive rate regions, and is a simpler protocol for transmission and reception.

Chapter 6

Uncoded Transmission with AF Protocol over Nonlinear Satellite Channel

In previous Chapters, we assume that there is an ideal linear amplifier at the satellite, with an average power constraint. However, real satellite channels are well-known to exhibit nonlinear behavior due to a high power amplifier (HPA). The near-saturation region is the preferred region of operation since more RF power is available at the output and the DC-to-RF efficiency is highest. Of course for non-constant envelope carriers, or multisignal cases, nonlinear impairments may degrade performance severely. Memory effects may be significant as well, induced by bandpass filtering and transmitter/receiver pulse shaping.

Uncoded single-user transmission over nonlinear channels has been widely studied (see [33]), but to our knowledge, there is relatively little work for the two-way relay nonlinear channel. Existing protocols proposed in [43–47] are either assuming channel is memoryless or requiring the nonlinear function to be invertible, both of which are unrealistic.

As a first step to study the nonlinear satellite channel, this Chapter will focus on the uncoded transmission of AF protocol over such nonlinear channel.

6.1 Problem model

Figure 6.1 provides a diagram of the two-way relay nonlinear system we consider. We now ignore the noise before the nonlinear amplifier assuming that the uplink is fairly strong, as a simplifying assumption. $h(t)$ represents transmit pulse shaping for spectrum control, and the nonlinear model for $g(.)$ we adopt is the familiar Saleh model [30] as described in Chapter 2.
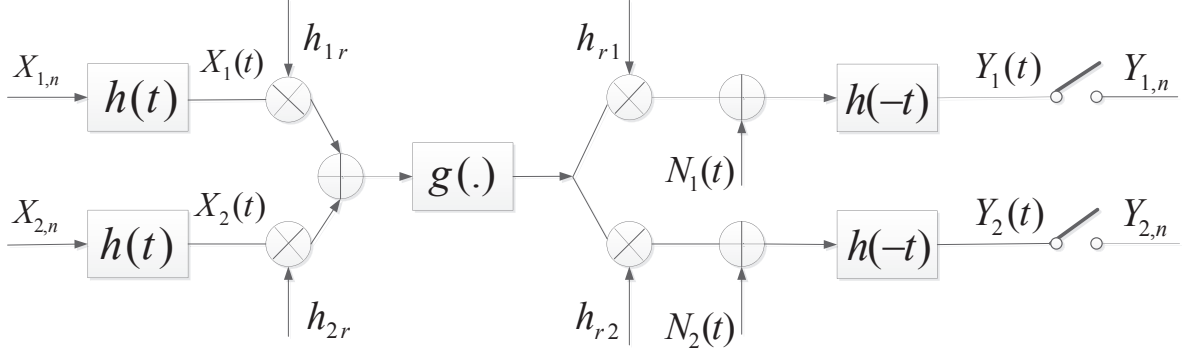
Figure 6.1: System diagram for two-way relay channel

Assume that the channel input message $\bar{X}_1 = \{X_{1,1}, X_{1,2}, ..., X_{1,n}\}$ (generated from T_1) and $\bar{X}_2 = \{X_{2,1}, X_{2,2}, ..., X_{2,n}\}$ (generated from T_2) are two complex sequences drawn from some 2-D constellation set $\mathcal{C}$, then the continuous-time outputs of the channel are [1]

$$Y_1(t) = h_{r1}h(-t)\bigstar g\left(h_{1r}\sum_l X_{1,l}h(t-lT) + h_{2r}\sum_l X_{2,l}h(t-lT)\right) + N_1(t)$$

$$Y_2(t) = h_{r2}h(-t)\bigstar g\left(h_{1r}\sum_l X_{1,l}h(t-lT) + h_{2r}\sum_l X_{2,l}h(t-lT)\right) + N_2(t)$$

where $Y_1(t)$ and $Y_2(t)$ are output signals at T_1 and T_2 respectively, and $N_1(t)$, $N_2(t)$ are white Gaussian noise. Matched filtering is adopted for the sake of practical convenience, though it does not necessarily provide sufficient statistics for decoding on a nonlinear channel [33]. The general discrete-time channel model can be expanded into Volterra series [31] whose exact form will have high nonlinear order and significant memory length. Though side information of $\bar{X}_1$ is still available as in the linear channel case, it is not trivial and efficient to do interference cancellation on the received signal as in an ideal linear channel, due to the nonlinearity of the channel.

---

[1] We assume two signals are perfectly synchronized

6.2 Symbol detection

Without loss of generality, we consider the detection of $\bar{X}_2$ at terminal T_1, with side information $\bar{X}_1$. The maximum likelihood sequence detector (MLSD) will find

$$\widehat{\bar{X}_2} = \{\widehat{X_{2,1}}, \widehat{X_{2,2}}, ..., \widehat{X_{2,n}}\} = \arg\min_{\widehat{\bar{X}_2} \in \mathcal{C}^n} \int |Y_1(t) - \widehat{Y_1(t)}|^2 dt$$

where

$$\widehat{Y_1(t)} = h_{r1}h(-t) \bigstar g\left(h_{1r}\sum_l X_{1,l}h(t-lT) + h_{2r}\sum_l \widehat{X_{2,l}}h(t-lT)\right)$$

However, the computational complexity of this estimator grows exponentially with the length of $\bar{X}_2$, which is too complicated to be implemented. One practical solution is to truncate the length of memory by absorbing the nonlinear effect into mappings denoted by μ, and the effective channel model becomes

$$Y_{1,n}|_{t=nT} = \mu_0(X_{1,n}, X_{2,n}) + Z_{0,n} \qquad \text{or} \tag{6.1}$$

$$Y_{1,n}|_{t=nT} = \mu_2(X_{1,n-1}, X_{1,n}, X_{1,n+1}, X_{2,n-1}, X_{2,n}, X_{2,n+1}) + Z_{2,n} \qquad \text{or} \tag{6.2}$$

$$Y_{1,n}|_{t=nT} = \mu_4(X_{1,n-2}, X_{1,n-1}, X_{1,n}, X_{1,n+1}, X_{1,n+2}, X_{2,n-2}, X_{2,n-1}, X_{2,n}, X_{2,n+1}, X_{2,n+2}) + Z_{4,n}$$

$$\tag{6.3}$$

where μ_0, μ_2, μ_4 are deterministic nonlinear functions for memory order 0, 2, and 4 respectively, and Z_0, Z_2, Z_4 are residuals of the model, aggregating residual nonlinear ISI and additive noise. Practically, $\mu_0(.)$, $\mu_2(.)$, $\mu_4(.)$ can be obtained from training or perhaps pre-calculation, and pre-stored in lookup tables.

We will present three practical sequence detectors in this Chapter, which are all based on some short memory discrete-time channel models. It will be shown later that a memory-2 model can greatly improve the symbol detection performance.

6.2.1 Symbol-by-symbol decisions

The most straight forward scheme is to make symbol-by-symbol decisions. Using the memoryless model

$$Y_{1,n} = \mu_0(X_{1,n}, X_{2,n}) + Z_{0,n} \tag{6.4}$$

the symbol-by-symbol estimator is described by

$$\widehat{X_{2,n}} = \arg\min_{\widehat{X_{2,n}} \in \mathcal{C}} |Y_{1,n} - \mu_0(X_{1,n}, \widehat{X_{2,n}})|^2$$

The memoryless model (6.4) consists of two symbols, so the the size of the μ-table is only $|\mathcal{C}|^2$ (16^2 for 16APSK). However, such model ignores the effect of memory, and it leaves strong nonlinear ISI into the residual of the model, so the corresponding symbol-by-symbol estimator is prone to make wrong decisions, and the overall performance of the symbol detector is not good, as will be shown later in this Chapter.

6.2.2 Improved symbol-by-symbol decisions

In the previous symbol-by-symbol detection, we did not make good use of the side-information $\bar{X}_1 = \{X_{1,1}, X_{1,2}, ..., X_{1,n}\}$. An improved scheme is to adopt a more accurate discrete-time model as below

$$Y_{1,n} = \mu_0'(X_{1,n-1}, X_{1,n}, X_{1,n+1}, X_{2,n}) + Z_{0,n}'$$

$$\tag{6.5}$$

which includes three symbols from the side information, and can capture some of the memory effect, but the size of the μ-table becomes $|\mathcal{C}|^4$ (16^4 for 16APSK). The corresponding estimator is described by

$$\widehat{X_{2,n}} = \arg\min_{\widehat{X_{2,n}} \in \mathcal{C}} |Y_{1,n} - \mu_0'(X_{1,n-1}, X_{1,n}, X_{1,n+1}, \widehat{X_{2,n}})|^2$$

Adding more symbols from the side information can make the model more accurate, and reduce the variance of the residual Z_n, but it also induces more computational and storage complexity.

6.2.3 Viterbi-based detection

In the previous described symbol-by-symbol detection schemes, we adopt some discrete-time models which are memoryless for $\bar{X}_2$. Such type of approximation can severely decrease the accuracy of the model, and hence degrade the detection performance. A more efficient way is to use Viterbi algorithm [56] which incorporates memory (for X_2). We will adopt a finite-state channel model (memory 2) as below

$$Y_{1,n} = \mu_2(X_{1,n-1}, X_{1,n}, X_{1,n+1}, X_{2,n-1}, X_{2,n}, X_{2,n+1}) + Z_{2,n}$$

$$(6.6)$$

and the corresponding μ-table size becomes $|\mathcal{C}|^6$ (16^6 for 16APSK). In the Viterbi algorithm, each state is represented by a pair of two consecutive symbols $(X_{2,n-1}, X_{2,n})$, and there are $|\mathcal{C}|^2$ states for each stage. For state transition $(X_{2,n-1}, X_{2,n}) \to (X_{2,n}, X_{2,n+1})$, the corresponding branch metric is calculated as

$$\lambda_{(X_{2,n-1},X_{2,n}) \to (X_{2,n},X_{2,n+1})} = |Y_{1,n} - \mu_2(X_{1,n-1}, X_{1,n}, X_{1,n+1}, X_{2,n-1}, X_{2,n}, X_{2,n+1})|^2$$

We will see below significant performance improvement by adopting this Viterbi detection scheme.

6.3 Simulation results

In this section, we will present the simulation results for each of the three symbol detection schemes. We adopt a RRC (root-raised cosine) filter with roll off factor 0.25, in the mid-range of {0.2, 0.25, 0.35} specified for DVB-S2 [57]. Modulation options in DVB-S2 include QPSK, 8PSK, 16APSK, and 32APSK. We choose 16APSK as the modulation scheme, which has non-

constant envelope, and provides relatively higher information rate (bits/symbol) than QPSK and 8PSK.

Figure 6.2 shows the performance for the symbol-by-symbol detection. For each of the four input backoffs (IBO) considered, there is an error floor, which means there are significant non-linear ISI residuals that can dominate noise at high SNR region. Increasing IBO (decreasing drive level at the amplifier) makes the amplifier work more linearly, so it can reduce the ISI residuals and lower the error floor. Notice that, a large IBO is preferred for the uncoded two-way relay nonlinear channel, while for the single-user case the amplifier is preferred to work in a near saturation region to get high DC/RF efficiency.
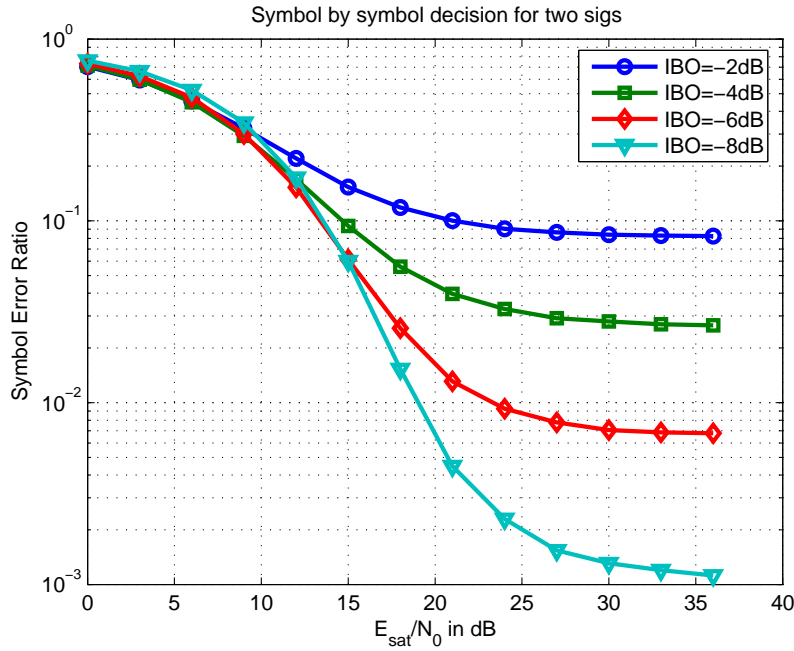
Figure 6.2: Symbol-by-symbol detection on nonlinear channel

Figure 6.3 presents the performance of the improved symbol-by-symbol detection scheme. Similar as in Figure 6.2, there is an error floor for each of the IBOs considered. For the same IBO, the error floor is lowered as compared with the previous memoryless detection, due to the improved channel model adopted. However, such improvement is small, and there still exist severe error floors. We conjecture that further performance improvement is available if more symbols from side information are added to the channel model, but such type of improvement will saturate very quickly.
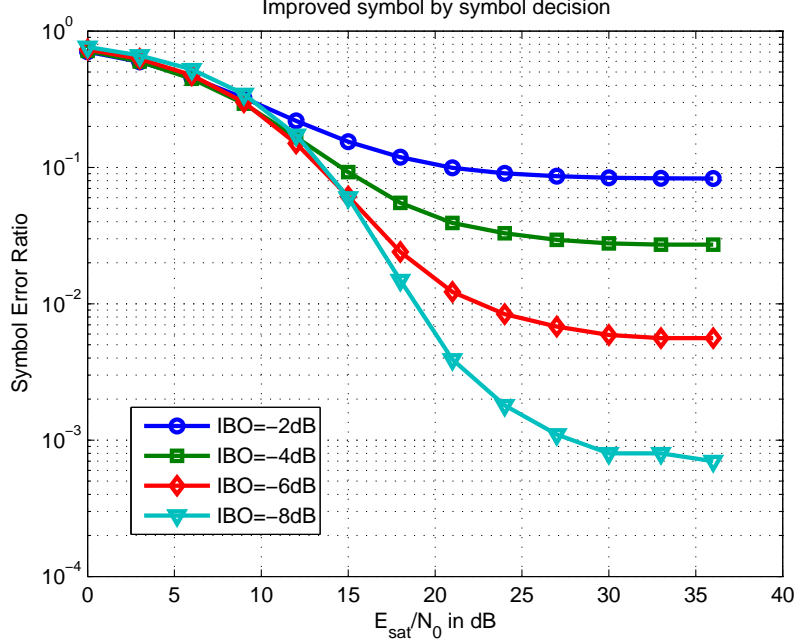
Figure 6.3: Improved symbol-by-symbol detection on nonlinear channel

Figure 6.4 shows the simulation performance for the Viterbi algorithm. Comparing with the two previous symbol-by-symbol detection scheme, there is significant performance improvement. For IBO=-4 or -6 dB, the error floor is lowered by more than one order of magnitude. For IBO=-8 dB, the error floor even vanishes. The conclusion is that for the two-way nonlinear relay channel, large IBO is preferred (for uncoded transmission), and in order to achieve substantial performance improvement, we need to adopt a more complicated channel model with memory.

6.4 Summary

In this Chapter, we study the uncoded transmission of AF protocol over a nonlinear channel. We first adopted a memoryless detection scheme. Simulation results show that such memoryless detection has severe error floor even at a large IBO (-8 dB). Then a memory-2 Viterbi detector is proposed, which can improve the performance by at least one order of magnitude for IBO=-4 and -6 dB, and there is no visible error floor for IBO=-8 dB. Studies of the uncoded transmission in this Chapter will benefit us in our later exploration of the coded case.

Figure 6.4: Viterbi detection on nonlinear channel

Chapter 7

Achievable Information Rates for Nonlinear Satellite Relaying Channels

The single-user nonlinear channel has been widely studied for uncoded transmission, (see [33]), but to our knowledge relatively little work exists on the information-theoretic potential of such channels, i.e. what performance is approachable if advanced channel coding and decoding is allowed? [39] indicates that under aver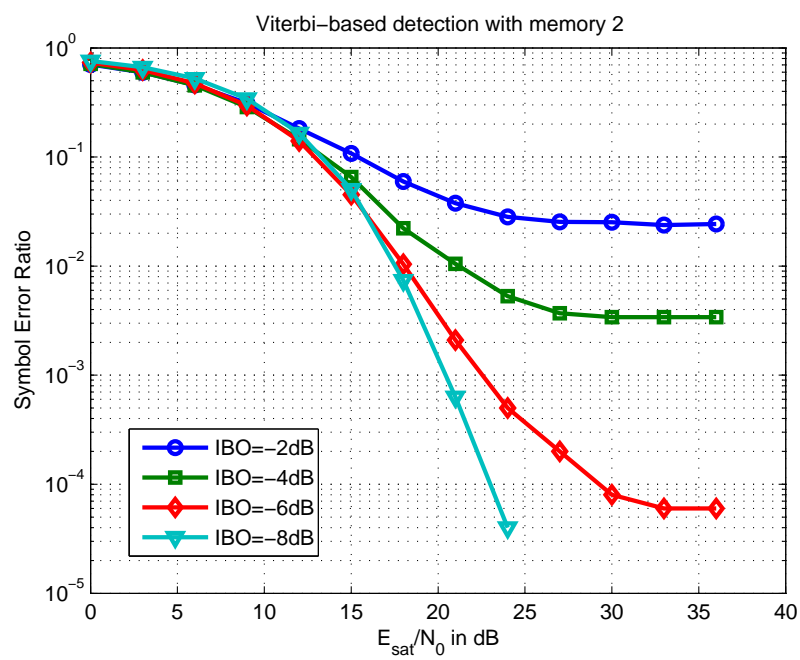age and peak power constraints, the capacity-achieving output distribution is discrete and possesses a finite number of mass points. Other related work in this area can be found in [40, 42].

In the following sections we formulate two transmission scenarios for non-regenerative satellite relays: 1) single-carrier, unidirectional transmission through the nonlinearity, and 2) bi-directional, 'carrier-on-carrier' relaying wherein side-information held at both receivers mitigates the mutual interference. In the first case, the analysis seeks the achievable information rate in bits/modulator symbol, as a function of SNR's and amplifier backoff. 16APSK modulation as in the DVB-S2 standard is employed as a vehicle for study. In the second scenario, we seek the region of rate pairs for which jointly-reliable transmission is allowed; again this region depends on link SNR's, total input backoff at the satellite, and uplink power adjustment.

For each scenario, we formulate two pulse shaping and reception models. In the first case, a rectangular (NRZ) pulse is employed. Though admittedly less practically-motivated, the resulting model is simple, and mutual information results are quickly obtained via numerical integration,

allowing easy experimentation with system parameters. In scenario 1, with proper receiver timing, a memoryless, nonlinear discrete-time model is obtained, while in scenario 2, under uplink asynchronism, nonlinear ISI with memory-1 results.

The second filtering assumption employs root-raised-cosine transmit and receive filtering, followed by symbol-rate sampling in the receiver, as a traditional approach to spectrum management and receiver filtering. The resulting nonlinear system now has memory (even though the pulse shaping produces a Nyquist characteristic on an ideal linear channel). Given a power series model for the PA and a choice of pulse shape, one can determine Volterra kernels that express the input-output behavior in discrete-time. Alternatively, via simulation, one can determine a tabular form of a finite-state machine that represents the nonlinear system with memory. In either case, we need to compute the entropy rate of the channel output, as described below.

7.1 System model

Figure 7.1 and 7.2 provides a diagram of the systems we consider. Either one or two digital uplink transmissions share a common nonlinear transponder at the same frequency. The transponder output is broadcast to either one or two earth terminals. We again neglect noise at the input to the nonlinearity, treating uplink SNR as large.
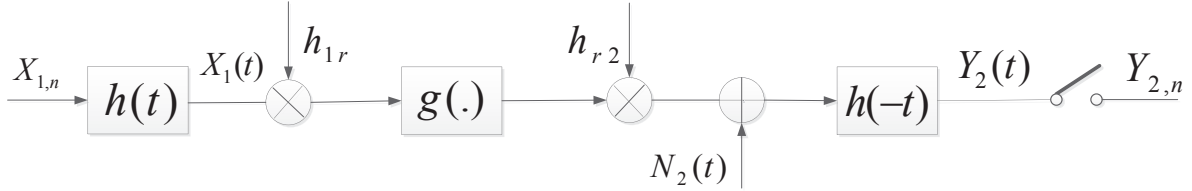
Figure 7.1: System diagram for single-user nonlinear channel

The input message(s) $\bar{X}_1$ or $\bar{X}_2$ is a complex sequence drawn from some 2-D constellation $\mathcal{C}$. Two pulse shaping cases are considered. The first employs an NRZ (rectangular) pulse shape at transmitter and receiver, in the single-user case, and leaves a memoryless nonlinear equivalent baseband discrete-time model from input to output, namely

$$Y_{2,n} = h_{r2}g(h_{1r}X_{1,n}) + n_{2,n}$$

80
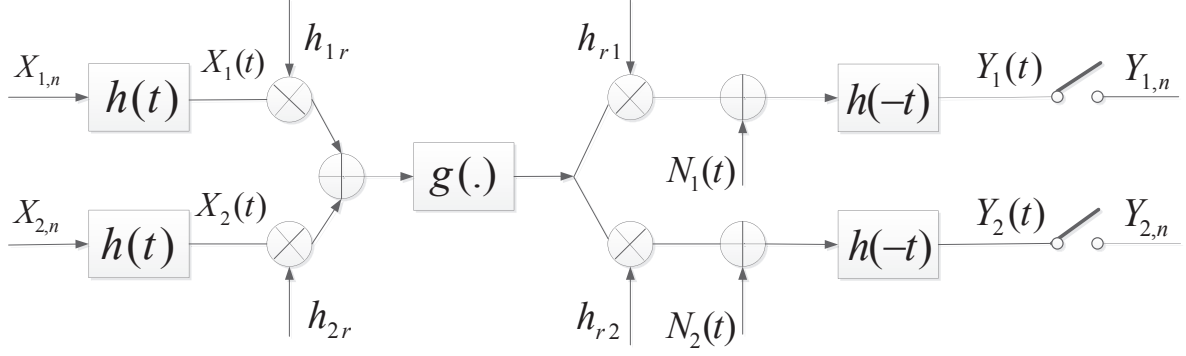
Figure 7.2: System diagram for two-way relay nonlinear channel

where $g(.)$ is a nonlinear complex function defining the amplifier at baseband, and we adopt the Saleh model throughout this thesis as described in Chapter 2.

For the two-user case, we have (still with NRZ and user synchronism)

$$Y_{2,n} = h_{r2}g(h_{1r}X_{1,n} + h_{2r}X_{2,n}) + n_{2,n}$$

$$Y_{1,n} = h_{r1}g(h_{1r}X_{1,n} + h_{2r}X_{2,n}) + n_{1,n}$$

Here one transmission interferes with the other, but because of side-information held at the transmitters in this bidirectional relaying scenario, the echo may be mitigated to some extent, though standard cancellation cannot be performed because of the nonlinearity.

In the pulse-shaped case we employ root-raised-cosine upsampled filters with rolloff factor 0.25. Here $g(.)$ must be applied to the pulse-shaped signal, then receiver filtering applied followed by sampling, to obtain a nonlinear finite-state model.

For simplicity, downlink SNR's are assumed equal, defined as E_{sat}/N_0, with E_{sat} equal to the received energy for a continuous wave signal at saturated power output. The complex noise variance at the receiver output is N_0, the one-sided noise spectral density. Further, we assume total signal power at the input to the satellite is much larger than satellite receiver noise, so noise at the relay may be ignored.

7.2 Achievable rates

A rate of transmission, measured in bits/modulator symbol, is said to be *achievable* if one can prove encoders and decoders exist that allow arbitrarily reliable communication, as complexity of encoding/decoding is allowed to grow, [7]. Such rates are related to the mutual information in bits/symbol that is attainable in the different scenarios, as SNR varies, and as the satellite operating point varies.

7.2.1 Single-user case

For the single-user case, the mutual information of interest is

$$I(X_1 \wedge Y_2) = \lim_{n \to \infty} \frac{1}{n} I(\overline{X_{1,n}} \wedge \overline{Y_{2,n}})$$

and it may be argued that all information rates satisfying $R_{12} < I(X_1 \wedge Y_2)$ are achievable. Notice we do not perform maximization of mutual information over all distributions on input sequences, but instead adopt an equiprobable, independent probability distribution on a standard constellation (here 16-APSK) for X_1. Thus we speak of achievable rates rather than channel capacity. We conjecture that this choice is in fact optimal as information rates increase toward the limiting value, as is true for linear ISI channels.

It is convenient to express mutual information as the difference between entropy rate $h(Y_2)$ and differential entropy of the additive Gaussian noise, $h(N_2) = \log_2(2\pi e \sigma^2)$, meaning that the key task is finding the entropy rate of the output of the nonlinear FSM observed in additive white Gaussian noise. This is equivalent to asking for the entropy rate of a hidden Markov model, for which no closed-form solutions exist. Thus, we appeal to either numerical integration or Monte Carlo methods [58] to evaluate the desired entropy rates, as described below.

The entropy rate for a stationary process is defined by

$$h(Y) = \lim_{n \to \infty} \frac{1}{n} h(Y_1, Y_2, ... Y_n)$$

In some cases, specifically when pulse-shaping is rectangular, we may compute $h(Y)$ by finding the first-order p.d.f. of the complex channel output, then numerically integrating the expression

$$h(Y) = -\int f(y) \log_2 f(y) dy$$

(Focusing upon the marginal p.d.f. alone suffices due to the memoryless model.)

In other cases, notably where memory is present, we employ the Monte Carlo method described in [58] to estimate the limiting form of the entropy rate from a long realization of the actual channel. Typically 10000 symbols were employed to obtain sufficient estimation accuracy [1].

7.2.2 Bidirectional case

In the bidirectional relaying case, we are interested in two coupled mutual information expressions:

$$I(X_1 \wedge Y_2|X_2) = \lim_{n \to \infty} \frac{1}{n} h(\overline{Y_{2,n}}|\overline{X_{2,n}}) - h(N_2) \tag{7.1}$$

$$I(X_2 \wedge Y_1|X_1) = \lim_{n \to \infty} \frac{1}{n} h(\overline{Y_{1,n}}|\overline{X_{1,n}}) - h(N_1) \tag{7.2}$$

For bidirectional communication, we define an *achievable rate region* (ARR) as the set of all bidirectional rate pairs for which joint reliability is possible. This region depends on input probability assignment as well as receiver processing. The two-user capacity region is the union of all ARR's.

As for the single-user case, numerical integration suffices for NRZ pulse shaping where a memoryless (or memory-1) model results. With RRC pulse shaping, we resort to simulation methods to estimate the desired entropy rate.

We study achievable rates under the realistic assumption of inputs drawn independently and equiprobably from some M-ary constellation. Slightly higher rates may be attainable by 1) unequal marginal input probabilities, and 2) proper input dependency over time, i.e. spectrum shaping.

---

[1]Thanks to John Peng for his work in calculating entropy rates by Monte Carlo simulations

7.3 Numerical results

7.3.1 Single-user transmission

In Figure 7.3 we plot achievable rate for NRZ shaping versus the downlink SNR E_{sat}/N_0, assuming uplink noise is negligible at the input to the nonlinearity. As expected for 16-APSK modulation, the high-SNR asymptote approaches 4 bits/symbol at all input backoffs (IBO), but the proper choice of input backoff can optimize resources. For example, with coded transmission onto 16-APSK at 3 bits/symbol, input backoff of -2 dB is optimal, and requires $E_{sat}/N_0 = 11$ dB. Notice also that for lower SNR (or rates), even closer to saturation is (slightly) better, but input backoff = -2 dB is remarkably robust. For comparison in Figure 7.3, we show the achievable rate for the unconstrained Gaussian channel and for 16-APSK on a linear channel, interpreting E_{sat} as *average* energy per symbol. Much of the gap between the linear 16-APSK result and the best nonlinear curve is due to the difference in average output power between linear and nonlinear cases, about 1.1 dB at IBO=-2 dB. Under an equal average power comparison, linear and nonlinear achievable rates are quite similar, indicating that nonlinear distortion is a relatively small effect compared to additive noise over the region of interest, provided proper receiver processing is employed.

As an aside, in this simple memoryless case, one may formulate the question of optimal input constellation of 16APSK, since the output constellation is merely a warped version of the input. We take the case of IBO= -2 dB for study, and vary $\gamma = r_{out}/r_{in}$ for some typical values used in DVB-S2 [57], which is the ratio of ring radii for 16-APSK. Figure 7.4 shows the impact of γ on information rate. As can be seen in Figure 3, for the 'optimal' IBO=-2 dB, there is little change between different γ. We believe this holds for other IBOs, so we fix $\gamma = 2.85$ throughout this Chapter, as in the DVB-S2 standard [57].

For single-user transmission with raised-cosine pulse shaping, the nonlinear model now has memory, introducing nonlinear ISI. Figure 7.5 illustrates a typical receiver output constellation at the best sampling time, in the absence of noise, when IBO=-2 dB, showing this ISI. One may observe in such plots that as IBO increases, the constellation shrinks, as does the nonlinear ISI, but the overall performance in noise degrades. On the other hand, harder drive of the amplifier only increases scatter of the received data, without increasing distance between clusters.

Figure 7.6 presents achievable rates for the root-raised-cosine case, showing similar behavior
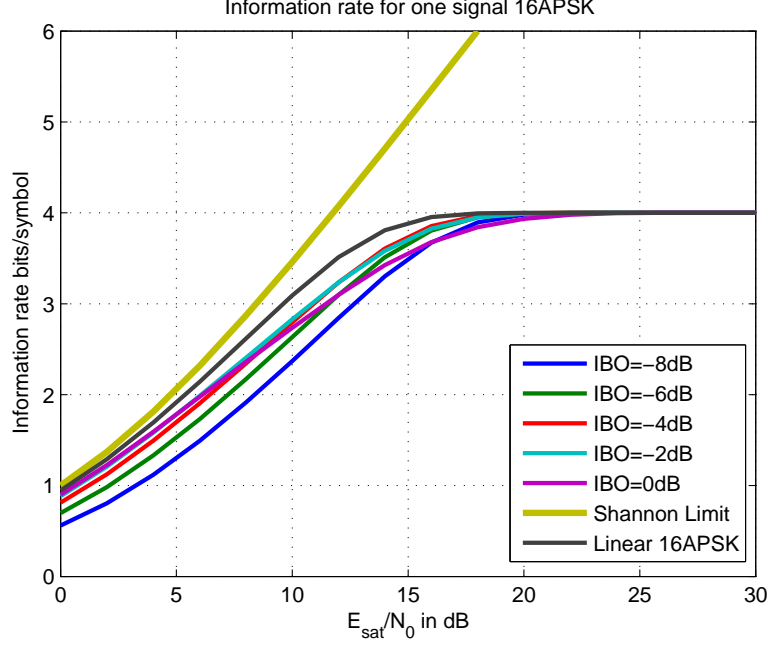
Figure 7.3: The impact of nonlinearity on single-user Gaussian channel, NRZ pulse
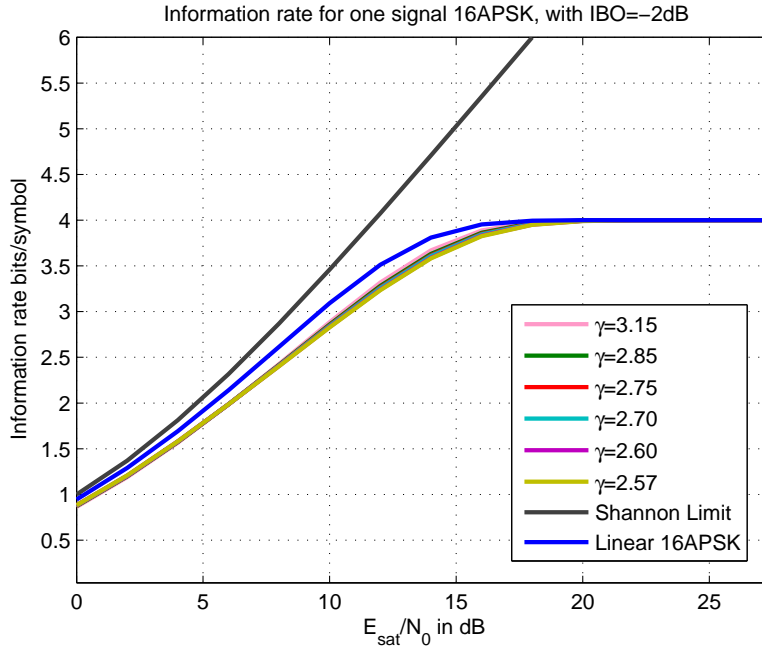
Figure 7.4: The effect of γ on information rate

with Figure 7.3. We determined that a memory-2 model was an accurate representation of the system, i.e. $Y_n = \mu(X_{n-1}, X_n, X_{n+1}) + Z_n$. This implies that a trellis-based computation of entropy rate following [58] requires 16^2 states. Close inspection shows that for signalling with

85
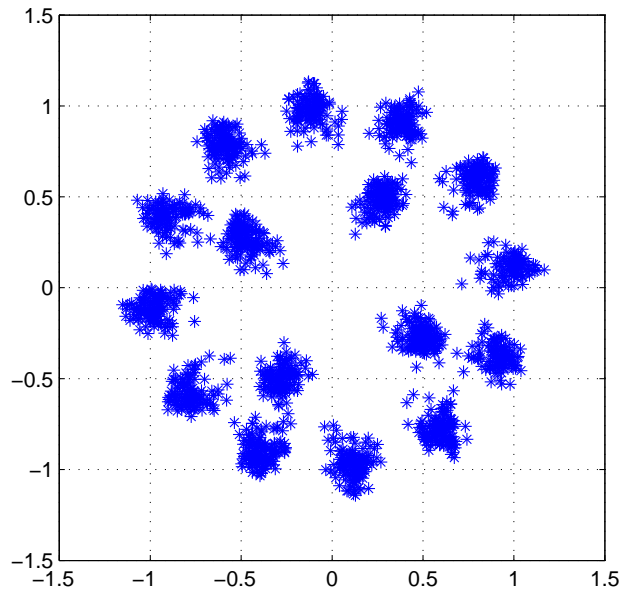
Figure 7.5: Received constellation of 16-APSK after matched filtering
with raised-cosine pulse shaping, IBO=-2 dB

root-raised-cosine pulses, the best IBO is again -2 dB, favoring a near saturation region.

Figure 7.6: Achievable rate for one signal pulse shaped by root-raised-cosine filter

7.3.2 Bidirectional transmission

We now have additional system variables that may affect achievable rates, specifically the fractional symbol delay between the two uplink transmissions, measured at the satellite. Carrier phase rotation is another. In the *fully-synchronous* case, the relative delay and phase shifts are zero, whereas in the *asynchronous* case, the symbol delay is 1/2 symbol, and carrier phase rotation is 15 degrees. These two cases are believed to present the two extremes.

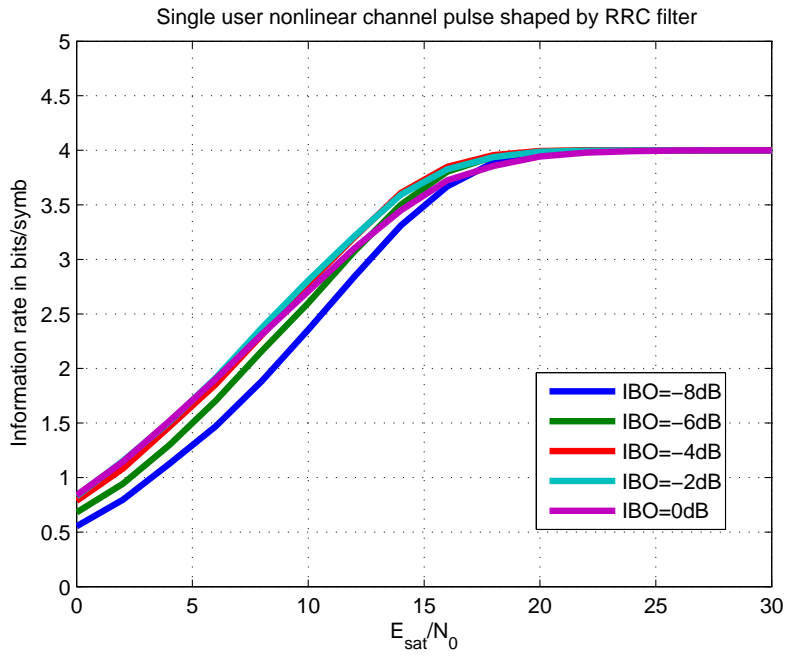With NRZ pulse shaping and synchronous signalling, a memoryless model is obtained in (7.1) and (7.2), i.e. $Y_n = \mu(X_{1,n},\ X_{2,n}) + Z_n$. In the asynchronous model two consecutive other-user symbols interfere with the desired symbol, so the corresponding model becomes $Y_n = \mu(X_{1,n},\ X_{1,n+1},\ X_{2,n}) + Z_n$, making computation of the p.d.f. for $h(Y)$ a little more difficult, but numerical integration is able to compute the result. Figure 7.7 and 7.8 show the information rate when $R_{12} = R_{21}$ vs SNR for both synchronous and asynchronous cases. Due to assumed rate symmetry, the two uplink signals are equally-strong, and share the satellite output power equally.
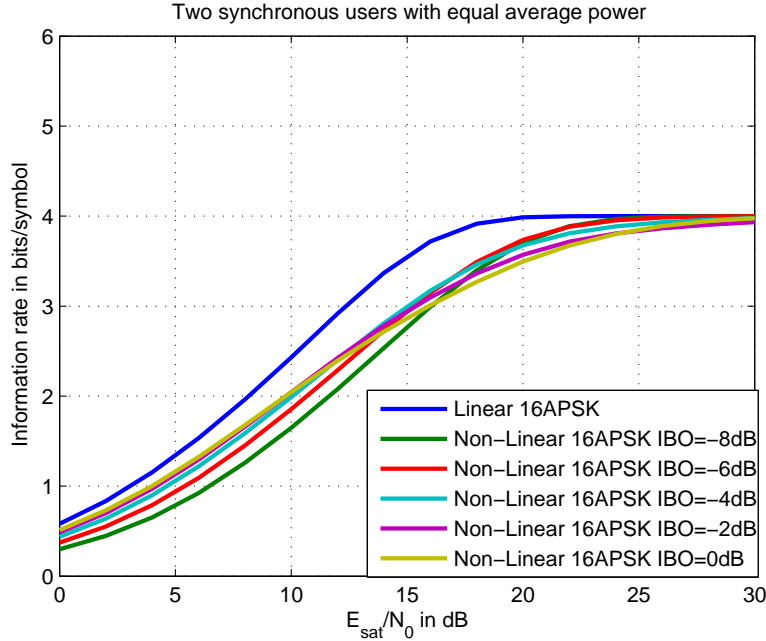

Figure 7.7: Achievable rate for two synchronous signals for different IBOs, NRZ pulse

We observe that the impact of asynchronism is quite small, and if anything, lack of common symbol timing and carrier phase rotation helps. Here the best IBO is about -4 dB, and this value is robust, i.e. for rate-symmetric or rate-asymmetric bidirectional relaying, IBO=-4 dB is preferred,
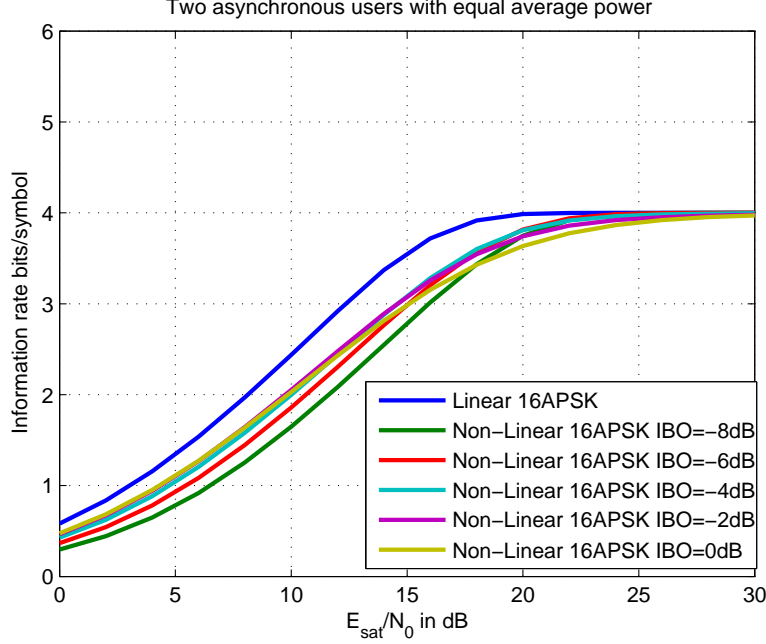
Figure 7.8: Achievable rate for two asynchronous signals for different IBOs, NRZ pulse

while for the case of single signal, it has been shown that IBO $= -2$ dB is 'optimal' for NRZ and root-raised-cosine pulse shaping. Finally we observe that to double the spectral efficiency (say 3 bits/symbol in both directions rather than one direction), we need to increase the downlink SNR from about 11 dB to 15 dB. An increase of 3 dB is expected, but the interaction of two signals in the nonlinear amplifier penalizes the system an extra 1 dB.

To display the complete achievable rate region for asymmetric rate cases, we must repeat the previous calculations for varying imbalances of uplink signal power. For example, if R_{12} is favored over R_{21}, the uplink power of the $1 \rightarrow 2$ signal should be increased (or the other reduced). The ARR then is the largest region of achievable rate pairs, when power imbalance is considered.

Figure 7.9, 7.10, 7.11, and 7.12 presents the ARR of both synchronous and asynchronous cases for varying backoff conditions. We considered two cases: $E_{sat}/N_0 = 5$ dB and $E_{sat}/N_0 = 15$ dB. As shown in Figure 7.9, 7.10, when the downlinks have low SNR (5 dB), IBO=0 dB is optimal. The reason is that here the additive noise effect dominates nonlinear distortion, so high drive level outperforms low drive level. However, when downlinks are strong ($E_{sat}/N_0 = 15$ dB), the impact of nonlinearity becomes more significant, and it is a better trade-off between additive noise and nonlinearity to operate with IBO=-4 dB, as can be seen in Figure 7.11 and 7.12. Finally, we

observe that for $E_{sat}/N_0 = 5$ dB and the same IBO, synchronous signals have slightly *larger* ARRs than that of asynchronous signals, while for $E_{sat}/N_0 = 15$ dB and the same IBO, asynchronous signals have slightly *larger* ARRs than synchronous signals. Our explanation is that for the case of $E_{sat}/N_0 = 5$ dB, noise plays a major part in the performance, asynchronous signals suffering from the loss of average power at the output of the NRZ matched filter (though synchronous and asynchronous signals have the same peak power), which is due to the misalignments of time and phase between the two signals, while for the case of $E_{sat}/N_0 = 15$ dB, noise is no longer the dominating factor, the output constellations of asynchronous signals are more Gaussian-like due to the misalignments of time and phase between the two signals, so it slightly outperforms synchronous signals.
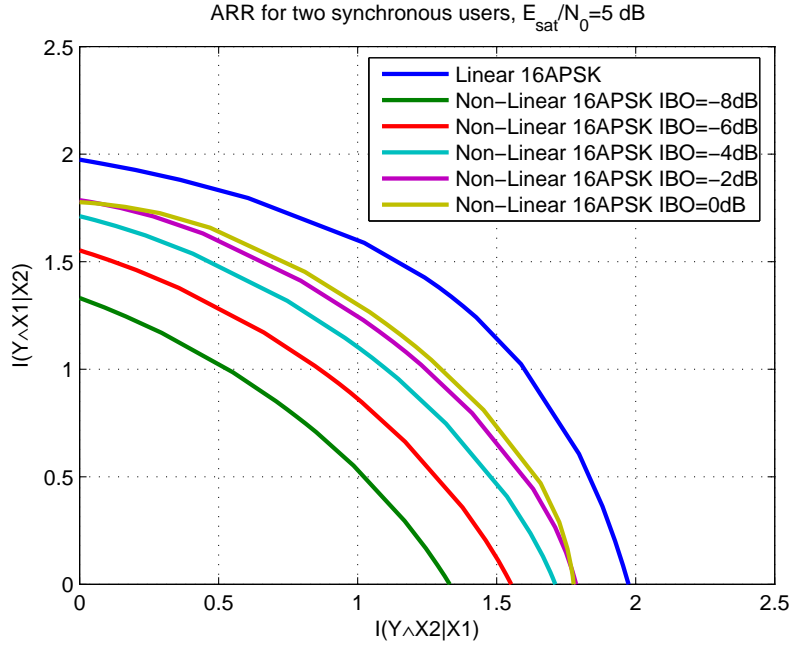
Figure 7.9: ARR for synchronous signals, with $E_{sat}/N_0 = 5$ dB

7.4 Summary

Achievable rate regions have been obtained for two nonlinear satellite transmission scenarios: traditional single-carrier per transponder transmission and bidirectional relaying, where spectrum reuse is utilized. The results are obtained by a mix of numerical integration and Monte Carlo methods, needed to obtain the entropy rate for the received process. Using the Saleh model for the non-
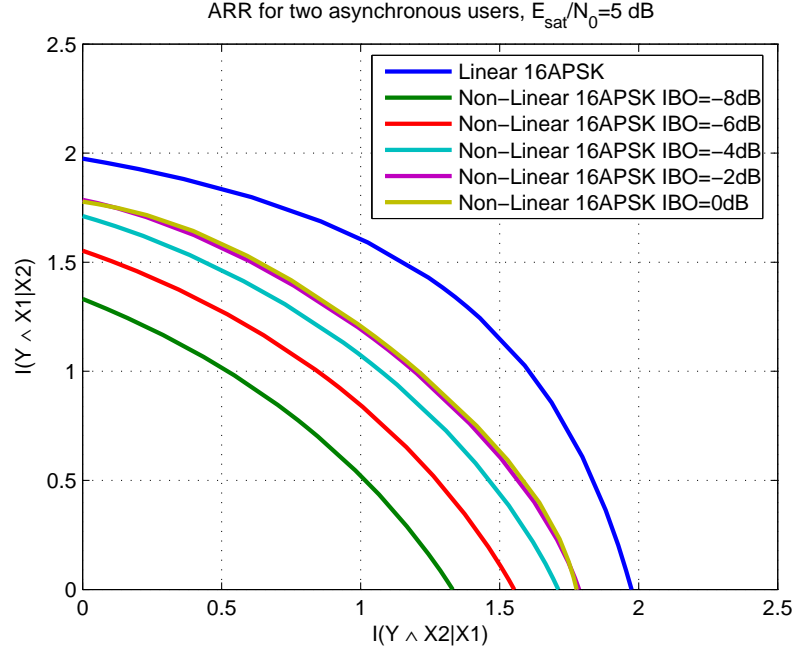
Figure 7.10: ARR for asynchronous signals, with $E_{sat}/N_0 = 5$ dB
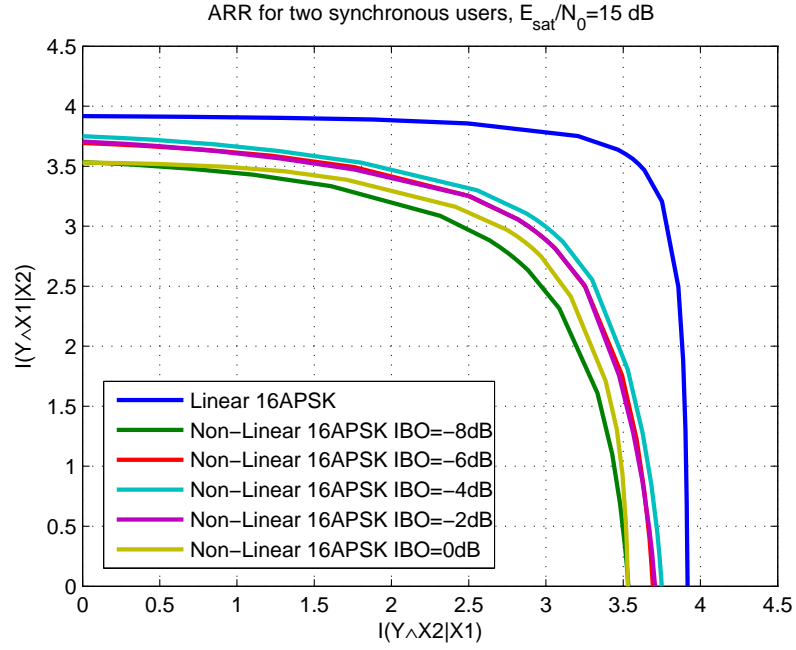
Figure 7.11: ARR for synchronous signals, with $E_{sat}/N_0 = 15$ dB

linear power amplifier, it appears that for single-user transmission input backoff of about -2 dB from saturation is near-optimal with NRZ pulse-shaping, and such IBO is also appropriate when root-raised-cosine signalling and matched filtering introduces memory to the system. Furthermore,
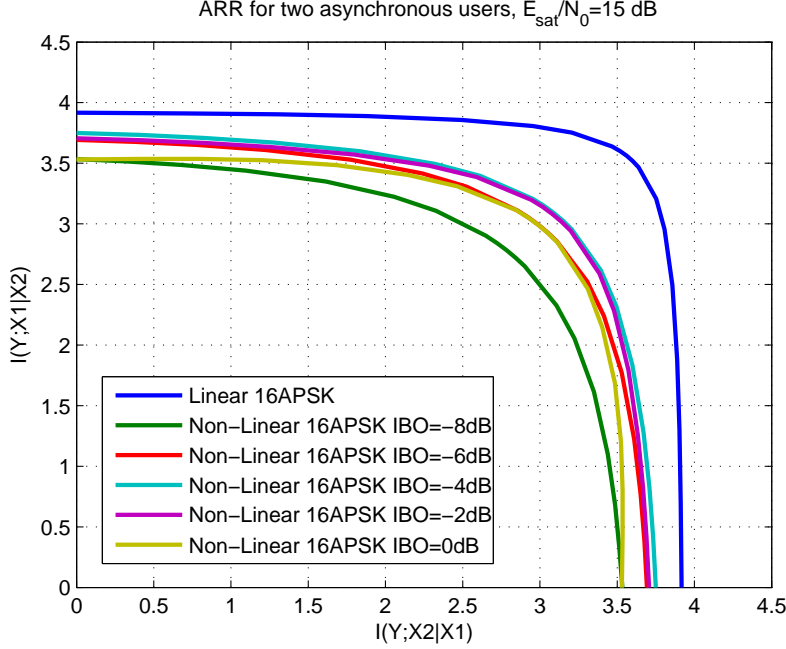
Figure 7.12: ARR for aynchronous signals, with $E_{sat}/N_0 = 15$ dB

the achievable rates are only slightly smaller than for a linear channel with the same *average* power at amplifier output, indicating a standard nonlinearity does not intrinsically compromise coding potential under average power constraints. Similar remarks hold for the bidirectional relaying case for the achievable rate region in 2-D.

Approaching these achievable rates implies sophisticated encoding and decoding operations. Our assumptions about the source sequence (iid, equiprobable) are consistent with traditional channel codes. Decoding on the other hand may need to appeal to algorithms that incorporate channel memory, in addition to incorporating the known nonlinear channel mechanism. This is taken up in the next Chapter. The results reported here are perhaps case-dependent, being affected by constellation choice, amplifier model, and rolloff factor. The methodology, however, is general, and we believe the major findings are likely to hold more generally.

Chapter 8

Coded Single-User Transmission Over Nonlinear Satellite Channel

The DVB-S2 standard is widely employed for satellite-aided networking and video broadcasting, [57, 59, 60]. The standard describes several modes of transmission with differing spectral efficiencies, typified by use of rate 3/4 LDPC coding combined with 16-APSK modulation to achieve 3 bits/symbol spectral efficiency. In these different modes, the standard defines constellations and bit labellings that appear to have been driven by operation over near-linear channels.

In this Chapter, we study single-user high spectral efficiency transmission over a nonlinear channel, in the context of DVB-S2. We adopt root-raised-cosine pulse shaping for spectrum control together with similar matched filtering in the receiver. The latter does not produce sufficient statistics for the decision problem, but is a common, and simple, choice for reception. From end-to-end, this produces a discrete-time model possessing nonlinear warping of the transmitted constellation points, plus a signal-dependent perturbation due to nonlinear ISI, as well as additive downlink noise. Outside of this inner channel sits an LDPC encoder and decoder, as specified by the DVB-S2 standard.

Previously presented approaches include using nonequiprobable distribution of signal constellation [61], predistorting the signal before the nonlinear amplifier [62], and joint channel equalization and soft decoding [63]. Some of these imply significant increase in complexity, or incompatibility with DVB-S2.

Our results in this Chapter are primarily the following. First, achievable rate calculations for the nonlinear channel with memory studied in Chapter 7 are found to correlate well with actual decoder performance, and thus can serve as a useful design tool. Second, an increase in the ring ratio of the 16-APSK constellation, together with proper IBO selection, allows nearly 0.4 dB decrease in downlink SNR to obtain a given decoder output error probability. Of course, the best choice of IBO and ring ratio depends on the actual channel, i.e. the parameters of the Saleh model and the RRC filtering. Finally, a proposed feedback decoding can further improve the performance by another 0.4 dB, by better decoding of the nonlinear ISI of the channel.

8.1 System model

Figure 8.1 provides a diagram of the two-link system we consider. We ignore the noise before the nonlinear amplifier assuming that the uplink is strong. The box denoted $E \ |\Pi| \ M$ represents channel encoding, then interleaving followed by modulation, $h(t)$ represents transmit pulse shaping for spectrum control, h_{1r} and h_{r2} represent complex channel gains on uplink and downlink, and D represents the channel decoder. The nonlinear model for $g(.)$ is again the familiar Saleh model [30] with the same parameters as previous in chapters.
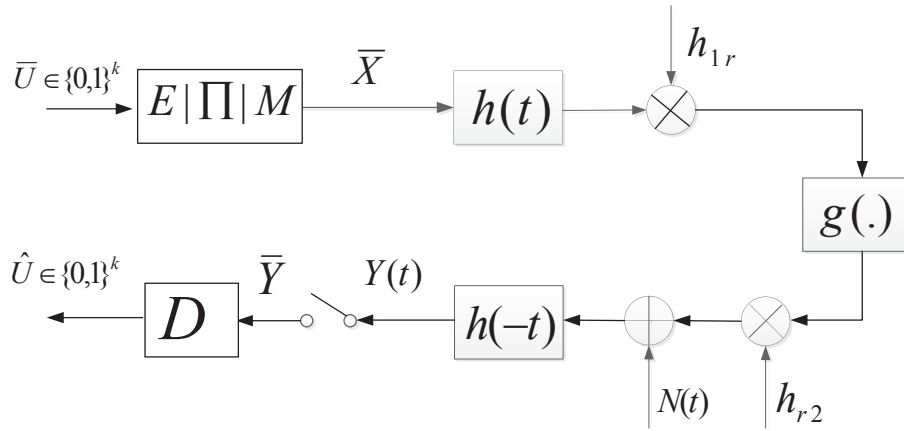

Figure 8.1: System model for single-user transmission in DVB-S2

93

8.1.1 The channel

The channel input message $\{X_1, X_2, ..., X_n\}$ is a complex sequence drawn from some 2-D constellation set $\mathcal{C}$. The continuous-time output of the channel is

$$Y(t) = h_{r2}h(-t) \star g\left(h_{1r}\sum_l X_l h(t - lT)\right) + N(t)$$

where $N(t)$ is white Gaussian noise.

Matched filtering is again adopted for the sake of practical convenience, though it does not necessarily provide sufficient statistics for decoding on a nonlinear channel [33]. We adopt a RRC (root raised cosine) filter with roll off factor 0.25, in the mid-range of $\{0.2, 0.25, 0.35\}$ specified for DVB-S2. The general discrete-time channel model can be expanded into Volterra series [31] whose exact form will have high nonlinear order and significant memory length. We may approximate the model by truncating the length of memory, and absorbing the nonlinear effect into non-causal mappings denoted by μ, and the effective channel model becomes

$$Y_n = \mu_0(X_n) + Z_{0,n} \qquad \text{or} \tag{8.1}$$

$$Y_n = \mu_2(X_{n-1}, X_n, X_{n+1}) + Z_{2,n} \qquad \text{or} \tag{8.2}$$

$$Y_n = \mu_4(X_{n-2}, X_{n-1}, X_n, X_{n+1}, X_{n+2}) + Z_{4,n} \tag{8.3}$$

where μ_0, μ_2, μ_4 are deterministic nonlinear functions for memory order 0, 2, and 4 respectively, and Z_0, Z_2, Z_4 are residuals of the model, aggregating residual nonlinear ISI and additive noise.

8.1.2 Coding and modulation

The original data bits are encoded, block-interleaved, and modulated before going into the channel. In DVB-S2 [60], LDPC code rates of 1/4, 1/3, 2/5, 1/2, 3/5, 2/3, 3/4, 4/5, 5/6, 8/9, and 9/10 are available depending on the selected constellation and the system application. Code block-lengths are 16200 or 64800 bits. In this paper, we focus on an intermediate case with rate 3/4 coding mapped to 16-APSK. Figure 8.2 shows the constellation and bit labeling of 16-APSK as used in DVB-S2. According to the DVB-S2 standard, the ring ratio $\gamma = r_{out}/r_{in}$ varies within $\{2.57, 2.60, 2.70, 2.75, 2.85, 3.15\}$, and $\gamma = 2.85$ is the suggested ring ratio for rate 3/4.
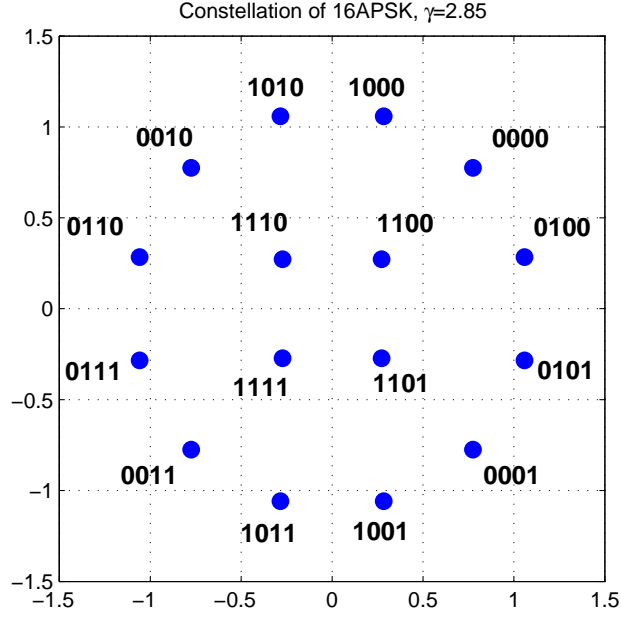
Figure 8.2: Constellation and bit labeling of 16APSK

8.2   Achievable information rate

In a similar way as described in Chapter 7, we are able to obtain Figures 8.3 and 8.4, which show achievable rates for memory-order 0 and memory-order 2 approximations to the channel, 16-APSK modulation is used with ring ratio $\gamma = 2.85$. The calculation assumes the residual Z_n sequence is independent and Gaussian, a further slight mismatch. The rate is shown versus downlink SNR E_{sat}/N_0 for various IBO choices. The μ functions in (8.1) and (8.2) above were obtained empirically for differing IBO from a long sequence of random data, as was the variance of the Z_n sequence. A point of interest from Figure 8.3 is that to obtain $R = 3$ bits/symbol, we require E_{sat}/N_0 of 11.2 dB, and at this point IBO = -2 dB seems optimum. Figure 8.4 shows a similar plot for memory-order 2, and the required SNR to obtain 3 bits/symbol drops by about 0.4 dB to 10.8 dB. This is a consequence of better modeling of the deterministic channel effect, and corresponding reduction in variance of Z_n.

Figure 8.3: Achievable information rate for memoryless model
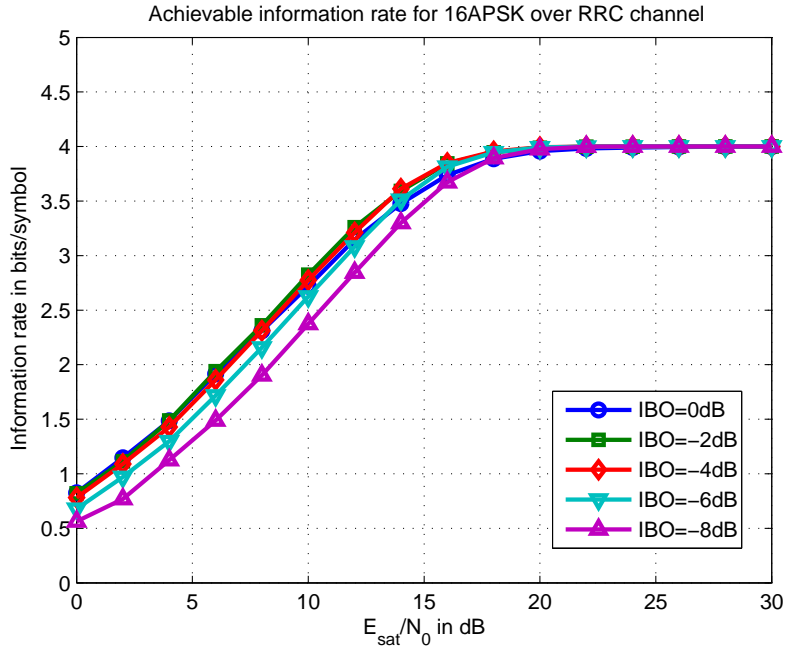
Figure 8.4: Achievable information rate for memory-2 model

8.3 LDPC coding performance in DVB-S2

From an information theory perspective, plots like those shown in Figure 8.3 and 8.4 tell us the best IBO as well as the lowest SNR limit for reliable transmission over the RRC nonlinear channel, provided maximum likel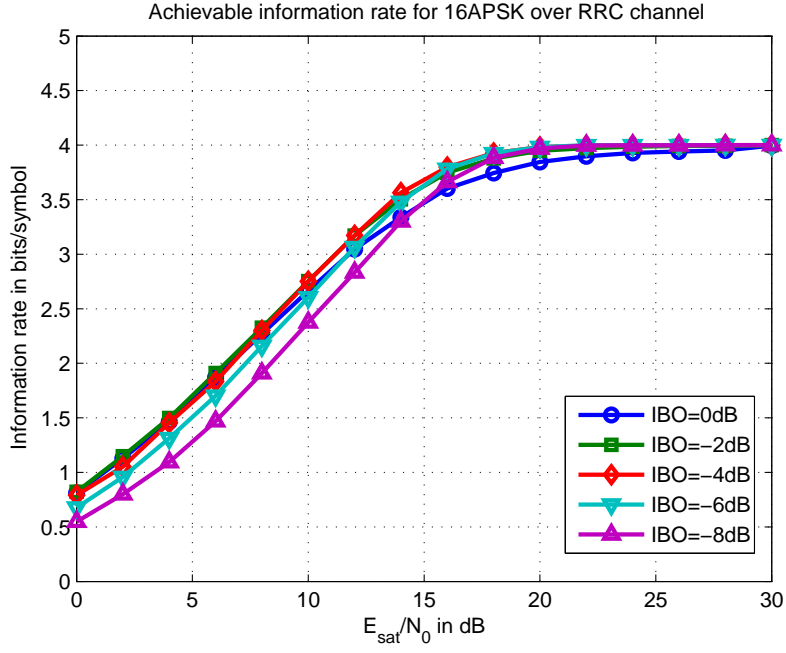ihood decoding is performed using the relevant channel model. Channel encoding/decoding is required to approach such a lowest SNR limit. In DVB-S2, LDPC codes are the adopted channel coding scheme. The original bit sequence $\bar{U}$ is first encoded into a valid binary codeword $\bar{C}$. $\bar{C}$ is interleaved, and mapped to a 16-APSK symbol sequence $\bar{X}$, based on the bit labeling shown in Figure 8.2. At the receiver, as shown in Figure 8.5 the complex-valued sequence $\bar{Y}$ from the output of the matched filter is first used to calculate the LLR (Log-likelihood Ratio) for each bit, and then this LLR information is deinterleaved and passed to the LDPC decoder to do soft decoding in iterative message-passing manner, [64].
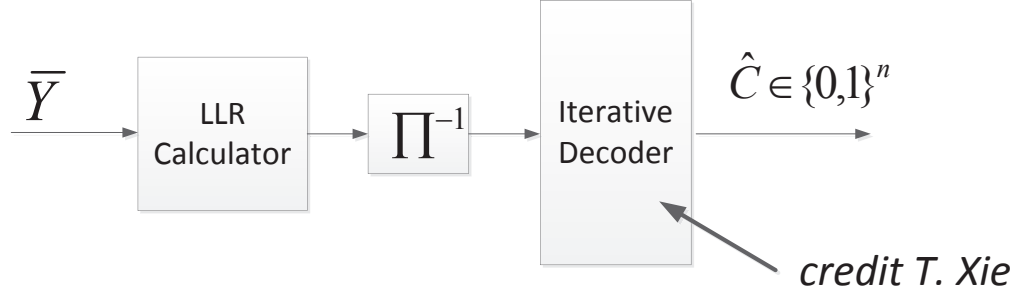
Figure 8.5: Diagram of single-user LDPC decoding

For bit C_i in the codeword C, let the corresponding symbol at the sender be X_n and the constellation set of 16-APSK be $\mathcal{C}$. Then if we adopt a memoryless model as in (8.1), a straightforward method to calculate the LLR for C_i is

$$LLR(C_i) = \log(\frac{P(C_i = 1|Y_n)}{P(C_i = 0|Y_n)})$$
$$= \log(\frac{\sum_{\{X_n|C_i=1, X_n \in \mathcal{C}\}} P(X_n|Y_n)}{\sum_{\{X_n|C_i=0, X_n \in \mathcal{C}\}} P(X_n|Y_n)})$$

97

Based on assumptions that $Z_{0,n}$ is white and Gaussian, and with equal probability of modulator symbols,

$$LLR(C_i) = \log(\frac{\sum_{\{X_n|C_i=1,X_n\in\mathcal{C}\}} f(Y_n|X_n)}{\sum_{\{X_n|C_i=0,X_n\in\mathcal{C}\}} f(Y_n|X_n)})$$

$$= \log(\frac{\sum_{\{X_n|C_i=1,X_n\in\mathcal{C}\}} e^{\frac{-|Y_n-\mu_0(X_n)|^2}{2\sigma^2}}}{\sum_{\{X_n|C_i=0,X_n\in\mathcal{C}\}} e^{\frac{-|Y_n-\mu_0(X_n)|^2}{2\sigma^2}}})$$

where σ^2 is the 1-D variance of the Gaussian residual, and $\mu_0(.)$ is a deterministic function for the memoryless model of RRC channel as stated in (8.1).

Figure 8.6 shows the decoding performance of LPDC in DVB-S2 with the memoryless channel model (8.1) [1]. We assume perfect synchronization and channel estimation. The code rate is 3/4 (or $R = 3$ bits/symbol), ring ratio $\gamma = 2.85$, and block length is 16200 bits as described in DVB-S2. From Figure 8.3, to achieve an information rate 3 bits/symbol, the lowest required SNR is $E_{sat}/N_0 = 11.2$ dB (with IBO=-2 dB), which is used as a threshold to indicate the limiting decoding performance of LDPC. This limit is shown as a vertical solid line on Figure 8.6.

Figure 8.6: LDPC decoding performance on the RRC nonlinear channel
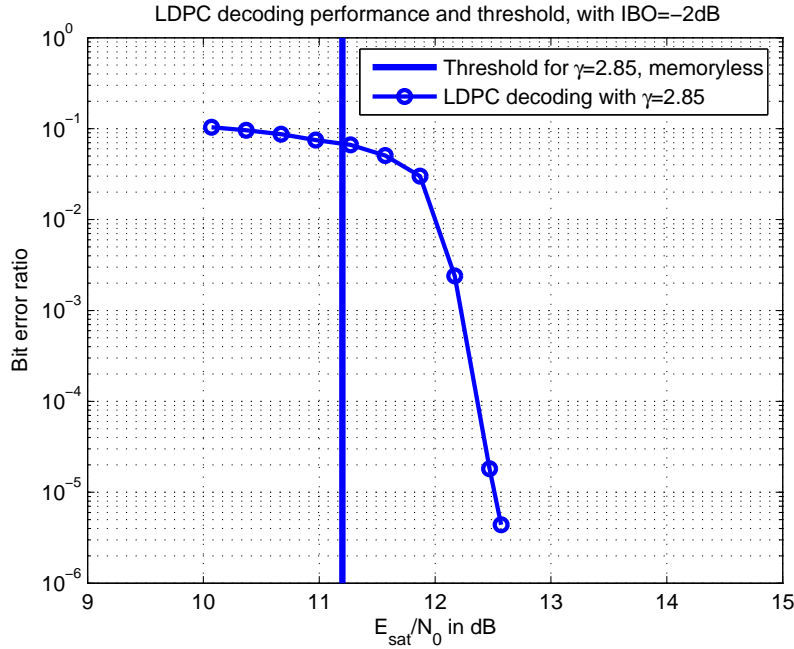
---

[1]Thanks to Tingjun Xie for his MATLAB source codes of LDPC encoding/decoding

Whereas the theoretical limit on SNR is $E_{sat}/N_0 = 11.2$ dB for $R = 3$, the practical decoding system requires E_{sat}/N_0 about =12.4 dB (at $P_b = 10^{-4}$), so there is a gap of about 1.2 dB between theory and practice. Most of this gap is due to the moderate code length of 4050 modulator symbols, and perhaps due to bit-interleaved coded modulation. (Our simulations of a linear channel with the same code show a gap of about 0.7 dB at this error probability.) Perhaps the 'profile' of the LDPC code in DVB-S2 could be improved as well.

8.4   Improving the performance of LDPC decoding

In this section, we present two methods to improve the decoding performance. The first adjusts the ring ratio of 16-APSK constellation, and by doing so, we can obtain a 0.4 dB improvement in simulation performance. (This gain is also reflected in achievable rate calculations, not shown here.) The second method is to adopt a decoder with memory by feeding back the hard decisions made by the LDPC decoder to a subsequent LLR calculation, which brings a separate 0.4 dB improvement.

8.4.1   Better design of 16-APSK constellation

Figure 8.7 shows the distribution of $\bar{Y}$ on the 2-D plane (we only plot the centroids of clusters) when $\gamma = 2.85$ and IBO=-2 dB. The warping and rotation of constellation is due to the nonlinear effect of the amplifier. One solution to mitigate such effect is using large IBO to make the amplifier work quasi-linearly. However, such large IBO suffers from inefficient usage of transmission power, and analysis in previous chapter from the information theory perspective suggests small IBO (IBO=-2 dB), so practically, a satellite transponder mainly works in a near- saturation region, and we will fix the IBO=-2 dB in this thesis.

One simple way to make the output constellation more uniformly-distributed is to change the ring ratio of 16-APSK. Such adjustment pre-compensates the nonlinear effect at the receiver without involving complicated signal processing technique like waveform predistortion, and it helps even in pulse-shaped cases. Figure 8.8 shows the plot of of $\bar{Y}$ on the 2-D plane (we again only plot the centroid of each cluster) for $\gamma = 4$ and IBO=-2 dB, which has slightly better distribution on the 2-D plane.
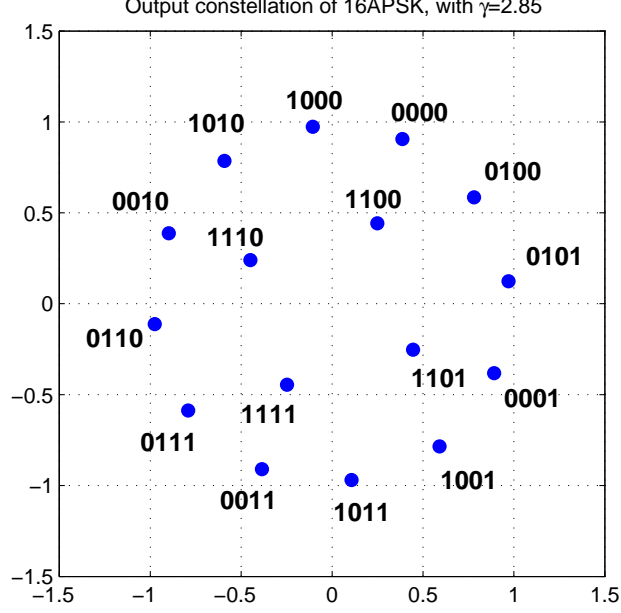
Figure 8.7: Output cluster centroids for $\gamma = 2.85$, and IBO=-2 dB

To test the actual decoding performance, we varied γ within $\{2, 4, 5, 6\}$ and ran a similar simulation on LDPC decoding as we did for $\gamma = 2.85$. Figure 8.9 presents the LDPC decoding performance for different ring ratios. $\gamma = 4$ turns out to be best ring ratio, which improves the performance by approximately 0.4 dB. For the sake of comparison, we show the SNR threshold to achieve a rate of 3 bits/symbol for $\gamma = 4$ obtained from achievable rate analysis. The minimum required SNR is $E_{sat}/N_0 = 11.3$ dB. The gap between threshold and LDPC decoding reduces to roughly 0.8 dB for $\gamma = 4$, but the absolute performance relative to $\gamma = 2.85$ is about 0.4 dB better.

8.4.2   Decoding with memory

To further improve the performance of decoding, we first describe how decoding traditionally operates. In the LDPC decoder as shown in Figure 8.1, the sample sequence $\bar{Y}$ is first passed to a bit LLR calculator. These LLR's are de-interleaved and passed to an iterative decoder, which outputs a tentative binary vector $\hat{\bar{U}}$. Viewing the channel as memoryless, the LLR calculator uses the memory-0 model in (8.1).

The actual channel has memory, so decoding this memory should benefit performance. One method is use a BCJR algorithm to calculate LLR information for each symbol, convert symbol
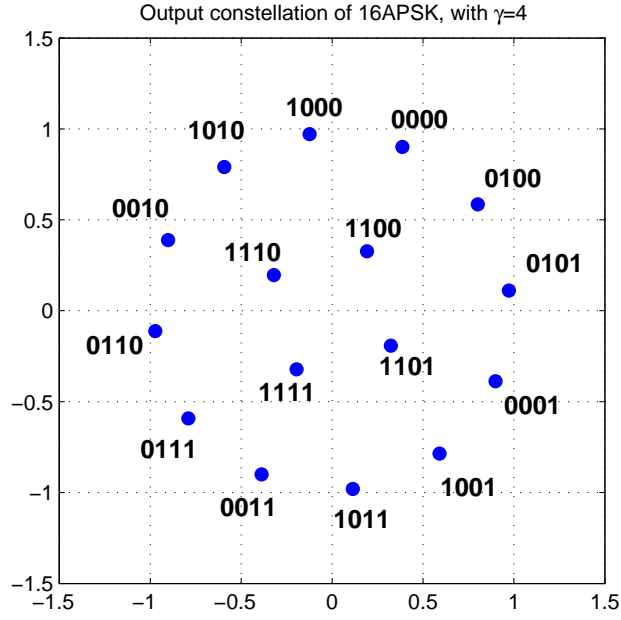
Figure 8.8: Output cluster centriods for $\gamma = 4$, and IBO=-2 dB
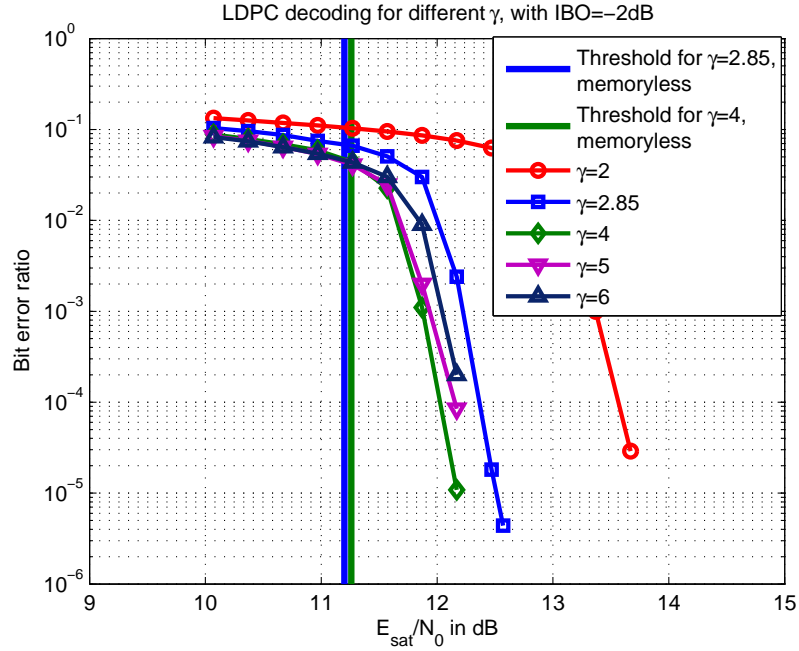

Figure 8.9: Decoding performance with γ varying

LLRs into bit LLRs, and pass them to the LDPC decoder. The overall procedure, depending on the memory length adopted in the discrete channel model, is generally very complicated.

Instead we propose a decoder with memory, which doesn't involve high computational complexity. The diagram of decoding is illustrated in Figure 8.10. Following initial decoding, the iterative decoder outputs $\bar{C}$, a concatenated sequence of $\hat{\bar{U}}$ and parity check bits, and indicates whether decoding has succeeded. If the decoding is unsuccessful, then $\bar{C}$ is reinterleaved and mapped to logical symbols. Under the assumption that the previous symbol and future symbol to X_n have been correctly found, we may appeal to a larger lookup table storing $\mu_2(X_{n-1}, X_n, X_{n+1})$ values with the earlier and future symbol pinned, to obtain a new set of LLR's for the current symbol. This recalculation (with memory 2) follows

$$LLR(C_i) = \log \left(\frac{\sum_{\{X_n | C_i = 1, X_n \in \mathcal{C}\}} e^{\frac{-|Y_n - \mu_2(\widehat{X_{n-1}}, X_n, \widehat{X_{n+1}})|^2}{2\sigma^2}}}{\sum_{\{X_n | C_i = 0, X_n \in \mathcal{C}\}} e^{\frac{-|Y_n - \mu_2(\widehat{X_{n-1}}, X_n, \widehat{X_{n+1}})|^2}{2\sigma^2}}} \right)$$

where $\mu_2(.)$ is memory-2 model for the RRC nonlinear channel as described in (8.2). Notice that this formula is the similar to the first pass, but aided by decoder feedback side-information on past and future symbols. Practically, $\mu_2(.)$ can be obtained from training data, and pre-stored in a lookup table of size $16^3 = 4096$ words.

The LLRs calculated again go into the iterative decoder, which outputs some new $\hat{C} \in \{0, 1\}^n$. If desired, $\hat{C} \in \{0, 1\}^n$ can again be fedback for another LLR calculation, though we have found no benefit to more than two-pass decoding.

Figure 8.11 presents the simulated performance for this decoder. We choose $\gamma = 4$, which has been shown to achieve best performance for a memoryless decoder, and we decode in three ways: memoryless decoding, decoding with feedback once, and decoding with feedback twice. As can be seen from Figure 8.11, there is a 0.4 dB improvement for such feedback-aided decoding, and the improvement from additional cycles is negligible.

Figure 8.12 presents the decoding performance when we adopt even longer memory length. We use the memory-4 model shown in (8.3), which is slightly more accurate than the memory-2 model. For the sake of comparison, we run simulations for the ideal case, in which the LLR(2) calculator uses the exact symbol information instead of hard decisions from the iterative decoder (which can be wrong). Such ideal case shows the best achievable performance for a memory decoder. As can be seen from Figure 8.12, the performance of memory-4 feedback decoding has no advantage
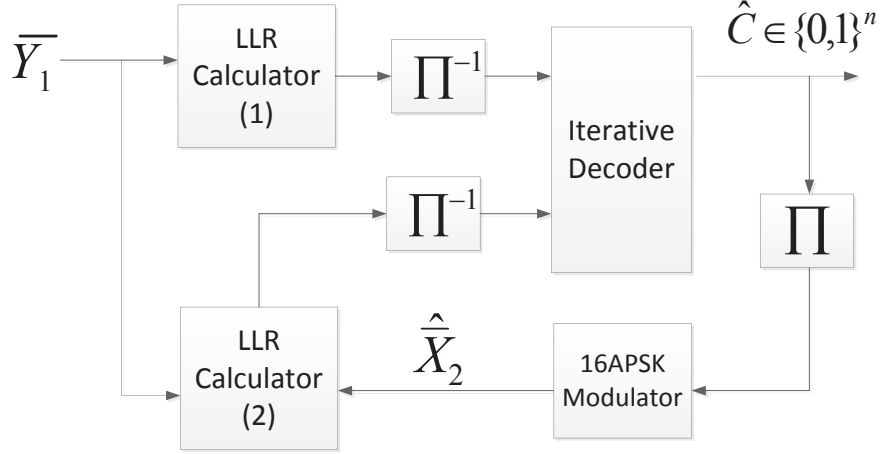
Figure 8.10: Diagram of LDPC decoding with feedback
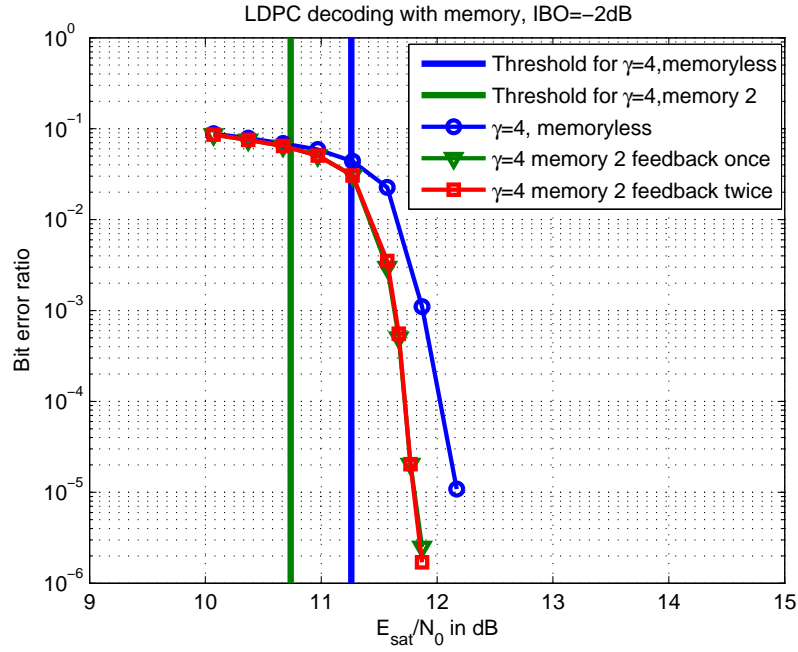
Figure 8.11: LDPC decoding performance with feedback

over the memory-2 model, indicating a memory-2 model sufficiently captures the memory of the nonlinear channel. In addition, for the memory 2 case, feedback once can achieve almost the same performance as the ideal memory-2 decoder, while there is about 0.1 dB gap to ideal decoding for the memory 4 case.

Finally, Figure 8.13 summarizes the performance improvement achieved in this paper that we can improve the performance by about 0.8 dB by adjusting ring ratio and using a memory-2 decoder.
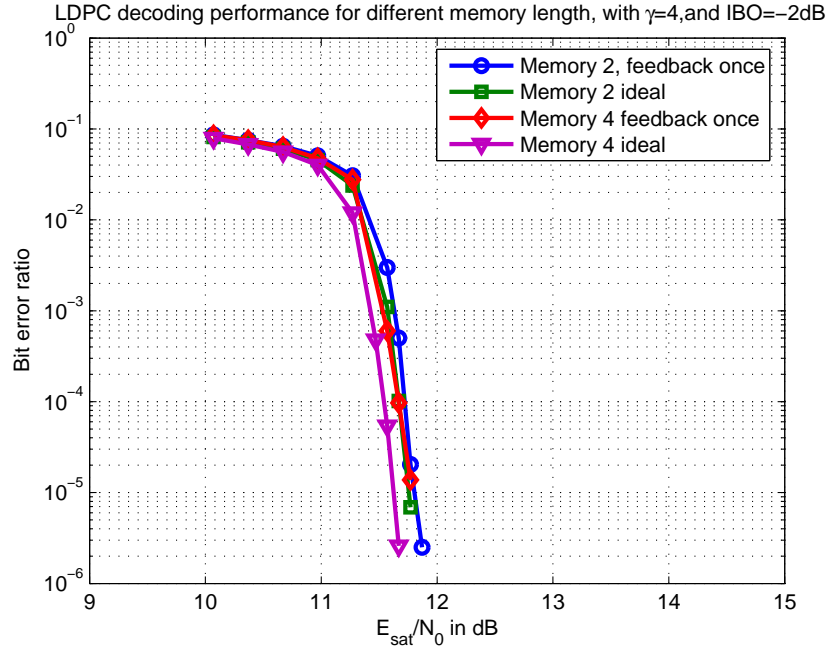
Figure 8.12: Effect of longer memory length for LDPC decoding

Figure 8.13 also presents the LDPC decoding performance for a linear channel, which indicates the limitation of blocklength 16200 LDPC coding.
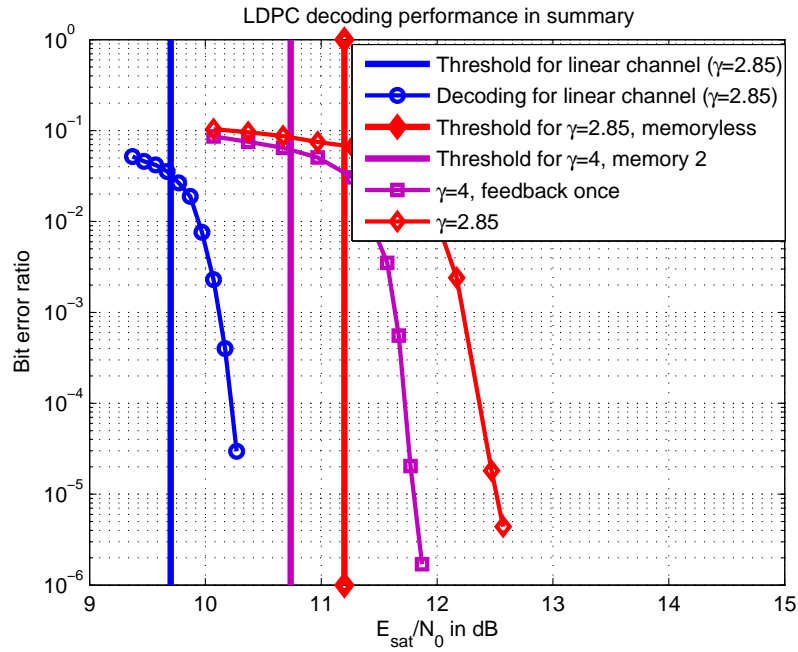
Figure 8.13: LDPC decoding performance summary

8.5 Summary

Higher-order modulation 16APSK with LDPC channel coding, as stated in DVB-S2, has been studied from both an information-theoretic perspective, as well as simulation of the actual standard coded modulation. We find that achievable rates are predictive of real performance, in terms of best ring ratio for 16-APSK as well as IBO. In addition, we showed how to improve the LDPC decoding performance, based on a feedback arrangement for exploiting the inherent nonlinear memory of the channel. Simulation results show that such feedback-based memory-2 decoding scheme provides 0.4 dB improvement, and there is less than 0.1 dB gap from an ideal (genie-aided) memory-2 decoding.

Chapter 9

Coded Two-way Communication over Nonlinear Satellite Channel

In Chapter 8, we studied the performance of LDPC decoding on a single-user nonlinear channel. In this Chapter, we will evaluate the performance of AF protocol for two users over the same nonlinear channel. Figure 9.1 shows the system diagram. The main difference from the diagram of one-user model is that each terminal can make use of side-information to decode the message it wants. In addition, the input to the amplifier is a combination of two pulse-shaped signals, so the downlink power at the satellite is "shared" by two signals, and the effective signal to noise ratio for each signal is therefore smaller.

Figure 9.1: Diagram for decoding on two-way relay nonlinear channel

The effective channel model between modulation and soft decoding is actually the same as what we have seen for the uncoded system. Assume that the channel input messages (after modulation) $\bar{X}_1 = \{X_{1,1}, X_{1,2}, ..., X_{1,n}\}$ (generated from T_1) and $\bar{X}_2 = \{X_{2,1}, X_{2,2}, ..., X_{2,n}\}$ (generated from T_2)

are complex sequences drawn from some 2-D constellation set $\mathcal{C}$. Then the continuous-time outputs of the channel are [1]

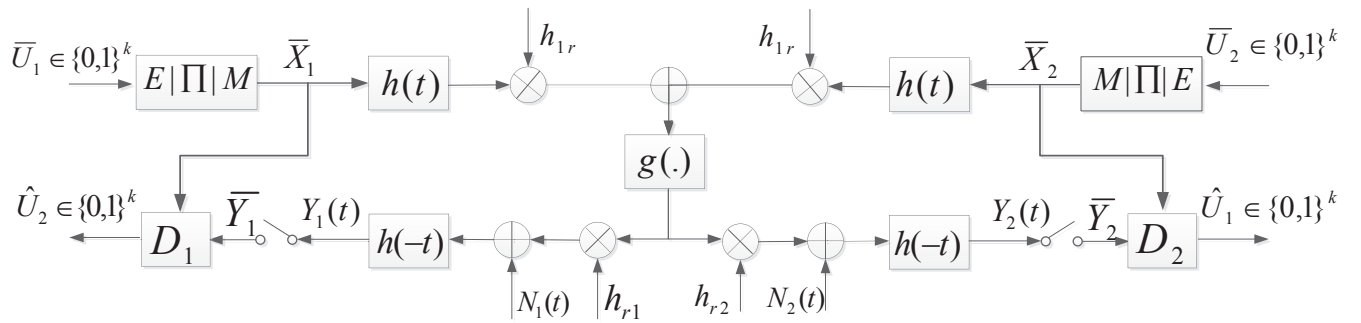$$Y_1(t) = h_{r1}h(-t) \bigstar g \left(h_{1r} \sum_l X_{1,l} h(t - lT) + h_{2r} \sum_l X_{2,l} h(t - lT) \right) + N_1(t)$$

$$Y_2(t) = h_{r2}h(-t) \bigstar g \left(h_{1r} \sum_l X_{1,l} h(t - lT) + h_{2r} \sum_l X_{2,l} h(t - lT) \right) + N_2(t)$$

where $Y_1(t)$ and $Y_2(t)$ are output signals at T_1 and T_2 respectively, and $N_1(t)$, $N_2(t)$ are white Gaussian noise. Matched filtering is adopted for the sake of practical convenience, though it does not necessarily provide sufficient statistics for decoding on a nonlinear channel.

Without loss of generality, we consider the decoding of X_2 at terminal T_1. At T_1, the output of the matched filter is a complex-valued sequence $\bar{Y}_1$. As stated in Chapter 6, it is not trivial or efficient to do interference compensation in this nonlinear system. The solutions presented later in this Chapter are based on some truncated memory models similarly as used in the uncoded symbol detection.

9.1 Decoding synchronous signals

9.1.1 Memoryless decoding

Similar to the uncoded symbol detection in Chapter 6, we adopt a short memory model to approximate the actual channel, in which the simplest is a memoryless model as below.

$$Y_{1,n}|_{t=nT} = \mu_0(X_{1,n}, X_{2,n}) + Z_{0,n} \tag{9.1}$$

where $Z_{0,n}$ subsumes additive noise and unmodeled nonlinear distortion.

For bit C_i in the codeword C, let the corresponding symbol at senders be $X_{1,n}$ and $X_{2,n}$ and the constellation set be $\mathcal{C}$ (16APSK in this Chapter). If we adopt the memoryless model as shown in (9.1), the LLR for C_i is

---

[1]We assume two signals are perfectly synchronized

$$LLR(C_i) = \log\left(\frac{P(C_i = 1|Y_{1,n})}{P(C_i = 0|Y_{1,n})}\right)$$

$$= \log\left(\frac{\sum_{\{X_{2,n}|C_i=1,X_{2,n}\in\mathcal{C}\}} P(X_{2,n}|Y_{1,n})}{\sum_{\{X_{2,n}|C_i=0,X_{2,n}\in\mathcal{C}\}} P(X_{2,n}|Y_{1,n})}\right)$$

Based on the approximation that $Z_{0,n}$ is white and Gaussian, and with equal probability of modulator symbols,

$$LLR(C_i) = \log\left(\frac{\sum_{\{X_{2,n}|C_i=1,X_{2,n}\in\mathcal{C}\}} f(Y_{1,n}|X_{2,n})}{\sum_{\{X_{2,n}|C_i=0,X_{2,n}\in\mathcal{C}\}} f(Y_{1,n}|X_{2,n})}\right)$$

$$= \log\left(\frac{\sum_{\{X_{2,n}|C_i=1,X_{2,n}\in\mathcal{C}\}} e^{\frac{-|Y_{1,n}-\mu_0(X_{1,n},X_{2,n})|^2}{2\sigma^2}}}{\sum_{\{X_{2,n}|C_i=0,X_{2,n}\in\mathcal{C}\}} e^{\frac{-|Y_{1,n}-\mu_0(X_{1,n},X_{2,n})|^2}{2\sigma^2}}}\right)$$

where σ^2 is the effective 1-D variance of the Gaussian residual $Z_{0,n}$. Figure 9.2 shows the diagram for the decoder.
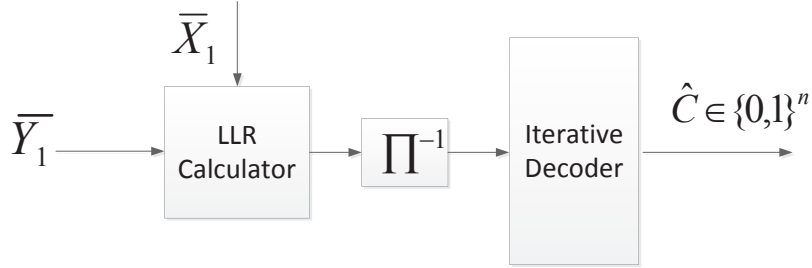

Figure 9.2: Diagram of 2-user memoryless LDPC decoding

9.1.2 Improved "memoryless" decoding

An improved "memoryless" decoding scheme is available, if we add two more symbols from the *side information* to the channel model as below:

$$Y_{1,n} = \mu_0'(X_{n-1}^1, X_n^1, X_{n+1}^1, X_n^2) + Z_{0,n}' \tag{9.2}$$

Then the LLR for bit C_i can be calculated as

$$LLR(C_i) = \log(\frac{P(C_i = 1|Y_{1,n})}{P(C_i = 0|Y_{1,n})})$$

$$= \log(\frac{\sum_{\{X_{2,n}|C_i=1, X_{2,n} \in \mathcal{C}\}} P(X_{2,n}|Y_{1,n})}{\sum_{\{X_{2,n}|C_i=0, X_{2,n} \in \mathcal{C}\}} P(X_{2,n}|Y_{1,n})})$$

$$= \log\left(\frac{\sum_{\{X_{2,n}|C_i=1, X_{2,n} \in \mathcal{C}\}} e^{\frac{-|Y_{1,n} - \mu_0'(X_{1,n-1}, X_{1,n}, X_{1,n+1}, X_{2,n})|^2}{2\sigma^2}}}{\sum_{\{X_{2,n}|C_i=0, X_{2,n} \in \mathcal{C}\}} e^{\frac{-|Y_{1,n} - \mu_0'(X_{1,n-1}, X_{1,n}, X_{1,n+1}, X_{2,n})|^2}{2\sigma^2}}}\right)$$

The diagram of the decoder is the same as shown in Figure 9.2 with the only difference being in the LLR calculation. The size of the LLR table is now $|\mathcal{C}|^3$. In the following section, we examine incorporation of memory in decoding the signal X_2.

9.1.3   Decoding with memory for desired message

In Chapter 6 and 8, we have seen the advantage of the using a memory-2 model in both uncoded two-way nonlinear channel, and coded single-user nonlinear channel. For the coded two-signal problem in this Chapter, we also want to decode while incorporating memory. The solution is similar to what we have presented for the single-user case, which is based on a hard-decision feedback mechanism, with the difference that side information of $\bar{X}_1$ is involved in the LLR calculation. The diagram of the decoder is shown in Figure 9.3. The problem of course is that we do not have prior knowledge about X_2.

The initial decoding pass is memoryless using the channel model (9.1). Assume the initial decoding outputs $\hat{C}$, a concatenated sequence of $\hat{\bar{U}}$ and parity check bits, and indicates whether decoding has succeeded. If the decoding is unsuccessful, then $\hat{C}$ is reinterleaved and mapped to logical symbols. Under the assumption that the previous symbol and future symbol to X_n have been correctly found, we adopt a more accurate memory-2 model as below, (requiring a bigger μ lookup table):

$$Y_{1,n} = \mu_2(X_{n-1}^1, X_n^1, X_{n+1}^1, X_{n-1}^2, X_n^2, X_{n+1}^2) + Z_{2,n} \tag{9.3}$$
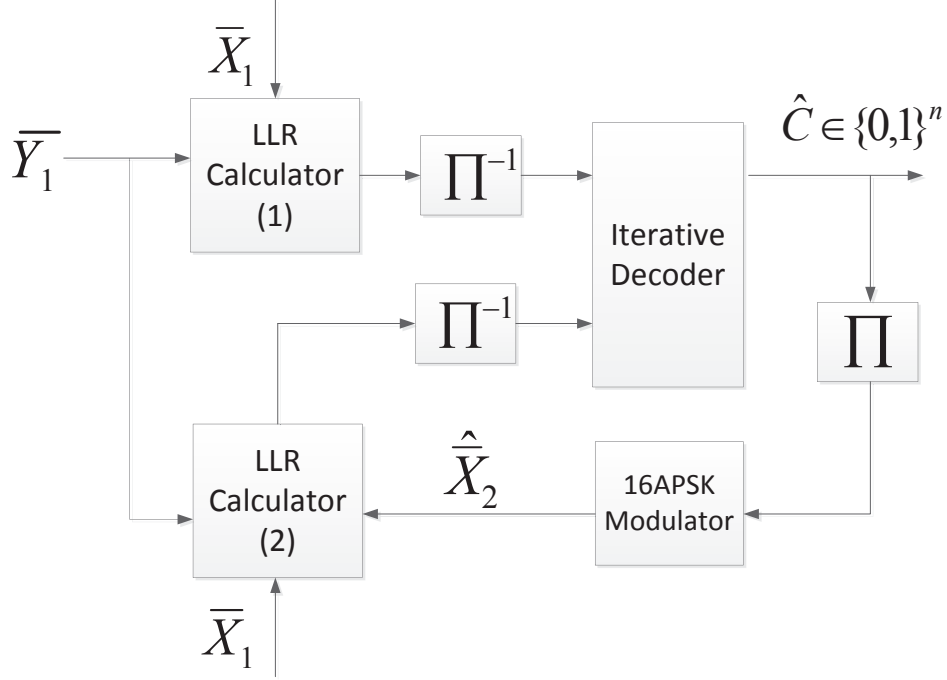
Figure 9.3: Diagram for decoding with feedback

The LLR for bit C_i is then calculated in LLR Calculator (2) as

$$
\begin{aligned}
LLR(C_i) &= \log\left(\frac{P(C_i = 1|Y_{1,n})}{P(C_i = 0|Y_{1,n})}\right) \\
&= \log\left(\frac{\sum_{\{X_{2,n}|C_i=1,X_{2,n}\in\mathcal{C}\}} P(X_{2,n}|Y_{1,n})}{\sum_{\{X_{2,n}|C_i=0,X_{2,n}\in\mathcal{C}\}} P(X_{2,n}|Y_{1,n})}\right) \\
&= \log\left(\frac{\sum_{\{X_{2,n}|C_i=1,X_{2,n}\in\mathcal{C}\}} e^{\frac{-|Y_{1,n}-\mu_2(X_{1,n-1},X_{1,n},X_{1,n+1},\hat{X_{2,n-1}},X_{2,n},\hat{X_{2,n+1}})|^2}{2\sigma^2}}}{\sum_{\{X_{2,n}|C_i=0,X_{2,n}\in\mathcal{C}\}} e^{\frac{-|Y_{1,n}-\mu_2(X_{1,n-1},X_{1,n},X_{1,n+1},\hat{X_{2,n-1}},X_{2,n},\hat{X_{2,n+1}})|^2}{2\sigma^2}}}\right)
\end{aligned}
$$

The LLRs calculated again go into the iterative decoder, which outputs some new $C \in \{0,1\}^k$. If desired, $C \in \{0,1\}^k$ can again be fedback for another LLR calculation, but we have found there is no benefit to have more than two-pass decoding, just as in single-user case.

9.1.4   Simulation results for synchronous signals

In this section, we will present the simulation results for the above-described three decoders. Similar as the one-signal case, we focus on an intermediate case with rate 3/4 coding mapped to 16-APSK,

and the code blocklength is 16200 bits. We fix the ring ratio $\gamma = 2.85$ as in the DVB-S2 standard,[2] and make two signals have equal power. We consider four IBO options, and for each IBO, we evaluate the performance in terms of bit error ratio versus SNR.

Figure 9.4 shows the decoding performance of the memoryless decoder using model (9.1). With channel coding, we are able to tolerate higher level of distortion. In contrast to the coded single-user case, larger IBO is preferred for the two-way relay channel. Two-user transmission is more vulnerable to nonlinear distortion, so increasing IBO makes the amplifier work more linearly, therefore the nonlinear distortion is reduced, and the decoding performance is improved. However, larger IBO also reduces the power of the output of the amplifier, which can degrade the decoding performance. IBO=-6 dB seems to be a good balance point for this trade-off.
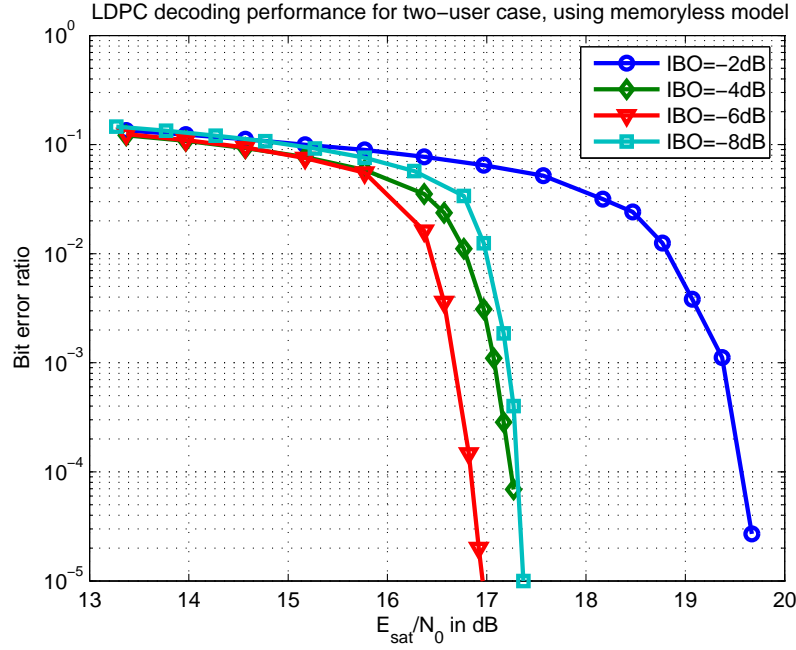
Figure 9.4: The performance of the memoryless decoder

Figure 9.5 presents the simulation results for the memory-2 decoder. Comparing with the memoryless decoder, the improvements are roughly 1.2 dB, 0.5 dB, 0.25 dB, 0.1 dB for IBO=-2 dB, -4 dB, -6 dB, -8 dB respectively. The reason that improvement decreases with IBO is that larger IBO reduces nonlinear distortion which also reduces the room for improvement. IBO=-6 dB is still the winner of all IBOs considered, but the gap between IBO=-6 dB and IBO=-4 dB

---

[2]For the two-way relay channel, there is negligible difference in decoding performance for different γ

is greatly reduced as compared with the memoryless decoding. The reason is that the memory-2 model better models the channel ISI than the memoryless model, so the advantage of IBO=-6 dB over IBO=-4 dB is no longer that significant as in a memoryless model. Therefore, for IBO=-4 dB, the pros of its slightly larger output power can almost compensate for its cons in less model accuracy.
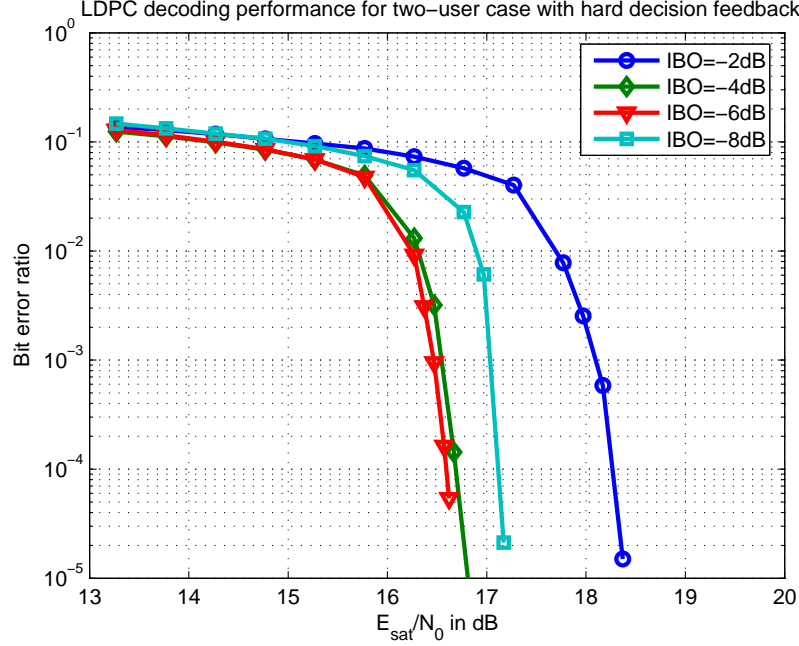
Figure 9.5: Performance of the memory-2 decoder

Figure 9.6 compares the best achievable performance between different decoders with IBO=-6 dB. The improved "memoryless" decoder performs almost the same as the memoryless decoder, which is similar to the uncoded case as shown in Chapter 6. The memory-2 decoder brings about 0.25 dB improvement as compared with the memoryless decoder. For reference, we run simulations for the ideal case, in which the LLR(2) calculator uses the exact symbol information instead of hard decisions from the iterative decoder (which can be wrong). Such an ideal decoder gives the best possible performance achievable for a memory-2 decoder. As can be seen from Figure 9.6, the performance gap between such ideal decoder and the feedback-based decoder is less than 0.1 dB, which means the feedback mechanism works well, though it is much less complicated than a BCJR-based decoder.
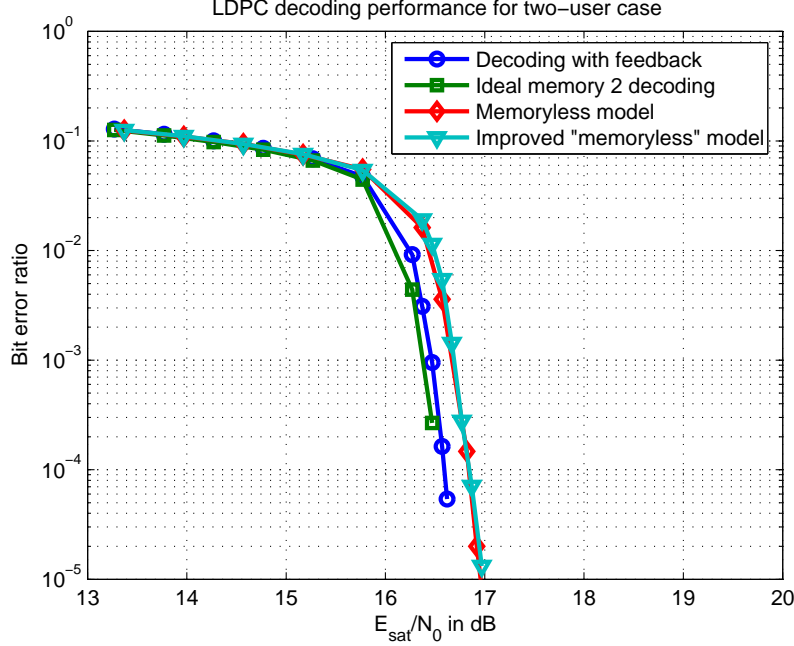
Figure 9.6: Performance for each decoder, with IBO=-6 dB

9.2   Decoding asynchronous signals

In previous sections, we have assumed that both signal are perfectly synchronized. However, practically two signals can be asynchronous due to propagation delay. Assuming the relative time delay between f_{12} and f_{21} be t_0, the continuous-time output of the channel at T_1 is

$$Y_1(t) = h_{r1}h(-t)\bigstar g\left(h_{1r}\sum_l X_{1,l}h(t - lT - t_0) + h_{2r}\sum_l X_{2,l}h(t - lT)\right) + N_1(t)$$

As in the synchronous case, we can truncate the length of memory to approximate the actual channel. Due to the existence of strong linear ISI from T_1's signal,[3] we adopt slightly different ways to absorb ISI into the mapping μ, and the effective channel models become

$$Y_{1,n}|_{t=nT} = \mu_{21}(X_{1,n}, X_{1,n+1}, X_{2,n}) + Z_{21,n} \tag{9.4}$$

$$Y_{1,n}|_{t=nT} = \mu_{41}(X_{1,n-1}, X_{1,n}, X_{1,n+1}, X_{1,n+2}, X_{2,n}) + Z_{41,n} \tag{9.5}$$

$$Y_{1,n}|_{t=nT} = \mu_{23}(X_{1,n}, X_{1,n+1}, X_{2,n-1}, X_{2,n}, X_{2,n+1}) + Z_{23,n} \tag{9.6}$$

---

[3]Receiver timing is set for the signal from T_2, hence timing offset for T_1's signal exists

9.2.1 Memoryless decoding

We may adopt a memoryless model as shown in (9.4) to approximate the actual channel. Notice that memoryless refers only to the desired signal X_2 but not for T_1's own signal. The corresponding LLR for bit $C_i \in C$ can be calculated as

$$
\begin{aligned}
LLR(C_i) &= \log\left(\frac{P(C_i=1|Y_{1,n})}{P(C_i=0|Y_{1,n})}\right) \\
&= \log\left(\frac{\sum_{\{X_{2,n}|C_i=1,X_{2,n}\in\mathcal{C}\}} P(X_{2,n}|Y_{1,n})}{\sum_{\{X_{2,n}|C_i=0,X_{2,n}\in\mathcal{C}\}} P(X_{2,n}|Y_{1,n})}\right) \\
&= \log\left(\frac{\sum_{\{X_{2,n}|C_i=1,X_{2,n}\in\mathcal{C}\}} e^{\frac{-|Y_{1,n}-\mu_{21}(X_{1,n},X_{1,n+1},X_{2,n})|^2}{2\sigma^2}}}{\sum_{\{X_{2,n}|C_i=0,X_{2,n}\in\mathcal{C}\}} e^{\frac{-|Y_{1,n}-\mu_{21}(X_{1,n},X_{1,n+1},X_{2,n})|^2}{2\sigma^2}}}\right)
\end{aligned}
$$

9.2.2 Improved memoryless decoding

We can improve the decoding performance by adopting a more accurate model as shown in (9.5), which includes two more symbols from side information. Then the corresponding LLR for bit $C_i \in C$ can be calculated as

$$
\begin{aligned}
LLR(C_i) &= \log\left(\frac{P(C_i=1|Y_{1,n})}{P(C_i=0|Y_{1,n})}\right) \\
&= \log\left(\frac{\sum_{\{X_{2,n}|C_i=1,X_{2,n}\in\mathcal{C}\}} P(X_{2,n}|Y_{1,n})}{\sum_{\{X_{2,n}|C_i=0,X_{2,n}\in\mathcal{C}\}} P(X_{2,n}|Y_{1,n})}\right) \\
&= \log\left(\frac{\sum_{\{X_{2,n}|C_i=1,X_{2,n}\in\mathcal{C}\}} e^{\frac{-|Y_{1,n}-\mu_{41}(X_{1,n-1},X_{1,n},X_{1,n+1},X_{1,n+2},X_{2,n})|^2}{2\sigma^2}}}{\sum_{\{X_{2,n}|C_i=0,X_{2,n}\in\mathcal{C}\}} e^{\frac{-|Y_{1,n}-\mu_{41}(X_{1,n-1},X_{1,n},X_{1,n+1},X_{1,n+2},X_{2,n})|^2}{2\sigma^2}}}\right)
\end{aligned}
$$

For the synchronous case, merely adding more symbols from side information $\bar{X}_1$ to the channel model helps little in decoding performance. However, for the asynchronous case, due to the existence of strong linear ISI from $\bar{X}_1$, the decoding performance can be significantly improved as will be shown below.

9.2.3 Memoryless decoding with ISI compensator

We can further improve the decoding performance by adding more symbols from $\bar{X}_1$ into the channel model. However, as more symbols are included in the model, the μ-table size grows exponentially. For example for 16APSK, the size of the μ-table is enlarged by 16 for each extra symbol added to the channel model. The distortion from these extra symbols are mainly linear ISI, so one possible way to improve the channel model without increasing the size of μ-table is to add a linear ISI compensator to cancel the strong linear ISI residue, and the effective channel model adopted is

$$Y_{1,n}|_{t=nT} = \mu_{41}(X_{1,n-1}, X_{1,n}, X_{1,n+1}, X_{1,n+2}, X_{2,n}) + \sum_{-4 \leq i \leq -2,\ 2 \leq i \leq 4} c_i X_{1,n+i} + Z'_{41,n} \tag{9.7}$$

where $\bar{c} = [c_{-4}\ c_{-3}\ c_{-2}\ c_2\ c_3\ c_4]$ are the coefficients for the linear terms that can be obtained from LMS (least mean square) training. The variance of $Z'_{41,n}$ saturated if more than 6 symbols are added to the original model, so we fix the length of the compensator to be 6. Figure 9.7 shows the diagram for such decoding scheme. Linear ISI is canceled from $\bar{Y}_1$ before it goes in the LLR calculator, and the LLR calculation is the same as the improved memoryless decoding shown in previous section.
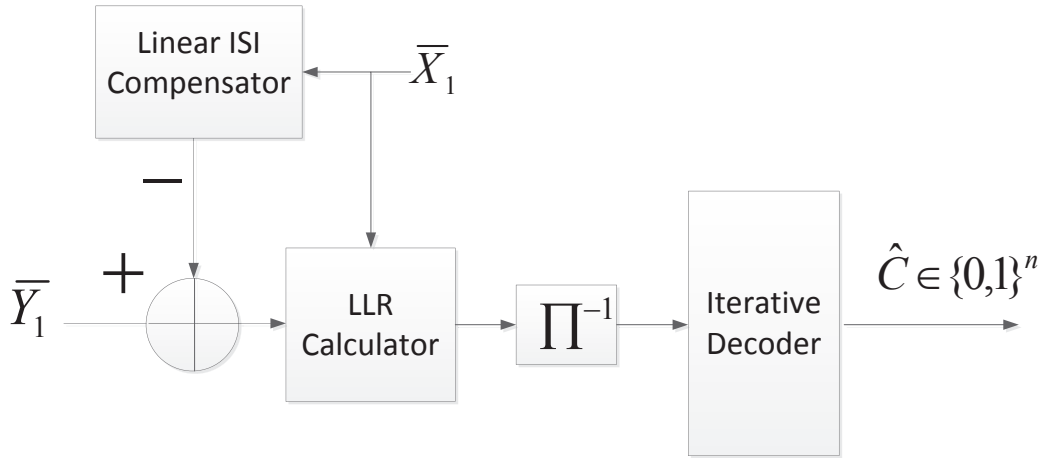

Figure 9.7: Memoryless decoding with linear ISI compensator

Decoding with memory

Studies in the synchronous case have shown that the performance can be further improved if a memory-2 model is used for decoding. For the asynchronous case, we can adopt the same idea used

in the synchronous case that hard decisions from the first pass is fedback to a new LLR calculator which adopts a channel model as below

$$Y_{1,n}|_{t=nT} = \mu_{23}(X_{1,n}, X_{1,n+1}, X_{2,n-1}, X_{2,n}, X_{2,n+1}) + Z_{23,n} \qquad (9.8)$$

and the new LLR information is passed to the iterative decoder.

However, due to the existence of strong linear ISI from $\bar{X}_1$, the variance of $Z_{23,n}$ is much larger than $Z'_{41,n}$ that the second pass to the iterative decoder can not bring any performance improvement. The solution again is to use a linear ISI compensator to cancel the strong linear ISI caused by $\bar{X}_1$, and the corresponding channel model adopted is

$$Y_{1,n}|_{t=nT} = \mu_{23}(X_{1,n}, X_{1,n+1}, X_{2,n-1}, X_{2,n}, X_{2,n+1}) + \sum_{-4 \leq i \leq -1,\ 1 \leq i \leq 4} c'_i X_{1,n+i} + Z'_{23,n}$$

where $\bar{c'} = [c'_{-4}\ c'_{-2}\ c'_{-3}\ c'_{-1}\ c'_1\ c'_2\ c'_3\ c'_4]$ are the coefficients for the linear terms that can be obtained from training. The variance of $Z'_{23,n}$ saturated if more than 8 symbols are added to the model shown in (9.7), so we fix the length of the compensator to be 8. Figure 9.8 presents the diagram for this memory-2 decoder.

9.2.4   Simulation results for asynchronous signals

This section presents the decoding performance for each of the four decoding methods proposed. The modulation scheme again is chosen to be 16APSK with $\gamma = 2.85$, and the block length of the LDPC encoder is 16200 bits as stated in DVB-S2, which is the same as in synchronous case. We make the two signals have the same average power, and choose the relative delay between them to be half symbol time. From the simulation result of the synchronous case, we know that IBO=-6 dB tends to be an optimal choice within the IBO options considered, so we fix the IBO=-6 dB for the simulation of asynchronous case. Figure 9.9 shows the simulation result.

There are four performance curves corresponding to the four decoding schemes proposed. The original memoryless decoding (Option 1) has a severe error floor, which means there is significant inadequacy of the channel model adopted as shown in (9.4). The reason is very clear that there exists
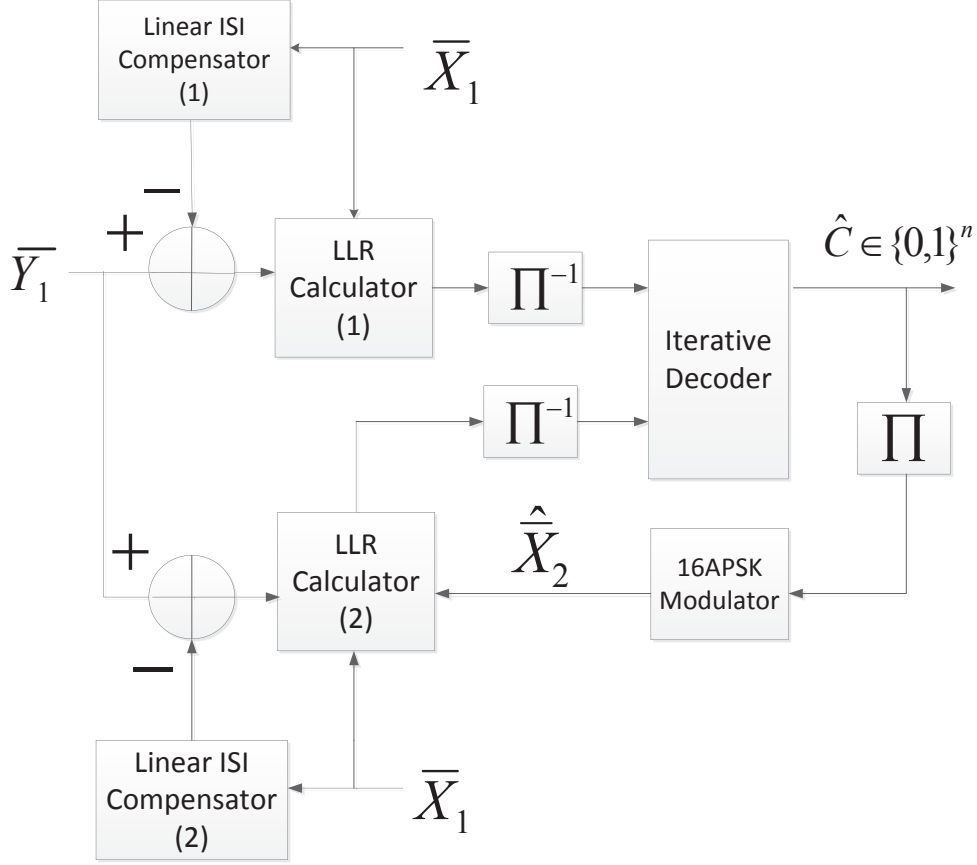
Figure 9.8: Memory-2 decoding for asynchronous signals

strong linear ISI from $\bar{X}_1$, which is not captured in channel model (9.4). In option 2 with model (9.5), two more symbols are added to make the channel model capture more linear ISI, and the performance is significantly improved as compared with option 1. For option 3, linear ISI is further captured in the channel model that, the performance is improved by roughly 1.1 dB as compared with option 2. Finally for option 4, by means of a memory-2 decoder, the performance is further improved by roughly 0.25 dB. By comparing the best performance achieved for both synchronous case and asynchronous, we find that there is no visible loss due to the asynchronization between the two signals.

9.3 Summary

This Chapter focuses on the coded transmission of bidirectional relaying with the AF protocol over a nonlinear satellite channel, in the context of DVB-S2. We first consider the case when both signals
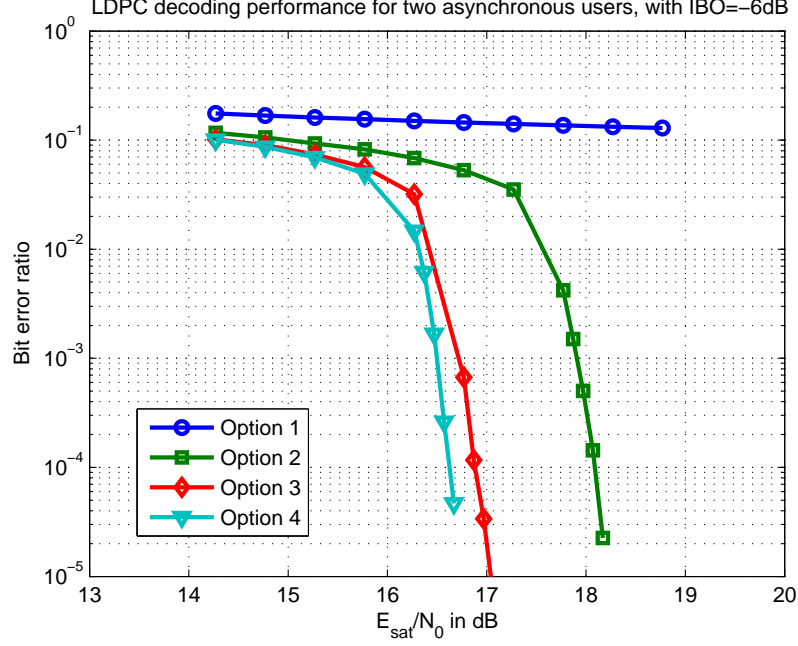
Figure 9.9: Decoding performance for asynchronous signals

are perfectly synchronized, and propose memoryless and memory-2 decoding schemes. Simulation results show that comparing with single-user transmission, the AF protocol is more vulnerable to nonlinear distortion, and a larger IBO like -6 dB is preferred. In addition, the memory-2 decoding scheme can improve the performance of memoryless decoding by 0.25 dB without involving too much computational complexity. We then move to the asynchronous case which induces strong linear ISI, and an original memoryless decoding including only 2 symbols from side information shows a severe error floor. However, by means of two linear ISI compensators we are able to achieve almost the same decoding performance as in synchronous case.

Chapter 10

Conclusions and Future Work

This thesis works on the two-way relay satellite channel, mainly from an information-theoretical perspective. We consider both linear and nonlinear channel scenarios.

For the linear satellite channel, we first summarize existing protocols including AF, MAC-BSI, MAC-XOR, MAC-NC, and NLC. By comparing their achievable rate regions, maximum rate sums and system complexity requirements, we find that the AF protocol can achieve a good, though not always best performance in all SNR and rate regions considered in this thesis. In addition, AF is flexible on signal synchronization, has minimum code knowledge requirements on each terminal, and is compatible with legacy satellites, so we adopt AF protocol for our later study.

Next, we consider the problem from a different view, where two terminals wish to communicate with some target rates that may be well less than what the full transponder resources can support. Then, the system operator would wish to lease a fraction of a transponder, and share the remaining resources with other non-interfering services. It is of practical interest to minimize the recurring cost. Assuming a generic pricing structure, we model it as a joint bandwidth and power optimization problem. A solution to this problem is given that determines the most efficient bandwidth and power resources requests to reduce the transponder cost.

Finally for a linear channel model, we extend the AF protocol to a three-terminal hub-remote exchange channel. Both three-flow and four-flow cases are studied. Theoretical analysis and numerical plots show that for the three-flow case, repetition of sending common forward message should be avoided, and at least one of the decoding schemes of three-user simultaneous signalling performs the best among all scenarios considered. For the four-flow case, three-user simultaneous signalling

only shows advantage when the hub terminal has much higher SNR than remote terminals or when remote terminal rate requirements are emphasized. In addition, the gains from three simultaneous signalling are generally smaller than one might expect based on multi-terminal information theory, due to the constraint on relay's power and power sharing.

For the nonlinear satellite channel, we first study the uncoded transmission of two-user AF in a nonlinear channel. Based on some finite-state channel models, we present memoryless and memory-2 symbol detection schemes. Simulation results show that in contrast to single-user transmission in nonlinear channel, a much larger IBO (-8 dB) is preferred for the AF protocol, which means AF is more vulnerable to nonlinear distortion, at least without coding. In addition, incorporating memory during demodulation and adopting a large IBO (-8 dB) can eliminate the visible error floor for uncoded transmission.

Next, we study the information-theoretic potential of nonlinear satellite channels, for both single-user per transponder operation and simultaneous two-way, on-frequency relaying to improve spectrum efficiency as much as 100%. Coded 16-APSK transmission is studied from the perspective of achievable information rate and channel simulation. We develop an analytic nonlinear model with finite memory for these two scenarios, and compute achievable rate regions, as a function of link SNR's and input backoff (IBO) of the nonlinear amplifier. These studies can tell us what is the maximum allowed information rate or minimum required SNR for a reliable transmission over NGC, and help us predict the best possible performance of channel coding.

Finally for the nonlinear satellite channel, we work on the channel coding performance over NGC in the context of DVB-S2. Both single-user and two-user (AF) scenarios are studied. For the single-user case, we show that modification to the constellation ring ratio for 16APSK can provide significant savings in SNR, at least for the adopted nonlinear Saleh model. In addition, a two-pass decoding that employs hard-decision feedback from the LDPC decoder is able to gain another 0.4 dB in performance without significant complexity increase. Simulation results in decoding performance is well predicted by our achievable rate analysis. For the two-user case (AF) an even larger backoff is preferred (-6 dB). We propose a memory-2 scheme which is based on the feedback mechanism and avoids the complicated calculation of the BCJR algorithm. Simulation results show that it brings 0.25 dB improvement as compared with a memoryless decoder, and there is only a small performance gap between the the proposed memory-2 decoder and an ideal memory-2 decoding.

We also consider the scenario when two signals are asynchronous, say there is half symbol delay between them. Such asynchronization brings strong linear ISI that a memoryless decoding has a severe error floor. By adopting two linear ISI compensators, we present a feedback-based memory-2 decoder which can achieve almost the same performance as in the synchronous case.

Future work will focus on the study of AF protocol over a nonlinear satellite channel, and it mainly consists of two parts. First, we plan to consider more modulation methods including QPSK, 8PSK, and 32APSK as stated in DVB-S2. QPSK and 8PSK both have constant envelopes, while 32APSK has three rings of constellation points. Each of the three modulation schemes corresponds to a different μ-table size for the same memory length. We believe that there will be more tradeoff issues between system complexity and decoding performance. Second, we want to work on a nonlinear fading channel. Satellite communication suffers from a slow channel fading, and we hope to work out some new methods that are able to track the channel fading under a nonlinear situation.

Bibliography

[1] G. Maral, M. Bousquet, and Z. Sun. *Satellite Communications Systems: Systems, Techniques, and Technology.* John Wiley and Sons, Inc, 2009.

[2] C. E. Shannon. Two-way communications channels. In *4th Berkeley Symp. Math. Stat. Prob., Chicago, IL*, 1961.

[3] A. Avestimehr, A. Sezgin, and D. Tse. Capacity of the two-way relay channel within a constant gap. *European Transactions on Telecommunications*, 21(4):363–374, April 2010.

[4] D. Gunduz, E. Tuncel, and J. Nayak. Rate regions for the separated two-way relay channel. In *Annual Allerton Conference on Communication Control and Computing*, Sep. 2008.

[5] Y. Hao, D. Goeckel, Z. Ding, D. Towsley, and K. Leung. Achievable rates for network coding on the exchange channel. In *Proc. IEEE Military Communications Conference, MILCOM*, 2007.

[6] T. M. Cover. Broadcast channels. In *IEEE Trans. Inform. Theory*, pages 2–14, 1972.

[7] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley and Sons, Inc, 2006.

[8] S. Avestimehr, A. Sezgin, and D. Tse. Capacity of the two way relay channel within a constant gap. In *European Transactions on Telecommunications*, Apr. 2010.

[9] R. Koetter and M. Medard. An algebraic approach to network coding. In *IEEE/ACM Transactions on Networking*, pages 782–795, Oct. 2003.

[10] S. Y. Li, R. Yeung, and N. Cai. Linear network coding. In *IEEE Transactions on Information Theory*, pages 371–381, Feb. 2003.

[11] T. Ho, M. Medard, J. Shi, M. Effros, and D. Karger. On randomized network coding. In *41st Annual Allerton Conference on Communication Control and Computing*, Oct. 2003.

[12] Y. Wu, P. Chou, and S. Kung. Information exchange in wireless networks with network coding and physical-layer broadcast. In *Conference on Information Sciences and Systems*, Mar. 2005.

[13] Y. Sagduyu and A. Ephremides. Joint scheduling and wireless network coding. In *Netcod*, 2005.

[14] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft. XORs in the air: Practical wireless network coding. In *Proc. ACM SIGCOMM*, 2006.

[15] S. J. Kim, P. Mitran, and V. Tarokh. Performance bounds for bidirectional coded cooperation protocols. *IEEE Transactions on Information Theory*, 54:5235–5241, Nov 2008.

[16] L. Xiao, T. E. Fuja, J. Kliewer, and D. J. Costello. Nested codes with multiple interpretations. In *IEEE 40th Annual Conference on Information Sciences and Systems*, pages 851–856, 2006.

[17] G. Kramer and S. Shamai. Capacity for classes of broadcast channels with receiver side information. In *IEEE Inf. Theory Workshop*, Sep. 2007.

[18] Y. Wu. Broadcasting when receivers know some messages a priori. In *IEEE ISIT*, pages 1141–1145, 2007.

[19] B. Rankov and A. Wittneben. Spectral efficient signaling for half-duplex relay channels. In *Asilomar Conf. Signals, Systems, and Computing, Pacific Grove, CA*, 2005.

[20] S. Katti, S. Gollakota, and D. Katabi. Embracing wireless interference: Analog network coding. In *Proc. ACM SIGCOMM*, 2007.

[21] R. Vaze and R. W. Heath. Optimal amplify and forward strategy for two-way relay channel with multiple relays. In *IEEE Information Theory Workshop, ITW*, 2009.

[22] M. Osato, H. Kobashi, R. Y. Omaki, T. Hara, and M. Okada. Development of signal canceller in the carrier super-positioning satellite networks. In *IEEE ICWMC*, Mar. 2007.

[23] K. Lu, T. Zhang, Y. Qian, and H. Chen. Cooperative wireless networks based on physical layer network coding. In *IEEE Wireless Communications*, volume 17, pages 86–95, Dec. 2010.

[24] S. Zhang, S. C. Liew, and P. P. Lam. Physical-layer network coding. In *ACM MOBICOM*, 2006.

[25] P. Popovski and H. Yomo. Bi-directional amplification of throughput in a wireless multi-hop network. In *IEEE VTC*, May. 2006.

[26] P. Popovski and H. Yomo. Physical network coding in two-way wireless relay channels. In *IEEE ICC*, June. 2007.

[27] M. P. Wilson, K. R. Narayanan, H. D. Pfister, and A. Sprintson. Joint physical layer coding and network coding for bi-directional relaying. In *IEEE Transactions on Information Theory*, Nov. 2007.

[28] W. Nam, S. Chung, and Y. H. Lee. Capacity of the Gaussian two-way relay channel to within 1/2 bit. In *IEEE Transactions on Information Theory*, volume 56, pages 5488–5494, Nov 2010.

[29] W. Guan and K. J. R. Liu. Two-way denoise-and-forward relaying with non-coherent differential modulation. In *IEEE Global Telecommunications Conference*, pages 1–5, Dec. 2011.

[30] A. A. M. Saleh. Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers. In *IEEE Transactions on Communications*, volume 29, pages 1715–1720, Nov. 1981.

[31] M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems.* John Wiley and Sons, Inc, 1980.

[32] S. A. Billings. Identification of nonlinear systems-a survey. In *Control Theory and Applications, IEE Proceedings D*, Nov. 1980.

[33] S. Benedetto and E. Biglieri. *Principles of Digital Transmission with Wireless Applications.* Kluwer Academic / Plenum Publishers, 1999.

[34] C. Eun and E. J. Powers. A new Volterra predistorter based on the indirect learning architecture. In *IEEE Trans. on Signal Processing*, Jan. 1997.

[35] L. Ding, R. Raich, and G. T. Zhou. A Hammerstein predistortion linearization design based on the indirect learning architecture. In *IEEE ICASSP*, Aug. 2002.

[36] L. Ding, G. Zhou, D. R. Morgan, Z. Ma, J. S. Kennedy, J. Kim, and C. R. Giardina. A robust digital baseband predistorter constructed using memory polynomials. In *IEEE Trans. on Communications*, Jan. 1980.

[37] J. Harmon and S.G. Wilson. Iterative approach to the indirect learning architecture for baseband digital predistortion. In *IEEE Globecom*, Dec. 2010.

[38] J. Harmon and S.G. Wilson. Complex nonlinear adaptive predistortion. In *Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2012.

[39] J. G. Smith. On the information capacity of peak and average power constrained Gaussian channels. In *Ph.D. dissertation, Dep. Elec. Eng, Univ. of California, Berkeley*, Dec. 1969.

[40] I. Abou-Faycal and J. Fahs. On the capacity of some deterministic non-linear channels subject to additive white Gaussian noise. In *IEEE International Conference on Telecommunications*, pages 63–70, 2010.

[41] S. Shamai and I. Bar-David. Upper bounds on capacity for a constrained Gaussian channel. In *IEEE Transactions on Information Theory*, volume 35, pages 1079–1084, Sep. 1989.

[42] S. Shamai and I. Bar-David. The capacity of average and peak-power-limited quadrature Gaussian channels. In *IEEE Transactions on Information Theory*, volume 41, pages 1060–1071, July. 1995.

[43] T. Uratani, R. Miyamoto, T. Hara, and M. Okada. Performance of an interference canceller for p-mp satellite networks with nonlinear TWTA. In *IEEE International Symposium on Communications, Control and Signal Processing*, Mar. 2008.

[44] T. Hara, K. Kubo H. Matsuda, and M. Okada. Performance improvement of interference canceller for carrier super-positioning by the nonlinearity compensation in satellite communications. In *IEEE International Conference on Wireless and Mobile Communications*, Sep. 2010.

[45] T. Hara, K. Kubo, and M. Okada. Adaptive back-off nonlinearity compensation for interference canceller in carrier super-positioning satellite communications. In *IEICE Transactions on Communications*, July. 2011.

[46] H. Matsuda, T. Ishiguro, T. Hara, and M. Okadaa. Nonlinearity compensation for super-positioning satellite system with interference canceller. In *Advanced International Conference on Telecommunications (AICT)*, Mar. 2011.

[47] T. Ishiguro, T. Hara, and M. Okada. Post-compensation technique for carrier superposed satellite channel including nonlinear TWTA. In *IEICE Transactions on Communications*, Nov. 2012.

[48] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate. In *IEEE Trans. on Information Theory*, volume 20, March 1974.

[49] S. J. Kim, N. Devroye, P. Mitran, and V. Tarokh. Comparison of bi-directional relaying protocols. In *IEEE Sarnoff Symposium*, pages 1–5, 2008.

[50] T. Cover, R. J. McEliece, and E. Posner. Asynchronous multiple-access channel capacity. In *IEEE Transactions on Information Theory*, pages 409–413, July 1981.

[51] B. Yang, K. Letaief, and R. Cheng. Timing recovery for OFDM transmission. In *IEEE J. Select. Areas Commun.*, volume 18, Nov 2000.

[52] H. Bolcske. Blind estimation of symbol timing and carrier frequency offset in wireless OFDM systems. In *IEEE Trans. on Commun.*, volume 49, June 2001.

[53] H. Minn, V. K. Bhargava, and K. B. Letaief. A robust timing and frequency synchronization for OFDM systems. In *IEEE Trans. on Wireless Commun.*, volume 2, July 2003.

[54] B. Park, H. Cheon, E. Ko, C. Kang, and D. Hong. A blind OFDM synchronization algorithm based on cyclic correlation. In *IEEE Signal Processing Lett*, volume 11, Feb 2004.

[55] Optimizing satellite communications using DoubleTalk Carrier-in-Carrier and CDM-625 advanced satellite modem. http://www.comtechefdata.com/products/modems/pcdm-625.asp.

[56] A. J. Viberbi. Bounds on the information rate of intersymbol interference channels based on mismatched receivers. In *IEEE Trans. on Information Theory*, volume 13, April. 1967.

[57] Second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broad-band satellite applications (DVB-S2). In *European Norm EN 302 307*.

[58] D. M. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng. Simulation-based computation of information rates for channels with memory. In *IEEE Transactions on Information Theory*, volume 52, pages 3498–3508, Aug. 2006.

[59] E. Casini, R. De Gaudenzi, and A. Ginesi. DVB-S2 modem algorithms design and performance over typical satellite channels. In *International Journal of Satellite Communications and Networking*, Jan. 2004.

[60] A. Morello and V. Mignore. DVB-S2: The second generation standard for satellite broadband services. In *Proceedings of the IEEE*, Jan. 2006.

[61] N. H. Ngo, S. A. Barbulescu, and S. S. Pietrobon. Performance of nonuniform M-ary QAM constellation on nonlinear channels. In *Australian Commun. Theory Workshop*, Feb. 2005.

[62] G. Karam and H. Sari. Analysis of predistortion, equalization and ISI cancellation techniques in digital radio systems with nonlinear transmit amplifiers. In *IEEE Transactions on Communications*, Dec. 1989.

[63] P. M. Olmos, J. J. Murillo-Fuentes, and F. Prez-Cruz. Joint nonlinear channel equalization and soft LDPC decoding with Gaussian processes. In *IEEE Transactions on Signal Processing*, Mar. 2010.

[64] T. K. Moon. *Error Correction Coding: Mathematical Methods and Algorithms*. John Wiley and Sons, Inc, 2005.

Publication List

1. **C. Xu** and S. G. Wilson. Minimizing transponder cost for two-way data exchange in satellite networking. *IEEE International Conference on Communications (ICC)*, pp.3250-3254, June 2012.

2. **C. Xu**, J. Peng, S. G. Wilson, and T. Berger. Achievable information rates for nonlinear satellite channels in unidirectional and bidirectional relaying. *IEEE LATINCOM*, Nov 2012.

3. **C. Xu** and S. G. Wilson. Achievable rate region for a hub-remote three-terminal satellite network. *IEEE International Conference on Communications (ICC)*, June 2013.

4. **C. Xu**, J. Peng, S. G. Wilson, and T. Xie. Improving decoding performance of DVB-S2 transmission on a nonlinear channel. *Accepted by 31st AIAA International Communications Satellite Systems Conference (ICSSC)*, Oct 2013.

5. J. Peng, **C. Xu**, S. G. Wilson, and T. Xie. Two-way (same-frequency) relaying: information rates and DVB-S2 performance. *Accepted by 31st AIAA International Communications Satellite Systems Conference (ICSSC)*, Oct 2013.