Design of Nonlinear Precoding and Estimation Schemes for Advanced Communication Systems

Universidad de Vigo

Departamento de Teoría de la Señal y Comunicaciones

Author
Marcos Álvarez Díaz

Advisors
Carlos Mosquera Nartallo
Roberto López Valcarce

2010

DOCTOR EUROPEUS

Dpto. de Teoría de la Señal y Comunicaciones
ETSE de Telecomunicación
Universidade de Vigo
Campus Universitario s/n
E-36310 Vigo

DOCTORAL THESIS

DOCTOR EUROPEUS

Design of Nonlinear Precoding and Estimation Schemes for Advanced Communication Systems

Author: **Marcos Álvarez Díaz**
Telecommunication Engineer

Advisors: **Carlos Mosquera Nartallo**
PhD in Telecommunication
Roberto López Valcarce
PhD in Telecommunication

April 2010

TESIS DOCTORAL

DOCTOR EUROPEUS

Design of Nonlinear Precoding and Estimation Schemes for Advanced Communication Systems

Autor: D. **Marcos Álvarez Díaz**

Directores: Dr. D. **Carlos Mosquera Nartallo**
Dr. D. **Roberto López Valcarce**

TRIBUNAL CALIFICADOR

Presidente: _____

Vocales: _____

_____

_____

Secretario: _____

CALIFICACIÓN: _____

Vigo, a de de 2010.

Acronyms

ACM Adaptive Coding and Modulation

APSK Amplitude and Phase Shift Keying

AWGN Additive White Gaussian Noise

BER Bit Error Ratio

CCDF Complementary Cumulative Density Function

CCI Co-Channel Interference

CER Constellation Expansion Ratio

CM Constant Modulus

CRB Cramér-Rao Lower Bound

CSI Channel State Information

DA Data-aided

DBF Digital Beamforming

DC/RF Direct Current to Radio Frequency

DD Decision-directed

DFE Decision Feedback Equalizer or Equalization

DFE Decision Feedback Equalizer

DPC Dirty Paper Coding

DRA Directly Radiating Array

DRAF Dual Reflector Array-Fed

EVB Envelope-based

FFT Fast Fourier Transform

FIM Fisher information matrix

GBBF Ground-Based Beamforming

GEO Geostationary Earth Orbit

HPA High Power Amplifier

HSGBF Hybrid Space-Ground Beamforming

IBO Input Back-Off

IF intermediate frequency

i.i.d. independent and identically distributed

IMR Intermediate Module Repeater

ISI Intersymbol Interference

LLF Log-likelihood Function

LP Linear Precoding

LDPC Low-Density Parity-Check

LN Linear-Nonlinear

LNL Linear-Nonlinear-Linear

LOS Line-of-Sight

LP Linear Precoding

LUT Look-Up Table

MIMO Multiple Input Multiple Output

MISO Multiple Input Single Output

ML Maximum Likelihood

M&M Mueller and Müller

MMSE Minimum Mean Square Error

MSE Mean Square Error

OBBF Onboard Beamforming

OBO Output Back-Off

OBP On-board Processing

O&M Oerder and Meyr

NDA Non-data-aided

NL Nonlinear-Linear

PAM Pulse Amplitude Modulation

PAR Peak-to-Average Power Ratio

pdf Probability Density Function

PLL Phase-Locked Loop

PSD Power Spectral Density

PSK Phase Shift Keying

QAM Quadrature Amplitude Modulation

QPSK Quadrature Phase Shift Keying

QoS Quality of Service

RC Raised Cosine

RF radio frequency

SER Symbol Error Rate

SINR Signal-to-Interference plus Noise Ratio

SISO Single Input Single Output

SNR Signal-to-Noise Ratio

SRRC Square Root Raised Cosine

SRF Single Reflector Feed

SRHT Single Reflector with Hybrid Transform

SSPA Solid State Power Amplifier

TDD Time Division Duplexing

TH Tomlinson-Harashima

THP Tomlinson-Harashima Precoding

TWTA Traveling Wave Tube Amplifier

WMF Whitened Matched Filter

ZF Zero-Forcing

Contents

List of Tables

List of Figures

Acknowledgments

The work contained in this thesis has been partially supported by the Spanish Goverment, Ministerio de Ciencia e Innovación, through the FPI grant BES-2005-8964, and through the projects TEC2004-02551/TCM (DIPSTICK), TEC2007-68094-C02-01/TCM (SPROACTIVE), and CONSOLIDER-INGENIO 2010 CSD2008-00010 (COMONSENS). Parts of this thesis have also been supported by the European Commission through the IST Programme under Contracts IST-507052 (SatNEx) and IST-027393 (SatNEx-II).

Chapter 1

Introduction

Contents

1.1 Outline of the Thesis and Contributions

Since the beginning of modern telecommunications there has been a constant evolution towards transmission systems with growing capacity and efficiency. One example is provided by the evolution of wireline communications. Consider the transmission speed of about 40 words/minute of wireline telegraph operators using Morse code back in 1840, and compare it with the 100 Gbits/s planned in the forthcoming IEEE 802.3ba standard for cable and fiber [92]. Such a dramatic speed increase has been possible thanks to the development of new transmission and signal processing technologies and techniques, and to the improvement of existing ones. The history of wireless communications is not different at all, but if we consider the particular case of satellite communications, the rapidness of the evolution is even more striking. Within a 50 year span, starting from the launches of the first artificial satellite in 1957 (Sputnik-I) and the first satellite specifically devoted to communications in 1962 (TelStar 1), the transmission capabilities of the satellite communication payloads have experienced a spectacular growth, to the extent that there are currently planned geostationary satellites which will offer an aggregate throughput of 100 Gbps [88]. In parallel, different satellite constellations have been deployed, e.g., at low orbit (Globastar, Iridium), highly elliptical orbit (Sirius), and geostationary orbit (Thuraya, Inmarsat BGAN) to provide a large range of voice, data or broadcasting services with continuously increasing bandwidths and data rates.

Key to the rapid and successful evolution of modern telecommunications has been the continued effort to provide the different communication systems with efficient physical layers. The ever growing demand for speed and reliability in the communication from

the end user's perspective has pushed the design of efficient physical layers featuring cutting edge technology and signal processing techniques. Because many different tasks are performed within the end-to-end communication chain of current systems, many are the blocks and operations that can be optimized towards an increased efficiency of the physical layer. Some examples are provided by the Adaptive Coding and Modulation (ACM) techniques used within the DVB-S2 standard [54], the Turbo and Low-Density Parity-Check (LDPC) codes currently implemented by different systems as error correction schemes, and the use of Multiple Input Multiple Output (MIMO) techniques available in the WiMAX protocol which, based on the IEEE 802.16 standard [91], allows simultaneous data transmission and reception over multiple antennas.

As a result of the application of the latest advances in physical layer techniques, novel communication systems are able to offer high data rates under high Signal-to-Noise Ratio (SNR) conditions, or lower data rates at very low SNRs. Oftentimes, the introduction of a novel, better performing technology or technique at some point within the communication chain shifts the critical point to other communication blocks which were originally designed to work under less challenging conditions. For example, it is well known that the parameter estimation and synchronization tasks performed at the receiver, in general, turn more difficult for increasing data rates and diminishing SNR operation points. Therefore, the move of current systems towards such working conditions requires the development of more capable estimation and synchronization techniques, or the optimization of the existing ones.

Satellite communications and, in particular, their physical layer issues have been the focus of much of the author's research work over the last years, and, as such, constitute the focus of this thesis to a great extent. Two of the chapters (2 and 3) are devoted to explore the novel application of Tomlinson-Harashima Precoding (THP) to satellite transmission systems under different settings. THP can be roughly described as a nonlinear preequalization technique which allows the efficient compensation of interference (not only of time-dispersive nature) already at transmitter side. THP is currently used in the 10 Gigabit Ethernet (10GBASE-T) standard [90], under quite different application conditions to those found in satellite systems. This motivates the introduction and study of THP in the satellite context.

Satellite communications have also provided the original motivation of Chapter 4. In this case, the focus is on the receiver, in particular, the SNR estimation task. The adoption by the DVB-S2 standard [54] of multilevel Amplitude and Phase Shift Keying (APSK) modulations such as 16-APSK and 32-APSK to accommodate higher data transmission rates, among other consequences, has posed difficulties to traditional SNR estimators which were originally designed for constant modulus modulations. Chapter 4 focuses on the SNR estimation from multilevel modulations. Since many terrestrial systems use multilevel modulations as well, the range of applications of the introduced estimators extends to systems other than satellite-based ones. Finally, parameter estimation from multilevel modulations is also the topic on which Chapter 5 focuses. In particular, this chapter deals with the problem of carrier phase recovery from a particular subset of Quadrature Amplitude Modulation (QAM) family known as cross-QAMs, whose application in restricted to terrestrial systems

The main contributions of this thesis are summarized next.

- **Chapter 2.** We consider the forward link of a unicast (point-to-point) satellite transmission system in which the geostationary satellite coverage in urban areas is complemented by the use of terrestrial gap-fillers reusing the same band as the satellite. Under such scenario, it is likely that the receiving terminals will experience multipath reception, and thus the desired signal will be corrupted by Intersymbol Interference (ISI). We propose and study the application of THP in such scenario as a means to precompensate the ISI already at the transmission stage (the satellite, or preferably, the gateway). The application of THP to satellite communications is novel, and must face some challenges not encountered in other scenarios. In particular, it turns out that THP signals are more affected by the nonlinear characteristics of the High Power Amplifier (HPA) onboard the satellite than traditional modulations. We focus on counteracting the nonlinear effect of the HPA onboard the satellite using signal predistortion. Since the performance of THP and predistortion is insufficient, we propose the use of shaping without scrambling (a combination of THP and signal shaping) to modify the dynamics of the precoded signal so as to make it more robust against nonlinear distortion. The content of this chapter stems from the work presented in [10, 11, 138].
- **Chapter 3.** The forward link of a multibeam satellite system is considered, in which a geostationary satellite delivers independent data streams to each beam. There is a single receiver per beam. In oder to avoid Co-Channel Interference (CCI) among near beams, multibeam systems typically implement some frequency reuse scheme which ensures that the same carrier frequency is never used by two adjacent beams. We propose the use of THP as a way to precompensate the CCI within a set of possibly contiguous beams without requiring any frequency reuse. We will evaluate several alternative THP implementations in a number of different application scenarios with different CCI levels. The results will show that the proposed system would yield considerable throughput gains with respect to current systems. This chapter is derived from the work presented in [7].
- **Chapter 4.** Two main contributions are contained in this chapter. First, we develop a novel family of blind Signal-to-Noise Ratio (SNR) estimators based on the moments of the envelope of the baud-rate samples. This family extends and generalizes several previous estimators. As a particular case of the new family, we analyze and optimize the performance of an estimator based on moments up to the eighth order, which turns out to outperform previous estimators of its kind for most practical modulations. As a second contribution, we study the problem of estimating the SNR from a data frame composed of symbols of different modulations. Several estimator combining strategies are proposed, and their performances are evaluated. References [6] and [9] constitute the basis of this chapter.
- **Chapter 5.** We introduce a new blind criterion for carrier phase recovery of QAM modulated data. The criterion maximizes a function of the ℓ_1-norm of the phase-compensated received data vector. Two iterative estimators with low complexity are derived from this criterion, and their performance is analyzed showing that they are competitive against higher-order estimators. As an additional advantage, the novel estimators exhibit robustness against quantization effects and finite implementation.

This chapter is based on the work presented in [5, 8]

1.2 Notation

Throughout this work, italic font will be used to denote scalars (x), whereas bold italic font will be used for vectors ($\boldsymbol{x}$, lower case) and matrices ($\boldsymbol{X}$, upper case). The elements of the vector $\boldsymbol{x}$ will be denoted either as $x_1, x_2, \ldots$ or $[\boldsymbol{x}]_1, [\boldsymbol{x}]_2, \ldots$, and similarly, the elements of the matrix $\boldsymbol{X}$ will be denoted either as $x_{11}, x_{12}, \ldots$ or $[\boldsymbol{X}]_{11}, [\boldsymbol{X}]_{12}, \ldots$ The estimated value of a signal or parameter y will be denoted as $\hat{y}$. For a complex-valued variable z, $\mathrm{Re}\{z\}$ and $\mathrm{Im}\{z\}$ will be used to represent its real and imaginary parts, respectively ($z = \mathrm{Re}\{z\} + j\mathrm{Im}\{z\}$), whereas $\arg\{z\}$ will denote the phase ($z = |z|\mathrm{e}^{j\arg\{z\}}$). For convenience we may also use the shorter notation $\bar{z}$, $\tilde{z}$ to denote the real and imaginary parts of z ($z = \bar{z} + j\tilde{z}$). The *complex sign* of z is defined as $\mathrm{csgn}(z) \doteq \mathrm{sgn}(\bar{z}) + j\mathrm{sgn}(\tilde{z})$, where for real x, $\mathrm{sgn}(x) = x/|x|$ if $x \neq 0$ and zero for $x = 0$.

Chapter 2

Application of Tomlinson-Harashima Precoding to Point-to-Point Satellite Communications

Contents

2.1 Introduction

The increasing demand of fixed multimedia applications is pushing the design of more and more efficient point-to-point satellite systems. In addition, the interest in providing satellite-based services to mobile or nomadic users is continuously growing [137]. In

order to accommodate such demands, system designers are being forced to improve the performance and efficiency of traditional systems. Different complementary approaches have been followed in the recent years to achieve this goal. One of them is the implementation of efficient fading mitigation techniques. Among such techniques we can mention the use of time or antenna diversity [41,141,160], power control, adaptive waveforms [32], or layer 2 methods coping with the temporal dynamics of the fade. Of special interest is the use of more efficient modulation and coding schemes, such as the ACM techniques included in the DVB-S2 standard [54]. Besides these techniques, larger throughputs can be obtained by operating in higher frequency bands (e.g. the Ku and Ka bands currently in use, or the foreseen Q and V bands [32,151]), and by deploying a larger number of beams within the coverage region of the satellite (e.g., about 200 beams in Inmarsat 4 [78], or over 500 beams in the Terrestar-1 system [156]).

In recent years, satellite overlay networks have shown their potential for delivering multimedia content both to rural and urban environment. In the latter case, ground-based gap-fillers, also known as Intermediate Module Repeaters (IMRs), have been installed or considered for systems in which direct satellite coverage cannot be integrally achieved due to unavoidable obstacles such as tunnels, buildings or hills. Examples of this approach can be found in the context of the SATIN and MAESTRO EC-funded projects [26,100] and in the forthcoming TerreStar service [155], where the use of IMRs was proposed to complement the provision of multimedia content by a satellite, and in the DVB-SH standard [55], which considers a hybrid satellite and terrestrial network to provide multimedia content to mobile and fixed terminals. Other examples of the use of gap-fillers in satellite systems can be found in [12,48,84,105,109,130]. The use of IMRs increases signal diversity, but may also introduce considerable dispersion, which invalidates the classic flat-fading assumption at the receiver. Indeed, the terminal may simultaneously receive the signals from several such gap-fillers, as well as the satellite, or even reflected or scattered versions of those signals as well. Under such conditions, the receiving terminal will experience multipath reception, which will translate into the existence of Intersymbol Interference (ISI) in the received signal. In order to successfully decode the received signal, some kind of compensation technique must be applied at the receiver, such as diversity combining or equalization.

In this chapter we propose the application of a nonlinear precoding technique called Tomlinson-Harashima Precoding (THP) to point-to-point satellite communications. This technique is related to the Dirty Paper Coding (DPC) principle originally introduced by Costa in [42], by virtue of which efficient communication can be achieved if the transmitter has knowledge of the interference seen by the receiver. In fact, in [42] it was proven that the capacity of a Gaussian channel affected by an additional Gaussian interference known to the transmitter is the same as the capacity of the channel without the additional interference. In order to achieve such capacity, the transmitter must not precancel the known interference, which would be inefficient, but try to adapt the transmitted signal to the known interference, so that the available transmit power is efficiently used. Although the original proposal of THP in [80, 81, 157] was done long before the introduction of the DPC principle, THP can be regarded as a suboptimal implementation of the DPC principle, as we will see in this chapter.

THP has the advantage that the processing requirements at the receiver side are simplified since the precoding operation is performed at the transmitter side. In particular, THP can be used to precompensate (i) the intersymbol interference introduced by time-dispersive channels or (ii) the co-channel interference existing in multiantenna or multibeam systems. The former application will be covered in this chapter, whereas the latter will be dealt with in Chapter 3.

The use of nonlinear precoding techniques for time-dispersive channels has been so far restricted to communication systems in which the amount of nonlinear distortion introduced by elements such as signal amplifiers is small (for instance 10GBASE-T technology [90]). This restriction is due, among other reasons, to the amplitude distribution of the precoded signal, which is not well suited for systems in which nonlinear distortion cannot be disregarded. Satellite communications systems belong to that class, since the HPAs onboard are typically pushed to operate close to saturation for cost and efficiency reasons. For this reason, the application of nonlinear precoding to other fields such as satellite communications has not been explored yet.

Chapter outline The chapter is structured as follows. In Section 2.2 we review the basic theory related to THP applied to Single Input Single Output (SISO) time dispersive channels, signal shaping, and signal predistortion. The satellite point-to-point scenario considered is described in Section 2.3. In Section 2.4 we propose the use of THP for equalizing point-to-point satellite transmissions over time dispersive SISO channels. As a means to reduce the impact of the nonlinear power amplifier onboard, we propose to combine THP with signal shaping and signal predistortion. The performance of the proposed systems is evaluated in Section 2.5. This chapter stems from the work presented in [10, 11, 138].

2.2 Theoretical Background

2.2.1 THP for Time-Dispersive Channels

2.2.1.1 Equalization and Precoding

Time-dispersive channels are commonly found in many communications systems. They are characterized by the fact that the desired signal, say the transmitted symbol at instant k, is not only corrupted with noise, but also with contributions of some past (or future) transmitted symbols: the Intersymbol Interference (ISI). In order to compensate for the effect of the channel, *equalization* has been typically employed at the receiver. The equalizer has the mission of removing the contributions of all symbols but the desired one at each instant.

Consider a discrete-time model of a time-dispersive channel, with transfer function $H(z) \doteq 1 + \sum_{n=1}^{N} h_n z^{-n}$. We assume that the impulse response is causal (there is no ISI coming from future samples) with $h_0 = 1$ (monic polynomial). As a first attempt,

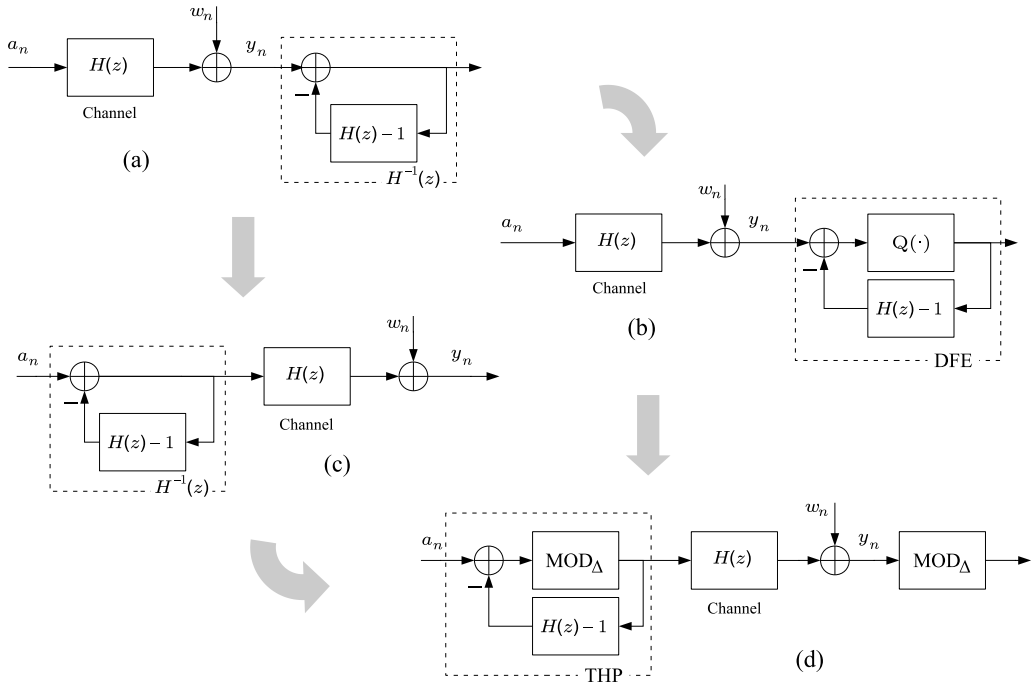


Figure 2.1: From linear equalization (a) to nonlinear precoding (d) via DFE (b) or linear preequalization (c).

and considering that the receiver has knowledge (e.g. through estimation) of $H(z)$, one could simply apply the inverse of the channel transfer function $1/H(z)$, as depicted in Figure 2.1a (the inverse function is realized using a negative feedback structure in which $H(z) - 1$ appears in the feedback branch). This yields a Zero-Forcing (ZF) linear equalizer, after which the decision device immediately follows. However, it is well known that such a scheme suffers from *noise enhancement*: the equalization step increases the noise power.

One way to avoid noise enhancement is to use a Decision Feedback Equalizer (DFE). As shown in Figure 2.1b, the decision device is placed in the feedforward branch of the channel inverter. When the decided symbols are correct, the DFE successfully cancels the ISI while avoiding noise enhancement (the signal that is fed back is noise-free). If there are wrong decisions instead, the ISI is not correctly removed, thus causing *error propagation*, which is one of the main drawbacks of the DFE. Another main disadvantage of the DFE is that coded modulation is not straightforwardly applicable: the DFE requires zero-delay decisions to work, which cannot be achieved with current coding schemes.[1]

Both problems can be avoided if the equalization procedure is moved to the transmitter side, as shown in Figure 2.1c. This is known as *preequalization*. Of course, a key requirement is that the channel transfer function be known at the transmitter. This can be accomplished by different means. If the transmitter entity acts also as receiver, and the channel in both directions can be considered equal, a local estimation of the transfer

[1] Admittedly, there are ways to overcome these problems, such as using interleaving and iterative processing [56], but they exhibit much higher complexity.

function would be possible. This is the case of Time Division Duplexing (TDD) systems with asymmetric processing (e.g. [16, 77]), where the channel impulse response can be estimated during the reception phase and then used to preequalize the signal during the transmission phase. Otherwise, a setup procedure at the beginning of the transmission can be used, through which the channel information is first estimated at the receiver and then passed back to the transmitter. Linear preequalization, however, shares the same noise enhancement problem as its receiver side counterpart. In general, the preequalizer increases the power of the transmitted signal. But the transmit power limitation which is inherent to every transmitter requires the transmitted signal to be scaled down enough to accommodate its highest peak value: this translates into an equivalent noise enhancement effect.

In this context, nonlinear *precoding* appears as an equalization scheme which maintains the advantages of the former schemes while avoiding their drawbacks. Several nonlinear precoding schemes exist such as THP and *flexible precoding* [59]. In general all precoding schemes share the same principle: they perform the equalization step at the transmitter while avoiding the noise enhancement problem by means of some nonlinear processing, as shown in Figure 2.1d for THP. The additional advantage of precoding with respect to receiver-based equalization is that the processing burden is moved to the transmitter side. This allows the deployment of simpler, and thus cheaper terminals. In this chapter THP will be the only nonlinear precoding scheme considered.

Duality between equalization and precoding The parallelism between receiver-based equalization and transmitter-based precoding can be seen as a sort of *duality* sharing the same principles. In this way, linear preequalization would be the dual of linear equalization, and THP would be the dual of DFE (both are nonlinear techniques).

2.2.1.2 Principles of Tomlinson-Harashima Precoding

THP was independently introduced by Harashima and Miyakawa [80, 81] and Tomlinson [157] between the late 1960s and early 1970s. A detailed analysis has been more recently provided by Fischer in [59, chapter 3].

The basic setup of a Tomlinson-Harashima (TH) precoder is shown in Figure 2.2. Here, the transfer function $H(z)$ is considered to be *causal* and *minimum-phase*, with $h_0 = 1$ (monic polynomial). Again, the channel inversion is implemented via a negative feedback structure, but now a modulo device is inserted into the feedforward branch, which introduces a nonlinearity in the precoding operation. In order to recover the (noisy) original symbols, the receiver must apply again the same modulo operation performed at the receiver. The fact that THP operates at the transmitter automatically prevents error propagation. As additional advantage, THP can be straightforwardly used in conjunction with current coding schemes (such as LDPC codes [90]). Similarly to linear preequalization, the channel transfer function needs to be known at the transmitter.

The operational principle of THP lies upon lattice theory: the use of modulo arithmetic allows us to represent one symbol not only by that lying in the original constellation set,

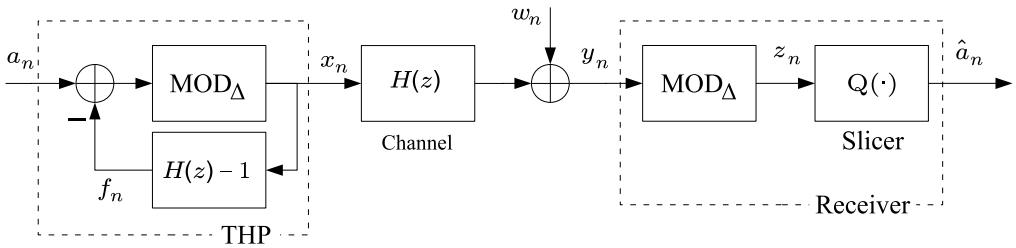


Figure 2.2: System using Zero-Forcing Tomlinson-Harashima precoding

but also by its correspondents in other (modulo-congruent) constellation replicas, as long as these replicas are distributed according to a well-constructed lattice. To illustrate how THP works, consider the signals a_n, f_n (the feedback signal corresponding to the ISI), y_n and z_n from Figure 2.2. We have

$$f_n = \sum_{k=1}^{p-1} h_k \, x_{n-k} \tag{2.1}$$

$$x_n = \text{MOD}_\Delta \left(a_n - f_n \right) \tag{2.2}$$

$$y_n = \sum_{k=0}^{p-1} h_k \, x_{n-k} + w_n = x_n + f_n + w_n \tag{2.3}$$

$$\begin{aligned} z_n &= \text{MOD}_\Delta \left(y_n \right) = \text{MOD}_\Delta \left(x_n + f_n + w_n \right) \\ &= \text{MOD}_\Delta \left(\text{MOD}_\Delta \left(a_n - f_n \right) + f_n + w_n \right) = \text{MOD}_\Delta \left(a_n + w_n \right). \end{aligned} \tag{2.4}$$

Therefore, in a noiseless case ($w_n = 0$) we have that $z_n = a_n$.

For real arguments, the modulo operation is defined as $\text{MOD}_\Delta \left(a \right) = a + m\Delta$, where m is the integer for which $\text{MOD}_\Delta \left(a \right)$ lies in $[-\frac{\Delta}{2}, \frac{\Delta}{2})$, and the *modulus* Δ is an appropriately chosen amplitude. Figure 2.3 shows the behavior of the real-valued modulo operation for $\Delta = 4$. For complex-valued arguments, as it is the case of square QAMs, the former real-valued modulo operation is separately applied to the real and imaginary parts of the argument. Table 2.1 provides the value of the modulus in terms of the minimum intersymbol distance d_{min} for PAM and square QAM constellations. In (2.4) the following property of the modulo operation was used:

$$\text{MOD}_\Delta \left(\text{MOD}_\Delta \left(a \right) + b \right) = \text{MOD}_\Delta \left(a + b \right). \tag{2.5}$$

An evident effect of the modulo operation at the precoder, see (2.2), is to limit the amplitude of the precoded signal x_n. Figure 2.4 illustrates the geometrical principles of THP transmission (a QPSK constellation was chosen for its simplicity). The gray square in the center, known as the *Voronoi region* in lattice theory, is the set of points where the precoded signal x_n takes values, and coincides with the set of all points that are invariant to the modulo operation, i.e., $\mathcal{R}_{\mathbf{p}} \doteq \{ z \in \mathbb{C}, \ \text{MOD}_\Delta \left(z \right) = z \}$.[2] For square QAMs, the

[2]More generally, given the lattice $\mathbf{\Lambda}$ of points which are modulo-congruent to 0, i.e., $\mathbf{\Lambda} = \{ z \in \mathbb{C}, \ \text{MOD}_\Delta \left(z \right) = 0 \}$, the Voronoi region is defined as the set of all points which are closer to $0 + j0$ than to any other point of the lattice. The small white circles in Figure 2.4 indicate the points of $\mathbf{\Lambda}$.

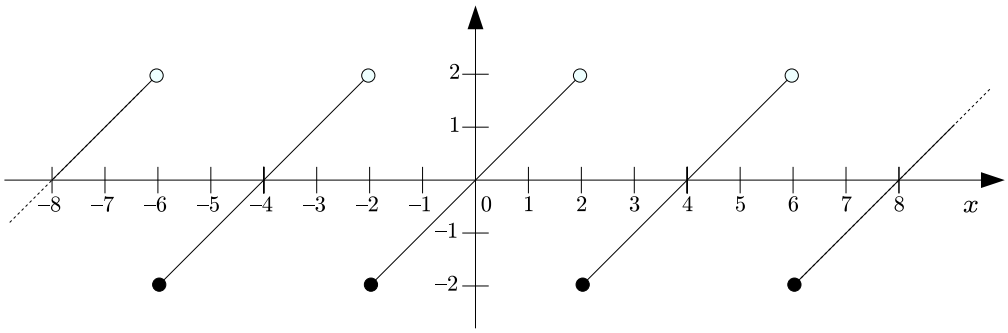


Figure 2.3: $\mathrm{MOD}_{\Delta}(x)$ for real-valued x and $\Delta = 4$.

Table 2.1: Definitions and corresponding modulus Δ for PAM and square QAM constellations in terms of their minimum intersymbol distance d_{min}.

Constellation	Symbols	Modulus
M-PAM, $M = 2q,\ q \in \mathbb{N}$	$\mathcal{A} = \left\{ \pm \frac{2i-1}{2} d_{\text{min}},\ i = 1, 2, \ldots, \frac{M}{2} \right\}$	$\Delta = M\ d_{\text{min}}$
M-QAM, $M = 2^{2q},\ q \in \mathbb{N}$	$\mathcal{A} = \{a_{\text{I}} + j a_{\text{Q}},\ \text{with } a_{\text{I}}, a_{\text{Q}} \in \mathcal{B}\}$, where $\mathcal{B} = \left\{ \pm \frac{2i-1}{2} d_{\text{min}},\ i = 1, 2, \ldots, \frac{\sqrt{M}}{2} \right\}$	$\Delta = \sqrt{M} d_{\text{min}}$

Voronoi region coincides with the half-open square given by $\mathcal{R}_{\mathbf{p}} = \{z_{\text{I}} + j z_{\text{Q}}, -\frac{\Delta}{2} \leq z_{\text{I}} < \frac{\Delta}{2},\ -\frac{\Delta}{2} \leq z_{\text{Q}} < \frac{\Delta}{2}\}$, as shown in Figure 2.4. After transmission of x_n, the channel adds the ISI f_n, and the signal at the receiver is a combination of the precoded symbol, ISI and noise, see (2.3). Disregarding the noise contribution, the noiseless received symbol $x_n + f_n$ will always fall in a point which is congruent to a_n under $\mathrm{MOD}_{\Delta}(\cdot)$ (such points are marked in Figure 2.4 as the darkest points of each QPSK replica). Therefore, as indicated in (2.4), THP allows to receive some modulo-congruent perturbation which will be automatically removed by the modulo operation at the receiver thanks to property (2.5). This allows to transmit the smallest (in magnitude) modulo-congruent difference between the desired symbol and the existing ISI, i.e., x_n.

In the light of the previous explanation it is not difficult to establish a similarity between the way THP works and the DPC principle mentioned in the introductory section of this chapter. The DPC principle states that the most efficient way to counteract an additive interference known by the transmitter is to adapt the transmitted signal to the existing interference, rather than the more "canonical" approach of precanceling the interference. In the case of THP, the "canonical" approach would be represented by linear precoding, which suffers from a power penalty, and therefore is inefficient. Instead, the modulo device of the TH precoder automatically selects the transmitted signal with the minimum amplitude that ensures the correct reception (of a modulo-congruent replica) of the original symbols. Hence, the transmitter takes advantage of the interference to efficiently use the available transmit power, as the DPC principle teaches.

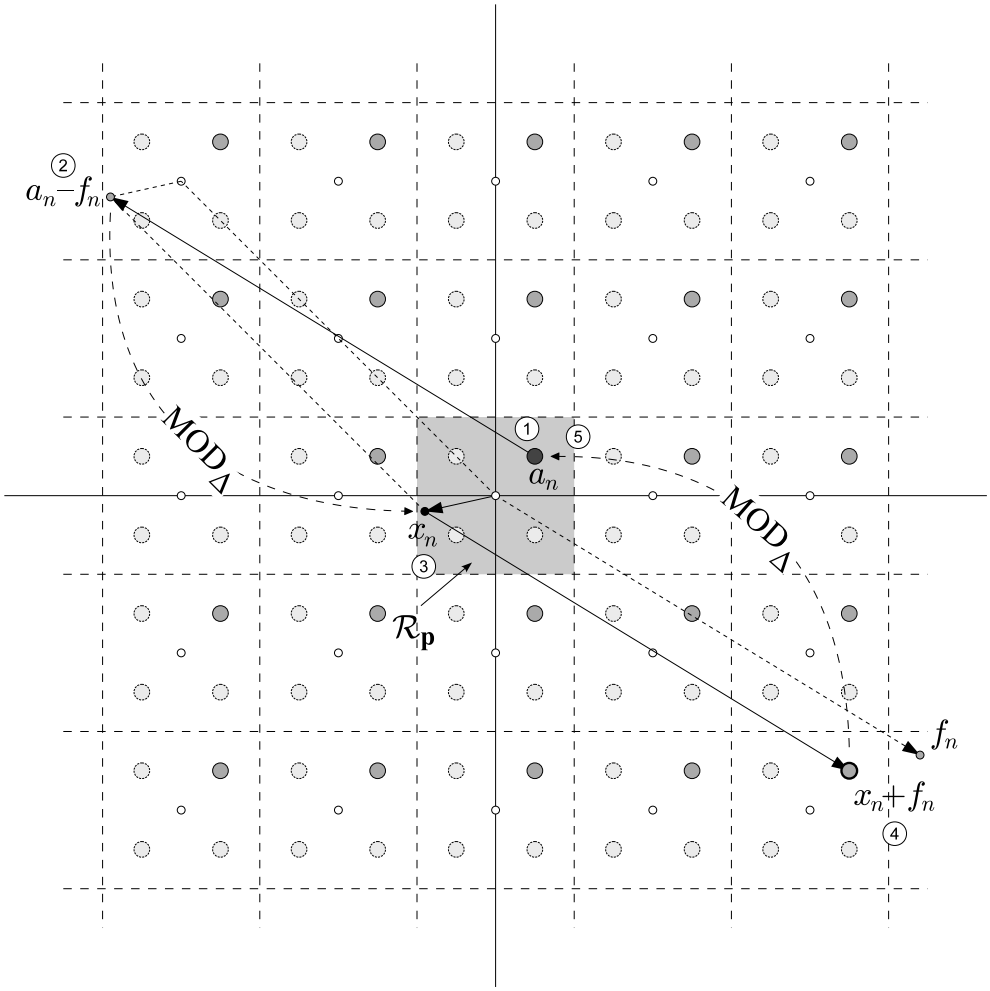


Figure 2.4: Example of transmission of a QPSK symbol using THP.
① The original QPSK symbol a_n is the darkest circle within the Voronoi region $\mathcal{R}_{\mathbf{p}}$ (gray square).
② The transmitter subtracts the known ISI f_n from a_n.
③ Then the result $a_n - f_n$ is modulo-reduced yielding the precoded symbol x_n to be transmitted.
④ At the receiver, the channel has added the ISI, so that (in absence of noise) the received symbol $x_n + f_n$ is a modulo-congruent replica of a_n.
⑤ Finally, the modulo operation at the receiver reduces $x_n + f_n$ to the Voronoi region yielding a_n, as desired.

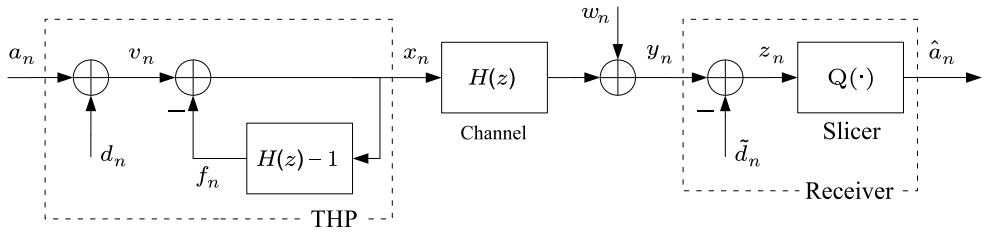


Figure 2.5: Linear equivalent model for THP.

Linearized THP model As already mentioned, the presence of the modulo operation renders the whole transmission model nonlinear, and this prevents the application of the results from the well-established theory of linear systems. However, it is possible to find a linearized equivalent version of the THP procedure, which will be useful later on. As shown in Figure 2.5, some new ad-hoc signals are needed, namely d_n (the *precoding sequence*), v_n (the *effective data sequence*, which is a noise-free version of the received signal y_n), and $\tilde{d}_n$. The following definitions and relationships will be useful:

$$d_n \doteq (a_n - f_n) - \mathrm{MOD}_\Delta\,(a_n - f_n) \tag{2.6}$$

$$v_n \doteq a_n + d_n \tag{2.7}$$

$$= x_n + f_n = \sum_{k=0}^{p-1} h_k x_{n-k} \tag{2.8}$$

$$y_n = v_n + w_n \tag{2.9}$$

$$\tilde{d}_n \doteq y_n - z_n = y_n - \mathrm{MOD}_\Delta\,(y_n) \tag{2.10}$$

From (2.6) and (2.10), it is clear that d_n and $\tilde{d}_n$ take values on the set of points that are modulo-congruent to 0 under the modulo operation $\mathrm{MOD}_\Delta\,(\cdot)$. (these points, which are marked as small white circles in Figure 2.4, constitute a lattice).

Description of the involved signals We summarize next the statistical features of the different discrete sequences involved in THP transmission, assuming that the channel introduces a sufficiently large amount of ISI [59]:

- a_n: independent and identically distributed (i.i.d.) symbols equiprobably drawn from a square QAM constellation $\mathcal{A}$ with M elements (Pulse Amplitude Modulation (PAM) signaling will not be further considered), with minimum intersymbol distance $d_{\min}$. Therefore, its Probability Density Function (pdf) is a discrete uniform distribution taking values at the symbol locations. The average energy of a square QAM (for which $M = 2^{2q}$, with $q \in \mathbb{N}$) is given by $\sigma_a^2 \doteq \mathrm{E}\{|a_n|^2\} = d_{\min}^2(M-1)/6$.
- d_n and $\tilde{d}_n$: from (2.6) and (2.10) it is clear that both sequences take values at points which are congruent to 0 under the modulo operation $\mathrm{MOD}_\Delta\,(\cdot)$. In fact, these points constitute the precoding lattice $\mathbf{\Lambda}$. In the case of a square QAM the lattice is given by

$$\mathbf{\Lambda} = \{d_{\mathrm{I}} + j d_{\mathrm{Q}},\ \text{with}\ d_{\mathrm{I}} = m\Delta,\ d_{\mathrm{Q}} = n\Delta,\ m,\ n \in \mathbb{Z}\}.$$

The small white circles in Figure 2.4 indicate the points of $\mathbf{\Lambda}$.

- v_n: the effective data sequence follows a *discrete Gaussian* pdf, which takes values at the expanded signal set with a Gaussian decay.[3] For square QAMs the support set is given by [112] $\mathcal{V} = \{v_{\text{I}} + j\, v_{\text{Q}}$ with $d_{\text{I}} = m\, d_{\text{min}}/2$, $d_{\text{Q}} = n\, d_{\text{min}}/2$, $m,\ n \in \mathbb{Z}\}$, and the associated probabilities are given by

$$f(v) = c \cdot \exp\left(-\frac{|v|^2}{\Delta^2 \eta}\right), \quad \text{for } v \in \mathcal{V}$$

  where $c > 0$ is a normalization constant and $\eta > 0$ is the so-called *granularity parameter*. For a fixed Δ, η accounts for the span of $f(v)$ (large η causes long tails). In general, η grows with the amount of ISI introduced by the channel.
- x_n: the precoded sequence is an almost i.i.d. sequence, uniformly distributed within the region delimited by the output of the modulo operation. The modulo operation has thus a whitening effect on its output. The average energy of the sequence is given by

$$\sigma_x^2 \doteq \mathrm{E}\left\{|x_n|^2\right\} = \Delta^2/6 \tag{2.11}$$

  for square regions of side Δ. For square QAMs, this variance can be written as $\sigma_x^2 = M\, d_{\text{min}}^2/6$.

THP presents an SNR penalty with respect to normal QAM transmission, due to a slight increase in the transmitted power. The so-called *precoding loss* is defined as [59]

$$\gamma_{\text{p}}^2 \doteq \frac{\sigma_x^2}{\sigma_a^2} = \frac{\mathrm{E}\left\{|x_n|^2\right\}}{\mathrm{E}\left\{|a_n|^2\right\}}. \tag{2.12}$$

Under sufficiently strong ISI conditions, the precoding loss is a function of the modulus Δ and the original constellation. In particular, for QAM signaling the precoding loss reduces to $\gamma_{\text{p}}^2 = M/(M-1)$, which indicates that the loss becomes negligible as the constellation size increases (i.e., for high-rate transmission). Table 2.2 illustrates this fact.

Table 2.2: Dependence of the precoding loss with the square QAM constellation size.

QAM size M	γ_{p}^2 (linear units)	γ_{p}^2 (dB)
4 (QPSK)	1.333	1.25
16	1.067	0.28
64	1.016	0.06
256	1.004	0.02
1024	1.001	0.00

[3] Such distribution is related to the Gaussian distribution of the ISI, f_k, although a formal justification of the discrete Gaussian nature of v_n is complicated due to the nonlinear effect of the modulo operation. Instead, such pdf can be easily checked using simulations.

2.2.1.3 ZF-THP and MMSE-THP

So far we have only considered the use of THP designed according to the ZF criterion. The goal of the ZF approach to equalization, which is based upon the inversion of the channel transfer function, is to *completely* remove the ISI introduced by the channel. As mentioned in Section 2.2.1.1, the application of this criterion in linear equalization may incur noise enhancement when there are spectral notches in the channel frequency response, but this problem is avoided by nonlinear techniques such as DFE and THP. An alternative solution is to use the *Minimum Mean Square Error (MMSE) criterion* (see e.g. [59]), which avoids the noise enhancement problem in linear equalization since it considers the noise variance in addition to the channel transfer function (both must be known in order to design the equalizer). In contrast to the ZF criterion, the MMSE criterion admits some residual interference after correction which helps prevent noise enhancement. Generally speaking, MMSE equalization offers better results than ZF equalization at low SNRs, whereas both criteria tend to converge asymptotically as the SNR grows (they coincide in the limit).

Although THP designed according to the ZF criterion (ZF-THP) does not suffer from noise enhancement, it turns out that the MMSE approach offers performance improvements over the ZF approach at low SNRs, as it does for linear equalization. Therefore, both approaches (ZF-THP and MMSE-THP) will be considered in this chapter. The difference between both approaches lies in how the filters are computed. We refer to Appendix 2.A for the details on the actual filter computation.

2.2.2 Combining THP and Signal Shaping

2.2.2.1 Introduction to Signal Shaping

Many communication systems typically work under quite demanding conditions in terms of power or interference limitations. In such scenarios, it is crucial to achieve good SNR efficiency in the transmission. Traditionally, SNR efficiency has been pursued by means of coding. However, it is well known today that the total improvement in SNR efficiency can be separated into two contributions: *coding gain* and *shaping gain* [65].[4] Whereas coding seeks to minimize the probability of error in the transmission, the goal of signal shaping is to find the transmitted signal that achieves the desired performance with a minimum cost. Typically, this *cost* is the average transmit power, which leads to energy-efficient transmissions, although other cost functions can be considered. In all cases, signal shaping modifies the mapping between the source bits and the transmitted symbols, so as to accommodate the pdf of the symbols to achieve the desired goal.

The *shaping gain* is defined as the transmit power reduction achieved by signal shaping with respect to the corresponding conventional scheme in which equiprobable signaling is used. In terms of shaping gain, it is desirable that the boundary of the transmitted signal be as close as possible to a circle (in two dimensions), or in general to an N-sphere

[4] This assertion is valid at high data rates, when dense constellations are used. For low rates both contributions are not fully separable.

(the generalization of the sphere in Nth-dimensional space, useful when more dimensions are used in the shaping operation).[5] In general, moving to higher dimensions allows to achieve larger shaping gains [62, 65], although there is an upper bound: the *ultimate shaping gain*, which amounts to $\pi e/6$ (or 1.53 dB).

Generally speaking, shaping gains cannot be achieved without negatively affecting other parameters of the transmitted signal. In particular, larger peak energies must be transmitted, and some degree of redundancy must be added to the transmitted signal (which is accommodated by increasing the number of symbols in the constellation) in exchange for the average energy reduction. We will use the following measures to quantify such effects:

- Peak-to-Average Energy Ratio (PAR), defined as the ratio of the peak energy in the constellation to the average energy;
- Constellation Expansion Ratio (CER), defined as the ratio between the size of the (augmented) constellation used for shaping and the baseline equiprobable constellation. It is always verified that CER ≥ 1.

Signal shaping can be implemented by several means. We cite here the most representative ones:

- *Shell Mapping*. This technique is related to a kind of fixed-rate vector quantizer used in source coding. Detailed descriptions can be found in [104] and [59, chapter 4].
- *Trellis Shaping*. In this technique, the transmitted sequence is found using a trellis decoder. The trellis is built according to a given code (the shaping code), and the decoder must minimize the aggregate cost (typically the average energy) of the transmitted sequence. Trellis shaping is well described in the original source [64], as well as in [59, chapter 4].

2.2.2.2 Shaping for THP: Shaping Without Scrambling

Unfortunately, THP is not well suited for direct combination with signal shaping, because the shape of the signal at the output of the modulo device typically differs considerably from that of the signal at its input. Instead, shaping can only be applied to THP signals when both operations are performed jointly. This can be achieved by means of two techniques: *trellis precoding* and *shaping without scrambling*. Both schemes are closely related: similarly to trellis shaping, both use a Viterbi decoder at the transmitter to select the sequence with minimum weight. However, shaping without scrambling presents some advantages over trellis precoding, one of which is that it is fully compatible with THP from the receiver point of view. For this we will exclusively consider shaping without scrambling for our proposed system, and no details will be given here on trellis precoding. The reader is referred to the original work [57], and to [59, chapter 5] for details on that technique.

[5] The circle (or N-sphere) is the geometrical shape of least average energy for a given volume in two dimensions (N dimensions).

As mentioned above, signal shaping requires that some redundancy be added to the transmitted signal. In shaping without scrambling this is implicitly done by enlarging the support range of the constellation (i.e., the region delimited by the modulo operation at the transmitter) to some extent, but keeping the same modulo operation at the receiver as in plain THP. If, in addition, the same minimum intersymbol distance $d_{\min}$ used in THP is maintained, then the enlarged support range for transmission translates into the use of extra symbols which lie outside the boundaries of the original constellation.

The reason why this works is the following. In plain THP, the modulo operation tightly encloses the boundaries of the original constellation. Therefore, recalling the linearized model of Figure 2.5, the precoding sequence d_n is *uniquely determined* by the modulo operation, as indicated by (2.6). This sequence yields a precoded sequence x_n that presents a uniform pdf within the support range of the modulo operation, as we have seen. Now, if we allow the use of a larger support range, say by an integer factor K (thus CER $= K$), but keeping the same modulo operation as in THP at the receiver, then the choice of the precoding sequence is not unique: there are K possible values for d_n at each instant that will yield K different precoded symbols x_n within the enlarged support range, and all of them will produce the same output after the modulo operation in the receiver. It is this redundancy that shaping takes advantage of: the actual sequence is adequately chosen to meet the desired requirement (e.g. reduce the average transmitted power). Interestingly, it can be shown [59, chapter 4] that CER $= 2$ (one bit redundancy) is sufficient to achieve most of the available shaping gain.

In Figure 2.6 we display the constellations and support regions used when applying shaping without scrambling in two dimensions with one bit redundancy to a 16-QAM constellation. Note that the support region for shaping (and therefore the modulo operation at the transmitter) remains square, but is rotated $45°$. Note also that the expanded constellation shown in Figure 2.6c (the union of a 16-QAM and a shifted version, with shift $\boldsymbol{\lambda}$) is only one possibility. The actual region of the transmitted sequence x_n is eventually determined by the modulo operation corresponding to the support region in 2.6d, regardless of the chosen expanded constellation, as long as the latter is well built.

The block diagram of a system using shaping without scrambling is shown in Figure 2.7. The receiver is identical to that of a THP system. The transmitter operates as follows. Recalling that one bit shaping is enough to achieve most of the available shaping gain, a constellation with twice the number of symbols of the original constellation (CER $= 2$) will be used for shaping, as shown in Figure 2.6c. With this, the role of the sequence selector in Figure 2.7 is to decide whether to precode the original constellation symbol a_n or its equivalent in the expanded replica, which we denote as $a_n + \boldsymbol{\lambda}$. Note that, as seen in Figure 2.6, the shift $\boldsymbol{\lambda}$ is modulo-congruent to 0 under the $\mathrm{MOD}_{\Delta}(\cdot)$ operation. Thus, the signal entering the precoding operation can be written as $p_n = a_n + b_n\boldsymbol{\lambda}$, where b_n is a binary sequence taking the values 1 or 0 depending on whether the shift $\boldsymbol{\lambda}$ takes place or not. The precoding operation is similar to that in plain THP, except for the modulo operation, which must now match the new support region used for shaping. We will denote this operation as $\mathrm{MOD}^{(\mathrm{s})}_{\sqrt{2}\Delta}(\cdot)$, where the superscript (s) differentiates it from the modulo operation performed at the receiver (and in plain THP).

As already explained, the sequence selection reduces to a binary decision for each

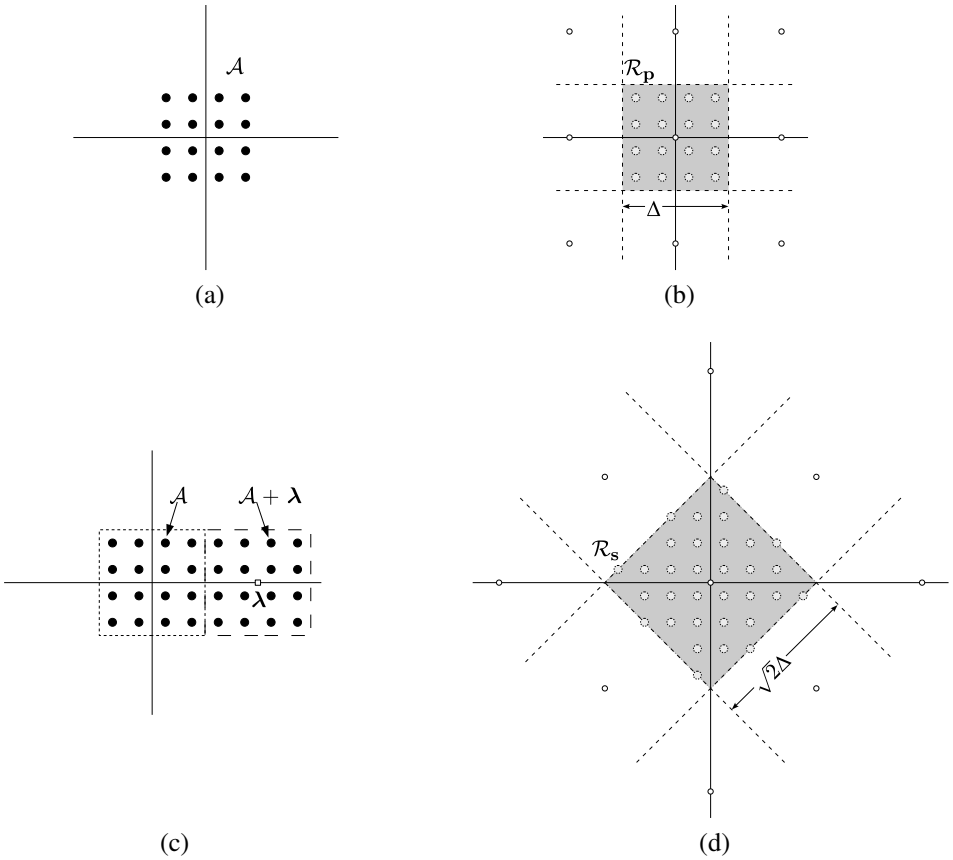


Figure 2.6: Constellations and support regions for shaping without scrambling with CER $= 2$: (a) original constellation, 16-QAM; (b) support region in THP; (c) possible expanded constellation with 32 symbols; (d) support region for shaping.

transmitted symbol (represented by the signal b_n). This selection is carried out by a Viterbi decoder. The reader is referred to the classical tutorial [63] for the basics of Viterbi decoding, which will be omitted here. Rather, we will focus on the key elements used in shaping. The Viterbi decoder operates on a trellis, which is built upon a virtual *convolutional scrambler*.[6] The choice of the actual scrambler used is not important, as long as the number of states is sufficient, as shown in [59, chapter 5]. Hence, a possible choice for the generator polynomial is (D-transform notation is used):

$$S(D) = 1 \oplus D \oplus D^s \tag{2.13}$$

where $s \in \mathbb{N}$ allows to control the number of states of the scrambler (2^s states). The Viterbi decoder is designed to minimize a given metric which is accumulated along the surviving sequences arriving at each of the 2^s states of the trellis. For example, in classical shaping, where the goal is to minimize the average transmit power, the branch metric used by the Viterbi decoder must account for the energy of the transmitted symbol associated

[6] A convolutional scrambler can be seen as a convolutional code of rate $1/1$. The scrambler is *virtual* because it does not physically exist at the transmitter; it is simply an algebraic tool to obtain a trellis representation which allows the use of Viterbi decoding for sequence selection.

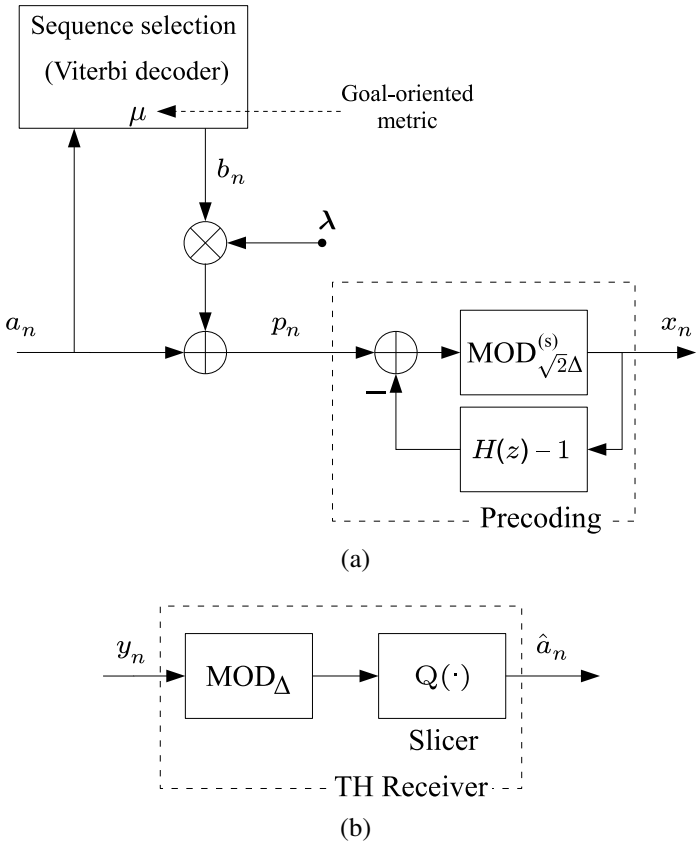


Figure 2.7: Block diagram of shaping without scrambling: (a) transmitter side; (b) receiver side.

to each state transition. The metric for state S reads

$$\mu_n = \left|x_n^{(S,b)}\right|^2 = \left|\text{MOD}^{(s)}_{\sqrt{2}\Delta}\left(a_n + b_n\boldsymbol{\lambda} - f_n^{(S)}\right)\right|^2 \tag{2.14}$$

where $x_n^{(S,b)}$ is the transmitted symbol associated to the transition leaving state S with the binary value b_n, and $f_n^{(S)} \doteq \sum_{k=1}^{p} h_k x_{n-k}^{(S)}$ is the ISI corresponding to the candidate sequence, which depends on the p (candidates to) past symbols $x_{n-k}^{(S)}$. Later in Section 2.4 we will propose other more elaborate metrics which seek other simultaneous goals such as peak-to-average ratio reduction.

2.2.3 Signal Predistortion

In a standard communication system, waveforms are generated at the transmitter at low levels and need to be amplified before they are radiated in order to close the link budget, i.e., in order to ensure they reach the receiver with at least the minimum required SNR. The amplification step is achieved by means of HPAs. These amplifiers are often an expensive part of the communication chain, both in terms of device cost and power consumption. Therefore, efficiency in HPA operation is a must in the design of most current transmitters. Satellite communications, where HPAs must typically yield large output

powers to compensate large free space losses, provide an extreme example of a communication system with demanding requirements in terms of power efficiency. However, there is a negative effect associated to high amplification: HPAs introduce nonlinear distortion in the amplified signal when they are forced to work close to their saturation points. Therefore a trade-off is established between the desire to achieve larger signal power and the undesired HPA distortion. Next we provide a brief overview of the distortion models and the compensation techniques existing in the literature.

2.2.3.1 Nonlinear Effect of the HPA

Accurately modeling the effect of the HPA allows the design of appropriate compensation methods. Different approaches exist to address the characterization of the HPA nonlinearities, depending on a number of factors such as the type of amplifier, or whether memory is included in the model. In this work we consider only memoryless models (Saleh and Rapp), which are the simplest to compensate. We briefly describe them next, including for completeness also the most representative models with memory found in the literature.

Saleh's model This is a memoryless model, originally introduced in [139] to model the effect of Traveling Wave Tube Amplifiers (TWTAs). The nonlinear distortion depends exclusively on the instantaneous envelope of the input signal. Let $x(t) = r(t)\cos(2\pi f_0 t + \phi(t))$ be the band-pass input signal, where $r(t)$ is the real-valued instantaneous envelope, f_0 the carrier frequency, and $\phi(t)$ the instantaneous phase. According to Saleh, the HPA output is modeled as

$$y(t) = A\left[r(t)\right]\cos\left(2\pi f_0 t + \phi(t) + \Phi[r(t)]\right) \tag{2.15}$$

where $A[\cdot]$ and $\Phi[\cdot]$ are, respectively, the AM/AM and AM/PM characteristics of the nonlinearity, given by

$$A[r] = \frac{\alpha_{\mathrm{a}}\, r}{\beta_{\mathrm{a}}\, r^2 + 1} \quad \text{(linear units)} \tag{2.16}$$

$$\Phi[r] = \frac{\alpha_{\mathrm{p}}\, r^2}{\beta_{\mathrm{p}}\, r^2 + 1} \quad \text{(rad)}. \tag{2.17}$$

The parameters α_{a}, β_{a}, α_{p}, β_{p} are chosen to match the characteristics of the particular amplifier, originally obtained through measurement.

If more accuracy is desired, Saleh's model can be straightforwardly extended by considering the ratio of full second-order polynomials both for the AM/AM and AM/PM characteristics:

$$A[r] = \frac{\alpha_{\mathrm{a},2}\, r^2 + \alpha_{\mathrm{a},1}\, r + \alpha_{\mathrm{a},0}}{\beta_{\mathrm{a},2}\, r^2 + \beta_{\mathrm{a},1}\, r + 1} \quad \text{(linear units)} \tag{2.18}$$

$$\Phi[r] = \frac{\alpha_{\mathrm{p},2}\, r^2 + \alpha_{\mathrm{p},1}\, r + \alpha_{\mathrm{p},0}}{\beta_{\mathrm{p},2}\, r^2 + \beta_{\mathrm{p},1}\, r + 1} \quad \text{(rad)}. \tag{2.19}$$

We will refer to (2.18)–(2.19) as the *generalized Saleh's model*.

Rapp's model This is also a memoryless model, first presented in [133], which describes the effect of Solid State Power Amplifiers (SSPAs). This model is simpler than the Saleh's model in that only AM/AM conversion is considered, since the amplifier is assumed to introduced no phase distortion [i.e., $\Phi[r] = 0$ in (2.15)]. The AM/AM characteristic under Rapp's model is given by

$$A[r] = \frac{v\, r}{\left[1 + \left(\frac{v\, r}{r_0}\right)^{2p}\right]^{1/(2p)}} \quad \text{(linear units)} \tag{2.20}$$

where the parameters v, r_0 and p take positive values and correspond to, respectively, the small signal gain, the limiting amplitude, and the smoothness factor controlling the transition from the linear region to the saturation region of the amplifier.

Models with memory Saleh's and Rapp's models are examples of memoryless models. If memory is to be accounted for in the model (e.g., in applications with large bandwidth such as W-CDMA or W-OFDM), then several approaches can be used. We cite next the most representative ones.

1. **Wiener model**. This model consists of the concatenation of a linear filter $g(t)$ and a nonlinear memoryless function $f(\cdot)$. The output signal can be written as

$$y(t) = f(x(t) * g(t)). \tag{2.21}$$

2. **Hammerstein model**. If the order of the blocks in the Wiener model is switched (i.e., memoryless nonlinearity first, then linear filter), then we have a Hammerstein model. Its output signal can be written as

$$y(t) = f(x(t)) * g(t). \tag{2.22}$$

3. **Memory polynomial**. Memory polynomials are another simple way to include memory effects in the model. The baseband version of the output signal is given by [50, 103]

$$y_n = \sum_{m=0}^{M} x_{n-m} \sum_{k=1}^{p_m} b_{m,k} |x_{n-m}|^{k-1} \tag{2.23}$$

   where $b_{m,k}$ are the coefficients of the model.
4. **Volterra series model**. The Volterra series is the most general nonlinear model, and comprises all previous models as particular cases. Its discrete-time expression is given by [22]

$$y_n = \sum_{k=1}^{\infty} \sum_{n_1} \cdots \sum_{n_{2k-1}} H^{(2k-1)}(n_1, \ldots, n_{2k-1}) \prod_{i=1}^{k} x_{n-n_i} \prod_{i=k+1}^{2k-1} x^*_{n-n_i} \tag{2.24}$$

   where $H^{(p)}(n_1, \ldots, n_p)$ is known as the p-th order Volterra kernel (for p odd).

2.2.3.2 Compensation Techniques

The compensation of the HPA-induced nonlinearity can be seen as a linearization problem: the ultimate goal is to remove the nonlinear contribution of the HPA. Such task is not fully achievable due to non-invertibility of the nonlinearity, although partial linearization can be achieved in general in a given range of the input-output characteristic, or at given sampling instants.

Several techniques have been reported in the literature as means to mitigate the nonlinear distortion introduced by the HPA:

- **Back-off**. The most straightforward choice is to back off from saturation. The Input Back-Off (IBO) is a useful system parameter defined as the ratio of the saturation power $P_{\text{sat,in}}$ to the average input power P_{in}, i.e.,

$$\text{IBO} \doteq \frac{P_{\text{sat,in}}}{P_{\text{in}}}. \tag{2.25}$$

  Large IBOs drive the working point into the HPA's linear region (the so-called *small signal* characteristic), thus preventing nonlinear effects. However this solution comes at the expense of (i) a reduced Direct Current to Radio Frequency (DC/RF) conversion efficiency, and (ii) a reduction of the available RF output power, which makes the link budget fulfillment difficult. This latter unwanted effect is measured by the Output Back-Off (OBO), which relates the maximum output power $P_{\text{max,out}}$ to the average output power P_{out}:

$$\text{OBO} \doteq \frac{P_{\text{max,out}}}{P_{\text{out}}}. \tag{2.26}$$

  In practice, some degree of IBO is typically used in combination with any of the compensation techniques described next. A trade-off must be established between the desired nonlinear interference mitigation and the undesired OBO.
- **Postdistortion**. A second possibility is the use of mitigation techniques at the receiver side. This can be efficiently achieved by using nonlinear equalizers [21], or by using maximum-likelihood sequence detection [119], which achieves optimum performance at high computational cost. The drawbacks of postdistortion are the augmented complexity of the receiver and the impossibility to prevent undesirable adjacent channel interference in the transmitted signal.
- **Predistortion**. In order to solve these latter issues, compensation can be introduced at the transmitter, before the HPA stage. In predistortion, the signal fed into the HPA has previously undergone a certain nonlinear function designed to yield a distortion-free signal at the HPA output (at least to some extent). Predistortion can be performed in two ways, depending on which signal it is applied to:
  - **Data predistortion** operates on the mapping between the binary information and the effective constellation points, prior to the conversion to waveform via the transmit filter. The underlying principle is to modify the location of the original constellation points in such a way that the effect of the HPA returns them close to their original location. Since this operation is performed before

the pulse shaping filter, the effects of this filter and the receive filter must be taken into account for the compensation.

- **Fractional predistortion** (also *waveform* or *signal predistortion*) is performed after the pulse shaping filter, thus needing to compensate for the HPA nonlinearity only. Waveform predistorters can be implemented both in analog and in digital form. In the analog case, they operate in radio frequency (RF) (or possibly intermediate frequency (IF)) and are typically made up of discrete components. In the digital case, they typically operate in baseband at oversampled rate. Digital predistorters have the advantage that they can be easily made adaptive to track system characteristics changes as a consequence of a variation in the signal characteristics (i.e., the pdf) or in the power amplifier characteristics, as a function of temperature, ageing or bias fluctuations.

For the aforementioned advantages, digital fractional predistortion combined with input back-off will be the compensation technique considered in this work.

2.3 Point-to-Point Scenario Description

Before giving the details of the proposed systems, we will provide the description of the scenario under which our solutions will be evaluated. We consider the forward link of a Geostationary Earth Orbit (GEO) satellite system such as that depicted in Figure 2.8. A fixed or nomadic ground-based terminal is located in an urban environment, and receives a data stream from the GEO satellite and one or several IMRs. The data stream, which is specifically intended to that given terminal (point-to-point or unicast transmission), is generated at a ground-based *gateway* station. The gateway sends the data stream to the satellite using the *feeder link*. Since it is likely that the satellite system must deliver some other content (e.g., other streams toward different users, broadcast TV service, etc), the gateway will be required to perform some kind of multiplexing (typically in frequency and polarization) of the signals before they are amplified and sent through the feeder link. The satellite will then perform the appropriate demultiplexing and frequency conversions to the signals before being transmitted to ground. The IMRs act as repeaters, i.e., they operate on the same band and use the same modulation as the direct signal coming from the satellite).

The following assumptions are made regarding the elements of the transmission chain.

As2.1) **Transparent IF and RF conversions.** All frequency conversions in the chain are ideal, i.e., they do not introduce any degradation in the transmitted signals.

As2.2) **Transparent (de)multiplexing.** The multiplexing at the gateway and the demultiplexing onboard do not alter the transmitted signals.

As2.3) **Transparent repeaters.** The IMRs act as ideal relays, i.e., they introduce neither any delay nor any degradation in the repeated signal.

As2.4) **No cochannel interference.** The signal received at the ground terminal is not affected by signals corresponding to other services or other possible receivers. This amounts to assuming some kind of frequency multiplexing in the data services offered by the satellite.

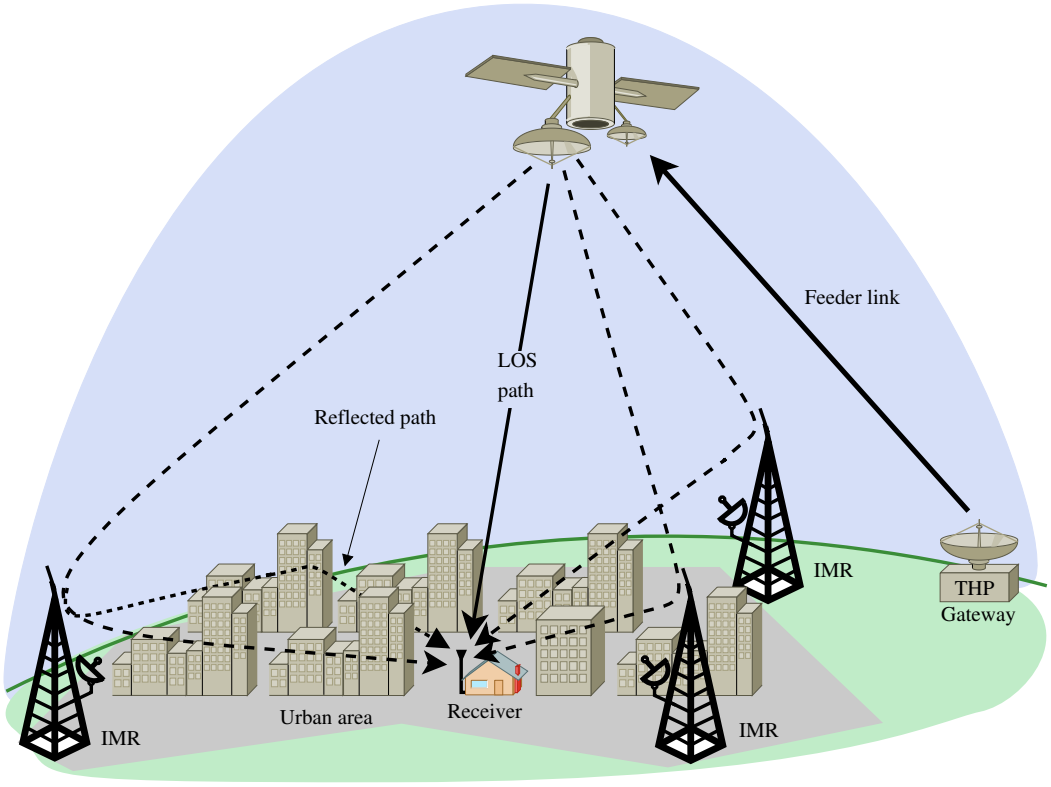


Figure 2.8: Point-to-point transmission scenario considered, in which THP is performed at the gateway.

As2.5) **Perfect timing and carrier recovery.** The receiver is assumed to implement sufficiently accurate timing, phase and frequency acquisition schemes, so that they do not have any impact in the received signal.

As2.6) **Nonlinear distortion.** The HPAs at the gateway and at the IMRs are assumed to operate under linear conditions thanks to, e.g., a generous back-off margin [39]. Thus, the only source of nonlinear distortion is the HPA onboard the satellite, whose characteristic will be known to the predistortion stage.

As2.7) **Perfect CSI knowledge at the transmitter.** The Channel State Information (CSI), i.e., the channel response of the forward link, is assumed to have a sufficiently slow variation (quasi-static channel) so that some channel estimation procedure can perfectly estimate and track the channel state at any moment. If such estimation is performed at the receiver, the existence of an ideal return channel is furthermore assumed so that the estimated CSI can be accurately passed back to the transmitter. Perfect CSI knowledge is required for the proper application of THP transmission as presented in Section 2.2. We refer to Appendix 3.E where we discuss the effects of inaccuracies in the estimated channel impulse response.

With the former assumptions, gateway and satellite act as a single transmitting unit from the receiver's point of view, and the signal received by the terminal will most likely be corrupted by ISI due to the multiple signals (and their possible reflections) coming from the satellite and the IMRs. In addition, such signal will suffer from nonlinear dis-

tortion introduced by the HPA onboard. In Section 2.4 we will propose the combination of THP and fractional predistortion as a way to counteract the ISI and the HPA-induced nonlinear distortion. Considering the above assumptions, both operations could be located either onboard the satellite or at the gateway. If THP and predistortion are to be performed onboard, the On-board Processing (OBP) requirements of the payload will be accordingly increased. This, in practice, translates into a heavier, more expensive, more power-consuming payload of the satellite system. Therefore, it seems preferable to perform such tasks at the gateway, where the processing power is not as limited as at the satellite.

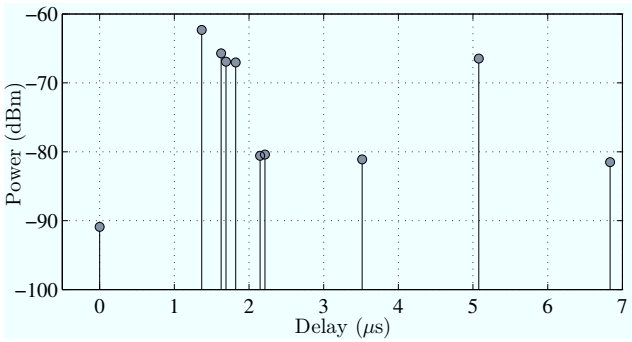


Figure 2.9: Power delay profile of the urban channel with IMRs used for evaluation. A symbol interval of $T = 0.26$ μs is considered.

The channel impulse response that will be used in the simulations, provided by the Digicomm research group from the University of Bologna, has been extracted from simulations using typical 3D urban terrain models. A data rate of 3.84 Mbaud was considered (symbol interval $T \approx 0.26$ μs). The power delay profile of the considered model, shown in Figure 2.9, was obtained for a terminal located in a street canyon with three IMRs in the surrounding area. The resulting channel, which is modeled as a tapped delay line, introduces a considerable amount of ISI due to the multiple direct incoming paths and their reflections (note that the delays of the echoes are much larger than the symbol interval T). The phases of each tap are fixed but arbitrary within the range $[-\pi, \pi)$. Complete details of the impulse response are given in Appendix 2.A. It can be seen in Figure 2.9 that the power of the zero-delay path (the Line-of-Sight (LOS) path) is lower than that of the rest of paths. This represents a typical situation in which the LOS path is more attenuated due to obstacles in the propagation path, or simply to the fact that the IMR signals arrive with more power.

The HPA characteristics, provided by the Digicomm group from the University of Bologna, have been extracted from measurements of a real TWTA located onboard. Its AM/AM and AM/PM characteristics are shown in Figure 2.10.

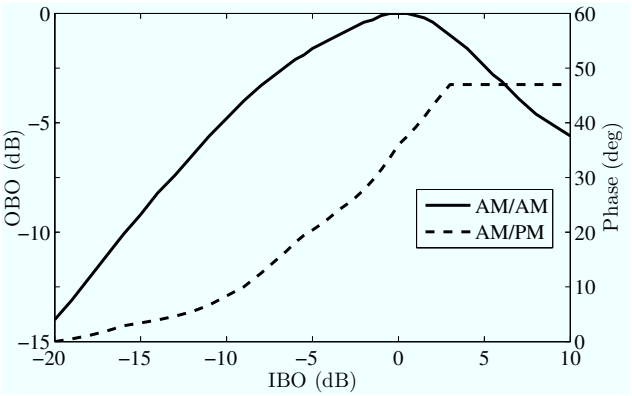


Figure 2.10: AM/AM and AM/PM characteristics of the TWTA onboard used for evaluation.

2.4 Proposed Schemes

2.4.1 Combined THP and Predistortion

We will study the combined performance of THP and fractional predistortion under uncoded transmission. A reference system with fractional predistortion and DFE at the receiver will be used for comparison.

The block diagram of the proposed system is shown in Figure 2.11. A bit sequence is mapped onto QAM symbols, which are then precoded using the standard TH scheme detailed in Section 2.2.1.2. The precoded symbols are converted to waveform by a pulse-shaping filter, which we consider a Square Root Raised Cosine (SRRC) filter of roll-off 0.3. The oversampling factor within the waveform domain is set to 4. The waveform is applied fractional predistortion before entering the HPA. The channel is a complex-valued tapped delay line as described in Section 2.3, after which complex AWGN is added. The receiver front end is a Whitened Matched Filter (WMF), which after baud rate sampling, yields a signal affected by ISI and Additive White Gaussian Noise (AWGN). The remaining ISI is removed by the modulo operation at the receiver, and then a QAM slicer produces an estimate of the transmitted symbol.

The reference system combining fractional predistortion and DFE is depicted in Figure 2.12. The only change with respect to our proposed system of Figure 2.11 is that the TH precoder and the modulo operation at the receiver are substituted by a DFE.

In this section, perfect timing and carrier recovery is assumed. We refer to Appendix 2.B, where some considerations are given regarding the timing recovery problem under THP transmission. Details on the computation of the filters involved in the system are given in Appendix 2.A.

Fractional predistorter We consider a fractional predistorter implemented using a Look-Up Table (LUT). The HPA characteristics provided in Section 2.3 (Figure 2.10), are approximated using the generalized Saleh's model given by (2.18)–(2.19). The measured

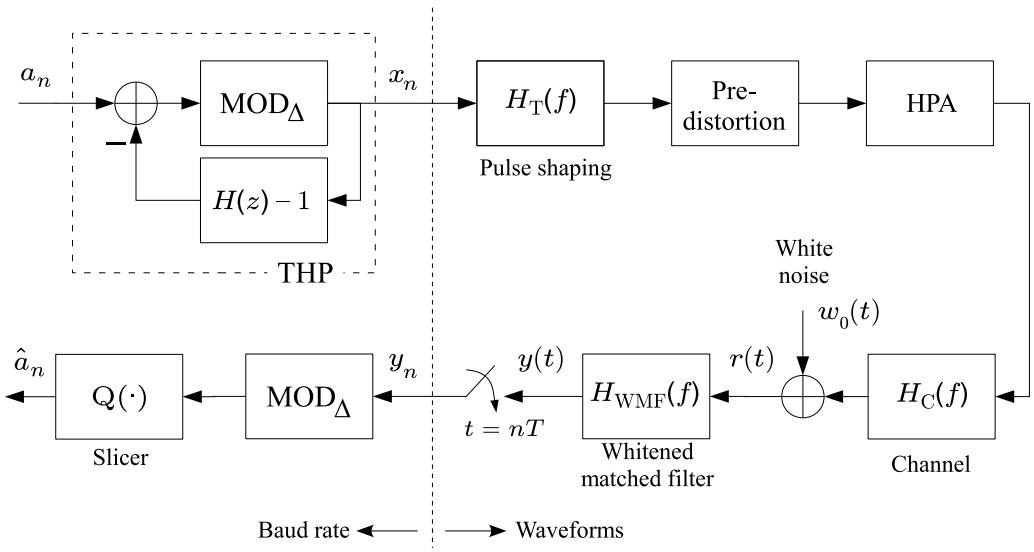


Figure 2.11: Block diagram of the proposed SISO system combining THP and fractional predistortion.

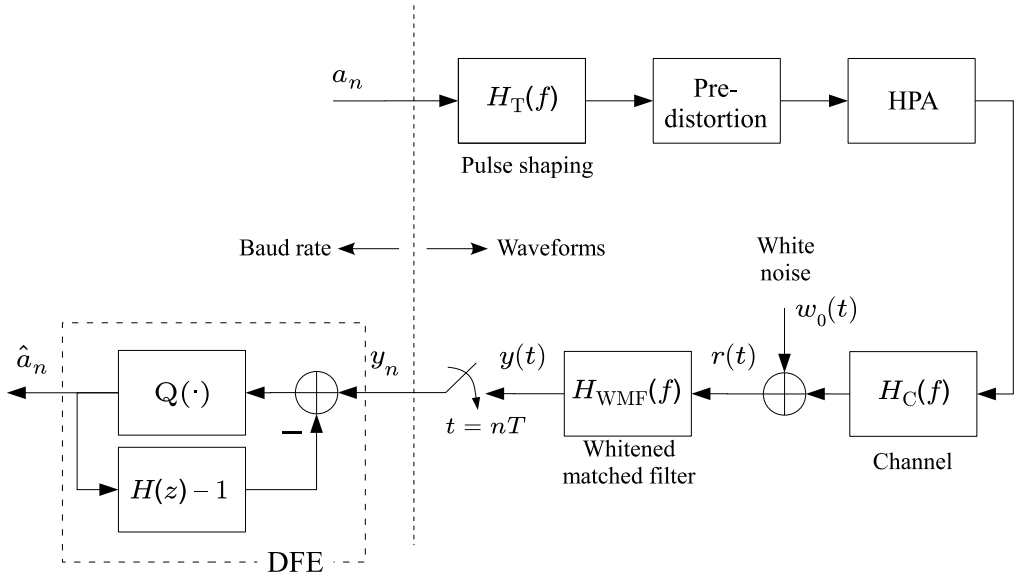


Figure 2.12: Block diagram of the proposed system combining fractional predistortion and DFE.

Table 2.3: Parameters of the generalized Saleh's model fitting of the HPA characteristics.

AM/AM	$\alpha_{a,2}$	$\alpha_{a,1}$	$\alpha_{a,0}$	$\beta_{a,2}$	$\beta_{a,1}$
	-0.0494	1.7564	0.0247	1.0393	-0.2649
AM/PM	$\alpha_{p,2}$	$\alpha_{p,1}$	$\alpha_{p,0}$	$\beta_{p,2}$	$\beta_{p,1}$
	0.3182	0.43898	-0.0324	0.6150	-0.4681

HPA characteristics are approximated using nonlinear least squares fitting. The resulting parameters are given in Table 2.3, and yield the result shown in Figure 2.13. The inverted AM/AM and AM/PM characteristics are stored in a LUT with 1024 entries indexed by the instantaneous signal power, as shown in Figure 2.14. LUT entries are uniformly spaced along the input signal power range (*linear in power indexing*), and contain the complex predistortion factors.

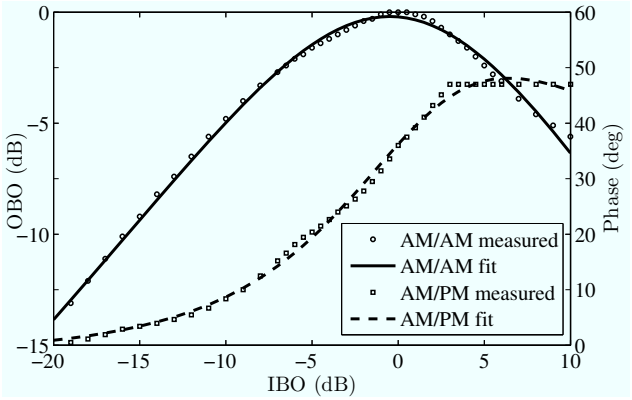


Figure 2.13: Approximation of the HPA AM/AM and AM/PM characteristics using the generalized Saleh's model.

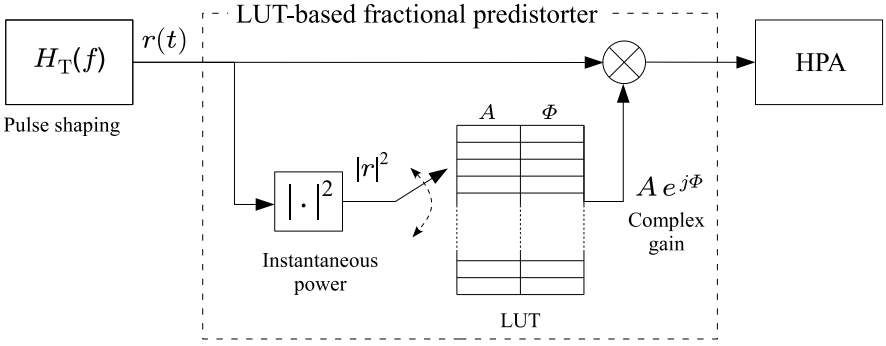


Figure 2.14: Block diagram of the LUT-based fractional predistorter with power indexing.

2.4.2 Combined Shaping Without Scrambling and Predistortion

As a way to improve the performance of the system with THP and predistortion proposed in the previous section, we consider here the combination of shaping without scrambling

and predistortion. As explained in Section 2.2.2, shaping without scrambling can yield a precoded signal with improved characteristics over plain THP.

As mentioned in Section 2.2.1.2, assuming that the channel introduces a sufficiently large amount of ISI, the output signal x_k of the TH precoder can be considered to follow a uniform distribution within the region delimited by the modulo operation (which is a square if square QAM constellations are considered). Such signal exhibits a large dynamic range, which after the SRRC filter translates into a waveform with a very disperse amplitude distribution, which is certainly undesirable considering the nonlinear amplification stage at the transmitter. In general, even if predistortion is applied, it is convenient that the distribution of amplitudes at the HPA input be as concentrated as possible so as to reduce the amount of nonlinear distortion. This is the reason why Constant Modulus (CM) constellations such as Phase Shift Keying (PSK) have been traditionally preferred for satellite communications to multilevel constellations such as QAM.

In this context, the use of shaping without scrambling can reduce the impact of the HPA by modifying the signal distribution at its input. Figure 2.15 displays the block diagram of the considered system with shaping without scrambling and predistortion. The blocks performing the shaping operation and the TH decoding operate as explained in Section 2.2.2.2 (see Figure 2.7). The generator polynomial used in the Viterbi decoder

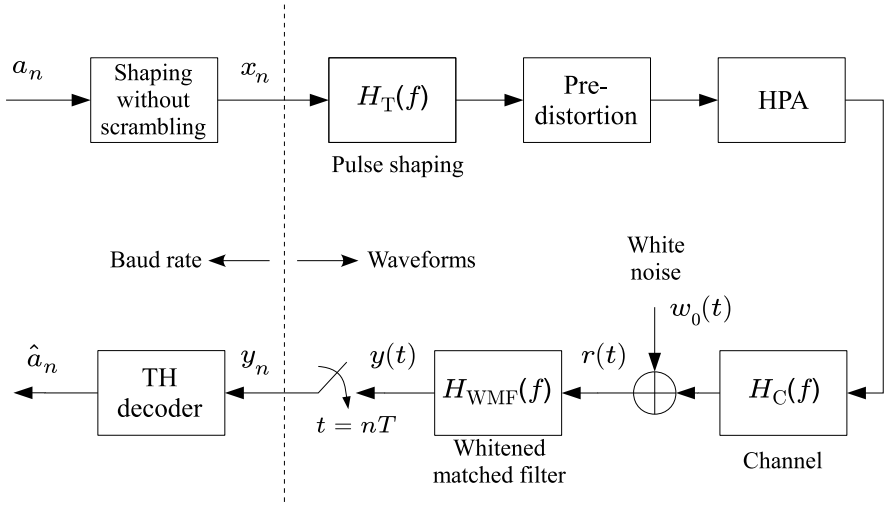


Figure 2.15: Block diagram of the proposed system combining shaping without scrambling and fractional predistortion.

is $S(D) = 1 \oplus D \oplus D^4$, and thus is a member of the family (2.13) with $s = 4$. The decoder operates therefore on $2^4 = 16$ states.

h

Metric Selection for Shaping Without Scrambling The choice of the metric used by the Viterbi decoder will determine the characteristics of the statistical distribution of the shaped signal. The ultimate design criterion of the shaping stage should be the overall improvement of the system in terms of, e.g., Bit Error Ratio (BER). However, the complexity

of the transmission chain, especially the inherent nonlinearity of the shaping, precoding, amplification and TH decoding stages, makes it difficult to map such ultimate goal onto a corresponding metric for the trellis precoder. Instead, one must resort to heuristic criteria for the selection of a shaping metric.

As a starting point we will consider the standard metric (2.14) which accounts for the instantaneous energy of the precoder output. As mentioned in Section 2.2.2.2, minimization of this metric yields the precoded sequence with minimum average energy. However the reduction in average energy (the *shaping gain*) comes at the expense of an increase in the Peak-to-Average Power Ratio (PAR) of the precoded sequence, which is undesirable considering the nonlinear amplification stage (large peak values would either incur large nonlinear distortion or require a large IBO to prevent such distortion, both undesirable effects). Exploiting the flexibility in the metric design allowed by the Viterbi decoder, we will next propose a number of alternative metrics which, heuristically, try to improve the overall performance.

1. *Minimization of average energy*. This is the metric (2.14), which we reproduce here for convenience

$$\mu_n^{(1)} = \left|x_n^{(S,b)}\right|^2 = \left|\mathrm{MOD}_{\sqrt{2}\Delta}^{(\mathrm{s})}\left(a_n + b_n\boldsymbol{\lambda} - f_n^{(S)}\right)\right|^2. \tag{2.27}$$

   Recall that $x_n^{(S,b)}$ is the precoded symbol associated to the trellis transition leaving state S with the binary value b_n, and $f_n^{(S)} \doteq \sum_{k=1}^{p} h_k x_{n-k}^{(S)}$ is the ISI corresponding to the candidate sequence, which depends on the p past symbols $x_{n-1}^{(S)}, \ldots, x_{n-p}^{(S)}$. Minimization of this metric yields a precoded sequence with minimum average energy, but large PAR.
2. *Joint minimization of average energy and Peak-to-Average Energy Ratio*. If instead of the quadratic order of the metric (2.27) we use a generic power q, we introduce one degree of freedom in the metric design that can be exploited. The metric to be minimized under this criterion is given by

$$\mu_n^{(2)} = \left|x_n^{(S,b)}\right|^q. \tag{2.28}$$

   Intuitively, if $q > 2$ is chosen, the optimal reduction of energy will not be fully achieved, but it will be possible to reduce the PAR since the higher signal amplitudes are being more penalized than under metric (2.27). Therefore, the extra degree of freedom q comes associated to a trade-off between average energy minimization ($q = 2$) and PAR reduction (which is larger as q grows). The analysis of such trade-off is difficult due to the nonlinearity of the modulo operation within the metric, so an empirical approach has instead been taken for the choice of q. The results are reported in Section 2.5.
   Metric (2.28) admits a refinement. By setting a certain amplitude threshold ρ, it is possible to apply a different exponent in the metric to each amplitude region. Thus, if the amplitude of the precoded symbol is below the threshold, it seems sensible to apply normal energy minimization applied ($q_{\text{low}} = 2$); instead, when the amplitude is over the threshold we could apply some other power $q_{\text{high}} > 2$, so as to reduce the

PAR. Such modified metric is given by

$$\mu_n^{(3)} = \begin{cases} \left| x_n^{(S,b)} \right|^{q_{\text{low}}} & \text{if } \left| x_n^{(S,b)} \right| \leq \rho \\ \left| x_n^{(S,b)} \right|^{q_{\text{high}}} & \text{if } \left| x_n^{(S,b)} \right| > \rho \end{cases} \tag{2.29}$$

with $q_{\text{low}} = 2$, and ρ and q_{high} to be empirically optimized. The results of such optimization are reported in Section 2.5.

3. *Minimization of the dispersion of the shaped signal with respect to a given amplitude* ρ. Considering the presence of the HPA, it would be desirable that the precoded signal amplitude be as concentrated as possible, as we have mentioned at the beginning of this section. Such is the case of CM constellations, which are the most suitable to undergo nonlinear amplification. We propose here the minimization of the following metric

$$\mu_n^{(4)} = \left| \left| x_n^{(S,b)} \right|^m - \rho^m \right|^q \tag{2.30}$$

in which the exponents m, q and the amplitude level ρ are to be optimized. Metric (2.30) aims at concentrating the shaped signal around a circle of fixed amplitude ρ. Such metric is expected to simultaneously drive the shape of the output signal towards that of a CM constellation, and reduce its PAR. Regarding the metric parameters, their optimization will be carried out in an empirical fashion, with the results shown in Section 2.5. It is nevertheless clear that the role of q is to control the dispersion around the circle of radius ρ.

4. *Minimizing the perturbation between consecutive symbols*. An alternative way to avoid the distortion introduced by the HPA can be to prevent large transitions between consecutive symbols of the precoded sequence. If the sequence to be transmitted presents few of such variations, it is likely that the HPA will introduce less distortion. With this purpose, we propose the minimization of the following transition metric:

$$\mu_n^{(5)} = \left| \left(x_n^{(S,b)} \right)^m - \left(x_{n-1}^{(S,b)} \right)^m \right|^q . \tag{2.31}$$

As in the previous cases, an analytical selection of the most appropriate pair (m, q) is not trivial, so we will use an empirical approach, with the results shown in Section 2.5.

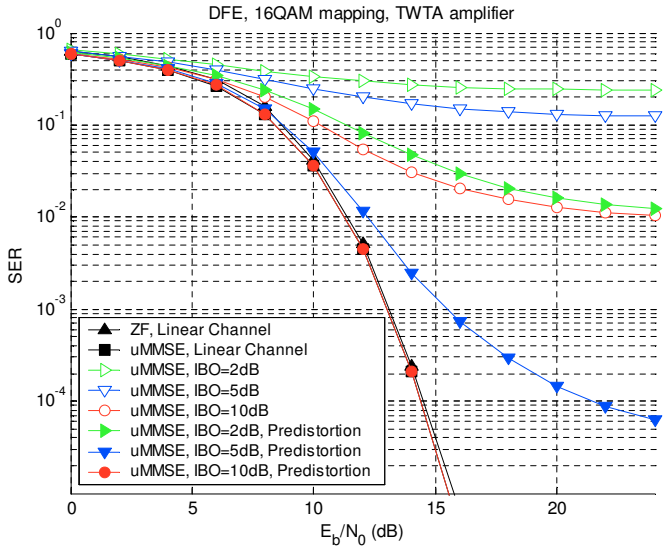
2.5 Evaluation and Discussion

We will present next the results of the evaluation of the systems proposed in the previous section.

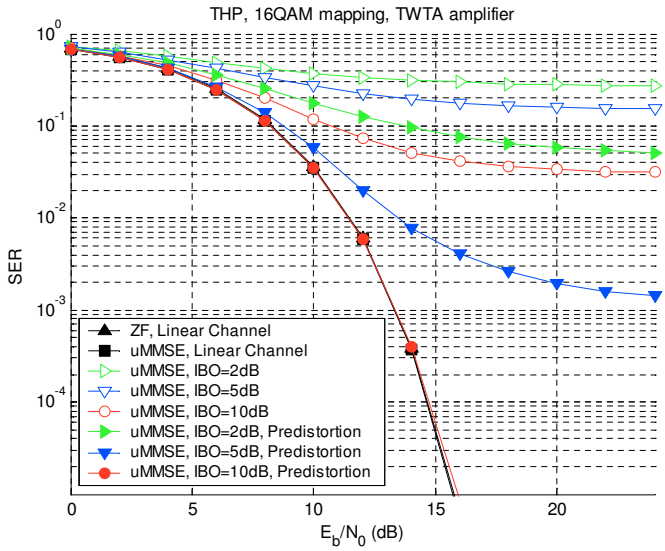
2.5.1 Combined THP and Predistortion

We have performed simulations of the systems combining THP with fractional predistortion, and DFE with predistortion, the latter used as benchmark. The details on the transmission scheme, including predistortion and HPA, are described in Section 2.4.1. For the

particular details on the discrete-time filter computation we refer to Appendix 2.A. Figure 2.16 displays the results obtained in terms of uncoded Symbol Error Rate (SER) for 16-QAM mapping by both systems using unbiased MMSE equalization, and for different IBOs. For reference we also show in both cases the results for ZF equalization, and for an ideal (linear) HPA.



(a)



(b)

Figure 2.16: Performance in terms of uncoded SER of (a) fractional predistortion and DFE, and (b) joint THP and fractional predistortion.

Several observations be extracted from Figures 2.16. To begin with, the advantage of using fractional predistortion is evident in both the DFE and THP cases. The nonlinear distortion introduced by the HPA translates into a high error floor, which can be slightly

reduced by increasing the IBO. For example, with IBO $= 2$ dB the floors are found to be SER $= 0.2$ (DFE) and SER $= 0.3$ (THP), whereas with IBO $= 10$ dB they reduce to SER $= 0.01$ (DFE) and SER $= 0.03$ (THP). Such behavior is as predicted from theory, since setting the HPA working point away from saturation (i.e., increasing the IBO) reduces the probability of saturation, and thus, the total amount of nonlinear distortion.

The addition of fractional predistortion proves quite useful in both systems. The use of predistortion is able to significantly reduce the impact of the HPA even with small IBOs. For example, with IBO $= 2$ dB, the floor is reduced from SER $= 0.2$ to 0.01 (DFE), and from SER $= 0.3$ to 0.05 (THP). However, predistortion is more efficient for larger IBOs. In particular, the combination of IBO $= 10$ dB and fractional predistortion completely removes the effects of the nonlinear distortion, both in the DFE and THP cases, at least above SER $= 10^{-5}$.

Although for the sake of clarity the performance of the DFE and THP appear in separate plots, comparison of Figures 2.16a and 2.16b reveals two differences between the performance of the DFE system and the THP system. On one hand, a small shift of about 0.3 dB can be observed for $E_\mathrm{b}/N_0 = 15$ dB between the performances for the linear case of DFE and THP. This shift, which favors the DFE system, is due to the precoding loss, which according to 2.2 amounts to $\gamma_\mathrm{p}^2 = 0.28$ dB for the 16-QAM.

The second difference between the performance of DFE and THP is that, for the same value of IBO, the system with DFE is less affected by the HPA than the system with THP. This is true whether predistortion is used or not. This fact can be explained recalling that the uniform distribution of the precoded signal is not well-suited to undergo nonlinear amplification. This has been already mentioned several times in this work, but let us justify it. In Figure 2.17 we display the Complementary Cumulative Density Function (CCDF) of the waveform amplitude at the output of the SRRC filter (roll-off 0.3) under different modulations including THP. The CCDFs were empirically obtained using a large oversampling factor, and all waveforms were normalized in power. Among all waveforms,

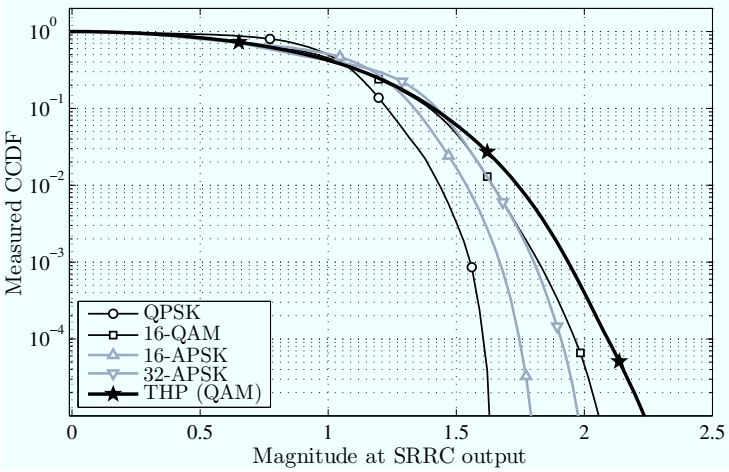


Figure 2.17: Complementary cumulative densities of the magnitude of the SRRC filter output (roll-off 0.3) under different modulations. All waveforms are normalized in power.

QPSK modulation is the one that offers the smallest peak magnitude, exhibiting a quickly

decreasing tail. The same effect is observed for the rest of modulations of the PSK family (not included in Figure 2.17). This is the reason why such modulations are the most suitable ones for undergoing nonlinear modulation: They minimize the peak amplitude of the waveform for a given average energy. The second best modulation in this sense is the 16-APSK, whereas the 16-QAM and 32-APSK modulations come next, with a similar decay at large amplitudes. The worst results are obtained for the THP-modulated waveform, which exhibits the largest peak and the slowest decay in large amplitudes.[7] This fact explains why THP signals suffer more nonlinear degradation than 16-QAM signals, as observed in Figure 2.16.

The effects of the HPA are illustrated in Figure 2.18, which displays scatter plots of the HPA output (oversampled waveform) and of the slicer input at the receiver (symbol rate) in the absence of noise. In general, it can be seen that the signal entering the slicer is less distorted in the DFE case than in the THP case (compare columns (2) and (4)). Also, the benefits of predistortion are clear if we compare slicer's inputs between rows (b) and (c) or between rows (d) and (e).

2.5.2 Combined Shaping Without Scrambling and Predistortion

The performance of the proposed system combining shaping without scrambling and predistortion was evaluated using Monte Carlo simulations. The modulation used in the simulations was 16-QAM, and only ZF precoding was considered for simplicity. Among the five proposed shaping metrics (2.27)–(2.31), only the energy-minimizing metric (2.27) has been reported in the literature. The remaining metrics (2.28)–(2.31) are novel, and all of them exhibit one or more free design parameters. Therefore, the goal of the simulations was twofold. On one hand, a first round of simulations allowed us to sweep over a sensible range of the different parameters and to identify the set of parameters which offered the best results. As mentioned in Section 2.4.2, such empirical approach was adopted in view of the difficulty of a theoretical analysis of the shaping operation due to the nonlinearities involved in the shaping and amplification stages. A second round of simulations was then performed to properly evaluate the performance of each metric with the best set of parameters. This allowed us to prove the usefulness of the shaping approach, and effectively identify which metrics are most suitable for implementation.

Table 2.4 lists the set of parameters for which each proposed metric obtained the best performance. The selection criterion was the lowest achieved BER, which in general corresponded to a good trade-off between the shaping gain G_s and the PAR. For each metric, Table 2.4 shows the achieved G_s and PAR_{99}, defined as the PAR considering the peak to be located at the 99% of the empirical cumulative density function of the shaped sequence. For reference, plain THP (no shaping) is included in the table. As expected from theory, the classical metric $\mu^{(1)}$ achieves the maximum shaping gain, but this comes associated to an increase of 0.9 dB in PAR_{99} over plain THP. The rest of metrics offer

[7] The curve shown for THP in Figure 2.17 applies to all QAM modulations in general. The effect of the precoding loss would yield different average powers depending on the QAM constellation, but it is compensated by the power normalization applied to the waveforms.

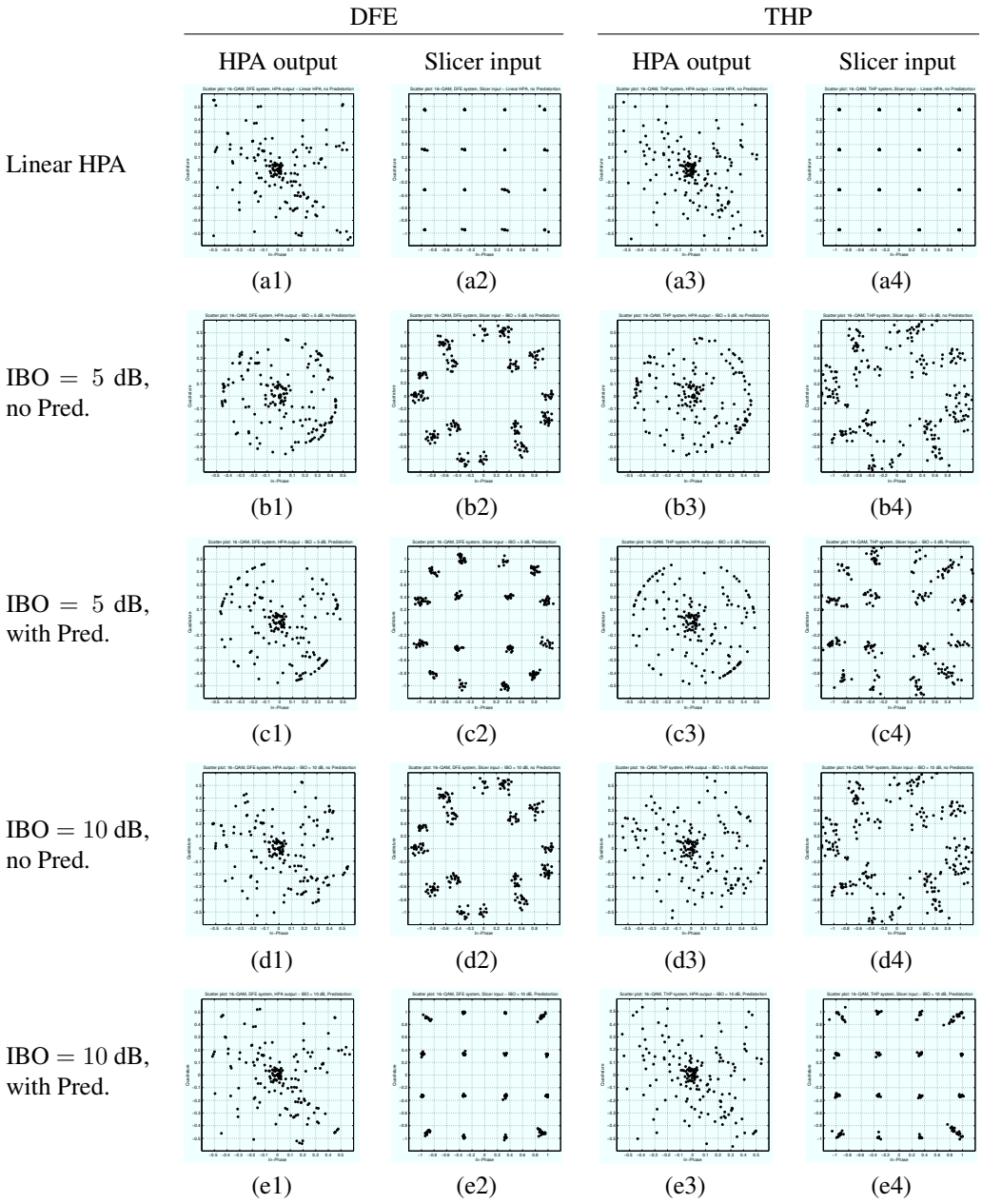


Figure 2.18: Scatter plots of the signals at the HPA output (columns 1 and 3) and at the slicer input (columns 2 and 4) in the system with DFE (columns 1–2) and the system with THP (columns 3–4). No noise was present in the simulations. Different IBO cases are shown in each row: (a) linear HPA; (b) and (c) IBO $= 5$ dB respectively without and with predistortion; (d) and (e) IBO $= 10$ dB respectively without and with predistortion.

lower shaping gains, but they are able to reduce the PAR_{99} between 0.5 and 0.8 dB with respect to $\mu^{(1)}$.

The performance of the shaping metrics in terms of BER for different IBOs is shown in Figure 2.19 for the cases without predistortion (top) and with predistortion (bottom). The performance of plain THP is shown for reference. We note that the results for metric

Table 2.4: Parameter selection for the proposed shaping metrics, and corresponding shaping gain G_{s} and PAR.

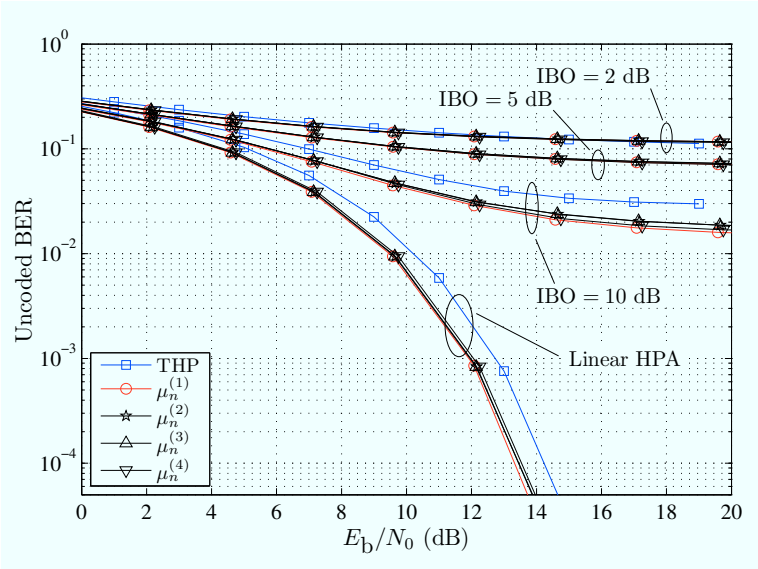
Metric	Eq. number	Parameters	G_{s} (dB)	PAR_{99} (dB)
THP	—	—	0	4.12
$\mu_n^{(1)}$	(2.27)	—	0.42	5.01
$\mu_n^{(2)}$	(2.28)	$q = 4$	0.34	4.49
$\mu_n^{(3)}$	(2.29)	$q_{\mathrm{low}} = 2$, $q_{\mathrm{high}} = 4$, $\rho = 0.6$	0.36	4.49
$\mu_n^{(4)}$	(2.30)	$m = 1$, $q = 4$, $\rho = 0.5$	0.25	4.25
$\mu_n^{(5)}$	(2.31)	$m = 5$, $q = 1$	0.24	4.22

$\mu^{(5)}$ are not included in Figure 2.19 since such metric is not able to achieve an acceptable BER despite offering a good trade-off between G_{s} and PAR_{99}.

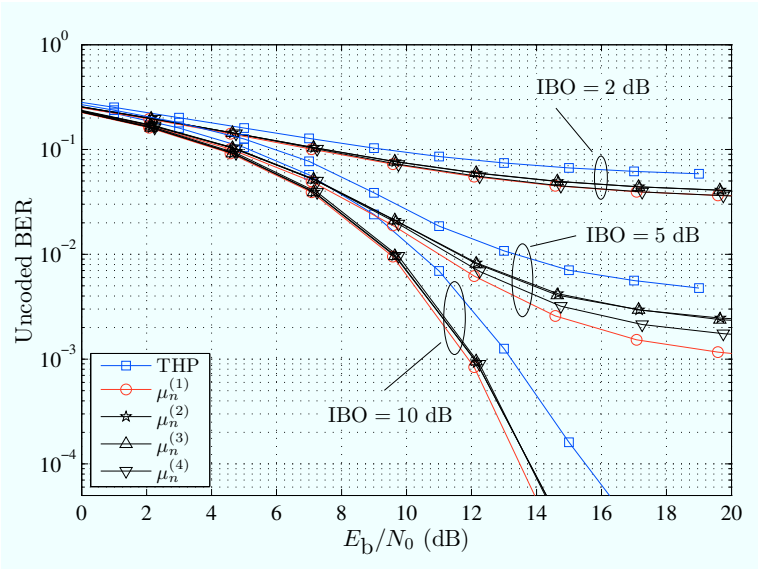
Considering the performance in absence of predistortion (Figure 2.19a), the HPA nonlinearity causes high performance floors, which can only be slightly reduced by increasing the IBO. Still, there is an evident improvement over plain THP achieved by means of shaping for the cases of linear HPA (about 1 dB gap) and IBO $= 10$ dB. When fractional predistortion is included (Figure 2.19b), the benefits of applying shaping without scrambling are even more evident as seen for IBO $= 2$, 5, and 10 dB. In particular, for IBO $= 10$ dB the performance of the system with shaping achieves the same performance as for a linear HPA (not shown in Figure 2.19b for clarity). A comparison with Figure 2.19a for IBO $= 2$ and 10 dB shows that the gap between shaping and plain THP increases when predistortion is used, meaning that predistortion is even more useful when combined with shaping without scrambling than when combined with plain THP. For example, the gap between shaping and plain THP for IBO $= 10$ dB reaches 2 dB at BER $= 10^{-4}$.

In what concerns the compared performance of the four shaping metrics evaluated in Figure 2.19, no meaningful differences in performance are observed among them when no predistortion is applied. However, differences can be observed when predistortion is included. For example, for IBO $= 5$ dB the classic energy-minimizing metric $\mu_n^{(1)}$ offers the best performance, followed by metric $\mu_n^{(4)}$ (minimization of the dispersion of the symbol magnitude), and then metrics $\mu_n^{(2)}$ and $\mu_n^{(3)}$, which offer similar performance. For IBO $= 10$ dB metric $\mu_n^{(1)}$ holds the best performance, with a gap of approximately 0.3 dB in E_{b}/N_0 over the rest of metrics at high SNR.

The fact that metric $\mu_n^{(1)}$ offers the best results provides some insight about the relative influence of the shaping gain and the PAR of the precoded sequence in the performance of the proposed transmission scheme. Recalling Table 2.4, it is interesting to note that metric $\mu_n^{(1)}$ achieves the highest shaping gain but also yields the largest PAR_{99}. The rest of metrics offer smaller PAR_{99} in exchange of a reduced G_{s}. Thus, it appears that a small increase in the shaping gain has a more beneficial effect in the BER than a similar decrease in the PAR of the precoded sequence when predistortion is combined with shaping without scrambling.



(a)



(b)

Figure 2.19: Performance in terms of uncoded BER of shaping without scrambling for the cases (a) without predistortion, and (b) with predistortion.

To conclude, we have seen that the joint application of shaping without scrambling and signal predistortion in the downlink of a point-to-point satellite system successfully counteracts the gap filler-induced ISI and the nonlinear distortion caused by the High Power Amplifier onboard. Therefore, this approach seems attractive for practical systems with fixed and nomadic terminals, since the ISI is precanceled at the transmitter (e.g., the gateway), which should simplify the complexity of the terminals.

Appendix 2.A Filter Computation for SISO THP

Consider Figure 2.20, which represents the continuous-time signal path between the symbol to waveform conversion at the transmitter and the sampling at the receiver. This continuous-time path corresponds to the waveform-domain chain of Figures 2.11 and 2.12, except for the nonlinear blocks (predistortion and HPA), which are not considered here (i.e., perfect predistortion is assumed). As shown in Figure 2.20, it is possible to describe the system behavior at baud rate by means of an equivalent discrete-time filter $H(z)$ and a discrete white noise source.

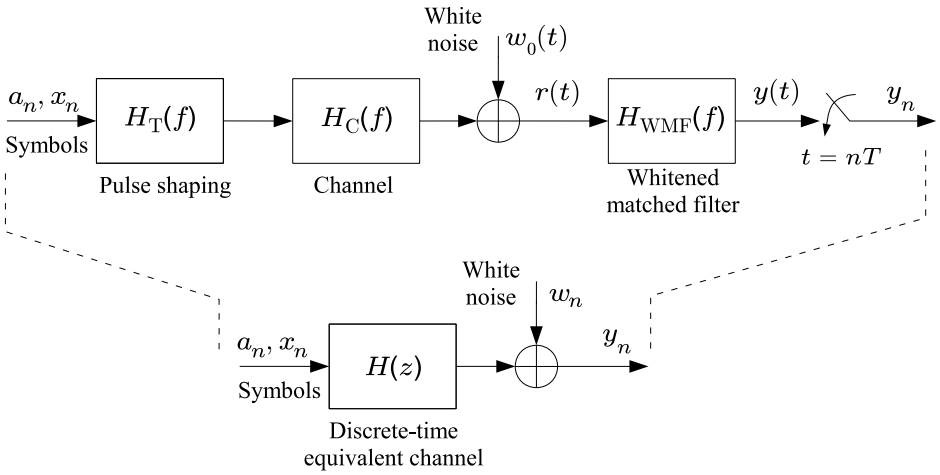


Figure 2.20: Continuous-time channel (top) and corresponding discrete-time equivalent channel (bottom) of the considered system.

We assume that the pulse shaping filter $H_{\mathrm{T}}(f)$ and the channel $H_{\mathrm{C}}(f)$ are fixed and known, and we want to find the required receive filter $H_{\mathrm{WMF}}(f)$ (the *whitened matched filter*) and the equivalent discrete-time filter $H(z)$ required by THP and DFE. An oversampling factor of 4 was used in the simulations to represent the continuous-time signals.

Pulse shaping filter As pulse shaping filter we consider a SRRC filter with roll-off $\alpha = 0.3$. The frequency response of the SRRC filter is the square root of that of the Raised Cosine (RC) filter of equal roll-off. The RC filter is a real-valued filter with even symmetry which features Nyquist characteristics (i.e., it introduces no ISI when correctly sampled at baud rate). The frequency response of a RC filter is given by

$$H_{\mathrm{RC}}(f) = \begin{cases} T, & 0 \le |f| \le \frac{1-\alpha}{2T} \\ \frac{T}{2}\left[1 + \cos\left(\frac{\pi T}{\alpha}\left(|f| - \frac{1-\alpha}{2T}\right)\right)\right], & \frac{1-\alpha}{2T} \le |f| \le \frac{1+\alpha}{2T} \\ 0, & |f| > \frac{1+\alpha}{2T} \end{cases} . \tag{2.32}$$

The frequency domain and time domain responses of the SRRC and RC filters are displayed in Figure 2.21.

Channel impulse response The channel response is chosen to comply with the power delay profile of Figure 2.9. A static tapped delay line is used as time-domain channel response (with 4 samples per symbol period), i.e.:

$$h_{\mathrm{C}}(t) = \sum_{k=1}^{10} \rho_k \, e^{j\phi_k} \delta(t - \tau_k) \tag{2.33}$$

where $\rho_k \, e^{j\phi_k}$ is the k-th complex tap of amplitude ρ_k and phase ϕ_k, and τ_k is its corresponding delay. Table 2.5 details the parameters of the time-domain impulse response of the channel (up to an amplitude scaling). As explained in Section 2.3, the tap phases ϕ_k were arbitrarily chosen.

Whitened matched filter The WMF is chosen as receiver front-end. The WMF is designed to be matched to the transmit chain and, when followed by proper baud rate sampling, provides samples affected by uncorrelated noise. As shown in Figure 2.22, The WMF can be thought as a single analog filter $H_{\mathrm{WMF}}(f)$ followed by baud rate sampling, or it can be seen as the concatenation of (i) a continuous-time *matched filter* $H_{\mathrm{MF}}(f) \doteq H_{\mathrm{C}}^*(f) \cdot H_{\mathrm{T}}^*(f)$, (ii) baud rate sampling, and (iii) a discrete-time *noise-whitening filter* $H_{\mathrm{NW}}(z)$. This latter filter removes the correlation that the matched filter introduces in the noise component.

Discrete-time channel and noise-whitening filters The noise-whitening filter $H_{\mathrm{NW}}(z)$ and the discrete-time equivalent filter $H(z)$ are necessarily related, see Figures 2.20 and 2.22. In fact, they are found as solutions of a spectral factorization problem applied to a certain

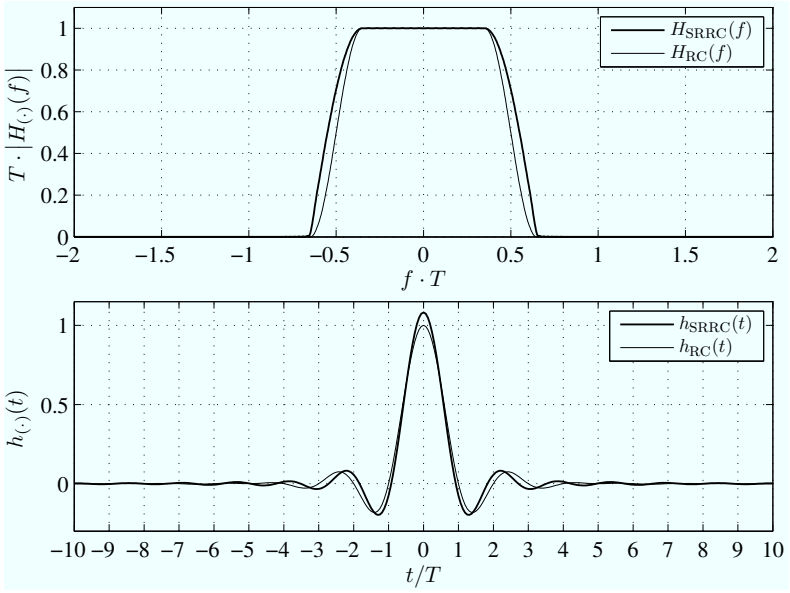


Figure 2.21: Frequency-domain (top) and time-domain (bottom) response of the SRRC and RC filters with roll-off $\alpha = 0.3$.

Table 2.5: Time-domain impulse response of the channel used for simulation. The assumed symbol rate is $1/T = 3.84$ Mbaud.

Tap Nr.	τ_k (μs)	τ_k/T	Power (dBm)	ρ_k	ϕ_k (rad)
1	0	0	-90.909	1	0
2	1.3672	$21/4$	-62.319	26.8844	0.2315
3	1.6276	$25/4$	-65.719	18.1761	-2.7677
4	1.6927	$26/4$	-66.929	15.8125	-2.5826
5	1.8229	$28/4$	-67.039	15.6135	-1.4369
6	2.1484	$33/4$	-80.579	3.2487	-0.5713
7	2.2135	$34/4$	-80.399	3.3535	-0.1631
8	3.5156	$54/4$	-81.119	3.0867	-1.5745
9	5.0781	$78/4$	-66.469	16.6725	-0.2711
10	6.7708	$104/4$	-81.509	2.9512	-0.8634

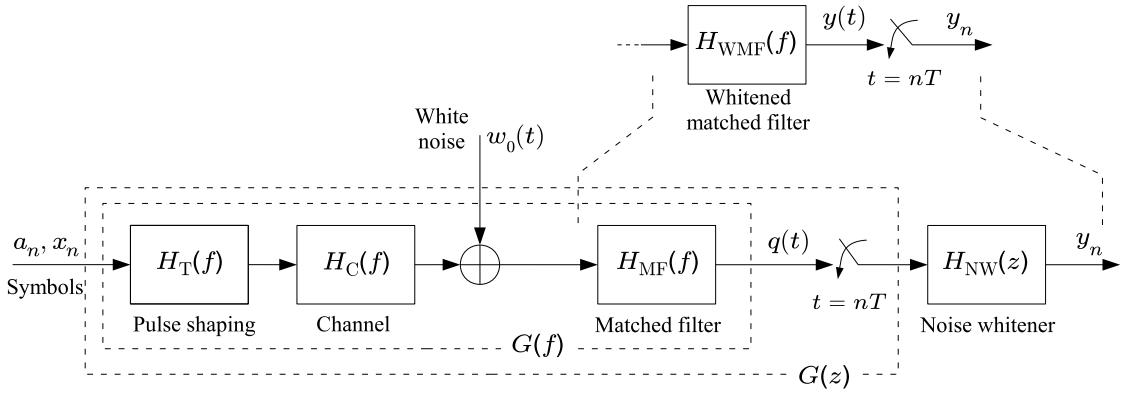


Figure 2.22: Two equivalent representations of the WMF: single continuous-time front-end (top), and concatenation of continuous-time matched filter and discrete-time noise-whitening (bottom).

discrete-time transfer function $U(z)$, whose definition changes according to the equalization criterion.[8] We consider in this chapter the ZF and MMSE criteria. The corresponding transfer functions to be factored are given by (subscripts indicate the criterion)

$$U_{\text{ZF}}(z) \doteq \sigma_a^2 \cdot G(z) \tag{2.34}$$

$$U_{\text{MMSE}}(z) \doteq \sigma_a^2 \cdot G(z) + \frac{N_0}{T} \tag{2.35}$$

where $\sigma_a^2 = \mathrm{E}\{|a_n|^2\}$, N_0 is the value of the Power Spectral Density (PSD) of the continuous-time white noise process $w_0(t)$, and T is the symbol period.[9] $G(z)$ is the discrete-time transfer function of the system as if the baud rate samples were taken at the output of the matched filter, see Figure 2.22. It is readily seen that $G(z)$ is proportional to the PSD of the discrete-time noise component of the baud rate samples at the matched filter output. If we define $G(f)$ as the continuous-time transfer function of the concatenation of transmit filter, channel, and matched filter, i.e.,

$$G(f) \doteq H_{\text{T}}(f) \cdot H_{\text{C}}(f) \cdot H_{\text{MF}}(f) = |H_{\text{T}}(f) \cdot H_{\text{C}}(f)|^2 \tag{2.36}$$

then $G(z)$ can be seen as the baud rate equivalent transfer function of $G(f)$. If $G(z)$ is evaluated in the unit circle, we obtain the discrete-time frequency response of the sampled system $G(e^{j2\pi T})$, which is related to $G(f)$ by

$$G(e^{j2\pi fT}) = \sum_{\mu} G\left(f - \tfrac{\mu}{T}\right) = \sum_{\mu} \left|H_{\text{T}}\left(f - \tfrac{\mu}{T}\right) \cdot H_{\text{C}}\left(f - \tfrac{\mu}{T}\right)\right|^2. \tag{2.37}$$

Figure 2.23 shows the magnitudes of the overall continuous-time transfer function $G(f)$ (top) and its periodic continuation in continuous-time obtained after baud rate sampling (middle). The time-dispersive nature of the channel response $H_{\text{C}}(f)$ is evident. The inverse of the periodic continuation of $G(f)$ is displayed at the bottom of Figure 2.23 to show the importance of the deep notch that $G(f)$ presents about $f \cdot T = 0.06$, which would cause severe noise enhancement if ZF linear equalization were used.

Regardless of the equalization criterion, $U(z)$ must be factored as follows:

$$\sigma_a^2 \cdot U(z) = \sigma_h^2 \cdot H(z) \cdot H^*(1/z^*) \tag{2.38}$$

where σ_h^2 is a positive constant (with the dimensions of variance), and $H(z)$ is the desired equivalent discrete-time impulse response required for THP and DFE. If we recall Section 2.2.1.2, $H(z)$ is required to be *causal* and *minimum-phase* with $h_0 = 1$. This implies that $U(z)$ must be factored into two transfer functions: $H(z)$ containing all the zeros of $U(z)$ lying inside the unit circle, and $H^*(1/z^*)$, which contains the zeros of $U(z)$ outside the unit circle. The factor σ_h^2 is introduced to ensure that $H(z)$ is a monic polynomial (i.e., $h_0 = 1$). Figure 2.24 displays $G(e^{j2\pi fT})$ as well as the location of the roots of $G(z)$, which coincide with the roots of $U_{\text{ZF}}(z)$ since both transfer functions are proportional.

[8] We note that the filters $H(z)$ and $H_{\text{NW}}(z)$ provided by the factorization procedure are valid for both the THP and DFE implementations of Figures 2.11 and 2.12.

[9] Note once again that $U_{\text{MMSE}}(z)$ converges asymptotically to $U_{\text{ZF}}(z)$ as the SNR approaches infinity (i.e., as $N_0 \to 0$).

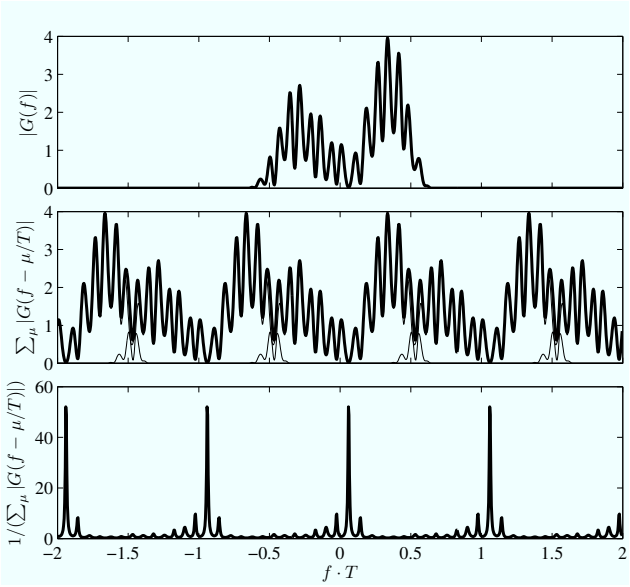


Figure 2.23: Magnitude (in linear units) of the overall transfer function $G(f)$ at the output of the matched filter (top), its baud rate sampled version (middle), and the inverse of the latter (bottom).

With the factorization (2.38) the discrete-time noise-whitening filter is obtained as

$$H_{\text{NW}}(z) \doteq \frac{\sigma_a^2}{\sigma_h^2} \cdot \frac{1}{H^*(1/z^*)} \tag{2.39}$$

where the scaling ensures that $H_{\text{NW}}(z)$ is a monic polynomial. Figure 2.25 shows the discrete-time impulse response h_n corresponding to $H(z)$ obtained in the ZF case.

In the MMSE case, $H(z)$ is related to $U_{\text{MMSE}}(z)$ by

$$H(z) = 1 - \frac{N_0/T}{U_{\text{MMSE}}(z)} \tag{2.40}$$

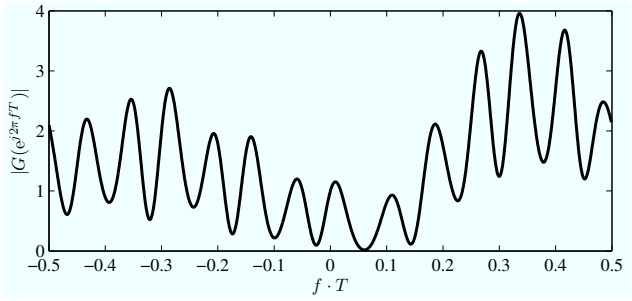
from which it can be seen that the coefficient h_0 is not one but

$$h_0 = 1 - \frac{N_0/T}{\sigma_h^2} < 1. \tag{2.41}$$

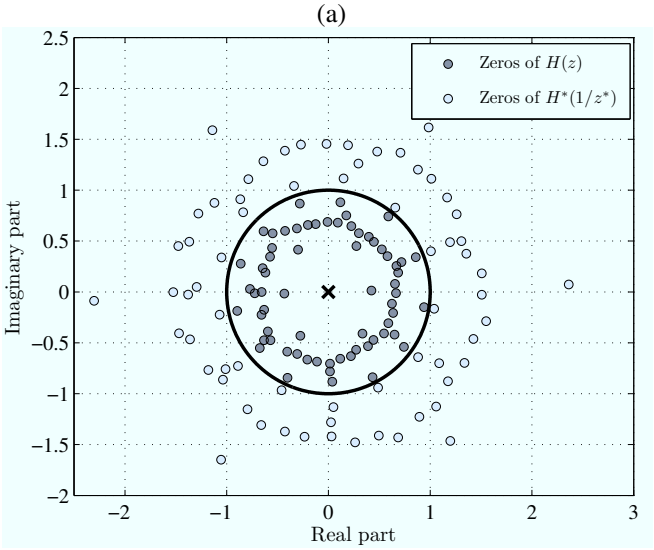
This fact translates into a bias in the signal entering the decision device at the receiver, which, if not corrected, would cause undesired decisions. In order to remove this bias, an appropriate scaling by h_0^{-1} must be applied to the signal entering the decision device when MMSE precoding is used. We refer to this approach as "unbiased MMSE".

Appendix 2.B Timing Recovery

Consider the SISO setup with the discrete-time equivalent linear model given in Figure 2.5. The received baud rate samples y_n (i.e., the signal at the input of the receiver's



(a)



(b)

Figure 2.24: Magnitude of the discrete-time frequency response $G(e^{j2\pi fT})$ in linear units (a), and spectral factorization of the zeros of $G(z)$ (b).

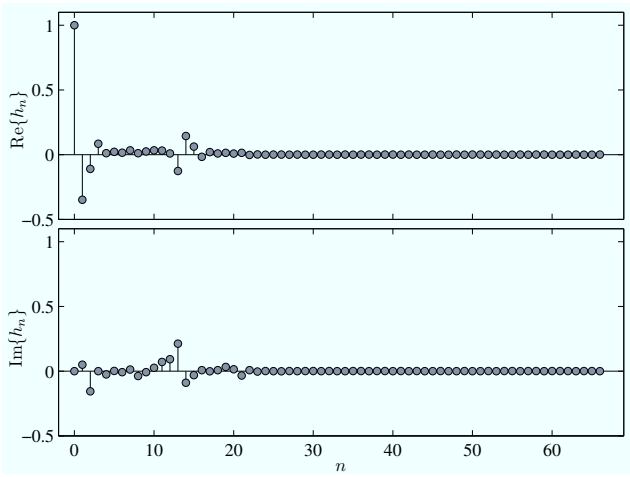


Figure 2.25: Discrete-time equivalent impulse response h_n obtained in the ZF case.

modulo device) turns out to be a noisy version of the effective data sequence v_n. As mentioned in Section 2.2.1.2, the effective symbols v_n follow a discrete Gaussian distribution taking values on the set of points which are modulo congruent to the original QAM constellation. Figure 2.26 shows a typical scatter plot for y_n assuming perfect timing recovery. Three different ISI conditions are shown, corresponding with three different channel transfer functions: (a) no ISI (i.e., $H(z) = 1$), (b) weak ISI conditions, and (c) strong ISI conditions. The transfer function providing strong ISI conditions is that considered for evaluation in the main text of this chapter, which is extracted from the time-domain oversampled impulse response provided in Table 2.5. The transfer function providing weak ISI conditions is extracted from the oversampled impulse response given in Table 2.6.[10] In Figure 2.26 it is observed that the received discrete Gaussian becomes more spread as the amount of ISI increases, as it was noted in Section 2.2.1.2. The application of THP turns out to pose important hurdles to the timing acquisition task. In this section we provide an overview of the timing acquisition problem in SISO THP systems.

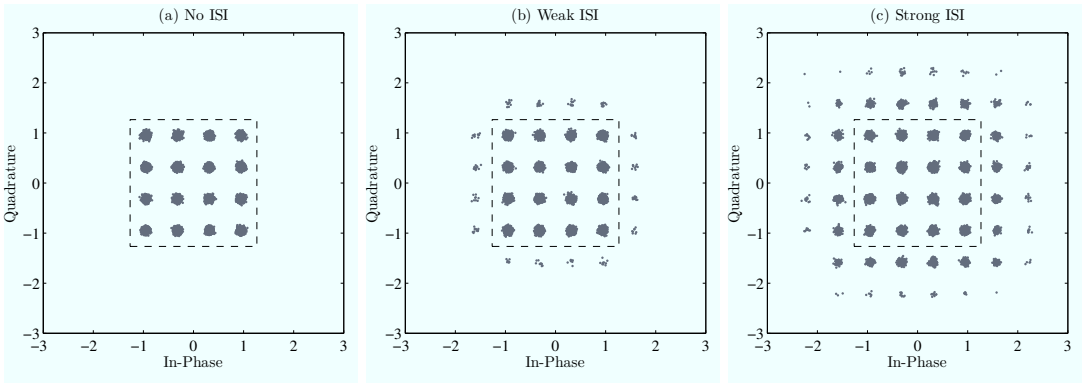


Figure 2.26: Typical scatter plot of the symbols y_n at the input of the modulo device of the receiver, for 16-QAM THP transmission with perfect timing synchronization and $E_\mathrm{b}/N_0 = 20$ dB. Three cases are shown: (a) ISI-free conditions; (b) weak ISI conditions; (c) strong ISI conditions. The dashed square represents the limits of the modulo operation.

Symbol timing recovery can be performed in continuous (or feedback) or in blockwise (or feedforward) fashion. In continuous mode the estimate is repeatedly updated after each newly received sample, whereas burst-mode estimators are suitable for operation over finite blocks of observed samples. Most practical recovery schemes use a feedback loop composed of a timing error detector followed by a loop filter to control the sampling instants of the incoming signal [74, 126]. Such configuration derives from Phase-Locked Loop (PLL)-based techniques traditionally used in analog transmission, and exhibits similar operational issues. One of such issues is the inherent trade-off between two antagonistic goals: on one hand, the ability to cope with quick timing offset variations requires a sufficiently large loop filter bandwidth; on the other hand, the jitter of the offset estimates

[10] As in the case of the reference channel response of Table 2.5, this channel impulse response has been kindly provided by the Digicomm research group of the DEIS Department at the University of Bologna. In this case, the channel response has been obtained using rural terrain data, which is typically associated to less multipath, and hence, less ISI.

Table 2.6: Time-domain impulse response of the channel corresponding to weak ISI conditions. The assumed symbol rate is $1/T = 3.84$ Mbaud.

Tap Nr.	τ_k (μs)	τ_k/T	Power (dBm)	ρ_k	ϕ_k (rad)
1	0	0	-91.949	1	0
2	0.1953	3/4	-106.329	0.1910	0.2315
3	0.2604	4/4	-110.129	0.1233	-2.7677
4	0.8464	13/4	-112.519	0.0936	-2.5826
5	1.1719	18/4	-110.179	0.1226	-1.4369
6	1.9531	30/4	-112.519	0.0936	-0.5713
7	2.8646	44/4	-112.519	0.0936	-0.1631

grows with the loop bandwidth. In practice, a two step approach is often adopted: first, a coarse timing recovery phase is performed using a large loop bandwidth in order to rapidly obtain a first timing estimate, after which a smaller loop bandwidth is used to provide a steady-state timing with low jitter. Feedforward schemes exhibit a similar trade-off between loop bandwidth and estimate error variance, but, in contrast to feedback schemes, they avoid the slow start-up phase. This fact makes them attractive for application under certain conditions, especially for burst-mode operation [106, 121, 128, 145, 165]. As for the sampling rate of the input signal, symbol timing recovery can be performed using the oversampled output of the matched filter [74, 106, 128, 145, 165], or the baud rate samples at the output of the matched filter [93, 126]. According to the knowledge of the transmitted data, we can distinguish among Data-aided (DA) timing recovery (known pilot symbols), Decision-directed (DD) timing recovery (symbol decisions at the receiver play the role of known symbols) [126], and Non-data-aided (NDA) mode (no knowledge is assumed) [106, 124, 125, 128, 145, 165].

The problem of timing acquisition in THP systems has been hardly covered in the open literature. In coincidence with the recent standardization of the 10GBASE-T technology [90], the first works on the topic are being published. To the best of the author's knowledge, the two contributions by Chien et *et al.* [36, 38] constitute the only references on the topic. Whereas in [36] the authors focus on the problem of joint timing recovery of four interdependent data streams[11], in [38] the authors present a DD baud rate timing recovery scheme particularly designed for SISO THP transmission. The algorithm is based on the classical Mueller and Müller (M&M) timing detector applied to the effective THP data sequence, and it apparently outperforms by far other timing recovery schemes not specifically derived for THP.

O&M timing recovery for THP signals In order to illustrate the difficulties that classical timing acquisition schemes face when applied to THP signals we consider here the Oerder and Meyr (O&M) scheme [128] applied to a SISO ZF-THP system. Figure 2.27

[11]The 10GBASE-T standard [90] considers, among other cablings, the use of four-pair copper lines, which may interfere with each other.

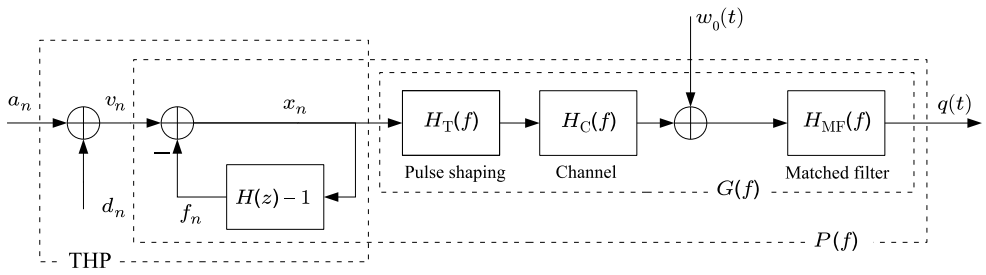


Figure 2.27: Linear signal model considered for timing acquisition.

shows the considered signal model. The O&M scheme operates on the oversampled output of the matched filter $q(t = kT_{\rm s})$, where $T_{\rm s}$ is the oversampling rate. $G(f)$ is the cascade of transmit, channel and matched filters, i.e., $G(f) = H_{\rm T}(f)H_{\rm C}(f)H_{\rm MF}(f)$ [recall (2.36)]. Hence, the output of the matched filter can be written as

$$q(t) = \sum_n x_n g(t - nT) + w(t) \tag{2.42}$$

where $g(t)$ is the time domain impulse response corresponding to $G(f)$, and $w(t)$ represents the colored Gaussian noise component at the output of the matched filter. Now, considering the relationship between v_n and x_n given by (2.8), i.e., $v_n = \sum_{k=0}^{p-1} h_k x_{n-k}$, let us define $h_n^{(-1)}$ as the discrete-time impulse response of the inverse of the channel, which allows to write $x_n = \sum_k h_k^{(-1)} v_{n-k}$. Alternatively, $h_n^{(-1)}$ can be considered as the inverse Z-transform of $1/H(z)$. Now, using $h_n^{(-1)}$ it is possible to rewrite the ISI-free signal (2.42) in terms of v_n as follows

$$\begin{aligned} q(t) &= \sum_n \left(\sum_m v_m h_{n-m}^{(-1)} \right) g(t - nT) + w(t) \\ &= \sum_m v_m \left(\sum_n h_{n-m}^{(-1)} g(t - nT) \right) + w(t) \\ &= \sum_m v_m \left(\sum_k h_k^{(-1)} g(t - kT - mT) \right) + w(t) \qquad (2.43) \\ &= \sum_m v_m p(t - mT) + w(t) \qquad (2.44) \end{aligned}$$

where in (2.43) we have used the change of index $k = n - m$. The continuous-time pulse $p(t) \doteq \sum_k h_k^{(-1)} g(t - kT)$ represents the equivalent impulse response of the cascade of the (linear) precoding operation, transmit filter, channel response, and matched filter, as indicated in Figure 2.27 by its frequency response $P(f)$.

The O&M scheme extracts the timing estimate $\hat{\epsilon}$ in a feedforward, blockwise manner as follows

$$\hat{\epsilon}_m = -\frac{1}{2\pi} \arg \left\{ \sum_{k=mN_{\rm s}L}^{(m+1)N_{\rm s}L-1} |q(kT_{\rm s})|^2 \exp\left(-j\frac{2\pi k}{N_{\rm s}} \right) \right\} \tag{2.45}$$

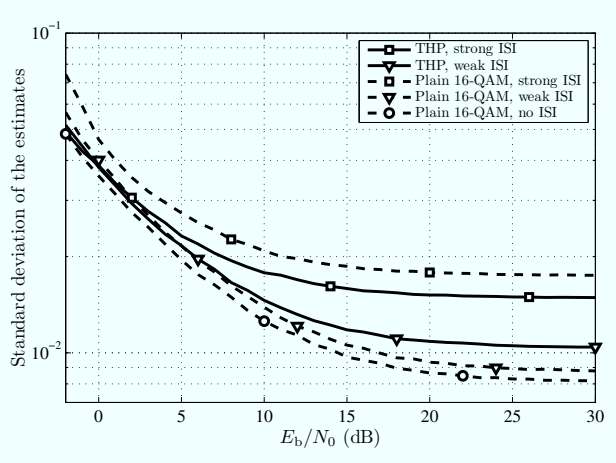


Figure 2.28: Standard deviation of the O&M estimates ($L = 100$, $N_s = 4$) under 16-QAM THP transmission through channels introducing different amounts of ISI.

with m the block index, L the block length in symbol intervals, N_s the oversampling factor, and $T_s = T/N_s$ the oversampling interval ($N_s \geq 2$). The estimate $\hat{\epsilon}_m$ is used as the sampling offset for all the symbols within the observation block. The variance of the O&M estimate is in general inversely proportional to the estimation block length L, and can be seen to be contributed by three components [128]

$$\mathrm{Var}\{\hat{\epsilon}\} \doteq \sigma^2_{\mathrm{O\&M}} = \sigma^2_{n\times n} + \sigma^2_{s\times n} + \sigma^2_{s\times s}$$

of which the first term $\sigma^2_{n\times n}$ is due to noise self-interaction, and dominates in the small SNR region; the term $\sigma^2_{s\times n}$ is due to the interaction between the signal and noise components, and dominates at intermediate SNRs; and the term $\sigma^2_{s\times s}$ is due to self-interaction of the signal component, and dominates in the high SNR region. The last term is independent of the SNR, and thus represents a performance floor at high SNRs.

In general, the variance of the O&M estimate, and especially its ($s \times s$) contribution, depends on the covariance of the symbols and the overall impulse response of the transmission chain [128]. In the case of THP transmission, the effective data sequence v_n is correlated through the equivalent transfer function $H(z)$, see (2.8). But since this transfer function depends on the overall impulse response $G(f)$, see Appendix 2.A, the contributions of the symbol covariance and the channel response are not fully separable.

Figure 2.28 displays the influence of the amount of ISI introduced by the channel in the performance achieved by O&M timing recovery in terms of standard deviation of the estimates. Three cases are shown corresponding to the three levels of ISI (strong, weak, nonexistent) of Figure 2.26 both for THP and plain 16-QAM transmission (in the ISI-free case, THP reduces to plain 16-QAM). The cases of plain 16-QAM transmission with ISI correspond to the case where sampling would be performed prior to (baud-rate) equalization. The curves were obtained through simulation with $L = 100$ symbols and $N_s = 4$. The bias is negligible for SNRs above 0 dB for THP transmission, but not for plain 16-QAM transmission under ISI conditions. Figure 2.28 shows that the standard deviation at high SNRs for the cases with ISI lies above that of the ISI-free case. Such

performance gap reveals a direct influence of the amount of ISI in the $(s \times s)$ contribution to the timing jitter. Surprisingly, a comparison between the THP and plain 16-QAM cases under equal ISI conditions shows that the performance of O&M under weak ISI is worse for THP than for plain 16-QAM, whereas under strong ISI it is better for THP than for plain 16-QAM. A variance analysis similar to that performed in [128] for the plain QAM case would be needed to fully explain and quantify the observed gaps, although the mathematical developments would be rather cumbersome. Although it may seem from Figure 2.28 that the performance floor increases with the amount of ISI, this rule does not hold in general.

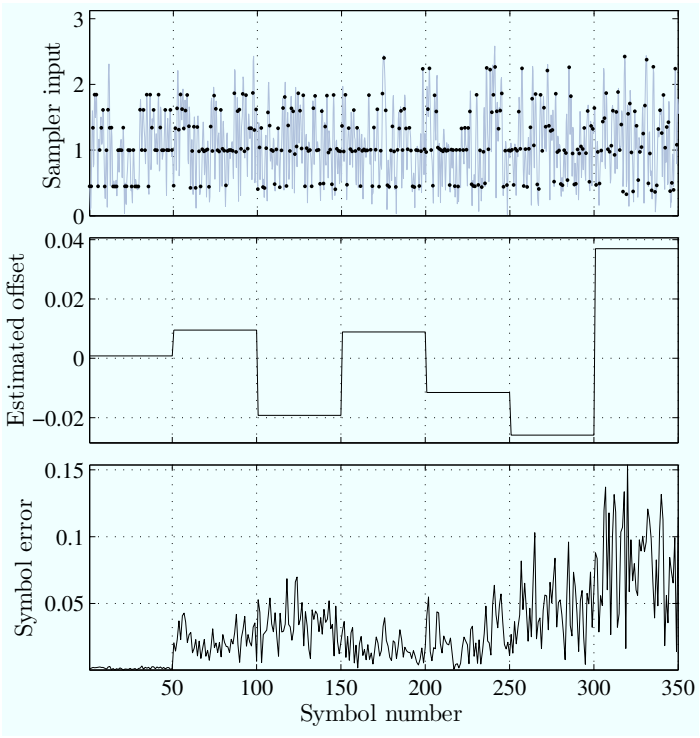


Figure 2.29: Time series of a 16-QAM THP received signal under strong ISI conditions and O&M timing recovery ($L = 50$ symbols, $N_\mathrm{s} = 4$). Shown are (a) the magnitude of the oversampled received signal $q(kT_\mathrm{s})$ (gray), and the magnitude of the baud rate samples (black dots), (b) the estimated offsets for each symbol block, and (c) the magnitude of the error between the true and the sampled symbols. No noise was present in the simulation, and the true offset was set to $\epsilon = 0$.

One of the reasons that may explain the worse performance of O&M under THP transmission is that the signal at the matched filter output (which amounts to a noisy version of v_k modulated by a raised cosine filter) does not seem to be cyclostationary.[12] Such non-cyclostationarity would be in agreement with the analysis performed in [158], where it is concluded that the precoded sequence x_k is only cyclostationary for some particular channels $H(z)$ with integer coefficients. In our case, the coefficients of $H(z)$ are not (complex) integers in general, so that the x_k, and thus v_k, can be considered non-

[12] A signal is said to be cyclostationary when its statistical characteristics vary periodically with time [75].

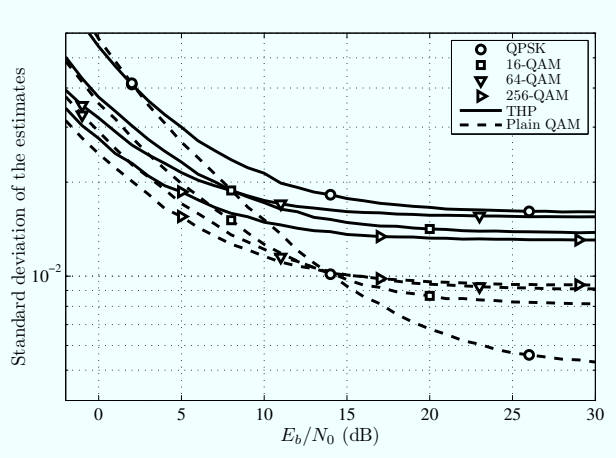


Figure 2.30: Standard deviation of the O&M estimates under THP transmission with different QAM constellations.

cyclostationary. This fact is important in terms of timing recovery since it could explain the sudden increase in the error of the timing estimates for some O&M estimation blocks. One such event can be observed in Figure 2.29, which represents a snapshot of the performance of the O&M estimator during several estimation blocks (of length $L = 50$). The largest offset error is obtained for the last block of symbols (about $0.04T$), which happens to exhibit a large number of effective symbols with large magnitude. In contrast, blocks with fewer such symbols provide more accurate offset estimates. In our simulations we have observed that the occasional appearance of a symbol with large amplitude can have a negative impact in the accuracy of the O&M timing estimate for its estimation block. However, we acknowledge that a complete statistical analysis would be needed in order to completely explain the behavior of O&M timing recovery under THP transmission. This task is left here as a topic for future research.

Figure 2.30 shows the performance in terms of standard deviation of the O&M estimates for different TH-precoded QAM modulations under strong ISI conditions. For comparison, the performance for plain ISI-free QAM transmission is also shown. The curves were obtained through simulation with $L = 100$ and $N_{\mathrm{s}} = 4$. The bias turns out to be negligible for SNRs above 0 dB in all tested cases. The high SNR gap of the O&M estimates for THP with respect to plain QAM can be observed for all constellations, although the gap decreases as the QAM constellation size increases.

Figure 2.31 displays the dependence of the O&M estimator with respect to the length of the estimation block, L. It can be observed that the performance loss of THP with respect to plain QAM is negligible at low SNRs, but not at high SNRs, where the standard deviation gap varies between a factor of 1.5 and 2.

The recent M&M-based timing recovery scheme presented in [38] successfully exploits the correlation in the effective data sequence, achieving a considerable improvement over other state-of-the art schemes, and thus appears to be the best current recovery scheme suitable for implementation in THP systems.

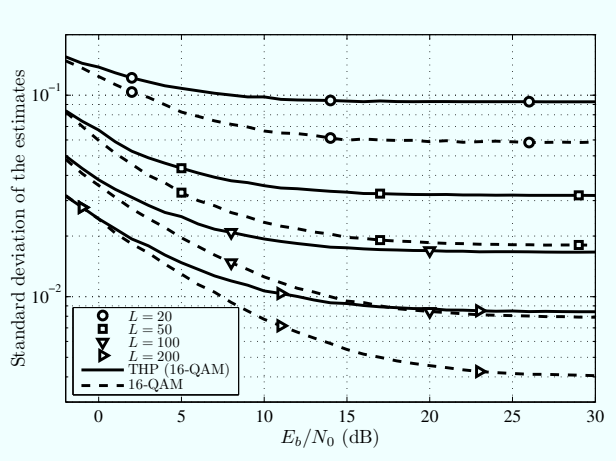


Figure 2.31: Standard deviation of the O&M estimates under 16-QAM THP transmission with different estimation block lengths.

Chapter 3

Application of Tomlinson-Harashima Precoding to Multibeam Satellite Communications

Contents

3.1 Introduction

Whereas in the previous chapter we have considered point-to-point transmissions affected by time dispersion, in this chapter we will focus on the case of multibeam satellite transmission affected by Co-Channel Interference (CCI). The combination of multibeam coverage with frequency reuse has proven a successful approach to increase the throughput of current GEO satellite communication systems. As examples we can mention the near 200 beams implemented by the Inmarsat 4 satellites [78], or the more than 500 beams provided by the TerreStar-1 satellite [156].

Figure 3.1 represents the main elements in the forward link of a GEO multibeam satellite system. The total coverage area is divided in smaller cells, which will be referred to as *spotbeams*. Each spotbeam is illuminated by a different beam coming from the satellite antenna, which is composed of typically between 100 and 200 radiating elements [44]. The beam pattern of the *service link* (satellite to terminals) is achieved by the combined action of the radiating elements in the satellite antenna and a beamforming stage which properly combines the signals to be transmitted before being fed to the radiating elements. Although the power radiated by each beam is maximum on its spotbeam, part of it spreads over other beams due to the sidelobes of the radiation pattern.

The data content towards the beams is provided by several *gateways* through the so-called *feeder links* (gateway to satellite). The number of gateways is in general smaller than the number of beams. The transmitted signals may vary from beam to beam due to, e.g., service differentiation among countries. Hence, a gateway must feed the satellite with several independent signals, which requires the implementation of some kind of multiplexing (typically in frequency and polarization). As a consequence of the service differentiation among beams, the signals from adjacent beams may become an unwanted source of disturbance (CCI).

Multibeam coverage benefits from the smaller size of the spotbeam, which allows to achieve higher radiation gains. In contrast, this may induce a higher CCI in adjacent beams. For this reason, most practical systems employ a regular frequency reuse pattern with N_{fr} frequencies, which prevents two adjacent beams from reusing the same frequency.[1] In this case, the *frequency reuse factor* N_{fr} (also known as *cluster size*) is of the form $N_{\text{fr}} = m^2 + mn + n^2$, with $m, n = 0, 1, 2, \ldots$ As a rule of thumb, the dependence of the throughput, C, of a multibeam system with the beam radius, R, and N_{fr} is of the form [113]

$$C \propto N_{\text{fr}}^{-1} \cdot R^{-2}.$$

In terms of throughput, it is therefore desirable that N_{fr} and R be as small as possible. However, reducing N_{fr} or R will cause an undesirable increase in the CCI level, so, in practice, a trade-off must be met between these two conflicting goals.

In this chapter we propose the application of Tomlinson-Harashima Precoding (THP) to precompensate the CCI existing in the forward link of a multibeam satellite transmission system, in which the satellite antenna is to provide different data streams to non-collaborative receivers located in different spotbeams. The same principles presented in

[1] In Figure 3.1 colors are used to represent the different frequencies used in each beam.

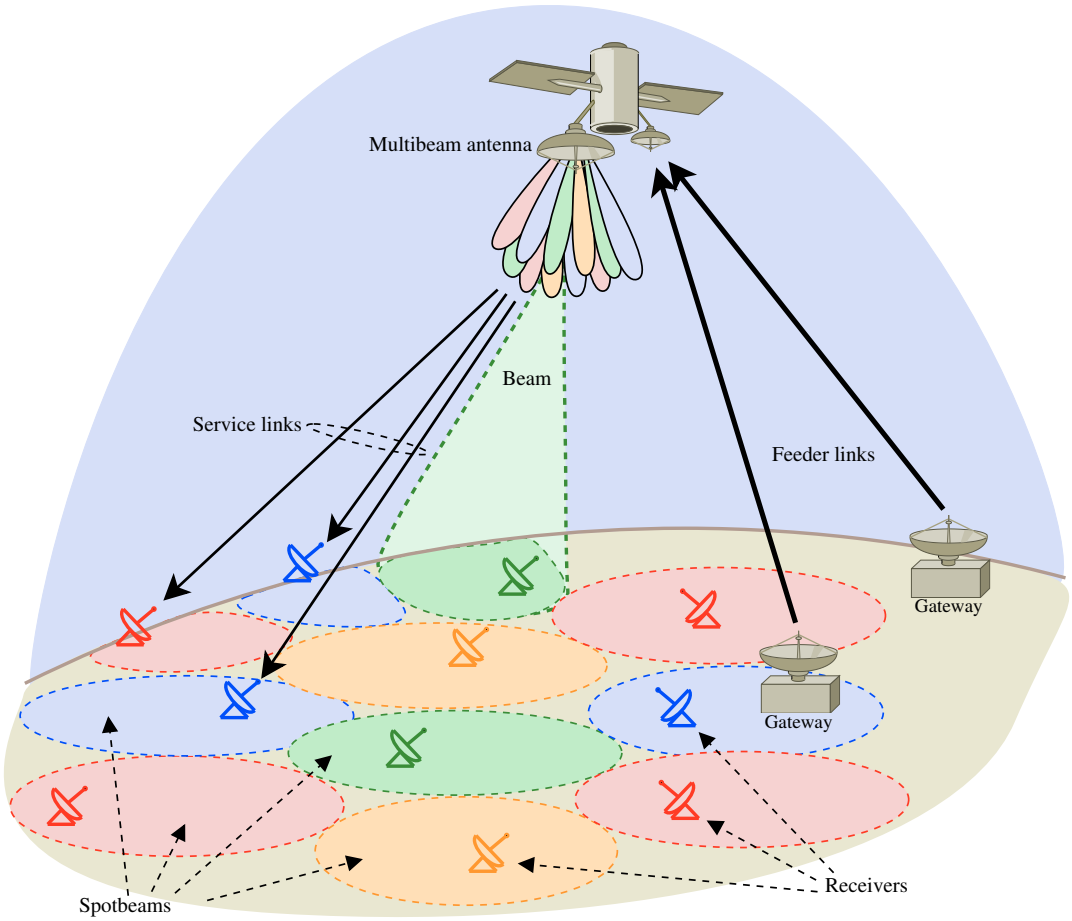


Figure 3.1: Simplified representation of the forward link of a multibeam satellite transmission system.

Chapter 2 apply in the multibeam scenario, with the CCI playing the role of the ISI, as we will see. The fact that the CCI is precompensated by the THP transmitter offers potential throughput increases, since the frequency reuse factor (or, equivalently, the spotbeam radius) could be reduced without sacrificing system performance. To the best of the author's knowledge, the application of THP to the forward link of a multibeam satellite system is a novelty in the related literature. A similar approach, but restricted to linear precoding techniques, was explored in [43, 67] with promising results. Furthermore, the problem of CCI cancellation in the reverse link of a multibeam system has been dealt with in [120, 123] using multiuser detection and decoding techniques.

Chapter outline The chapter is structured as follows. In Section 3.2 we provide the basic principles of THP applied to multibeam systems. The particular realistic multibeam system used for reference is described in Section 3.3, and several application scenarios are selected for evaluation. In Section 3.4 we propose several possible THP implementation schemes suitable to be applied to the described scenario. Such systems are evaluated in Section 3.5 by means of numerical simulations. This chapter is based on the work

presented in [7].

3.2 Theoretical Background: THP for Multibeam Systems

In the system we consider, the satellite antenna generates a certain number of beams, each one carrying an independent data stream towards its spotbeam, and there is only one receiving terminal per spotbeam thanks, for example, to the use of some orthogonal multiplexing scheme.[2] The receivers are assumed to have a single antenna and to be noncollaborative. Therefore, from each receiver's point of view, this transmission system is composed of a single transmitter (the satellite) sending a number of independent data streams (one per beam) which are received with different attenuations at the receiver. Strictly speaking, such transmission scheme can be regarded as a Multiple Input Single Output (MISO) system, but not a MIMO system.[3] MISO or MIMO models are typically applied to transmission schemes using *multiple antennas*. Although multibeam antennas are in fact composed of many radiating elements (or *feeds*), in terms of signal modeling we are more interested in the multiplicity of beams which those feeds conform, so we will speak of *beams* rather than *antennas* in our MISO model. Figure 3.2 represents the general concept of MISO and MIMO transmission, with a multiple transmitter and several receivers which may collaborate (MIMO) or not (MISO). Note that the effect of the channel is to attenuate and linearly combine the multiple transmitted signals into each receiver. In the remaining of this section we will refer to MISO transmission only, although most concepts apply to MIMO systems as well.

3.2.1 Signal Model

We assume a quasi-static flat-fading signal model (no ISI caused by time dispersion) which will be justified in Section 3.3 when we describe the particular application scenario. When considering MISO and MIMO systems it is common to use vector notation in order to group all signals corresponding to the same time instant. With this, the flat-fading discrete-time signal model of a general MISO system affected by AWGN is given by

$$\boldsymbol{y}_n = \boldsymbol{H}\boldsymbol{a}_n + \boldsymbol{w}_n \tag{3.1}$$

where $\boldsymbol{a}_n$ is a $N_{\mathrm{Tx}} \times 1$ vector (N_{Tx} is the number of beams) containing the transmitted symbols at time n, which we assume to be drawn from a QAM constellation; $\boldsymbol{y}_n$ is a $N_{\mathrm{Rx}} \times 1$ vector which groups the received samples by all receivers at time n (N_{Rx} receivers); and $\boldsymbol{w}_n$ is a $N_{\mathrm{Rx}} \times 1$ vector with the noise contribution at each receiver. The symbol and noise

[2] In the context of Information Theory, this transmission scheme is referred to as a *broadcast channel*. However, we will not use the term *broadcast* with this meaning throughout this chapter to avoid confusion with the concept of satellite broadcast service (e.g., TV and radio).

[3] Since the receivers do not cooperate, one cannot speak of MIMO transmission, where joint processing of the received signals is possible.

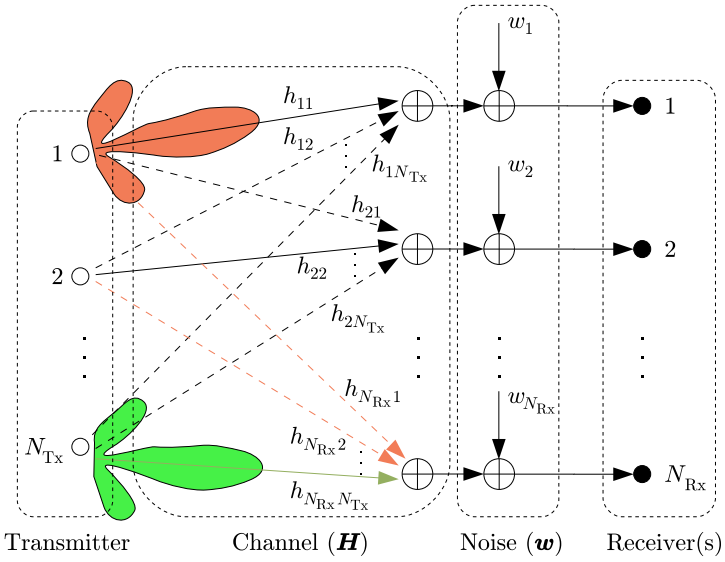


Figure 3.2: Representation of MIMO and MISO transmission. MIMO: collaborative receivers, or single receiver with multiple antennas. MISO: non collaborative single antenna receivers.

sequences have zero mean and are white in the time domain and also in the space domain (i.e., $\mathrm{E}\left\{\boldsymbol{a}\boldsymbol{a}^{\mathrm{H}}\right\} = \sigma_a^2\boldsymbol{I}$ and $\mathrm{E}\left\{\boldsymbol{w}\boldsymbol{w}^{\mathrm{H}}\right\} = \sigma_w^2\boldsymbol{I}$). The QAM constellation is assumed to be normalized in energy ($\sigma_a^2 = 1$).

The flat-fading channel, which comprises the effects of the beam radiation patterns and the propagation, is represented by the $N_{\mathrm{Rx}} \times N_{\mathrm{Tx}}$ matrix $\boldsymbol{H}$ whose elements h_{ij} contain the complex attenuations from the j-th beam to the i-th receiving terminal, with $i = 1, 2, \ldots, N_{\mathrm{Rx}}$, and $j = 1, 2, \ldots, N_{\mathrm{Tx}}$.[4] For example, consider the i-th element of $\boldsymbol{y}_n$, which is given by

$$[\boldsymbol{y}_n]_i = \sum_{j=1}^{N_{\mathrm{Tx}}} h_{ij}\,[\boldsymbol{a}_n]_j + [\boldsymbol{w}_n]_i\,. \tag{3.2}$$

It is clear that the effect of the channel is to *mix* the signals coming from all beams into each receiving terminal.

In our case, the previous general model can be slightly simplified since we consider that there is one receiver per beam. Letting N_{B} be the number of beams, we have that $N_{\mathrm{Tx}} = N_{\mathrm{Rx}} = N_{\mathrm{B}}$, and hence $\boldsymbol{x}$ and $\boldsymbol{y}$ are $N_{\mathrm{B}} \times 1$ vectors, and $\boldsymbol{H}$ is an $N_{\mathrm{B}} \times N_{\mathrm{B}}$ square matrix.

In MISO systems the precoding operation can be considered as a spatial-domain version of time-domain preequalization already introduced in Section 2.2.1.1. By "space" we are referring to the beams. Whereas in time-domain preequalization a time-domain

[4] If the flat-fading assumption were not valid, then one should consider a matrix transfer function of the form $\boldsymbol{H}(z) = \sum_k \boldsymbol{H}_k \cdot z^{-k}$, whose entries are scalar transfer functions, i.e., $[\boldsymbol{H}]_{ij}(z) = \sum_k [\boldsymbol{H}_k]_{ij} z^{-k}$. Precoding schemes designed to counteract both time and spatial dispersion exist in the literature, e.g. [96], although we restrict our analysis to the flat-fading model.

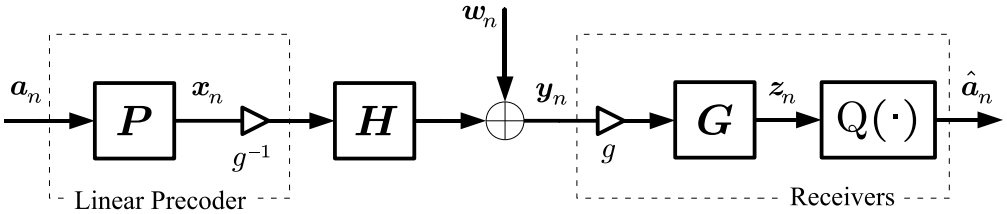


Figure 3.3: Generic block diagram of linear precoding for multiantenna transmission and noncollaborative receivers.

filter is used to somehow invert the effect of the time-dispersive channel, in MISO precoding a "spatial filter" is used to precompensate for the effect of a "space-dispersive" channel. This analogy allows to introduce the application of THP to MISO transmission starting from the more natural concept of Linear Precoding (LP), as it was done for SISO time-dispersive channels in Section 2.2.1.1.

In order to make explicit the similarity with the time-dispersive scenario, let $[\boldsymbol{a}_n]_i$ be the sequence intended for the i-th receiver. It is possible to rewrite (3.2) as

$$[\boldsymbol{y}_n]_i = h_{ii}\,[\boldsymbol{a}_n]_i + \sum_{\substack{j=1 \\ j\neq i}}^{N_\mathrm{B}} h_{ij}\,[\boldsymbol{a}_n]_j + [\boldsymbol{w}_n]_i \tag{3.3}$$

where the desired message appears scaled by the complex gain h_{ii}, and distorted by two additive terms: the first one gathers the signals intended for the rest of receivers, and the second one is AWGN. The first term can be seen as undesired *spatial* ISI, hence the analogy with the traditional time-dispersive scenario. Within this context, we will refer to the spatial ISI as the CCI.

Since the receivers cannot perform joint processing, some alternative transmitter-based *equalization* procedure must be designed to allow each receiver to recover its corresponding message. Different solutions exist for this challenge, which share analogies with the transmitter-based solutions covered in Section 2.2.1.1 in the context of equalization for time-dispersive channels. Next we review two such solutions: linear transmitter precoding and THP for MIMO systems.[5]

3.2.2 Linear Transmitter Precoding

Linear Precoding (LP) is a linear scheme to compensate for the interfering effect of the channel.[6] Since it is a transmitter-based technique, the channel matrix $\boldsymbol{H}$ must be known at the transmitter. The block diagram of LP is shown in Figure 3.3. The original symbols

[5]These techniques correspond respectively to linear preequalization and THP for SISO ISI channels.

[6]In the context of Information Theory, linear precoding is commonly referred to as *linear beamforming*. However, in this chapter we will reserve the term *beamforming* to denote the signal processing stage which, in combination with the satellite antenna, is needed to achieve the desired multibeam radiation pattern. In Appendix 3.C we provide an overview of beamforming techniques suitable for multibeam satellite systems.

are fed to a linear spatial filter (matrix $\boldsymbol{P}$, which in our case is $N_{\mathrm{B}} \times N_{\mathrm{B}}$) whose role is to precompensate the effect of the channel (matrix $\boldsymbol{H}$). The positive scalar factor g^{-1} at the transmitter is applied in order to meet the transmit power limitation (represented by E_{TX}) present in a practical transmitter.[7] Its effect is undone at the receiver by applying the inverse scalar factor g. We consider here two approaches to LP (see [99] for a complete derivation):

- **Zero-forcing LP (ZF-LP)**. The transmit filter and scaling constant are given by

$$\boldsymbol{P}_{\text{ZF-LP}} = \boldsymbol{H}^{+} \doteq \boldsymbol{H}^{\mathrm{H}}(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}})^{-1} \tag{3.4}$$

$$g_{\text{ZF-LP}} = \sqrt{\frac{\mathrm{Tr}\left((\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}})^{-1}\sigma_a^2\right)}{E_{\mathrm{TX}}}} \tag{3.5}$$

  where $\boldsymbol{H}^{+}$ is the so-called *Moore-Penrose pseudoinverse* of $\boldsymbol{H}$. Note that if $\boldsymbol{H}$ is invertible, then $\boldsymbol{H}^{+}$ reduces to $\boldsymbol{H}^{-1}$, which clearly reveals the ZF nature of this approach.
- **Minimum mean square error LP (MMSE-LP)**. Under this approach, the noise variance σ_w^2 (which, in addition to $\boldsymbol{H}$, must be known by the transmitter) is taken into account so as to improve the performance of the ZF approach in the low SNR region. In this case the expressions of the transmit filter and scaling constant are given by:

$$\boldsymbol{P}_{\text{MMSE-LP}} = (\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \xi\boldsymbol{I})^{-1}\boldsymbol{H}^{\mathrm{H}} \tag{3.6}$$

$$g_{\text{MMSE-LP}} = \sqrt{\frac{\mathrm{Tr}\left((\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \xi\boldsymbol{I})^{-2}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\sigma_a^2\right)}{E_{\mathrm{TX}}}} \tag{3.7}$$

  where $\xi \doteq N_{\mathrm{B}}\sigma_w^2/E_{\mathrm{TX}}$. As expected, the MMSE and ZF solutions coincide asymptotically as the SNR grows (i.e., when $\sigma_w^2 \to 0$).

In general, the main drawback of LP is the increase in average energy of the precoded signal with respect to the original signal due to the channel inversion. This effect becomes stronger as the level of CCI increases. Considering the satellite scenario, where transmit power is a precious resource, this effect should be avoided. Not surprisingly, such drawback is shared by the ZF approach to preequalization for SISO ISI channels, as we have seen in Section 2.2.1.1.

3.2.3 Nonlinear Transmitter Precoding: THP

This latter problem of LP can be avoided if nonlinear precoding is used at the transmitter. Similarly to the SISO case of Section 2.2.1, the main advantage of nonlinear precoding (and in particular, THP) is that the energy of the precoded signal is kept almost at the same level as that of the original signal. Again, perfect knowledge of the channel matrix $\boldsymbol{H}$ is assumed (full Channel State Information knowledge). We refer to Appendix 3.E for details on the channel estimation procedure.

[7] Although its effect could well be integrated within the transmit filter $\boldsymbol{P}$, we have intentionally represented both operations separately to make explicit the transmit power requirements.

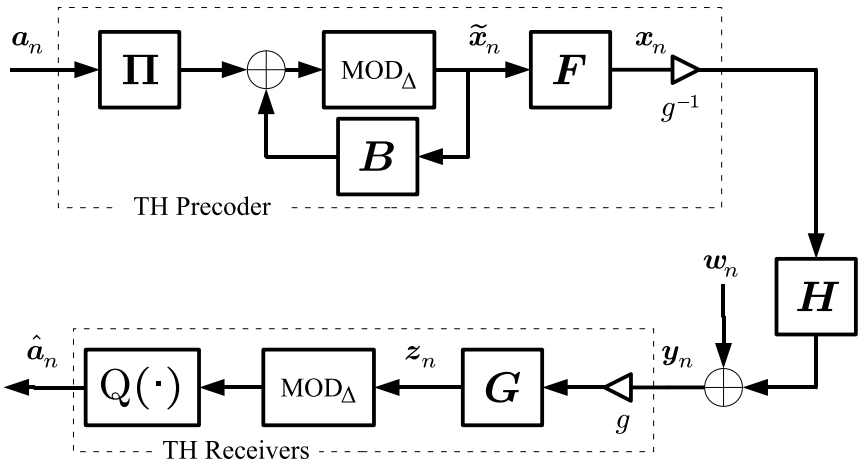


Figure 3.4: Generic block diagram of THP applied to a multiantenna transmission and noncollaborative receivers.

Figure 3.4 depicts the generic block diagram of THP applied to a MIMO system. A block-by-block description follows:

- **Signal reordering**. First, the rows of the original QAM vector $\boldsymbol{a}_n$ are reordered by means of the $N_\mathrm{B} \times N_\mathrm{B}$ permutation matrix $\boldsymbol{\Pi}$. This operation, which is a novelty with respect to LP, selects the order in which the different components of $\boldsymbol{a}_n$ will be precoded, since the overall performance depends in general on the precoding order. Since the reordering is transparent to the receivers, it can be seen as a degree of freedom available in the precoder design.
- **Nonlinear feedback loop**: The reordered vector $\boldsymbol{\Pi a}_n$ goes through a feedback loop with a modulo operation in the forward branch, and a strictly lower triangular $N_\mathrm{B} \times N_\mathrm{B}$ matrix $\boldsymbol{B}$ (i.e., with zeros in the main diagonal) in the backward branch. The matrix $\boldsymbol{B}$ is obtained from the channel matrix $\boldsymbol{H}$. The vector modulo operation is defined as the elementwise application of the same MOD$_\Delta(\cdot)$ operation as defined in Section 2.2.1.2 for THP in ISI channels. Observing the structure of $\boldsymbol{B}$ we can recognize the similarity with the nonlinear feedback loop in THP for ISI channels (cf. Figure 2.2 in Section 2.2), where the negative feedback filter $H(z) - 1$ presubtracts the ISI caused by previous symbols. Here, the role of the feedback loop is to presubtract the (spatial) ISI introduced by previously precoded symbols, which play the role of *past* symbols. Hence, denoting the output of the modulo operation by $\tilde{\boldsymbol{x}}_n$, the precoding operation is to be performed in sequential order as follows (the time index n is dropped in $\boldsymbol{a}_n$ and $\tilde{\boldsymbol{x}}_n$ for convenience, and b_{ij}

represents the i, j element of $\boldsymbol{B}$):

$$[\tilde{\boldsymbol{x}}]_1 = \mathrm{MOD}_\Delta\left([\boldsymbol{\Pi a}]_1\right) = [\boldsymbol{\Pi a}]_1 \tag{3.8a}$$

$$[\tilde{\boldsymbol{x}}]_2 = \mathrm{MOD}_\Delta\left([\boldsymbol{\Pi a}]_2 + b_{21}\,[\tilde{\boldsymbol{x}}]_1\right) \tag{3.8b}$$

$$\vdots$$

$$[\tilde{\boldsymbol{x}}]_{N_\mathrm{B}} = \mathrm{MOD}_\Delta\left([\boldsymbol{\Pi a}]_{N_\mathrm{B}} + \sum_{j=1}^{N_\mathrm{B}-1} b_{N_\mathrm{B}j}\,[\tilde{\boldsymbol{x}}]_j\right). \tag{3.8c}$$

Note that the first precoded symbol does not receive any interference, whereas the second one receives only the interference of the first one, and so on.

- **Feedforward filter**. The feedforward matrix $\boldsymbol{F}$ ($N_\mathrm{B} \times N_\mathrm{B}$) is applied to the output of the modulo yielding the precoded signal $\boldsymbol{x}$. $\boldsymbol{F}$ is obtained from $\boldsymbol{H}$, and, in general, is a full matrix, i.e., without zeros.
- **Transmit power limitation**. Similarly to LP, a positive scalar factor g^{-1} is applied to the transmitted signal to comply with the transmit power requirements. The first step at the receiver is to rescale the received signal with the corresponding factor g.
- **Gain control**. The received signal, affected by channel and noise, is first rescaled (g) and then applied an extra gain which is represented by the $N_\mathrm{B} \times N_\mathrm{B}$ *diagonal matrix* $\boldsymbol{G}$. Note that although all scalar factors $[\boldsymbol{G}]_{i,i}$ are grouped into one matrix for mathematical convenience, they correspond to gain control stages which are separately performed by each non-cooperative receiver.[8] If $\boldsymbol{G}$ does not have a constant diagonal, then the SNR observed by each receiver will not be the same. Some THP solutions in the literature exploit this feature to deliberately provide different SNRs to different users according to some design criterion or quality of service requirements [140].
- **Modulo reduction and decision**. Finally the signal undergoes the same modulo reduction as in the precoder, and then the estimate $\hat{\boldsymbol{a}}$ of the original signal is computed by the decision device $\mathrm{Q}(\cdot)$.

MIMO THP is affected by the same *precoding loss* as SISO THP for time-dispersive channels. Recall that the precoding loss depends on the modulation only, and for square M-QAM constellations it evaluates to $\gamma_p = M/(M-1)$ (see Section 2.2.1.2). We note that LP can be seen as a degenerate case of a THP system, where $\boldsymbol{\Pi} = \boldsymbol{I}$, and the modulo operations are replaced by shortcircuits. In this case, simple algebra shows that the feedforward matrix $\boldsymbol{P}$ of LP and the matrices $\boldsymbol{B}$, $\boldsymbol{F}$ of THP are related by $\boldsymbol{P} = \boldsymbol{F}(\boldsymbol{I} - \boldsymbol{B})^{-1}$.

The performance of THP depends on the precoding order given by $\boldsymbol{\Pi}$, so this matrix should be carefully chosen. The optimum ordering depends ultimately on the channel matrix $\boldsymbol{H}$ and the particular precoding scheme, and can only be found by exhaustive search among the $N_\mathrm{B}!$ possible orderings. A less demanding alternative is to use the "*best-first*" [97] principle which, though suboptimal, usually achieves a close approximation of the optimal solution with much less computational complexity. In Appendix 3.A we offer details on the "best-first" method.

[8] Note also that the rescaling (g) and gain control ($\boldsymbol{G}$) operations can be performed in one single step. However, we have kept them apart for illustrative reasons.

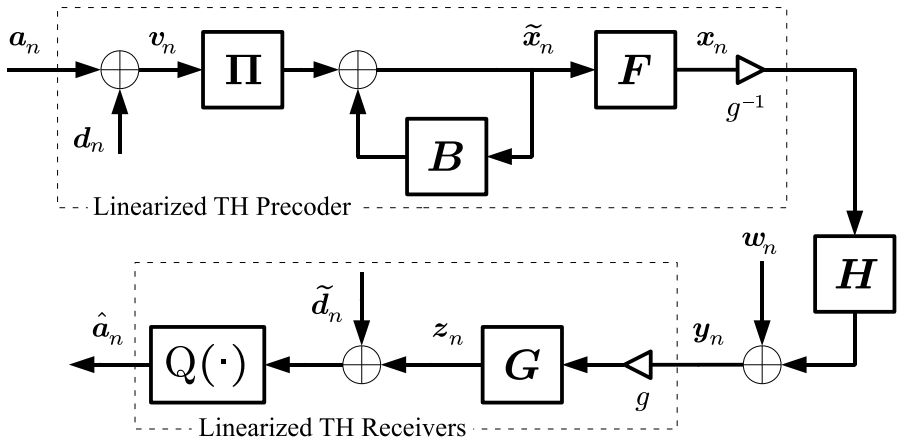


Figure 3.5: Equivalent linearized model of MIMO THP transmission and reception.

Figure 3.5 displays the equivalent linearized block diagram of THP for MIMO transmission. The modulo operations in the THP chain (cf. Figure 3.4) are replaced by additions of sequences $\boldsymbol{d}_n$ and $\tilde{\boldsymbol{d}}_n$ which take values on the set of vectors which are modulo-congruent to zero under $\mathrm{MOD}_{\Delta}(\cdot)$. Then, the sequence $\boldsymbol{v}_n$ is considered the *effective data sequence* which undergoes linear processing at the transmitter. The corresponding noisy version of $\boldsymbol{v}_n$ at the receiver is $\boldsymbol{z}_n$.[9]

There is no unique implementation of MIMO-THP, but a variety of them. The matrices $\boldsymbol{B}$, $\boldsymbol{F}$, $\boldsymbol{G}$ and the scalar g are obtained as functions of some factorization of $\boldsymbol{H}$ and some performance constraints, such as the transmit power limitation. Depending on how $\boldsymbol{H}$ is factored and the performance constraints imposed to the system different implementations of THP are obtained. Some of them will be compared later in this section under a set of application scenarios that we next describe.

Duality of MIMO THP and MIMO DFE In Section 2.2.1.1 we have mentioned the existing duality between THP and DFE in SISO transmission. Such duality still holds in MIMO transmission.[10] In this case the MIMO DFE receiver applies nonlinear *sequential detection*: once the symbol of a given stream has been decided, it is subtracted (*fed back*) from the received sequence prior to deciding the symbol of the next stream, and so on. Such detection procedure is the dual of the precoding operation of (3.8). Similarly to MIMO THP, the detection order has an impact on the overall performance of the transmission scheme.

3.3 Multibeam Downlink Scenario Description

As particular application case, we consider the forward link of a practical GEO satellite system located at 19.2° E, which provides multimedia content over Europe, as proposed

[9]Note the parallelism with the equivalent linearized model of THP for SISO transmission of Figure 2.5.
[10]MIMO DFE is commonly known as *successive interference cancellation* in the literature.

in [79]. As depicted in Figure 3.1, the forward link of the satellite system comprises a number of gateway stations providing the data content, the satellite itself which generates the beams and delivers the data streams, and one receiving station per spotbeam. No ground-based gap-fillers are considered in the system, and the receiver terminals are assumed to be fixed and equipped with a directive antenna able to reject any possible reflected or scattered signal contributions other than the LOS signal. The service area is covered by 96 spotbeams, see Figure 3.7 in page 62. The downlink carrier frequency is 20 GHz (Ka band).

3.3.1 Multibeam coverage

We consider a tapered-aperture satellite antenna which generates beams with one-sided half-power width $\theta_{3\text{dB}} = 0.2°$, and a radiation pattern given by [27, 113]:

$$G(\theta) = G_{\max} \cdot \left(\frac{J_1(u)}{2u} + 36\frac{J_3(u)}{u^3} \right)^2 \tag{3.9}$$

where $u = 2.07123 \cdot \sin\theta / \sin\theta_{3\text{dB}}$, $G_{\max}$ is the maximum beam gain, θ is the angle between the spotbeam center and the receiver location as seen from the satellite, and J_1 and J_3 are the Bessel functions of the first kind of order 1 and 3 respectively. Figure 3.6 displays the beam radiation pattern.

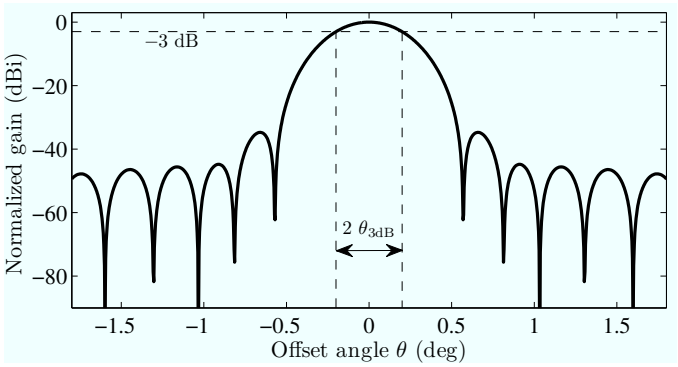


Figure 3.6: Beam radiation pattern as given by (3.9) with one-sided half-power width $\theta_{3\text{dB}} = 0.2°$.

The resulting coverage map is displayed in Figure 3.7 considering the particular case of the frequency reuse plan with $N_{\text{fr}} = 4$ frequencies proposed in [79]. However, we will consider different frequency plans for the particular scenarios used for evaluation, as we will describe below. The data content of each beam is provided by a number of gateway stations possibly located at different geographical positions. Each gateway provides data content to a subset of beams which is disjoint from the subsets fed by other gateways. Such subsets do not necessarily correspond to the clusters of N_{fr} beams as given by the frequency reuse plan, but are formed according to other criteria such as, e.g., geographical proximity. The data streams, which are assumed to be independent among beams, are

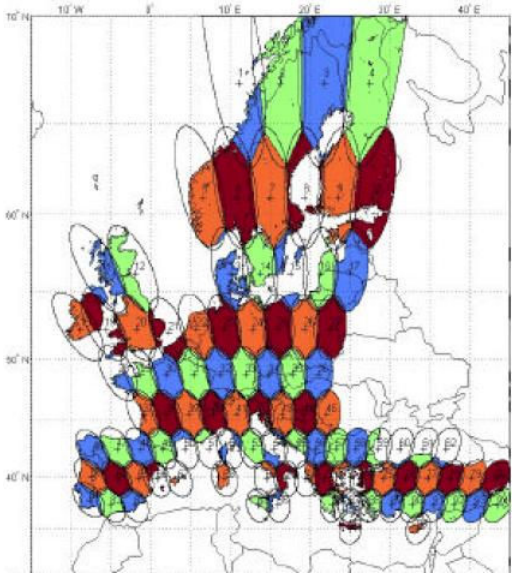


Figure 3.7: Coverage and spotbeam distribution of the considered multibeam system with a sample frequency reuse factor of $N_{\text{fr}} = 4$ frequencies (marked in different colors), as proposed in [79].

multiplexed (e.g. in frequency and polarization) before being sent to the satellite through the feeder link. At the satellite, the data streams from all gateways are demultiplexed and distributed to the appropriate beams.

3.3.2 Satellite system details

Although the original system proposed in [79] considers a satellite with onboard processing and switching capabilities, such assumption is not needed, so satellites with no switching capabilities and with limited or no processing capabilities would be also valid for implementation of the proposed system. When the satellite cannot perform onboard processing we speak of a *bent-pipe architecture* in which the effect of the satellite is transparent. Figure 3.8 displays the most representative blocks of a sample forward link transmission chain, from the gateway to the receiving terminal. The block diagram is not intended to include *all* stages of a real satellite system, but only those needed to justify our signal model. The beamforming blocks represent the linear signal preprocessing stages required to adapt the signals to be transmitted to the satellite antenna, so that the desired multibeam radiation pattern is generated, and each beam carries the desired signals. The beamforming task can be completely performed at the gateway or onboard the satellite, or jointly at both places (for this reason there is a beamforming block at both places). The choice between these alternatives depends on the processing capabilities of the payload onboard. The beamforming stage is transparent from the receiver's point of view, so it will not be considered for the signal model. We refer to Appendix 3.C, where we provide an overview of the different approaches to the beamforming task.

Regarding the rest of blocks, the next assumptions will be made in order to simplify

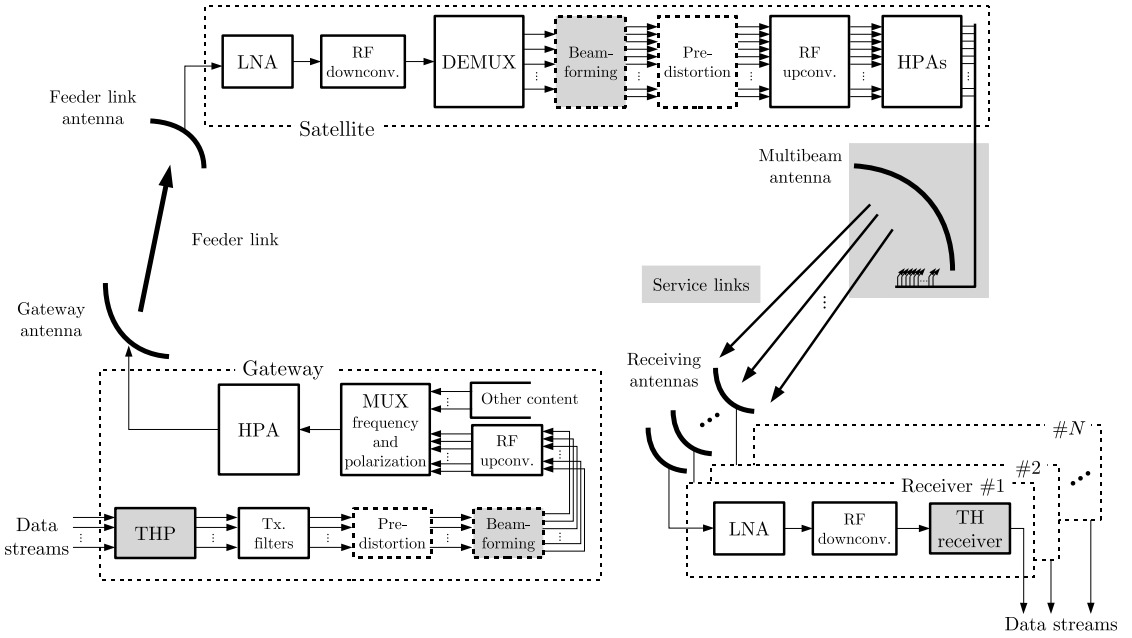


Figure 3.8: Block diagram of a multibeam satellite system (forward link) implementing THP, including gateway, satellite and receivers. Dashed blocks might be implemented either at the gateway or onboard. White blocks are "transparent" under our signal model from the receivers' point of view.

our analysis:

As3.1) **Transparent IF and RF conversions.** All frequency conversions in the chain are ideal, i.e., they do not introduce any degradation in the transmitted signals.

As3.2) **Transparent (de)multiplexing.** The multiplexing at the gateway and the demultiplexing onboard do not alter the transmitted signals.

As3.3) **Perfect timing and carrier recovery.** The receiver is assumed to implement sufficiently accurate timing, phase and frequency acquisition schemes, so that they do not impact negatively the received signal.

As3.4) **Linear HPAs.** The HPAs at the gateway and onboard the satellite are assumed to operate under linear conditions, so that they do not introduce nonlinear distortion.

As3.5) **Perfect CSI knowledge.** The Channel State Information (CSI), i.e., the channel response, is assumed to have a sufficiently slow variation (quasi-static channel) so that some channel estimation procedure can perfectly estimate and track the channel state at any moment. The existence of an ideal return channel is furthermore assumed so that the estimated CSI can be accurately passed back to the transmitter. Perfect CSI knowledge is required for the proper application of THP transmission as presented in Section 3.2.

The linear assumption for the gateway HPA is reasonable since simple countermeasures against nonlinear distortion can be employed in the gateway (e.g. a sufficient back-off [39]) with low impact in the overall system cost. However, assuming a linear HPA onboard seems unrealistic unless proper countermeasures such as predistortion are applied (large back-offs are prohibitive). Since our goal in this chapter is to evaluate the

system performance under ideal conditions, we will consider that such countermeasures are applied. We note, however, that the application of predistortion in multibeam transmission may not be as simple as in the case of SISO transmission seen in Section 2.3. We refer to Appendix 3.D for a initial analysis on this issue.

Similarly, the perfect CSI knowledge assumption is motivated by the need of simplifying the analysis as much as possible. In general, any channel estimation scheme will offer imperfect estimates whose effect in the performance of THP must be properly considered. We refer to Appendix 3.E where we provide an overview of the techniques available in the literature for dealing with imperfect CSI.

Since the receivers are fixed terminals equipped with directive antennas, it can be considered that there is no multipath in the signals received at each terminal. Considering the above assumptions, the narrowband signal model presented in Section 3.2.1 applies under such quasi-static flat-fading conditions, with the (i, j) element of the channel matrix $\boldsymbol{H}$, h_{ij}, representing the channel gain of the j-th beam towards the i-th receiver. In fact, the value of the channel gain h_{ij} is a function of the radiation pattern (given by (3.9) and assumed equal for all beams) and the relative position of the i-th receiver with respect to the center of the j-th beam. The only source of interference are then the signals corresponding to other beams reusing the same frequency (CCI), which partially reach outside their spotbeams due to the sidelobes of the radiation pattern. As mentioned in Section 3.2, from each receiver's point of view such a communication scenario can be regarded as a MISO scheme. On one side of the link, the gateway-satellite chain acts as a single transmitter delivering a given number of simultaneous transmitted streams (beams). At the other end, a terminal (one per spotbeam, and independent of other terminals) receives its signal corrupted with AWGN and CCI from other beams reusing the same frequency.

3.3.3 Particular evaluation scenarios

For the sake of evaluation we have selected a set of sample scenarios based on the coverage pattern of Figure 3.7. Each scenario exhibits different features so as to cover a meaningful range of application examples. The varying features among the selected scenarios are (i) the location of the receivers within their spotbeams, (ii) the frequency reuse factor, and (iii) the total number of beams controlled by a single gateway (N_{B}). This yields scenarios with different levels of CCI and processing complexity. We have selected four scenarios, which will be labeled as Si, with $i = 1, 2, 3, 4$. Figure 3.9 represents the spotbeams considered in each scenario and the location of the receivers. The beams marked in gray are assumed to cause no interference. We refer to Appendix 3.B where we provide the details on the geographical location of the spotbeam centers and the receivers for reproducibility purposes. Next we provide a short description of each scenario:

S1) A gateway controls $N_{\mathrm{B}} = 10$ *adjacent* beams which reuse the same frequency. Receivers are deliberately located near the beam borders, and very close to each other, so that many of them will experience a high CCI level.

S2) The scenario is similar to S1, with a gateway controlling $N_{\mathrm{B}} = 10$ adjacent beams, but with the receivers more sparsely located, which yields a moderate to high level of CCI. Figure 3.10 displays the coverage map corresponding to this scenario.

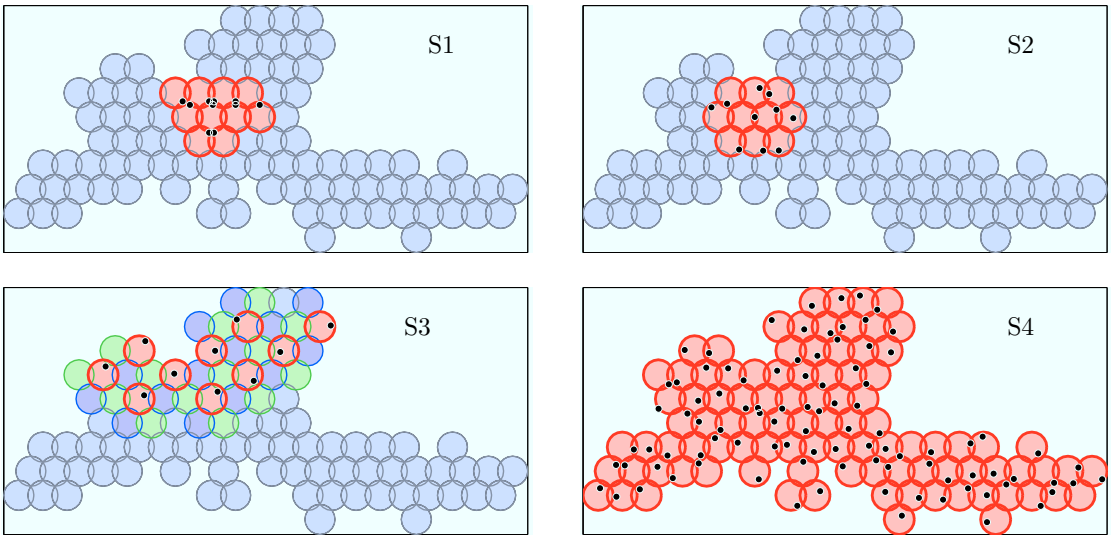


Figure 3.9: Considered beams (thick red circles) and receiver locations (black dots) for scenarios S1, S2, S3, and S4. Beams marked in gray are assumed to cause no interference.

S3) In this scenario we consider a multibeam system with frequency reuse factor $N_{\text{fr}} = 3$, in which one gateway controls $N_{\text{B}} = 10$ beams of common frequency. This scenario provides a low level of CCI since the beams are not adjacent. In Figure 3.9 we have marked some of the adjacent spotbeams in two different colors to represent the frequency reuse.

S4) All beams of Figure 3.7 ($N_{\text{B}} = 96$) are controlled by the same gateway and reuse the same frequency. Receivers are located so as to receive a moderate to high degree of CCI, similarly to scenario S2. Although it is unlikely that a current practical gateway could manage such a large number of beams due to the large bandwidth required in the feeder link (gateway to satellite), this scenario is included in the study in order to show the potential gain of such a system when precoding is used.

The differences among the above scenarios translate into different channel matrices $\boldsymbol{H}$. The channel matrix is found by considering the relative gains with which each terminal receives the signals of all considered beams. Under our signal model, these gains are functions of the antenna pattern and the location of the terminal. From the location of the terminal we obtain the angle θ (for the beam gain), and the link distance, which determines the propagation loss (we consider only free space loss). In practice, a scenario with low CCI will exhibit a channel matrix whose elements outside the main diagonal are small in magnitude. In turn, a high CCI level amounts to a number of off-diagonal elements having a considerable magnitude.

3.4 Considered THP Implementations

Our goal is to study the performance of THP transmission in the described scenarios. For that we have chosen three of the most representative THP implementations found

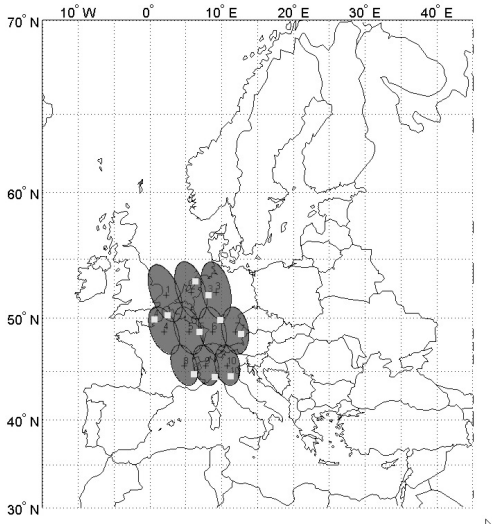


Figure 3.10: Coverage map with the considered beams (marked in dark) and receiver locations (marked with light squares) of scenario S2.

in the recent literature. Next we provide a description for which the block diagrams of Figures 3.4 and 3.5 will be useful.

ZF-THP by Fischer *et al.*, "ZF-THP-Fi". We will consider the ZF solution proposed by Fischer *et al.* in [60], and more detailed in [59]. In this scheme the channel matrix is decomposed as

$$\boldsymbol{H} = \boldsymbol{G}_{\text{ZF-Fi}}^{-1}(\boldsymbol{I} - \boldsymbol{B}_{\text{ZF-Fi}})\boldsymbol{F}_{\text{ZF-Fi}}^{\text{H}} \tag{3.10}$$

where $\boldsymbol{F}_{\text{ZF-Fi}}$ is unitary. The diagonal matrix $\boldsymbol{G}_{\text{ZF-Fi}}$ and the feedback matrix $\boldsymbol{B}_{\text{ZF-Fi}}$ are obtained through

$$\boldsymbol{R} = \boldsymbol{G}_{\text{ZF-Fi}}^{-1}(\boldsymbol{I} - \boldsymbol{B}_{\text{ZF-Fi}}) \tag{3.11}$$

where $\boldsymbol{R}$ is a lower triangular matrix which can be obtained through the Cholesky factorization

$$\boldsymbol{H}\boldsymbol{H}^{\text{H}} = \boldsymbol{R}\boldsymbol{R}^{\text{H}}. \tag{3.12}$$

The diagonal matrix $\boldsymbol{G}_{\text{ZF-Fi}}$ does not present a constant diagonal in general, i.e., $\boldsymbol{G}_{\text{ZF-Fi}} \neq c \cdot \boldsymbol{I}$. This implies that under this precoding scheme each receiver sees a different level of SNR. Although no special ordering is considered in [59, 60], we will use the "best-first" method to improve the performance of this scheme. In Section 3.A of the Appendix we detail such ordering procedure.

THP by Joham *et al.*, "ZF-THP-Jo" and "MMSE-THP-Jo". We will also consider the ZF and MMSE solutions proposed by Joham *et al.* in [97, 98]. These solutions are obtained through minimization of Mean Square Error (MSE) between the transmitted and received effective data sequences, $\text{E}\left\{\|\boldsymbol{v}_n - \boldsymbol{z}_n\|^2\right\}$, under the ZF and MMSE criteria.

Assuming that a given permutation matrix $\boldsymbol{\Pi}$ is applied, the ZF solution can be obtained via the Cholesky factorization

$$\boldsymbol{\Pi}\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{\Pi}^{\mathrm{T}} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{H}} \tag{3.13}$$

with $\boldsymbol{L}$ lower triangular. Such decomposition allows to write the matrix filters and scaling constant as follows

$$\boldsymbol{F}_{\text{ZF-Jo}} = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{\Pi}^{\mathrm{T}}\left(\boldsymbol{L}^{\mathrm{H}}\right)^{-1}\boldsymbol{D} \tag{3.14a}$$

$$\boldsymbol{B}_{\text{ZF-Jo}} = \boldsymbol{I} - \boldsymbol{L}\boldsymbol{D} \tag{3.14b}$$

$$\boldsymbol{G}_{\text{ZF-Jo}} = \boldsymbol{I} \tag{3.14c}$$

$$g_{\text{ZF-Jo}} = \sqrt{\frac{\mathrm{E}\left\{\|\boldsymbol{x}_n\|^2\right\}}{E_{\text{TX}}}} \tag{3.14d}$$

where $\mathrm{E}\left\{\|\boldsymbol{x}_n\|^2\right\}$ is the average energy of the precoded sequence at the output of the feedforward filter $\boldsymbol{F}$ (see Figure 3.3), and can be computed as

$$\mathrm{E}\left\{\|\boldsymbol{x}_n\|^2\right\} = \mathrm{Tr}\left(\boldsymbol{D}^2\tilde{\boldsymbol{X}}\right) \tag{3.15}$$

where $\boldsymbol{D}$ is a diagonal matrix whose diagonal elements are the inverses of those in the diagonal of $\boldsymbol{L}$, i.e., $\boldsymbol{D} \doteq \mathrm{diag}\left(([\boldsymbol{L}]_{11})^{-1}, \ldots, ([\boldsymbol{L}]_{N_{\mathrm{B}}N_{\mathrm{B}}})^{-1}\right)$. The matrix $\tilde{\boldsymbol{X}}$ in (3.14d) is the covariance matrix of the sequence $\tilde{\boldsymbol{x}}_n$ at the output of the modulo operator. Matrix $\tilde{\boldsymbol{X}}$ can be considered diagonal since the modulo operation decorrelates the entries of $\tilde{\boldsymbol{x}}_n$. We will assume that all precoded streams (entries) of $\tilde{\boldsymbol{x}}_n$ have the same variance $\frac{\Delta^2}{6}$, recall (2.11), except the one processed in the first place. Indeed, as indicated by (3.8a), the first processed entry $[\boldsymbol{\Pi}\boldsymbol{a}]_1$ remains unchanged, so the former variance does not apply. Thus, $\tilde{\boldsymbol{X}}$ is a diagonal matrix with its nonzero entries given by

$$[\tilde{\boldsymbol{X}}]_{ii} \doteq \sigma^2_{[\tilde{\boldsymbol{x}}]_i} = \begin{cases} \sigma_a^2 & \text{for } i = 1 \\ \frac{\Delta^2}{6} & \text{otherwise.} \end{cases} \tag{3.16}$$

Note that $\sigma_a^2 = 1$, since we have assumed that the original constellation is normalized in energy. Thus, it is possible to rewrite (3.15) as

$$\mathrm{E}\left\{\|\boldsymbol{x}_n\|^2\right\} = \frac{1}{([\boldsymbol{L}]_{11})^2} + \frac{\Delta^2}{6}\sum_{i=2}^{N_{\mathrm{B}}}\frac{1}{([\boldsymbol{L}]_{ii})^2}. \tag{3.17}$$

As for the permutation matrix $\boldsymbol{\Pi}$, it is obtained using "best-first" ordering. We refer to the Section 3.A in the Appendix for the details.

Regarding the MMSE solution, unfortunately it is not possible to write it in terms of some matrix factorization. Still, it is possible to write the solution in terms of the feedforward matrix $\boldsymbol{F}_{\text{MMSE-Jo}}$ as follows:

$$[\boldsymbol{B}_{\text{MMSE-Jo}}]_{ij} = \begin{cases} 0 & \text{for } i \leq j \\ -[\boldsymbol{\Pi}\boldsymbol{H}\boldsymbol{F}_{\text{MMSE-Jo}}]_{ij} & \text{for } i > j \end{cases}, \quad 1 \leq i, j \leq N_{\mathrm{B}} \tag{3.18a}$$

$$\boldsymbol{G}_{\text{MMSE-Jo}} = \boldsymbol{I} \tag{3.18b}$$

$$g_{\text{MMSE-Jo}} = \sqrt{\frac{\mathrm{E}\left\{\|\boldsymbol{x}_n\|^2\right\}}{E_{\text{TX}}}} \tag{3.18c}$$

with $\xi \doteq (N_{\mathrm{B}} \cdot \sigma_w^2)/E_{\mathrm{TX}}$, i.e., the ratio of aggregate received noise power to total transmitted power. The permutation matrix $\boldsymbol{\Pi}$ is obtained using "best-first" ordering as detailed in Section 3.A of the Appendix. The feedforward matrix $\boldsymbol{F}_{\text{MMSE-Jo}}$ admits a closed form. However, its mathematical expression is somewhat intricate, and requires some auxiliary matrices which are also used in "best-first" ordering. Therefore, we provide the expression of $\boldsymbol{F}_{\text{MMSE-Jo}}$ in Section 3.A of the Appendix.

The value of $\mathrm{E}\left\{\|\boldsymbol{x}_n\|^2\right\}$ (the average energy at the output of $\boldsymbol{F}$) can be written in terms of the feedforward matrix $\boldsymbol{F}_{\text{MMSE-Jo}}$ as follows ($\boldsymbol{f}_i$ represents the i-th column of $\boldsymbol{F}_{\text{MMSE-Jo}}$)

$$\mathrm{E}\left\{\|\boldsymbol{x}_n\|^2\right\} = \sum_{i=1}^{N_{\mathrm{B}}} \sigma^2_{[\tilde{\boldsymbol{x}}]_i} \|\boldsymbol{f}_i\|^2 = \|\boldsymbol{f}_1\|^2 + \frac{\Delta^2}{6} \sum_{i=2}^{N_{\mathrm{B}}} \|\boldsymbol{f}_i\|^2 \tag{3.19}$$

where we have used that $\sigma^2_{[\tilde{\boldsymbol{x}}]_1} = \sigma_a^2 = 1$.

We note that $\boldsymbol{G} = \boldsymbol{I}$ in both ZF and MMSE solutions, and thus the SNR is equally balanced among all receivers.

THP by Sanguinetti *et al.*, "ZF-THP-Sa". We have also considered the ZF solution proposed by Sanguinetti *et al.* in [140]. This solution can be seen as a generalization of ZF-THP-Jo, which allows the provision of different levels of Quality of Service (QoS) (i.e., SNR) to each receiver. The different SNR levels are represented by a real-valued diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_{N_{\mathrm{B}}})$. The computation of the matrices $\boldsymbol{B}$ and $\boldsymbol{F}$ and the gain g proceeds similarly to the "ZF-THP-Jo" case, but incorporating matrix $\boldsymbol{\Lambda}$ in the proper place. Assume that $\boldsymbol{\Lambda}$ and the permutation matrix $\boldsymbol{\Pi}$ are given. Then, application of Cholesky factorization to $\boldsymbol{\Pi H}$ yields

$$\boldsymbol{\Pi H H}^{\mathrm{H}} \boldsymbol{\Pi}^{\mathrm{T}} = \boldsymbol{L L}^{\mathrm{H}} \tag{3.20}$$

with $\boldsymbol{L}$ lower triangular. With this, the precoding system is determined by

$$\boldsymbol{F}_{\text{ZF-Sa}} = \boldsymbol{H}^{\mathrm{H}} \boldsymbol{\Pi}^{\mathrm{T}} \left(\boldsymbol{L}^{\mathrm{H}}\right)^{-1} \boldsymbol{D \Lambda} \tag{3.21a}$$

$$\boldsymbol{B}_{\text{ZF-Sa}} = \boldsymbol{I} - \boldsymbol{\Lambda}^{-1} \boldsymbol{L D \Lambda} \tag{3.21b}$$

$$\boldsymbol{G}_{\text{ZF-Sa}} = \boldsymbol{I} \tag{3.21c}$$

$$g_{\text{ZF-Sa}} = \sqrt{\frac{\mathrm{E}\left\{\|\boldsymbol{x}_n\|^2\right\}}{E_{\mathrm{TX}}}} \tag{3.21d}$$

with the permutation matrix $\boldsymbol{\Pi}$ obtained through "best-first" ordering (see Section 3.A for the details). The diagonal matrix $\boldsymbol{D}$ is defined similarly to the case "ZF-THP-Jo", and $\mathrm{E}\left\{\|\boldsymbol{x}_n\|^2\right\}$ incorporates now the effect of the matrix $\boldsymbol{\Lambda}$ as follows

$$\mathrm{E}\left\{\|\boldsymbol{x}_n\|^2\right\} = \left(\frac{\lambda_1}{[\boldsymbol{L}]_{11})}\right)^2 + \frac{\Delta^2}{6} \sum_{i=2}^{N_{\mathrm{B}}} \left(\frac{\lambda_i}{[\boldsymbol{L}]_{ii})}\right)^2 \tag{3.22}$$

where again the fact $\sigma^2_{[\tilde{\boldsymbol{x}}]_1} = 1$ has been used.

The diagonal matrix $\mathbf{\Lambda}$ is thus a further design parameter suitable to be tuned according to the desired SNR differentiation among receivers. Its effect on the received signals is clearly seen analyzing the SNR experienced by each receiver. The SNR seen by the i-th receiver can be written as

$$\text{SNR}_i = \lambda_i^2 \cdot \left(\frac{1}{\sigma_w^2 \cdot g_{\text{ZF-Sa}}^2} \right), \quad 1 \leq i \leq N_{\text{B}} \tag{3.23}$$

where the term inside brackets, which is directly proportional to E_{TX} through (3.21d), is common to all receivers. With this scheme, an appropriate choice of the matrix $\mathbf{\Lambda}$ allows to obtain any set of *relative* SNRs among the different receivers.[11] If $\mathbf{\Lambda} = \lambda \boldsymbol{I}$ is chosen for some real-valued scalar λ (i.e., all users experience the same SNR), the resulting scheme boils down to "ZF-THP-Jo".

3.5 Results and Discussion

We have evaluated the performance of all proposed precoding schemes for all selected scenarios. For a fair comparison, all schemes transmit the same power. Therefore, the comparison is made in terms of uncoded BER vs. E_{b}/N_0, where E_{b} is the average *transmitted* (instead of the received) energy per bit and N_0 is the one-sided noise power spectral density.

Figures 3.11 and 3.12 show the performance of precoding in a scenario S1 (high level of CCI) using QPSK and 16-QAM signaling respectively. Clearly, if no precoding is used the performance is very poor, reaching a high BER floor already at very low SNRs. The reason for this is that the CCI dominates over the noise, and the level of Signal-to-Interference plus Noise Ratio (SINR) remains the same despite a decrease in the noise power. LP schemes are seen to improve the performance, but, as expected, THP outperforms LP in the high SNR region due to the transmitted energy penalty of the LP (around 12 dB with respect to THP in both cases). Also as expected, MMSE approaches yield better results than their ZF counterparts in the low SNR region, whereas the performances of both approaches tend to converge as the SNR increases. In the QPSK case, it is remarkable that MMSE-LP performs better than ZF-THP for SNRs below 15 dB. A similar effect is observed also for 16-QAM, but in this case for SNRs below 10 dB. Finally, Figs. 3.11 and 3.12 confirm the equal performances of ZF-THP-Jo and ZF-THP-Sa when the same SNR is provided to all receivers in the latter implementation. Therefore, we drop ZF-THP-Sa from subsequent comparisons. Finally, it can be seen in both cases that LP and THP offer the same performance at very low SNRs. This is in accordance with the results in [94], where it is proved that no additional gain can be expected from DPC or any of its suboptimum approximations in the low SNR region.

Figures 3.13, 3.14 and 3.15 show the same as Figure 3.12 but for scenarios S2, S3 and S4 respectively. In Figures 3.13 and 3.15 (intermediate-to-high CCI) the results are similar

[11] We speak of *relative* SNRs because the absolute values are ultimately constrained by the transmit power limitation E_{TX}.

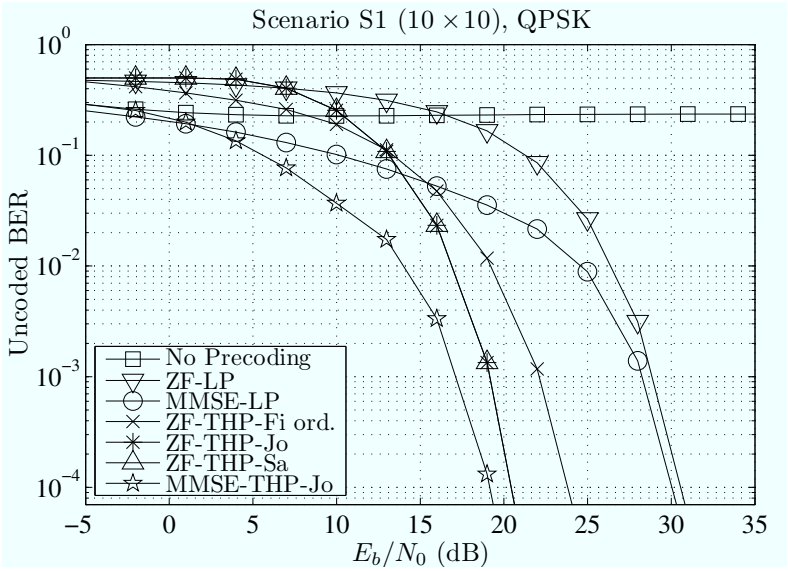


Figure 3.11: Performance in terms of uncoded BER of all precoding schemes for S1 using QPSK.

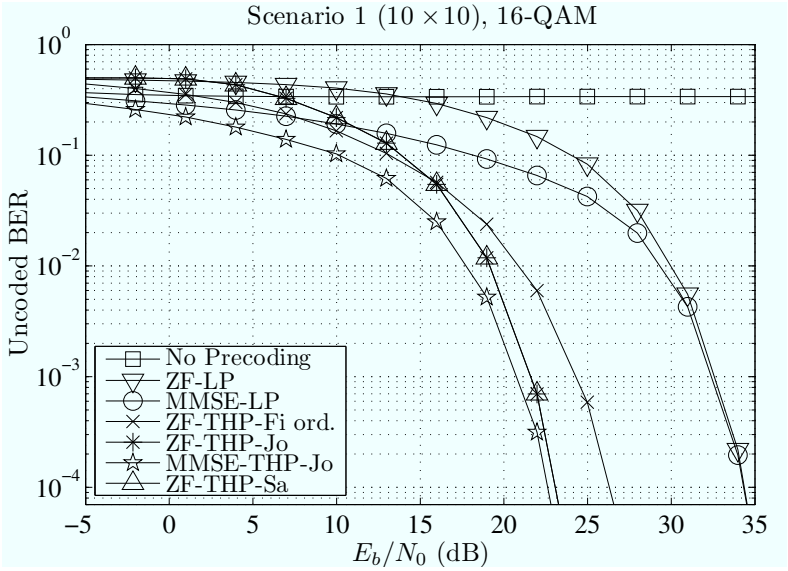


Figure 3.12: Performance in terms of uncoded BER of all precoding schemes for S1 using 16-QAM.

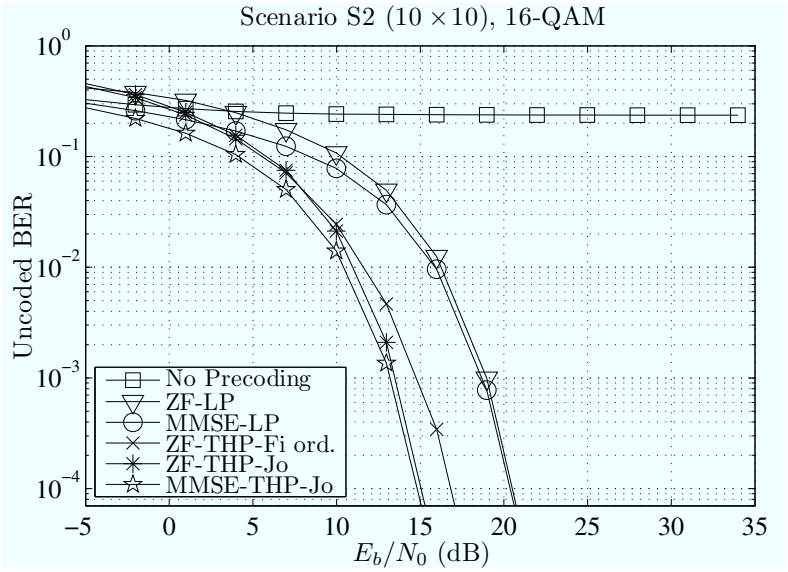


Figure 3.13: Performance in terms of uncoded BER for S2 using 16-QAM.

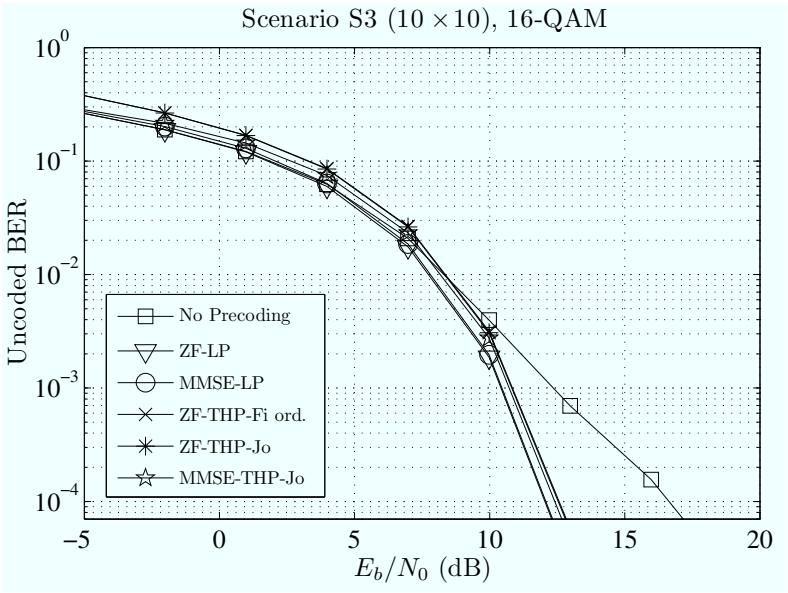


Figure 3.14: Performance in terms of uncoded BER of all precoding schemes for S3 using 16-QAM.

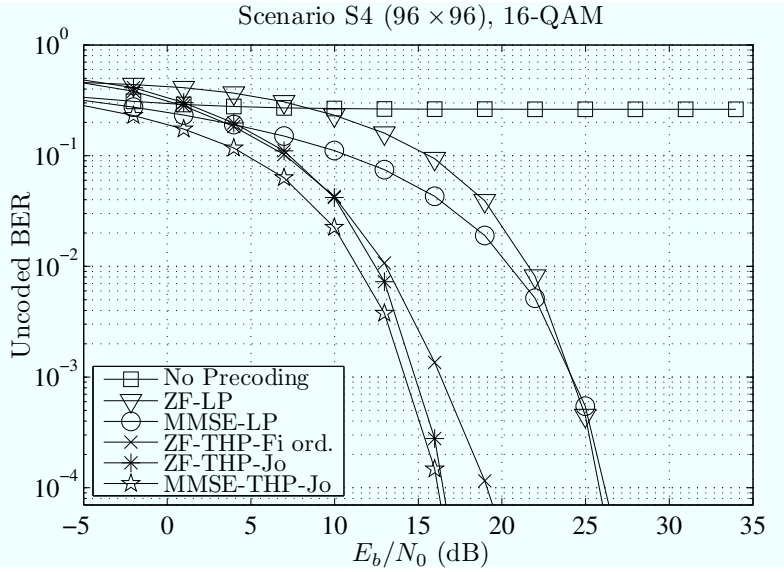


Figure 3.15: Performance in terms of uncoded BER for S4 using 16-QAM.

to those found for scenario S1, with THP outperforming LP. The difference with respect to scenario S1 is that the gap between LP and THP is smaller due to the lower level of CCI. As for Figure 3.14 (scenario S3, low CCI), it is noticeable that the performance of plain 16-QAM transmission is close to that of the schemes with beamforming and precoding. Only for high SNRs the performance of plain 16-QAM begins to degrade. Furthermore, the curves for the schemes with beamforming and THP overlap almost completely. This is so because the transmit power penalty of LP is negligible in this case due to the low level of CCI. The small shift observed between the curves for THP and LP (in favor of the latter scheme) is due to the precoding loss, which for 16-QAM equals $\gamma_p^2 = 16/15$ (0.28 dB). It is evident that the performance of all schemes is in this case dominated by the noise (rather than by the CCI). This suggests that, in this case, a simpler scheme such as LP would be preferable to THP at very high SNRs (above 10 dB), whereas no precoding would be needed at lower SNRs.

Figure 3.16 shows the performance of MMSE-THP-Jo under the considered scenarios. For reference, the performance of plain 16-QAM transmission in scenario S3 is shown. Scenarios S1 to S4 offer channels with different amounts of CCI: low CCI in scenario S3, moderate to high CCI in scenarios S2 and S4, and high CCI (worst case) in scenario S1. It can be seen that an increase in the amount of CCI translates into a shift of the BER curve towards the right. It is remarkable that the performances are similar for the scenarios with moderate to high CCI, S2 and S4 (10 and 96 beams respectively).

To wrap up, we have seen that the adoption of THP in the downlink of a multibeam satellite system can offer remarkable benefits in terms of system efficiency. The fact that the CCI can be effectively canceled by the THP scheme can be used to reduce the frequency reuse factor required in current multibeam systems to prevent an excess of CCI. This is especially true when the gateway performs the precoding for a large number of beams. The application of Linear Precoding (LP) to similar scenarios was analyzed

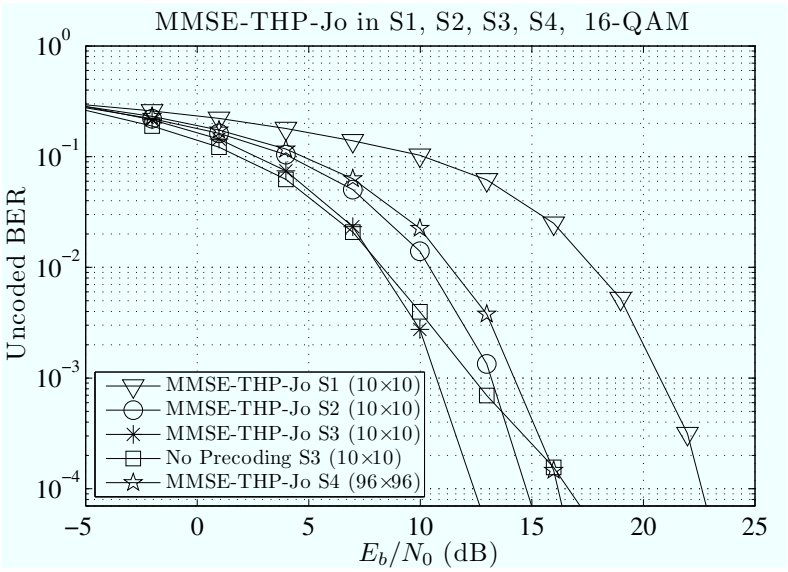


Figure 3.16: Performance in terms of uncoded BER of MMSE-THP-Jo for all scenarios using 16-QAM.

in detail in [43, 67] offering interesting potential improvements in terms of throughput with respect to traditional systems (above 80%). Our simulations show that even greater improvements can be expected if THP is used, mainly due to the removal of the transmit power penalty inherent to LP schemes.

Appendix 3.A Filter Computation for MIMO THP

3.A.1 "Best-first" Ordering

As mentioned in Section 3.2, the overall performance of MIMO precoding depends on the particular precoding order. Recall that the precoding operation is performed sequentially, as indicated in (3.8a)–(3.8c). Following the work in [97, 98], as optimality criterion we consider the minimization of the MSE between the transmitted and received *effective data sequences*. In reality, the optimal precoding order can only be determined by extensive search. It would be necessary to test all of the $N_{\mathrm{B}}!$ possible orderings and then choose the one yielding the minimum MSE. Such task is computationally intensive, so some suboptimal but feasible procedure would be desirable. One such scheme is the "best first" ordering procedure, which is simpler but yields orderings which usually perform close to optimality.

We note that in all precoding schemes proposed in Section 3.4, matrices $\boldsymbol{B}$ and $\boldsymbol{F}$, and the scaling constant g depend on the permutation matrix $\boldsymbol{\Pi}$. Therefore, the precoding order must in general be found in the first place, and then proceed to the computation of the matrices and the constant. It is possible, however, to jointly compute the columns of $\boldsymbol{\Pi}$ and $\boldsymbol{F}$, by applying a sequential algorithm as detailed in [97, 98], although we will not provide the details here.

We describe next how best-first ordering works following [97, 98]. The basic principle is to minimize the MSE between the N_{B} received and transmitted effective data sequences. Each MSE is minimized assuming a fixed ordering of the streams to be precoded next. Depending on the precoding scheme (ZF or MMSE), the ordering procedure changes. Omitting the details (see [97, 98] for the derivation), the ordering procedure is as follows ($\mathcal{O}_{(\cdot)}$ represents the set of ordered indices):

$$\begin{aligned}\mathcal{O}_{\mathrm{ZF}} &= \{o_1, o_2, \ldots, o_{N_{\mathrm{B}}}\} \quad \text{with} \\ o_i &= \underset{o \in \{1,\ldots,N_{\mathrm{B}}\}\backslash\{o_{i+1},\ldots,o_{N_{\mathrm{B}}}\}}{\arg\min} \left[\boldsymbol{\Pi}_i \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{\Pi}_i\right]^{+}_{oo} \quad \text{for } i = N_{\mathrm{B}}, \ldots, 1 \end{aligned} \tag{3.24}$$

for ZF-THP, and

$$\begin{aligned}\mathcal{O}_{\mathrm{MMSE}} &= \{o_1, o_2, \ldots, o_{N_{\mathrm{B}}}\} \quad \text{with} \\ o_i &= \underset{o \in \{1,\ldots,N_{\mathrm{B}}\}\backslash\{o_{i+1},\ldots,o_{N_{\mathrm{B}}}\}}{\arg\min} \left[\boldsymbol{\Pi}_i \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{\Pi}_i + \xi\boldsymbol{I}\right]^{-1}_{oo} \quad \text{for } i = N_{\mathrm{B}}, \ldots, 1 \end{aligned} \tag{3.25}$$

for MMSE-THP, with $\xi \doteq \frac{N_{\mathrm{B}}\cdot\sigma_w^2}{E_{\mathrm{TX}}}$. In both cases, the projection matrices $\boldsymbol{\Pi}_i$ are diagonal matrices which, given a precoding order $\mathcal{O} = \{o_1, o_2, \ldots, o_{N_{\mathrm{B}}}\}$, are defined as

$$\boldsymbol{\Pi}_{N_{\mathrm{B}}} = \boldsymbol{I}, \tag{3.26a}$$

$$[\boldsymbol{\Pi}_i]_{kk} = \begin{cases} 0 & \text{if } k = o_i \\ [\boldsymbol{\Pi}_{i+1}]_{kk} & \text{otherwise} \end{cases}, \quad \text{for } i = N_{\mathrm{B}} - 1, \ldots, 1. \tag{3.26b}$$

The permutation matrix $\boldsymbol{\Pi}$ is then a pure function of the obtained ordering, $\mathcal{O}_{\mathrm{ZF}}$ or $\mathcal{O}_{\mathrm{MMSE}}$.

Procedure (3.24) is valid for all the proposed ZF schemes: "ZF-THP-Fi", "ZF-THP-Jo", and "ZF-THP-Sa". In the latter case, the only change is that the modified matrix $\boldsymbol{\Lambda}^{-1}\boldsymbol{H}$ must be used instead of $\boldsymbol{H}$.

Note that in procedures (3.24) and (3.25) the first obtained index refers to the data stream to be precoded in the last place, followed by the second last, and so on until the index of the stream to be precoded first, which is obtained in the last place.

3.A.2 Feedforward Matrix Computation for "MMSE-THP-Jo"

In Section 3.4 we indicate the solution to the "MMSE-THP-Jo" case as a function of the feedforward matrix $\boldsymbol{F}_{\text{MMSE-Jo}}$, which is not provided. Here we provide its closed form expression, for which we will take advantage of the projection matrices $\boldsymbol{\Pi}_i$, already defined by (3.26). Instead of a single expression for $\boldsymbol{F}_{\text{MMSE-Jo}}$, it is more convenient to provide the expression of its columns. By $\boldsymbol{f}_i$ we will denote the i-th column of $\boldsymbol{F}_{\text{MMSE-Jo}}$, and we will assume that the precoding order has already been found and is given by $\mathcal{O} = \{o_1, o_2, \ldots, o_{N_\text{B}}\}$. We introduce the matrix

$$\boldsymbol{A}^{(i)} \doteq \boldsymbol{H}^{\text{H}}\boldsymbol{\Pi}_i(\boldsymbol{\Pi}_i\boldsymbol{H}\boldsymbol{H}^{\text{H}}\boldsymbol{\Pi}_i + \xi\boldsymbol{I})^{-1} \tag{3.27}$$

with $\xi = \frac{N_\text{B}\cdot\sigma_w^2}{E_\text{TX}}$, and whose j-th column we denote by $\boldsymbol{a}_j^{(i)}$, i.e.,

$$\boldsymbol{A}^{(i)} = [\boldsymbol{a}_1^{(i)}, \boldsymbol{a}_2^{(i)}, \ldots, \boldsymbol{a}_{N_\text{B}}^{(i)}].$$

Then $\boldsymbol{F}_{\text{MMSE-Jo}}$ is defined columnwise as follows

$$\boldsymbol{F}_{\text{MMSE-Jo}} = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_{N_\text{B}}] \quad \text{with} \tag{3.28a}$$

$$\boldsymbol{f}_i = \boldsymbol{a}_{o_i}^{(i)}, \quad 1 \leq i \leq N_\text{B} \tag{3.28b}$$

i.e., the i-th column of $\boldsymbol{F}_{\text{MMSE-Jo}}$ is given by the o_i-th column of $\boldsymbol{A}^{(i)}$. Note that the computation of each column of $\boldsymbol{F}_{\text{MMSE-Jo}}$ requires the computation of a different matrix $\boldsymbol{A}^{(i)}$.

Appendix 3.B Receiver Locations for Scenarios S1, S2 and S3

For the sake of reproducibility, we include in this appendix the coordinates of the 96 spotbeam centers of the multibeam system detailed in Section 3.3 (see Figure 3.7). We also provide the receiver locations for scenarios S1, S2 and S3 (see Figure 3.9). Scenario S4 is left out for economy of space.

Table 3.1 lists the coordinates of the 96 spotbeam centers. The spotbeams are numbered from 1 to 96, from West to East and North to South. For example, the spotbeams in the most northern row are numbered from 1 to 4; those in the second most northern row are numbered from 5 to 10, and so on. Tables 3.2 provides the coordinates of the receiver locations for scenarios S1, S2 and S3. We include in the table the index of the spotbeam which the receiver belongs to.

Table 3.1: Location of the spotbeam centers considered in Section 3.3. Spotbeams are numbered from 1 to 96. Partial horizontal lines indicate where a new row of spotbeams starts.

Nr.	Center location		Nr.	Center location		Nr.	Center location	
1	67.00°N	11.00°E	33	48.75°N	11.90°E	65	39.75°N	1.50°W
2	67.00°N	16.25°E	34	48.75°N	14.80°E	66	39.75°N	0.75°E
3	67.00°N	21.75°E	35	48.75°N	17.80°E	67	39.75°N	3.50°E
4	67.00°N	27.50°E	36	48.75°N	21.00°E	68	39.75°N	9.25°E
5	61.00°N	6.50°E	37	45.50°N	0.90°W	69	39.75°N	14.50°E
6	61.00°N	10.75°E	38	45.50°N	2.00°E	70	39.75°N	17.00°E
7	61.00°N	15.00°E	39	45.50°N	4.80°E	71	39.75°N	19.50°E
8	61.00°N	19.25°E	40	45.50°N	7.80°E	72	39.75°N	21.75°E
9	61.00°N	23.50°E	41	45.50°N	10.80°E	73	39.75°N	24.00°E
10	61.00°N	27.75°E	42	45.50°N	13.70°E	74	39.75°N	26.50°E
11	56.20°N	6.25°W	43	45.50°N	16.50°E	75	39.75°N	29.25°E
12	56.20°N	2.00°W	44	45.50°N	19.30°E	76	39.75°N	32.00°E
13	56.20°N	10.00°E	45	45.50°N	22.10°E	77	39.75°N	34.75°E
14	56.20°N	13.75°E	46	42.25°N	7.25°W	78	39.75°N	37.50°E
15	56.20°N	17.50°E	47	42.50°N	4.25°W	79	39.75°N	40.25°E
16	56.20°N	21.25°E	48	42.50°N	1.25°W	80	39.75°N	43.20°E
17	56.20°N	25.00°E	49	42.50°N	1.50°E	81	37.75°N	7.75°W
18	52.25°N	8.90°W	50	42.50°N	4.25°E	82	37.25°N	5.00°W
19	52.25°N	5.65°W	51	42.50°N	7.25°E	83	37.25°N	2.00°W
20	52.25°N	1.85°W	52	42.50°N	10.00°E	84	37.25°N	13.50°E
21	52.00°N	2.20°E	53	42.50°N	12.75°E	85	37.25°N	16.00°E
22	52.25°N	5.75°E	54	42.50°N	15.50°E	86	37.25°N	22.75°E
23	52.25°N	9.25°E	55	42.50°N	18.25°E	87	37.25°N	25.00°E
24	52.25°N	12.75°E	56	42.50°N	20.75°E	88	37.25°N	27.50°E
25	52.25°N	16.00°E	57	42.50°N	23.25°E	89	37.25°N	30.25°E
26	52.25°N	19.25°E	58	42.50°N	25.75°E	90	37.25°N	32.75°E
27	52.25°N	22.50°E	59	42.50°N	28.50°E	91	37.25°N	35.50°E
28	48.75°N	4.00°W	60	42.50°N	31.25°E	92	37.25°N	38.00°E
29	48.75°N	1.25°W	61	42.50°N	34.25°E	93	37.25°N	40.75°E
30	48.75°N	2.00°E	62	42.50°N	37.00°E	94	37.25°N	43.50°E
31	48.75°N	5.40°E	63	39.90°N	7.50°W	95	35.00°N	25.00°E
32	48.75°N	8.75°E	64	39.75°N	4.25°W	96	35.00°N	33.50°E

Table 3.2: Receiver locations for scenarios S1, S2 and S3. "SB Nr." stands for the receiver's spotbeam number, as listed in Table 3.1.

S1			S2			S3		
SB Nr.	Receiver location		SB Nr.	Receiver location		SB Nr.	Receiver location	
22	51.00°N	7.00°E	21	50.20°N	2.25°E	7	63.00°N	12.50°E
23	51.00°N	11.00°E	22	53.50°N	6.00°E	10	61.00°N	30.00°E
24	51.00°N	11.50°E	23	52.00°N	8.00°E	12	58.50°N	3.00°W
25	51.00°N	14.50°E	30	50.00°N	1.00°E	13	56.60°N	11.00°E
32	50.50°N	9.00°E	31	48.75°N	7.00°E	16	56.40°N	20.20°E
33	50.50°N	11.90°E	32	49.50°N	10.00°E	19	54.00°N	6.50°W
34	50.50°N	15.00°E	33	48.50°N	12.50°E	22	52.50°N	5.10°W
35	50.50°N	16.00°E	39	45.00°N	6.50°E	25	51.00°N	17.30°E
41	46.75°N	11.80°E	40	44.50°N	9.25°E	30	49.50°N	2.50°E
42	46.75°N	12.70°E	41	44.50°N	11.50°E	33	50.00°N	12.50°E

Appendix 3.C Beamforming Implementation

An important issue not addressed in the main text of this chapter is the implementation of the antenna beamforming. By *beamforming* we refer to the techniques used to shape the radiation pattern of the satellite antenna so that the coverage area is illuminated as desired. We provide here a brief overview of the state-of-the-art antenna technology and beam conformation techniques, and their possible relationship with the precoding stage.

Typically, a GEO satellite broadcasting TV over a region implements a single beam. The shape of the beam is adapted to the region of coverage with some kind of beamforming technology in order not to waste transmission power over undesired regions such as oceans. In contrast, in multibeam systems the desired coverage is achieved through the superposition of a number of smaller beams of circular or elliptical shape. In principle, each of these beams can be fed with a signal which is completely independent of the signals from the rest of beams. This fact is exploited in the THP-based transmission system considered in this chapter.

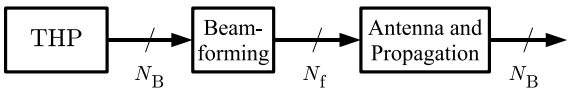


Figure 3.17: Simplified block diagram of a multibeam transmission scheme using THP.

In general, beam conformation is achieved in two steps, as illustrated in Figure 3.17. The eventual radiation pattern is determined by the combined action of the former stages:

- First, the **beamforming stage** is a signal preprocessing block which adapts the signals to be transmitted (say N_B beam signals) to the N_f antenna feed. This preprocessing stage has been traditionally implemented in the analog domain, although digital processing techniques are possible and may be considered for future systems. The beamforming stage has been typically located onboard the satellite, al-

though several systems have been recently proposed or implemented in which the preprocessing is (at least partially) performed on ground, as we will see below.
- The **satellite antenna** is the last stage before the signals exit the satellite. By *antenna* we refer to both the feeds and the possible reflector.

Multibeam Antenna Technologies We briefly provide a classification of the satellite antenna systems suitable to implement multibeam coverage. Four main possibilities exist [108, 116]:

- **Single Reflector Feed (SRF) system**. It consists of one reflector which is illuminated by a feed array, with each feed generating a *constituent* beam. *Actual* beams are formed by superposition of several constituent beams.
- **Single Reflector with Hybrid Transform (SRHT) system**. It is basically an SRF system in which an hybrid transformer or a Butler matrix[12] is placed directly before the feeds with the purpose of distributing the energy evenly among the feeds.
- **Dual Reflector Array-Fed (DRAF) system**. This system uses a smaller subreflector located between the feeds and the main reflector. The subreflector produces an extra signal transformation which yields a uniform amplitude output.
- **Directly Radiating Array (DRA) system**. No reflector is used under this configuration, so the antenna elements radiate directly towards the coverage region on the Earth surface. The size of the array must be comparable to that of the main reflectors of the above configurations. Depending on the total power to be radiated, an active amplifier module must be connected to each single feed or to a small group of them.

Roughly speaking, the number of feeds is proportional to the number of beams to be deployed. Current multibeam systems deploy antennas with more than one hundred feeds [44, 61, 156]

Beamforming schemes The task of the beamforming stage is to provide the adequate signals to the antenna feeds so that the desired beams are formed. In general, the number of beams N_{B} will not coincide with the number of the antenna feeds (N_{f}). Thus, the preprocessing stage must apply some kind of combination of its inputs, implicitly performing a dimensional transformation from N_{B} to N_{f}. The beamforming operation amounts to a weighted linear combination of the input signals, so that it can be represented using matrix notation. Figure 3.18 shows this, representing the beamforming operation through the $N_{\mathrm{f}} \times N_{\mathrm{B}}$ matrix $\boldsymbol{M}_{\mathrm{BF}}$, and the joint effect of antenna and propagation by means of the $N_{\mathrm{B}} \times N_{\mathrm{f}}$ matrix $\boldsymbol{H}_{\mathrm{rad}}$. The cascade of $\boldsymbol{M}_{\mathrm{BF}}$ and $\boldsymbol{H}_{\mathrm{rad}}$ amounts to the channel matrix $\boldsymbol{H}$ considered in the signal model of Section 3.2.1, i.e., $\boldsymbol{H} = \boldsymbol{H}_{\mathrm{rad}} \cdot \boldsymbol{M}_{\mathrm{BF}}$.

Beamforming techniques, originally performed in the analog domain, evolved significantly with the application of digital processing techniques. In practice both approaches can be used together, but it is the digital part which provides a degree of flexibility unreachable by pure analog beamforming. Whereas analog beamforming can take place at

[12] Butler matrices are analog beamforming networks whose processing is equivalent to that of a Fast Fourier Transform (FFT), but performed in the analog domain.

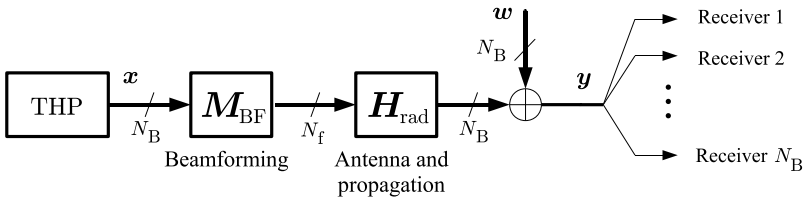


Figure 3.18: Matrix representation of beamforming and propagation.

RF or IF frequencies, Digital Beamforming (DBF) is performed in the baseband domain. In general, beams are conformed through a weighted combination of the individual signals to be transmitted. The antenna feeds are fed by one different linear combination each. Beamforming can be applied at the transmitter or at the receiver (in which case it consists of a linear combination of the signals coming from the antenna receiving elements which yields the beams as result).

The beamforming stage has been traditionally located onboard the satellite, approach known as Onboard Beamforming. Nevertheless, there exist other possibilities such as Ground-Based Beamforming and Hybrid Space-Ground Beamforming, which (at least partially) locate the beamforming process at the ground unit. Next we describe them, assuming that the feeder link is transparent in the sense that it does not alter the beamforming process.

- **Onboard Beamforming (OBBF)**. Under this approach the whole beamforming task is performed onboard the satellite. The data signals are sent from the gateway through the feeder link using some kind of multiplexing (i.e., frequency and polarization). Such signals are first demultiplexed onboard, and then serve as inputs to the beamforming processor onboard. These input signals may be already the beam signals, but not necessarily. This configuration is the most demanding in terms of satellite payload complexity. An example of this approach can be found in [117].
- **Ground-Based Beamforming (GBBF)**. In order to reduce the processing burden onboard the satellite, the beamforming stage can be performed at the gateway. Under this configuration, the already beamformed signals are multiplexed through the feeder link at the cost of losing flexibility if the coarse beamforming stage is fixed. The satellite onboard unit is only required to demultiplex them and then feed the antenna elements with the signals received from ground. Although some antenna related hardware must exist on the satellite, the onboard hardware is greatly simplified. Furthermore, the complexity of the beamforming stage is in principle unlimited, since the ground-based equipment is not subject to strict weight, space and power consumption limitations as it is the case for the satellite payload. The possible drawback of this approach is the potentially large bandwidth required for the feeder link, which amounts to the aggregate bandwidth of the beamformed signals. This approach was originally proposed in [25], and more recently updated in [14, 24, 162], and it is being currently deployed in satellite systems such as the ICO-G1 [89] and TerreStar-I [156].
- **Hybrid Space-Ground Beamforming (HSGBF)**. This configuration lies midway between the two former schemes. This approach seeks to reduce the hardware re-

quirements onboard the satellite with respect to the OBBF approach, and simultaneously reduce the bandwidth required for the feeder link with respect to the GBBF scheme. The beamforming task is divided between the gateway and the satellite payload. The gateway performs most of the beamforming process, and the task is completed by an onboard unit with "coarse" beamforming capabilities (e.g., an FFT matrix). By doing this it is possible to reduce the number of signals sent through the feeder link. This beamforming scheme has been proposed in [13, 134].

As already mentioned, a THP system with one receiver per beam such as that considered in this chapter apparently calls for the use of beam-space beamforming, since the precoding operation already provides one separate data stream per beam. A key issue for the correct performance of the THP multibeam system is the minimization of the impact of the nonlinear elements of the transmission chain, including the gateway and the satellite. Among possible sources of nonlinear distortion are frequency mismatches in RF and IF mixers, signal couplings, and the nonlinear characteristics of the gateway and onboard HPAs. The possible effects of the HPA-induced nonlinearities are dealt with in the next section. In general, the presence of nonlinearities will entangle the signals intended to different users, and therefore will harm the performance of the precoding scheme.

Appendix 3.D Nonlinear Distortion in the Multibeam Downlink Scenario

In this section we focus on the multibeam downlink scenario since the compensation of the satellite HPA nonlinearity via predistortion for the SISO case was analyzed in Chapter 2. The baud rate signal model assumed in Section 3.2 for the proposed precoded system (see Figure 3.4) does not consider the nonlinear effects that the HPAs may introduce in the transmission chain. If such nonlinearities are disregarded, then the chain gateway-satellite-receiver is linear, so that the channel matrix $\boldsymbol{H}$ includes the combined effects of the beamforming stage onboard, antenna and propagation. However, if nonlinearities cannot be neglected then the linear model is insufficient. The aim of this section is to provide a brief analysis of the effects of the HPA nonlinearity within the multibeam precoding setup. Only the effect of the satellite HPA is considered. The HPA at the gateway is assumed to introduce negligible nonlinear distortion, assuming that proper countermeasures are employed in the gateway (e.g. back-off [39]), and the RF up- and down-converters at the gateway, satellite and receivers are supposed to introduce no further distortion.

Figure 3.19 represents the resulting block diagram of the system including THP, beamforming and HPA onboard. The RF upconversion block represent the corresponding operation performed e.g. at the gateway, but we assume they work ideally. Its inclusion in Figure 3.19 is to indicate where the limits lie between baseband and RF processing. The HPA block represents in fact one HPA per antenna feed.[13] We will assume that the HPAs

[13]This model is in principle valid even when multiport amplifiers are used. Multiport amplifiers [53, 115, 154] are composed of a number (usually a power of two) of SSPAs working in parallel and two RF mixing matrices (e.g., analog Butler matrices) of the same dimension located at the input and the output of the

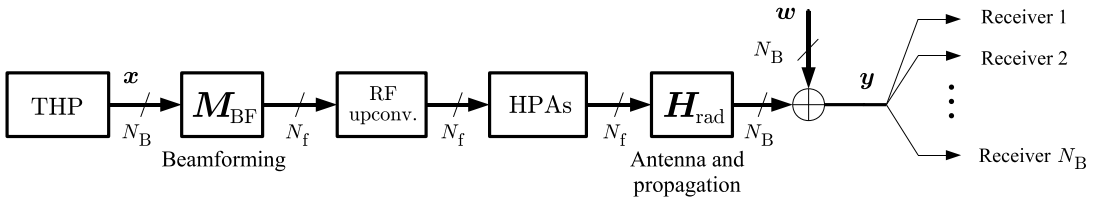


Figure 3.19: Simplified block diagram of the transmission chain including THP, beamforming and HPA onboard.

do not introduce "cross-talk", i.e., no fraction of their input signals is driven into any feed other than the correct one.

The main issue observed in Figure 3.19 is that the nonlinear HPAs are located between two linear blocks. If we assume ideal (linear) amplifier behavior, then the HPA block becomes transparent, and the cascade of beamforming and propagation matrices behaves as a single channel matrix $\boldsymbol{H} = \boldsymbol{H}_{\text{rad}} \cdot \boldsymbol{M}_{\text{BF}}$, which corresponds in fact with the channel matrix $\boldsymbol{H}$ of the signal model considered in Section 3.2.1. However, when the HPA behavior departs from linearity this simplified model does not apply anymore. The main consequence of the presence of the HPAs between the two linear blocks is that each beam signal is not only affected by self-induced nonlinear distortion but also from (nonlinear) contributions from signals belonging to other beams (nonlinear interbeam interference), including cross terms between signals belonging to different beams. This is due to the beamforming stage, represented by $\boldsymbol{M}_{\text{BF}}$, whose effect is to spread the beam signals among all feeds. Therefore, if the HPA associated to one feed introduces nonlinear distortion, such distortion will be affecting the signals corresponding to all beams.

The question that arises now is: *Is it possible to achieve nonlinear interference-free transmission (including interbeam contributions) in a multibeam satellite system using THP while maintaining non-cooperativeness among receivers?*

To the best of the author's knowledge, this specific topic does not seem to have been addressed in the open literature. Possible solutions must necessarily consider system-specific features such as the employed beamforming scheme, the particular satellite HPA models and the possibility of applying adequate processing onboard the satellite, which may not be guaranteed depending on the affordable or already onboard located payload. We leave the actual solution here as a future research line, although in the remainder of this section we will provide a shallow overview of different possible approaches to the solution.

3.D.1 Possible Approaches to Predistortion

Initially we can consider solutions based on applying some processing either at transmitter side (predistortion) or at receiver side (postdistortion), or at both locations. However, the

amplifiers. The output matrix is placed immediately before the feeds. The effect of the input and output matrices could be properly accounted for within, respectively, the matrices $\boldsymbol{M}_{\text{BF}}$ and $\boldsymbol{H}_{\text{rad}}$ of Figure 3.19.

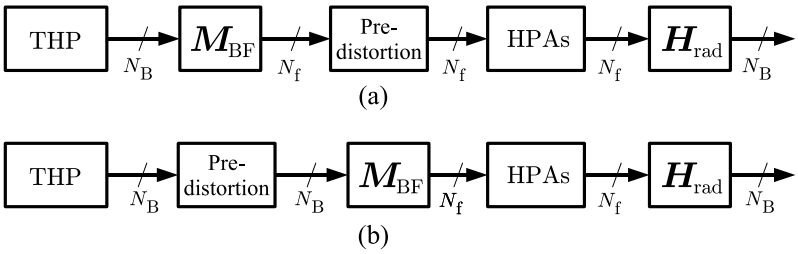


Figure 3.20: Two possible approaches to predistortion: (a) after the beamforming stage, (b) before the beamforming stage.

constraint imposed by the non-cooperativeness of the receivers makes it impossible to solve the problem only with processing at the receiver. This would require the application of some kind of nonlinear equalization scheme based upon the knowledge of the signals corresponding to other beams. Thus, the solution should consist of a precorrecting stage necessarily located at some point before the HPA block. In principle, an optional post-processing stage located at each receiver could be used to complete the task, although this stage could only take care of the self-induced components of the nonlinear interference. For simplicity of the analysis, we seek here solutions in which the whole task is performed before the HPAs.

Recalling Figure 3.19, and disregarding the effect of the RF upconversion onboard, the origin of the problem lies in the cascade of the beamforming matrix $\boldsymbol{M}_{\text{BF}}$, the HPA nonlinearity, and the radiation matrix $\boldsymbol{H}_{\text{rad}}$. Such configuration is typically referred to as cascaded Linear-Nonlinear-Linear (LNL) model (also *sandwich* model), and appears often in problems of Control Theory [35]. Other typical configurations are the Hammerstein model [Nonlinear-Linear (NL) cascade] and the Wiener model [Linear-Nonlinear (LN) cascade]. In what regards the predistortion operation, it suffices to assume that the nonlinear behavior of the satellite HPAs can be modeled with enough accuracy. We briefly discuss here two possible approaches to predistortion which differ in the location of the predistortion stage within the transmission chain: (i) between the beamforming stage and the HPAs, and (ii) before the beamforming stage. Figure 3.20 represents both approaches (all other elements in the transmission chain are assumed transparent).

Signal Predistortion after the beamforming stage. Under this configuration, predistortion is to be applied to the signals corresponding to the feeds, i.e., between the beamforming stage and the HPAs. Recalling Figure 3.19, it seems more convenient to predistort the beamformer outputs in the baseband domain, i.e., before the RF upconversion is performed. Note that this approach requires one predistorter per feed. Assuming this, the different predistortion strategies given in Section 2.2.3.2 apply, so, for example, the predistorter could combine fractional predistortion and a certain back-off. The physical location of the predistortion stage depends in any case on the location of the beamforming stage. If beamforming is performed onboard the satellite, either partially (HSGBF) or completely (OBBF), then the predistortion block can only be placed onboard. If, instead, beamforming is performed at the gateway (GBBF), then the predistortion block can be placed either onboard or at the gateway.

The possibility of performing the predistortion at the gateway is quite appealing, considering the presumably large number of signals to be predistorted (typically in the order of one hundred, according to the current antenna technology). Indeed, gateway-based predistortion has the advantages of adding no extra processing requirements to the satellite payload, and, simultaneously, allowing the implementation of advanced predistortion techniques (thanks to the larger processing capacity at the gateway), which would not be affordable onboard. When performed at the gateway, the baseband beamformed signals should be predistorted before being RF upconverted and multiplexed in frequency and polarization for the feeder link.

Signal predistortion before the beamforming stage. An alternative to the above scheme is to perform predistortion before the beamforming operation. Under this configuration the predistortion stage would be preferably located at the gateway, irrespective of where the beamforming task is performed. This scheme would be useful in the case that OBBF or HSGBF are used, and there is no possibility of locating the predistortion stage onboard the satellite. Basically, the problem here is how to precompensate (invert) a nonlinearity that appears after a linear block (matrix $\boldsymbol{M}_{\text{BF}}$).

Tentatively, one could think of building the predistorter as an LNL cascade composed of $\boldsymbol{M}_{\text{BF}}$, a nonlinear mapping inverting the HPA characteristic, and a matrix $\boldsymbol{M}_{\text{BF}}^{+,\text{r}}$ representing the right pseudoinverse of $\boldsymbol{M}_{\text{BF}}$. We note however that $\boldsymbol{M}_{\text{BF}}$ is typically a "tall" matrix, i.e., it has more rows than columns, since the number of feeds is typically larger than the number of data streams, and therefore $\boldsymbol{M}_{\text{BF}}$ does have a left pseudoinverse but not a right pseudoinverse.[14] In general, the precorrection of the nonlinearity in a LN cascade seems quite a challenging problem, which may compromise this predistortion approach.

Appendix 3.E CSI Knowledge

Throughout Chapters 2 and 3 we have assumed a precise knowledge of the CSI at the transmitter. In the SISO THP setup of Chapter 2 this translates into precise knowledge of the channel response $H_{\text{c}}(f)$ (see Figure 2.11), and therefore of the discrete-time equivalent transfer function $H(z)$. As for the multibeam THP setup of Chapter 3 (see Figure 3.4), full CSI knowledge amounts to perfect knowledge of the channel matrix $\boldsymbol{H}$. Such assumption allows to identify the performance of the proposed schemes under ideal conditions. Nevertheless, perfect acquisition of the CSI may not be possible in many cases. In this section we briefly describe possible techniques of channel estimation for both cases, and review how THP can be applied when only partial CSI is available.

The performance of precoded transmission when there are inaccuracies in the channel estimate has been studied e.g. in [150] (time-dispersive SISO THP) and [127] (MIMO case). The performance degrades consistently with the degree of mismatch between the estimated and the actual channel. The mismatches between true and estimated CSI can

[14] In other words, when $\boldsymbol{M}_{\text{BF}}$ is "tall" it is possible to find a matrix $\boldsymbol{M}_{\text{BF}}^{+}$ (left pseudoinverse) such that $\boldsymbol{M}_{\text{BF}}^{+} \cdot \boldsymbol{M}_{\text{BF}} = \boldsymbol{I}$, but not a matrix $\boldsymbol{M}_{\text{BF}}^{+,\text{r}}$ such that $\boldsymbol{M}_{\text{BF}} \cdot \boldsymbol{M}_{\text{BF}}^{+,\text{r}} = \boldsymbol{I}$.

be due to estimation and quantization errors, or time variations of the channel response. Channel estimation for both SISO and MISO setups must necessarily consider the use of DA schemes (e.g., pilot symbols or training sequences) to be used (at least) in the initialization phase of the transmission. This startup phase, during which THP cannot be yet applied, can be used to perform other tasks simultaneously, such as timing recovery or frame acquisition. In addition, a complementary calibration scheme will be required to aid in the frequency recovery, phase recovery and gain control tasks. Such calibration scheme would be useful not only for channel estimation but also for the beamforming task. Examples of calibration schemes adequate for these purposes can be found in [24, 162].

In the SISO case, if the transmission can be considered reciprocal (i.e., both transmission ends perceive the same channel channel response), channel estimation may be performed at one end, and then fed back to the other end. In this case, errors in the feedback transmission or in the quantization of the channel response may introduce degradation in the channel estimate, and should be taken into account. As a practical example, the 10GBASE-T standard [90] establishes a transmission startup phase during which both transmission ends exchange their channel estimates in the form of quantized taps.

Apart from the channel estimation required for the transmission start, some channel tracking scheme may be required in cases where the channel response varies with time. In this case, the channel estimates must be regularly updated by some channel tracking scheme which should incorporate as much useful information as possible, including the statistics of the channel coefficients and their variation. If the channel presents sufficiently fast variations, the tracking scheme may not be able to track such variations, and precoding may eventually fail. For this reason the application of THP transmission towards mobile receivers is in principle compromised.

Considering the CSI uncertainty caused by the quantized, possibly delayed feedback, several authors have proposed different precoding schemes designed to work under a certain degree of channel uncertainty (partial CSI knowledge), which exploit the statistical knowledge of the channel variations to improve the performance of classical. Examples of such schemes can be found in [37, 107] for the SISO setup, and [33, 34, 49, 87, 95] for the MIMO setup.

Chapter 4

SNR Estimation

Contents

4.1 Introduction

Estimation of the SNR at the receiver side is an important task in existing and emerging communication systems, as many of them require SNR knowledge in order to achieve different goals, such as power control or adaptive coding and modulation [4, 17], and soft decoding [153]. Whereas SNR estimation is a relatively easy task with simple modulation formats such as Quadrature Phase Shift Keying (QPSK), the increasing complexity of the modulation schemes used in modern systems poses significant challenges to current methods. Satellite communications provide a good example of a system where the traditionally used PSK modulations are leaving room to multilevel modulations such as the APSK family used e.g. within the DVB-S2 standard [54].

Existing SNR estimators can be classified according to a number of criteria. DA estimators can be used when the receiver has knowledge of the transmitted symbols, in contrast to NDA estimators, which do not require such knowledge. DA estimators can be used in DD mode by substituting the true transmitted symbols by the outputs of the decoder. Under a different classification, I/Q-based estimators make use of both the in-phase and quadrature components of the received signal, and thus require coherent detection; in contrast, Envelope-based (EVB) estimators only make use of the received signal magnitude, and thus can be applied even if the carrier phase has not been completely acquired. This is important in applications in which the SNR must be estimated even if its value is so low as to preclude accurate synchronization and decoding. Concerning the sampling rate of the received signal, most estimators operate on baud-sampled data. However, several estimators exist for oversampled data such as those covered in [131], and others which exploit different features of the signal such as properties of its moments [143] or the cyclostationarity of the received signal [136]. Most approaches focus on the SISO channel with AWGN and (quasi)static flat fading [23, 47, 52, 58, 66, 68, 71, 72, 102, 110, 111, 118, 131, 135, 144, 147–149, 166], although [159] and [167] address the static frequency selective channel and time-varying flat-fading channel cases respectively. Recently, SNR estimators have been proposed for multiantenna receivers [18, 152], which happen to be also applicable to multicarrier transmission.

The Maximum Likelihood (ML) approach has been previously applied in this context, for either Constant Modulus (CM) [46, 66, 102, 131, 166, 167] or multilevel constellations [47] (I/Q-based), [111] (EVB). Although ML estimators provide good statistical performance, they tend to be computationally intensive. This motivates the development of simpler, suboptimal approaches such as those based on the moments of the envelope of the received signal, here referred to as M_n (n is the order). These methods belong to the class of NDA EVB estimators, requiring neither accurate carrier recovery, nor knowledge of the transmitted symbols. This flexibility, together with implementation simplicity, makes these estimators attractive for practical applications.

Early examples of moments-based methods are two estimators originally proposed for CM constellations [23, 118] built respectively upon the first- and second-order moments (M_1M_2), and the second- and fourth-order moments (M_2M_4). More recently, [68] proposed a family of estimators for multilevel constellations based on pairs of moments, containing M_1M_2 and M_2M_4 as particular cases. As it turns out, the performance of M_1M_2 and M_2M_4 is close to the Cramér-Rao Lower Bound (CRB) for CM constellations [23, 131]. However, this is not so for multilevel constellations, for which the estimation variance considerably departs from the CRB as the SNR increases [68].

Recent efforts have been made in order to improve the quality of moments-based SNR estimators in these challenging settings. In [71], the observations are partitioned in different subsets corresponding to symbols of equal modulus, and then M_2M_4 is applied to one of these subsets. Although the resulting performance is good for sufficiently high SNR, it degrades substantially for low SNR due to errors in the partition step. Alternatively, a sixth-order statistics-based estimator (henceforth referred to as M_6) was proposed in [110] which makes use of M_2, M_4 and M_6, with a significant performance improvement over M_2M_4 for multilevel constellations at intermediate and high SNR. In addition, a scalar parameter allows to tune the estimator to any particular constellation. The use of higher-order statistics has also been considered in other related contexts, such as modulation classification (see [82], [45] and the survey in [51]).

Chapter structure This chapter addresses the NDA SNR estimation problem under the AWGN SISO channel model. This setting applies either to single carrier systems or to multicarrier systems in which the SNR is to be separately estimated at each carrier. The focus is on the estimation from unknown symbols belonging to multilevel constellations, such as the 16-APSK (two levels) and 32-APSK (three levels) found in the DVB-S2 standard, since the traditional estimators originally designed for PSK constellations fail for multilevel constellations. The contributions of this chapter can be divided into two main blocks:

- In Section 4.3 we develop a novel family of NDA EVB SNR estimators based on even-order moments of the received signal. The presented family of estimators turns out to generalize previous estimators based on even order moments, and features the interesting property that its estimators can be implemented without memory-costly look-up tables. We show how useful estimators can be built and optimized for improved performance with multilevel constellations. Two different estimator optimization criteria are proposed and seen to yield good estimators, especially for modulations with two and three levels. Interestingly, the proposed optimization criteria are readily applicable irrespective of the (even) statistical order of the moments used in the estimator. This general approach lays the foundations of a new framework for the development of new estimators, which is further generalized by introducing a family of estimators based on moments of any (possibly odd) order. This section is based on the work contained in [6].
- As second block, in Section 4.4 we study the problem of estimating the SNR from a frame composed of symbols of two or more constellations. Such case is commonly found in practice, for example in DVB-S2 frames, whose header and possible pilots

belong to a CM constellation, whereas the data symbols can belong to multilevel constellations. We propose and evaluate several (NDA) estimation schemes which use all the available information to yield improved estimates over more straightforward schemes. This section stems from the work originally presented in [9].

4.2 Review of SNR Estimators

We assume a quasistatic flat-fading channel model, with perfect timing and carrier synchronization, so the symbol-rate samples at the matched filter output are given by

$$r_k = \sqrt{S}a_k + n_k, \qquad k = 1, \ldots, K \tag{4.1}$$

where a_k are the complex-valued transmitted symbols, $\sqrt{S}$ is the unknown channel gain, and n_k are complex circular i.i.d. Gaussian noise samples with unknown variance N. The symbols are i.i.d., drawn from a constellation which is known to the receiver and has I different amplitudes, being R_i and P_i, respectively, the i-th amplitude and its associated probability ($i = 1, \ldots, I$). The *constellation moments* are denoted $c_p \doteq \mathrm{E}\left\{|a_k|^p\right\} = \sum_{i=1}^{I} P_i R_i^p$. Without loss of generality, an energy-normalized constellation is assumed, i.e., $c_2 = 1$.

If the symbols are known to the receiver (i.e., training symbols are available), then the DA ML estimator represents a good choice provided that coherent reception is available (I/Q-based estimator). The DA ML SNR estimator [66, 131], also known as SNORE by some authors [4], is given by

$$\hat{\rho}_{\text{DA-ML}} = \frac{\left[\dfrac{1}{K_t}\displaystyle\sum_{k=1}^{K_t} \mathrm{Re}\left\{r_k^* a_k\right\}\right]^2}{\dfrac{1}{K_t}\displaystyle\sum_{k=1}^{K_t} |r_k|^2 - \left[\dfrac{1}{K_t}\displaystyle\sum_{k=1}^{K_t} \mathrm{Re}\left\{r_k^* a_k\right\}\right]^2} \tag{4.2}$$

where K_t is the number of training symbols, and $\frac{1}{K_t}\sum_{k=1}^{K_t} |a_k|^2 = 1$. The bias and variance properties of this estimator have been studied in [40].

Moments-based SNR estimation is a useful alternative when the transmitted symbols and the carrier phase are not known. The estimators belonging to this family stem from the approach known as *method of moments* [101]. In our case, the goal is to find a function of the *sample moments* of the envelope,

$$\hat{M}_p \doteq \frac{1}{K}\sum_{k=1}^{K} |r_k|^p \tag{4.3}$$

from which an estimate of the SNR $\rho \doteq S/N$ or the *normalized SNR* $z \doteq \rho/(1+\rho)$ can be derived.[1] $\hat{M}_p$ is a consistent, unbiased estimator of the *true moment* of the envelope,

[1] $z \in (0,1)$ is a one-to-one transformation of $\rho \in (0, \infty)$, so that for estimation purposes it is equivalent to derive an estimator of z and then undo the transformation to find ρ.

given by[2] [68, 132]

$$\begin{aligned} M_p &\doteq \mathrm{E}\left\{|r_k|^p\right\} \\ &= N^{\frac{p}{2}} \sum_{i=1}^{I} P_i \Gamma\left(\frac{p}{2}+1\right) e^{-\rho R_i^2} {}_1F_1\left(\frac{p}{2}+1; 1; \rho R_i^2\right) \end{aligned} \tag{4.4}$$

with $\Gamma(\cdot)$ the gamma function and ${}_1F_1(\cdot;\cdot;\cdot)$ the confluent hypergeometric function. For p even, (4.4) can be seen to admit a simpler form which depends only on the even constellation moments up to order p and the two unknowns ρ, N [86, 110]:

$$M_{2n} = N^n \sum_{m=0}^{n} \frac{(n!)^2}{(n-m)!(m!)^2} c_{2m} \rho^m. \tag{4.5}$$

Note that (4.5) is a polynomial in ρ, and that in (4.4) and (4.5) the noise power N appears as a multiplicative factor. This latter fact allows to obtain pure functions of ρ through ratios of moments in which N vanishes. For example, [68] proposes the family of functions

$$f_{k,l}(\rho) \doteq \frac{M_k^l}{M_l^k}, \quad \text{for } k \neq l. \tag{4.6}$$

An SNR estimate can be then obtained from the sample moments by inversion of $f_{k,l}$:

$$\hat{\rho}_{k,l} = f_{k,l}^{-1}\left(\frac{\hat{M}_k^l}{\hat{M}_l^k}\right). \tag{4.7}$$

In [68] the authors state that $f_{k,l}$ in (4.6) is monotonic.[3] This property is a requirement for any function $f(\rho)$ from whose inverse an estimator is to be derived. The $\mathrm{M_1M_2}$ and $\mathrm{M_2M_4}$ estimators are obtained for $(k,l) = (2,1)$ and $(4,2)$ respectively.

4.2.1 Cramér-Rao Bounds for SNR Estimation

The CRB represents the lower variance floor of any unbiased estimator, so it is commonly used as a benchmark for assessing the competitiveness of the proposed estimators. The closed form of the CRB for DA I/Q-based estimation was derived in [131] for complex-valued flat-fading channels, and reads

$$\mathrm{CRB_{DA}} = \frac{1}{K}(2\rho + \rho^2). \tag{4.8}$$

This CRB, which has been recently shown to be also valid for time-varying channels [167], is independent of the modulation scheme. In turn, the CRB for the NDA case does depend on the modulation and, unfortunately, does not admit a closed form in general. The

[2] There is typographical error in Eqs. (3) and (4) of [68]. The complex noise power is defined therein as $N = 2\sigma^2$, but the factor 2 does not appear in the expression of the moments.

[3] Although no proof is provided in [68], monotonicity holds for all constellations and all pairs (k,l) we have tested. Moreover, we have observed that $f_{k,l}$ is monotonically decreasing when $k > l$, and increasing when $k < l$.

NDA I/Q-based CRB for BPSK and QPSK was derived in [3], and further discussed on [166, 167]. NDA I/Q-based CRBs for M-PSK, APSK and 16- and 32-QAM were independently presented in [69] and [47]. A general form for the NDA I/Q-based CRB for square QAM modulations has been presented in [19], and subsequently applied to derive the CRB for SNR estimation in multicarrier systems [20]. Recently, this bound has been shown to be independent of the knowledge of the carrier phase and frequency [73]

On the other hand, the NDA EVB CRB has been recently presented in [72] for M-PSK, APSK and 16-QAM modulations. In this section we focus on the EVB setup, so we will use the NDA EVB CRB as a benchmark when evaluating the performance of the proposed estimators.

As a rule of thumb, it can be said that the NDA CRB converges to its correspondent DA CRB for large SNRs, whereas for low SNRs the NDA CRB can be considerably larger than its DA counterpart. Furthermore, EVB CRBs are at least a factor of 2 larger than the corresponding I/Q-based CRBs. This last feature follows from the fact that the EVB approach only considers the magnitude but not the phase, which amounts to discarding half of the information contained in the samples.

4.3 Higher-Order-Statistics-Based Non-Data-Aided SNR Estimation

4.3.1 Motivation and Section Outline

The good performance of the sixth-order-based estimator of [110] for two-level constellations suggests that further improvements could be expected if higher-order moments are allowed in the construction of the estimator. With this in mind, this section investigates general higher-order approaches for the moments-based SNR estimation problem. The content of this section is detailed as follows:[4]

- In Section 4.3.1.1 we identify a novel family of estimators based on ratios of even-order moments which features an interesting property: all estimators in this family can be implemented via direct or iterative computations, thus avoiding the need for lookup tables.
- In Section 4.3.2 we propose a general estimator built upon a linear combination of members of the previous family, which generalizes the sixth-order approach of [110] to any (even) order. The weights of the linear combination can be tuned to each particular constellation and nominal SNR operation range.
- To illustrate the potential of this method, in Section 4.3.3 we develop a statistical analysis of the particular case in which moments of up to eighth order are allowed, whereas Section 4.3.4 presents two criteria for weight optimization. These approaches can be readily generalized to higher-order moments.

[4] Supporting *Mathematica* and *MATLAB* code for some of the algebraic derivations is available online at [146].

- As a result, we provide an eighth-order estimator with rather competitive performance for non-CM constellations, especially for two- and three-level constellations, as shown in Section 4.3.5.
- A generalization of the former even-order-based approach to possibly odd-order moments is proposed in Appendix 4.B, including its statistical analysis.

4.3.1.1 A New Family of NDA Moments-Based Estimators

The analytical inversion of the function $f_{k,l}$ given by (4.6) is intractable in general, as one may expect from the complexity of (4.4) (an exception is the $\mathrm{M_2M_4}$ estimate, which admits a closed-form expression, as we will see below). In such cases one must resort to precomputed LUTs, as in [23, 68] for the $\mathrm{M_1M_2}$ estimator. Although LUTs are computationally efficient, their storage requirements, which are proportional to the desired accuracy, become an issue in hardware implementations with small memory space. This motivates the search for SNR estimators which do not rely on LUTs.

Consider the following family of functions built upon even-order moments:

$$f(\rho) \doteq \left(\prod_{i=1}^{U} M_{2n_i}^{p_i} \right) \Big/ M_2^Q, \quad n_i, p_i \in \mathbb{N} \tag{4.9}$$

where $Q = \sum_{i=1}^{U} n_i p_i$ in order to ensure that the ratio (4.9) does not depend on N. We refer to $2Q$ as the *statistical order* of (4.9). We assume that M_2 does not appear in the numerator (i.e., $n_i > 1$ for all i). Note that (4.9) can be written in terms of the functions $f_{k,l}$ from (4.7) as $f(\rho) = \prod_{i=1}^{U} f_{2n_i,2}^{p_i/2}(\rho)$, so it inherits decreasing monotonicity from the $f_{k,l}$ family (since $2n_i > 2$ and $p_i > 0$). In addition, this function exhibits the following property, whose proof is given in the Appendix:

Property 1. *The ratio $f(\rho)$ as defined in* (4.9) *boils down to a polynomial in z of degree Q:*

$$f(\rho) = \sum_{k=0}^{Q} F_k \, z^k \doteq F(z). \tag{4.10}$$

Moreover, the coefficient F_1 in (4.10) *is always zero.*

Now, one can construct SNR estimators by equating $F(\hat{z})$ to the sample-moments version of (4.9), i.e.,

$$F(\hat{z}) = \left(\prod_{i=1}^{U} \hat{M}_{2n_i}^{p_i} \right) \Big/ \hat{M}_2^Q \tag{4.11}$$

and then solving for $\hat{z}$ in $(0, 1)$. Monotonicity of $f(\rho)$ guarantees uniqueness of the solution. This procedure is in fact a polynomial root-finding problem, which may well be implemented by means of LUTs. Nevertheless, other less memory-costly choices are possible. On one hand, there exist standard direct algebraic solutions for the roots of polynomials of degree 4 or less ($2Q \leq 8$); on the other hand, iterative polynomial root-finding algorithms can be used, regardless of the order.

The popular $\mathrm{M_2M_4}$ estimator is a particular case of the family (4.9): it is a function of M_4/M_2^2, which by Property 1 reduces to $F(z) = (c_4 - 2)z^2 + 2$. The estimate is hence given by the positive root of $F(\hat{z}) = \hat{M}_4/\hat{M}_2^2$:

$$\hat{z} = \sqrt{\frac{\hat{M}_4/\hat{M}_2^2 - 2}{c_4 - 2}}. \tag{4.12}$$

The ratio M_4/M_2^2 is the lowest-order ($2Q = 4$) member of this family. The next members, in growing statistical order, are M_6/M_2^3 ($2Q = 6$), M_4^2/M_2^4 and M_8/M_2^4 ($2Q = 8$), M_6M_4/M_2^5 and M_{10}/M_2^5 ($2Q = 10$), and so on.

4.3.2 A New Moments-Based Estimator

The appealing property of ratios of moments of the form (4.9) is that they can be reduced to a polynomial in z, so that estimator implementation becomes computationally simple. This applies as well to *any linear combination* of these ratios. This fact has been exploited in [110], where the linear combination $f(\rho) = M_6/M_2^3 - bM_4/M_2^2$ yields a sixth-order statistics-based estimate with improved behavior for multilevel constellations, as the weight b can be adjusted depending on the constellation. The good results in terms of variance attained by this estimator suggest that further improvements could be obtained by allowing higher-order ratios from the family (4.9) in the linear combination. Two questions arise at this point. First, how should the ratios featuring in the linear combination be selected? And second, once these ratios are somehow chosen, how should one select the weights?

Regarding the first question, there is no straightforward recipe for this "basis selection" problem. In practice, implementation complexity and finite precision effects would favor the selection of lower-order moments. On the other hand, including higher-order terms may be beneficial in terms of estimation bias and/or variance. Therefore, a tradeoff must be reached balancing these two conflicting goals. As for how to adjust the weights, it is sensible to optimize them in terms of statistical performance. This issue will be discussed in Section 4.3.4.

Since the fourth- and sixth-order combinations have been already analyzed in [68] and [110] respectively, henceforth we focus on the weighted linear combination of *all* ratios up to order eight (EOS stands for *eighth-order statistics*):

$$f_{\text{EOS}}(\rho) \doteq \beta\frac{M_4}{M_2^2} + \gamma\frac{M_6}{M_2^3} + \delta\frac{M_4^2}{M_2^4} + \epsilon\frac{M_8}{M_2^4} \tag{4.13}$$

for some weights $\boldsymbol{\alpha} = [\beta \;\; \gamma \;\; \delta \;\; \epsilon]^{\mathrm{T}} \in \mathbb{R}^4$. Note that $f_{\text{EOS}}(\rho)$ includes as particular cases $\mathrm{M_2M_4}$ (by setting $\beta \neq 0$, $\gamma = \delta = \epsilon = 0$) and $\mathrm{M_6}$ ($\beta = -b$, $\gamma = 1$, and $\delta = \epsilon = 0$). For further reference we also define the vector $\boldsymbol{m} \doteq [M_2 \;\; M_4 \;\; M_6 \;\; M_8]^{\mathrm{T}} \in \mathbb{R}^4$ containing the true moments appearing in (4.13), and the function $h_{\text{EOS}}(\boldsymbol{m}) \doteq f_{\text{EOS}}(\rho)$ resulting from taking $\boldsymbol{m}$ as the independent variable in (4.13).

After applying the variable change $\rho = z/(1-z)$, one arrives at

$$F_{\text{EOS}}(z) = F_4\, z^4 + F_3\, z^3 + F_2\, z^2 + F_0 \tag{4.14a}$$
$$= z^2(F_4\, z^2 + F_3\, z + F_2) + F_0 \tag{4.14b}$$

where the coefficients F_k are linear in the weights, and are given by

$$F_4 = \delta(c_4 - 2)^2 + \epsilon[72(c_4 - 1) - 16c_6 + c_8] \tag{4.15a}$$
$$F_3 = (\gamma + 16\epsilon)(12 - 9c_4 + c_6) \tag{4.15b}$$
$$F_2 = (\beta + 9\gamma + 4\delta + 72\epsilon)(c_4 - 2) \tag{4.15c}$$
$$F_0 = 2(\beta + 3\gamma + 2\delta + 12\epsilon). \tag{4.15d}$$

The estimator is found by solving for $\hat{z}$ in

$$F_{\text{EOS}}(\hat{z}) = \beta\frac{\hat{M}_4}{\hat{M}_2^2} + \gamma\frac{\hat{M}_6}{\hat{M}_2^3} + \delta\frac{\hat{M}_4^2}{\hat{M}_2^4} + \epsilon\frac{\hat{M}_8}{\hat{M}_2^4} = h_{\text{EOS}}(\hat{\boldsymbol{m}}) \tag{4.16}$$

where $\hat{\boldsymbol{m}} \doteq [\hat{M}_2 \;\; \hat{M}_4 \;\; \hat{M}_6 \;\; \hat{M}_8]^{\text{T}} \in \mathbb{R}^4$ is the vector of sample moments. This amounts to finding the roots of the quartic polynomial $F_{\text{EOS}}(\hat{z}) - h_{\text{EOS}}(\hat{\boldsymbol{m}})$, in which only the independent term $F_0 - h_{\text{EOS}}(\hat{\boldsymbol{m}})$ depends on the observations. The roots of quartic polynomials can be algebraically found, although the procedure is somewhat intricate (see e.g. [1, pp.18-19]). Alternatively, one can resort to root-finding algorithms. For example, (4.14b) suggests the following iterative rule:

$$\hat{z}^{(0)} = 1, \qquad \hat{z}^{(n+1)} = \sqrt{\frac{h_{\text{EOS}}(\hat{\boldsymbol{m}}) - F_0}{F_4(\hat{z}^{(n)})^2 + F_3\,\hat{z}^{(n)} + F_2}}. \tag{4.17}$$

Note that the polynomial $F_{\text{EOS}}(z) - h_{\text{EOS}}(\boldsymbol{m})$ should have a unique root in $(0, 1)$, i.e., $f_{\text{EOS}}(\rho)$ should be monotonic. Given that the roots depend on the constellation moments c_p as well as on the weights $\boldsymbol{\alpha}$, one might ask whether conditions on $\boldsymbol{\alpha}$ can be given ensuring monotonicity of $f_{\text{EOS}}(\rho)$; the answer to this question is not trivial. In the next sections we show how $\boldsymbol{\alpha}$ can be optimized in terms of statistical performance, for a given constellation and SNR operation range. Let us advance that the optimal weights seem to yield monotonic functions $f_{\text{EOS}}(\rho)$ in the majority of tested cases, with some exceptions to be discussed in sections 4.3.4 and 4.3.5.

To close this section, let us remark that this EOS approach can be readily generalized to higher-order linear combinations of the form similar to (4.13). The presence of statistics of tenth or greater order will lead to polynomials $F(z)$ of degree 5 or greater. In such cases, direct computation of the root is not possible, but a root-finding rule of the form (4.17) always exists, since $F_1 = 0$ by Property 1.

4.3.3 Statistical Analysis

Next we present a small-error analysis to obtain approximate expressions for the variance and bias of the EOS-based estimate introduced in Section 4.3.2. Our goal is to expose the dependence of these performance measures with the weight vector $\boldsymbol{\alpha}$, in order to address weight optimization in Section 4.3.4. We note that the analytical approach of this section can be readily generalized to estimators based on higher-order ratios of the form (4.9).

4.3.3.1 Variance

From (4.16), the estimator is given by $\hat{z} = F_{\text{EOS}}^{-1}(h_{\text{EOS}}(\hat{\boldsymbol{m}})) \doteq g(\hat{\boldsymbol{m}})$ and $\hat{\rho} = \hat{z}/(1-\hat{z}) = g(\hat{\boldsymbol{m}})/(1-g(\hat{\boldsymbol{m}})) \doteq t(\hat{\boldsymbol{m}})$, where g is implicitly given by $F_{\text{EOS}}(g) - h_{\text{EOS}}(\hat{\boldsymbol{m}}) = 0$. Following a standard procedure (see e.g. [101, Sec. 9.5]) we consider a first-order Taylor expansion of the estimator $\hat{\rho}$ about $\hat{\boldsymbol{m}} = \boldsymbol{m}$, which yields the approximation $\hat{\rho} \approx \rho + \boldsymbol{v}^{\mathrm{T}}(\hat{\boldsymbol{m}} - \boldsymbol{m})$, with $\boldsymbol{v} \doteq \nabla t|_{\hat{\boldsymbol{m}}=\boldsymbol{m}}$. Therefore,

$$\operatorname{Var}\{\hat{\rho}\} \approx \boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} \tag{4.18}$$

with $\boldsymbol{C}$ the covariance matrix of $\hat{\boldsymbol{m}}$, whose elements are given by[5]

$$[\boldsymbol{C}]_{ij} = (M_{2(i+j)} - M_{2i}M_{2j})/K, \quad i,\, j \in \{1,2,3,4\}. \tag{4.19}$$

The computation of (4.18) is outlined in the Appendix; eventually, one arrives at

$$\operatorname{Var}\{\hat{\rho}\} \approx \frac{1}{K}\frac{A^{(\text{Var})}(\rho)}{B^{(\text{Var})}(\rho)} = \frac{1}{K}\frac{\sum_{n=0}^{10} A_n^{(\text{Var})}\rho^n}{\sum_{n=2}^{6} B_n^{(\text{Var})}\rho^n} \tag{4.20}$$

with $A^{(\text{Var})}(\rho) \doteq \sum_{n=0}^{10} A_n^{(\text{Var})}\rho^n$ and $B^{(\text{Var})}(\rho) \doteq \sum_{n=2}^{6} B_n^{(\text{Var})}\rho^n$. The coefficients of the former polynomials turn out to be quadratic in $\boldsymbol{\alpha}$:

$$A_n^{(\text{Var})} = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}_n^{(\text{Var})}\boldsymbol{\alpha} \quad \Rightarrow \quad A^{(\text{Var})}(\rho) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}^{(\text{Var})}(\rho)\boldsymbol{\alpha} \tag{4.21}$$

$$B_n^{(\text{Var})} = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{B}_n^{(\text{Var})}\boldsymbol{\alpha} \quad \Rightarrow \quad B^{(\text{Var})}(\rho) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{B}^{(\text{Var})}(\rho)\boldsymbol{\alpha} \tag{4.22}$$

where the 4×4 matrices $\boldsymbol{A}_n^{(\text{Var})}$, $\boldsymbol{B}_n^{(\text{Var})}$ are functions of the constellation moments only, and $\boldsymbol{A}^{(\text{Var})}(\rho) = \sum_{n=0}^{10} \boldsymbol{A}_n^{(\text{Var})}\rho^n$, $\boldsymbol{B}^{(\text{Var})}(\rho) = \sum_{n=2}^{6} \boldsymbol{B}_n^{(\text{Var})}\rho^n$. Note that (4.20) predicts that the variance is $\mathcal{O}(\rho^4)$ at high SNR, similarly to previous results for other estimators based on ratios of moments [68, 110].

The matrices in (4.21) and (4.22) exhibit interesting properties regardless of the constellation, which we summarize next for further reference. The proofs can be found in the Appendix.

Property 2. *The 4×4 matrix $\boldsymbol{A}^{(\text{Var})}(\rho)$ has rank not exceeding three.*

Property 3. *The 4×4 matrix $\boldsymbol{A}_{10}^{(\text{Var})}$ has rank not exceeding three. In addition, rank $\left(\boldsymbol{A}_{10}^{(\text{Var})}\right) \leq \min\{3, I-1\}$ (recall that I is the number of amplitude levels of the constellation).*

Property 4. *The 4×4 matrix $\boldsymbol{B}^{(\text{Var})}(\rho)$ is of the form $\boldsymbol{B}^{(\text{Var})}(\rho) = \boldsymbol{b}^{(\text{Var})}(\rho)(\boldsymbol{b}^{(\text{Var})}(\rho))^{\mathrm{T}}$ for some 4×1 nonzero vector $\boldsymbol{b}^{(\text{Var})}(\rho)$, and therefore its rank is one.*

Property 5. *The 4×4 matrix $\boldsymbol{B}_6^{(\text{Var})}$ is of the form $\boldsymbol{B}_6^{(\text{Var})} = \boldsymbol{b}_6^{(\text{Var})}(\boldsymbol{b}_6^{(\text{Var})})^{\mathrm{T}}$ for some 4×1 nonzero vector $\boldsymbol{b}_6^{(\text{Var})}$, and therefore its rank is one.*

[5] Based on the independence of the envelope samples, it is straightforward to show that the variance of the sample moment $\hat{M}_p$ ($p \in \mathbb{N}$) is given by $\mathrm{E}\{(\hat{M}_p - M_p)^2\} = (M_{2p} - M_p^2)/K$. Similarly, the covariance of two sample moments $\hat{M}_p$, $\hat{M}_q$ ($p, q \in \mathbb{N}$) is seen to be $\mathrm{E}\{(\hat{M}_p\hat{M}_q - M_pM_q)^2\} = (M_{p+q} - M_pM_q)/K$.

4.3.3.2 Bias

In order to obtain an approximation for the estimation bias, we use a second-order Taylor expansion of $\hat{\rho}$ about $\hat{\boldsymbol{m}} = \boldsymbol{m}$: $\hat{\rho} \approx \rho + \boldsymbol{v}^{\mathrm{T}}(\hat{\boldsymbol{m}} - \boldsymbol{m}) + \frac{1}{2}(\hat{\boldsymbol{m}} - \boldsymbol{m})^{\mathrm{T}}\boldsymbol{H}(\hat{\boldsymbol{m}} - \boldsymbol{m})$, where $\boldsymbol{H}$ is the Hessian matrix of $t(\hat{\boldsymbol{m}})$ evaluated at $\hat{\boldsymbol{m}} = \boldsymbol{m}$. Straightforward algebra yields the approximation

$$\mathrm{Bias}\,\{\hat{\rho}\} \approx \frac{1}{2}\mathrm{Tr}\,(\boldsymbol{H}\,\boldsymbol{C})\,. \tag{4.23}$$

After a few steps, as outlined in the Appendix, one finds

$$\begin{aligned}
\mathrm{Bias}\,\{\hat{\rho}\} &\approx \frac{1}{K}\,L(\rho) + J(\rho)\,\mathrm{Var}\,\{\hat{\rho}\} && \text{(4.24a)}\\
&= \frac{1}{K}\frac{A^{(L)}(\rho)}{B^{(L)}(\rho)} + \frac{A^{(J)}(\rho)}{B^{(L)}(\rho)}\,\frac{1}{K}\,\frac{A^{(\mathrm{Var})}(\rho)}{B^{(\mathrm{Var})}(\rho)} && \text{(4.24b)}\\
&= \frac{1}{K}\frac{\sum_{n=0}^{12} A_n^{(\mathrm{Bias})}\rho^n}{\sum_{n=3}^{9} B_n^{(\mathrm{Bias})}\rho^n} = \frac{1}{K}\frac{A^{(\mathrm{Bias})}(\rho)}{B^{(\mathrm{Bias})}(\rho)} && \text{(4.24c)}
\end{aligned}$$

where $L(\rho) \doteq A^{(L)}(\rho)/B^{(L)}(\rho)$, with $A^{(L)}(\rho) = \sum_{n=0}^{6} A_n^{(L)}\rho^n$, $B^{(L)}(\rho) = \sum_{n=1}^{4} B_n^{(L)}\rho^n$, and $J(\rho) \doteq A^{(J)}(\rho)/B^{(L)}(\rho)$, with $A^{(J)}(\rho) = \sum_{n=0}^{3} A_n^{(J)}\rho^n$ (note that $L(\rho)$ and $J(\rho)$ share the same denominator). Furthermore, the former polynomials and their coefficients are linear in $\boldsymbol{\alpha}$:

$$\begin{aligned}
A_n^{(L)} = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{a}_n^{(L)} \quad &\Rightarrow \quad A^{(L)}(\rho) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{a}^{(L)}(\rho) && \text{(4.25)}\\
B_n^{(L)} = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{b}_n^{(L)} \quad &\Rightarrow \quad B^{(L)}(\rho) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{b}^{(L)}(\rho) && \text{(4.26)}\\
A_n^{(J)} = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{a}_n^{(J)} \quad &\Rightarrow \quad A^{(J)}(\rho) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{a}^{(J)}(\rho) && \text{(4.27)}
\end{aligned}$$

where the 4×1 vectors $\boldsymbol{a}_n^{(L)}$, $\boldsymbol{b}_n^{(L)}$, $\boldsymbol{a}_n^{(J)}$ are functions of the constellation moments only, and $\boldsymbol{a}^{(L)}(\rho) = \sum_{n=0}^{6} \boldsymbol{a}_n^{(L)}\rho^n$, $\boldsymbol{b}^{(L)}(\rho) = \sum_{n=1}^{4} \boldsymbol{b}_n^{(L)}\rho^n$, $\boldsymbol{a}^{(J)}(\rho) = \sum_{n=0}^{3} \boldsymbol{a}_n^{(J)}\rho^n$. Therefore, $A^{(\mathrm{Bias})}$ and $B^{(\mathrm{Bias})}$ are *cubic* in $\boldsymbol{\alpha}$. Note that, at high SNR, $L(\rho)$ is $\mathcal{O}(\rho^2)$ whereas $J(\rho)$ is $\mathcal{O}(\rho^{-1})$, so (4.24) predicts that the bias is $\mathcal{O}(\rho^3)$ at high SNR.

4.3.3.3 MSE

Recalling that the minimum square error (MSE) of an estimator is given by $\mathrm{MSE}\,\{\hat{\rho}\} = \mathrm{Var}\,\{\hat{\rho}\} + \mathrm{Bias}^2\{\hat{\rho}\}$, and using (4.20) and (4.24), it is found that

$$\mathrm{MSE}\,\{\hat{\rho}\} \approx \frac{1}{K}\frac{\sum_{n=0}^{10} A_n^{(\mathrm{Var})}\rho^n}{\sum_{n=2}^{6} B_n^{(\mathrm{Var})}\rho^n} + \frac{1}{K^2}\left(\frac{\sum_{n=0}^{12} A_n^{(\mathrm{Bias})}\rho^n}{\sum_{n=3}^{9} B_n^{(\mathrm{Bias})}\rho^n}\right)^2. \tag{4.28}$$

Therefore the MSE is the sum of two terms, the first of which (the variance) is $\mathcal{O}(\rho^4)$ at high SNR and inversely proportional to K, and the second one (the squared bias) is $\mathcal{O}(\rho^6)$ at high SNR and inversely proportional to K^2. The dependence of the second term of (4.28) with $\boldsymbol{\alpha}$ is thus of *sixth order*.

4.3.4 Weight Optimization

We focus now on the selection of the weight vector $\boldsymbol{\alpha}$. Note that the four weights in $\boldsymbol{\alpha}$ only provide in fact *three* degrees of freedom, since the solutions to (4.16) are invariant to scalings in $\boldsymbol{\alpha}$. This is further emphasized by the fact that the approximations (4.20) and (4.24) for the variance and the bias are also invariant under this operation. In general, the number of degrees of freedom in this kind of estimators equals the number of ratios of moments in the linear combination minus one; for instance, M_6 has one degree of freedom, whereas $\mathrm{M}_2\mathrm{M}_4$ has none. We present two weight selection procedures, according to two different optimization criteria. Ideally, one would like to obtain an unbiased estimator with variance close to the theoretical limit dictated by the Cramér-Rao Bound (CRB). This cannot be achieved with the limited degrees of freedom available, and therefore one should settle for a less ambitious goal. The question that arises then is how to trade off bias and variance, and where (i.e., in which SNR operating region).

For example, in [110] the only free parameter in M_6 was optimized to yield minimum variance in the high SNR region. This approach can be generalized to the case in which several degrees of freedom are available, taking the bias into account as well; we refer to this criterion as "C1". Yet other approaches are possible. For instance, systems using adaptive coding and modulation commonly use each particular constellation within a limited SNR range (see e.g. [54]). It seems then reasonable to optimize the performance for some nominal SNR within that range, with the hope that the estimator will still perform well in a neighborhood of this nominal value. We will refer to this criterion as "C2". Note that either C1 or C2 could reasonably be applied to any estimator derived from a linear combination of the form of (4.13).

4.3.4.1 Criterion C1: Weight Optimization for High SNR

Asymptotically as $\rho \to \infty$, all polynomial divisions $A^{(\cdot)}(\rho)/B^{(\cdot)}(\rho)$ in (4.20) and (4.24) can be accurately approximated by the first few terms of the quotient polynomial:

$$\begin{aligned}
\mathrm{Var}\,\{\hat{\rho}\} &\approx \frac{1}{K}\frac{A^{(\mathrm{Var})}(\rho)}{B^{(\mathrm{Var})}(\rho)} \doteq \frac{1}{K}C(\rho) \\
&\approx \frac{1}{K}\left(C_4\,\rho^4 + C_3\,\rho^3 + C_2\,\rho^2 + C_1\,\rho + C_0\right) && (4.29)\\
L(\rho) &\approx L_2\,\rho^2 + L_1\,\rho + L_0 && (4.30)\\
J(\rho) &\approx J_{-1}\,\rho^{-1} + J_{-2}\,\rho^{-2} + J_{-3}\,\rho^{-3}. && (4.31)
\end{aligned}$$

It makes sense now to use the available degrees of freedom to minimize (or, if possible, cancel) the magnitude of the highest-order coefficients of the variance and the terms of the bias.[6] As for the bias, see (4.24a), it seems convenient to separately minimize $L(\rho)$, $J(\rho)$ and the variance. In this regard, it must be noted that $J_{-1} = A_3^{(J)}/B_4^{(L)} = 2$ independently

[6] Alternatively, one could attempt to minimize the highest-order coefficients of the high-SNR approximation of the MSE. We do not follow this approach, however, in view of the sixth-order dependence with the weights of some of the terms in (4.28).

of $\boldsymbol{\alpha}$ and the constellation. This suggests that the minimization efforts should focus on $L(\rho)$ and the variance. Regarding $L(\rho)$, its highest-order coefficient is given by

$$L_2 = \frac{A_6^{(L)}}{B_4^{(L)}} = \frac{\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{a}_6^{(L)}}{\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{b}_4^{(L)}}. \tag{4.32}$$

Observe that unless $\boldsymbol{a}_6^{(L)}$ and $\boldsymbol{b}_4^{(L)}$ are collinear (which is not the case with typical constellations), (4.32) can be made zero. Thus, we propose the general principle for weight optimization in the high SNR region:

Criterion C1: Spend one degree of freedom in $\boldsymbol{\alpha}$ to achieve $L_2 = 0$. Spend the remaining degrees of freedom in order to minimize C_4. If C_4 can be made zero and there are degrees of freedom to spare, then proceed to minimize C_3, and so on.

With this approach, $L(\rho)$ becomes $\mathcal{O}(\rho)$. Note that $\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{a}_6^{(L)} = 0$ implies that $\boldsymbol{\alpha} = \boldsymbol{F}\bar{\boldsymbol{\alpha}}$ for some 3×1 vector $\bar{\boldsymbol{\alpha}}$, where $\boldsymbol{F}$ is a 4×3 matrix satisfying $\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F} = \boldsymbol{I}$ and $\boldsymbol{F}^{\mathrm{T}} \boldsymbol{a}_6^{(L)} = \boldsymbol{0}$. The columns of $\boldsymbol{F}$ can be selected as the three eigenvectors associated to the three nonzero eigenvalues of the projection matrix

$$\boldsymbol{W} \doteq \boldsymbol{I} - \frac{\boldsymbol{a}_6^{(L)} (\boldsymbol{a}_6^{(L)})^{\mathrm{T}}}{||\boldsymbol{a}_6^{(L)}||^2}. \tag{4.33}$$

Now, the highest-order coefficient of the variance is given by [cf. (4.20)–(4.22) and Property 5]

$$C_4 = \frac{A_{10}^{(\mathrm{Var})}}{B_6^{(\mathrm{Var})}} = \frac{\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{A}_{10}^{(\mathrm{Var})} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{b}_6^{(\mathrm{Var})} (\boldsymbol{b}_6^{(\mathrm{Var})})^{\mathrm{T}} \boldsymbol{\alpha}}. \tag{4.34}$$

Thus, for $\boldsymbol{\alpha} = \boldsymbol{F}\bar{\boldsymbol{\alpha}}$, C_4 becomes

$$C_4 = \frac{\bar{\boldsymbol{\alpha}}^{\mathrm{T}} \boldsymbol{P} \bar{\boldsymbol{\alpha}}}{\bar{\boldsymbol{\alpha}}^{\mathrm{T}} \boldsymbol{q}\boldsymbol{q}^{\mathrm{T}} \bar{\boldsymbol{\alpha}}} \tag{4.35}$$

where $\boldsymbol{P} \doteq \boldsymbol{F}^{\mathrm{T}} \boldsymbol{A}_{10}^{(\mathrm{Var})} \boldsymbol{F}$ and $\boldsymbol{q} \doteq \boldsymbol{F}^{\mathrm{T}} \boldsymbol{b}_6^{(\mathrm{Var})}$. Note that (4.35) is a generalized Rayleigh quotient with a rank-one denominator matrix. It can be checked that the solution $\bar{\boldsymbol{\alpha}}_\star$ to the minimization of (4.35) (subject to $\bar{\boldsymbol{\alpha}}^{\mathrm{T}} \boldsymbol{q} \neq \boldsymbol{0}$) is as follows:

- If $\boldsymbol{P}$ is invertible, then $\bar{\boldsymbol{\alpha}}_\star = \boldsymbol{P}^{-1}\boldsymbol{q}$. The minimized value of C_4 is strictly positive.
- If $\boldsymbol{P}$ is singular and $\mathrm{rank}\,([\boldsymbol{P}\ \boldsymbol{q}]) = \mathrm{rank}\,(\boldsymbol{P})$, then there exist infinitely many solutions, but all of them yield the same minimized value of C_4, which again is strictly positive. One of these solutions is $\bar{\boldsymbol{\alpha}}_\star = \boldsymbol{P}^{+}\boldsymbol{q}$, where $\boldsymbol{P}^{+}$ denotes the pseudoinverse of $\boldsymbol{P}$.
- If $\boldsymbol{P}$ is singular and $\mathrm{rank}\,([\boldsymbol{P}\ \boldsymbol{q}]) > \mathrm{rank}\,(\boldsymbol{P})$, then C_4 can be made zero. The solutions can be given in terms of the Singular Value Decomposition (SVD) of $\boldsymbol{P}$:

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{V} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \\ & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}^{\mathrm{T}} \\ \boldsymbol{V}^{\mathrm{T}} \end{bmatrix} \tag{4.36}$$

where $\boldsymbol{\Sigma}$ is positive definite, and $\boldsymbol{U}$, $\boldsymbol{V}$ have orthonormal columns. Then all the vectors $\bar{\boldsymbol{\alpha}}_\star$ such that $\boldsymbol{P}\bar{\boldsymbol{\alpha}}_\star = \boldsymbol{0}$ and $\bar{\boldsymbol{\alpha}}_\star^{\mathrm{T}}\boldsymbol{q} \neq \boldsymbol{0}$ are given (up to a scaling) by

$$\bar{\boldsymbol{\alpha}}_\star = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{q} + \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\left(\boldsymbol{I} - \frac{\boldsymbol{q}\boldsymbol{q}^{\mathrm{T}}}{||\boldsymbol{V}^{\mathrm{T}}\boldsymbol{q}||^2}\right)\boldsymbol{V}\boldsymbol{s} \tag{4.37}$$

where $\boldsymbol{s}$ is arbitrary.

Let $r \doteq \mathrm{rank}\,(\boldsymbol{A}_{10}^{(\mathrm{Var})})$, and note that $\mathrm{rank}\,(\boldsymbol{P}) \leq r$. Although in general $\mathrm{rank}\,(\boldsymbol{P})$ could be strictly smaller than r, for all constellations tested it holds that $\mathrm{rank}\,(\boldsymbol{P}) = r$. Also, recall that $r \leq \min\{3, I-1\}$ by Property 3. These facts imply that, depending on the number I of amplitude levels in the constellation, one has the following:

- If $I = 1$ (CM constellations), then C_4 is automatically zero.
- If $I = 2$, then C_4 can be made zero spending just one degree of freedom.
- If $I = 3$, then C_4 can be made zero spending two degrees of freedom.
- If $I > 3$, then C_4 cannot be made zero.

Thus, estimation performance is expected to degrade as the number of levels in the constellation increases. More details are given on each specific case (in terms of I) in Section 4.3.5. For reference, Table 4.1 lists the optimal weights $\boldsymbol{\alpha}_\star = \boldsymbol{F}\bar{\boldsymbol{\alpha}}_\star$ under Criterion C1 for a number of practical multilevel constellations.[7]

Criterion C1 provides valid estimators for almost all practical constellations, in the sense that the resulting function $f_{\mathrm{EOS}}(\rho)$ in (4.13) is monotonic; the performance obtained is good in the medium to high SNR range. The only exception we have found is the 16-QAM constellation, for which the weights provided by Criterion C1 cause the MSE to exhibit a sharp peak at $\rho \approx 19$ dB (the denominators $\boldsymbol{\alpha}_\star^{\mathrm{T}}\boldsymbol{b}^{(\mathrm{Var})}(\rho)$ and $\boldsymbol{\alpha}_\star^{\mathrm{T}}\boldsymbol{b}^{(L)}(\rho)$ become zero for this SNR value). Thus, Criterion C1 is not well suited to 16-QAM.

4.3.4.2 Criterion C2: Weight Optimization for a Nominal SNR Value

Considering a target SNR ρ_0 and using Property 4, one can write (4.24) and (4.20) as

$$\mathrm{Bias}\,\{\hat{\rho}\}|_{\rho=\rho_0} \approx \frac{1}{K}L(\rho_0) + J(\rho_0)\ \mathrm{Var}\,\{\hat{\rho}\}|_{\rho=\rho_0} \tag{4.38}$$

$$= \frac{1}{K}\frac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{a}^{(L)}(\rho_0)}{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{b}^{(L)}(\rho_0)} + \frac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{a}^{(J)}(\rho_0)}{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{b}^{(L)}(\rho_0)}\ \mathrm{Var}\,\{\hat{\rho}\}|_{\rho=\rho_0} \tag{4.39}$$

$$\mathrm{Var}\,\{\hat{\rho}\}|_{\rho=\rho_0} \approx \frac{1}{K}\frac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}^{(\mathrm{Var})}(\rho_0)\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\mathrm{T}}\,\boldsymbol{b}^{(\mathrm{Var})}(\rho_0)\,[\boldsymbol{b}^{(\mathrm{Var})}(\rho_0)]^{\mathrm{T}}\,\boldsymbol{\alpha}}. \tag{4.40}$$

Therefore, it is natural to consider the following way of selecting the weights:

Criterion C2: Given a target SNR value ρ_0, minimize the approximate variance (4.40) in terms of $\boldsymbol{\alpha}$, under the constraint that $L(\rho_0)$ in (4.38) be zero.

[7] The *Mathematica* and *MATLAB* code used to derive these optimal weights is available online at [146].

Table 4.1: Optimal weights ($\epsilon = 1$) under Criterion C1. APSK constellations are as defined in [54] for the specified code rates

Constellation	I	β	γ	δ
16-APSK (2/3)	2	5.9396	−2.8400	−1.4325
16-APSK (3/4)	2	6.0768	−2.8769	−1.4572
16-APSK (4/5)	2	6.1331	−2.8918	−1.4676
16-APSK (5/6)	2	6.1637	−2.8999	−1.4734
16-APSK (8/9)	2	6.2306	−2.9173	−1.4860
16-APSK (9/10)	2	6.2522	−2.9229	−1.4901
32-APSK (3/4)	3	12.3187	−3.8576	−2.7445
32-APSK (4/5)	3	12.0919	−3.8541	−2.7024
32-APSK (5/6)	3	11.9647	−3.8527	−2.6762
32-APSK (8/9)	3	11.6972	−3.8501	−2.6102
32-APSK (9/10)	3	11.6684	−3.8498	−2.6027
16-QAM	3	9.9411	−5.28	−0.2807
32-QAM	5	10.2400	−3.8552	−2.1227
64-QAM	9	10.7991	−4.3509	−1.8525
128-QAM	16	10.7081	−4.1297	−2.0170
256-QAM	32	10.4846	−4.4251	−1.6505

This criterion amounts to minimizing a generalized Rayleigh quotient with a rank-one denominator matrix, subject to a linear constraint. This is analogous to the case encountered for Criterion C1. The only difference is that now the objective function (4.40) cannot be made zero, as it corresponds to the estimator variance (and not just one coefficient in its Taylor series expansion). Therefore, the solution is given (up to a scaling) by

$$\boldsymbol{\alpha}_\star(\rho_0) = \boldsymbol{F}\left(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{A}^{(\mathrm{Var})}(\rho_0)\boldsymbol{F}\right)^{-1}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{b}^{(\mathrm{Var})}(\rho_0) \tag{4.41}$$

where the columns of the 4×3 matrix $\boldsymbol{F}$ are now given by the three eigenvectors associated to the three nonzero eigenvalues of

$$\boldsymbol{W} = \boldsymbol{I} - \frac{\boldsymbol{a}^{(L)}(\rho_0)[\boldsymbol{a}^{(L)}(\rho_0)]^{\mathrm{T}}}{||\boldsymbol{a}^{(L)}(\rho_0)||^2}. \tag{4.42}$$

This SNR-dependent solution $\boldsymbol{\alpha}_\star(\rho_0)$ proves in most cases competitive within a few dBs about ρ_0. Nevertheless, one must evaluate its performance in the entire SNR range for each particular constellation and target SNR, as $\boldsymbol{\alpha}_\star(\rho_0)$ could yield an invalid estimator in some SNR interval (i.e. $f_{\mathrm{EOS}}(\rho)$ in (4.13) could turn out to be non-monotonic). Of all constellations tested, only the 16-APSK family suffers from this problem, but only within some limited SNR intervals, as shown in the next section.

Note that in (4.38)-(4.39) the factor $J(\rho_0)$ has the same form as $L(\rho_0)$, and therefore, in contrast to Criterion C1, it could be made zero using one degree of freedom in $\boldsymbol{\alpha}$.

This suggests an alternative criterion, in which one of the degrees of freedom used for minimizing the variance under Criterion C2 would be used instead for canceling $J(\rho_0)$ (and therefore completely canceling the bias at ρ_0). However, the solutions provided by this alternative criterion turn out to be problematic in practice ($f_{\text{EOS}}(\rho)$ is not monotonic for most constellations; sharp peaks appear in bias and variance at certain SNRs), and hence it will not be further considered here.

4.3.5 Performance Results

The performance of the proposed estimators strongly depends on the number of levels of the constellation. Next we discuss the results achieved with criteria C1 and C2 with respect to existing estimates of the same kind: M_2M_4 and M_6. We show results for the dependence of the bias with the SNR, and the dependence of the SNR-normalized MSE with the SNR and the sample size K. The SNR-normalized MSE is defined as $\text{NMSE}\{\hat{\rho}\} \doteq \text{MSE}\{\hat{\rho}\}/\rho^2$. The NDA-EVB CRB (numerically evaluated as in [72]) is provided as benchmark. Analytical results are completed with empirical results obtained through simulations, in which each point was averaged over 10 000 realizations.

4.3.5.1 CM constellations

CM constellations (for which $c_p = 1$, for all p) constitute a special case, as it turns out that $A_6^{(L)} = 0$ whereas $A_5^{(L)} \neq 0$ (with $B_4^{(L)} \neq 0$), and that $A_{10}^{(\text{Var})} = A_9^{(\text{Var})} = 0$ whereas $A_8^{(\text{Var})} \neq 0$ (with $B_6^{(\text{Var})} \neq 0$). Therefore, the bias and variance are respectively linear and quadratic in ρ for general $\boldsymbol{\alpha}$. Interestingly, the highest-order coefficients of $L(\rho)$ and of the variance turn out to be independent of $\boldsymbol{\alpha}$: $A_5^{(L)}/B_4^{(L)} = 1$ and $A_8^{(\text{Var})}/B_6^{(\text{Var})} = 2$. It makes sense then to use a modified version of Criterion C1, focusing for example on the lowest-order coefficient of the variance $A_0^{(\text{Var})}/B_2^{(\text{Var})}$, which dominates in the region of *low* SNR. This term is minimized for $\boldsymbol{\alpha}_\star = [\beta \quad 0 \quad \delta \quad 0]^{\text{T}}$, where β and δ cannot be both zero; the attained variance is $\text{Var}\{\hat{\rho}\} \approx (2\rho^4 + 8\rho^3 + 10\rho^2 + 6\rho + 1)/(K\rho^2)$, independently of β and δ, and coincides with that of the M_2M_4 estimator for CM constellations [68]. The predicted MSE turns out to remain quite close to the CRB throughout the entire SNR range.

4.3.5.2 Two-level Constellations

When $I = 2$, application of Criterion C1 with the available degrees of freedom results in $L_2 = C_4 = C_3 = 0$. In this case, it is possible to obtain closed-form expressions for the optimal weights in terms of the probabilities $p = P_1$, $1 - p = P_2$, and the ring ratio

$w = R_2/R_1$:

$$\beta = \frac{(1+w^2)\left[3w^6 + w^4 + 3p^2(1+w^2)(1+w^4) - 2p(3w^6+2w^4+w^2)\right]}{\left[p^2-(1-p)^2w^4\right]^2} \cdot \epsilon$$

$$\gamma = -\frac{2(1+w^2)}{p+(1-p)w^2} \cdot \epsilon$$

$$\delta = -\frac{p+(1-p)w^4}{\left[p-(1-p)w^2\right]^2} \cdot \epsilon, \quad \epsilon \neq 0.$$

As a result, Criterion C1 achieves $\mathcal{O}(\rho^2)$ in the variance and the MSE, which is the lowest possible order as dictated by the CRB [72]. In contrast, the variances of M_6 and $\mathrm{M}_2\mathrm{M}_4$ are respectively $\mathcal{O}(\rho^3)$ and $\mathcal{O}(\rho^4)$, and their respective MSEs are $\mathcal{O}(\rho^4)$ and $\mathcal{O}(\rho^6)$.

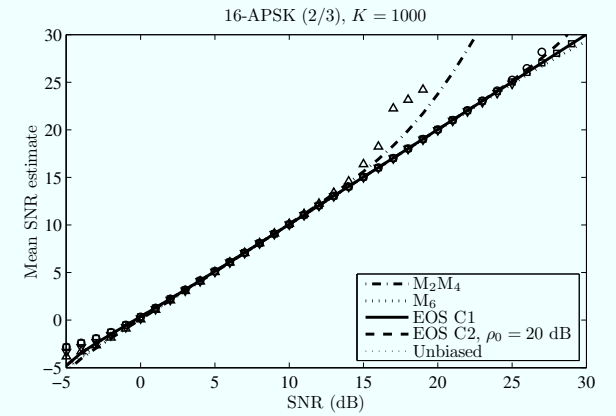
Figures 4.1a and 4.1b show the bias and the NMSE vs. the SNR (for $K = 1000$ samples) for one of the 16-APSK constellations specified in [54] ($P_1 = 1/4$, $P_2 = 3/4$, $R_2/R_1 = 3.15$). The bias under Criterion C1 remains small, and the improvement in NMSE at high SNR is evident. Still, the gap to the CRB is not negligible. Note that for $\rho < 22$ dB, M_6 achieves a slightly lower NMSE than Criterion C1, but its performance quickly degrades beyond this point.

Results for Criterion C2 with $\rho_0 = 20$ dB are also displayed in Figure 4.1b, showing that this is in fact a good strategy in the vicinity of ρ_0 (it outperforms the rest of estimators between 15 and 22 dB), but also for $\rho < \rho_0$, where its NMSE is never larger than that provided by Criterion C1 (this seems to hold regardless of the value of ρ_0). From a designer's perspective, setting the target SNR ρ_0 near the upper limit of the SNR operation range seems an appropriate choice.

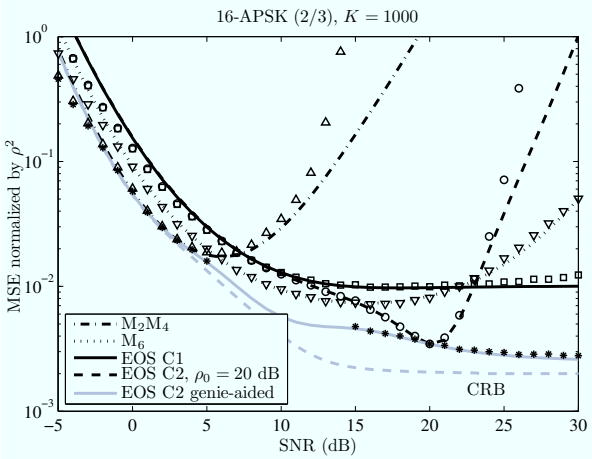
For reference, Figure 4.1b also shows the "genie-aided" NMSE curve, obtained by choosing the optimal $\boldsymbol{\alpha}_\star(\rho)$ under Criterion C2 for each ρ. This gives a lower bound on the NMSE for the proposed eighth-order estimator. Note that the genie-aided curve approaches the CRB as the SNR increases. Hence, within this region, the estimator designed under Criterion C2 is near optimal at the target SNR. In the low SNR region (below 5 dB), $\mathrm{M}_2\mathrm{M}_4$ attains the genie-aided curve; however, neither $\mathrm{M}_2\mathrm{M}_4$, nor M_6, nor Criterion C1 can reach genie-aided performance above 5 dB. In general, good agreement is observed between theory and simulations.

As mentioned in Section 4.3.4.2, Criterion C2 may yield invalid estimators for certain values of ρ_0. For the 16-APSK constellation of Figure 4.1b, this happens for ρ_0 between 5 and 15 dB. In this case one could fall back on some suboptimal solution, e.g., use the valid C2 estimator for $\rho_0 = 15$ dB, which offers good performance (not shown in Figure 4.1) in this SNR interval.

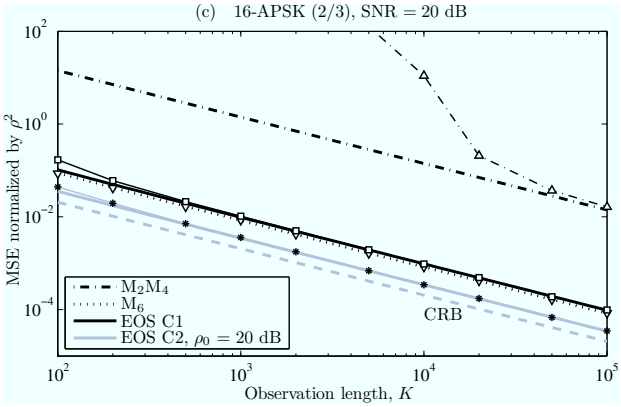
Figure 4.1c shows the variation of the NMSE with the number of observed samples K, for SNR $= 20$ dB. It is seen that, for all estimators, the empirical results agree with the predicted theoretical values as long as the corresponding NMSE is sufficiently small (which is an indicator of the accuracy of the Taylor series approximation used to obtain the theoretical expressions). Note that the new EOS estimators outperform the $\mathrm{M}_2\mathrm{M}_4$ and M_6 schemes in the sense that they can achieve the same NMSE with significantly fewer samples. This holds true also for other SNR values.



(a)



(b)



(c)

Figure 4.1: Theoretical (lines) and empirical (markers) performance of the proposed estimator in terms of (a) bias vs. SNR, (b) NMSE vs. SNR,, and (c) NMSE vs. K for a 16-APSK with ring ratio $R_2/R_1 = 3.15$. Markers are as follows: $\triangle$ for M_2M_4, $\triangledown$ for M_6, $\square$ for C1, $\circ$ for C2 with $\rho_0 = 20$ dB, and $*$ for C2 genie-aided.

4.3.5.3 Three-level Constellations

For $I = 3$, application of Criterion C1 with the available degrees of freedom results in $L_2 = C_4 = 0$. The resulting variance and MSE are respectively $\mathcal{O}(\rho^3)$ and $\mathcal{O}(\rho^4)$, a remarkable improvement over M_2M_4 and M_6, which are both $\mathcal{O}(\rho^4)$ (variance) and $\mathcal{O}(\rho^6)$ (MSE). This can be seen in Figures 4.2a and 4.2b, which display the bias, and the NMSE vs. the SNR (for $K = 1000$ samples) for one of the 32-APSK constellation specified in [54] ($P_1 = 1/4$, $P_2 = 3/8$, $P_3 = 1/2$, $R_2/R_1 = 2.84$, $R_3/R_1 = 5.27$). Good agreement is observed between theory and simulations. M_6 and Criterion C2 with $\rho_0 = 20$ dB remain close to the genie-aided case between 5 and 10 dB, and between 12 and 21 dB, respectively. Note that the genie-aided curve does not approach the CRB in this case. Figure 4.3c displays the dependence of the NMSE with K (for SNR = 20 dB). As observed for 16-APSK, the theoretical analysis proves useful for sufficiently small NMSE, and the EOS estimators outperform M_2M_4 and M_6 in terms of K.

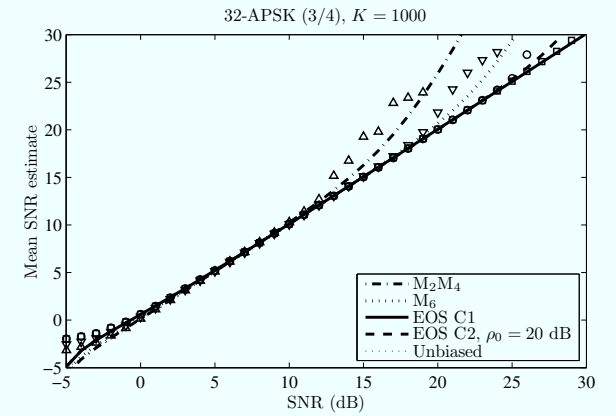
As mentioned in Section 4.3.4.1, Criterion C1 fails to provide a good estimator for the 16-QAM constellation, see Figure 4.3. The estimator designed under Criterion C1 is severely biased above 14 dB, and below that SNR its NMSE is outperformed by those of M_2M_4 and M_6 (observe also the peak in the theoretical NMSE at about 19 dB). From Figures 4.3b and 4.3c it is seen that M_6 almost achieves genie-aided performance above 10 dB, which suggests that the EOS approach is not advantageous for 16-QAM. Comparison with Figure 4.2 shows that the best NMSEs obtained at intermediate-to-high SNR for 16-QAM are substantially higher than those obtained for 32-APSK, *for all moments-based methods.* Thus, the arrangement of signal points for 16-QAM seems to be particularly challenging to this class of estimators.

4.3.5.4 Constellations with More than Three Levels

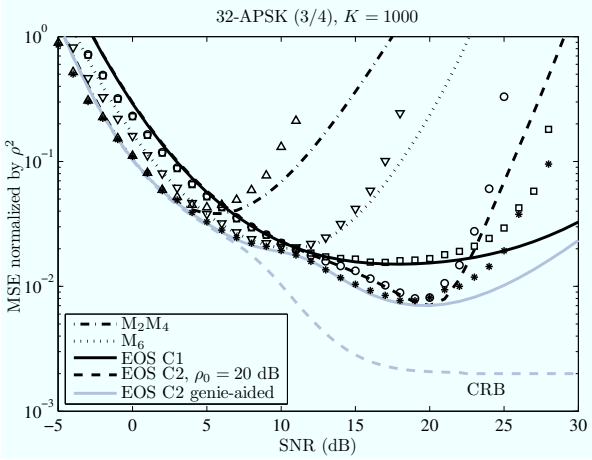
For $I \geq 4$, application of Criterion C1 with the available degrees of freedom results in $L_2 = 0$ and $C_4 > 0$. The resulting variance and MSE are therefore $\mathcal{O}(\rho^4)$ and $\mathcal{O}(\rho^6)$, respectively. Nevertheless, this approach is still useful, as shown in Figure 4.4, which shows the results obtained with 32-QAM ($I = 5$). In terms of NMSE, (Figure 4.4b, $K = 1000$ samples), Criterion C1 clearly improves over M_2M_4 and M_6 in the high SNR region. M_2M_4 achieves close-to-genie-aided performance below 6 dB, whereas M_6 takes over between 7 and 13 dB. For higher SNR, criteria C1 and C2 ($\rho_0 = 20$ dB), which show almost identical performance, are the most competitive. The results in terms of K (Figure 4.4c) confirm the same trend observed for 16-APSK and 32-APSK.

The behavior for other QAM constellations with more than three levels is similar to that of Figure 4.4. We must note that the performance improvement obtained with eighth-order estimators at high SNR is observed to be greater for cross-QAM constellations (32-, 128-QAM) than for square QAMs (64-, 256-QAM).

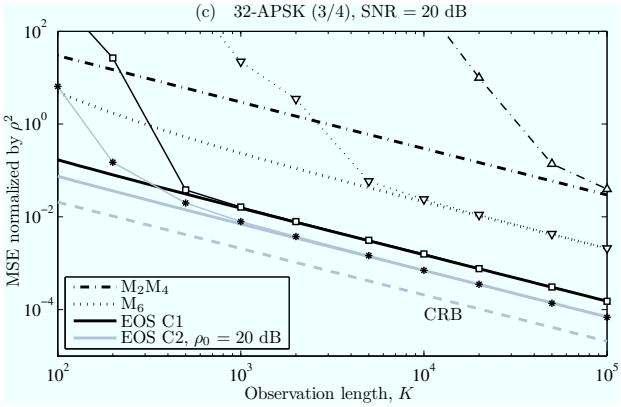
Summarizing, we have seen that the use of higher-order moments is an interesting approach for estimator design. The proposed eighth-order SNR estimator features three degrees of freedom which have been used to optimize the performance of the estimator according to two different criteria. Both criteria have proven useful to yield an improved



(a)

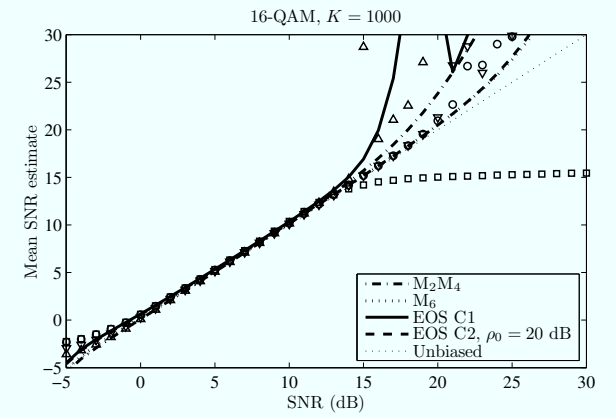


(b)

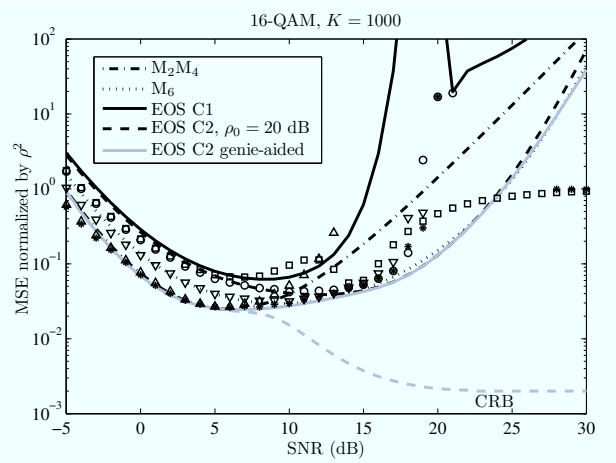


(c)

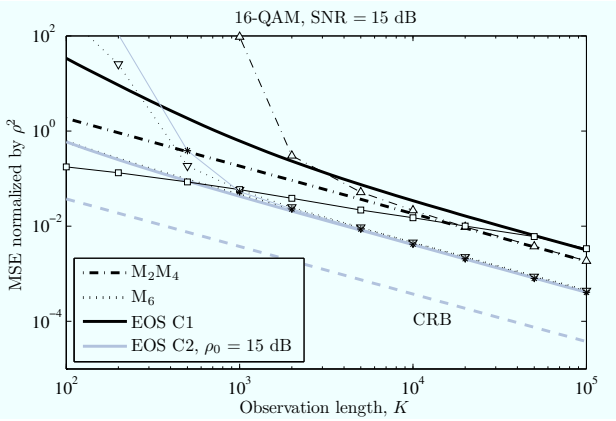
Figure 4.2: Theoretical (lines) and empirical (markers) performance of the proposed estimator in terms of (a) bias vs. SNR, (b) NMSE vs. SNR, and (c) NMSE vs. K for a 32-APSK with ring ratios $R_2/R_1 = 2.84$, $R_3/R_1 = 5.27$. Markers are as follows: $\triangle$ for M_2M_4, $\triangledown$ for M_6, $\square$ for C1, $\circ$ for C2 with $\rho_0 = 20$ dB, and $*$ for C2 genie-aided.



(a)

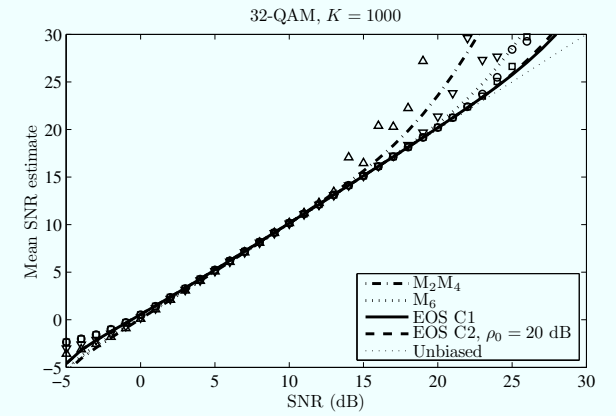


(b)

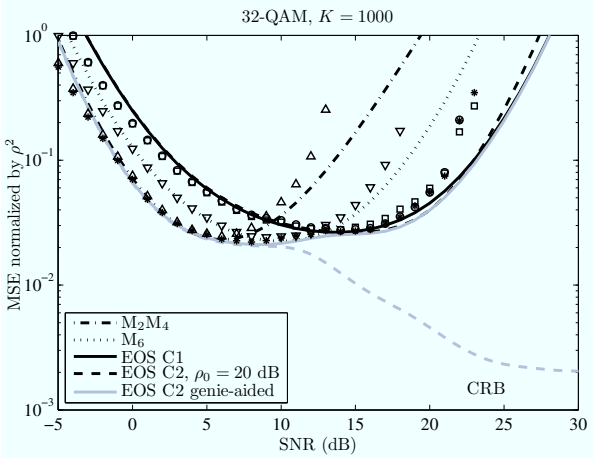


(c)

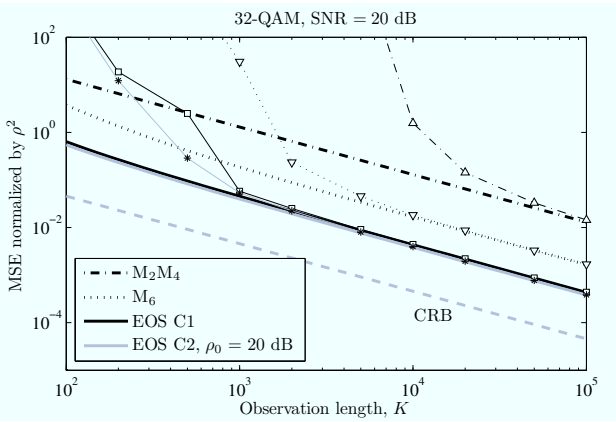
Figure 4.3: Theoretical (lines) and empirical (markers) performance of the proposed estimator in terms of (a) bias vs. SNR, (b) NMSE vs. SNR, and (c) NMSE vs. K for 16-QAM. Markers are as follows: $\triangle$ for M_2M_4, $\triangledown$ for M_6, $\square$ for C1, $\circ$ for C2 with $\rho_0 = 20$ dB, and $*$ for C2 genie-aided.



(a)



(b)



(c)

Figure 4.4: Theoretical (lines) and empirical (markers) performance of the proposed estimator in terms of (a) bias vs. SNR, (b) NMSE vs. SNR, and (c) NMSE vs. K for 32-QAM. Markers are as follows: $\triangle$ for M_2M_4, $\triangledown$ for M_6, $\square$ for C1, $\circ$ for C2 with $\rho_0 = 20$ dB, and $*$ for C2 genie-aided.

performance over previous moments-based estimators. The improvement is remarkable in the case of two-level constellations and most of three-level constellations.

4.4 SNR Estimation with Heterogeneous Frames

4.4.1 Introduction

Whereas most NDA SNR estimators exploit the knowledge of the underlying constellation, in certain situations the observed frame consists of several subframes, with the particular feature that the symbols in different subframes are drawn from different constellations. We shall refer to this kind of frames as *heterogeneous*, as opposed to homogeneous frames composed of symbols from one single constellation. For example, the DVB-S2 standard [54] allows frames in which the data symbols belong either to QPSK, 8-PSK, 16-APSK or 32-APSK. In all of those cases, the frame header and, optionally, some extra pilot symbols within the frame can be assumed to be drawn from a QPSK constellation. Since available NDA SNR estimators are tailored to a given constellation, it is not immediate how to generalize their use to these heterogeneous frame scenarios. Each constituent subframe contains a different quantity of information depending on its number of symbols and the constellation used. One obvious approach is to use only one of the subframes, but this results in inefficient use of the available data. This motivates the search for techniques making use of the whole observation set, in order to improve estimation performance. In our study we will consider both NDA EVB SNR estimators and DA estimators.

In this section we analyze the estimation problem with heterogeneous frames. We derive the CRB for the problem and we propose a number of estimators (Section 4.4.3). Then we evaluate the proposed solutions in some practical scenarios (Section 4.4.4). The evaluation is performed both analytically and empirically through Monte Carlo simulations.

4.4.2 Signal Model for a Heterogeneous Frame

We assume the same flat-fading signal model given in (4.1), with the exception that the original symbols belong to different constellations. Hence, the baud rate sampled output of the matched filter is given by

$$r_k = \sqrt{S}a_k + n_k, \qquad k \in \{1, \ldots, K\} = \mathcal{K}_1 \cup \mathcal{K}_2 \tag{4.43}$$

where a_k are statistically independent complex valued transmitted symbols, $\sqrt{S}$ is the unknown channel gain, and n_k are independent zero-mean circular Gaussian noise samples with unknown variance N. The transmitted symbol sequence $\{a_k,\ k \in \mathcal{K}_1 \cup \mathcal{K}_2\}$ conforms a heterogeneous frame. The disjoint sets $\mathcal{K}_1$, $\mathcal{K}_2$ (known by the receiver) contain the indices of, respectively, the K_1 and K_2 symbols $(K_1 + K_2 = K)$ drawn from constellations $\mathcal{A}_1$ and $\mathcal{A}_2$, which have respectively I_1 and I_2 amplitude levels (e.g. $I_1 = 1$ if

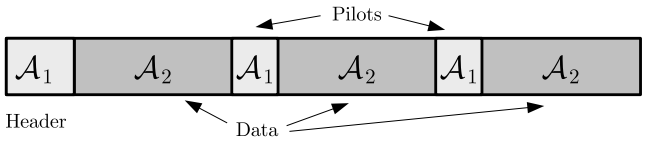


Figure 4.5: Example of a heterogeneous frame containing symbols from two different constellations $\mathcal{A}_1$, $\mathcal{A}_2$.

$\mathcal{A}_1$ is M-PSK and $I_2 = 3$ if $\mathcal{A}_2$ is 32-APSK). The n-th order moment of $\mathcal{A}_i$ is denoted as $c_n^{(i)} \doteq \mathrm{E}\{|a_k|^n\}$, where $k \in \mathcal{K}_i$. In this section we will use the superindex $^{(i)}$ to denote that the estimation is performed over the symbols corresponding to constellation $\mathcal{A}_i$. It will be useful for us to define the ratio $p \doteq K_1/(K_1 + K_2)$. An example of a heterogeneous frame containing symbols from two different constellations is shown in Figure 4.5.

For reference throughout this section we have selected two practical cases of application with different heterogeneous frames which are based on the framing structure provided by the DVB-S2 standard [54]. This standard admits two frame lengths: short frames (carrying 16200 data bits) and normal frames (carrying 64800 bits). The inclusion of pilots is optional. The structure of a DVB-S2 frame is shown in Figure 4.6, and can be broken down as follows:

- *Header*. The header is composed of 90 $\pi/2$ BPSK symbols.[8] The first 26 symbols constitute the SOF (Start of Frame), and have a fixed value. The remaining 64 symbols constitute the so-called PLSCODE (Physical Layer Signaling code) which indicates the frame length, the possible use of pilots, and the MODCOD (modulation and coding rate) used in the data blocks.
- *Data bearing blocks*. Data symbols are placed in slots of 90 symbols. The total number of slots depends on the frame length and the code rate. The following modulations are allowed for the data symbols: QPSK, 8PSK, 16-APSK and 32-APSK. A number of code rates are valid for each modulation.
- *Pilot blocks*. When pilots are used, one block of 36 pilot symbols is inserted each 16 data slots. Again, the total number of pilot symbols depends on the frame length and code rate. All pilot symbols take the value $(1 + j)/\sqrt{2}$. Note that such value corresponds to the first quadrant symbol of a QPSK constellation normalized in energy.

The two heterogeneous frames considered for evaluation, namely HF1 and HF2, are described as follows:

HF1) This frame is a DVB-S2 short frame in which $\mathcal{A}_1 =$ QPSK, $K_1 = 162$, $\mathcal{A}_2 =$ 32-APSK (code rate $3/4$), $K_2 = 3240$ (therefore $p \approx 0.0475$), with $K_t = 98$ known training symbols. In the K_1 QPSK symbols we include not only the pilots inserted

[8]The $\pi/2$ BPSK modulation is composed of complex-valued symbols carrying only one information bit. Assuming energy normalization, odd symbols take values in $\{(1 + j)/\sqrt{2}, (-1 - j)/\sqrt{2}\}$, whereas even symbols take values in $\{(1 - j)/\sqrt{2}, (-1 + j)/\sqrt{2}\}$. Hence, disregarding this alternating structure, $\pi/2$ BPSK symbols can be perfectly considered as QPSK symbols.

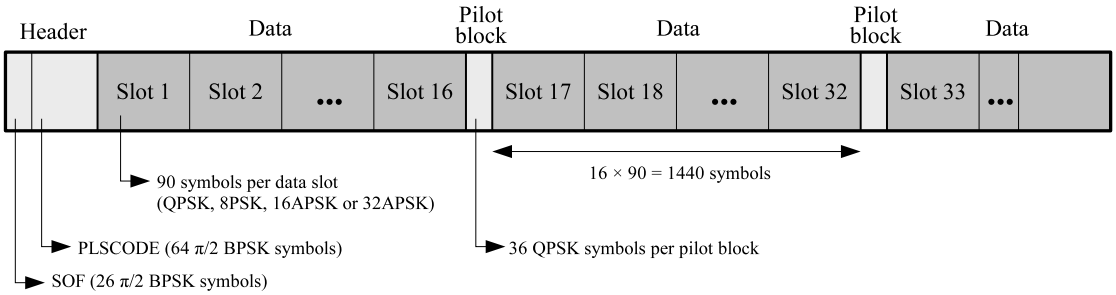


Figure 4.6: Structure of a DVB-S2 frame.

within the frame, but also the $\pi/2$ BPSK symbols of the frame header, which for SNR estimation purposes can be regarded as QPSK.

HF2) This is not a DVB-S2 compliant frame containing pilots with $\mathcal{A}_1 =$ QPSK, $K_1 = 162$, $\mathcal{A}_2 =$ 16-APSK (equivalent to that used in the DVB-S2 standard for code rate $3/4$), $K_2 = 1458$ (therefore $p = 0.1$), $K_t = 98$.

The reason why $K_t \neq K_1$ in the frames described above is the following. Consider the case HF1, where $\mathcal{K}_1$ contains the indices of the QPSK symbols. According to the DVB-S2 standard, a short 32-APSK frame using pilots has 90 header symbols and 72 pilots. From the 90 symbols in the header, however, only those 26 conforming the SOF are known to the receiver, and thus the difference of 64 symbols between K_t and K_1. This has been assumed for case HF2 too. The existence of training symbols allows the use of DA SNR estimation schemes. This has been taken into account in the results reported in Section 4.4.4.

4.4.3 Proposed SNR Estimators

With heterogeneous frames, the choice of an SNR estimator is not straightforward. If one is to apply an available estimator for each subframe and then select the one with lower MSE, the final choice would depend on the true SNR, the constellations, and the sample sizes. For example, suppose $K_1 = 1$, $K_2 > 1$, and $K_2 \gg K_1$. This last fact could suggest that a single estimator over the subset $\mathcal{K}_2$ be employed. However, the performance of standard NDA SNR estimators severely degrades with multilevel constellations, as we have seen in Section 4.3, and hence using the data from subset $\mathcal{K}_1$ may result in significant improvement even if the corresponding number of samples is relatively small.

Another possibility is to use a single estimator for the whole set, upon defining the *aggregate constellation* $\mathcal{A}_0 = \mathcal{A}_1 \cup \mathcal{A}_2$ with new moments $c_p^{(0)}$. This approach, however, does not exploit knowledge of the index sets $\mathcal{K}_i$; moreover, since the I_1 levels of $\mathcal{A}_1$ and the I_2 levels of $\mathcal{A}_2$ need not coincide, $\mathcal{A}_0$ will present a larger number of levels $I_0 = I_1 + I_2$, which can be expected to further degrade estimator performance. This motivates alternative techniques exploiting all available data. In the rest of this section we first provide the procedure to obtain the CRB for this problem, and then we present a number of SNR estimators which make use of the whole set of received samples. For

further reference, $\hat{\rho}_{24}$ will be used to denote the M_2M_4 estimator, and $\hat{\rho}_6$ will denote the sixth-order estimator of [110].

4.4.3.1 Cramér-Rao Lower Bound

Let $\boldsymbol{\phi} = [\rho, N]$ be the vector of unknown parameters. Since the observations are statistically independent, the Fisher information matrix (FIM) satisfies $\boldsymbol{F}(\boldsymbol{\phi}) = \boldsymbol{F}_1(\boldsymbol{\phi}) + \boldsymbol{F}_2(\boldsymbol{\phi})$, where $\boldsymbol{F}_i(\boldsymbol{\phi})$ is the FIM for the estimation problem using only $\{r_k,\ k \in \mathcal{K}_i\}$. Now $\boldsymbol{F}_i(\boldsymbol{\phi}) = K_i \bar{\boldsymbol{F}}_i(\boldsymbol{\phi})$, where $\bar{\boldsymbol{F}}_i(\boldsymbol{\phi})$ does not depend on K_i. Hence,

$$\boldsymbol{F}^{-1}(\boldsymbol{\phi}) = \frac{1}{K}\left[p\bar{\boldsymbol{F}}_1(\boldsymbol{\phi}) + (1-p)\bar{\boldsymbol{F}}_2(\boldsymbol{\phi})\right]^{-1} \tag{4.44}$$

and $\mathrm{Var}\{\hat{\rho}\} \geq [\boldsymbol{F}^{-1}(\boldsymbol{\phi})]_{11}$ for any unbiased estimate $\hat{\rho}$. Depending on the information available, the FIM should be computed differently. In the coherent DA case, one has

$$\boldsymbol{F}_t(\boldsymbol{\phi}) = K_t \begin{bmatrix} \dfrac{1}{2\rho} & \dfrac{1}{2N} \\ \dfrac{1}{2N} & \dfrac{2+\rho}{2N^2} \end{bmatrix}. \tag{4.45}$$

For the NDA EVB case, the FIM must be computed numerically; see e.g. [72].

4.4.3.2 Expectation-Maximization (EM) Estimator

In principle, one could attempt to maximize the Log-likelihood Function (LLF) of the parameters, which is the sum of the individual LLFs over each subframe. This Maximum Likelihood (ML) approach requires numerical optimization and is computationally expensive. As a cheaper alternative, the EM algorithm can be applied. In [72], the EM algorithm for homogeneous frames was derived in the EVB NDA setting. This approach can be readily modified to fit heterogeneous frames, as it allows to incorporate information about the underlying constellation for each individual symbol in the frame. The fact that the conditional pdf is not Gaussian (since only the envelope is considered) makes necessary to use suitable high-SNR approximations, so that the performance of the EM SNR estimator degrades as the SNR decreases [111]. In particular, a non-negligible bias shows up when applied to cases HF1, HF2, and therefore we do not include its performance in this work.

4.4.3.3 Convex Combination of Estimates (CCE)

Suppose that two unbiased estimators $\hat{\rho}^{(i)}$ obtained from samples in subframes $i = 1, 2$, are available. Then the convex combination

$$\hat{\rho}_{\text{cce}} = \lambda_{\text{cce}} \cdot \hat{\rho}^{(1)} + (1 - \lambda_{\text{cce}}) \cdot \hat{\rho}^{(2)} \tag{4.46}$$

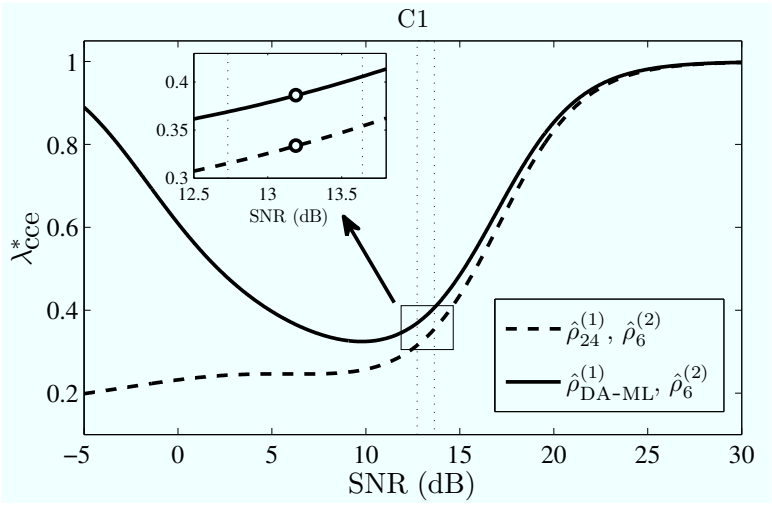


Figure 4.7: $\lambda^{\star}_{\text{cce}}$ as a function of the SNR for CCE: case HF1, with $\{\hat{\rho}^{(1)}_{24}, \hat{\rho}^{(2)}_{6}\}$ and $\{\hat{\rho}^{(1)}_{\text{DA-ML}}, \hat{\rho}^{(2)}_{6}\}$. The inset zooms in the region of operation of the DVB-S2 32-APSK $3/4$ MODCOD. The trade-off values for λ_{cce}, marked as circles, are taken from the center of the working SNR interval (delimited by vertical dotted lines).

is also an unbiased estimator. Its variance is readily found to be

$$\sigma^2_{\text{cce}} = \lambda^2_{\text{cce}}\sigma^2_1 + (1-\lambda_{\text{cce}})^2\sigma^2_2 \tag{4.47}$$

where $\sigma^2_i \doteq \text{Var}\left\{\hat{\rho}^{(i)}\right\}$. This variance has a minimum value

$$\sigma^2_{\text{cce},\star} = \frac{\sigma^2_1\,\sigma^2_2}{\sigma^2_1+\sigma^2_2}$$

which is achieved by the optimal choice

$$\lambda^{\star}_{\text{cce}} = \frac{\sigma^2_2}{\sigma^2_1+\sigma^2_2}. \tag{4.48}$$

This approach is motivated by the following fact: suppose that $\hat{\boldsymbol{\phi}}^{(i)}$ were efficient estimates restricted to each subframe $i = 1, 2$ [i.e. their covariance matrices are $\boldsymbol{F}_i^{-1}(\boldsymbol{\phi})$]. Then $\hat{\boldsymbol{\phi}} = \boldsymbol{W}\hat{\boldsymbol{\phi}}^{(1)} + (\boldsymbol{I}-\boldsymbol{W})\hat{\boldsymbol{\phi}}^{(2)}$ is an efficient estimate (considering all the frame), i.e. its covariance matrix is given by (4.44). The weight $\boldsymbol{W}$ depends in general on the true parameter and is given by $\boldsymbol{W} = \boldsymbol{F}_2^{-1}(\boldsymbol{\phi})\left[\boldsymbol{F}_1^{-1}(\boldsymbol{\phi}) + \boldsymbol{F}_2^{-1}(\boldsymbol{\phi})\right]^{-1}$. Thus, convex combinations of estimates are attractive, although in practice the constituent estimators need not be efficient.

In order to choose λ_{cce}, the true variances σ^2_1, σ^2_2 are needed. Since in general they depend on ρ, a trade-off must be reached in order to obtain acceptable performance over the whole operation range. To illustrate this, let us consider case HF1, and assume an NDA-NDA CCE with $\{\hat{\rho}^{(1)}_{24},\ \hat{\rho}^{(2)}_{6}\}$, and a DA-NDA CCE with $\{\hat{\rho}^{(1)}_{\text{DA-ML}},\ \hat{\rho}^{(2)}_{6}\}$ are used

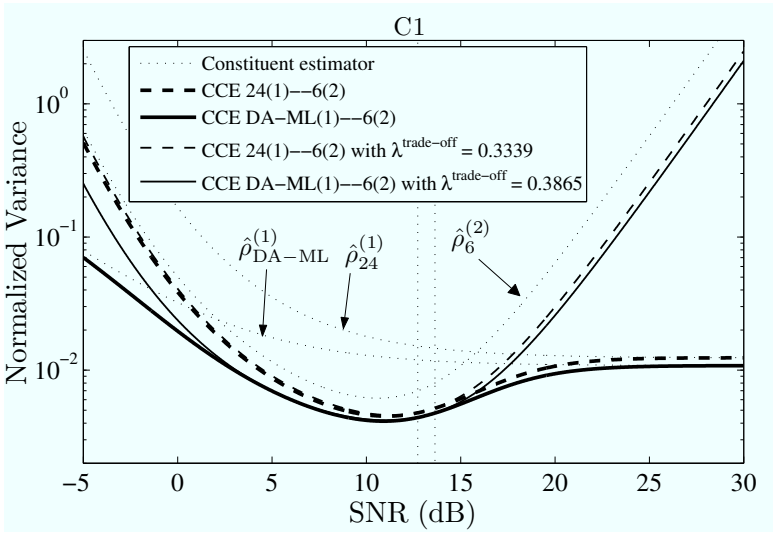


Figure 4.8: Case HF1: Analytical variances of $\hat{\rho}^{(1)}_{24}$, $\hat{\rho}^{(1)}_{\text{DA-ML}}$, $\hat{\rho}^{(2)}_{6}$ and the combined estimators $\hat{\rho}_{\text{cce}}$ for the optimal SNR dependent values $\lambda^{\star}_{\text{cce}}$ and for the trade-off values $\lambda^{\text{trade-off}}_{\text{cce},24-6} = 0.3339$, $\lambda^{\text{trade-off}}_{\text{cce,DA-ML}-6} = 0.3865$.

in the corresponding subframes.[9] Since analytical approximate expressions are known for the variances of $\hat{\rho}^{(1)}_{24}$, $\hat{\rho}^{(1)}_{\text{DA-ML}}$ and $\hat{\rho}^{(2)}_{6}$ [40, 68, 110], it is possible to plot $\lambda^{\star}_{\text{cce}}$ in terms of ρ, as in Figure 4.7. According to the DVB-S2 standard, the MODCOD 32-APSK $3/4$ is expected to be used in the SNR range $[12.73 \text{ dB}, 13.64 \text{ dB}]$, shown in the inset of Figure 4.7, and where tentative trade-off values for λ_{cce} are marked. These values $\lambda^{\text{trade-off}}_{\text{cce},24-6} = 0.3339$, $\lambda^{\text{trade-off}}_{\text{cce,DA-ML}-6} = 0.3865$ have been chosen as the value of $\lambda^{\star}_{\text{cce}}$ at the center of the nominal SNR working range of the selected DVB-S2 MODCOD.

Figure 4.8 shows the analytical variance of $\hat{\rho}_{\text{cce}}$ using the optimal value $\lambda^{\star}_{\text{cce}}(\rho)$, marked with thicker lines, and the analytical variance obtained with $\lambda^{\text{trade-off}}_{\text{cce}}$. The variances of the two constituent estimators are shown for comparison. It can be seen that the CCE estimators consistently outperform the better of the two constituent estimators over a wide SNR range. In general, the largest improvement is achieved when the variances of the original estimators are equal; in this case $\lambda^{\star}_{\text{cce}} = 0.5$ and the CCE method yields a 3 dB reduction in estimation variance. This can be observed around SNR $\approx 15 - 16$ dB in Figure 4.8. Also, it can be seen that the selected $\lambda^{\text{trade-off}}_{\text{cce}}$ perform very close to the optimal value not only in the desired SNR range, but also outside of it.

We remark that the CCE approach is not limited to constituent moment-based estimators. Any two independent unbiased estimators can be used, as long as their true variances are known or can be analytically or empirically approximated, in order to select λ_{cce}.

[9] We note that in the DA-NDA approach, only K_t of the K_1 symbols of subframe 1 are being used by the DA-ML estimator; thus there are $K_1 - K_t = 64$ symbols that remain unused. In this case, one could define a third subframe with those 64 QPSK symbols to which an NDA estimator should be applied, yielding an overall DA-NDA-NDA estimator over three subframes. The CCE approach can be straightforwardly extended to more than two subframes, although we restrict our analysis to two subframes for the sake of simplicity.

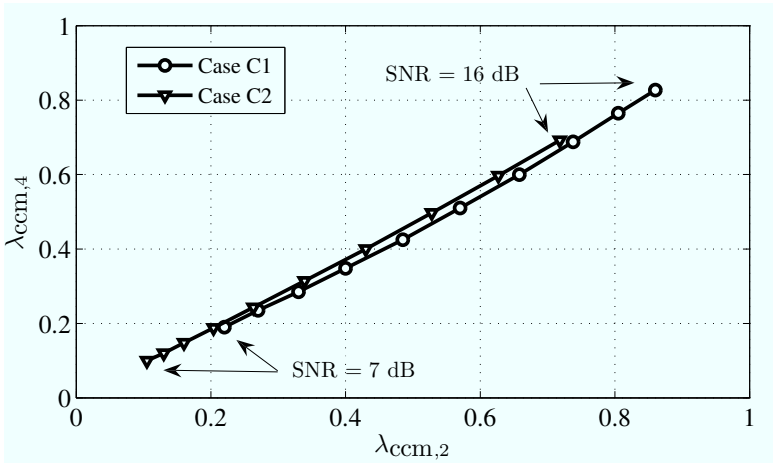


Figure 4.9: Optimal weights $\lambda^*_{\mathrm{ccm},2}$, $\lambda^*_{\mathrm{ccm},4}$ for $\hat{\rho}^{\mathrm{ccm}}_{24}$ as a function of the SNR.

4.4.3.4 Convex Combination of Sample Moments (CCM)

Instead of combining estimators, one could combine sample moments corresponding to different subframes *before* computing the estimate. For instance, we could use

$$\hat{M}^{\mathrm{ccm}}_n = \lambda_{\mathrm{ccm},n} \cdot \hat{M}^{(1)}_n + (1 - \lambda_{\mathrm{ccm},n}) \cdot \hat{M}^{(2)}_n \tag{4.49}$$

with $\lambda_{\mathrm{ccm},n} \in [0, 1]$. These are unbiased estimates of the convex combination of the true moments $M^{\mathrm{ccm}}_n \doteq \lambda_{\mathrm{ccm},n} \cdot M^{(1)}_n + (1 - \lambda_{\mathrm{ccm},n}) \cdot M^{(2)}_n$. Restricting ourselves to $n = 2,\ 4$ for simplicity, the convexly combined moments are given by

$$M^{\mathrm{ccm}}_2 = c^{\mathrm{ccm}}_2 S + N \tag{4.50}$$

$$M^{\mathrm{ccm}}_4 = c^{\mathrm{ccm}}_4 S^2 + 4c^{\mathrm{ccm}}_2 SN + 2N^2 \tag{4.51}$$

with $c^{\mathrm{ccm}}_n \doteq \lambda_{\mathrm{ccm},n} \cdot c^{(1)}_n + (1 - \lambda_{\mathrm{ccm},n}) \cdot c^{(2)}_n$. An estimate $\hat{\rho}^{\mathrm{ccm}}_{24}$ is obtained in the same way as the original $\mathrm{M_2M_4}$ estimator, see (4.12), but considering the new combined sample and constellation moments.

The weights $\lambda_{\mathrm{ccm},2}$, $\lambda_{\mathrm{ccm},4}$ should be chosen to minimize the variance of the resulting estimate. An approximation to this variance can be analytically obtained following the guidelines in [68], resulting in a quadratic function of $\lambda_{\mathrm{ccm},2}$, $\lambda_{\mathrm{ccm},4}$ whose coefficients depend in a complicated manner on the SNR ρ, the ratio p, and the moments of the constellations.

Figure 4.9 shows the values of the optimal weights as a function of the SNR for Cases HF1 and HF2. For low SNR both weights approach 0: the estimate gives more weight to subframe 2, which has more samples. As the SNR increases, however, the optimal weights approach 1, i.e. exploiting subframe 1 becomes more beneficial. Note that the optimal parameters are always close to satisfying $\lambda^*_{\mathrm{ccm},2} = \lambda^*_{\mathrm{ccm},4}$, and thus it makes sense to impose this condition *a priori* and then minimize the estimator variance. In practice, adequate values for $\lambda_{\mathrm{ccm},n}$ can be obtained empirically for the desired SNR range. In the Appendix 4.C we propose a simpler, suboptimal approach to weight selection for the CCM estimator.

4.4.3.5 Weighted Least Squares (WLS) Approach

Using (4.5) we can obtain estimates of S and N by matching the theoretical and sample second and fourth order moments for the two subframes: for $i = 1, 2$,

$$\hat{M}_2^{(i)} = c_2^{(i)}\hat{S} + \hat{N}, \tag{4.52a}$$

$$\hat{M}_4^{(i)} = c_4^{(i)}\hat{S}^2 + 4c_2^{(i)}\hat{S}\hat{N} + 2\hat{N}^2. \tag{4.52b}$$

These four equations in two unknowns suggest a weighted LS approach to estimate S and N.[10] However, this is a nonlinear problem with no straightforward closed-form solution. However, if we take $\hat{N} = \hat{M}_2^{(i)} - c_2^{(i)}\hat{S}$ in (4.52a) and substitute this back in (4.52b), we obtain two linear equations in $\hat{S}^2$:

$$\left[c_4^{(i)} - 2\big(c_2^{(i)}\big)^2\right]\hat{S}^2 = \hat{M}_4^{(i)} - 2\big(\hat{M}_2^{(i)}\big)^2, \qquad i = 1, 2. \tag{4.53}$$

Let $w_i,\ i = 1, 2$ be the weights associated to each subframe, with $w_2 = 1 - w_1$. Then the WLS solution to (4.53) is

$$\hat{S}_{\text{WLS}} = \sqrt{\frac{w_1\gamma_1\delta_1 + w_2\gamma_2\delta_2}{w_1\gamma_1^2 + w_2\gamma_2^2}} \tag{4.54}$$

with $\delta_i = \hat{M}_4^{(i)} - 2\big(\hat{M}_2^{(i)}\big)^2$ and $\gamma_i = c_4^{(i)} - 2\big(c_2^{(i)}\big)^2$. With this value, we now obtain the WLS solution to $\hat{M}_2^{(i)} = c_2^{(i)}\hat{S}_{\text{WLS}} + \hat{N}$, $i = 1, 2$, which is

$$\hat{N}_{\text{WLS}} = [w_1\hat{M}_2^{(1)} + w_2\hat{M}_2^{(2)}] - [w_1c_2^{(1)} + w_2c_2^{(2)}]\hat{S}_{\text{WLS}}. \tag{4.55}$$

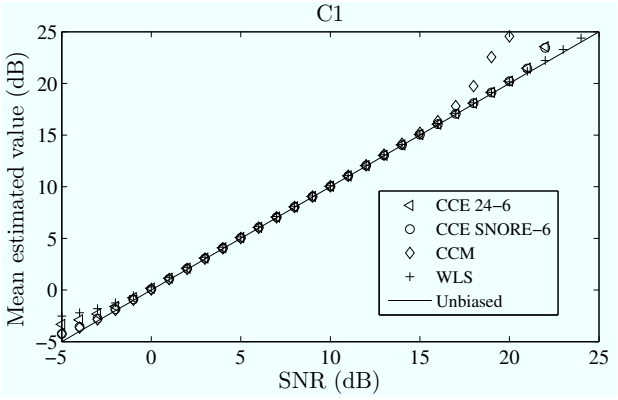
Note that $w_1 = 1$ yields $\hat{\rho}_{24}^{(1)}$, and $w_1 = 0$ yields $\hat{\rho}_{24}^{(2)}$. The weights w_i should be chosen to yield minimum variance. Similarly to the CCM case, the optimal weights should be obtained through the analytical variance of the WLS estimate, whose derivation is quite cumbersome. In practice, empirically obtained weights can be used.

4.4.4 Performance Comparison

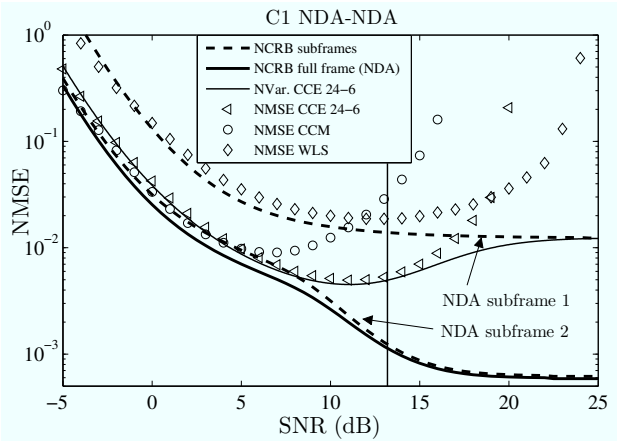
In this section we offer a comparison among the proposed NDA estimators for the cases of application defined in Section 4.4.2. For completeness, we include the performance of the DA ML (SNORE) estimator (4.2), which is of course applied over the set of training symbols of length K_t. For each case, the CRB normalized by ρ^2 (NCRB) is shown, as well as the analytical variance (normalized by ρ^2) of the CCE estimators. Using Monte Carlo simulations (100,000 repetitions per SNR point) we have measured the mean estimated value and the NMSE $\doteq \mathrm{E}\left\{(\hat{\rho} - \rho)^2\right\}/\rho^2$ (Mean Square Error normalized by ρ^2) of the proposed estimators for chosen trade-off values of the weights λ_{cce}, $\lambda_{\text{ccm},n}$ and w_1.

Case HF1 is reported in Figure 4.10. Figure 4.10a displays the mean estimated value of the proposed heterogeneous SNR estimators. The CCM estimator presents a significant

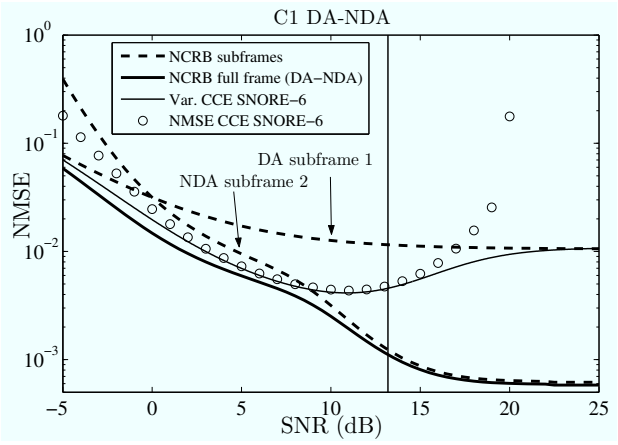
[10] Given the estimates of the signal power $\hat{S}$ and noise power $\hat{N}$, the SNR estimate is simply obtained as $\hat{\rho} = \hat{S}/\hat{N}$.



(a)



(b)



(c)

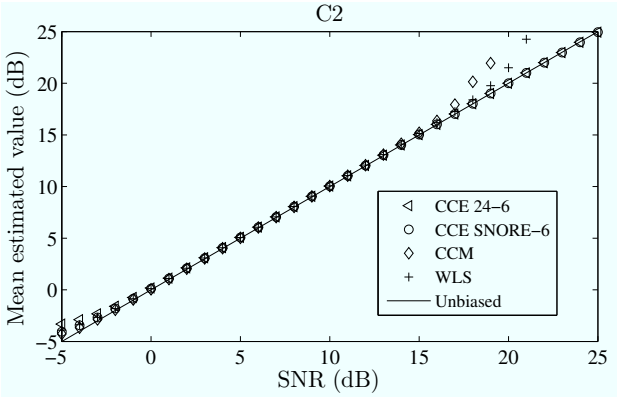
Figure 4.10: Case HF1, trade-off SNR $= 13.19$ dB: (a) mean estimated value; (b) NDA case: NCRBs, analytical variances and NMSEs of the proposed pure NDA estimators; (c) DA-NDA case: NCRBs, analytical variances and NMSEs of the DA-NDA CCE.

bias above SNR $= 16$ dB, while both CCE estimators and the WLS estimator present no noticeable bias between 0 and 20 dB. Figure 4.10b displays the case of combination of two NDA estimates. Shown are the NDA NCRBs for each subframe and for the heterogeneous frame, as well as the analytical variance of the NDA-NDA CCE of $\hat{\rho}_{24}^{(1)}$, $\hat{\rho}_{6}^{(2)}$. Finally, we show also the NMSE obtained through simulations for the NDA-NDA CCE estimator ($\lambda_{\text{cce},24-6}^{\text{trade-off}} = 0.3339$), the CCM estimator ($\lambda_{\text{ccm},2} = 0.2069$, $\lambda_{\text{ccm},4} = 0.2254$), and the WLS estimator ($w_1 = 0.85$). The former weights have been calculated taking SNR $= 13.19$ dB as a reference [marked as a vertical solid line in Figures 4.10b and 4.10c]. Figure 4.10c displays the DA-NDA case. We show the DA NCRB over the K_t training symbols, the NDA NCRB over subframe 2, and the joint DA-NDA NCRB. We also show the analytical variance of the DA-NDA CCE $\hat{\rho}_{\text{DA-ML}}^{(1)}$, $\hat{\rho}_{6}^{(2)}$, as well as the empirical NMSE obtained for the DA-NDA CCE with $\lambda_{\text{cce,DA-ML}-6}^{\text{trade-off}} = 0.3865$.

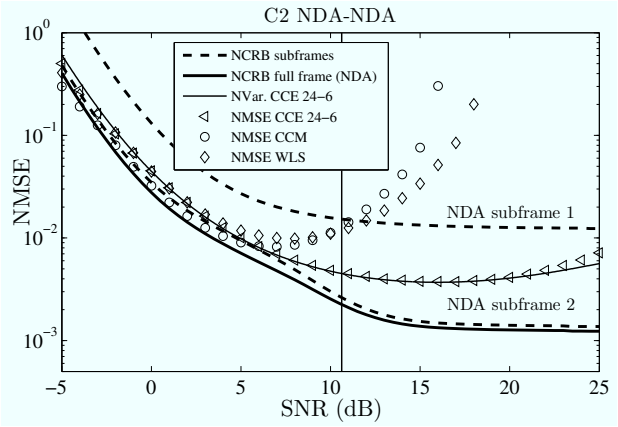
It is remarkable that both the NDA-NDA and the DA-NDA full-frame NCRBs are close to the NDA NCRBs over subframe 2, especially for moderate to high SNRs; this means that the higher number of samples in subframe 2 is more statistically meaningful than the few QPSK symbols (or training symbols) of subframe 1. We can also observe that both CCE estimators perform close to their minimum variance not only for the SNR of reference but also for a wide range of SNRs. The CCE of DA-ML and $\hat{\rho}_6$ performs slightly better than the CCE of $\hat{\rho}_{24}$ and $\hat{\rho}_6$, which is explained due to the lower variance of the DA estimator with respect to the NDA estimator over subframe 1. Both CCE estimators depart from the NCRB for SNRs above 10 dB, similarly to the behavior of the constituent estimator $\hat{\rho}_6$.

Figure 4.11 presents the same results as Figure 4.10 but for case HF2. The estimators have been simulated considering the reference SNR $= 10.62$ dB [marked as a vertical solid line in Figures 4.11b and 4.11c], for which the following parameters were obtained: $\lambda_{\text{cce},24-6}^{\text{trade-off}} = 0.2632$, $\lambda_{\text{cce,DA-ML}-6}^{\text{trade-off}} = 0.3259$, $\lambda_{\text{ccm},2} = 0.2038$, $\lambda_{\text{ccm},4} = 0.1919$, and $w_1 = 0.3$. Roughly speaking, the same comments made for case HF1 apply for case HF2. Among the differences, we note the bias of the WLS estimator above 18 dB, while both CCE estimators can be considered unbiased between 0 and 25 dB. Also, unlike in HF1, here the WLS performs similarly to CCM in terms of NMSE. However, both CCE estimators show again the lowest NMSE. This is due to $\hat{\rho}_6$, which outperforms the estimators over subframe 2 because it is well suited to the 16-APSK constellation (notice that the gap between NCRB and NMSE of CCE is smaller for case HF2 than for case HF1).

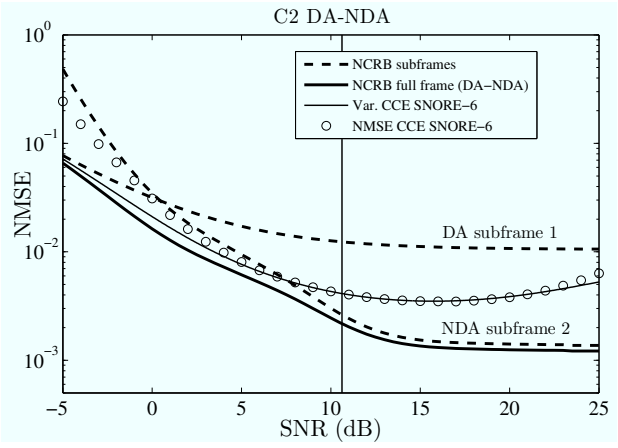
In general, our results show that the CCE approach consistently outperforms the CCM and WLS approaches. The limitation of the CCM and WLS approaches lies in the fact that they are restricted to using only second and fourth order statistics. Therefore, they cannot be expected to perform well with multilevel constellations, as shown in our results. On the other hand, the CCE approach is not bound to use a particular estimator; instead, any pair of (unbiased) estimators can be combined to yield a new estimator with lower variance (under an appropriate choice of λ_{cce}). This includes DA schemes, which turns into a further advantage of the CCE scheme when training symbols (pilots, partially known header) are available in the heterogeneous frame.. The two considered CCE estimators perform well over a wide SNR range; in particular, the hybrid DA-NDA CCE estimator performs slightly better than the NDA-NDA estimator, which is explained by the lower



(a)



(b)



(c)

Figure 4.11: Case HF2, trade-off SNR $= 10.62$ dB: (a) mean estimated value; (b) NDA case: NCRBs, analytical variances and NMSEs of the proposed pure NDA estimators; (c) DA-NDA case: NCRBs, analytical variances and NMSEs of the DA-NDA CCE.

variance of the DA estimator over subframe 1.

Appendix 4.A SNR Estimation Using Higher-Order Moments

In this appendix section we provide the proofs of the results given in Section 4.3.

4.A.1 Proof of Property 1

Recalling (4.5), it is possible to write the numerator and denominator of (4.9) as

$$\prod_{i=1}^{U} M_{2n_i}^{p_i} = N^Q D(\rho) \tag{4.56}$$

$$M_2^Q = N^Q (1+\rho)^Q \tag{4.57}$$

with $D(\rho)$ a polynomial of degree Q. As expected, the ratio of (4.56) to (4.57) does not depend on N. Now, applying the change of variable $\rho = z/(1-z)$, one has $\rho^m/(1+\rho)^Q = z^m(1-z)^{Q-m}$, for $m = 0, 1, \ldots, Q$, which shows that

$$f(\rho) = \sum_{m=0}^{Q} D_m \, z^m (1-z)^{Q-m} = F(z) \tag{4.58}$$

is a polynomial of degree Q in z. We note that this property holds also for any linear combination of ratios of the form of (4.9). We will prove now that F_1, the coefficient in z of $F(z) = f(\rho)$, is always zero. Since $f(\rho)$ is a product of terms of the form M_{2n}/M_2^n, it suffices to show that any such term is a polynomial in z with zero first-order coefficient. To see this, note from (4.5) and $\rho = z/(1-z)$ that $M_{2n}/M_2^n = \sum_{m=0}^{n} G_m^{(n)} z^m (1-z)^{n-m}$, where $G_m^{(n)} \doteq c_{2m}(n!)^2/[(n-m)!(m!)^2]$. Since $z^m(1-z)^{n-m} = z^m - (n-m)z^{m+1} + \cdots$, it follows that the coefficient in z of M_{2n}/M_2^n is $G_1^{(n)} - nG_0^{(n)} = n(n!)c_2 - n(n!)c_0 = 0$, since $c_0 = c_2 = 1$.

4.A.2 Computation of the Variance (4.20)

From the definitions of $t(\hat{\boldsymbol{m}})$, $g(\hat{\boldsymbol{m}})$ and $h_{\text{EOS}}(\hat{\boldsymbol{m}})$ we have

$$\nabla t = \frac{\nabla g}{(1-g)^2} \tag{4.59}$$

$$\nabla g = \frac{\nabla h_{\text{EOS}}}{F'_{\text{EOS}}(g)} = \frac{\nabla h_{\text{EOS}}}{4F_4\, g^3 + 3F_3\, g^2 + 2F_2\, g}. \tag{4.60}$$

∇h_{EOS} can be readily computed from (4.13), obtaining

$$\nabla h_{\text{EOS}} = \frac{1}{\hat{M}_2^5} \hat{\boldsymbol{R}}\, \boldsymbol{\alpha} \tag{4.61}$$

$$\hat{\boldsymbol{R}} \doteq \begin{bmatrix} -2\hat{M}_4\hat{M}_2^2 & -3\hat{M}_6\hat{M}_2 & -4\hat{M}_4^2 & -4\hat{M}_8 \\ \hat{M}_2^3 & 0 & 2\hat{M}_4\hat{M}_2 & 0 \\ 0 & \hat{M}_2^2 & 0 & 0 \\ 0 & 0 & 0 & \hat{M}_2 \end{bmatrix}. \tag{4.62}$$

Now, noting that $g|_{\hat{\boldsymbol{m}}=\boldsymbol{m}} = z = \rho/(1+\rho)$, the evaluation of (4.59)–(4.62) at $\hat{\boldsymbol{m}} = \boldsymbol{m}$ yields

$$\boldsymbol{v} = \frac{1}{N^5}\frac{\boldsymbol{R}\,\boldsymbol{\alpha}}{X(\rho)} \tag{4.63}$$

where $\boldsymbol{R} \doteq \hat{\boldsymbol{R}}|_{\hat{\boldsymbol{m}}=\boldsymbol{m}}$. The polynomial $X(\rho) \doteq X_3\,\rho^3 + X_2\,\rho^2 + X_1\,\rho$ is linear in $\boldsymbol{\alpha}$: we can write $X_3 \doteq 4F_4 + 3F_3 + 2F_2 = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x}_3$, $X_2 \doteq 3F_3 + 4F_2 = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x}_2$, and $X_1 \doteq 2F_2 = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x}_1$, where the 4×1 vectors $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$ depend on the constellation moments only. Upon defining $\boldsymbol{x}(\rho) \doteq \sum_{n=1}^{3} \boldsymbol{x}_n\,\rho^n$, one has that $X(\rho) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x}(\rho)$, and the approximation of the variance can be written as

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} = \frac{1}{N^{10}}\frac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{R}\,\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x}(\rho)[\boldsymbol{x}(\rho)]^{\mathrm{T}}\boldsymbol{\alpha}}. \tag{4.64}$$

Observe that $[\boldsymbol{C}]_{ij} = (M_{2(i+j)} - M_{2i}M_{2j})/K$ has the form of (4.56) for $Q = i + j$, and that the elements of the i-th row of $\boldsymbol{R}$ have the form of (4.56) as well, with $Q = 5 - i$, $(i = 1, 2, 3, 4)$. It is then possible to show that all the nonzero elements of $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{R}$ are of the form of (4.56) with $Q = 10$. With this, the factor N^{10} in (4.64) cancels out and (4.20) follows by simply identifying $A^{(\mathrm{Var})}(\rho) = K/N^{10}\,\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{R}\,\boldsymbol{\alpha}$, and $B^{(\mathrm{Var})}(\rho) = X^2(\rho) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x}(\rho)[\boldsymbol{x}(\rho)]^{\mathrm{T}}\boldsymbol{\alpha}$, which are polynomials in ρ of degrees 10 and 6 respectively.

4.A.3 Proofs of Properties 2–5

That $\boldsymbol{B}^{(\mathrm{Var})}(\rho)$ has rank one (Property 4) is evident from (4.64), if we identify $\boldsymbol{x}(\rho)[\boldsymbol{x}(\rho)]^{\mathrm{T}}$ with $\boldsymbol{B}^{(\mathrm{Var})}(\rho)$ of (4.40) [and therefore $\boldsymbol{x}(\rho) = \boldsymbol{b}^{(\mathrm{Var})}(\rho)$]. Similarly, identifying $\boldsymbol{b}_6^{(\mathrm{Var})}$ with $\boldsymbol{x}_3$, so that $\boldsymbol{B}_6^{(\mathrm{Var})} = \boldsymbol{x}_3\boldsymbol{x}_3^{\mathrm{T}}$, then Property 5 follows.

In order to prove that $\boldsymbol{A}^{(\mathrm{Var})}(\rho)$ is rank deficient (Property 2), we note that $\boldsymbol{R}$ has rank 3 since its first and third columns are linearly dependent, see (4.62). It follows that $\mathrm{rank}\left(\boldsymbol{A}^{(\mathrm{Var})}(\rho)\right) = \mathrm{rank}\left(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{R}\right) \leq 3$ for all ρ. The first part of Property 3 immediately follows by noting that $\boldsymbol{A}_{10}^{(\mathrm{Var})} = \lim_{\rho\to\infty} \boldsymbol{A}^{(\mathrm{Var})}(\rho)/\rho^{10}$. Finally, for constellations with $I = 1$, 2 or 3 amplitude levels, it is straightforward (but rather tedious) to explicitly compute $\boldsymbol{A}_{10}^{(\mathrm{Var})}$, as well as to check that its rank does not exceed 0, 1, and 2, respectively.

4.A.4 Computation of the Bias (4.24)

The Hessian matrix of $t(\hat{\boldsymbol{m}})$, $\nabla^2 t$, can be readily computed from (4.59)–(4.61) as

$$\begin{aligned}\nabla^2 t &= \frac{1}{(1-g)^2 F'_{\mathrm{EOS}}(g)}\,\nabla^2 h \\ &\quad + \frac{2F'_{\mathrm{EOS}}(g) - (1-g)F''_{\mathrm{EOS}}(g)}{(1-g)^3[F'_{\mathrm{EOS}}(g)]^2}\,\nabla h\,(\nabla h)^{\mathrm{T}} \qquad (4.65)\\ &= \frac{1}{(1-g)^2 F'_{\mathrm{EOS}}(g)}\,\frac{1}{\hat{M}_2^6}\,\hat{\boldsymbol{Y}} + J_t(g)\,\nabla t\,(\nabla t)^{\mathrm{T}} \qquad (4.66)\end{aligned}$$

where

$$J_t(g) \doteq \frac{(1-g)\ [2F'_{\text{EOS}}(g) - (1-g)F''_{\text{EOS}}(g)]}{F'_{\text{EOS}}(g)} \tag{4.67}$$

and $\hat{\boldsymbol{Y}}$ is a 4×4 symmetric matrix whose only nonzero elements are given by

$$[\hat{\boldsymbol{Y}}]_{11} \doteq 6\beta\hat{M}_4\hat{M}_2^2 + 12\gamma\hat{M}_6\hat{M}_2 + 20(\delta\hat{M}_4^2 + \epsilon\hat{M}_8) \tag{4.68a}$$
$$[\hat{\boldsymbol{Y}}]_{12} = [\hat{\boldsymbol{Y}}]_{21} \doteq -2\beta\hat{M}_2^3 - 8\delta\ \hat{M}_4\hat{M}_2 \tag{4.68b}$$
$$[\hat{\boldsymbol{Y}}]_{13} = [\hat{\boldsymbol{Y}}]_{31} \doteq -3\gamma\hat{M}_2^2 \tag{4.68c}$$
$$[\hat{\boldsymbol{Y}}]_{14} = [\hat{\boldsymbol{Y}}]_{41} \doteq -4\epsilon\hat{M}_2 \tag{4.68d}$$
$$[\hat{\boldsymbol{Y}}]_{22} \doteq 2\delta\hat{M}_2^2. \tag{4.68e}$$

The evaluation of (4.66) at $\hat{\boldsymbol{m}} = \boldsymbol{m}$ yields

$$\boldsymbol{H} \doteq \nabla^2 t|_{\hat{\boldsymbol{m}}=\boldsymbol{m}} = \frac{1}{N^6}\frac{1}{(1+\rho)X(\rho)}\boldsymbol{Y} + 2J(\rho)\ \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} \tag{4.69}$$

where $\boldsymbol{Y} = \hat{\boldsymbol{Y}}|_{\hat{\boldsymbol{m}}=\boldsymbol{m}}$, $X(\rho)$ has been defined in Section 4.A.2 of this Appendix, and $J(\rho) \doteq \frac{1}{2}J_t(\frac{\rho}{1+\rho})$. Regarding the first term of (4.69), its denominator is clearly linear in $\boldsymbol{\alpha}$: $(1+\rho)X(\rho) = \boldsymbol{\alpha}^{\mathrm{T}}[(1+\rho)\boldsymbol{x}(\rho)]$, from where we can identify $\boldsymbol{b}^{(L)}(\rho)$ of (4.26) as

$$\boldsymbol{b}^{(L)}(\rho) = (1+\rho)\boldsymbol{x}(\rho) = (1+\rho)\boldsymbol{b}^{(\text{Var})}(\rho). \tag{4.70}$$

Regarding the factor $J(\rho)$ in second term of (4.69), it is readily seen to be of the form $J(\rho) = A^{(J)}(\rho)/[(1+\rho)X(\rho)] = [\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{a}^{(J)}(\rho)]/[\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{b}^{(L)}(\rho)]$, where $A^{(J)}(\rho) \doteq \sum_{n=0}^{3} A_n^{(J)}\rho^n = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{a}^{(J)}(\rho)$, and $\boldsymbol{a}^{(J)}(\rho)$ depends on ρ and the constellation moments.

Recalling the approximation (4.23), simple algebra shows that

$$\frac{1}{2}\mathrm{Tr}\,(\boldsymbol{H}\ \boldsymbol{C}) = \frac{1}{2}\frac{\mathrm{Tr}\,(\boldsymbol{Y}\ \boldsymbol{C})}{N^6\ (1+\rho)\ X(\rho)} + J(\rho)\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} \tag{4.71}$$

where the second term turns out to be proportional to the approximation for the variance (4.18). Using (4.20)-(4.22) and (4.70) we can write the second term of (4.71) as a quotient of cubic forms in $\boldsymbol{\alpha}$:

$$J(\rho)\ \boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} = \frac{1}{K}\frac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{a}^{(\mathrm{J})}(\rho)\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}^{(\text{Var})}(\rho)\boldsymbol{\alpha}}{(1+\rho)[\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{b}^{(\text{Var})}(\rho)]^3}. \tag{4.72}$$

As for the computation of the first term of (4.71), which amounts to $L(\rho)$ in (4.24a), observe that the nonzero $[\boldsymbol{Y}]_{ij}$ have the form of (4.56) with $Q = 6-(i+j)$. We are eventually interested in the elements of the diagonal of $\boldsymbol{HC}$, which are given by

$$[\boldsymbol{HC}]_{ii} = \frac{\sum_{j=1}^{4}[\boldsymbol{Y}]_{ij}\,[\boldsymbol{C}]_{ji}}{N^6(1+\rho)X(\rho)} \tag{4.73}$$

for $i = 1,2,3,4$. All terms $\left([\boldsymbol{Y}]_{ij}\,[\boldsymbol{C}]_{ji}\right)$ have then the form of (4.56) with $Q = 6$. Clearly, N^6 cancels out in (4.73), so that the first term in (4.24) follows if we identify $L(\rho) = A^{(L)}(\rho)/B^{(L)}(\rho)$, where $A^{(L)}(\rho) = K\ \sum_{i=1}^{4}\sum_{j=1}^{4}[\boldsymbol{Y}]_{ij}\,[\boldsymbol{C}]_{ji}\,/(2N^6)$ and $B^{(L)}(\rho) = (1+\rho)X(\rho)$. Besides, $A^{(L)}(\rho)$ can be readily seen to be linear in $\boldsymbol{\alpha}$, since it is a linear combination of the elements of $\boldsymbol{Y}$, which are themselves linear in $\boldsymbol{\alpha}$ [cf. (4.66)]; this allows to write $A^{(L)}(\rho) = \sum_{n=0}^{6} A_n^{(L)}\rho^n = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{a}^{(L)}(\rho)$ as in (4.25), for some $\boldsymbol{a}^{(L)}(\rho) = \sum_{n=0}^{6}\boldsymbol{a}_n^{(L)}\rho^n$, in which $\boldsymbol{a}_n^{(L)}$ depends only on the constellation moments for all n.

Appendix 4.B Generalization: SNR Estimation from Ratios of Moments of Any Order

The approach to SNR estimation developed in Section 4.3 is based on ratios of moments of the form of (4.9). Such ratios are built with exclusively even-order moments in the numerator and M_2 in the denominator. Such moment selection allows to develop SNR estimators with the desirable feature that they can be implemented avoiding LUTs via some root-finding algorithm. If we drop this feature, i.e., if we allow LUTs in the implementation, then a wider choice of ratios is available for estimator design. For example one could try to develop estimators based on linear combinations of ratios of the family (4.6) proposed in [68]. However, a more general family can be considered.

In this Appendix we propose the development of SNR estimators based on of ratios of moments of the following form

$$f(\rho) = \frac{\prod_{k=1}^{U} M_{p(k)}^{r(k)}}{\prod_{k=1}^{V} M_{q(k)}^{s(k)}} \quad \text{such that} \tag{4.74a}$$

$$2Q \doteq \sum_{k=1}^{U} p(k)r(k) = \sum_{k=1}^{V} q(k)s(k) \tag{4.74b}$$

where we maintain the notation "$2Q$" used in Section 4.3 to denote the statistical order of numerator and denominator, which must be the same in order to ensure that the noise power N cancels out in the ratio [recall (4.4)]. Without loss of generality, we will assume that the numerator and denominator of (4.74) do not feature common moments, and that $p(k) \neq p(j)$, $q(k) \neq q(j)$ for $k \neq j$. Therefore (4.74) features a total of $U + V$ different moments. We note that the moment orders $p(k)$ and $q(k)$ can be odd. Examples of ratios belonging to this family are

$$\frac{M_1 M_3}{M_2^2}\ (2Q = 4),\ \frac{M_3 M_2}{M_1 M_4}\ (2Q = 5),\ \frac{M_5 M_3}{M_4 M_2 M_1^2}\ (2Q = 8), \ldots$$

The family of ratios (4.74) represents a general form which includes the ratios (4.6) and (4.9) as particular cases. The number of possible ratios within a given statistical order increases considerably with respect to those subfamilies. Table 4.2 shows this by comparing the total number of possible ratios within each (sub)family for each statistical order up to eighth-order. This fact represents a potential improvement over the family analyzed in Section 4.3: one can build a potentially competitive SNR estimator without moving to high statistical orders. For example, if we allow up to the third statistical order ($2Q \leq 3$) we have the following:

(i) there are no available ratios of the form of (4.9);
(ii) the family (4.6) offers four ratios: M_2/M_1^2, M_3/M_1^3, and their inverses;

Table 4.2: Number of existing ratios in the family (4.74), and the subfamilies (4.6) from [68], and (4.9)

Order $2Q$	Number of existing ratios*		
	(4.74)	(4.6)	(4.9)
2	2	2	0
3	4	2	0
4	12	2	1
5	16	2	0
6	48	4	1
7	56	2	0
8	134	4	2

*A ratio and its inverse are considered as different ratios except for the subfamily (4.9).

(iii) the general family (4.74) offers six ratios: M_2/M_1^2, M_3/M_1^3, $M_3/(M_1M_2)$ and their inverses.

We propose to estimate the SNR ρ using a linear combination of the form

$$\begin{aligned} f(\rho) &= \alpha_1 f_1(\rho) + \alpha_2 f_2(\rho) + \ldots + \alpha_t f_t(\rho) \\ &= \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{f}(\rho) \end{aligned} \tag{4.75}$$

with $\boldsymbol{\alpha} \doteq [\alpha_1 \; \alpha_2 \; \ldots \; \alpha_t]^{\mathrm{T}}$ a vector of t real-valued weights, and $\boldsymbol{f}(\rho) \doteq [f_1(\rho) \; f_2(\rho) \; \ldots \; f_t(\rho)]^{\mathrm{T}}$ the vector of the t ratios of moments belonging to the family (4.74). Similarly to Section 4.3.3, we introduce the vector of true moments $\boldsymbol{m} \doteq [m_1 \; m_2 \; \ldots \; m_\mu]^{\mathrm{T}}$ containing the μ different moments appearing in (4.75) ($\hat{\boldsymbol{m}}$ is the corresponding vector of sample moments), and the function $h(\boldsymbol{m}) = \alpha_1 h_1(\rho) + \alpha_2 h_2(\rho) + \ldots + \alpha_t h_t(\rho) \doteq f(\rho)$ which considers the moments as the independent variables in (4.75). The SNR estimate is then given by

$$\hat{\rho} = f^{-1}(h(\hat{\boldsymbol{m}})) \tag{4.76}$$

with $f^{-1}(\cdot)$ the inverse of (4.75), which in general will not have a closed-form (and thus the need of an LUT for implementing it).

Example For illustrative reasons, consider the linear combination of all ratios up to order three

$$f(\rho) = \alpha_1 \frac{M_2}{M_1^2} + \alpha_2 \frac{M_1^2}{M_2} + \alpha_3 \frac{M_3}{M_1^3} + \alpha_4 \frac{M_1^3}{M_3} + \alpha_5 \frac{M_3}{M_1 M_2} + \alpha_6 \frac{M_1 M_2}{M_3}. \tag{4.77}$$

We identify $t = 6$ (number of ratios), $\boldsymbol{m} = [M_1 \; M_2 \; M_3]^{\mathrm{T}}$ (and hence $\mu = 3$), and the obvious definition of the vector of ratios $\boldsymbol{f}(\rho)$. Considering the last ratio $f_6(\rho) = (M_1M_2)/M_3$, comparison with (4.74) allows to identify $U = 2$, $V = 1$, $p(1) = 1$, $p(2) = 2$, $q(1) = 3$, and $r(1) = r(2) = s(1) = 1$.

4.B.1 Statistical Analysis

We briefly sketch the statistical analysis of the proposed estimator (4.76), which proceeds similarly to the analysis performed in Section 4.3.3. Let us first define the function $g(\hat{\boldsymbol{m}}) \doteq f^{-1}(h(\hat{\boldsymbol{m}})) = \hat{\rho}$, whose implicit definition is given by $f(g) = h(\hat{\boldsymbol{m}})$.

Variance We approximate the variance through $\mathrm{Var}\{\hat{\rho}\} \approx \boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v}$, see (4.18), where in this case $\boldsymbol{v} \doteq \nabla g|_{\hat{\boldsymbol{m}}=\boldsymbol{m}}$ and $\boldsymbol{C}$ is the covariance matrix of $\hat{\boldsymbol{m}}$. After some algebraic steps it is found that

$$\nabla g = \frac{\nabla h}{f'(g)} \tag{4.78}$$

$$\nabla h = \hat{\boldsymbol{R}}\boldsymbol{\alpha} \tag{4.79}$$

with $\hat{\boldsymbol{R}}$ a $\mu \times t$ matrix with elements $[\hat{\boldsymbol{R}}]_{ij} = \partial h_j / \partial \hat{m}_i$. Hence,

$$\boldsymbol{v} = \frac{\boldsymbol{R}\boldsymbol{\alpha}}{f'(\rho)} \tag{4.80}$$

with $\boldsymbol{R} \doteq \hat{\boldsymbol{R}}|_{\hat{\boldsymbol{m}}=\boldsymbol{m}}$, so the approximate variance eventually results in

$$\mathrm{Var}\{\hat{\rho}\} \approx \frac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{R}\boldsymbol{\alpha}}{[f'(\rho)]^2}. \tag{4.81}$$

Note that, similarly to (4.20), the variance is a quotient of quadratic forms in $\boldsymbol{\alpha}$,[11] which is inversely proportional to the observation length K (due to $\boldsymbol{C}$).

Bias Recalling (4.23), the bias is approximated by $\mathrm{Bias}\{\hat{\rho}\} \approx \frac{1}{2}\mathrm{Tr}\,(\boldsymbol{H}\,\boldsymbol{C})$, with $\boldsymbol{H} \doteq \nabla^2 g|_{\hat{\boldsymbol{m}}=\boldsymbol{m}}$. After some algebra, and using the definitions $\hat{\boldsymbol{Y}} \doteq \nabla^2 h$ and $\boldsymbol{Y} \doteq \hat{\boldsymbol{Y}}|_{\hat{\boldsymbol{m}}=\boldsymbol{m}}$ it is found that

$$\nabla^2 g = \frac{1}{f'(g)}\hat{\boldsymbol{Y}} - \frac{f''(g)}{f'(g)}\nabla g(\nabla g)^{\mathrm{T}} \tag{4.82}$$

$$\boldsymbol{H} = \frac{1}{f'(\rho)}\boldsymbol{Y} - \frac{f''(g)}{f'(g)}\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}. \tag{4.83}$$

With this, the approximation of the bias reads

$$\mathrm{Bias}\{\hat{\rho}\} \approx \frac{1}{2}\frac{\mathrm{Tr}\,(\boldsymbol{Y}\boldsymbol{C})}{f'(\rho)} - \frac{1}{2}\frac{f''(\rho)}{f'(\rho)}\mathrm{Var}\{\hat{\rho}\} \tag{4.84}$$

which shares the structure of the approximation (4.24): the first term is a quotient of linear forms in $\boldsymbol{\alpha}$, and the second one is a quotient of cubic forms in $\boldsymbol{\alpha}$.

[11] Since $f(\rho)$ is linear in $\boldsymbol{\alpha}$, see (4.75), and so is $f'(\rho)$.

Weight optimization The optimization of the weight vector $\boldsymbol{\alpha}$ must be performed for each particular linear combination of ratios. In any case, the principles that inspired the optimization criteria proposed in Section 4.3.4, and the mathematical tools used therein are valid in this case as well.[12] The only difference with respect to the estimators proposed in Section 4.3 is that, in general, the elements of the vectors and matrices involved in (4.81) and (4.84) will not be polynomials in ρ, but more complicated functions of ρ. This fact may yield the optimization procedure somewhat more cumbersome. Intuitively, increasing t (number of ratios in the linear combination) should translate into a performance improvement in terms of MSE.[13] Based on this, the proposed family of estimators has the potential to outperform the eight-order estimator of Section 4.3 with statistics of lower order. However, this is a conjecture that needs corroboration, and is left here as a future research line.

Appendix 4.C SNR Estimation with Heterogeneous Frames

4.C.1 A Suboptimal Approach to Weight Selection for the CCM Estimator

A simpler, suboptimal approach is to choose $\lambda_{\text{ccm},n}$ so as to minimize the variance of $\hat{M}_n^{\text{ccm}}$, since this should intuitively translate into a smaller variance of $\hat{\rho}_{24}^{\text{ccm}}$. The variance of the combined moments is given by

$$\text{Var}\left\{\hat{M}_n^{\text{ccm}}\right\} = \lambda_{\text{ccm},n}^2\left(\sigma_n^{(1)}\right)^2 + (1-\lambda_{\text{ccm},n})^2\left(\sigma_n^{(2)}\right)^2 \tag{4.85}$$

where the variances in the right hand side are those of the constituent sample moments, and are given by

$$\left(\sigma_n^{(i)}\right)^2 \doteq \text{Var}\left\{\hat{M}_n^{(i)}\right\} = \left[M_{2n}^{(i)} - \left(M_n^{(i)}\right)^2\right]/K_i. \tag{4.86}$$

With this we find that the value minimizing (4.85) is given by

$$\lambda_{\text{ccm},n} = \frac{\left(\sigma_n^{(1)}\right)^2}{\left(\sigma_n^{(1)}\right)^2 + \left(\sigma_n^{(2)}\right)^2}. \tag{4.87}$$

Notice the similarity with (4.48). For $n = 2,\ 4$ the variance ratios in (4.87) equal

$$\frac{\left(\sigma_2^{(1)}\right)^2}{\left(\sigma_2^{(2)}\right)^2} = \frac{K_2}{K_1}\frac{A^{(1)}(\rho)}{A^{(2)}(\rho)}, \qquad \frac{\left(\sigma_4^{(1)}\right)^2}{\left(\sigma_4^{(2)}\right)^2} = \frac{K_2}{K_1}\frac{B^{(1)}(\rho)}{B^{(2)}(\rho)} \tag{4.88}$$

[12] The actual optimization criterion must be tailored to each particular case, considering the number of scalar weights available for optimization.

[13] Recall that the dimension t of the weight vector $\boldsymbol{\alpha}$ represents in fact $t-1$ degrees of freedom available for estimator design, since the variance (4.81) and the bias (4.84) are invariant under scalings in $\boldsymbol{\alpha}$.

where

$$A^{(i)}(\rho) \doteq \left[c_4^{(i)} - \left(c_2^{(i)}\right)^2\right]\rho^2 + 2c_2^{(i)}\rho + 1, \tag{4.89}$$

$$B^{(i)}(\rho) \doteq \left[c_8^{(i)} - \left(c_4^{(i)}\right)^2\right]\rho^4 + 8\left[2c_6^{(i)} - c_2^{(i)}c_4^{(i)}\right]\rho^3 + \left[68c_4^{(i)} - 16\left(c_2^{(i)}\right)^2\right]\rho^2 + 80c_2^{(i)}\rho + 80. \tag{4.90}$$

Notice that for low SNR values both quotients in (4.88) approach 1, which leads to $\lambda_{\text{ccm},n} \approx K_1/(K_1 + K_2) = p$. On the other hand, for high SNR the terms in ρ^2 (resp. ρ^4) dominate in $A^{(i)}(\rho)$ [resp. $B^{(i)}(\rho)$], and thus $\lambda_{\text{ccm},n} \approx K_1/(K_1 + \alpha_n K_2)$, where

$$\alpha_n \doteq \frac{c_{2n}^{(1)} - \left(c_n^{(1)}\right)^2}{c_{2n}^{(2)} - \left(c_n^{(2)}\right)^2} \qquad \text{for } n = 2, 4.$$

Chapter 5

Phase Estimation for Cross-QAM Constellations

Contents

5.1 Introduction

Carrier phase recovery is a key element in bandpass digital communication receivers. Many authors have addressed this topic over the last years, yielding a wide variety of recovery methods depending on the modulation scheme and application constraints [76,

142]. The evolution of digital communications has fostered the use of denser constellations such as Quadrature Amplitude Modulation (QAM) at higher transmission rates, posing new challenges to the receiver operation in general; and in particular to the estimation of the unknown carrier phase offset, a task that can be performed resorting to either trained or non-data-aided (NDA, or *blind*) methods. Our focus is on NDA phase estimation schemes (since they provide better spectral efficiency) for QAM communication systems.

Existing NDA phase estimation methods can be roughly classified into two categories. Histogram-based schemes [76, 129] estimate the probability density function (pdf) of the phase of a suitable nonlinear transformation of the observed data. The phase offset affects this pdf as a cyclic shift, from which it can be estimated. Although these methods present good performance, their computational load is relatively high. In addition, they require previous estimation of the channel gain [76] or the SNR operation point [129].

On the other hand, higher-order statistics (HOS)-based methods obtain the phase estimate as a function of the sample averages of nonlinear transformations of the data. This class includes the classical fourth-power (4P) estimator [122], the Viterbi and Viterbi (V&V) family of estimators [161], which were originally proposed for Phase-Shift Keying (PSK) modulation, and eighth-order estimates such as that of Cartwright (C8) [29] and the so-called concentration ellipse orientation (CEO) method [28]. All of these apply fixed nonlinear transformations to the observed data, and do not require previous gain control or knowledge of the SNR. Generalizations of the V&V family were proposed in [163, 164], where the nonlinearity is matched to the particular constellation and optimized in terms of the estimation variance. This approach provides good performance, but similarly to histogram-based methods, it requires knowledge of the overall gain and of the effective SNR.

Low computational complexity makes the standard 4P estimator very attractive for practical implementation. Its asymptotic variance was analyzed in [122, 142], showing that for a given sample size, it does not decrease with increasingly high SNR (except in the case of a QPSK constellation). This error floor is induced by the so-called *self-noise* due to the multimodulus nature of QAM constellations. Despite this drawback, performance is considered acceptable for square QAM constellations. In fact, it is known that the NDA Maximum Likelihood estimator reduces to the 4P estimator as the SNR approaches zero [122]. However, its self-noise-induced error floor is much higher for cross-QAM constellations, such as the 32-QAM and 128-QAM displayed in Figure 5.1, in which several corner symbols in the constellation (those with larger modulus) are missing. For these cases, eighth-order methods such as C8 and CEO constitute a better choice, since they provide lower variance floors at medium to high SNR, at the expense of an increase in complexity.

Like other dense constellations, cross-QAM constellations such as the 32-QAM and 128-QAM are very sensitive to phase errors. This is illustrated in Figure 5.2, which shows the Symbol Error Rate (SER) versus the SNR per bit (E_b/N_0) curves for uncoded 32-QAM and 128-QAM in additive white Gaussian noise (AWGN) under a number of static phase offsets.[1] Clearly, even relatively small phase errors may incur a substantial penalty

[1] A closed-form expression for the SER of static-phase-rotated cross-QAM in AWGN is provided in [2].

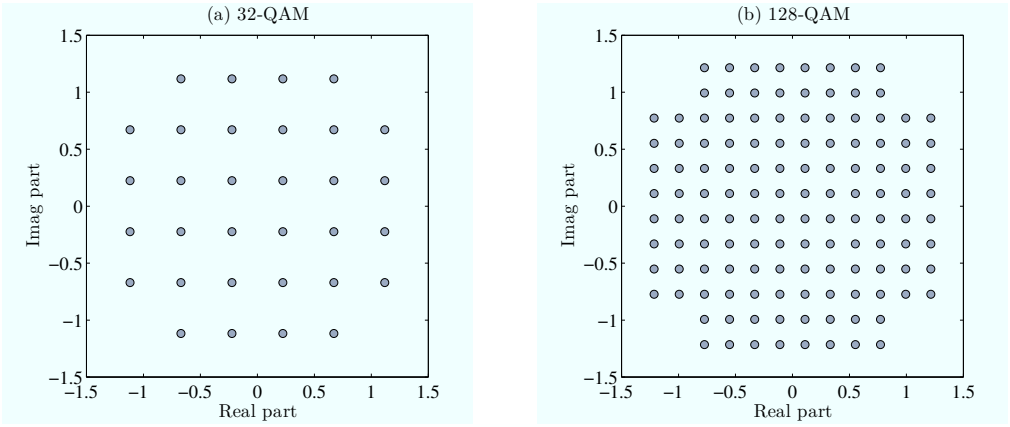
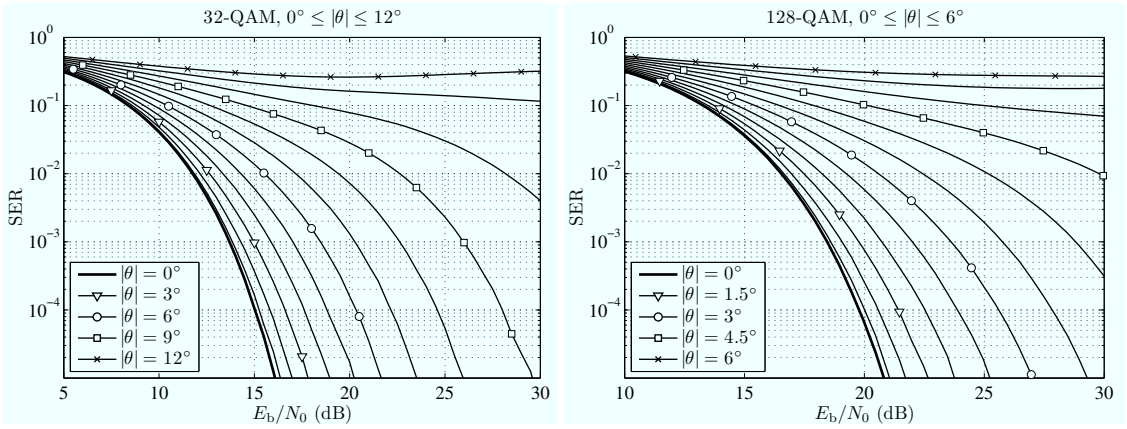


Figure 5.1: Examples of two cross-QAM constellations normalized in energy: (a) 32-QAM, (b) 128-QAM.



(a) (b)

Figure 5.2: SER versus E_b/N_0 for (a) 32-QAM and (b) 128-QAM with different static phase offsets.

in terms of E_b/N_0 (e.g. 1.25 and 5.25 dB for 32-QAM and 128-QAM respectively, for a phase offset of $3°$ at a target raw SER of 10^{-3}).

Practical implementation of digital receivers must usually face hardware limitations imposed by cost, size, speed, and power consumption constraints. In those situations, fixed point arithmetic devices may be the only available choice. Similarly, if high-speed analog-to-digital converters (ADCs) are required, the available resolution may not be sufficiently high so as to ignore quantization effects in the design. In particular, these effects will impact different phase estimation algorithms in a different way, so that the choice of an estimator will be determined by its behavior with quantized data and fixed point arithmetic processing, rather than by its theoretical asymptotic variance under infinite pre-

We note that the SER curves under realistic phase detection will differ in general due to the randomness of the residual phase offset after correction, although Figure 5.2 can provide a rough idea of the harmful effects of small phase offsets on the SER.

cision. Although estimation performance is difficult to analyze under finite wordlength constraints, intuitively one would expect that estimators requiring fewer, lower-order operations on the observed data be more robust in these situations. For example, the performance of the 4P phase estimate is expected to degrade more gracefully than that of eighth-order methods such as C8 and CEO.

These considerations motivate the search for low-complexity NDA phase estimators for cross-QAM systems. In this chapter we develop two such schemes by maximizing certain cost functions arising from geometrical considerations based on the square-like shape of QAM constellations. Although it is not possible to derive closed-form expressions for these maximizers, we present simple iterations which locally converge to the desired values, and which can be initialized by any suitable scheme providing a coarse phase estimate. The choice of the initializer should be tailored to the operation environment (constellation size, number of samples, availability of training data, etc.).

Chapter structure The chapter is organized as follows. Section 5.2 presents the signal model and reviews the classical one-shot estimators that will be considered in the chapter. Maximization of the ℓ_1-norm as a criterion for phase estimation is justified in Section 5.3, whereas the two proposed estimators are derived in Section 5.4. Section 5.5 provides an analysis in terms of variance, initialization and computational complexity. Simulation results, for both floating point and fixed point implementations, are shown in Section 5.6, and conclusions are drawn in Section 6.1. The contents of this chapter derive from the work presented in [5, 8].

5.2 System Model and Classical Estimators

Consider the receiver of QAM system, in which the received signal is sampled at the baud rate after front-end processing. Assuming that carrier frequency recovery has been previously established, the observed data can be written as

$$r_k = a_k \mathrm{e}^{j\theta} + n_k, \qquad k = 0, 1, \ldots, L-1 \tag{5.1}$$

where $\{a_k\}$ are the transmitted symbols, drawn equiprobably from a constellation $\mathcal{A}$ with size M and variance E_a, and $\{n_k\}$ are the complex-valued noise samples. The noise is assumed zero-mean, circular white Gaussian with variance σ^2, and independent of the symbols. The phase offset θ is assumed to vary slowly; thus, it can be taken as constant within the block of L samples. The goal is to identify θ without knowledge of the symbols a_k or the variances E_a, σ^2. Due to the quadrant symmetry of QAM constellations, this phase offset can only be blindly identified up to a four-fold ambiguity; hence, we assume that $|\theta| < \pi/4$.

The V&V family of estimators [161] embraces a wide range of choices. Of particular interest are power-law estimators

$$\hat{\theta}_{\text{VV-}p} = \frac{1}{4}\arg\left\{ -\sum_{k=0}^{L-1} |r_k|^p \mathrm{e}^{j4\arg\{r_k\}} \right\} \tag{5.2}$$

where p is an integer. For $p = 0$, the corresponding estimate $\hat{\theta}_{\text{VV-0}}$ discards the envelope information of the received signal. On the other hand, for $p = 4$ one obtains the 4P estimator [122, 142]:

$$\hat{\theta}_{\text{4P}} = \frac{1}{4}\arg\left\{-\sum_{k=0}^{L-1} r_k^4\right\}. \tag{5.3}$$

Several alternatives have been proposed to improve the performance of the 4P method with cross-QAM constellations. One of them advocates the use of (5.2) with negative values of p, although the resulting estimators are competitive in the high SNR regime only [31]. An alternative approach with better performance at lower SNRs is Cartwright's eighth-order estimator [29]

$$\hat{\theta}_{\text{C8}} = \frac{1}{4}\tan^{-1}\left(\frac{\sum_k A_k B_k \sum_l A_l C_l - \sum_k A_k^2 \sum_l B_l C_l}{\sum_k A_k B_k \sum_l B_l C_l - \sum_k B_k^2 \sum_l A_l C_l}\right) \tag{5.4}$$

where $A_k \doteq \text{Re}\{r_k^4\}$, $B_k \doteq \text{Im}\{r_k^4\}$ and $C_k \doteq |r_k|^4$, and all summations in k and l in (5.4) run from 0 to $L-1$. We note that the four quadrant inverse tangent function is required in (5.4). The price to pay for the reduced variance of (5.4) with respect to that of (5.3) is a higher computational load. A different eighth-order estimator (CEO) was proposed in [28]; its computational complexity and performance with cross-QAM constellations are similar to those of (5.4).

5.3 Phase Estimation Based on ℓ_1-norm Maximization

Consider the mapping $f : \mathbb{C} \to \mathbb{R}^2$ that assigns to each complex number $z = \bar{z} + j\tilde{z}$ the real vector $\boldsymbol{z} \doteq f(z) = [\bar{z} \quad \tilde{z}]^{\text{T}}$. Under f, multiplication by a phase term $\text{e}^{j\theta}$ in $\mathbb{C}$ becomes multiplication by an orthogonal rotation matrix in $\mathbb{R}^2$. Given a candidate estimate $\hat{\theta}$, let us define the de-rotated samples

$$y_k = \text{e}^{-j\hat{\theta}} r_k = \bar{y}_k + j\tilde{y}_k \tag{5.5}$$

and consider the effect of such de-rotation on the vector $\boldsymbol{y}_k = f(y_k) = [\bar{y}_k \quad \tilde{y}_k]^{\text{T}}$. In particular, we focus on the resulting ℓ_1-norm of the phase-compensated vector

$$||\boldsymbol{y}_k||_1 = |\bar{y}_k| + |\tilde{y}_k|. \tag{5.6}$$

We claim that maximization of the ℓ_1-norm with respect to $\hat{\theta}$ is a meaningful criterion for obtaining a phase estimate with QAM constellations. To see this, it is instructive to consider the geometry of the simplest case, i.e. a QPSK constellation ($M = 4$) $\mathcal{A} = \{\sqrt{E_a}e^{j(\frac{\pi}{4}+n\frac{\pi}{2})},\ 0 \le n \le 3\}$ in the noiseless case. Recall that the ℓ_1-ball defined by $||\boldsymbol{y}||_1 = \gamma$ is a diamond with corners at $[0 \ \pm\gamma]^{\text{T}}$ and $[\pm\gamma \ \ 0]^{\text{T}}$. It is clear from Figure 5.3 that any rotation of $\mathcal{A}$ by an angle that is not of the form $n\pi/2$ with n integer can only result in a smaller average ℓ_1-norm of the rotated constellation. That is, the ℓ_1-norm is

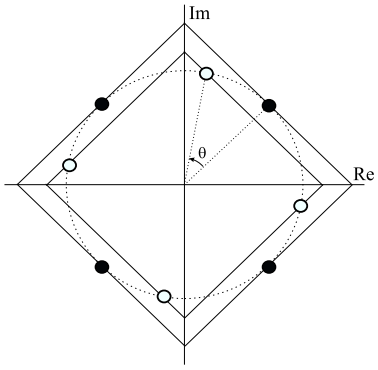


Figure 5.3: Effect of a rotation on the average ℓ_1-norm of a QPSK constellation. Solid dots: original constellation; empty dots: rotated constellation.

maximized when the positioning of the constellation is restored to its original regular grid.

Observe that for QPSK the average ℓ_1-norm of the constellation coincides with the ℓ_1-norm of any of its elements, for all phase offsets. This is because this constellation can be obtained by picking any of its elements and all possible rotations by an integer multiple of $\pi/2$ rad, and the ℓ_1-norm is preserved by these operations, as it can be easily checked.[2] For larger constellations this property does not hold anymore, although it can be checked that the average ℓ_1-norm still attains its maximum for a zero phase offset in absence of noise. This is shown in Figures 5.4 and 5.5, which plot the cost functions

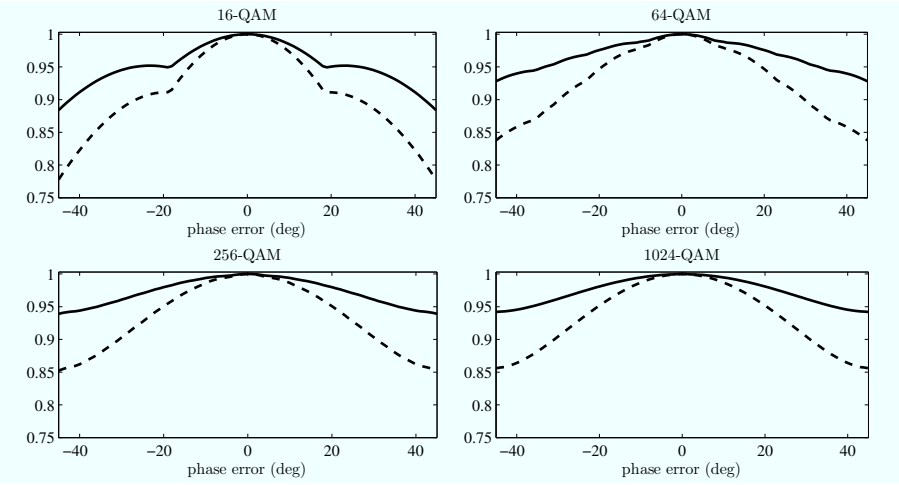


Figure 5.4: Square QAM constellations: behavior of the ℓ_1-norm based cost functions $J_q(\hat{\theta})$ (normalized to their maximum value) in absence of noise. $q = 1$ (solid) and 2 (dashed).

$$J_q(\hat{\theta}) \doteq \mathrm{E}\left\{||\boldsymbol{y}||_1^q\right\} = \mathrm{E}\left\{(|\bar{y}| + |\tilde{y}|)^q\right\} \tag{5.7}$$

[2]Conjugation is another ℓ_1-norm-preserving operation.

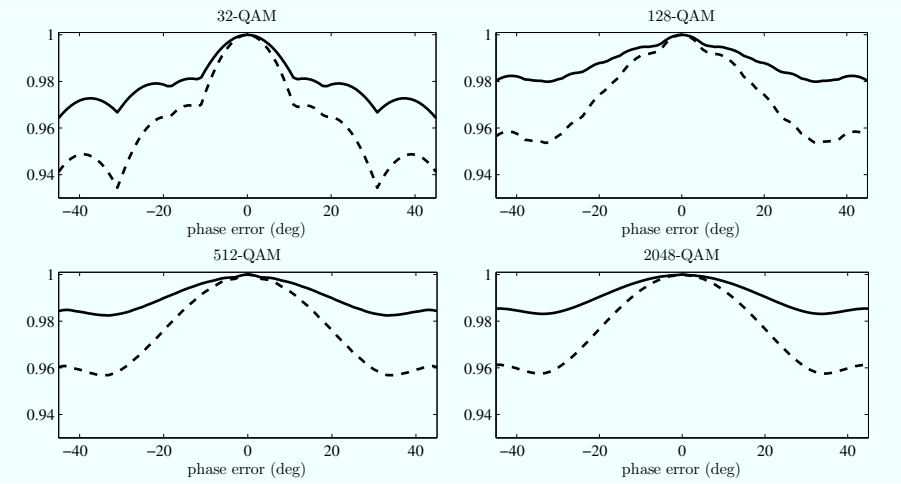


Figure 5.5: Cross QAM constellations: behavior of the ℓ_1-norm based cost functions $J_q(\hat{\theta})$ (normalized to their maximum value) in absence of noise. $q = 1$ (solid) and 2 (dashed).

for $q = 1$ and 2, in terms of the residual phase error $\theta - \hat{\theta}$, and in absence of noise.[3] The jagged appearance of J_1 and J_2 is due to the fact that the ℓ_1-norm is not everywhere differentiable. Besides the desired global maximum at zero phase offset, these cost functions present additional local maxima for 16-QAM as well as for cross-QAM constellations; for square QAM with $M > 16$, both are unimodal.

In the Appendix we prove that an alternative approach based on the maximization of the ℓ_∞-norm based costs $K_q(\hat{\theta}) \doteq \mathrm{E}\left\{||\boldsymbol{y}||_\infty^q\right\}$ is equivalent to the minimization of the ℓ_1-norm based costs in terms of phase estimation.

5.4 Fixed Point Iterations for Phase Estimation

Since no closed-form expressions for the maxima of $J_q(\hat{\theta})$ are available, one must resort to numerical optimization strategies, e.g. gradient or subgradient ascent, or Newton's method. We propose alternative fixed point iterations derived from the maximization of the proposed cost functions, which are computationally simple and do not require stepsize tuning. Our focus is on $J_q(\hat{\theta})$ for $q = 1$ and 2, as their lower order yields computationally simpler schemes robust to finite precision effects, as will be shown in Section 5.6.

[3]The noiseless cost function J_q for QPSK is straightforwardly found to be $J_q(\hat{\theta}) = J_1^q(\hat{\theta}) = C \cdot [|\cos(\theta_\mathrm{e}) - \sin(\theta_\mathrm{e})| + |\cos(\theta_\mathrm{e}) + \sin(\theta_\mathrm{e})|]^q$, where C is an irrelevant constant, and θ_e is the phase error defined as $\theta_\mathrm{e} = \theta - \hat{\theta}$.

5.4.1 Maximization of $J_1(\hat{\theta})$

The goal is to find a maximum of $J_1(\hat{\theta}) \doteq \mathrm{E}\{|\bar{y}_k| + |\tilde{y}_k|\}$. If J_1 is differentiable at such point, then $\partial J_1(\hat{\theta})/\partial\hat{\theta}$ must vanish. From (5.5), note that

$$\frac{\partial y_k}{\partial\hat{\theta}} = -jy_k \quad \Rightarrow \quad \frac{\partial\bar{y}_k}{\partial\hat{\theta}} = \tilde{y}_k,\ \frac{\partial\tilde{y}_k}{\partial\hat{\theta}} = -\bar{y}_k. \tag{5.8}$$

Therefore, at the points at which J_1 is differentiable,

$$\frac{\partial J_1(\hat{\theta})}{\partial\hat{\theta}} = \mathrm{E}\left\{\mathrm{sgn}\left(\bar{y}_k\right)\tilde{y}_k - \mathrm{sgn}\left(\tilde{y}_k\right)\bar{y}_k\right\}. \tag{5.9}$$

If $\hat{\theta}_*$ is a maximum of J_1, then (5.9) equals zero (if it exists), i.e.

$$\mathrm{E}\left\{\mathrm{sgn}\left(\bar{y}_k\right)\tilde{y}_k\right\} = \mathrm{E}\left\{\mathrm{sgn}\left(\tilde{y}_k\right)\bar{y}_k\right\}. \tag{5.10}$$

Now, if we write $r_k = \bar{r}_k + j\tilde{r}_k$, then from (5.5)

$$\bar{y}_k = \bar{r}_k\cos\hat{\theta} + \tilde{r}_k\sin\hat{\theta}, \quad \tilde{y}_k = \tilde{r}_k\cos\hat{\theta} - \bar{r}_k\sin\hat{\theta} \tag{5.11}$$

and we arrive at

$$\tan\hat{\theta}_* = \frac{\mathrm{E}\left\{\mathrm{sgn}\left(\bar{y}_k\right)\tilde{r}_k - \mathrm{sgn}\left(\tilde{y}_k\right)\bar{r}_k\right\}}{\mathrm{E}\left\{\mathrm{sgn}\left(\bar{y}_k\right)\bar{r}_k + \mathrm{sgn}\left(\tilde{y}_k\right)\tilde{r}_k\right\}} \tag{5.12}$$

$$= -\frac{\mathrm{E}\left\{\mathrm{Im}\left\{\mathrm{csgn}\left(y_k\right)\cdot r_k^*\right\}\right\}}{\mathrm{E}\left\{\mathrm{Re}\left\{\mathrm{csgn}\left(y_k\right)\cdot r_k^*\right\}\right\}} \tag{5.13}$$

$$= \tan\left(-\mathrm{arg}\left\{\mathrm{E}\left\{\mathrm{csgn}\left(y_k\right)\cdot r_k^*\right\}\right\}\right). \tag{5.14}$$

Note that (5.14) characterizes the extrema of J_1 only implicitly, since the right-hand side of (5.14) depends on $\hat{\theta}_*$.[4] Nevertheless, this condition suggests a fixed point iteration to obtain $\hat{\theta}_*$. Given the observations $\{r_k\}$ and a suitable initialization $\hat{\theta}_0^{(1)}$, we substitute the expectation in (5.14) by a sample mean and then iteratively compute

$$\hat{\theta}_{n+1}^{(1)} = -\mathrm{arg}\left\{\sum_{k=0}^{L-1}\mathrm{csgn}\left(r_k\cdot\mathrm{e}^{-j\hat{\theta}_n^{(1)}}\right)\cdot r_k^*\right\}. \tag{5.15}$$

5.4.2 Maximization of $J_2(\hat{\theta})$

The cost function J_2 can be written as

$$J_2(\hat{\theta}) = \mathrm{E}\left\{\left(|\bar{y}_k| + |\tilde{y}_k|\right)^2\right\} = \mathrm{E}\left\{|y_k|^2\right\} + 2\cdot\mathrm{E}\left\{|\bar{y}_k|\cdot|\tilde{y}_k|\right\}. \tag{5.16}$$

[4] We note that a different expression resembling (5.14) can be found in [83, Eq. (8)], yet in a gain-control-based context.

Since $\mathrm{E}\{|y_k|^2\}$ is invariant under phase rotations, maximizing $J_2(\hat{\theta})$ amounts to maximizing $\mathrm{E}\{|\bar{y}_k\tilde{y}_k|\}$. Using basic trigonometric relations, it is found that

$$\bar{y}_k\tilde{y}_k = \bar{r}_k\tilde{r}_k\cos(2\hat{\theta}) - \frac{1}{2}(\bar{r}_k^2 - \tilde{r}_k^2)\sin(2\hat{\theta}). \tag{5.17}$$

Therefore, at the points at which J_2 is differentiable,

$$\begin{aligned}\frac{\partial J_2(\hat{\theta})}{\partial\hat{\theta}} &= -\mathrm{E}\left\{\mathrm{sgn}\left(\bar{y}_k\tilde{y}_k\right)\ (\bar{r}_k^2 - \tilde{r}_k^2)\right\}\cos(2\hat{\theta}) \\ &\quad - \mathrm{E}\left\{\mathrm{sgn}\left(\bar{y}_k\tilde{y}_k\right)\ 2\bar{r}_k\tilde{r}_k\right\}\sin(2\hat{\theta}).\end{aligned} \tag{5.18}$$

Observe that

$$\bar{r}_k^2 - \tilde{r}_k^2 = \mathrm{Re}\left\{r_k^2\right\}, \quad 2\bar{r}_k\tilde{r}_k = \mathrm{Im}\left\{r_k^2\right\}, \tag{5.19}$$

$$s_k(\hat{\theta}) \doteq \mathrm{sgn}\left(\bar{y}_k\tilde{y}_k\right) = \mathrm{sgn}\left(\mathrm{Im}\left\{r_k^2\mathrm{e}^{-j2\hat{\theta}}\right\}\right). \tag{5.20}$$

If $\hat{\theta}_*$ is a maximum of J_2 at which J_2 happens to be differentiable, then (5.18) must be zero, i.e.,

$$\tan(2\hat{\theta}_*) = -\frac{\mathrm{E}\left\{\mathrm{Re}\left\{r_k^2\right\}\, s_k(\hat{\theta}_*)\right\}}{\mathrm{E}\left\{\mathrm{Im}\left\{r_k^2\right\}\, s_k(\hat{\theta}_*)\right\}} \tag{5.21}$$

$$= \tan\left(\arg\left\{-j\mathrm{E}\left\{r_k^2\, s_k(\hat{\theta}_*)\right\}\right\}\right). \tag{5.22}$$

Since the right-hand side of (5.22) depends on $\hat{\theta}_*$, this condition does not provide the desired solution in closed form, but similarly to (5.15), it suggests a fixed point iteration. Given the observations $\{r_k\}$ and a suitable initialization $\hat{\theta}_0^{(2)}$, the expectation in (5.22) is replaced by a sample mean and then $\hat{\theta}_n^{(2)}$ is iteratively computed as follows:

$$s_k(\hat{\theta}_n^{(2)}) = \mathrm{sgn}\left(\mathrm{Im}\left\{r_k^2\mathrm{e}^{-j2\hat{\theta}_n^{(2)}}\right\}\right),\ k = 0, \ldots, L-1 \tag{5.23a}$$

$$\hat{\theta}_{n+1}^{(2)} = \frac{1}{2}\arg\left\{\sum_{k=0}^{L-1}\left[r_k^2 s_k(\hat{\theta}_n^{(2)})\right]\right\} - \frac{\pi}{4}. \tag{5.23b}$$

5.5 Discussion and Statistical Analysis

5.5.1 Algorithm Initialization

The cost functions $J_q(\hat{\theta})$ are multimodal in general, due to finite sample effects; in the case of cross-QAM, multimodality holds even as $L \to \infty$, see Figure 5.5. Hence the proposed iterative schemes may experience convergence to undesirable solutions. Initialization must be good enough to set the starting point within the domain of attraction of the global maximum. Potential initializers include those methods discussed in Section 5.2. The use of the novel iterative methods as refinements of the standard estimators of Section 5.2 is justified only if they are able to provide better performance than the initializer, and at a reasonable computational cost. Both issues are discussed next.

5.5.2 Computational Complexity

Consider the standard 4P estimator (5.3). Obtaining the averaged value $\sum_k r_k^4$ requires $5L$ real multiplications and $4L$ real additions [31]. On the other hand, computation of the C8 estimate (5.4) requires $11L$ real multiplications and $8L$ real additions. Thus, the C8 estimate is about twice as expensive as the standard fourth-phase estimate.

However, quantifying the computational load associated to a given estimator is not always such a straightforward task. Consider for instance the estimate $\hat{\theta}_{\text{VV-0}}$ from (5.2), which must compute $\sum_k \mathrm{e}^{j4\arg\{r_k\}} = (\sum_k \cos(4\arg\{r_k\}) + j\sum_k \sin(4\arg\{r_k\})$. This in turn requires (i) extraction of the phase of the complex numbers r_k, i.e., Cartesian-to-Polar (C2P) conversion; (ii) computation of the terms $\cos(4\arg\{r_k\})$ and $\sin(4\arg\{r_k\})$; and (iii) averaging these values. The last step amounts to $2L$ real additions; however, the first two steps are likely to be implemented as look-up table (LUT) operations. Thus, complexity comparisons of the 4P, C8, and V&V-0 estimators will depend on the relative costs assigned to memory area, hardware multipliers, etc.[5]

In a similar vein, iteration (5.15) requires L phase rotations (in order to compute $\mathrm{csgn}\left(r_k \cdot \mathrm{e}^{-j\hat{\theta}_n^{(1)}}\right)$) and $2L$ real additions per iteration. Phase rotations can be performed very efficiently (with shift-and-add operations only) with CORDIC hardware [85, 114]. Alternatively, with a previous C2P module providing $|r_k|$ and $\arg\{r_k\}$, determination of the csgn terms only requires L additions per iteration. The scheme (5.23) is slightly more costly than (5.15) due to the need to precompute r_k^2, although it can also benefit from C2P/CORDIC modules for the computation of $s_k(\hat{\theta}_n^{(2)})$. Complexity evaluation of the proposed estimators should also take into account the need for proper initialization, as well as the fact that some of the partial results of the initialization stage can be reused in subsequent iterations.

5.5.3 Asymptotic Variances

Using a small-error analysis, it is possible to obtain closed-form expressions for the asymptotic variances of the estimates $\hat{\theta}^{(q)} \doteq \lim_{n\to\infty} \hat{\theta}_n^{(q)}$, $q = 1, 2$. Details are given in Appendix 5.B, and the results are summarized in the following theorem.

Theorem 1. *Assuming that the iterations* (5.15) *and* (5.23) *are initialized sufficiently close to the true phase offset θ and that the signal-to-noise ratio $\rho \doteq E_a/\sigma^2$ is sufficiently high, the estimates $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ are asymptotically unbiased, and their asymptotic variances are given by*

$$\lim_{L\to\infty} L\,\mathrm{E}\left\{(\hat{\theta}^{(1)} - \theta)^2\right\} = \frac{1}{2d_1^2}\left((2 - d_2) + \frac{2}{\rho}\right) \tag{5.24}$$

$$\lim_{L\to\infty} L\,\mathrm{E}\left\{(\hat{\theta}^{(2)} - \theta)^2\right\} = \frac{1}{4\tilde{g}_2^2}\left(\frac{c_4 + \bar{b}_4}{2} + \frac{2}{\rho} + \frac{1}{\rho^2}\right) \tag{5.25}$$

[5]Note that the operation counts given above for the 4P and C8 schemes will change if a previous LUT-based C2P stage is applied to the observed data, which makes direct comparison with V&V-0 even more difficult.

Table 5.1: Asymptotic variance ratios (5.26) for several QAM constellation sizes.

	Square QAMs				Cross-QAMs			
M	16	64	256	1024	32	128	512	2048
α_1	2.01	0.93	0.81	0.79	0.06	0.06	0.05	0.05
α_2	2.01	0.98	0.86	0.84	0.09	0.08	0.08	0.08

where b_4, c_4, g_2, d_n ($n = 1, 2$) are constellation-dependent constants:

$$b_4 \doteq \frac{\mathrm{E}\left\{a_k^4\right\}}{E_a^2} \qquad c_4 \doteq \frac{\mathrm{E}\left\{|a_k|^4\right\}}{E_a^2}$$
$$g_2 \doteq \frac{\mathrm{E}\left\{a_k^2 \operatorname{sgn}\left(\operatorname{Im}\left\{a_k^2\right\}\right)\right\}}{E_a} \qquad d_n \doteq \frac{\mathrm{E}\left\{(a_k^* \operatorname{csgn}\left(a_k\right))^n\right\}}{E_a^{n/2}} \quad n = 1, 2.$$

The asymptotic variances (5.24)-(5.25) depend on the constellation and the SNR, but not on the value of θ; this is also the case for the 4P estimate, whose asymptotic variance is given in [30, 142]. Using that expression and (5.24)-(5.25), one can obtain the ratio of limits (as the SNR tends to infinity) of the asymptotic variances of the new estimates to that of the 4P estimate

$$\alpha_q \doteq \frac{\lim_{\rho,L\to\infty} L\,\mathrm{E}\left\{(\hat{\theta}^{(q)} - \theta)^2\right\}}{\lim_{\rho,L\to\infty} L\,\mathrm{E}\left\{(\hat{\theta}_{\mathrm{4P}} - \theta)^2\right\}}. \tag{5.26}$$

These ratios are given in Table 5.1, showing that the new schemes perform similarly to the 4P estimator with square QAM, but they consistently offer a sizable improvement (about one order of magnitude reduction of the asymptotic variance) with cross-QAM constellations.

5.6 Simulation Results

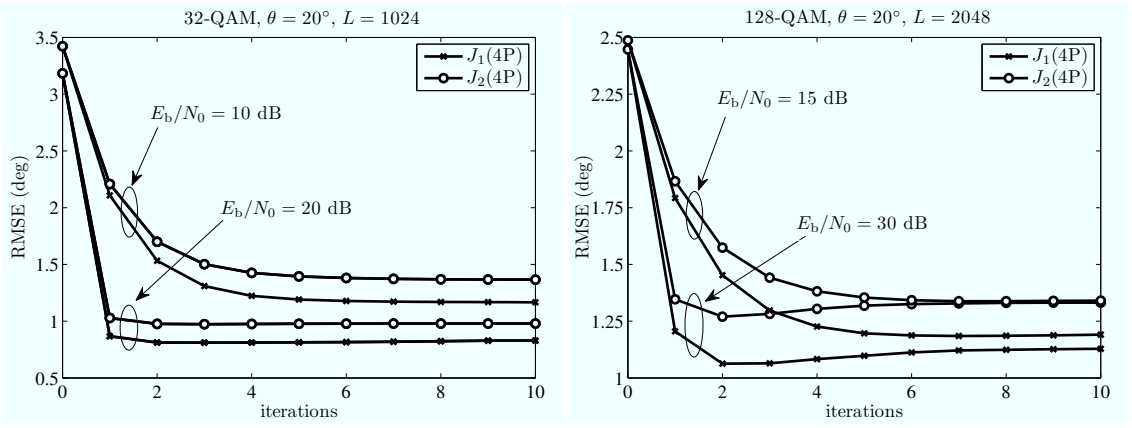
We present the results obtained with the novel iterative estimators, using Monte Carlo simulations, under both floating point and fixed point implementations. The focus is on 32-QAM and 128-QAM cross constellations. Results are presented in terms of bias and Root Mean Square Error (RMSE), defined as

$$\mathrm{RMSE} \doteq \sqrt{\mathrm{E}\left\{\min\left\{\left(\hat{\theta} - \theta\right)^2, \left(\hat{\theta} - \theta + \frac{\pi}{2}\right)^2, \left(\hat{\theta} - \theta - \frac{\pi}{2}\right)^2\right\}\right\}}$$

where the $\min\{\cdot\}$ function represents the fact that a practical receiver should be able to remove the $\pi/2$-ambiguity of the QAM phase estimator. Bias and RMSE are estimated via the Monte Carlo method.

5.6.1 Floating Point Precision

Convergence properties were investigated first, with a phase offset $\theta = 20°$. Figure 5.6 shows the RMSE as a function of the number of iterations. Both estimators were initialized using the 4P method (the RMSE shown for iteration zero corresponds to $\hat{\theta}_{4P}$), and different values of the SNR per bit [defined as $E_b/N_0 \doteq E_a/(\sigma^2 \log_2 M)$] are considered. Convergence slows down with decreasing E_b/N_0 as expected, but nevertheless, for practical situations a few iterations suffice.



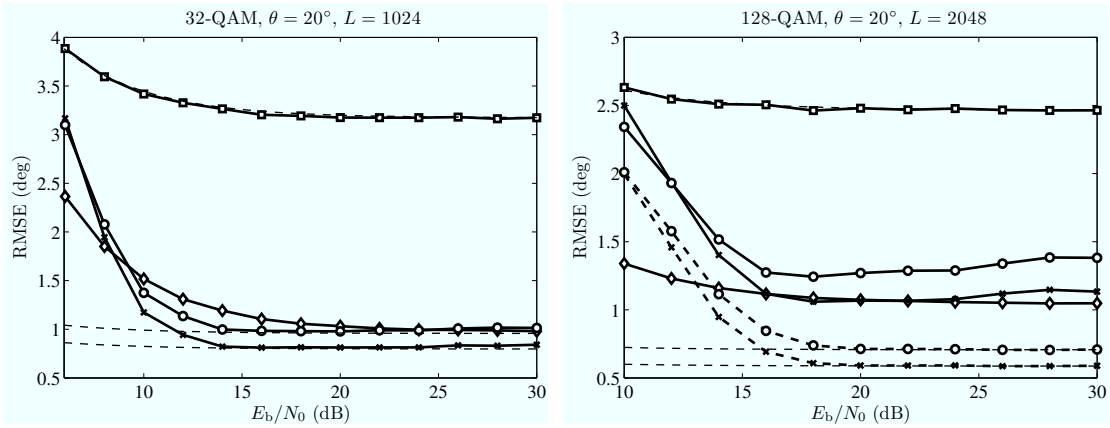
(a) (b)

Figure 5.6: RMSE versus number of iterations for (a) 32-QAM and (b) 128-QAM. The method in parentheses after J_1 and J_2 indicates which estimator was used as initializer for the respective iteration.

Note from Figure 5.6 that for 128-QAM and $E_b/N_0 = 30$ dB, the RMSE actually *increases* after the second iteration. This effect is due to the fact that, occasionally, the initializer ($\hat{\theta}_{4P}$ in this case) will fail to set the starting point within the domain of attraction of the desired solution, so that convergence to an undesired local maximum of the cost function J_q takes place. The nonzero probability of these events introduces a bias in the estimates, which in turn pulls up the final RMSE.

Figure 5.7 shows the RMSE of the proposed estimators (fixing the number of iterations to five), together with those of $\hat{\theta}_{4P}$ and $\hat{\theta}_{C8}$, as a function of the SNR per bit. For 32-QAM, the iterative estimators are initialized at $\hat{\theta}_{4P}$; they perform similarly to $\hat{\theta}_{C8}$ in the high E_b/N_0 region, although $\hat{\theta}_{C8}$ degrades more gracefully as the noise increases. Good agreement with the theoretical variance predicted by (5.24)-(5.25) is observed for $E_b/N_0 \geq 14$ dB.

Whereas $\hat{\theta}_{4P}$ seems to be a sufficiently good initializer with 32-QAM, the situation is different for 128-QAM: occasional convergence to local maxima prevents the RMSE of the iterative estimators from reaching their theoretical asymptotic values. Another option is to use $\hat{\theta}_{C8}$ as initializer. Note from Figure 5.7 that the theoretical RMSE of the iterative estimators with 128-QAM in high SNR lies below the observed RMSE of $\hat{\theta}_{C8}$, so that



(a) (b)

Figure 5.7: RMSE versus SNR per bit for (a) 32-QAM and (b) 128-QAM. Markers are as follows: 4P (□), C8 (◇), J_1 (×), J_2 (○). J_1 and J_2 are initialized with 4P (solid line) and C8 (dashed line). Dashed thin lines indicate the theoretical RMSE values for estimators $\hat{\theta}_{4P}$, $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$.

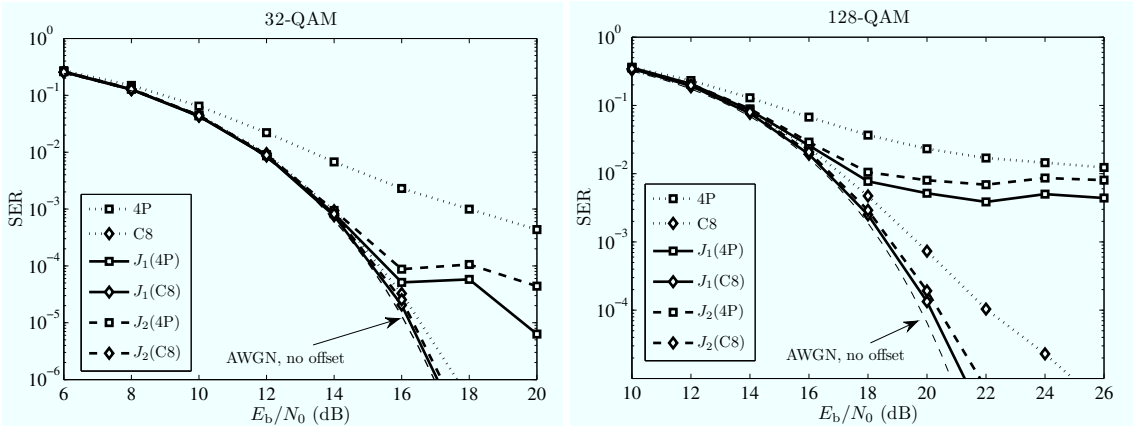
it makes sense to apply these schemes as refinements to $\hat{\theta}_{C8}$. As shown, this initializer avoids the problem of convergence to local maxima.

This same trend can be observed in Figure 5.8, which displays the SER achieved after application of the proposed estimators in comparison with the one-shot estimates (4P and C8) which serve as intializers. Considering the 32-QAM, the poor performance of 4P as initializer poses difficulties to the iterative methods, whose SER degrades above $E_b/N_0 = 16$ dB due to the convergence to local maxima. However, 4P seems a good initializer below $E_b/N_0 = 14$ dB (i.e., for SER $\leq 10^{-3}$). Initialization with a better estimate (C8) allows both J_1 and J_2 to stay close to reference SER with no phase offset. This is more clearly seen in the case of the 128-QAM, where the improvement of the iterative schemes over the initializers is evident. It is noticeable that J_1 and J_2 are able to significantly reduce the E_b/N_0 penalty of the best initializer at high SNRs (C8).

Of course, all these considerations depend on the number of samples L used for the estimation; too low a value of L could result in $\hat{\theta}_{C8}$ not being a good initializer either. But in general, it can be said that the novel estimators have the potential to improve the performance of either fourth-order or eighth-order schemes with low computational complexity. This potential is even greater under finite precision constraints, as discussed next.

5.6.2 Fixed Point Implementation

The performance of the different estimators is evaluated under quantization and fixed point arithmetic conditions. The goal is to test the robustness of each method under realistic implementation constraints. The receiver ADC quantizes each real-valued (in-phase and quadrature) data sample to B bits. The binary point is set to the right of the first bit,



(a) (b)

Figure 5.8: SER versus E_b/N_0 after phase correction for (a) 32-QAM, $L = 1024$ samples, and (b) 128-QAM, $L = 2048$ samples, both with $\theta = 20°$. The thin dashed line indicates the SER under no phase offset.

which is the sign bit. The full-scale value of the ADC (i.e., the maximum representable amplitude) is chosen in order to set the probability of clipping at 10^{-3} under AWGN, measured at a reference SNR for which a raw SER of 10^{-1} is obtained.[6]

When working with fixed point systems, care must be taken in order to minimize the effects of underflow and overflow. Appropriate scalings must be introduced at certain stages in order to obtain a suitable tradeoff between the probabilities of these two undesirable events, denoted by P_u and P_o respectively. The locations and values (usually powers of two, for efficiency reasons) of the scaling factors must be determined by careful study of P_u and P_o at each of the intermediate computations. Of the estimators considered in this chapter, it was found that scaling was beneficial for the C8 and 4P schemes when $B \leq 18$ and 9 bits respectively, but not for the proposed iterative methods. This confirms what one would expect by considering the number of products involved in each case.

In the simulations, wordlengths ranging from $B = 30$ down to 8 bits have been tested, and the scalings were optimized for each estimator. When B is sufficiently high, the convergence and the RMSE of all estimators follow the floating point results shown in the previous section. However, the performance of the methods tested degrades below a certain value of B, which depends on the estimator and the constellation. It was observed that in general $\hat{\theta}_{C8}$ is more sensitive to finite wordlength effects than the rest of estimators, whereas $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ are much more robust in this sense. Next we review in detail the results of our fixed point simulations in terms of bias and RMSE for the 32-QAM and 128-QAM constellations.

[6]The probability of clipping is the probability that the noisy received in-phase or quadrature components exceed the full-scale value. Since this probability depends on the noise variance, some reference noise variance must be considered. Such noise variance is provided by the SNR point for which the uncoded SER of the constellation equals 10^{-1}.

5.6.2.1 Fixed Point Bias

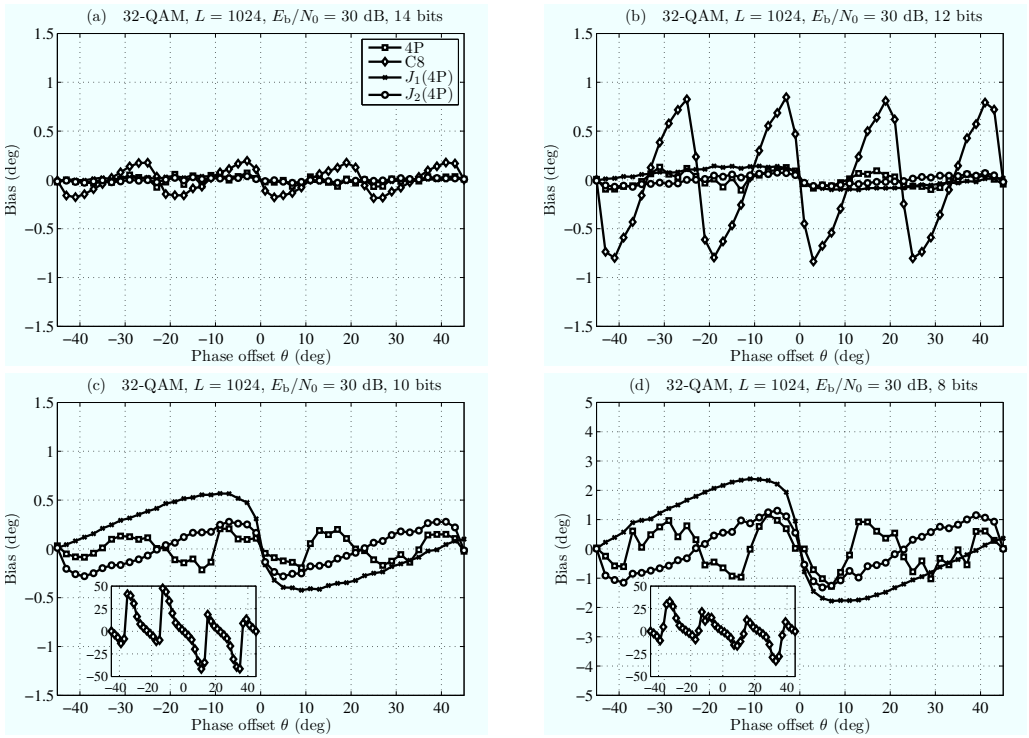


Figure 5.9: Bias versus θ. 32-QAM, $L = 1024$, $E_b/N_0 = 30$ dB. (a) $B = 14$, (b) $B = 12$, (c) $B = 10$, and (d) $B = 8$ bits. Markers are common to all figures as indicated in (a).

All estimators are unbiased for sufficiently large wordlengths, but as B is decreased an offset-dependent bias appears, whose magnitude grows inversely to B. Figures 5.9 and 5.10 show the bias for 32-QAM and 128-QAM respectively in terms of θ, at $E_b/N_0 = 30$ dB and for different wordlengths. At such high SNR values, the bias can only be attributable to finite precision effects. Clearly, $\hat{\theta}_{C8}$ is more severely affected than the rest of estimators. For 32-QAM, $\hat{\theta}_{C8}$ already shows bias at $B = 14$ bits, and for $B \leq 10$ bits it becomes useless. The remaining estimators degrade more gracefully, with $\hat{\theta}_{4P}$ and $\hat{\theta}^{(2)}$ exhibiting the smallest peak bias (about $0.3°$ and $1.0°$ for $B = 10$ and 8 bits respectively). For 128-QAM, $\hat{\theta}_{C8}$ is seen to have a large peak bias of $2°$ already for $B = 14$ bits. The bias of the other schemes becomes noticeable only for $B \leq 12$ bits.

5.6.2.2 Fixed Point RMSE

Results for 32-QAM are given in Figure 5.11(a)–(c) for $L = 1024$ symbols, $\theta = 20°$ and different values of B. With $B > 12$ bits, all methods perform similarly to the floating point case. For $B = 12$ bits, only $\hat{\theta}_{C8}$ shows a small degradation, which becomes very large for smaller wordlengths. In contrast, $\hat{\theta}_{4P}$ is not as sensitive, similarly to the novel estimators $\hat{\theta}^{(1)}$, $\hat{\theta}^{(2)}$, which achieve a much smaller asymptotic RMSE. We note that the small decrease in RMSE with respect to the finite precision curve is achieved in exchange for a slight bias.

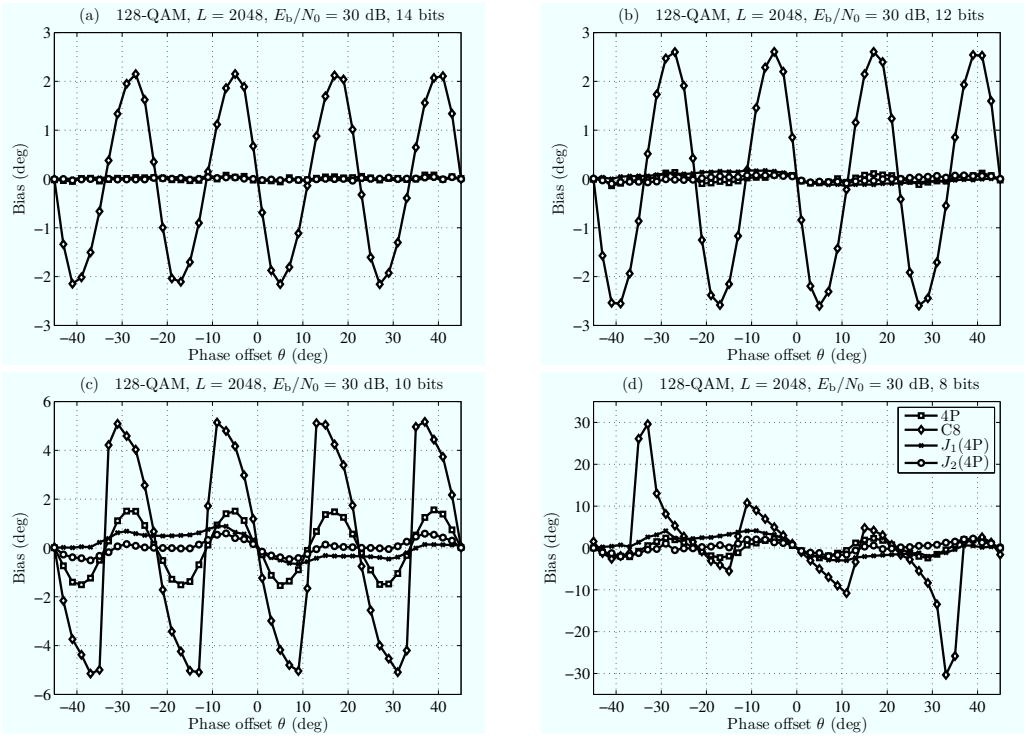


Figure 5.10: Bias versus θ. 128-QAM, $L = 2048$, $E_b/N_0 = 30$ dB. (a) $B = 14$, (b) $B = 12$, (c) $B = 10$, and (d) $B = 8$ bits. Markers are common to all figures as indicated in (d).

Results for 128-QAM, see Figure 5.11(d)–(f), roughly follow those for 32-QAM, but with all estimators suffering larger performance losses. For $B = 14$ and 12 bits, $\hat{\theta}_{\text{C8}}$ exhibits a large RMSE of about 2° (due to its large bias, see Figure 5.10). Nevertheless, it still provides a good initializer for $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$, which reach their asymptotic variances at high E_b/N_0. For $B = 10$ the high RMSE of $\hat{\theta}_{\text{C8}}$ (3.3°) pulls up the RMSEs of $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ when they start from $\hat{\theta}_{\text{C8}}$. The performance of $\hat{\theta}_{\text{4P}}$ is not affected by finite wordlengths as small as $B = 10$ bits, and thus when this estimate is used to initialize $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$, these achieve the same RMSE as in the floating point implementation (see Figure 5.7b).

To wrap up, the two proposed phase estimators present advantages over existing higher-order methods in terms of finite precision implementation. Whereas the higher-order metgods are clearly seen to suffer from finite precision issues, the proposed estimators are able to hold their good performance under tight wordlength constraints thanks to their lower statistical order. This makes them appealing choices for practical receiver design under e.g. fixed point implementation. Regarding their initialization, we have seen that standard one-shot methods can provide adequate coarse estimators for this purpose, upon which the proposed methods are able to provide refined estimates.

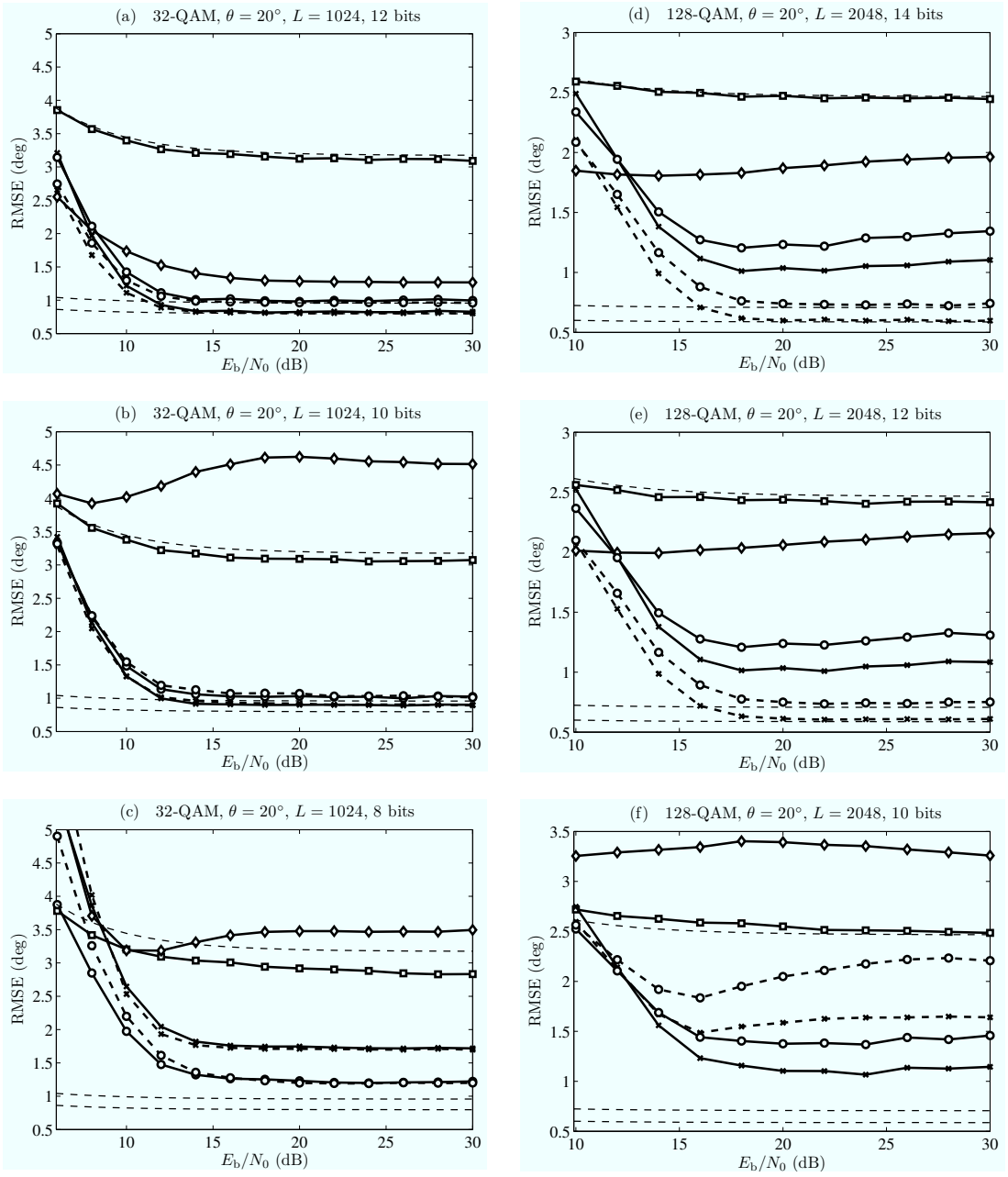


Figure 5.11: RMSE versus E_b/N_0 under fixed point arithmetic.
32-QAM, $L = 1024$, $E_b/N_0 = 30$ dB: (a) $B = 12$, (b) $B = 10$, and (c) $B = 8$ bits.
128-QAM, $L = 2048$, $E_b/N_0 = 30$ dB: (d) $B = 14$, (e) $B = 12$, and (f) $B = 10$ bits.
Markers are as follows: 4P (□), C8 (◊), J_1 (×), J_2 (○). J_1 and J_2 are initialized with 4P (solid line) and C8 (dashed line). Dashed thin lines indicate the theoretical RMSE values for estimators $\hat{\theta}_{4P}$, $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$.

Appendix 5.A Phase Estimation Based on ℓ_∞-norm

In $\mathbb{R}^2$, the ℓ_∞-norm $\|\boldsymbol{y}\|_\infty = \max\{|\bar{y}|, |\tilde{y}|\}$ turns out to be related to the ℓ_1-norm by the following property:

Property 6. *Let $\boldsymbol{x} = [\bar{x} \quad \tilde{x}]^{\mathrm{T}}$ and $\boldsymbol{z} = [\bar{z} \quad \tilde{z}]^{\mathrm{T}}$ be two vectors in $\mathbb{R}^2$, respectively corresponding to the complex numbers $x = \bar{x} + j\,\tilde{x}$ and $z = x\,e^{j\frac{\pi}{4}}$. For all $x \in \mathbb{C}$, it holds that*

$$\|\boldsymbol{x}\|_1 = \sqrt{2}\,\|\boldsymbol{z}\|_\infty. \tag{5.27}$$

This result holds as well for any rotation $z = x\,e^{j(m\frac{\pi}{2}+\frac{\pi}{4})}$ with $m \in \mathbb{Z}$.

For the proof, it suffices to realize that one can write the right hand side of (5.27) as $\sqrt{2}\,\|\boldsymbol{z}\|_\infty = \max\{|\bar{x} - \tilde{x}|, |\bar{x} + \tilde{x}|\}$, which turns out to be equal to $|\bar{x}| + |\tilde{x}| = \|\boldsymbol{x}\|_1$. This property states that the ℓ_1-norm of a $\mathbb{R}^2$ vector equals a scaled ℓ_∞-norm of the same vector rotated $\frac{\pi}{4}$ (or $\frac{\pi}{4}$ plus any integer multiple of $\frac{\pi}{2}$). It follows from this result that the ℓ_∞-norm-based cost function $K_q(\hat{\theta}) = \mathrm{E}\{\|\boldsymbol{y}\|_\infty^q\}$ is just a scaled and shifted (by $\frac{\pi}{4}$) version of the cost function $J_q(\hat{\theta})$ given in (5.7), i.e.:

$$K_q(\hat{\theta}) = \left(\frac{1}{\sqrt{2}}\right)^q J_q\left(\hat{\theta} - \frac{\pi}{4}\right). \tag{5.28}$$

Hence, both cost functions are equivalent for phase estimation purposes. This is illustrated by Figure 5.12, which represents the (normalized) costs J_q and K_q for 32-QAM and 128-QAM with $q = 1, 2$.

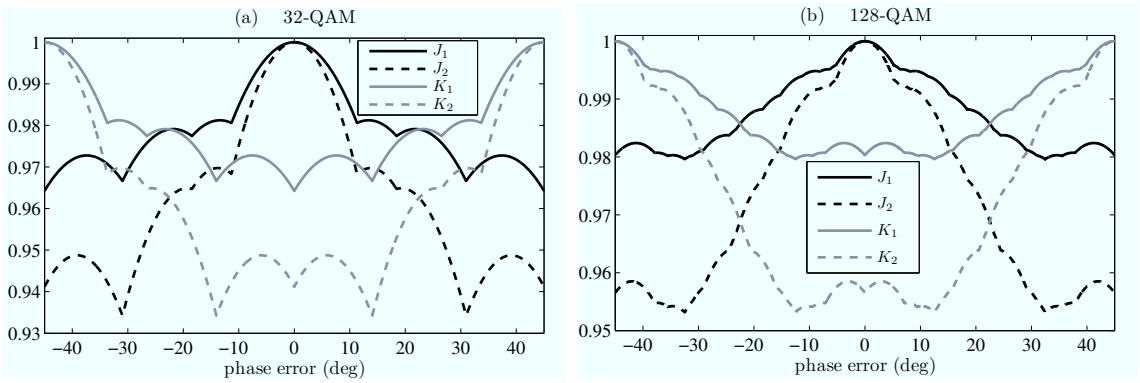


Figure 5.12: Behavior of the ℓ_1-norm based cost functions $J_q(\hat{\theta})$ and the ℓ_∞-norm based cost functions $K_q(\hat{\theta})$ (normalized to their maximum value) in the noiseless case, for $q = 1, 2$. (a) 32-QAM, (b) 128-QAM.

Since K_q presents absolute maxima at $\theta - \hat{\theta} = \pm\frac{\pi}{4}$, note that if K_q were used for phase estimation, the way to proceed should be to seek *the maxima* of the cost functions, and then subtract (or add) $\frac{\pi}{4}$ to the obtained values.

Appendix 5.B Proof of Theorem 1

We will prove (5.24), as the proof of (5.25) follows completely analogous steps. Let $\hat{\theta}^{(1)}$ be a fixed point of (5.15). In order to prove (5.24), let us define

$$\hat{\eta} \doteq \frac{1}{L} \sum_{k=0}^{L-1} r_k^* \operatorname{csgn}\left(r_k \, \mathrm{e}^{-j\hat{\theta}^{(1)}}\right). \tag{5.29}$$

Then $\hat{\theta}^{(1)}$ satisfies $\hat{\theta}^{(1)} = -\arg\{\hat{\eta}\}$. Let us also define

$$\eta \doteq \mathrm{E}\left\{r_k^* \operatorname{csgn}\left(r_k \, \mathrm{e}^{-j\theta}\right)\right\} = \bar{\eta} + j\tilde{\eta} \tag{5.30}$$

where θ is the true phase offset. Then, similarly to [142], we can write

$$\hat{\theta}^{(1)} = -\tan^{-1}\left(\frac{\operatorname{Im}\{\hat{\eta}\}}{\operatorname{Re}\{\hat{\eta}\}}\right) = -\tan^{-1}\left(\frac{\operatorname{Im}\{\eta\} - \operatorname{Im}\{\hat{\eta} - \eta\}}{\operatorname{Re}\{\eta\} - \operatorname{Re}\{\hat{\eta} - \eta\}}\right) \tag{5.31}$$

so that a first-order approximation of the argument in (5.31) yields

$$\hat{\theta}^{(1)} \approx -\tan^{-1}\left(\frac{\tilde{\eta}}{\bar{\eta}} + \epsilon\right) \tag{5.32}$$

$$\epsilon \doteq \frac{1}{\bar{\eta}}\left(\operatorname{Im}\{\hat{\eta} - \eta\} - \frac{\tilde{\eta}}{\bar{\eta}} \operatorname{Re}\{\hat{\eta} - \eta\}\right). \tag{5.33}$$

Note that $r_k \, \mathrm{e}^{-j\theta} = a_k + m_k$, where $m_k = n_k \, \mathrm{e}^{-j\theta}$ is a noise process with the same statistical properties as n_k. Hence, if the SNR is sufficiently high, then

$$\operatorname{csgn}\left(r_k \, \mathrm{e}^{-j\theta}\right) = \operatorname{csgn}\left(a_k + m_k\right) \approx \operatorname{csgn}\left(a_k\right)$$

so that

$$\eta \approx \mathrm{E}\left\{r_k^* \operatorname{csgn}\left(a_k\right)\right\} \tag{5.34}$$

$$= \underbrace{\mathrm{E}\left\{a_k^* \operatorname{csgn}\left(a_k\right)\right\}}_{=\sqrt{E_a}\, d_1} \mathrm{e}^{-j\theta} + \underbrace{\mathrm{E}\left\{n_k^* \operatorname{csgn}\left(a_k\right)\right\}}_{=0} \tag{5.35}$$

and the approximation error in (5.34) decays exponentially fast with the SNR. The statistic d_1 (defined in Theorem 1) turns out to be real and positive for QAM constellations, and therefore $-\tan^{-1}(\tilde{\eta}/\bar{\eta}) \to \theta$ as the SNR goes to infinity. Then a first-order expansion of (5.32) yields

$$\hat{\theta}^{(1)} \approx \theta - \epsilon \cdot \cos^2\theta. \tag{5.36}$$

With the asymptotic approximation $\eta \approx \sqrt{E_a}\, d_1 \, \mathrm{e}^{-j\theta}$, (5.33) becomes

$$\epsilon \approx \operatorname{Im}\left\{(\hat{\eta} - \eta) \, \mathrm{e}^{j\theta}\right\} / (\bar{\eta} \, \cos\theta). \tag{5.37}$$

Note that if the SNR is sufficiently large and the phase error $\theta - \hat{\theta}^{(1)}$ is sufficiently small, then $\hat{\eta}$ is an unbiased estimate of η:

$$\begin{aligned}
\mathrm{E}\left\{\hat{\eta}\right\} &= \frac{1}{L}\sum_{k=0}^{L-1} \mathrm{E}\left\{r_k^* \,\mathrm{csgn}\left(r_k\, \mathrm{e}^{-j\hat{\theta}^{(1)}}\right)\right\} \\
&= \frac{1}{L}\sum_{k=0}^{L-1} \mathrm{E}\left\{r_k^* \,\mathrm{csgn}\left(a_k\, \mathrm{e}^{j(\theta-\hat{\theta}^{(1)})} + n_k\, \mathrm{e}^{-j\hat{\theta}^{(1)}}\right)\right\} \\
&\approx \frac{1}{L}\sum_{k=0}^{L-1} \mathrm{E}\left\{r_k^* \,\mathrm{csgn}\left(a_k\right)\right\} \;\approx\; \eta. \qquad (5.38)
\end{aligned}$$

Hence, from (5.36)-(5.38), $\hat{\theta}^{(1)}$ is asymptotically unbiased.

In order to compute the asymptotic variance, some straightforward manipulations show that

$$\mathrm{Im}^2\left\{(\hat{\eta}-\eta)\,\mathrm{e}^{j\theta}\right\} = \frac{1}{2}\left[|\hat{\eta}-\eta|^2 - \mathrm{Re}\left\{(\hat{\eta}-\eta)^2\,\mathrm{e}^{j2\theta}\right\}\right]. \qquad (5.39)$$

Therefore, from (5.36)-(5.37) and (5.39),

$$\begin{aligned}
\lim_{L\to\infty} L\,\mathrm{E}\left\{(\hat{\theta}^{(1)}-\theta)^2\right\} &= \cos^4\theta \lim_{L\to\infty} L\,\mathrm{E}\left\{\epsilon^2\right\} \qquad (5.40) \\
&= \frac{\cos^2\theta}{2\bar{\eta}^2} \lim_{L\to\infty} L\left[\mathrm{E}\left\{|\hat{\eta}-\eta|^2\right\} - \mathrm{Re}\left\{\mathrm{E}\left\{(\hat{\eta}-\eta)^2\right\}\mathrm{e}^{j2\theta}\right\}\right] \qquad (5.41)
\end{aligned}$$

where the expectations in (5.41) are given by

$$\begin{aligned}
\mathrm{E}\left\{|\hat{\eta}-\eta|^2\right\} &= \mathrm{E}\left\{|\hat{\eta}|^2 - |\eta|^2\right\} \\
&\approx \frac{1}{L^2}\sum_{k=0}^{L-1}\sum_{l=0}^{L-1} \mathrm{E}\left\{r_k^*\, r_l \,\mathrm{csgn}\left(r_k\,\mathrm{e}^{-j\theta}\right)\,\mathrm{csgn}^*\left(r_l\,\mathrm{e}^{-j\theta}\right)\right\} - |\eta|^2 \qquad (5.42) \\
&= \frac{(L^2-L)}{L^2}\left|\mathrm{E}\left\{r_k^* \,\mathrm{csgn}\left(r_k\,\mathrm{e}^{-j\theta}\right)\right\}\right|^2 \\
&\quad + \frac{1}{L^2}\sum_{k=0}^{L-1} \mathrm{E}\left\{|r_k|^2\,\left|\mathrm{csgn}\left(r_k\,\mathrm{e}^{-j\theta}\right)\right|^2\right\} - |\eta|^2 \qquad (5.43) \\
&= \left(1-\frac{1}{L}\right)|\eta|^2 + \frac{2}{L}\,\mathrm{E}\left\{|r_k|^2\right\} - |\eta|^2 \qquad (5.44) \\
&= \frac{1}{L}\left(2E_a + 2\sigma^2 - |\eta|^2\right) \qquad (5.45)
\end{aligned}$$

$$
\begin{aligned}
\mathrm{E}\left\{(\hat{\eta}-\eta)^2\right\} &= \mathrm{E}\left\{\hat{\eta}^2-\eta^2\right\} \\
&\approx \frac{1}{L^2}\sum_{k=0}^{L-1}\sum_{l=0}^{L-1}\mathrm{E}\left\{r_k^*\, r_l^*\,\mathrm{csgn}\left(r_k\,\mathrm{e}^{-j\theta}\right)\,\mathrm{csgn}\left(r_l\,\mathrm{e}^{-j\theta}\right)\right\}-\eta^2 \qquad (5.46)\\
&= \frac{(L^2-L)}{L^2}\mathrm{E}^2\left\{r_k^*\,\mathrm{csgn}\left(r_k\,\mathrm{e}^{-j\theta}\right)\right\} \\
&\quad + \frac{1}{L^2}\sum_{k=0}^{L-1}\mathrm{E}\left\{\left(r_k^*\,\mathrm{csgn}\left(r_k\,\mathrm{e}^{-j\theta}\right)\right)^2\right\}-\eta^2 \qquad (5.47)\\
&\approx \left(1-\frac{1}{L}\right)\eta^2+\frac{1}{L}\mathrm{E}\left\{\left(r_k^*\,\mathrm{csgn}\left(a_k\right)\right)^2\right\}-\eta^2 \qquad (5.48)\\
&= \frac{1}{L}\left(E_a\, d_2\,\mathrm{e}^{-j2\theta}-\eta^2\right). \qquad (5.49)
\end{aligned}
$$

Note that d_2 is defined in Theorem 1. Now, substituting (5.45) and (5.49) in (5.41), and noting that $\bar{\eta}^2 = E_a\, d_1^2\,\cos^2\theta$, the final result (5.24) is obtained.

Chapter 6

Conclusions and Future Lines

Contents

6.1 Conclusions

We proposed the joint use of shaping without scrambling (i.e., the combination of Tomlinson-Harashima Precoding (THP) and signal shaping) and signal predistortion in the context of a downlink point-to-point satellite service making use of gap fillers to complete the satellite coverage. The signals coming from the gap fillers introduce diversity in the received signal, which, unless properly processed, translates into undesired Intersymbol Interference (ISI). We saw that the combination of shaping without scrambling and signal predistortion successfully counteracts the gap filler-induced ISI and the nonlinear distortion caused by the High Power Amplifier onboard the satellite. The role of shaping without scrambling is to modify the shape of the precoded signal so that the HPA-induced distortion is minimum. With this goal, different shaping metrics have been tested with similar net results in terms of BER. All in all, the proposed scheme may constitute an attractive solution for practical systems with fixed and nomadic terminals, since the ISI is precanceled at the transmitter (e.g., the gateway), which should simplify the complexity of the terminals.

THP was also considered for the downlink of a satellite multibeam satellite system providing independent data streams to each beam. The THP operation, which is to be preferably performed at the gateway, was used to cancel the Co-Channel Interference (CCI) that each terminal receives from the neighboring beams reusing the same frequency. Several application scenarios were tested, based on a realistic beam pattern, and for several terminal locations and frequency reuse factors. We saw that the adoption of THP in this context can achieve the proper removal of the unwanted CCI. This, in turn, may

reduce the frequency reuse factor required in current multibeam systems to prevent an excess of CCI, especially when a single centralized gateway performs the precoding for a large number of beams. The application of Linear Precoding (LP) to similar scenarios was analyzed in detail in [43, 67] offering interesting potential improvements in terms of throughput with respect to traditional systems. Our work shows that even greater improvements can be expected if THP is used, mainly due to the removal of the transmit power penalty inherent to LP schemes. This fact has attracted the interest of recent surveys devoted to identify novel transmission technologies for future satellite systems, e.g. [15].

In what regards the problem of Signal-to-Noise Ratio (SNR) estimation, we saw that the use of higher-order moments is an interesting approach for estimator design, particularly considering the SNR estimation from higher modulations such as the two- and three-level APSK constellations used in the DVB-S2 standard. We developed an SNR estimator based on eighth-order statistics, which can be efficiently implemented without look-up tables. The estimator is built upon a linear combination of ratios of moments, whose weights can be tuned according to the constellation and the SNR operation range. Two weight optimization criteria were proposed, which yield good estimators with improved performance over existing methods. The new higher-order approach, which generalizes many of the previous moments-based estimators existing in the literature, can be easily extended to any statistical order and to any modulation order. The analysis and optimization of such extensions can be carried out using the same approach and tools that were applied to the eighth-order estimator.

We also studied the problem of estimating the SNR from heterogeneous frames, i.e., frames containing two (or more) subframes, each one formed by symbols drawn from a different constellation. Such kind of frames are commonly found in current systems in which the data-bearing symbols can belong to higher modulations, whereas simpler modulations are chosen for the header and the pilot symbols. The DVB-S2 standard provides a practical case of application of our approach. By exploiting the knowledge of the symbol set we obtained the CRB for the problem and we proposed a number of possible heterogeneous estimators of which the so called convex combination of estimates (CCE) proved to be the best choice. Analysis of the MSEs of the proposed estimators showed that the CCE estimator outperforms the constituent estimators, and also the rest of the heterogeneous estimators considered. Furthermore, CCE estimators are rather flexible since they can combine any type of constituent estimators, including data-aided ones. We have seen that this last feature is a further advantage of the CCE scheme when training symbols (pilots, partially known header) are available in the heterogeneous frame.

As for the problem of phase estimation, we introduced two iterative low-order phase estimators for cross-QAM constellations (e.g. 32-QAM, 128-QAM) whose performance achieves that of higher-oder schemes specifically derived for such constellations. The proposed estimators do not require neither gain control nor data knowledge, and feature a low statistical order, which turns into an advantage when they are to be implemented under finite precision conditions. Whereas higher-order methods are clearly seen to suffer from finite precision issues, the proposed estimators are able to hold their good performance under tight wordlength constraints, which makes them appealing choices for practical receiver design under e.g. fixed point implementation. A key issue in the proposed it-

erative estimators is their initialization. We saw that standard one-shot methods (eighth- and fourth-order) can provide adequate coarse estimators for this purpose, upon which the proposed methods are able to provide refined estimates. Different initialization strategies are also possible: for example, the low complexity of the Viterbi&Viterbi estimators of 0-th and negative orders makes them attractive initializers at low and high SNRs, respectively. Yet another possibility arises whenever training symbols are inserted in the transmitted frame, as it is usually the case in practice. A data-aided phase estimator based just on these training symbols may well present a variance too large for decoding purposes (due to a limited number of training data), but sufficiently low to initialize the proposed iterative schemes. Such *semiblind* approach would significantly improve the quality of the plain data-aided phase estimate.

6.2 Future Research Lines

The performance analysis of THP developed in Chapters 2 and 3 has been possible thanks to the adoption of certain simplifying assumptions. The list of assumptions for each chapter can be found in pages 23 and 63. Among these assumptions, the most demanding ones are probably the following: perfect timing and carrier recovery, perfect CSI knowledge at the transmitter, and linear HPAs. The two former assumptions are common to the proposed SISO and multibeam systems, whereas the latter was used in the multibeam case only. In Appendices 2.B, 3.E, and 3.D we have provided an overview of these issues, as well as some of the approaches available in the open literature. A completely realistic system analysis would require to evaluate the impact in performance of practical timing and carrier recovery schemes, practical CSI retrieval schemes, and practical HPA linearization schemes (e.g. predistortion). Furthermore, no coding was considered in the THP systems analyzed in Chapters 2 and 3. The introduction of coding (e.g. the LDPC codes considered in 10GBASE-T) would allow a complete system analysis in terms of throughput.

Under a more fundamental approach, it would be interesting to consider the use of hexagonal Voronoi regions in the THP operation. In comparison with the squares provided by the standard modulo operation, the hexagons automatically provide certain shaping gain at the cost of a more complex precoding (and decoding) operation. It would be also interesting to consider the application of THP to staggered modulation formats such as offset-QPSK or offset-QAM. Such modulations produce waveforms with reduced envelope fluctuations, which would complement the effect of predistortion as a means to mitigate the impact of the nonlinear HPA. To conclude the THP-related future research lines, now focusing on the receiver operation, the problem of timing recovery from THP signals does not yet seem to have received enough interest within the research community. In this sense, it seems possible to improve the performance of current NDA timing recovery schemes when applied to THP signals. This line could be complemented with the computation of the corresponding CRB for the problem.

Regarding the SNR estimation problem, two topics appear as the natural continuation of the novel moments-based approach. First, the eighth-order estimator proposed in Sec-

tion 4.3 could be extended to higher orders (e.g. extending the linear combination with tenth-order statistics), which would presumably bring further performance improvements thanks to the use of more ratios of moments. Second, the generalized approach to SNR estimation based on moments of any order, as presented in Appendix 4.B, could be used to develop novel SNR estimators of low order (e.g., third or fourth) with presumably good performance for multilevel constellations at intermediate and high SNRs. In addition to these two lines, another interesting topic of further research is the problem of estimating the SNR from oversampled data. In this context, several estimators have been recently proposed [70, 136, 143], although there seems to be room for new estimators as well as for finding the CRB for the problem.

As for the proposed phase estimation schemes, the potential local convergence issues caused by the multimodality of the proposed ℓ_1-norm-based cost functions makes it desirable to find other cost functions suitable for phase estimation avoiding such issues. From a more general perspective, an interesting problem is the design of joint schemes for SNR estimation and carrier recovery. Although it has been recently shown that the CRBs for SNR estimation and carrier recovery (phase and frequency) are decoupled [73], this does not mean that a suboptimal estimation scheme cannot benefit from a joint approach. In fact, it turns out that optimal (CRB-achieving) NDA estimation schemes are computationally intensive, so the seek for suboptimal but simpler estimation schemes would be justified.

Capítulo 7

Resumen en castellano

Contents

Desde el inicio de las telecomunicaciones, la evolución hacia sistemas de transmisión de mayor capacidad y eficiencia ha sido una constante. Generalmente, la aparición de una nueva tecnología de transmisión o de una técnica de procesado de señal suscita la creación de nuevas aplicaciones en las que estas técnicas resultan útiles y eficientes. Tras una fase de investigación y optimización, dichas tecnologías, ya maduras, acaban aportando una mayor capacidad de transmisión a los sistemas reales en los que se emplean. Sin embargo, la demanda eficiencia y capacidad de transmisión es siempre creciente, y tan pronto las nuevas técnicas son asimiladas y explotadas no tardan en aparecer otras nuevas que mejoran a sus predecesoras. Como ejemplo, obsérvese la evolución de las comunicaciones por satélite, en donde en poco más de 50 años se ha pasado de la puesta en órbita del primer satélite artificial (Sputnik-I, 1957) y el primer satélite dedicado a comunicaciones (TelStar 1, 1962) a planear la puesta en órbita de satélites con una tasa de transferencia total de 100 Gbps [88].

Esta tesis se centra en gran parte en diversos aspectos relacionados con la capa física de la transmisión vía satélite. Por una parte, los capítulos 2 y 3 se centran en aspectos relacionados con el transmisor. En particular, se propone la aplicación de la técnica de transmisión conocida como *precodificación Tomlinson-Harashima* (THP, del inglés *Tomlinson-Harashima precoding*) a la transmisión por satélite. Así, en el Capítulo 2 se explora la aplicación de THP en sistemas de transmisión punto a punto, mientras que el Capítulo 3 se ocupa de la aplicación a sistemas con cobertura multihaz. Por su parte, los capítulos 4 y 5 se centran en la mejora de técnicas de procesado realizadas en el receptor, en particular, en métodos estimación de parámetros de señal. En particular, el Capítulo 4 trata de la estimación del la *relación señal a ruido* (SNR, del inglés *signal-to-noise ratio*)

en el receptor, y tiene una vocación de aplicación general, incluyendo las modulaciones empleadas actualmente en transmisión por satélite. El Capítulo 5 trata de la estimación de fase en el receptor para modulaciones QAM en cruz, que son modulaciones en amplitud y cuadratura (QAM, del inglés *quadrature-amplitude modulation*) utilizadas exclusivamente en sistemas de comunicaciones terrestres. A continuación se resumirá el contenido de los cuatro capítulos principales de esta tesis.

7.1 Aplicación de THP a la transmisión punto a punto por satélite

En los últimos años ha surgido un gran número de servicios de comunicación que emplean un satélite geoestacionario como un medio eficaz para dar cobertura a grandes regiones. Sin embargo, esa cobertura puede no ser todo lo completa que sería deseable debido a que la señal proveniente del satélite no alcanza a todos los posibles receptores. Esto sucede especialmente en zonas urbanas densamente pobladas, en donde las edificaciones, especialmente las de gran altura, representan obstáculos insalvables que bloquean completamente la propagación de la señal del satélite. Ante esta situación, una de las soluciones es el empleo de *repetidores modulares intermedios* (IMRs, del inglés *intermediate module repeaters*, también conocidos como *gap-fillers*). Los IMRs actúan como repetidores de la señal del satélite, y dan cobertura a las zonas a las que no llega la señal del satélite. Ejemplos del uso de IMRs se pueden encontrar en [12, 48, 55, 84, 105, 109, 130]. El empleo de IMRs es positivo en el sentido de que introduce *diversidad* en la señal recibida por los terminales. Sin embargo, esta diversidad viene asociada a una dispersión temporal en la señal recibida, que se traduce en la existencia de *interferencia entre símbolos* (ISI, del inglés, *intersymbol interference*). Para una correcta recepción de la señal deseada, el receptor deberá aplicar algún tipo de medida de compensación adecuada, como por ejemplo combinación de diversidad, o igualación de canal, como por ejemplo la igualación mediante realimentación de decisiones (DFE, del inglés *decision feedback equalization*).

En el Capítulo 2 de esta tesis se propone el empleo de THP para la compensación de la ISI presente en sistemas de transmisión punto a punto por satélite que emplean IMRs en el enlace descendente para completar la cobertura. En [59] se puede encontrar una introducción muy completa a la transmisión THP. La figura 7.1 se muestra el modelo básico en tiempo discreto de la transmisión THP, en donde $H(z)$ es la función de transferencia discreta equivalente del canal, y $\mathrm{MOD}_{\Delta}(\cdot)$ representa una operación módulo, por tanto no lineal. Los símbolos de entrada, a_n, que supondremos pertenecientes a constelaciones QAM cuadradas, son precodificados mediante el lazo indicado en la figura, obteniéndose la secuencia precodificada $x_n = \mathrm{MOD}_{\Delta}\left(a_n - \sum_{k=1}^{p-1} h_k\, x_{n-k}\right)$. Esta operación puede verse como una suerte de igualación de canal que se realiza en el transmisor, en vez de en el receptor. Muy similar a la preigualación lineal de canal, THP introduce una operación no lineal en la preigualación ($\mathrm{MOD}_{\Delta}(\cdot)$) que permite evitar el principal inconveniente de la preigualación lineal: la penalización de potencia transmitida. Además, el hecho de que la operación se realice en transmisión evita problemas como la propagación de errores que ocurren en DFE para bajas SNRs. Para una precorreción adecuada de la ISI, el transmisor

THP necesita tener una estimación precisa del canal discreto equivalente hacia el receptor, que le será proporcionada mediante algún esquema de calibración y estimación de canal.

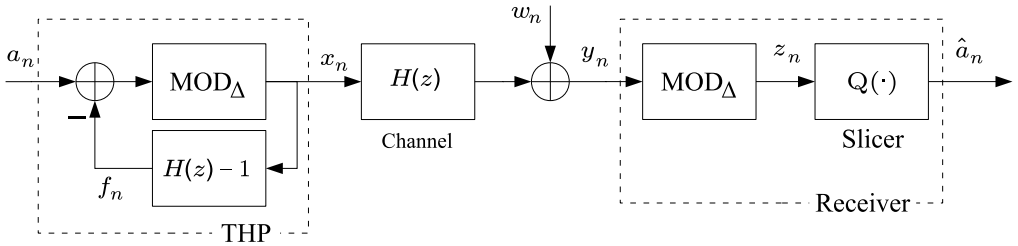


Figura 7.1: Esquema básico de transmisión THP

Hasta ahora, la aplicación más relevante de THP se encuentra en la tecnología 10GBASE-T de transmisión por cable, según se especifica en el estándar [90], pero nunca se ha empleado en transmisión por satélite. Una de las razones que dificultan la aplicación de THP a las comunicaciones por satélite es la presencia del amplificador de alta potencia (HPA, del inglés *high power amplifier*) a bordo del satélite. Estos amplificadores deben compensar las importantes pérdidas por propagación debidas a la gran distancia existente entre satélite y terminal terrestre, y por razones de coste y eficiencia de consumo eléctrico suelen trabajan en puntos de amplificación no muy lejanos a la saturación. Esto conlleva que la señal transmitida por el satélite sufra cierto grado de distorsión no lineal, que de no ser corregida, mina las prestaciones del sistema de transmisión.

El problema de la distorsión no lineal es, en mayor o menor grado, común a todos los esquemas de transmisión por satélite. En la práctica, sus efectos se pueden reducir mediante el empleo de técnicas de compensación de la distorsión como el alejamiento del punto de la saturación (técnica conocida comúnmente como IBO, del inglés *input back-off*), o la *predistorsión*, que trata de corregir la señal en algún punto previo a la entrada del amplificador aplicando una aproximación de la característica inversa del amplificador. En la práctica, siempre se utiliza cierto grado de IBO. Además, el impacto del amplificador se reduce con el empleo de modulaciones con cierta tolerancia a la distorsión no lineal, como son las modulaciones en desplazamiento de fase (PSK, del inglés *phase shift keying*) o las modulaciones en amplitud y desplazamiento de fase (APSK, del inglés *amplitude and phase shift keying*), las cuales tienen la virtud de generan formas de onda con una baja amplitud de pico. Todo lo contrario sucede con la señal precodificada según THP, x_n, cuya distribución uniforme en el cuadrado delimitado por la operación $\mathrm{MOD}_\Delta(\cdot)$ genera una señal con una amplitud de pico mucho mayor. Esto se puede observar en la Figura 7.2, que muestra las funciones de densidad acumulada complementarias obtenidas mediante simulación para varias modulaciones además de para THP.

Un valor de pico grande se traduce en una probabilidad mayor de que el HPA entre en saturación, y por tanto, provoque mayor distorsión no lineal. Por esta razón, en el Capítulo 2 se ha propuesto el uso conjunto de THP y *predistorsión* () de la señal a la entrada del HPA. Esta operación de predistorsión opera, por tanto, en el dominio de las formas de onda. En general, el diseño de la etapa de predistorsión se basa en un modelo previo que se ajusta al comportamiento no lineal del HPA. A partir de este modelo, la etapa

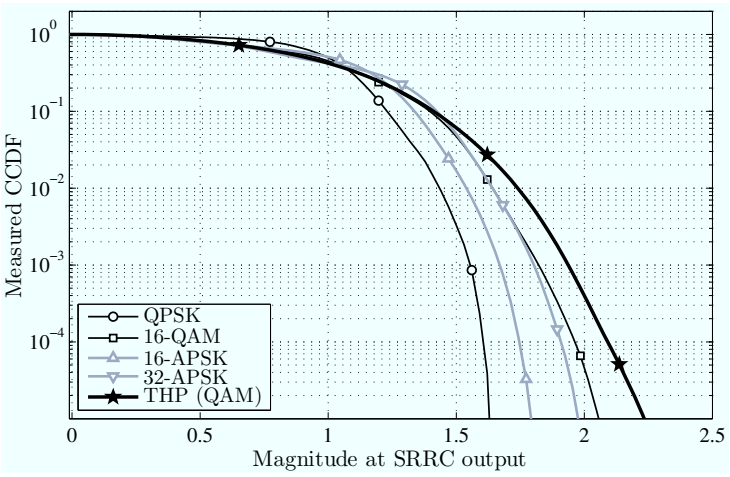


Figura 7.2: Densidades acumuladas complementarias de la magnitud de la forma de onda a la salida del filtro de transmisión para diferentes modulaciones. Todas las formas de onda están normalizadas en potencia.

de predistorsión trata de implementar una función inversa a la característica del HPA, de forma que la cascada predistorsión–HPA ofrezca la mínima distorsión residual posible. En general, una inversión perfecta de los efectos del HPA no suele ser posible. En el Capítulo 2 se han considerado HPAs de la familia TWTA (del inglés *traveling wave tube amplifier*), cuya característica, que se muestra en la Figura 7.3 se ha aproximado mediante un modelo de Saleh generalizado diez parámetros. Éste es un modelo sin memoria en el que los efectos del amplificador dependen únicamente de la envolvente instantánea de la señal de entrada. Así, si $x(t) = r(t)\,\cos(2\pi f_0 t + \phi(t))$ es la señal de radiofrecuencia (RF) a la entrada del amplificador, con $r(t)$ la envolvente instantánea, la salida viene dada por $y(t) = A\,[r(t)]\cos\left(2\pi f_0 t + \phi(t) + \Phi(r(t))\right)$, en donde las características AM/AM ($A[r]$) y AM/PM ($\Phi[r]$) se modelan mediante

$$A[r] = \frac{\alpha_{\mathrm{a},2}\, r^2 + \alpha_{\mathrm{a},1}\, r + \alpha_{\mathrm{a},0}}{\beta_{\mathrm{a},2}\, r^2 + \beta_{\mathrm{a},1}\, r + 1} \quad \text{(unidades lineales)} \tag{7.1}$$

$$\Phi[r] = \frac{\alpha_{\mathrm{p},2}\, r^2 + \alpha_{\mathrm{p},1}\, r + \alpha_{\mathrm{p},0}}{\beta_{\mathrm{p},2}\, r^2 + \beta_{\mathrm{p},1}\, r + 1} \quad \text{(rad)}. \tag{7.2}$$

Como implementación práctica de la etapa de predistorsión se ha considerado una tabla de búsqueda (LUT, del inglés *look-up table*) de 1024 entradas en las que los valores de la función de predistorsión se pueden almacenar con antelación. De esta manera sustituye el cálculo analítico de los valores de predistorsión para cada instante por operaciones de acceso a memoria, con el consiguiente ahorro de carga computacional. La tabla está indexada por el valor instantáneo de la envolvente de la entrada, como se indica en la Figura 7.4.

El esquema completo del sistema con THP y predistorsión considerado se muestra en la Figura 7.5. Para evaluar las prestaciones de este sistema, como sistema de referencia se ha considerado un sistema igual al de la Figura 7.5 en donde el bloque THP del transmisor se sustituye por un DFE en el receptor. Los resultados en forma de tasa de error de símbolo

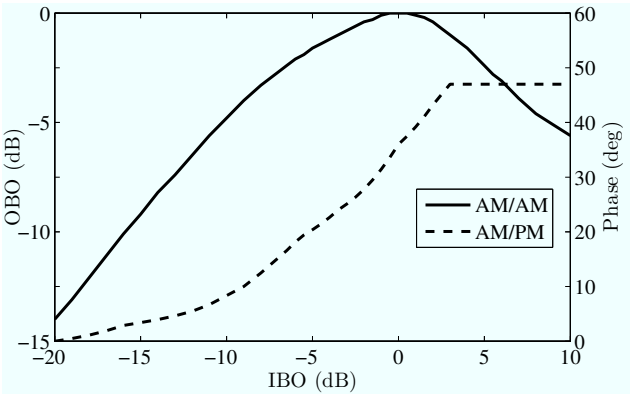


Figura 7.3: Características AM/AM y AM/PM consideradas para el TWTA a bordo del satélite.

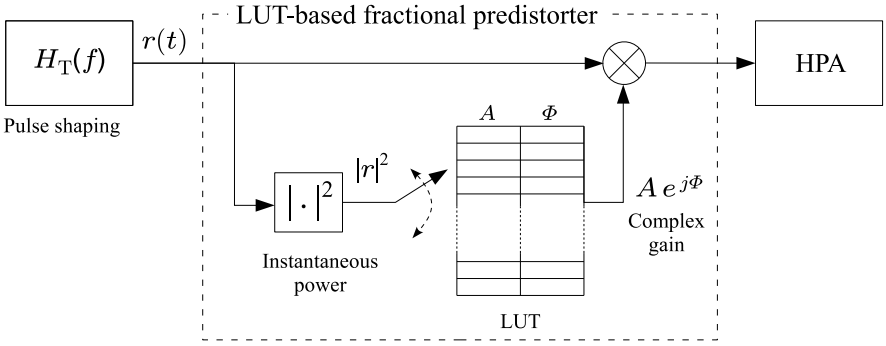


Figura 7.4: Esquema de la etapa de predistorsión considerada implementada mediante una tabla de búsqueda indexada por el valor de potencia instantánea de la señal de entrada..

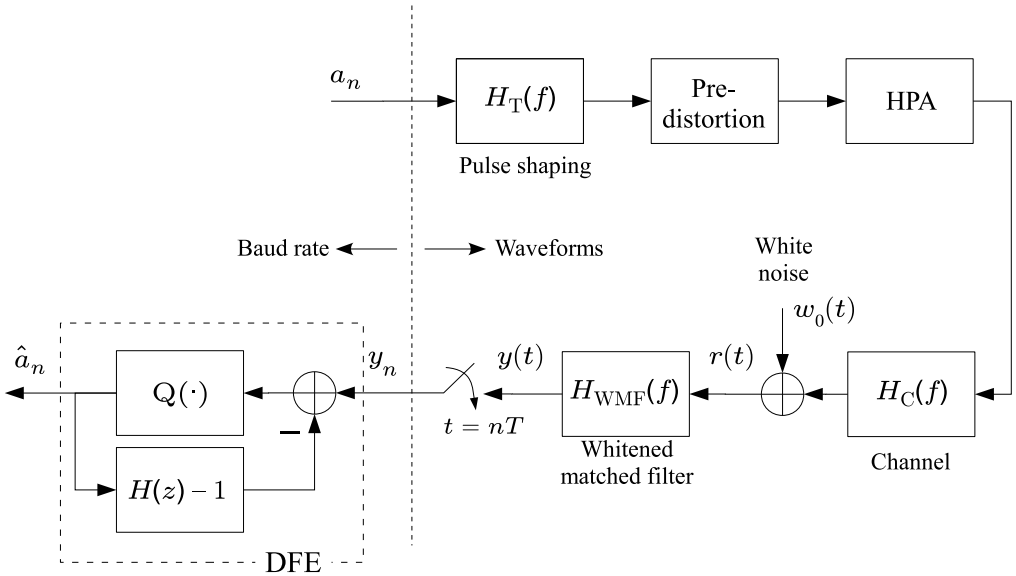
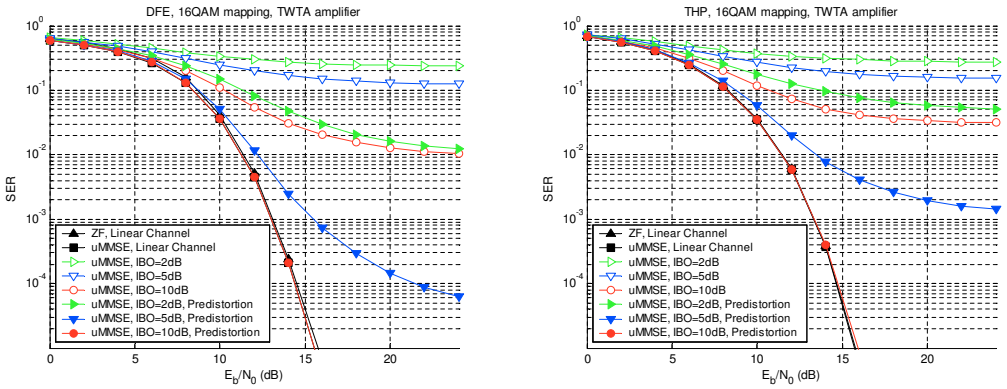


Figura 7.5: Esquema del sistema propuesto que combina THP y predistorsión.

(SER, del inglés *symbol error rate*) de ambos sistemas se muestran en la Figura 7.6, para diferentes valores de IBO. Como se puede observar, el empleo de predistorsión es



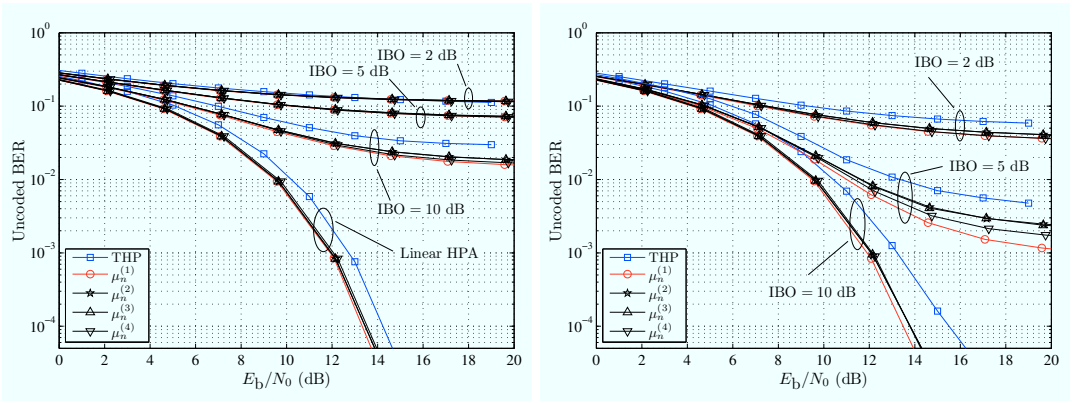
(a) (b)

Figura 7.6: Prestaciones en términos de SER no codificada para (a) predistorsión combinada con DFE, y (b) THP combinada con predistortion.

claramente positivo en los dos sistemas. Por otra parte, el sistema con DFE se ve menos afectado por la distorsión no lineal residual que el sistema con THP. Esto se explica por la distribución de densidad de la señal precodificada, que como se muestra en la Figura 7.2 genera una forma de onda con un valor de pico de amplitude muy alto. Sin embargo, la modulación QAM (que es la empleada en el sistema con DFE) ofrece un valor de pico menor que THP, y por ello la SER obtenida para un mismo IBO es menor para DFE que para THP.

Este resultado sugiere que para mejorar las prestaciones del sistema con THP se debe reducir la amplitud de pico de la señal precodificada. Esto se puede conseguir con el uso de técnicas de "moldeado de señal", comúnmente conocidas como *signal shaping*, que consiguen modificar la distribución de densidad de los símbolos transmitidos de cara a mejorar alguna característica de la señal [59]. Tradicionalmente, el objetivo del *signal shaping* ha sido la minimización de la energía media de la secuencia transmitida, aunque en general es una técnica flexible que permite alcanzar otros objetivos. La combinación de *signal shaping* con THP no es sencilla: no basta con concatenar ambas operaciones, sino que ambas tareas se deben realizar en una única operación. Esto se puede conseguir mediante dos técnicas distintas: *trellis shaping* y *shaping sin barajado* (en inglés, *shaping without scrambling*) [59]. La técnica escogida en el Capítulo 2 ha sido el *shaping* sin barajado debido a que no obliga a realizar modificaciones en el receptor THP, mientras que *trellis shaping* sí requiere de un receptor específico. En *shaping* sin barajado se emplea un decodificador de Viterbi para escoger la secuencia precodificada que minimice una métrica particular. La métrica se diseña de forma que la señal precodificada resultante presente ciertas características deseadas (como por ejemplo la mínima energía media transmitida).

En el segundo sistema propuesto en el Capítulo 2 el transmisor implementa *shaping* sin barajado y la misma etapa de predistorsión considerada anteriormente. Los resultados obtenidos en términos de tasa de error de bit (BER, del inglés *bit error rate*) se recogen en



(a) (b)

Figura 7.7: Prestaciones en términos de BER no codificada del *shaping* sin barajado para los casos (a) sin predistorsión, y (b) con predistorsión.

la Figura 7.7. Aparte de métrica que minimiza la energía media de las señal precodificada (indicada como $\mu_n^{(1)}$), se han probado otras métricas alternativas: minimización conjunta de energía media y energía de pico ($\mu_n^{(2)}$ y $\mu_n^{(3)}$), y minimización de la dispersión de la señal precodificada con respecto a una amplitud determinada ($\mu_n^{(4)}$). En líneas generales la BER ofrecida por la clásica métrica $\mu_n^{(1)}$ resulta ser la menor en todos los casos, aunque todas las métricas mejoran sensiblemente los resultados del sistema THP sin *shaping*.

7.2 Aplicación de THP en sistemas de transmisión multihaz por satélite

La aplicación de THP no ser restringe a la transmisión punto a punto a través de canales con ISI, sino que también se puede aplicar a la cancelación de la interferencia intracanal (CCI, del inglés *cochannel interference*) en esquemas de transmisión múltiple. En el Capítulo 3 proponemos la aplicación de THP a uno de estos sistemas, a saber, el enlace descendente en un sistema de comunicación por satélite con cobertura multihaz. En este tipo de sistemas, la cobertura de la zona deseada se consigue mediante la yuxtaposición de una serie de estrechos haces generados por la antena multihaz a bordo del satélite, que está compuesta por un gran número de elementos radiantes (del orden de la centena en sistemas actuales). Como sistema particular de referencia se ha escogido el propuesto en [79], en el que 96 haces dan cobertura al centro y oeste de Europa, tal y como muestra la Figura 7.8. La estrechez de los haces en los sistemas multihaz permite obtener mayores ganancias de radiación, pero esto acarrea la aparición de CCI en los las regiones correspondientes a haces adyacentes debido a los lóbulos laterales del diagrama de radiación de cada haz. En la práctica, esta CCI se consigue reducir mediante la utilización de frecuencias distintas en los haces adyacentes. Esto se consigue habitualmente mediante el empleo de patrones regulares de reutilización de frecuencia como el mostrado en la Figura 7.8, que utiliza un factor de reutilización de $N_{\text{fr}} = 4$ frecuencias.

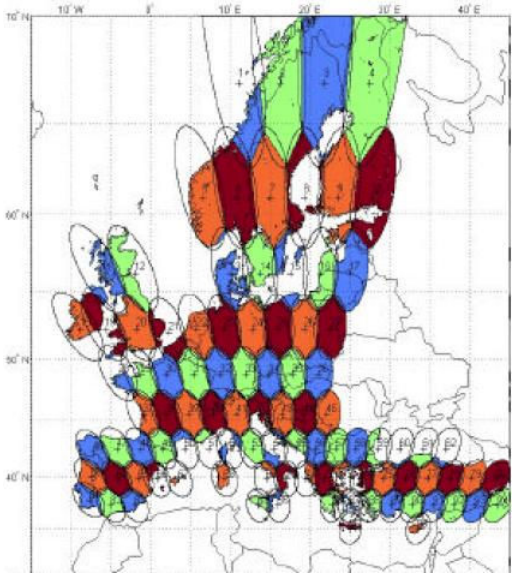


Figura 7.8: Cobertura y distribución de haces en el sistema multihaz considerado. Los cuatro colores representan el factor de reutilización de frecuencia $N_{\text{fr}} = 4$ propuesto en [79].

En el sistema estudiado en el Capítulo 3 se supone que existe un solo terminal receptor con una antena directiva en cada uno de los N_{B} haces. Los receptores no cooperan entre sí. Independientemente del factor de reutilización de frecuencia empleado, cada receptor recibirá, además de la señal correspondiente al propio haz, las señales no deseadas, aunque atenuadas, correspondientes a los demás haces, como se representa en el esquema de la Figura 7.9. Utilizando notación vectorial como es habitual cuando se trata de transmisión múltiple con salida simple (MISO, del inglés *multiple-input single-output*), el modelo de señal para el sistema considerado se puede escribir como

$$\boldsymbol{y}_n = \boldsymbol{H}\boldsymbol{a}_n + \boldsymbol{w}_n \tag{7.3}$$

en donde $\boldsymbol{y}_n$ agrupa las señales recibidas por los terminales en el instante n, $\boldsymbol{a}_n$ agrupa los símbolos transmitidos por cada haz, $\boldsymbol{w}_n$ es ruido aditivo blanco gaussiano, y $\boldsymbol{H}$ es la matriz que representa el comportamiento del canal como filtro espacial: modela las combinaciones lineales de los distintos haces que recibe cada terminal. Por ejemplo la señal recibida por el terminal i-ésimo se puede escribir como

$$[\boldsymbol{y}_n]_i = \sum_{j=1}^{N_{\text{Tx}}} h_{ij}\,[\boldsymbol{a}_n]_j + [\boldsymbol{w}_n]_i = h_{ii}\,[\boldsymbol{a}_n]_i + \sum_{\substack{j=1 \\ j\neq i}}^{N_{\text{B}}} h_{ij}\,[\boldsymbol{a}_n]_j + [\boldsymbol{w}_n]_i \tag{7.4}$$

en donde el segundo término por la derecha representa la CCI recibida por el i-ésimo terminal.

Al igual que el caso punto a punto del Capítulo 2, THP para transmisión MISO es una operación no lineal (debido de nuevo a la operación módulo) que tiene su equivalente lineal conocido como *precodificación lineal*. La figura 7.10 representa el esquema general

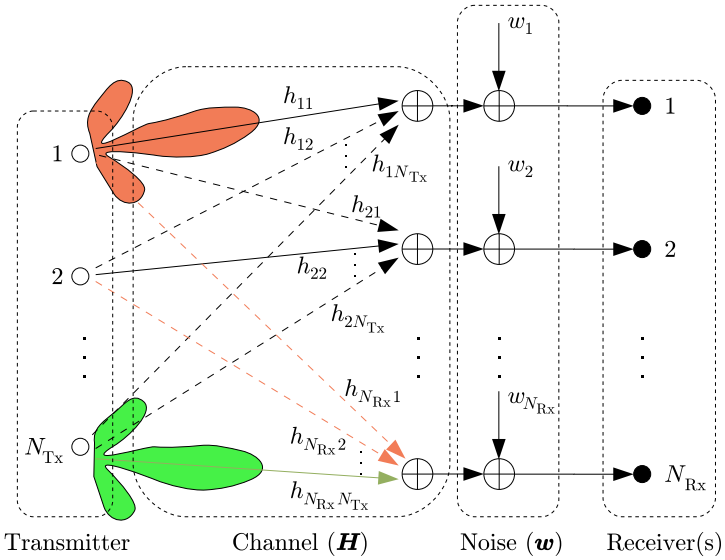


Figura 7.9: Representación de un sistema de transmisión MISO, en donde los receptores no colaboran y tienen una sola antena.

de un sistema de transmisión THP MISO. A continuación detallamos su funcionamiento bloque a bloque:

- En primer lugar las filas del vector original $\boldsymbol{a}$ (que supondremos compuesto por símbolos QAM) se reordenan por medio de la matriz de permutación $\boldsymbol{\Pi}$. Las prestaciones de THP dependen en general del orden en que se precodifiquen los elementos de $\boldsymbol{a}$. Por tanto, la operación de reordenamiento pretende seleccionar el orden adecuado de precodificación. Dado que el reordenamiento es transparente desde el punto de vista de los receptores, se puede ver como un grado de libertad adicional en el diseño de la operación de precodificación.
- A continuación se encuentra el bucle realimentado no lineal formado por la operación módulo en la rama hacia adelante y una matriz triangular inferior estricta $\boldsymbol{B}$ en la rama de realimentación. La operación de precodificación se realiza de forma secuencial empezando por el primer elemento del vector reordenado $\boldsymbol{\Pi a}$, después el segundo al que se le sustrae la interferencia causada por el primero, y así sucesivamente, como se indica a continuación

$$[\tilde{\boldsymbol{x}}]_1 = \mathrm{MOD}_\Delta\left([\boldsymbol{\Pi a}]_1\right) = [\boldsymbol{\Pi a}]_1 \tag{7.5a}$$

$$[\tilde{\boldsymbol{x}}]_2 = \mathrm{MOD}_\Delta\left([\boldsymbol{\Pi a}]_2 + b_{21}\,[\tilde{\boldsymbol{x}}]_1\right) \tag{7.5b}$$

$$\vdots$$

$$[\tilde{\boldsymbol{x}}]_{N_{\mathrm{B}}} = \mathrm{MOD}_\Delta\left([\boldsymbol{\Pi a}]_{N_{\mathrm{B}}} + \sum_{j=1}^{N_{\mathrm{B}}-1} b_{N_{\mathrm{B}}j}\,[\tilde{\boldsymbol{x}}]_j\right). \tag{7.5c}$$

- A continuación la precodificación se completa mediante la aplicación del filtro hacia adelante $\boldsymbol{F}$.

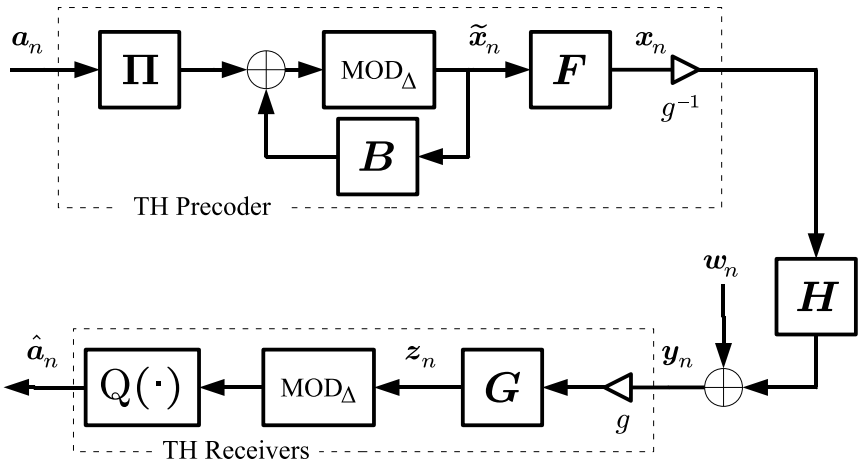


Figura 7.10: Esquema genérico de THP aplicado a transmisión multihaz con receptores no cooperantes.

- El escalado g^{-1} representa la limitación de potencia transmitida que existe en el sistema de transmisión. Su valor depende del las matrices $\boldsymbol{B}$ y $\boldsymbol{F}$ y de la potencia media permitida al transmisor.
- Tras el paso por el canal $\boldsymbol{H}$ y el añadido de ruido blanco, el receptor aplica un reescalado por g para devolver la señal a su nivel adecuado. La matriz diagonal $\boldsymbol{G}$ representa una posible operación de control de ganancia que podría ser necesaria si el esquema THP no ofrece el mismo nivel de señal a cada receptor.
- Finalmente cada receptor aplica la operación módulo tras la cual el decisor ofrece las estimas de los símbolos recibidos.

Las matrices $\boldsymbol{B}$ y $\boldsymbol{F}$ y el escalar g, que se obtienen a partir de cierta factorización de la matriz de canal $\boldsymbol{H}$, no son únicos sino que dependen de la factorización empleada y del criterio de cancelación de interferencia, ZF o MMSE (del inglés *zero-forcing*, cero forzado, y *minimum mean square error*, mínimo error cuadrático medio). En el Capítulo 3 se han considerado los esquemas THP presentados en [59, 98, 140], aunque aquí omitiremos los detalles por razones de espacio.

Para la evaluación de los sistemas escogidos se han escogido cuatro escenarios de aplicación variando la localización de los receptores dentro de su haz, el factor de reutilización de frecuencia N_{fr} y el número de haces considerados N_{B}. De entre los cuatro escenarios, referidos como S1, S2, S3 y S4, en los escenarios S1 a S3 se consideran 10 haces y 10 usuarios, mientras que en el escenario S4 se consideran 96 haces y usuarios. En términos de nivel de CCI, el escenario S1 ($N_{\text{fr}} = 1$)ofrece un nivel alto, los escenarios S2 y S4 ($N_{\text{fr}} = 1$ ambos) ofrecen un nivel intermedio/alto, mientras que el escenario S3 ofrece un nivel bajo de CCI debido a que reutiliza $N_{\text{fr}} = 3$ frecuencias. La Figura 7.11 representa esquemáticamente los haces considerados y la posición de los receptores para cada escenario.

En las Figuras 7.12 y 7.13 mostramos un extracto de los resultados obtenidos para los escenarios S1 (10 haces, CCI alta) y S4 (96 haces, CCI intermedia/alta) en términos de BER sin codificar para una 16-QAM. Como referencia se incluyen los sistemas de

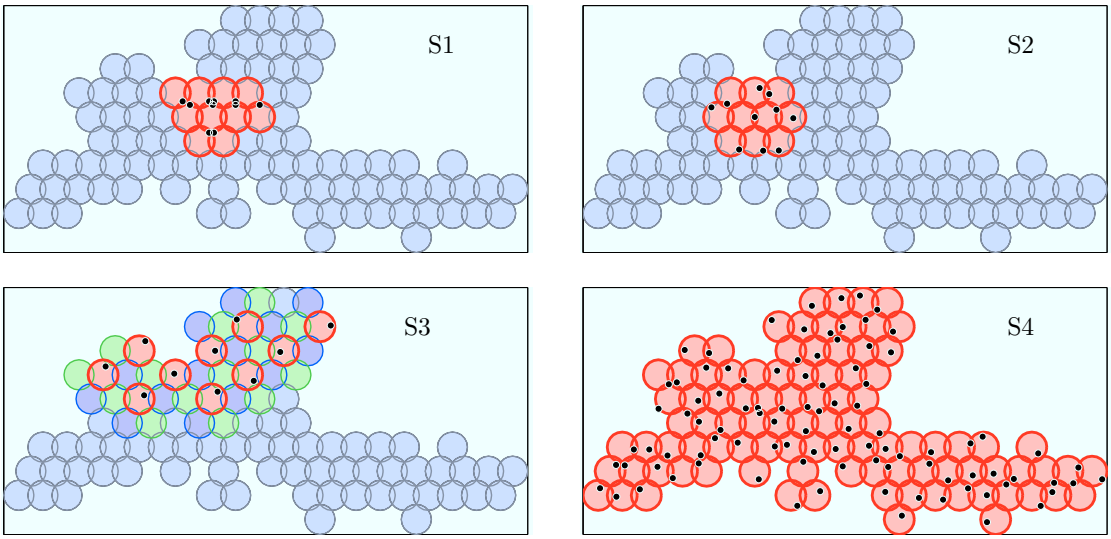


Figura 7.11: Haces considerados (circunferencias rojas) y posición de los receptores (puntos negros) en los escenarios S1, S2, S3, and S4. Los haces de color gris no causan interferencia.

precodificación lineal ZF y MMSE [99]. En líneas generales, se observa que los esquemas MMSE son mejore que los ZF para SNRs bajas, y sobre todo, que THP ofrece unas prestaciones muy superiores a la precodificación no lineal. Por ejemplo, para el escenario S1 el mejor sistema THP alcanza una BER de 10^{-4} para $E_{\mathrm{b}}/N_0 = 23$ dB, mientras que los esquemas lineales necesitan $E_{\mathrm{b}}/N_0 = 34$ dB para alcanzar la misma BER. Esta diferencia, que aparece en mayor o menor medida siempre que el nivel de CCI no es despreciable, se debe a la penalización de de potencia transmitida que sufren las técnicas de precodificación lineal.

Recientemente, en [43, 67] se ha estudiado la aplicación de precodificación lineal a escenarios multihaz similares a los considerados en este capítulo, con resultados prometedores en términos de disponibilidad y tasa de transmisión agregada del sistema. Al igual que THP, la precodificación lineal alcanza a eliminar la CCI sin necesidad de reutilizar frecuencias adicionales, lo cual aumenta la capacidad de transmisión potencial del sistema. El hecho de que THP evite la penalización de potencia transmitida propia de la precodificación lineal, sugiere que el empleo de THP aportaría beneficios aún mayores en términos de capacidad de transmisión del sistema que los ofrecidos por los sistemas estudiados en [43, 67].

7.3 Estimación de SNR

El Capítulo 4 se proponen métodos de estimación de la SNR aplicables en principio en un rango amplio de situaciones prácticas. La estimación de SNR es necesaria en muchos de los actuales sistemas de comunicaciones que implementan codificación blanda [153], esquemas de control de potencia o de codificación y modulación adaptativa [4, 17]. Existe una gran variedad de estimadores de SNR que responde a distintas condiciones de

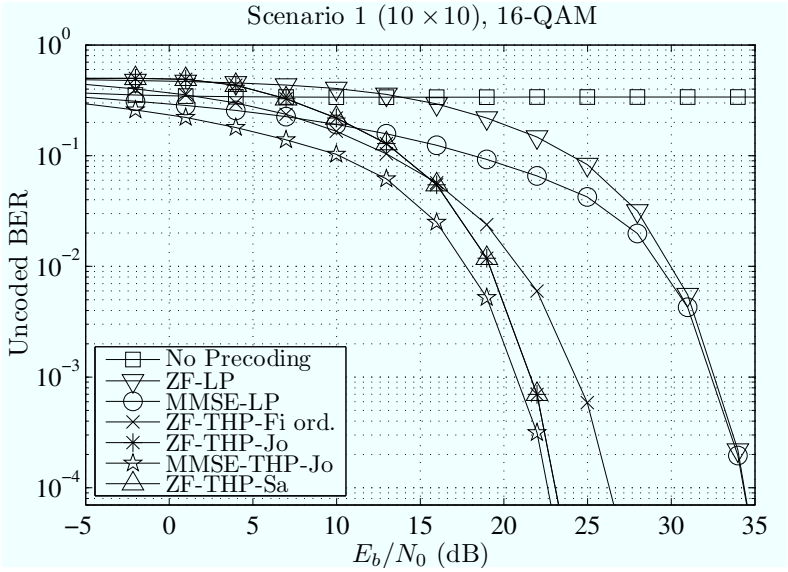


Figura 7.12: BER sin codificar de los sistemas de precodificación considerados para el escenario S1 y modulación 16-QAM.

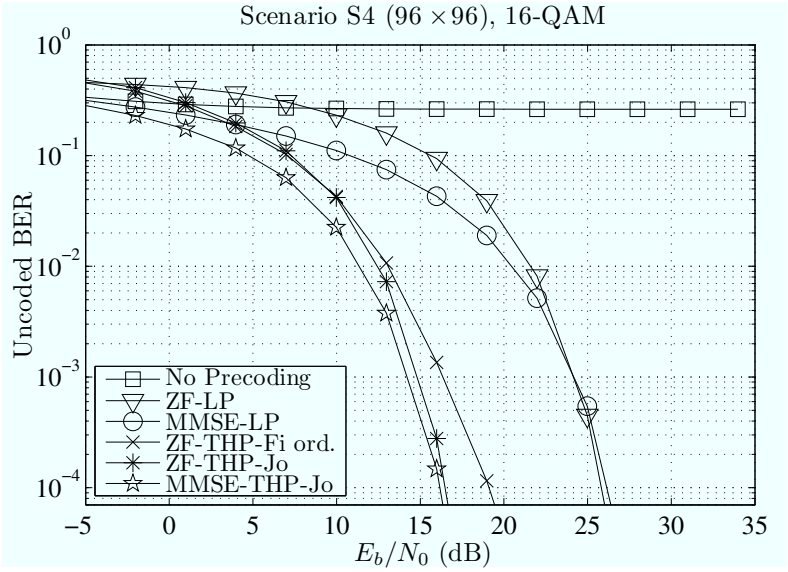


Figura 7.13: BER sin codificar de los sistemas de precodificación considerados para el escenario S4 y modulación 16-QAM.

aplicación. Los estimadores considerados en el Capítulo 4 son estimadores que operan a tasa de símbolo y no ayudados por datos (NDA, del inglés *non-data aided*) principalmente. Existen otro tipo de estimadores que operan a partir de señales sobremuestreadas [131, 136, 143], o con conocimiento de los datos (DA, del inglés *data-aided*). De entre los estimadores NDA, el capítulo se centra principalmente en los estimadores basados en la envolvente de las señal (EVB, del inglés *envelope-based*). Dentro de los estimadores EVB, los estimadores basados en momentos representan opciones atractivas dada su sencillez de implementación. Como ejemplos de estos estimadores se pueden citar el estimador $\mathrm{M_1M_2}$ [23] o el estimador $\mathrm{M_2M_4}$ [118] que utilizan los momentos M_1, M_2 y M_4. Estos estimadores funcionan bien para modulaciones de módulo constante (CM, del inglés *constant modulus*) como las PSK, pero sus prestaciones se degradan para modulaciones con varios niveles (APSK, QAM). Recientemente se han propuesto algunos estimadores nuevos basados en momentos que mejoran a los anteriores cuando se aplican a dichas constelaciones. En particular, en [110] se presenta un nuevo estimador basado en los momentos M_2, M_4 y M_6 cuyas prestaciones mejoran mucho las de $\mathrm{M_2M_4}$ para constelaciones de dos niveles.

Asumiendo un modelo estático de señal con desvanecimiento plano, y con sincronización perfecta de portadora y muestreo, las muestras a tasa de símbolo a la salida del filtro adaptado vienen dadas por

$$r_k = \sqrt{S}x_k + n_k, \qquad k = 1, \ldots, K \tag{7.6}$$

siendo x_k los símbolos complejos transmitidos, n_k las muestras de ruido blanco gaussiano complejo circular con varianza desconocida N, y $\sqrt{S}$ la ganancia desconocida del canal. Los símbolos x_k pertenecen a una constelación con I amplitudes, con R_i y P_i la amplitude i-ésima y su probabilidad asociada. Los *momentos de la constelación* se definen como $c_p \doteq \mathrm{E}\left\{|x_k|^p\right\} = \sum_{i=1}^{I} P_i R_i^p$. Los *momentos muestrales* de la envolvente, que se calculan como

$$\hat{M}_p \doteq \frac{1}{K}\sum_{k=1}^{K} |r_k|^p, \tag{7.7}$$

son estimadores insesgados de los *momentos verdaderos* de la envolvente, que en general vienen dados por

$$\begin{aligned} M_p &\doteq \mathrm{E}\left\{|r_k|^p\right\} \\ &= N^{\frac{p}{2}} \sum_{i=1}^{I} P_i \Gamma\left(\frac{p}{2}+1\right) e^{-\rho R_i^2} {}_1F_1\left(\frac{p}{2}+1; 1; \rho R_i^2\right) \end{aligned} \tag{7.8}$$

siendo ρ la SNR en unidades naturales, $\Gamma(\cdot)$ la función gamma y ${}_1F_1(\cdot;\cdot;\cdot)$ la función hipergeométrica confluente. Para orden p par (7.8) admite una expresión más sencilla que depende de ρ, N y los momentos de la constelación [86, 110]:

$$M_{2n} = N^n \sum_{m=0}^{n} \frac{(n!)^2}{(n-m)!(m!)^2} c_{2m} \rho^m. \tag{7.9}$$

Muchos de los estimadores basados en momentos se construyen a partir de cocientes de potencias de momentos en donde N se cancela. Por ejemplo, en [68] se propone la familia de funciones

$$f_{k,l}(\rho) \doteq \frac{M_k^l}{M_l^k}, \quad \text{para } k \neq l \tag{7.10}$$

a partir de la cual, empleando momentos muestrales en lugar de los verdaderos, se obtienen las estimas de SNR mediante inversión de (7.10)

$$\hat{\rho}_{k,l} = f_{k,l}^{-1}\left(\frac{\hat{M}_k^l}{\hat{M}_l^k}\right). \tag{7.11}$$

En la Sección 4.3 se propone una nueva familia de estimadores basada en cocientes de momentos que generaliza tanto la familia propuesta en [68] como el estimador de sexto orden propuesto en [110]. La familia propuesta se compone de combinaciones lineales de cocientes de la forma

$$f(\rho) \doteq \left(\prod_{i=1}^{U} M_{2n_i}^{p_i}\right) \Big/ M_2^Q, \quad n_i, p_i \in \mathbb{N} \tag{7.12}$$

en donde $Q = \sum_{i=1}^{U} n_i p_i$ para cumplir que N desaparece del cociente ($2Q$ es el orden estadístico del cociente). Estos cocientes se componen exclusivamente de momentos de orden par, y presentan la propiedad de que tras aplicar el cambio de variable $z = \rho/(1+\rho)$ (SNR normalizada), pueden reducirse a un polinomio $F(z)$ de orden Q. En particular, en la Sección 4.3 se propone el estimador compuesto de todos los cocientes de esta familia de orden igual o inferior a 8:

$$f_{\text{EOS}}(\rho) \doteq \beta\frac{M_4}{M_2^2} + \gamma\frac{M_6}{M_2^3} + \delta\frac{M_4^2}{M_2^4} + \epsilon\frac{M_8}{M_2^4} \tag{7.13}$$

siendo $\boldsymbol{\alpha} = [\beta \;\; \gamma \;\; \delta \;\; \epsilon]^{\mathrm{T}} \in \mathbb{R}^4$ el vector de pesos de la combinación lineal. Este estimador generaliza otros anteriores como $\mathrm{M_2M_4}$, que se construye a partir de $f(\rho) = M_4/M_2^2$, o el estimador de sexto orden propuesto en [110], que se extrae de $f(\rho) = -bM^4/M_2^2 + M_6/M_2^3$. La combinación lineal (7.13) se puede reducir a

$$F_{\text{EOS}}(z) = F_4\, z^4 + F_3\, z^3 + F_2\, z^2 + F_0 \tag{7.14}$$

en donde los coeficientes F_i son funciones exclusivas de los momentos de la constelación y de los pesos de $\boldsymbol{\alpha}$. El estimador de la SNR normalizada $\hat{z} \in (0,1)$ viene dado por la solución de

$$F_{\text{EOS}}(\hat{z}) = \beta\frac{\hat{M}_4}{\hat{M}_2^2} + \gamma\frac{\hat{M}_6}{\hat{M}_2^3} + \delta\frac{\hat{M}_4^2}{\hat{M}_2^4} + \epsilon\frac{\hat{M}_8}{\hat{M}_2^4} = h_{\text{EOS}}(\hat{\boldsymbol{m}}) \tag{7.15}$$

con $\hat{\boldsymbol{m}} \doteq [\hat{M}_2 \;\; \hat{M}_4 \;\; \hat{M}_6 \;\; \hat{M}_8]^{\mathrm{T}} \in \mathbb{R}^4$ el vector de momentos muestrales. Para encontrar la estima se puede aplicar el método algebraico de cálculo de raíces de polinomios de cuarto grado [1, pp.18-19], o bien optar por un método iterativo, como por ejemplo:

$$\hat{z}^{(0)} = 1, \qquad \hat{z}^{(n+1)} = \sqrt{\frac{h_{\text{EOS}}(\hat{\boldsymbol{m}}) - F_0}{F_4(\hat{z}^{(n)})^2 + F_3\,\hat{z}^{(n)} + F_2}}. \tag{7.16}$$

El análisis estadístico del estimador basado en aproximaciones de Taylor ofrece que la varianza, $\mathrm{Var}\{\hat{\rho}\}$ y el sesgo $\mathrm{Bias}\{\hat{\rho}\}$ vienen dados por

$$\mathrm{Var}\{\hat{\rho}\} \approx \boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{v} = \frac{1}{K}\frac{\sum_{n=0}^{10} A_n^{(\mathrm{Var})}\rho^n}{\sum_{n=2}^{6} B_n^{(\mathrm{Var})}\rho^n} \tag{7.17}$$

$$\mathrm{Bias}\{\hat{\rho}\} \approx \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{H}\,\boldsymbol{C}\right) = \frac{1}{K}\,L(\rho) + J(\rho)\,\mathrm{Var}\{\hat{\rho}\} = \frac{1}{K}\frac{\sum_{n=0}^{12} A_n^{(\mathrm{Bias})}\rho^n}{\sum_{n=3}^{9} B_n^{(\mathrm{Bias})}\rho^n} \tag{7.18}$$

en donde $\boldsymbol{C}$ es la matriz de covarianza de $\hat{\boldsymbol{m}}$, $\boldsymbol{v}$ y $\boldsymbol{H}$ dependen de $\boldsymbol{\alpha}$, ρ y los momentos de la constelación, al igual que los cocientes de polinomios $L(\rho) \doteq A^{(L)}(\rho)/B^{(L)}(\rho)$, con $A^{(L)}(\rho) = \sum_{n=0}^{6} A_n^{(L)}\rho^n$, $B^{(L)}(\rho) = \sum_{n=1}^{4} B_n^{(L)}\rho^n$, y $J(\rho) \doteq A^{(J)}(\rho)/B^{(L)}(\rho)$, con $A^{(J)}(\rho) = \sum_{n=0}^{3} A_n^{(J)}\rho^n$. Los polinomios $A^{(\mathrm{Var})}(\rho)$, $B^{(\mathrm{Var})}(\rho)$ son cuadráticos en $\boldsymbol{\alpha}$, mientras que $A^{(L)}(\rho)$, $B^{(L)}(\rho)$ y $A^{(J)}(\rho)$ son lineales en $\boldsymbol{\alpha}$.

En general, tanto la varianza como el sesgo son funciones de los momentos de la constelación y los pesos $\boldsymbol{\alpha}$, además de la SNR ρ. Por tanto, $\boldsymbol{\alpha}$ representa un conjunto de grados de libertad en el diseño del estimador. Parece lógico emplear dichos grados de libertad en optimizar las prestaciones del estimador resultante en términos de sesgo y varianza. Para ello, en la Sección 4.3 se proponen dos criterios de optimización alternativos:

- **C1: optimización para SNR alta.** Usar un grado de libertad en $\boldsymbol{\alpha}$ en anular el coeficiente de mayor orden de $L(\rho)$, y el resto en minimizar el coeficiente de mayor orden de $\mathrm{Var}\{\hat{\rho}\}$. Si éste se puede anular, y sobran grados de libertad, usar estos en minimizar el siguiente coeficiente de $\mathrm{Var}\{\hat{\rho}\}$, y así sucesivamente.
- **C2: optimización para un valor nominal de SNR.** Dado un valor particular de la SNR ρ_0, minimizar la varianza (7.17) bajo la restricción de que $L(\rho)$ en (7.18) sea cero.

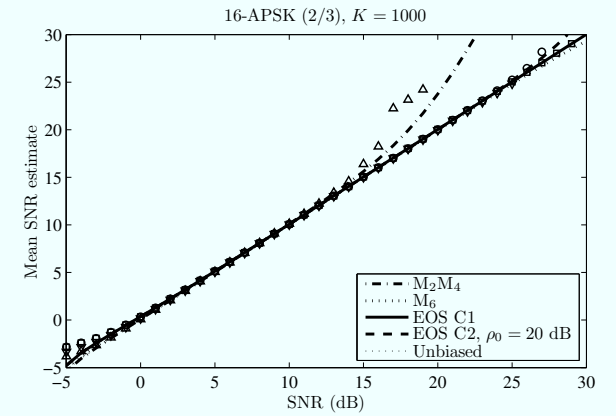
La aplicación de estos criterios permite obtener estimadores cuyas prestaciones para constelaciones de varios niveles mejoran sustancialmente la de los estimadores previos basados en momentos. Por ejemplo, los resultados para una constelación 16-APSK (dos niveles) se muestran en la Figura 7.14, tanto en sus valores teóricos como empíricos obtenidos mediante simulación.

En la Sección 4.4 se trata del problema de estimar la SNR a partir de conjuntos de muestras pertenecientes a varias constelaciones. Este caso puede darse por ejemplo si se pretende estimar la SNR a partir de las muestras de una trama DVB-S2 [54], en donde los símbolos de la cabecera y los símbolos pilotos pertenecen a una constelación distinta a la empleada para los datos (trama *heterogénea*), como se puede ver en la Figura 7.15 El modelo de señal viene dado por

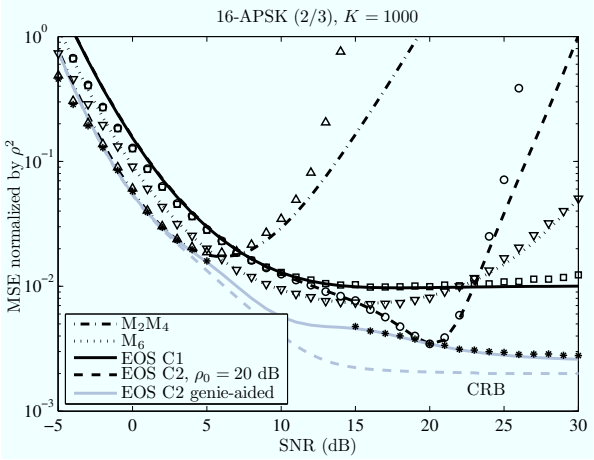
$$r_k = \sqrt{S}x_k + n_k, \qquad k \in \{1,\ldots,K\} = \mathcal{K}_1 \cup \mathcal{K}_2 \tag{7.19}$$

en donde $\mathcal{K}_1$ y $\mathcal{K}_2$ son conjuntos disjuntos de índices que indexan los símbolos de la trama que pertenecen a dos constelaciones distintas, y son conocidos por el receptor. (7.6)

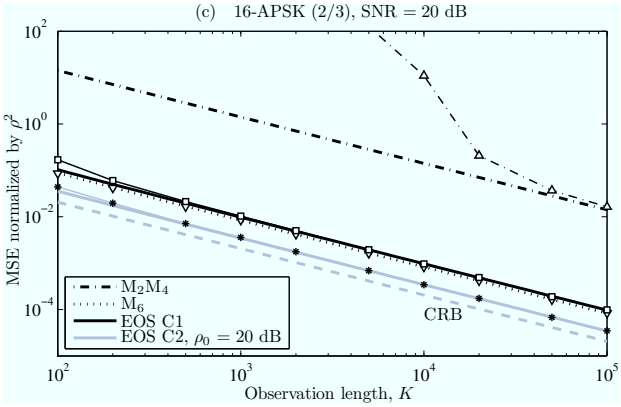
Para estimar la SNR a partir de esta trama, podría plantearse el empleo de un estimador NDA directamente sobre toda la trama, o bien sobre los símbolos de una de las subtramas.



(a)



(b)



(c)

Figura 7.14: Prestaciones teóricas (líneas) y empíricas (marcadores) del estimador propuesto en términos de (a) sesgo frente a SNR, (b) NMSE frente a SNR y (c) NMSE frente a K para una 16-APSK con cociente entre radios $R_2/R_1 = 3{,}15$. Leyenda: $\triangle$ para M_2M_4, ∇ para M_6, $\square$ para C1, $\circ$ para C2 con $\rho_0 = 20$ dB, y $*$ para C2 "genie-aided".

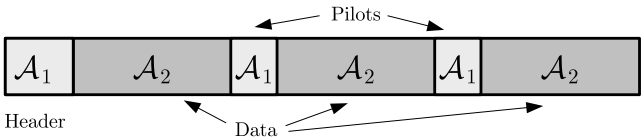


Figura 7.15: Ejemplo de una trama heterogénea que contiene símbolos de dos constelaciones distintas $\mathcal{A}_1$, $\mathcal{A}_2$.

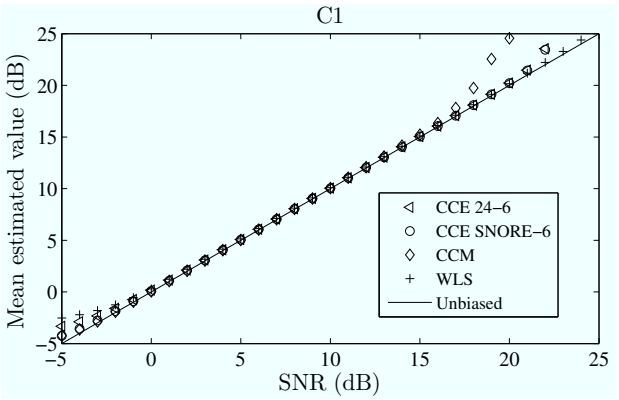
Ninguna de estas dos opciones sería óptima en el sentido de que no emplean toda la información disponible. Por ello, en la Sección 4.4 se proponen tres esquemas alternativos de estimación:

- **Combinación convexa de estimas, CCE** (del inglés *convex combination of estimates*). En este esquema el estimador se construye a partir de dos estimadores distintos que se aplican cada uno a los símbolos de una constelación $\hat{\rho}_{\text{cce}} = \lambda_{\text{cce}} \cdot \hat{\rho}^{(1)} + (1 - \lambda_{\text{cce}}) \cdot \hat{\rho}^{(2)}$. Los estimadores de partida $\hat{\rho}^{(1)}$ y $\hat{\rho}^{(2)}$ pueden ser de cualquier tipo, pero si son insesgados, el estimador resultante también lo será. El parámetro λ_{cce} se puede optimizar para minimizar la varianza del estimador resultante.
- **Combinación convexa de momentos, CCM** (del inglés *convex combination of moments*). En este esquema se utilizan estimadores basados en momentos que se construyen a partir de *momentos modificados*. Estos momentos modificados son combinaciones convexas de los momentos de cada uno de los dos subconjuntos: $\hat{M}_n^{\text{ccm}} = \lambda_{\text{ccm},n} \cdot \hat{M}_n^{(1)} + (1 - \lambda_{\text{ccm},n}) \cdot \hat{M}_n^{(2)}$, en donde los parámetros $\lambda_{\text{ccm},n}$ se pueden optimizar para minimizar la varianza resultante.
- **Mínimos cuadrados ponderados, WLS** (del inglés *weighted least squares*). Este esquema busca un estimador a partir de los momentos M_2 y M_4 aplicados a cada una de las dos subtramas. Las cuatro ecuaciones se combinan linealmente y de esa combinación se extrae un estimador de la SNR. Los pesos de la combinación se optimizan para minimizar la varianza.

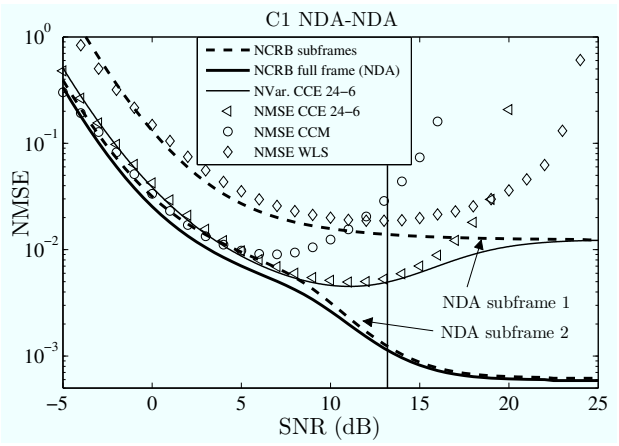
En la Figura 7.16 se muestran los resultados obtenidos con los tres esquemas propuestos para la estimación a partir de una trama corta DVB-S2 formada por una QPSK (162 símbolos) y una 32-APSK (3240 símbolos). En general, el esquema CCE, usando los estimadores M_2M_4 y M_6 para cada subtrama, ofrece con diferencia mucho mejores prestaciones en términos de sesgo y NMSE que los esquemas CCM y WLS. Además, el esquema CCE presenta la flexibilidad de poderse emplear con cualesquier estimadores disponibles, incluyendo estimadores DA. El empleo de estimadores DA tiene interés práctico, dado que comúnmente algunos de lo símbolos en cada trama son conocidos (inicio de trama, símbolos pilotos). En la Figura 7.16c se representan las prestaciones del esquema CCE que combina el estimador DA ML (SNORE) [66, 131] con M_6.

7.4 Estimación de fase para modulaciones QAM en cruz

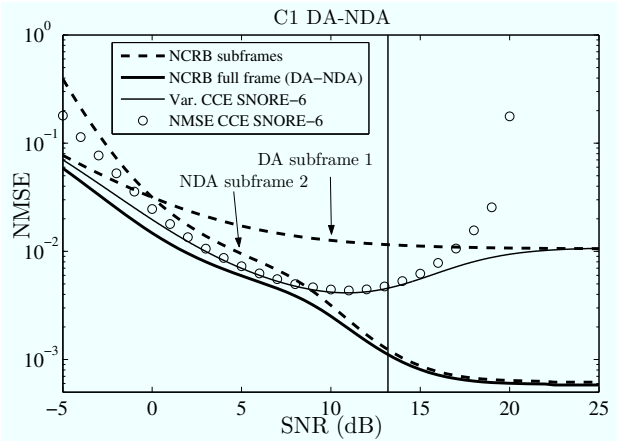
El Capítulo 5 propone dos estimadores NDA de fase para constelaciones QAM en cruz, como por ejemplo la 32-QAM y la 128-QAM. Métodos NDA clásicos como los estima-



(a)



(b)



(c)

Figura 7.16: Prestaciones de los estimadores propuestos en términos de (a) valor medio estimado (a), y NMSEs para combinación de estimadores (b) NDA y (c) DA-NDA. Se muestran varianzas analíticas, CRBs normalizadas y NMSEs obtenidos mediante simulación.

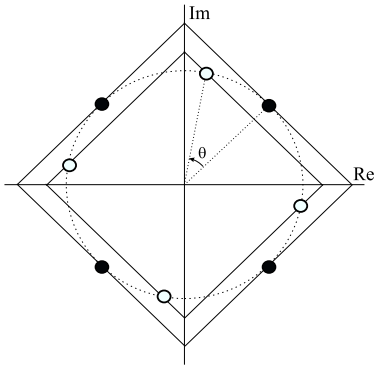


Figura 7.17: Efecto de una rotación en la norma ℓ_1 media de una constelación QPSK. •: constelación original; ○: constelación rotada.

dores Viterbi y Viterbi (V&V) basados en potencias [161], que incluyen el método de la cuarta potencia (4P) [122], funcionan bien cuando se aplican a modulaciones PSK o QAM cuadradas. Sin embargo, sus prestaciones se degradan sustancialmente cuando se aplican a modulaciones QAM en cruz. En los últimos años se han presentado algunos estimadores que mejoran mucho las prestaciones para dichas constelaciones, pero que, o bien requieren del empleo de estadísticos de octavo orden [28, 29] o bien sólo funcionan bien para SNRs altas [31]. De cara a la implementación práctica (posiblemente en punto fijo) de un estimador, interesa evitar los estadísticos de orden alto, dado que en general conllevan un mayor número de operaciones, lo cual aumenta el riesgo de desbordamiento o pérdida de precisión.

Con el objeto de ofrecer estimadores con buenas prestaciones para QAMs en cruz, y que además sean atractivos para implementación en punto fijo, en el Capítulo 5 se proponen dos estimadores basados en el empleo de la norma ℓ_1. Partiendo de las muestras a tasa de símbolo (asumiendo sincronización de portadora adquirida)

$$r_k = a_k \mathrm{e}^{j\theta} + n_k, \qquad k = 0, 1, \ldots, L-1 \tag{7.20}$$

en donde a_k son los símbolos transmitidos con varianza E_a, n_k es ruido gaussiano complejo circular independiente de los símbolos y de varianza σ^2, y θ es el desfase desconocido, que se asume de variación lenta. Dada una posible estima $\hat{\theta}$, definimos las muestras *desrotadas* $y_k \cdot \mathrm{e}^{-j\hat{\theta}} r_k = \bar{y}_k + j\tilde{y}_k$, y consideramos el efecto de esa desrotación en la norma ℓ_1 de vector $\boldsymbol{y}_k \cdot f(y_k) = [\bar{y}_k \quad \tilde{y}_k]^{\mathrm{T}}$, que viene dada por

$$||\boldsymbol{y}_k||_1 = |\bar{y}_k| + |\tilde{y}_k|. \tag{7.21}$$

Como se observa en la Figura 7.17, cualquier rotación no exacta (excepto errores múltiplos de $\pi/2$) disminuye el valor medio de la norma ℓ_1. Intuitivamente se puede conjeturar que la maximización de la norma ℓ_1 puede ser un buen criterio de partida para obtener un estimador de fase. Con esta idea, se propone la siguiente familia de funciones de coste

$$J_q(\hat{\theta}) \doteq \mathrm{E}\left\{||\boldsymbol{y}||_1^q\right\} = \mathrm{E}\left\{(|\bar{y}| + |\tilde{y}|)^q\right\}. \tag{7.22}$$

La Figura 7.18 muestra la variación de las funciones de coste con $q = 1$ y $q = 2$ para varias constelaciones en cruz. Como se puede observar, J_1 y J_2 son funciones multimodales, con un máximo absoluto para error de fase cero. 3 A partir de la maximización de J_1 es posible

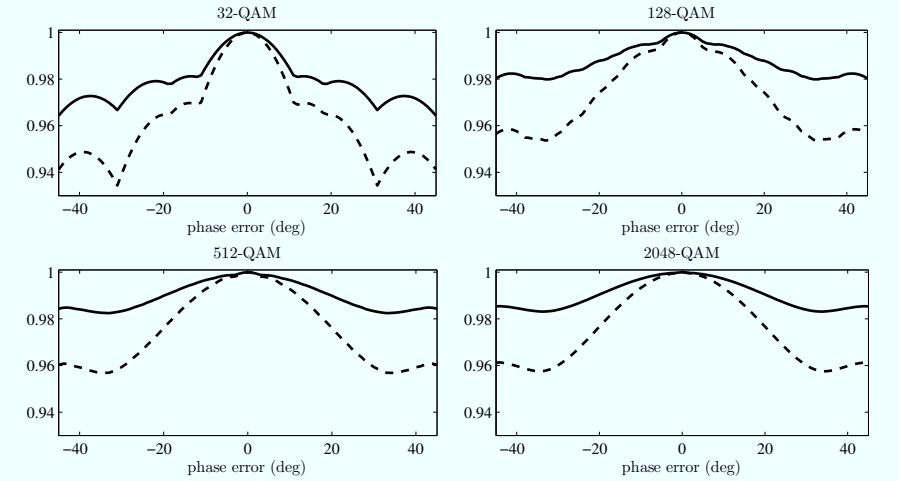


Figura 7.18: Constelaciones QAM en cruz: comportamiento de las funciones de coste basadas en la norma ℓ_1 $J_q(\hat{\theta})$ (normalizadas por su valor máximo) en ausencia de ruido. $q = 1$ (línea continua) y 2 (línea discontinua).

obtener el siguiente estimador

$$\hat{\theta}_{n+1}^{(1)} = -\arg\left\{\sum_{k=0}^{L-1} \operatorname{csgn}\left(r_k \cdot \mathrm{e}^{-j\hat{\theta}_n^{(1)}}\right) \cdot r_k^*\right\}. \tag{7.23}$$

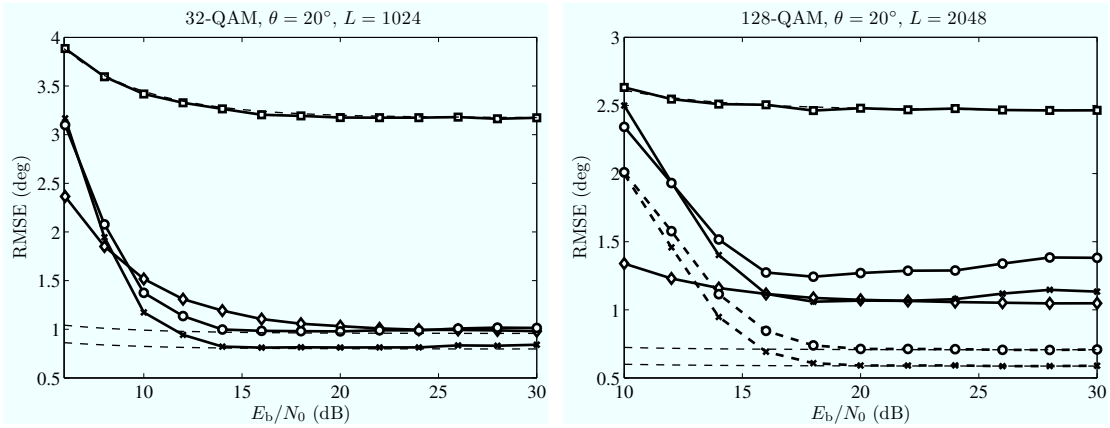
Asimismo, la maximización de J_2 ofrece el siguiente estimador

$$s_k(\hat{\theta}_n^{(2)}) = \operatorname{sgn}\left(\operatorname{Im}\left\{r_k^2 \mathrm{e}^{-j2\hat{\theta}_n^{(2)}}\right\}\right),\ k = 0, \ldots, L-1 \tag{7.24a}$$

$$\hat{\theta}_{n+1}^{(2)} = \frac{1}{2}\arg\left\{\sum_{k=0}^{L-1}\left[r_k^2 s_k(\hat{\theta}_n^{(2)})\right]\right\} - \frac{\pi}{4}. \tag{7.24b}$$

Ambos estimadores dependen de una estima inicial $\hat{\theta}_0^{(q)}$ que ha de tomarse de algún estimador existente, como por ejemplo, el estimador 4P [122] o el estimador de Cartwright de octavo orden (C8) [29]. La estima ha de ser lo suficientemente buena como para colocar la iteración dentro del dominio de atracción del máximo absoluto. Como se puede observar, las operaciones realizadas por los estimadores (7.23) y (7.24) son como máximo de segundo orden estadístico, lo cual supone una ventaja apreciable respecto a 4P y C8 de cara a la implementación en punto fijo. En la Sección 5.5 se obtienen las varianzas de ambos estimadores para los casos asintóticos de SNR suficientemente alta y longitud de las observaciones suficientemente larga.

Las prestaciones de ambos estimadores inicializados con 4P y C8 se muestran en la Figura (7.19) para el caso de implementación en punto flotante tras cinco iteraciones. Como se puede observar ambos estimadores ofrecen un RMSE comparable (32-QAM)



(a) (b)

Figura 7.19: RMSE frente a la SNR por bit para (a) 32-QAM y (b) 128-QAM. Leyenda: 4P (□), C8 (◇), J_1 (×), J_2 (○). J_1 y J_2 se inicializan con 4P (línea continua) y C8 (línea discontinua). Las líneas finas discontinuas muestran los valores de RMSE teóricos de los estimadores 4P, $\hat{\theta}^{(1)}$ y $\hat{\theta}^{(2)}$.

o incluso menor (128-QAM) que el ofrecido por el estimador C8, incluso cuando son inicializados por la peor estima (4P).

Los mismos resultados, pero obtenidos con la implementación en punto fijo, se muestran en la Figura 7.20. Los resultados para la constelación 32-QAM (con longitudes de palabra de 12, 10 y 8 bits) dan muestra de la robustez de los métodos propuestos frente a los existentes de mayor orden (4P y C8). A medida que disminuye la longitud de palabra, los métodos de mayor orden dejan de funcionar correctamente, mientras que los métodos iterativos propuestos se ven mucho menos afectados. Lo mismo se observa para el caso de la 128-QAM.

Estos resultados permite concluir que los estimadores propuestos no sólo mejoran las prestaciones de sus inicializadores, sino que mantienen sus prestaciones en condiciones exigentes de punto fijo, lo cual los hace atractivos de cara a la implementación práctica. En cuanto a su inicialización, se ha visto que los estimadores 4P y C8 funcionan razonablemente bie, aunque otros esquemas como los V&V o incluso esquemas DA (en el caso de que el bloque de datos contenga símbolos de entrenamiento) podrían ser igualmente útiles.

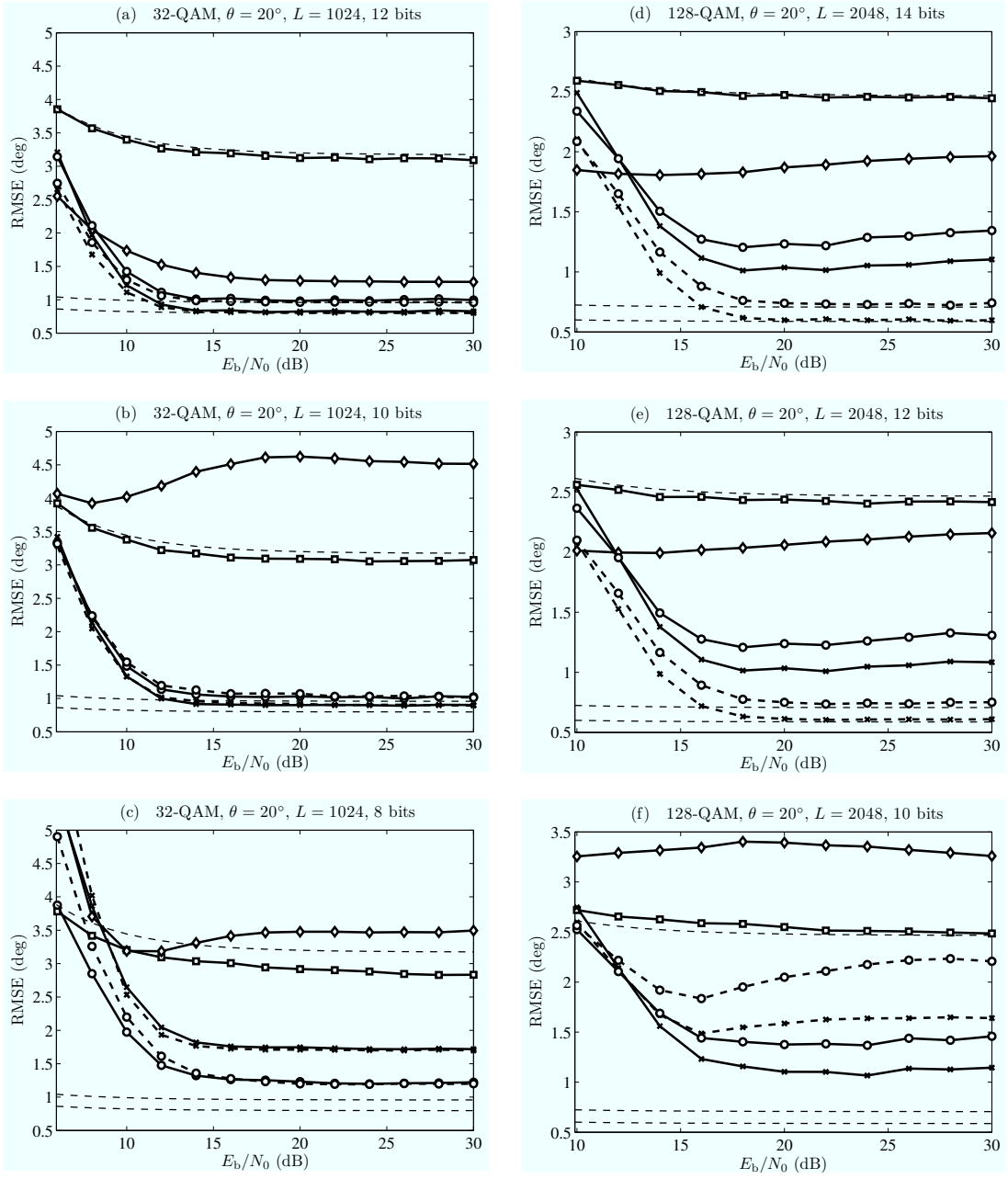


Figura 7.20: RMSE frente a E_b/N_0 bajo aritmética de punto fijo.
32-QAM, $L = 1024$, $E_b/N_0 = 30$ dB: (a) $B = 12$, (b) $B = 10$, y (c) $B = 8$ bits.
128-QAM, $L = 2048$, $E_b/N_0 = 30$ dB: (d) $B = 14$, (e) $B = 12$, y (f) $B = 10$ bits.
Leyenda: 4P (□), C8 (◇), J_1 (×), J_2 (○). J_1 y J_2 se inicializan con 4P (línea continua) y C8 (línea discontinua). Las líneas finas discontinuas muestran los valores de RMSE teóricos de los estimadores 4P, $\hat{\theta}^{(1)}$ y $\hat{\theta}^{(2)}$.

Bibliography

[1] M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions*, 9th ed. New York: Dover, 1972.

[2] S. Abrar, "Blind carrier phase recovery in cross QAM communication systems," in *Proc. International Conference on Emerging Technologies ICET '06*, 13–14 Nov. 2006, pp. 1–7.

[3] N. Alagha, "Cramer-Rao bounds of SNR estimates for BPSK and QPSK modulated signals," *IEEE Commun. Lett.*, vol. 5, no. 1, pp. 10–12, Jan. 2001.

[4] G. Albertazzi, S. Cioni, G. Corazza, M. Neri, R. Pedone, P. Salmi, A. Vanelli-Coralli, and M. Villanti, "On the adaptive DVB-S2 physical layer: Design and performance," *IEEE Wireless Commun. Mag.*, vol. 12, no. 6, pp. 62–68, Dec. 2005.

[5] M. Álvarez-Díaz, R. López-Valcarce, and I. Gómez Alonso, "Blind ℓ_1-norm-based phase estimation for cross-QAM robust to finite precision effects," *Signal Processing*, accepted for publication, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.sigpro.2010.04.019

[6] M. Álvarez-Díaz, R. López-Valcarce, and C. Mosquera, "SNR estimation for multilevel constellations using higher-order moments," *IEEE Trans. Signal Process.*, vol. 58, no. 3, Part 2, pp. 1515–1526, March 2010.

[7] M. Álvarez-Díaz, N. Courville, C. Mosquera, G. Liva, and G. E. Corazza, "Non-linear interference mitigation for broadband multimedia satellite systems," in *Proc. of International Workshop on Satellite and Space Communications 2007 (IWSSC'07)*, Salzburg, Austria, September 13-14 2007, pp. 61–65.

[8] M. Álvarez-Díaz and R. López-Valcarce, "Diamond contour-based phase recovery for (cross)-QAM constellations," in *Proc. IEEE/SP 13th Workshop on Statistical Signal Processing*, Bourdeaux, France, July 17-20 2005, pp. 669–674.

[9] M. Álvarez-Díaz, R. López-Valcarce, and C. Mosquera, "SNR estimation with heterogeneous frames," in *4th Advanced Satellite Mobile Systems Conference*, Bologna, Italy, August 26-28 2008.

[10] M. Álvarez-Díaz, M. Neri, C. Mosquera, and G. E. Corazza, "Joint precoding and predistortion techniques for satellite telecommunication systems," in *Proc. of International Workshop on Satellite and Space Communications 2005 (IWSSC'05)*, Siena, Italy, September 8-9 2005, pp. 688–692.

[11] ——, "Trellis shaping techniques for satellite telecommunication systems," in *Proc. of International Workshop on Satellite and Space Communications 2006 (IWSSC'06)*, Leganés, Spain, September 14-15 2006, pp. 148–152.

[12] M. Álvarez-Díaz, S. Scalise, G. Sciascia, R. Mura, P. Conforto, and H. Ernst, "DVB-S air interface over railroad satellite channel: Performance and extensions," in *Proc. of Sixth Baiona Workshop on Signal Processing in Communications*, Baiona, Spain, September 8-10 2003.

[13] P. Angeletti and N. Alagha, "Space/ground beamforming techniques for emerging hybrid satellite terrestrial networks," in *Proc. 27th AAIA International Communications Satellite Systems Conference*, Edinburgh, Scotland, UK, June 1–4 2009.

[14] P. Angeletti, G. Gallinaro, M. Lisi, and A. Vernucci, "On-ground digital beamforming techniques for satellite smart antennas," in *Proc. 19th AIAA International Communications Satellite Systems Conference*, Toulouse, France, April 17–20 2001.

[15] P.-D. Arapoglou, *Deliverable D1: Preliminary Selection of Most Promising MIMO Techniques Applicable to Study Case 1*. ESA Contract No. 21591/08/NL/AT, Wireless & Satellite Communications Group, National Technical University of Athens, 2009.

[16] R. E. Badra and B. Daneshrad, "Asymmetric physical layer design for high-speed wireless digital communications," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1712–1724, Oct. 1999.

[17] K. Balachandran, S. R. Kadaba, and S. Nanda, "Channel quality estimation and rate adaptation for cellular mobile radio," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 7, pp. 1244–1256, July 1999.

[18] F. Bellili, A. Stephenne, and S. Affes, "SNR estimation of QAM-modulated transmissions over time-varying SIMO channels," in *Proc. IEEE International Symposium on Wireless Communication Systems 2008 (ISWCS '08)*, Reykjavik, Iceland, 21–24 Oct 2008, pp. 199–203.

[19] ——, "Cramér-Rao bound for NDA SNR estimates of square QAM modulated signals," in *Proc. IEEE Wireless Communications and Networking Conference WCNC 2009*, 5–8 April 2009, pp. 1–5.

[20] ——, "Cramér-Rao lower bounds for subcarrier SNR estimates over multicarrier channels," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 891–896, Feb. 2010.

[21] S. Benedetto and E. Biglieri, "Nonlinear equalization of digital satellite channels," *IEEE J. Sel. Areas Commun.*, vol. 1, no. 1, pp. 57–62, Jan 1983.

[22] S. Benedetto, E. Biglieri, and R. Daffara, "Modeling and performance evaluation of nonlinear satellite links-A Volterra series approach," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-15, no. 4, pp. 494 –507, july 1979.

[23] T. Benedict and T. Soong, "The joint estimation of signal and noise from the sum envelope," *IEEE Trans. Inf. Theory*, vol. 13, no. 3, pp. 447–454, Jul 1967.

[24] T. Benjamin, M. Griffin, and W. W. Chapman, "Methods and systems for satellite communications employing ground-based beam forming with spatially distributed hybrid matrix amplifiers," U.S. Patent 2007/9 281 612 A1, December 6, 2007, patent application.

[25] D. Brandel, W. Watson, and A. Weinberg, "NASA's advanced tracking and data relay satellite system for the years 2000 and beyond," *Proc. IEEE*, vol. 78, no. 7, pp. 1141 –1151, jul 1990.

[26] C. Caini and G. Corazza, "Satellite downlink reception through intermediate module repeaters: power delay profile analysis," in *Proc. EMPS/ASMS Conference*, ESA/ESTEC, Noordwijk, The Netherlands, September 21-22 2004.

[27] C. Caini, G. E. Corazza, G. Falciasecca, M. Ruggieri, and F. Vatalaro, "A spectrum- and power-efficient EHF mobile satellite system to be integrated with terrestrial cellular systems." *IEEE J. Sel. Areas Commun.*, vol. 10, no. 8, pp. 1315–1325, 1992.

[28] P. Campisi, G. Panci, S. Colonnese, and G. Scarano, "Blind phase recovery for QAM communication systems," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1348–1358, 2005.

[29] K. V. Cartwright, "Blind phase recovery in cross QAM communication systems with eighth-order statistics," *IEEE Signal Process. Lett.*, vol. 8, no. 12, pp. 304–306, Dec. 2001.

[30] K. V. Cartwright and E. J. Kaminsky, "Asymptotic performance of the Pth power-law phase estimator," in *Proc. IEEE Global Telecommunications Conference GLOBECOM '05*, vol. 1, 28 Nov.–2 Dec. 2005, p. 6pp.

[31] ——, "A simple improvement to the Viterbi and Viterbi monomial-based phase estimators," in *Proc. IEEE Global Telecommunications Conference GLOBECOM '06*, Nov. 27 2006–Dec. 1 2006, pp. 1–6.

[32] L. Castanet, A. Bolea-Alamanac, and M. Bousquet, "Interference and fade mitigation techniques for Ka and Q/V band satellite communication systems," in *Proc. of the Joint Cost 272/280 Workshop*. Noordwijk, the Netherlands: ESA/ESTEC, May 26-28 2003.

[33] P. M. Castro, M. Joham, L. Castedo, and W. Utschick, "Robust MMSE linear precoding for multiuser MISO systems with limited feedback and channel prediction,"

in *Proc. IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Cannes, France, September 2008.

[34] P. M. Castro, M. Joham, W. Utschick, and L. Castedo, "Optimized CSI feedback for robust THP design," in *Proc. 41st Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, California, USA, October-November 2007.

[35] H.-W. Chen, "Modeling and identification of parallel nonlinear systems: structural classification and parameter estimation methods," *Proc. IEEE*, vol. 83, no. 1, pp. 39 –66, jan 1995.

[36] Y.-R. Chien, W.-L. Mao, and H.-W. Tsao, "Design of a robust multi-channel timing recovery system with imperfect channel state information for 10GBASE-T," *IEEE Trans. Circuits Syst. I*, no. 99, 2009, to be published.

[37] Y.-R. Chien, H.-W. Tsao, and W.-L. Mao, "Adaptive two-stage equalisation and FEXT cancellation architecture for 10GBASE-T system," *IET Communications*, vol. 3, no. 7, pp. 1124 –1135, july 2009.

[38] Y.-R. Chien, W.-L. Mao, and H.-W. Tsao, "A novel baud-rate timing error detector design for baseband transmission system using Tomlinson-Harashima precoder," *IEEE Signal Process. Lett.*, vol. 15, pp. 673 –676, 2008.

[39] S. Cioni, G. E. Corazza, M. Neri, and A. Vanelli-Coralli, "On the use of OFDM radio interface for satellite digital multimedia broadcasting systems," *Int. J. Satellite Communications and Networking*, vol. 24, no. 2, pp. 153–167, 2006.

[40] S. Cioni, G. E. Corazza, and M. Bousquet, "An analytical characterization of maximum likelihood signal-to-noise ratio estimation," in *Wireless Communication Systems, 2005. 2nd International Symposium on*, Siena, Italy, 5–7 Sep. 2005, pp. 827–830.

[41] S. Cioni, C. Párraga Niebla, G. Seco Granados, S. Scalise, A. Vanelli-Coralli, , and M. A. Vázquez Castro, "Advanced fade countermeasures for DVB-S2 systems in railway scenarios," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, 2007.

[42] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[43] L. Cottatellucci, M. Debbah, G. Gallinaro, R. Mueller, M. Neri, and R. Rinaldo, "Interference mitigation techniques for broadband satellite systems," in *Proc. 24th AIAA International Communications Satellite Systems Conference (ICSSC) and 4th Annual International Satellite & Communications (ISC) Conference and Expo*, San Diego (CA), USA, June 11-14 2006.

[44] A. Couchman, A. Dawkins, A. Russell, S. Stirland, F. Coromina, and X. Maufroid, "Defocused array fed reflector antennas for Ka, broad band satellites," in *Proc.*

28th ESA Antenna Workshop, ESTEC, Noordwijk, The Netherlands, 31 May–2 June 2005.

[45] W. Dai, Y. Wang, and J. Wang, "Joint power estimation and modulation classification using second- and higher statistics," in *Proc. WCNC2002 Wireless Communications and Networking Conference 2002 IEEE*, vol. 1, 17–21 March 2002, pp. 155–158.

[46] M. A. Dangl and J. Lindner, "How to use a priori information of data symbols for SNR estimation," *IEEE Signal Process. Lett.*, vol. 13, no. 11, pp. 661–664, Nov. 2006.

[47] A. Das, "NDA SNR estimation: CRLBs and EM based estimators," in *Proc. IEEE Region 10 Conference TENCON 2008*, Hyderabad, India, 18–21 Nov. 2008, pp. 1–6.

[48] R. De Gaudenzi and F. Giannetti, "Analysis of an advanced satellite digital audio broadcasting system and complementary terrestrial gap-filler single frequency network," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 194–210, May 1994.

[49] F. A. Dietrich, P. Breun, and W. Utschick, "Robust Tomlinson-Harashima precoding for the wireless broadcast channel," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 631 –644, feb. 2007.

[50] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Jan 2004.

[51] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *Communications, IET*, vol. 1, no. 2, pp. 137–156, April 2007.

[52] S. Dolinar, "Exact closed-form expressions for the performance of the split-symbol moments estimator of signal-to-noise ratio," Telecommunications and Data Acquisition Prog. Rep. 42-100, Pasadena, CA., 1990.

[53] S. Egami, "A power-sharing multiple-beam mobile satellite in Ka band," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 2, pp. 145 –152, feb 1999.

[54] EN 300 307 V1.1.2, *Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications*. ETSI, June 2006.

[55] EN 302 583 V1.1.1, *Digital Video Broadcasting (DVB); Framing Structure, channel coding and modulation for Satellite Services to Handheld devices (SH) below 3 GHz*. ETSI, March 2008.

[56] M. V. Eyuboğlu, "Detection of coded modulation signals on linear, severely distorted channels using decision-feedback noise prediction with interleaving," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 401–409, April 1988.

[57] M. V. Eyuboğlu and J. Forney, G. D., "Trellis precoding: combined coding, precoding and shaping for intersymbol interference channels," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 301–314, March 1992.

[58] Y. Feria, "A complex symbol signal-to-noise ratio estimator and its performance," Telecommunications and Data Acquisition Prog. Rep. 42-116, Pasadena, CA., 1994.

[59] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York, NY: John Wiley & Sons, 2002.

[60] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "MIMO precoding for decentralized receivers," in *Proc. ISIT 2002*, Lausanne, Switzerland, June–July 2002, p. 496.

[61] M. Forest, S. Richard, and C. A. McDonach, "ACeS antenna feed arrays," in *Proc. of the 5th International Mobile Satellite Conference (IMSC'97)*, Pasadena (CA), USA, June 1997, pp. 387–391.

[62] J. Forney, G., R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 5, pp. 632–647, Sep 1984.

[63] J. Forney, G. D., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, March 1973.

[64] ——, "Trellis shaping," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 281–300, March 1992.

[65] J. Forney, G. D. and L. F. Wei, "Multidimensional constellations– I. Introduction, figures of merit, and generalized cross constellations," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 877–892, Aug. 1989.

[66] R. M. Gagliardi and C. M. Thomas, "PCM data reliability monitoring through estimation of signal-to-noise ratio," *IEEE Trans. Commun.*, vol. 16, no. 3, pp. 479–486, Jun 1968.

[67] G. Gallinaro, G. Caire, M. Debbah, L. Cottatellucci, R. Mueller, and R. Rinaldo, "Perspectives of adopting interference mitigation techniques in the context offset broadband multimedia satellite systems," in *Proc. 23rd AIAA International Communications Satellite Systems Conference (ICSSC 2005)*, Rome, Italy, 25–28 September 2005.

[68] P. Gao and C. Tepedelenlioğlu, "SNR estimation for nonconstant modulus constellations," *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 865–870, March 2005.

[69] W. Gappmair, "Cramer-Rao lower bound for non-data-aided SNR estimation of linear modulation schemes," *IEEE Trans. Commun.*, vol. 56, no. 5, pp. 689–693, May 2008.

[70] W. Gappmair, M. Flohberger, and O. Koudelka, "Moment-based estimation of the signal-to-noise ratio for oversampled narrowband signals," in *Proc. 16th IST Mobile and Wireless Communications Summit*, Budapest, Hungary, 1–5 July 2007, pp. 1–4.

[71] W. Gappmair and O. Koudelka, "Moment-based SNR estimation of signals with non-constant envelope," in *Proc. 3rd Advanced Satellite Mobile Systems Conf.*, Herrsching, Germany, May 2006, pp. 301–303.

[72] W. Gappmair, R. López-Valcarce, and C. Mosquera, "Cramer-Rao lower bound and EM algorithm for envelope-based SNR estimation of nonconstant modulus constellations," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1622–1627, June 2009.

[73] W. Gappmair, R. López-Valcarce, and C. Mosquera, "Joint NDA estimation of carrier frequency/phase and SNR for linearly modulated signals," *IEEE Signal Process. Lett.*, 2010, accepted for publication.

[74] F. M. Gardner, "A BPSK/QPSK timing-error detector for sampled receivers," *IEEE Trans. Commun.*, vol. 34, pp. 399–406, May 1986.

[75] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal Processing*, vol. 86, no. 4, pp. 639 – 697, 2006.

[76] C. N. Georghiades, "Blind carrier phase acquisition for QAM constellations," *IEEE Trans. Commun.*, vol. 45, no. 11, pp. 1477–1486, Nov. 1997.

[77] M. R. Gibbard and A. B. Sesay, "Asymmetric signal processing for indoor wireless lans," *IEEE Trans. Veh. Technol.*, vol. 48, no. 6, pp. 2053–2064, Nov. 1999.

[78] R. F. E. Guy, C. B. Wyllie, and J. R. Brain, "Synthesis of the Inmarsat 4 multibeam mobile antenna," in *Proc. of 12th International Conference on Antennas and Propagation*, vol. 1, Exeter, UK, April 2003, pp. 90–93.

[79] C. Haardt and N. Courville, "Internet switching by satellite: an ultra fast processor with radio burst switching," in *Proc. Disruption in Space, First Symposium on Potentially Disruptive Technologies and their Impact in Space Programs*, Marseille, France, 4-6 July 2005.

[80] H. Harashima and H. Miyakawa, "A method of code conversion for digital communications channels with intersymbol interference," *Transactions of the Institute of Electronics and Communications Engineers of Japan*, vol. 52-A, pp. 272–273, 1969, (In Japanese.).

[81] ——, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, pp. 774–780, 1972.

[82] G. Hatzichristos and M. P. Fargues, "A hierarchical approach to the classification of digital modulation types in multipath environments," in *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, vol. 2, 4–7 Nov. 2001, pp. 1494–1498.

[83] A. Heidari and M. Nasiri-Kenari, "A family of simple blind phase recovery algorithms," in *Proc. Int. Symp. on Intelligent Signal Processing and Communications (ISPACS '00)*, Honolulu, Hawaii, USA, Nov. 2000, p. 4 pages. [Online]. Available: www.cst.uwaterloo.ca/~reza/doc/ispacs00.pdf

[84] S. Hirakawa, N. Sato, and H. Kikuchi, "Broadcasting satellite services for mobile reception," *Proc. IEEE*, vol. 94, no. 1, pp. 327–332, Jan. 2006.

[85] Y. H. Hu, "CORDIC-based VLSI architectures for digital signal processing," *IEEE Signal Process. Mag.*, vol. 9, no. 3, pp. 16–35, Jul. 1992.

[86] C.-Y. Huang and A. Polydoros, "Likelihood methods for MPSK modulation classification," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 1493–1504, Feb–Mar–Apr 1995.

[87] M. Huang, S. Zhou, and J. Wang, "Analysis of Tomlinson-Harashima precoding in multiuser MIMO systems with imperfect channel state information," *IEEE Trans. Veh. Technol.*, vol. 57, no. 5, pp. 2856 –2867, sept. 2008.

[88] Hughes Network Systems, LLC, "Hughes to launch 100 Gbps high throughput satellite in 2012," Press release, June 16, 2009. [Online]. Available: http://www.hugheseurope.com/files/press/HTS%20Satellite%20Final(5)%2006-16-09.pdf

[89] ICO Global Communications, "ICO G1 space segment testing complete," Press release, January 7, 2009. [Online]. Available: http://investor.ico.com/releasedetail.cfm?ReleaseID=357858

[90] *IEEE Std 802.3an-2006 (Amendment to IEEE Std 802.3-2005)*, IEEE Computer Society Std., September 2006.

[91] *IEEE Std 802.16-2009, Air Interface for Broadband Wireless Access Systems*, IEEE Computer Society and the IEEE Microwave Theory and Techniques Society Std., 29 May 2009.

[92] IEEE P802.3ba 40Gb/s and 100Gb/s Ethernet Task Force. (2010) Available online at http://www.ieee802.org/3/ba/public/index.html. [Online]. Available: http://www.ieee802.org/3/ba/public/index.html

[93] N. Jablon, "Joint blind equalization, carrier recovery and timing recovery for high-order QAM signal constellations," *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1383 –1398, jun 1992.

[94] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1783–1794, may 2005.

[95] M. Joham, P. M. Castro, J. Zhen, L. Castedo, and W. Utschick, "Joint design of limited feedback and multiuser precoding based on a precoding MSE metric," in *Proc. 42nd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, USA, November 2008.

[96] M. Joham, D. Schmidt, J. Brehmer, and W. Utschick, "Finite-length MMSE Tomlinson–Harashima precoding for frequency selective vector channels," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3073 –3088, june 2007.

[97] M. Joham, "Optimization of linear and nonlinear transmit signal processing," Ph.D. dissertation, Technical University Munich, 2004.

[98] M. Joham, J. Brehmer, and W. Utschick, "MMSE approaches to multiuser spatio-temporal Tomlinson-Harashima precoding," in *Proc. ITG SCC'04*, January 2004, pp. 387–394.

[99] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, August 2005.

[100] M. Karaliopoulos, K. Narenthiran, B. Evans, P. Henrio, M. Mazzella, W. de Win, M. Dieudonne, P. Philippopoulos, D. Axiotis, I. Andrikopoulos, I. Mertzanis, G. Corazza, A. Vanelli-Coralli, N. Dimitriou, and A. Polydoros, "Satellite radio interface and radio resource management strategy for the delivery of multicast/broadcast services via an integrated satellite-terrestrial system," *IEEE Commun. Mag.*, vol. 42, no. 9, pp. 108 – 117, sept. 2004.

[101] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[102] R. B. Kerr, "On signal and noise level estimation in a coherent PCM channel," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-2, no. 4, pp. 450–454, 1966.

[103] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electronics Letters*, vol. 37, no. 23, pp. 1417–1418, 8 Nov 2001.

[104] R. Laroia, N. Farvardin, and S. A. Tretter, "On optimal shaping of multidimensional constellations," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1044–1056, July 1994.

[105] S.-J. Lee, S. Lee, K.-W. Kim, and J.-S. Seo, "Personal and mobile satellite DMB services in Korea," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 179–187, March 2007.

[106] S. J. Lee, "A new non-data-aided feedforward symbol timing estimator using two samples per symbol," *IEEE Commun. Lett.*, vol. 6, no. 5, pp. 205 –207, may 2002.

[107] A. Liavas, "Tomlinson-Harashima precoding with partial channel knowledge," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 5 – 9, jan. 2005.

[108] J. Litva and T. Kwok-Yeung Lo, *Digital Beamforming in Wireless Communications*. Norwood, MA, USA: Artech House, 1996.

[109] G. Liva, N. Riera Díaz, S. Scalise, B. Matuz, C. Párraga Niebla, J.-G. Ryu, M.-S. Shin, and H.-J. Lee, "Gap filler architectures for seamless DVB-S2/RCS provision in the railway environment," in *Proc. IEEE Vehicular Technology Conference VTC Spring 2008*, 11–14 May 2008, pp. 2996–3000.

[110] R. López-Valcarce and C. Mosquera, "Sixth-order statistics-based non-data-aided SNR estimation," *IEEE Commun. Lett.*, vol. 11, no. 4, pp. 351–353, April 2007.

[111] R. López-Valcarce, C. Mosquera, and W. Gappmair, "Iterative envelope-based SNR estimation for nonconstant modulus constellations," in *Proc. IEEE 8th Workshop on Signal Processing Advances in Wireless Communications SPAWC 2007*, Helsinki, Finland, 17–20 June 2007, pp. 1–5.

[112] R. López-Valcarce and C. Mosquera, "Blind carrier phase estimation for non-equiprobable constellations," in *Proc. IEEE 7th Workshop on Signal Processing Advances for Wireless Communications Workshop (SPAWC'06)*, Cannes, France, 2–5 July 2006.

[113] E. Lutz, M. Werner, and A. Jahn, *Satellite Systems for Personal Broadband Communications*. Berlin: Springer, 2000.

[114] K. Maharatna, A. Troya, S. Banerjee, and E. Grass, "Virtually scaling-free adaptive CORDIC rotator," *IEE Proceedings -Computers and Digital Techniques*, vol. 151, no. 6, pp. 448–456, 18 Nov. 2004.

[115] A. Mallet, A. Anakabe, J. Sombrin, and R. Rodriguez, "Multiport-amplifier-based architecture versus classical architecture for space telecommunication payloads," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 12, pp. 4353 –4361, dec. 2006.

[116] G. Maral and M. Bousquet, *Satellite communications systems: systems, techniques, and technology*, 4th ed. Chichester, UK: John Wiley & Sons, 2002.

[117] M. Marinelli and R. Giubilei, "A regenerative payload for satellite multimedia communications," *IEEE Multimedia*, vol. 6, no. 4, pp. 30 –37, oct-dec 1999.

[118] R. Matzner, "An SNR estimation algorithm for complex baseband signals using higher order statistics," *Facta Universitatis (Niš)*, vol. 6, no. 1, pp. 41–52, 1993.

[119] M. Mesiya, P. McLane, and L. Campbell, "Maximum likelihood sequence estimation of binary sequences transmitted over bandlimited nonlinear channels," *IEEE Trans. Commun.*, vol. 25, no. 7, pp. 633–643, Jul 1977.

[120] J. P. Millerioux, M. L. Boucheret, C. Bazile, and A. Ducasse, "Iterative interference cancellation and channel estimation in multibeam satellite systems," *Int. J. Satellite Communications and Networking*, vol. 25, no. 3, pp. 263–283, 2007.

[121] M. Moeneclaey and G. de Jonghe, "Tracking performance comparison of two feedforward ML-oriented carrier-independent NDA symbol synchronizers," *IEEE Trans. Commun.*, vol. 40, no. 9, pp. 1423 –1425, sep 1992.

[122] ——, "ML-oriented NDA carrier synchronization for general rotationally symmetric signal constellations," *IEEE Trans. Commun.*, vol. 42, no. 8, pp. 2531–2533, Aug. 1994.

[123] M. L. Moher, "Multiuser decoding for multibeam systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1226–1234, July 2000.

[124] M. Morelli, A. D'Andrea, and U. Mengali, "Feedforward ML-based timing estimation with PSK signals," *IEEE Commun. Lett.*, vol. 1, no. 3, pp. 80 –82, may 1997.

[125] ——, "Correction to "Feedforward ML-based timing estimation with PSK signals"," *IEEE Commun. Lett.*, vol. 2, no. 1, pp. 24 –24, jan 1998.

[126] K. Mueller and M. Müller, "Timing recovery in digital synchronous data receivers," *IEEE Trans. Commun.*, vol. 24, no. 5, pp. 516 – 531, may 1976.

[127] A. Narula, M. Lopez, M. Trott, and G. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1423 –1436, oct 1998.

[128] M. Oerder and H. Meyr, "Digital filter and square timing recovery," *IEEE Trans. Commun.*, vol. 36, no. 5, pp. 605–612, May 1988.

[129] G. Panci, S. Colonnese, S. Rinauro, and G. Scarano, "Gain-control-free near-efficient phase acquisition for QAM constellations," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2849–2864, 2008.

[130] G. Pasolini and A. Giorgetti, "DVB-S gap fillers for railway tunnels," in *Proc. VTC-2006 Fall Vehicular Technology Conference 2006 IEEE 64th*, 25–28 Sept. 2006, pp. 1–6.

[131] D. Pauluzzi and N. Beaulieu, "A comparison of SNR estimation techniques for the AWGN channel," *IEEE Trans. Commun.*, vol. 48, no. 10, pp. 1681–1691, Oct. 2000.

[132] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

[133] C. Rapp, "Effects of HPA-nonlinearity on 4-DPSK/OFDM-signal for a digital sound broadcasting system," in *Proc. 2nd European Conference on Satellite Communications*, vol. 2, Liege, Belgium, Oct. 1991, pp. 179–184.

[134] V. S. Reinhardt, "Hybridized space/ground beamforming," U.S. Patent 6 571 081 B1, May 27, 2003.

[135] G. Ren, Y. Chang, and H. Zhang, "A new SNR's estimator for QPSK modulations in an AWGN channel," *IEEE Trans. Circuits Syst. II*, vol. 52, no. 6, pp. 336–338, June 2005.

[136] J. Riba, J. Villares, and G. Vázquez, "A non-data-aided SNR estimation technique for multilevel modulations exploiting signal cyclostationarity," *IEEE Trans. Signal Process.*, under review, 2009. [Online]. Available: http://gps-tsc.upc.es/comm/gregori/papers/SNR_DRAFT.pdf

[137] J. J. Rivera, E. Trachtman, and M. Richharia, "The BGAN extension programme," *ESA Bulletin*, vol. 124, pp. 62–68, Nov. 2005.

[138] A. A. Rontogiannis, M. Álvarez-Díaz, M. Casadei, V. Dalakas, A. Duverdier, F. J. González Serrano, M. Iubatti, T. Javornik, L. Lapierre, M. Neri, and P. Salmi, "Distortion countermeasures," in *Digital Satellite Communications*, G. E. Corazza, Ed. New York: Springer, April 2007, ch. 8.

[139] A. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Commun.*, vol. 29, no. 11, pp. 1715–1720, Nov 1981.

[140] L. Sanguinetti and M. Morelli, "Non-linear pre-coding for multiple antenna multi-user downlink transmissions with different QoS requirements," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 852–856, March 2007.

[141] G. Sciascia, S. Scalise, H. Ernst, and R. Mura, "Statistical characterization of the railroad satellite channel at Ku-band," in *Proc. of the International Workshop of Cost Actions 272 and 280*, Noordwijk, The Netherlands, 26–29 May 2003.

[142] E. Serpedin, P. Ciblat, G. B. Giannakis, and P. Loubaton, "Performance analysis of blind carrier phase estimators for general QAM constellations," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1816–1823, Aug. 2001.

[143] J. F. Sevillano, I. Vélez, M. Leyh, S. Lipp, A. Irizar, and L. Fontán, "In-service SNR estimation without symbol timing recovery for QPSK data transmission systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 9, pp. 3202–3207, 2007.

[144] B. Shah and S. Hinedi, "The split symbol moments SNR estimator in narrow-band channels," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 5, pp. 737–747, 1990.

[145] K. Shi, Y. Wang, and E. Serpedin, "On the design of a digital blind feedforward, nearly jitter-free timing-recovery scheme for linear modulations," *IEEE Trans. Commun.*, vol. 52, no. 9, pp. 1464 – 1469, sept. 2004.

[146] (2009) Signal Processing in Communications Group, University of Vigo. Vigo, Spain. *Mathematica* and *MATLAB* code supporting some of the derivations in Section 4.3. Available online at http://www.gts.tsc.uvigo.es/˜valcarce/snr_hos.html. [Online]. Available: http://www.gts.tsc.uvigo.es/~valcarce/snr_hos.html

[147] M. Simon and S. Dolinar, "Signal-to-noise ratio estimation for autonomous receiver operation," in *Proc. IEEE Global Telecommunications Conference GLOBECOM '04*, vol. 1, Dallas, TX, USA, 29 Nov.–3 Dec. 2004, pp. 282–287.

[148] ——, "Improving SNR estimation for autonomous receivers," *IEEE Trans. Commun.*, vol. 53, no. 6, pp. 1063–1073, June 2005.

[149] M. Simon and A. Mileant, "SNR estimation for the baseband assembly," Telecommunications and Data Acquisition Prog. Rep. 42-85, Pasadena, CA., 1986.

[150] J. Smee and S. Schwartz, "Adaptive compensation techniques for communications systems with Tomlinson-Harashima precoding," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 865 – 869, june 2003.

[151] C. Stallo, M. Lucente, T. Rossi, E. Cianca, M. Ruggieri, A. Paraboni, C. Cornacchini, A. Vernucci, M. Nocerino, A. Ceccarelli, L. Bruca, G. Codispoti, and M. De Sanctis, "TRANSPONDERS: Research and analysis for the development of telecommunication payloads in Q/V bands," in *Proc. 2009 IEEE Aerospace conference*, March 2009, pp. 1 –11.

[152] A. Stephenne, F. Bellili, and S. Affes, "Moment-based SNR estimation for SIMO wireless communication systems using arbitrary QAM," in *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers ACSSC 2007*, Pacific Grove, CA, USA, 4–7 Nov. 2007, pp. 601–605.

[153] T. A. Summers and S. G. Wilson, "SNR mismatch and online estimation in turbo decoding," *IEEE Trans. Commun.*, vol. 46, no. 4, pp. 421–423, April 1998.

[154] M. Tanaka and S. Egami, "Reconfigurable multiport amplifiers for in-orbit use," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 228 – 236, jan. 2006.

[155] TerreStar Networks, "FCC grants TerreStar ATC authority," Press release, January 14, 2010. [Online]. Available: http://www.terrestar.com/press/20100114.html

[156] ——, "TerreStar announces ground based beam forming testing success," Press release, February 22, 2010. [Online]. Available: http://www.terrestar.com/press/20100222.html

[157] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electronic Letters*, vol. 7, pp. 138–139, 1971.

[158] T. Trump and U. Forssén, "On the statistical properties of Tomlinson filters," in *IEEE Winter Workshop on Nonlinear Digital Signal Processing*, Tampere, Finland, Jan. 17–20 1993, pp. P–10.1 –P–10.4.

[159] M. Türkboylari and G. Stüber, "An efficient algorithm for estimating the signal-to-interference ratio in TDMA cellular systems," *IEEE Trans. Commun.*, vol. 46, no. 6, pp. 728–731, June 1998.

[160] A. Vanelli-Coralli, G. E. Corazza, G. K. Karagiannidis, P. T. Mathiopoulos, D. S. Michalopoulos, C. Mosquera, S. Papaharalabos, and S. Scalise, "Satellite communications: Research trends and open issues," in *Proc. International Workshop on Satellite and Space Communications IWSSC '07*, 13–14 Sept. 2007, pp. 71–75.

[161] A. J. Viterbi and A. M. Viterbi, "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission," *IEEE Trans. Inf. Theory*, vol. IT-29, pp. 543–551, Jul. 1983.

[162] J. L. Walker, R. Menendex, D. Burr, and G. Dubellay, "Ground-based beamforming for satellite communication systems," U.S. Patent 2008/0 051 080 A1, February 28, 2008, patent application.

[163] Y. Wang and E. Serpedin, "A class of blind phase recovery techniques for higher order QAM modulations: Estimators and bounds," *IEEE Signal Process. Lett.*, vol. 9, no. 10, pp. 301–304, Oct. 2002.

[164] Y. Wang, E. Serpedin, and P. Ciblat, "Optimal blind nonlinear least-squares carrier phase and frequency offset estimation for general QAM modulations," *IEEE Trans. Wireless Commun.*, vol. 2, no. 5, pp. 1040–1054, 2003.

[165] ——, "An alternative blind feedforward symbol timing estimator using two samples per symbol," *IEEE Trans. Commun.*, vol. 51, no. 9, pp. 1451 – 1455, sept. 2003.

[166] A. Wiesel, J. Goldberg, and H. Messer, "Non-data-aided signal-to-noise-ratio estimation," in *Proc. IEEE International Conference on Communications ICC 2002*, vol. 1, New York City, NY, USA, 28 April–2 May 2002, pp. 197–201.

[167] A. Wiesel, J. Goldberg, and H. Messer-Yaron, "SNR estimation in time-varying fading channels," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 841–848, May 2006.